Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Decision making study: methods and applications of evidential reasoning and judgment analysis

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Yixing Shan

PUBLISHER STATEMENT

This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Shan, Yixing. 2019. “Decision Making Study: Methods and Applications of Evidential Reasoning and Judgment Analysis”. figshare. https://hdl.handle.net/2134/17330.

DECISION MAKING STUDY: METHODS AND APPLICATIONS OF EVIDENTIAL REASONING AND JUDGMENT ANALYSIS

by

Yixing Shan

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of Loughborough University

March 2015

© Yixing Shan, 2015

Table of Contents

Table of Contents 2

List of Figure 6

List of Tables 8

List of Abbreviations 9

Acknowledgements 11

Abstract 12

Chapter 1 : Introduction 14

1.1 Background 14

1.2 Research Questions & Objectives 17

1.3 Research methodology 21

1.4 Envisaged contributions 23

1.5 Thesis outline 24

Chapter 2 : Literature review 26

2.1 Multiple Criteria Decision Making (MCDM) 26

2.2 Uncertainty Management & Evidential Reasoning (ER) approach 28

2.2.1 The Dempster-Shafer theory of evidence 31

2.2.2 The ER algorithm 36

2.2.3 Utility based ranking for belief structure 39

2.2.4 Rule and utility based transformation technique for evaluation grades 40

2.3 Naturalistic Decision Making (NDM) 44

2.3.1 Characteristics of NDM 47

2.4 Social Judgment Theory (SJT) 50

2.4.1 Brunswik's probabilistic functionalism 51

2.4.2 Social Judgment Theory: Hammond's further development of Probabilistic Functionalism 54
2.4.3. Judgment Analysis: a technique used by SJT researchers 56
2.5 Heuristic decision making: Adaptive Toolbox (AT) program 60
2.5.1 Origin 60
2.5.2 Main topics 61
2.5.3 Characteristics of AT 67
2.6 Conclusion 71
Chapter 3 : Research Methodology 73
3.1 Evidence-based Trade-Off (EBTO) 73
3.1.1 Model construction 75
3.1.2 Data collection 77
3.1.3 Assessment normalization 79
3.1.4 Assessment aggregation & choice of decision alternative 80
3.2 Judgment Analysis with Heuristic Modelling 82
3.2.1 Phase 1: Formulating the decision problem 84
3.2.2 Phase 2: Formulating the decision profile 87
3.2.3 Phase 3: Constructing the decision making task 92
3.2.4 Phase 4: Delivering the decision making task to participants 96
3.2.5 Phase 5: Data analysis 97
3.3 Summary 105
Chapter 4 : Case study of Soldier System decision making 108
4.1 Soldier system decision making 108
4.2 The assessment model for a Soldier system 111

4.3 Data collection 113
4.3.1 Importance weighting of criteria 113
4.3.2 Assessment at the bottom level criteria 116
4.4 Assessment aggregation & Choice of Soldier system 118
4.5 Summary 122
Chapter 5 : Case study of fire emergency decision making 124
5.1 The problem 124
5.2 Background 126
5.2.1 Incident Command System (ICS) in UK FRS 126
5.2.2 Rank structure of FRS in UK 128
5.3 Formulation of decision profile 128
5.4 Construction of decision making task 136
5.5 Participants and Data collection process 137
5.6 Analysis and results 139
5.6.1 Raw data and data alignment 139
5.6.2 Statistical characteristics of the data 140
5.6.3 ICs' decision making policies 142
5.7 Summary 160
Chapter 6 : Discussions, Conclusions and Future Research 163
6.1 Summary and discussion of research findings 163
6.2 Research contributions 173
6.3 Limitations and future research 175
References: 178
Appendix A: Field context of high rise apartment fire 211

Appendix B: Interview guide 228

Appendix C: Hypothetical cases in a decision making task 230

Appendix D: Data collected from the questionnaires 233

Appendix E: A sample questionnaire 237

Appendix F: List of participants of interviews and questionnaires 253

Appendix G: Matlab scripts used for data analysis 255

1. Main function 256

2. Function of calculating cue validities and threshold values 258

3. Predicting data on derivation sample with Matching Heuristic 260

4. Predicting data on cross-validation sample with Matching Heuristic 261

5. Predicting data on derivation sample with Franklin's rule 262

6. Predicting data on cross-validation sample with Franklin's rule 263

7. Predicting data on derivation sample with Dawes' rule 264

8. Predicting data on cross-validation sample with Dawes' rule 265

9. Predicting 3-point ordinal data with MH-3R 266

10. Predicting 4-point ordinal data with MH-4R 268

11. Predicting 3-point ordinal data with WADD-3R 270

12. Predicting 4-point ordinal data with WADD-4R 272

List of Figure

Figure 2-1, Recognition-Primed Decision (RPD) model (Klein, 1993) 47
Figure 2-2, Adapted Lens Model. From "The conceptual framework of psychology" (p.20), E.Brunswik, 1952, Chicago, IL: University of Chicago Press. Copyright 1952 by University of Chicago Press. 53
Figure 2-3, Flow diagram of take-the-best heuristic (Gigerenzer & Goldstein, 1996) 65
Figure 3-1, The framework of EBTO method 75
Figure 3-2, Work flow of JA-HM: A variant of JA method for decision making task under extreme time pressure and dynamic shifting situations 84
Figure 3-3, Schematic illustration of how two models fit existing data and how they predict new data. Model A overfits the existing data and is not as accurate as Model B in predicting new data (Pitt, Myung, & Zhang, 2002). 104
Figure 4-1, Soldier system hierarchical model 112
Figure 4-2, Overall assessment of Soldier systems for patrol mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2) 119
Figure 4-3, Overall assessment of Soldier systems for assault mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2) 119
Figure 4-4, Utility interval of Soldier systems for patrol mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2) 121
Figure 4-5, Utility interval of Soldier systems for assault mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2) 121
Figure 5-1, An example of decision profile 135
Figure 5-2, Comparison between mean subjective weights of cues assigned by WMs and CMs 142
Figure 5-3, Calculation of the cue validity: supposing a binary cue (0-1) and a dataset containing binary decision data of 32 cases in total 144
Figure 5-4, An example of defining the threshold value of the Franklin's rule for a dataset with 32 cases 146

Figure 5-5, Calculations of Franklin's rule, Dawes' rule, and MH where K=2, on an example case 147
Figure 5-6, Flowchart of Matching Heuristic (K=2) 149
Figure 5-7, Flowchart of MH-3R (K=2) 152
Figure 5-8, Flowchart of MH-4R (K=2) 153
Figure 5-9, Comparison of cue rank orders applied by WMs 159
Figure 5-10, Comparison of cue rank orders applied by CMs 160
Figure A-1, Sector structure of high rise fire (HM Government, 2008a) 224
Figure A-2, Fire development in a compartment (Hartin, 2009) 226

List of Tables

Table 2-1, Combination of two piece of evidence 33
Table 2-2, The combination of conflict evidence by the D-S combination rule (Wang, Yang, Xu, & Chin, 2006) 35
Table 2-3, The combination of conflict evidence by the ER algorithm (Wang, Yang, Xu, & Chin, 2006). 38
Table 2-4, The combined belief degrees under different sets of relative weights (Wang, Yang, Xu, & Chin, 2006) 39
Table 2-5, Broad categorization of Judgment Analysis research contexts using the dimensions of task familiarity and task congruence (Cooksey, 1996, p.89) 59
Table 4-1, The Soldier systems to be considered 110
Table 4-2, Relative importance of the criteria 114
Table 4-3, Assessments of bottom level criteria 117
Table 5-1, Six varied cues 132
Table 5-2, Thirteen constant cues 133
Table 5-3, Mean percentage of binary decision data predicted correctly by the models 156
Table 5-4, Mean percentage of 3 & 4-point ordinal decision data predicted correctly by the models 157

List of Abbreviations

AFD	Automatic Fire Detection
ALP	Aerial Ladder Platform
AM	Area Manager
APA	American Psychological Association
AT	Adaptive Toolbox
BA	Breathing Apparatus
BAECO	Breathing Apparatus Entry Control Officer
BM	Brigade Manager
C41	Command, Control, Communications, Computers and Intelligence
CC	Crew Commander
CDM	Critical Decision Method
CM	Crew Manager
CS	Command Support
DEBA	Deterministic Elimination-by-Aspects
DM	Decision Maker
DRA	Dynamic Risk Assessment
D-S	Dempster-Shafer
EBA	Elimination-by-Aspects
EBTO	Evidence-based Trade-off
ER	Evidential Reasoning
FRS	Fire and Rescue Service
G	Good
GM	Group Manager
HB	Heuristics & Biases
HFES	Human Factors and Ergonomics Society
IC	Incident Commander
ICS	Incident Command System
IDS	Intelligent Decision System
JA	Judgment Analysis
JA-HM	Judgment Analysis with Heuristic Modelling
LFRS	Leicestershire Fire and Rescue Service
MCDM	Multiple Criteria Decision Making
MDT	Mobile Data Terminal
MH	Matching Heuristic
MH-3R	Matching Heuristic for 3-point Ranking
MH-4R	Matching Heuristic for 4-point Ranking
MOP	Multiobjective programming
N	Neutral
NDM	Naturalistic Decision Making
NOS	National Occupational Standards
OC	Operations Commander
P	Poor
PDA	Pre-determined Attendance
PDR	Pre-determined Response
PMM	Probabilistic Mental Model

PPE	Personal Protective Equipment
PRM	Premise Risk Management
PT	Probability Theory
RPD	Recognition-Primed Decision
SA	Situation Awareness
SaaS	Soldier as a System
SC	Sector Commander
SJT	Social Judgment Theory
SIPE	Soldier Integrated Protective Ensemble
SM	Station Manager
SME	Subjective Matter Expert
SO	Safety Officer
SOP	Standard Operation Procedure
SSRI	Site Specific Risk Information
UK	United Kingdom
VG	Very Good
VP	Very Poor
WADD-3R	Weighted Additive Linear Model for 3-point Ranking
WADD-4R	Weighted Additive Linear Model for 4-point Ranking
WM	Watch Manager

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Lili Yang for her continuous guidance, advice and support throughout the study. I would also like to extend my deepest gratitude to Professor Shuang-Hua Yang and Dr. Weiwei He. Without their recommendations, I would not get the opportunity to conduct this study.

The research presented in this thesis was sponsored by Research School of Systems Engineering and School of Business & Economics, Loughborough University. I would like to express my sincere thanks to the two departments and Professor Roy Kalawsky for the financial support.

I am indebted to the support extended by the Leicestershire Fire and Rescue Service for allowing me to conduct interviews and questionnaires within the department. I especially appreciate those fire officers who participated in this research and spare their valuable time to facilitate the study.

Finally, I would like to thank my family and my friends for their continuous love, support, and encouragement throughout this research journey.

Abstract

Decision making study has been the multi-disciplinary research involving operations researchers, management scientists, statisticians, mathematical psychologists and economists as well as others. This study aims to investigate the theory and methodology of decision making research and apply them to different contexts in real cases.

The study has reviewed the literature of Multiple Criteria Decision Making (MCDM), Evidential Reasoning (ER) approach, Naturalistic Decision Making (NDM) movement, Social Judgment Theory (SJT), and Adaptive Toolbox (AT) program. On the basis of these literatures, two methods, Evidence-based Trade-Off (EBTO) and Judgment Analysis with Heuristic Modelling (JA-HM), have been proposed and developed to accomplish decision making problems under different conditions.

In the EBTO method, we propose a novel framework to aid people's decision making under uncertainty and imprecise goal. Under the framework, the imprecise goal is objectively modelled through an analytical structure, and is independent of the task requirement; the task requirement is specified by the trade-off strategy among criteria of the analytical structure through an importance weighting process, and is subject to the requirement change of a particular decision making task; the evidence available, that could contribute to the evaluation of general performance of the decision alternatives, are formulated with belief structures which are capable of capturing various format of uncertainties that arise from the absence of data, incomplete information and subjective judgments.

The EBTO method was further applied in a case study of Soldier system decision making. The application has demonstrated that EBTO, as a tool, is able to provide a holistic analysis regarding the requirements of Soldier missions, the physical conditions of Soldiers, and the capability of their equipment and weapon systems, which is critical in domain.

By drawing the cross-disciplinary literature from NDM and AT, the JA-HM extended the traditional Judgment Analysis (JA) method, through a number of novel methodological procedures, to account for the unique features of decision making tasks under extreme time pressure and dynamic shifting situations. These novel methodological procedures include, the notion of decision point to deconstruct the dynamic shifting situations in a way that decision problem could be identified and formulated; the classification of routine and non-routine problems, and associated data alignment process to enable meaningful decision data analysis across different decision makers (DMs); the notion of composite cue to account for the DMs' iterative process of information perception and comprehension in dynamic task environment; the application of computational models of heuristics to account for the time constraints and process dynamics of DMs' decision making process; and the application of cross-validation process to enable the methodological principle of competitive testing of decision models.

The JA-HM was further applied in a case study of fire emergency decision making. The application has been the first behavioural test of the validity of the computational models of heuristics, in predicting the DMs' decision making during fire emergency response. It has also been the first behavioural test of the validity of the non-compensatory heuristics in predicting the DMs' decisions on ranking task. The findings identified extend the literature of AT and NDM, and have implications for the fire emergency decision making.

Key Words: Multiple Criteria Decision Making (MCDM), Evidential Reasoning (ER), Judgment Analysis (JA), Heuristic Modelling, Social Judgment Theory (SJT), Naturalistic Decision Making (NDM)

Chapter 1 : Introduction

1.1 Background

Decision making is a critical cognitive activity of people about judging, deciding, and choosing. It has been the subject of active and multi-disciplinary research involving operations researchers, management scientists, statisticians, mathematical psychologists and economists as well as others.

The study of decision making is normally formalized within a decision environment, which is defined as the collection of information, alternatives, values, and goals available at the time of the decision.

Information is the knowledge about the decision. It could be “ecological” attributes that contribute to the context of the decision. Or it could be “physical” attributes that contribute to the measurement of alternatives. It is objective, and is independent of a decision maker’s personal preference.

Alternatives are a collection of possible course of action that could be chosen during the decision making. They indicate the solutions of a decision making activity.

Values indicate the preferences of the decision maker (DM), and are therefore, subjective. It could be the DM’s estimation of the effect of attributes. Or it could specify how desirable a particular value of the alternative is.

Goals indicate the overall objective that a DM wants to achieve. The identification of the decision goals could be cognitively demanding and be an essential part of a decision making activity.

According to the specification of a decision environment, the decision making could be further classified as: choice oriented, and process oriented (Orasanu & Connolly, 1993).

In choice oriented decision making, the information, alternative, and goals are well-defined and available, typically specified by the researcher in advance. The decision

making is then working on identifying the value preferences of the DM and how these preferences functionally prescribe the choice of alternatives. The study of this type of decision making is mostly conducted by mathematical psychologists and economists.

In process oriented decision making, all of the factors of the decision environment would not be necessarily well-defined in advance. The decision making is then involving active individual efforts of DMs, on searching information, identifying alternative and values, and formalizing the decision goals. It could involve choice of alternatives as the final part of the process. But the quality of the decision making is determined by every aspects of the decision making process as a whole, rather than solely by the final choice itself. The study of this type of decision making is mostly conducted by operations researchers, management scientists, and cognitive psychologists.

Perspectives in studying the decision making could be different, and normally be generalized into two categories: normative and descriptive (Hammond et al., 1980).

Normative perspective concerns with identifying the best decision to take, assuming an ideal DM who is fully informed, able to compute with perfect accuracy, and fully rational. The starting point of the study has been in theoretical derivation through axioms and logical system. One example of this perspective of decision study is the rational choice theory conducted by mathematical economists (Edwards, 1954). The theory states that people make decisions by determining the likelihood of a potential outcome, the value of the outcome and then multiplying the two. For instance, with a 50% chance of winning $20 or a 100% chance of winning $10, people's choice would depend on the values he or she assigned to $20 and $10. Once the values are determined, a rational choice would be prescribed as the only appropriate way of people making the decision. The "rationality" here means that an individual acts as if balancing costs against benefits to arrive at action that maximizes personal advantage (Friedman, 1953). In other words, a prescription would normally be associated with the normative analysis of decision making, which indicates how people should behave when they are confronting a specified problem.

Descriptive perspective, however, concerns with how people actually behave when they are confronting a specified problem. The starting point of the study has been in empirical observations, either from real world or through experiments.

It has been repeatedly demonstrated that DMs only infrequently behave as they "should" (e.g., Allais, 1953; Ellsberg, 1961; Kahneman & Tversky, 1979). This lack of conformity between normative prescription and descriptive validity has prompted essential responses from decision researchers. According to these responses, the decision making study could also be classified into three categories (Beach & Lipshitz, 1993).

The first category of response attributes the gap between normative theory and descriptive behaviour to the cognitive limitation of human, and strives to reduce the gap by changing the behaviour. The decision making study is then working on the practical aids to help people make their decision making processes conform to normative prescription. This is mostly conducted by operations researchers and management scientists, such as the disciplines of Decision Analysis (Von Winterfeldt & Edwards, 1986), and Multiple Criteria Decision Making (MCDM) (Figueira, Greco, & Ehrgott, 2005).

The second category of response has been to retain the general logic and structure of normative theory but to make modifications of some of the theory's components and operations in light of the empirical evidence of behaviour. This is the position taken by behavioural economics, of which prospect theory (Kahneman & Tversky, 1979) is perhaps the most famous example.

The third category of response attempts to more accurately describe the people's decision making by searching for the cognitive mechanisms that contribute to the decisions they made. The underlying notion is that, any prescription that violates the DMs' deep cognitive concerns would hardly get adopted and used in field settings (Klein, 1998). The decision making research, therefore, should aim to describe the people's decision making behaviour before prescribing changes to improve it. In other words, scholars who have provided this response are almost solely interested in understanding behaviour. Cognitive psychologists led the way in the formulation of this third response,

such as the Naturalistic Decision Making (NDM) movement (Zsambok & Klein, 1997), and Adaptive Toolbox (AT) program (Gigerenzer et al., 1999; Gigerenzer & Selten, 2001).

Finally, according to the type of subjects who are making the decision, the study of decision making could be classified into two streams: laypeople decision making and expert (domain practitioner) decision making. The study of expert decision making sees experience as an essential basis of people's decision making and works on identifying how the experience come into play of the decision making process. Laypeople decision making study, on the other hand, considered experience as a complicating factor. Subjects who know something about the task may have preconceived notions that could get in the way, or their personal strategies could distort the results. Therefore, subjects are given totally novel tasks to make sure all of them start with the same level of experience: zero (Klein, 1998).

This thesis studies the process oriented decision making of domain experts. Both normative and descriptive elements could be found in the study, the balance between which depends on the specific conditions of the decision making task.

1.2 Research Questions & Objectives

The conditions of decision making task indicate the state of environment in which a decision is made. Literature has theoretically and empirically demonstrated that specific conditions have essential influence on the people's decision making process, and in turn, shape the research theories and methodologies on the decision making topic (e.g., Simon 1955; 1990; Klein, Orasanu, Calderwood, and Zsambok, 1993).

Decision making task could be under time pressure, in which people have to make decision under certain time constraint. This may be at the level of needing action in minutes or seconds. For example, chess players under blitz conditions make move in six seconds on average (Klein et al., 1995). Fire incident commanders (ICs) have also been identified that they make around 80 percent of their decisions in less than one minute (Klein, Calderwood, & Clinton-Cirocco, 1988). Under extreme time pressure, the DMs'

thinking will shift, characteristically, in the direction of using less complicated reasoning strategies (Payne, Bettman, & Johnson, 1993). The DMs tend to be a satisficer, one seeking a satisfactory solution rather than the optimal one (Gigerenzer & Selten, 2001). Deliberate decision making, characterized by the extensive evaluation of multiple options recommended from normative analysis, is simply not feasible (Orasanu & Connolly, 1993). Decision aids, based on such deliberate analysis and normative theory of how decision should be made, are disappointing and rarely used by DMs (Klein, 1998).

Decision making could be under dynamic shifting situations, in which the dynamic change of situation would require the major shift in the way the DMs understand the situation (Klein, 1998). New information may be received, or old information invalidated, and the goals can become radically transformed. For example, as the fire develops, the IC's goal may shift from saving lives to protecting the fire spread. In Klein (1998)'s research with fire ICs, they estimated that the situation changed an average of five times per incident. Their work with U.S. Navy commanders showed the same thing. The dynamic shifting situations demands new models to account for the process dynamics of people's decision making (e.g., information search, shift in information emphasis or judgment behaviour) (Cooskey, 1996), and new methods to accomplish the investigation (Crandall, Klein, & Hoffman, 2006).

Decision making could be under imprecise goal, in which the decision making is driven by multiple conflicted objectives (Stewart, 1991). For example, in deciding the managerial plan of most businesses and public services, two objectives are mostly considered: "minimize cost" and "optimize the quality of service". However, better service can often only be attained for a price, in which two objectives conflict between each other. The conflict indicates that at some point the DM will be faced with the proposition that further achievement on one objective can only be accomplished at the expense of achievement on the other. Formal procedures are demanded to help DMs overcome their cognitive limitation in trading-off among multiple factors (Keeney & Raiffa, 1993).

Decision making could also be under uncertainty, in which a DM is lack of complete and perfect information to make decisions (Zimmermann, 2000). He may have information about some part of the problem, but not about others (incompleteness) (Xu & Yang, 2001). He may have to use his personal judgment on the information that he is not sure about, which is subjective and ambiguous in nature (randomness) (Xu & Yang, 2001). The uncertainty demands novel tools to numerically and symbolically represent and reason it with other knowledge (Halpern, 2003).

The four characteristics illustrated above fall short of providing a complete description of all possible decision making conditions in real world. They are listed here to suggest the types of decision conditions of interest in this thesis, which we believe that there have been essential gaps to be filled. It is easy to find real-world examples that embody several of the characteristics in a single task, which makes the decision making even complicated.

Literature of decision making under extreme time pressure and dynamic shifting situations is qualitative in nature (Endsley, 2004; Klein, 1998; Lipshitz, Klein, Orasanu, & Salas, 2001a), trying to describe and explain how people think when deciding strategies through notions like mental simulation, pattern matching, story building, etc. That is, while we think we have some idea of the mechanisms at play in how people make decisions, we have very little ability to determine a priori what decision a DM will make under a given situation, or to predict the ways in which one system design of a decision support system will affect the DMs' decisions as compared to another. Only through formal quantitative model, would prediction of DMs' decisions be possible.

Literature of decision making under imprecise goal and uncertainty suggests that common difficulties in making decisions with such task conditions are: the human cognitive limitations of in dealing with multiple factors, the lack of universally agreed scale to represent the uncertainty knowledge in a numerical way, and the need to combine different type of scales on the factors in a consistent manner (Keeney & Raiffa, 1993; Liu et al., 2002). The correspondent research is then focused on developing formal

procedures to help the DMs tackle these difficulties so as to support their decision making.

Accordingly, three questions are addressed to be answered in the present research.

Research Question 1: What is the proper methodology to help DMs making decision under uncertainty and imprecise goal?

Research Question 2: What is the proper methodology to quantitatively describe and predict the DMs' decision making process when they are under extreme time pressure and dynamic shifting situations?

Research Question 3: What is the applicability of the proposed methodologies in addressing the domain problems?

In order to provide answers to the research questions, the study aims to investigate the theory and methodology of decision making and apply them to different contexts in real cases. It is detailed as the following objectives to be achieved:

1) **Carry out literature review on the theory and methodology of the decision making study.**

2) **Investigate the theory and methodology used for different conditions in decision making.**

3) **Choose proper cases for the applications of the research**

4) **Discuss and propose an alternative framework to aid people's decision making under uncertainty and imprecise goal, and apply it to a case such as Soldier system decision making.**

5) **Discuss and propose a methodology to investigate DMs' decision making process when they are under extreme time pressure and dynamic shifting situations, and apply it to a case such as fire emergency decision making.**

1.3 Research methodology

A research methodology is the general research strategy that outlines the way in which the research questions are to be answered and, among other things, identifies the methods to be used in it. These methods, described in the methodology, define the means or modes of data collection and how a specific result is to be produced.

In addition to the basic methods like interview, questionnaire, and statistical techniques, four general approaches are used complementarily to accomplish the investigation of this research. They are briefed in the following. They together facilitate two novel methods being developed and applied in our research: Evidence-based Trade-Off (EBTO) and Judgment Analysis with Heuristic Modelling (JA-HM). They will be detailed in chapter 3.

Firstly, Multiple Criteria Decision Making (MCDM) approach is one major approach that facilitates the development of EBTO of our research. MCDM refers to making decisions in the presence of multiple, usually conflicting, criteria. Criterion here is defined as a tool allowing comparison of decision alternatives according to a particular significance axis or point of view (Bouyssou, 1990). The identification of the criteria is relevant to a particular context and is part of the modelling and problem formulation process. The approach provides a formal tool in tackling with decision making problem characterized by imprecise goal, in which the imprecise decision goal is, in a hierarchical fashion, decomposed into a set of measurable criteria so that justified decisions could be made which is consistent with the preferences at those measurable criteria.

Secondly, Evidential Reasoning (ER) algorithm is applied in our EBTO, as a tool for aggregating DMs' preferences at isolated criteria into general prescription of decision. The ER algorithm (Yang and Singh 1994; Yang and Sen 1994; Yang 2001; Yang and Xu 2002a,b) employs a belief structure to represent a preference judgment as a distribution. For example, the quality of a car engine can be assessed through a belief structure in the following form: {(Excellent, 20%), (Good, 40%), (Average, 30%), (Poor, 5%), (Worst, 5%)}, which is read as "Excellent with 20% of belief degree, Good with 40% of belief degree,

Average with 30% belief degree, Poor with 5% of belief degree, and Worst with 5% of belief degree". Through such belief structure, the algorithm is able to deal with MCDM problems with uncertainties and hybrid nature of information (see detailed elaboration in Section 2.2).

Thirdly, an experiment method, Judgment Analysis (JA), is modified to facilitate the development of JA-HM method of our research. An experiment is an orderly procedure carried out with the goal of verifying, refuting, or establishing the validity of a hypothesis. Experimental method allows for the precise testing of cause-and-effect (Berkowitz & Donnerstein, 1982) by demonstrating what outcome occurs when a particular factor is manipulated. It has been argued as "most effective when the goal of the research is to test theories in a deductive fashion or to study situations that rarely or never arise in natural situations" (Lind & Tyler, 1988). JA method (Cooskey, 1996) involves asking individuals to make decisions on a set of cases that might be either real or hypothetical and which comprise a combination of cues (information). Each individual's decision making process is then inferred from his or her judgments through some computational models. This method attempts to avoid the pitfalls of verbal protocol methods, like direct interview, in the context of hypothesis testing (Biggs et al., 1993; Nisbett & Wilson, 1977). Our research further extends the JA to accomplish the decision making problem under extreme time pressure and dynamic shifting situations.

Finally, heuristic modelling approach is applied in our JA-HM method to predict a DM's decision making process. A heuristic, according to the definition in (Pearl, 1983), is "a strategy using readily accessible, though loosely applicable, information to control problem solving in human beings and machines" (p.vii). Since the late 1990s, there has been a research program working on precisely specified step-by-step process models of heuristic to describe and predict the DMs' decision making process (Gigerenzer et al., 1999; Gigerenzer, Hertwig, & Pachur, 2011). The heuristics formalized in the program goes beyond the classical assumption that a heuristic trades off some accuracy for less effort (e.g., Payne, Bettman, & Johnson, 1993). Rather, people's decision making

strategies, either through optimization methods or heuristics, were taken equally as one part of their adaptive toolbox. Heuristics are efficient and sometimes optimal as long as they exploit the environment characteristics (see detailed elaboration in Section 2.5). Our research develops new heuristics and applies the heuristic modelling in a new professional domain.

1.4 Envisaged contributions

By accomplishing the research objectives outlined above, the present study will contribute to the literature of decision making research as follows:

- Through the employment of ER algorithm, we would propose a novel framework to aid DMs making decisions under imprecise goal and uncertain information.
- Through applying the framework to the case study of Soldier system decision making, our method EBTO would be demonstrated to be able to provide the holistic analysis regarding the requirements of Soldier missions, the physical conditions of Soldiers, and the capability of their equipment and weapon systems, which is critical to the domain.
- The same application would be the pioneering employment of ER algorithm in Soldier system decision making, which contribute to the empirical study of ER literature.
- The JA-HM developed in our research would provide a novel variant of JA method for the investigation of people's decision making process under extreme time pressure and dynamic shifting conditions, which contribute to the JA literature.
- The application of JA-HM to the case of fire emergency decision making would produce a panel of information requirement of ICs' resource demand decision at the on-arrival stage of high rise apartment fire incident. The application would also identify a number of characteristics of ICs' resource demand decisions, like the decision consistency, post-decisional confidence, self-insight of decision

making policy, etc. These findings would contribute to the field to better understand the ICs' resource decision making of high rise apartment incident.

- The case study of fire emergency decision making would be the first behavioural test of the validity of the computational models of heuristics, in predicting the DMs' decision making during fire emergency response. The evidence of ICs' use of matching heuristic in binary classification process of decision making would extend the literature of people's use of non-compensatory strategy in professional decision making.
- The case study of fire emergency decision making would also be the first behavioural test of the validity of the non-compensatory heuristics in predicting the DMs' decisions on ranking tasks. Our study would develop two novel non-compensatory heuristics for this type of task (MH-3R & MH-4R). They would together extend the candidate heuristics in AT literature.
- Finally, by comparing our findings of ICs' decision making process with the qualitative account suggested in the literature of NDM, this study would demonstrate how the decision making studies from different perspective (i.e., higher order information processing perspective VS. lower order mental mechanisms perspective) could complement and validate each other. The study would therefore, contribute to the cross-disciplinary research of AT and NDM.

1.5 Thesis outline

The remaining chapters of this thesis are organized in the following way.

In chapter 2, the literature review of decision making research would cover the topics on Multiple Criteria Decision Making (MCDM), Evidential Reasoning (ER) algorithm, Naturalistic Decision Making (NDM) movement, Brunswik's probabilistic functionalism approach, Social Judgment Theory (SJT), and Adaptive Toolbox (AT) program. They together contribute to the theoretical and methodological basis of our research.

In chapter 3, the literature gaps of decision making under different conditions would be further elaborated and two methods, Evidence-based Trade-Off (EBTO) and Judgment

Analysis with Heuristic Modelling (JA-HM), would be developed to address the literature gaps.

In chapter 4, a case study of Solder system decision problem is conducted to address the field gap and demonstrate the applicability of EBTO method.

In chapter 5, a case study of fire emergency decision making is presented to address the field gap and demonstrate the applicability of the JA-HM method.

Finally, the research will be concluded in chapter 6, in which the contributions, limitations, and future directions of the research will be discussed.

Chapter 2 : Literature review

Decision making study has been the multi-disciplinary research involving operations researchers, management scientists, statisticians, mathematical psychologists and economists as well as others. As briefed in Chapter 1, our research studies the process oriented decision making of domain practitioners. Both normative and descriptive elements have been involved in the study, the balance between which depends on the specific conditions of the decision making task. This chapter reviews a number of literatures of decision making study in the light of this general theme. We first brief the literature of Multiple Criteria Decision Making (MCDM). This will be followed with an elaboration of the Evidential Reasoning (ER) approach, which has been the latest development in MCDM literature. The chapter then reviews the Naturalistic Decision Making (NDM) program. The program represents a movement in literature that studies the people's decision making under tough conditions. Following the NDM program, an experimental approach of decision making study, Social Judgment Theory (SJT), is reviewed. This is followed by the review of a heuristic approach of decision making study, Adaptive Toolbox (AT) program. These literatures together contribute to the theoretical and methodology basis of the thesis.

2.1 Multiple Criteria Decision Making (MCDM)

Multiple Criteria Decision Making (MCDM) refers to making choice among decision alternatives in the presence of multiple, usually conflicting, criteria (Xu & Yang, 2001). It is a sub-discipline of operations research that explicitly considers multiple criteria of decision making tasks.

Decision alternatives here are action plans or options to be assessed or evaluated. Formal MCDM methodologies and algorithms normally purport to set up a fixed decision space, in which a set of decision alternatives are pre-defined explicitly, or defined implicitly through a mathematical programming structure.

Criterion here is defined as a tool allowing comparison of decision alternatives according to a particular significance axis or point of view (Bouyssou, 1990). The identification of the criteria is relevant to a particular context and is part of the modelling and problem formulation process. It is normally precise in itself but conflicting among each other, so that some general but imprecise goal statement could be reflected. For example, in evaluating the performance of most businesses and public services, two criteria are mostly applied: “minimize cost” and “optimize the quality of service”. However, better service can often only be attained for a price, in which two criteria conflict between each other. Sometimes due to the complexity or the scale of the problem, some criteria, in a hierarchical fashion, may break down further into lower level sub- and sub-sub criteria, to refine the problem in a more precise way. It is generally assumed that each bottom level criterion would be able to be measured in some numerical way to represent the consequences arising from implementation of any particular decision alternative.

According to Figueira, Greco, & Ehrgott (2005)’s most recent survey, MCDM methods could be classified into three general categories: value or utility based approaches (e.g., Keeney & Raiffa, 1993; Saaty, 1980; 1994), outranking approaches (e.g., Brans & Vincke, 1985; Roy, 1990), and multiobjective mathematical programming (e.g., Romero, 1986).

The value or utility based approaches work on establishing some means of associating a numerical score or value with each decision alternative, after which choice of the optimal alternative becomes automatic. The utility here refers to a real number representing the preferability of the considered decision alternative. The approaches give a sense of objectivity to the process, transparency of the method logic to the decision maker, and ease of use by non-experts. They are therefore especially appealing to those quantitatively oriented managers and the context in which “the rationale for choices must be clearly documented, and justice must be seen to be done in the sense that criteria might refer to different members of the community being served, and proper consideration of each interest must be demonstrated” (Stewart, 1992, p.571).

The outranking class of approaches identify a decision alternative through pairwise outranking relation between alternatives. The outranking relation is defined as a binary relation S on the set of decision alternatives **A** such that aSb if there is a "sufficiently strong argument" in favour of the assertion that decision alternative a is at least as good as decision alternative b (Roy, 1990). This type of approach tends to assist in understanding and visualizing the key hard judgmental choices which have to be made, and can generate tentative partial orderings of the alternatives. It is especially useful in situation where the number of decision alternatives under consideration has been reduced to a relatively small number.

Multiobjective programming (MOP) is a part of mathematical programming dealing with decision problems in which multiple and conflicting objective functions are to be optimized. Decision alternatives in MOP are implicitly defined through constraints in the form of mathematical functions. In some senses, the approach can be seen as an operationalization of Simon's "satisficing" concept (Stewart, 1992). The MOP assumes that there are some absolute target levels, above which the decision maker will always be satisfied. It is also implicitly assumed to be very unlikely that such ideal solution exists, satisfying all the targets. The aim of MOP is then to find a decision alternative that is as near as possible to the ideal solution. The process then is normally associated with additional definition of concepts like, "distance" (measure of discrepancy) from the target, reference point, dominated & non-dominated solutions (Wierzbicki, 1980), etc. The approach is a valuable means of screening a large (or even infinite) number of alternatives down to a shortlist, when the number of criteria is large (Stewart, 1992).

2.2 Uncertainty Management & Evidential Reasoning (ER) approach

When the information is uncertain for a MCDM problem, the criterion is difficult and sometimes impossible to be simply assessed through a single number.

One solution of this circumstance is, through Probability Theory (PT) (Pearl, 1988), to represent objective frequency (Pate-Cornell, 1996) or subjective degree of belief (Ng & Abramson, 1990), based on available evidence. Then the current knowledge about the

uncertainty could be represented by a probability distribution on a proposition space, and new knowledge is supposed to be learned by conditionalization (Pearl, 1988).

However, two essential limitations have been identified in PT. First, PT is unable to represent the ignorance properly (Sentz & Ferson, 2002). In PT, ignorance is represented by assigning equal probabilities to all possible states. Such representation fails to differentiate the ignorance from randomness. Because the equal beliefs can be attributed to either complete ignorance or to an equal belief in all possible states. Secondly, the sum probability of all possible states in PT must equal to 1, which imply the reinforcement of belief in one state would be associated with a decrease of belief in other states. This is not necessarily the case in real life (Zadeh, 1965). Considering the case of medical diagnosis, a positive result on some test may increase the belief that the patient has some malady, however it does not necessarily decrease the belief that the patient has any other disease.

Dempster-Shafer theory of evidence (D-S theory, Shafer, 1976), as another general framework in representing and reasoning the uncertainty, was proposed to overcome the two limitations above. But when it is applied in MCDM problem of aggregating conflict evidence, irrational results may be produced (Murphy, 2000; see details in Section 2.2.1).

The third general alternative in dealing with uncertainty is through fuzzy set theory to represent imprecise information with fuzzy membership functions (Zadeh, 1965; 1973). The fuzzy logic starts with a set of objects, U. If A is a fuzzy subset of U, then there is a function $\mu_A(u)$ which maps the elements of U into A by numbers between 0 and 1. These numbers represent the degrees of membership of the elements in set A. In this way, the fuzzy set theory is able to model the partial belonging of elements of U to a set A. If $\mu_A(u) = 1$, then the membership is absolute. And $\mu_A(u) = 0$ indicates non-membership. It is originally developed to deal with linguistic vagueness, and as a tool, is especially useful in dealing with imprecise information. However, the main concern of the theory was the meaning of information rather than its measurement (Negoita,

Zadeh, & Zimmermann, 1978). When applied in the measurement oriented problem, the final fuzzy set, which is aggregated from individual assessments, is difficult to generate an accurate prescription (Kangari & Riggs, 1989).

Finally, as the most recent development in the MCDM literature, the D-S theory was modified in the Evidential Reasoning (ER) approach to provide a rigorous reasoning process for aggregating conflict information (Yang and Singh 1994; Yang and Sen 1994; Yang 2001; Yang and Xu 2002a,b). It is considered to be a powerful alternative to overcome the limitations of PT and D-S theory in dealing with uncertainty (Liu et al, 2002).

The ER approach belongs to the category of value/utility based approach in general. But it differs from those traditional approaches in that it employs a belief structure to represent the measurement at the measureable criteria as a distributed assessment rather than a single number. For example, the distributed assessment of the quality of a public service could be {(Excellent, 20%), (Good, 40%), (Average, 30%), (Poor, 5%), (Worst, 5%)}, which means the quality of the service is assessed to be Excellent with 20% of belief degree, Good with 40% of belief degree, Average with 30% belief degree, Poor with 5% of belief degree, and Worst with 5% of belief degree. Through such belief structure, the approach is able to deal with MCDM problems with uncertainties and hybrid nature of information.

The uncertainties here refer to three types of nature. They are:

- Absence of data situation where there is no data available to assess a criterion. If this is the case, the total sum of belief degrees in the distributed assessment of that criterion will be 0.
- Incomplete assessment of a criterion situation where data for measuring a criterion is partially available. If this is the case, the total sum of belief degrees in the distributed assessment of the criterion will be between 0% and 100%.

- Random nature of a criterion situation where there exists no generally accepted probabilities being attached to states of a criterion. In such situation, personal judgments are taking place in a format of probability distribution. Distributed assessment in ER could capture such subjective judgments through transforming the probability distribution into degrees of belief distribution of the criterion.

The hybrid nature of information here includes:

- Mixture of data from incommensurable criteria;
- Mixture of data from qualitative and quantitative criteria;
- Mixture of data from deterministic and probabilistic criteria.

Since its development, the approach has been widely applied in multiple criteria decision problems like new product design assessment (Chin, Yang, Guo, & Lam, 2009), quality function deployment (Chin, Wang, Yang, & Poon, 2009), environmental impact assessment (Wang, Yang, & Xu, 2006), pipeline leak detection (Xu et al., 2007), maritime security assessment (Yang et al., 2009), fault prediction (Si et al., 2010), engineering system safety analysis (Liu et al., 2005), etc.

As mentioned above, the ER approach is developed on the basis of the Dempster-Shafer theory of evidence. In the following, the Dempster-Shafer theory of evidence is firstly briefed. The core algorithm of the ER approach is then explained in detail. To be noted, the core algorithm of the ER approach has been automated in an intelligent decision support system (IDS) (Yang & Xu, 2000), in which the lengthy and tedious model building and result analysis could be easily conducted.

2.2.1 The Dempster-Shafer theory of evidence

The Dempster-Shafer theory of evidence, or the D-S theory in short, was formally established by Dempster (1967) and Shafer (1976) as a mathematical tool for reasoning with incomplete and random information.

The D-S theory is based on two ideas: the idea of obtaining degrees of belief for one question (hypothesis) from subjective probabilities for a related question (hypothesis);

and Dempster's rule for combining such degrees of belief when they are based on independent pieces of evidence (Shafer 1990; 1992).

Illustrative example

To illustrate, consider the following example adapted from Beynon, Curry, & Morgan (2000).

Mr. Jones has been murdered, and it is confirmed that the murderer was one of three suspicious people: Peter, Paul and Mary. They compose a set of hypotheses, named "frame of discernment", {Peter, Paul, Mary}. If the only available evidence is a witness who is 80% sure that the killer is man, i.e., P(man) = 0.8. This measure is termed as a "basic probability assignment" (bpa), functioned as $m_1(\{Peter, Paul\}) = 0.8$. Since we know nothing about the remaining probability, it is allocated to the whole of the frame of discernment, i.e., $m_1(\{Peter, Paul, Mary\}) = 0.2$.

If another witness testifies that he is 60% confident that Peter was not in the country when the killing happened, this would be understood as another basic probability assignment, i.e., $m_2(\{Paul, Mary\}) = 0.6$. Again, since there is no further information about the remaining probability, it is given to the whole frame of discernment, i.e., $m_2(\{Peter, Paul, Marry\}) = 0.4$.

As these two independent sources of evidence provide different assessments to the same situation, aggregating them is crucial for generating an overall assessment that takes into consideration all points of view. Following the Dempster's rule of evidence combination, the two sources of evidence could be aggregated into a new collection of evidence as depicted in the Table 2-1 below (To be noted, the combination here is achieved through a simple multiplication rule. A more complex multiplication rule would be required and formalized in the following section to account for the more generalized approach in D-S theory).

Table 2-1, Combination of two piece of evidence

	$m_1(\{Peter, Paul\}) = 0.8$	$m_1(\{Peter, Paul, Mary\}) = 0.2$
$m_2(\{Paul, Mary\}) = 0.6$	$m_{12}(\{Paul\}) = 0.48$	$m_{12}(\{Paul, Mary\}) = 0.12$
$m_2(\{Peter, Paul, Mary\}) = 0.4$	$m_{12}(\{Peter, Paul\}) = 0.32$	$m_{12}(\{Peter, Paul, Mary\}) = 0.08$

The aggregated evidence has a more spread-out allocation of probabilities to varying subsets of the frame of discernment. We can bring together this evidence to find some level of belief (Bel): the belief in any set is the sum of all the probabilities of all the subsets of that set. For example:

$$Bel(\{Paul, Mary\}) = m_{12}(\{Paul\}) + m_{12}(\{Mary\}) + m_{12}(\{Paul, Mary\})$$
$$= 0.48 + 0 + 0.12 = 0.6$$

Formal definition

The concepts and rules illustrated above could be given in a more formal way. They are elaborated in the following.

Assuming a Universe of Discourse $\theta = \{H_1, \dots, H_N\}$, also called a Frame of Discernment, which is a set of collectively exhaustive and mutually exclusive hypotheses. A basic probability assignment (bpa) is a function $m: 2^\theta \rightarrow [0,1]$, called a mass function and satisfying

$$m(\emptyset) = 0 \text{ and } \sum_{A \subseteq \theta} m(A) = 1, \quad (1)$$

where $\emptyset$ is an empty set, A is any subset of θ, and 2^θ is the power set of θ, which consists of all the subsets of θ, i.e. $2^\theta = \{\emptyset, \{H_1\}, \dots, \{H_N\}, \{H_1 \cup H_2\}, \dots, \{H_1 \cup H_N\}, \dots, \theta\}$. The assigned probability (also called probability mass) $m(A)$ measures the belief exactly assigned to A and represents how strongly the evidence supports A. All assigned probabilities sum to unity and there is no belief in the empty set $\emptyset$. The probability assigned to θ i.e. $m(\theta)$, is called the degree of ignorance.

Associated with this structure, a belief measure (Bel) and a plausibility measure (Pl) are introduced to represent the upper and lower bounds of an interval for every subset A of θ.

The measure of belief is a mapping $Bel: 2^{\theta} \rightarrow [0,1]$ such that for any subset A of θ

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B). \quad (2)$$

It is shown in Shafer (1976) that m can be uniquely recovered from Bel as follows:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B). \quad (3)$$

The measure of plausibility is a mapping $Pl: 2^{\theta} \rightarrow [0,1]$ such that for any subset A of θ

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \quad (4)$$

$Bel(A)$ represents the exact support to A, i.e. the belief of the hypothesis A being true. $Pl(A)$ represents the possible support to A, i.e. the total amount of belief that could be potentially placed in A. Thus $[Bel(A), Pl(A)]$ constitutes the interval of support to A and can be seen as the lower and upper bounds of the probability to which A is supported. The two functions can be further connected by the equation

$$Pl(A) = 1 - Bel(\bar{A}), \quad (5)$$

where $\bar{A}$ denotes the complement of A. The difference between the belief and the plausibility of a set A describes the ignorance of the assessment for the set A (Shafer, 1976).

When there involves more than one source of evidence, the Dempster's rule would then be applied to combine them. The rule assumes that the information sources are independent and uses the orthogonal sum to combine multiple assigned probability masses: $m = m_1 \oplus m_2 \oplus \ldots \oplus m_k$, where $\oplus$ represents the operator of combination. With two assigned probability masses m_1, m_2, the Dempster's rule of combination is defined as

$$m_{12}(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K}, A \neq \emptyset, m_{12}(\emptyset) = 0, \quad (6)$$

where $K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$. The denominator $1-K$ is called the normalization factor, K is called the degree of conflict which measures the conflict between the pieces of evidence.

This combination rule satisfies the commutative and associative properties, where $m_1 \oplus m_2 = m_2 \oplus m_1, m_1 \oplus (m_2 \oplus m_3) = (m_1 \oplus m_2) \oplus m_3)$.

Limitation

However, the crude application of the D-S theory and the combination rule can lead to irrational conclusions in the aggregation of multiple pieces of evidence in conflict (Murphy, 2000). Take for example, examine the following two pieces of evidence in conflict:

$$m_1(A) = 0.99, m_1(B) = 0.01, m_1(C) = 0,$$

$$m_2(A) = 0, m_2(B) = 0.01, m_2(C) = 0.99.$$

Before combination, the two pieces of evidence show that B is an unlikely event as only a probability of 1% is assigned to it. After the implementation of the original D-S combination rule, B becomes a certain event and is assigned a probability of 100%, which does not make sense. Table 2-2 shows the results generated by employing the original D-S combination rule.

Table 2-2, The combination of conflict evidence by the D-S combination rule (Wang, Yang, Xu, & Chin, 2006).

Belief structure	A	B	C
m_1	0.99	0.01	0
m_2	0	0.01	0.99
$m_1 \oplus m_2$ (before normalization)	0	0.0001	0
$m_1 \oplus m_2$ (after normalization)	0	1	0

In addition, due to the need to enumerate all subsets or supersets of a given subset A of θ, the direct use of the combination rule will result in an exponential increase in computational complexity (Buchanan & Shortliffe, 1984). These issues are addressed in

the ER approach by a distributed modelling framework – belief structure, and re-distributing the unassigned probability mass to the criteria according to their relative importance (i.e. weights).

2.2.2 The ER algorithm

The core algorithm of ER approach is briefly explained here. Suppose that a MCDM problem has M alternatives $a_l (l = 1, \dots, M)$, one upper level criterion, referred to as a general criterion, and L lower level measureable criteria $e_i (i = 1, \dots, L)$, called basic criteria. Then a decision matrix can be constructed as follows. Firstly, weightings are assigned to the L basic criteria to show their relative importance over one another: $W = \{w_i, i = 1, \dots, L\}$ and the weights are normalized so that $\sum_{i=1}^{L} w_i = 1$, where $0 \leq w_i \leq 1$. Secondly, to assess alternatives on each basic criterion, a set of mutually exclusive and collectively exhaustive evaluation grades is defined: $H = \{H_n, n = 1, \dots, N\}$. A MCDM problem can then be modelled as follows:

$$S\big(e_i(a_l)\big) = \{\big(H_n, \beta_{n,i}(a_l), n = 1, \dots, N\big\}, i = 1, \dots, L, l = 1, \dots, M, \tag{7}$$

$\beta_{n,i}(a_l) \geq 0$ and $\sum_{n=1}^{N} \beta_{n,i}(a_l) \leq 1$. $\beta_{n,i}(a_l)$ represents a degree of belief. The expression reads that an alternative l with respect to criterion e_i is assessed to a grade H_n with a degree of belief $\beta_{n,i}(a_l)$ $(n = 1, \dots, N)$, which is a distributed assessment and referred to as a belief structure. If $\sum_{n=1}^{N} \beta_{n,i}(a_l) = 1$, the assessment is said to be complete; otherwise, it is incomplete. Note that $\sum_{n=1}^{N} \beta_{n,i}(a_l) = 0$ denotes total ignorance about the assessment of a_l on e_i.

To combine the evidence, the belief degrees are first transformed into a basic probability mass by multiplying the relative weights with the following equations as shown below:

$$m_{n,i} = m_i(H_n) = w_i \beta_{n,i}(a_l), n = 1, \dots, N; i = 1, \dots, L, \tag{8}$$

$$m_{H,i} = m_i(H) = 1 - \sum_{n=1}^{N} m_{n,i} = 1 - w_i \sum_{n=1}^{N} \beta_{n,i}(a_l), i = 1, \dots, L, \tag{9}$$

$$\bar{m}_{H,i} = \bar{m}_i(H) = 1 - w_i, i = 1, \dots, L, \tag{10}$$

$$\widetilde{m}_{H,i} = \widetilde{m}_i(H) = w_i(1 - \sum_{n=1}^{N} \beta_{n,i}(a_l)),\ i = 1, \dots, L, \tag{11}$$

with $m_{H,i} = \overline{m}_{H,i} + \widetilde{m}_{H,i}$ and $\sum_{i=1}^{L} w_i = 1$.

Note that the probability mass assigned to the whole set H, $m_{H,i}$, which is currently unassigned to any individual grades, is split into two parts: $\overline{m}_{H,i}$ and $\widetilde{m}_{H,i}$, where $\overline{m}_{H,i}$ is caused by the relative importance of the criterion e_i and $\widetilde{m}_{H,i}$ is caused by the incompleteness of the assessment on e_i for a_l.

The basic probability masses generated above are then be aggregated into the combined probability assignments with D-S combination rule in a recursive fashion as follows:

$$\{H_n\}: m_{n,I(i+1)} = K_{I(i+1)}[m_{n,I(i)}m_{n,i+1} + m_{n,I(i)}m_{H,i+1} + m_{H,I(i)}m_{n,i+1}],$$

$$m_{H,I(i)} = \overline{m}_{H,I(i)} + \widetilde{m}_{H,I(i)},\ n = 1, \dots, N, \tag{12}$$

$$\{H\}: \widetilde{m}_{H,I(i+1)} = K_{I(i+1)}[\widetilde{m}_{H,I(i)}\widetilde{m}_{H,i+1} + \widetilde{m}_{H,I(i)}\overline{m}_{H,i+1} + \overline{m}_{H,I(i)}\widetilde{m}_{H,i+1}], \tag{13}$$

$$\{H\}: \overline{m}_{H,I(i+1)} = K_{I(i+1)}[\overline{m}_{H,I(i)}\overline{m}_{H,i+1}], \tag{14}$$

$$K_{I(i+1)} = [1 - \sum_{n=1}^{N} \sum_{t=1,t\neq n}^{N} m_{n,I(i)}m_{t,i+1}]^{-1},\ i = 1, \dots, L-1, \tag{15}$$

In the above equations, $m_{n,I(i)}$ denotes the combined probability mass generated by aggregating the first i criteria; $m_{n,I(i)}m_{n,i+1}$ measures the relative support to the hypothesis that the general criterion should be assessed to the grade H_n by both the first i critera and the $(i+1)$th criterion; $m_{n,I(i)}m_{H,i+1}$ measures the relative support to the hypothesis by the first i criteria only; $m_{H,I(i)}m_{n,i+1}$ measures the relative support to the hypothesis by the $(i+1)$th criterion only. It is assumed in the above equations that $m_{n,I(1)} = m_{n,1}(n = 1, \dots, N)$, $m_{H,I(1)} = m_{H,1}$, $\overline{m}_{H,I(1)} = \overline{m}_{H,1}$ and $\widetilde{m}_{H,I(1)} = \widetilde{m}_{H,1}$. Note that the aggregation process does not depend on the order in which criteria are combined.

Finally, the combined probability masses are normalized into the degrees of belief on the general criterion, with the following equations:

$$\{H_n\}: \beta_n = \frac{m_{n,I(L)}}{1-\bar{m}_{H,I(L)}}, n = 1, \dots, N, \tag{16}$$

$$\{H\}: \beta_H = \frac{\tilde{m}_{H,I(L)}}{1-\bar{m}_{H,I(L)}}. \tag{17}$$

where β_n and β_H represent the overall degree of belief of the combined basic probability mass assigned to the assessment grades H_n and H, respectively. The combined assessment can be denoted by $S\big(y(a_l)\big) = \{\big(H_n, \beta_n(a_l)\big), n = 1, \dots, N\}$. It has been proved that $\sum_{n=1}^{N} \beta_n + \beta_H = 1$ (Yang and Xu, 2002a).

In above ER algorithm, Eqs. (12)-(15) are the direct implementation of the D-S combination rule within the belief decision matrix; the weight normalization process, the assignment of the basic probability masses shown in Eqs. (8)-(11) and the normalization of the combined probability masses shown in Eqs. (16) and (17) are developed to ensure that the ER algorithm can process conflicting evidence rationally that satisfies common sense rules for criteria aggregation in MCDM (Yang and Xu, 2002a).

Now, the example of conflict evidence aggregation with D-S combination rule shown in Table 2-2 can be re-organized with ER as follows:

Table 2-3, The combination of conflict evidence by the ER algorithm (Wang, Yang, Xu, & Chin, 2006).

$m_1 \oplus m_2$	m_1			
	$m_1(A) = 0.99w_1$	$m_1(B) = 0.01w_1$	$m_1(C) = 0$	$m_1(\overline{H}) = 1 - w_1$
m_2				
$m_2(A) = 0$	0	0	0	0
$m_2(B) = 0.01w_2$	$0.0099w_1w_2$	$0.0001w_1w_2$	0	$0.01(1-w_1)w_2$
$m_2(C) = 0.99w_2$	$0.9801w_1w_2$	$0.0099w_1w_2$	0	$0.99(1-w_1)w_2$
$m_2(\overline{H}) = 1 - w_2$	$0.99w_1(1-w_2)$	$0.01w_1(1-w_2)$	0	$(1-w_1)(1-w_2)$
Combined basic probability assignment (before normalization)	$0.99w_1\ (1-w_2)$	0.01 - $0.0199w_1w_2$	$0.99(1-w_1)w_2$	$(1-w_1)(1-w_2)$
Combined basic probability assignment (after normalization)	$\frac{0.99w_1(1-w_2)}{1-0.9999w_1w_2}$	$\frac{0.01-0.0199w_1w_2}{1-0.9999w_1w_2}$	$\frac{0.99(1-w_1)w_2}{1-0.9999w_1w_2}$	$\frac{(1-w_1)(1-w_2)}{1-0.9999w_1w_2}$
Combined belief degrees	$\frac{0.99w_1(1-w_2)}{1-1.9999w_1w_2}$	0.01	$\frac{0.99(1-w_1)w_2}{1-1.9999w_1w_2}$	0

Note: w_1 and w_2 are the relative weights of the evidence m_1 and m_2 satisfying $w_1 + w_2 = 1$.

The combined belief degrees depend to a large extent on the assignment of the relative weights to the two pieces of evidence. Different weights lead to different belief degrees as shown in Table 2-4 below:

Table 2-4, The combined belief degrees under different sets of relative weights (Wang, Yang, Xu, & Chin, 2006).

Relative weights		$(m_1 \oplus m_2)(A)$	$(m_1 \oplus m_2)(B)$	$(m_1 \oplus m_2)(C)$
$w_1 = 0,$	$w_2 = 1$	0	0.01	0.99
$w_1 = 0.2,$	$w_2 = 0.8$	0.058	0.01	0.932
$w_1 = 0.4,$	$w_2 = 0.6$	0.305	0.01	0.685
$w_1 = 0.5,$	$w_2 = 0.5$	0.495	0.01	0.495
$w_1 = 0.6,$	$w_2 = 0.4$	0.685	0.01	0.305
$w_1 = 0.8,$	$w_2 = 0.2$	0.932	0.01	0.058
$w_1 = 1,$	$w_2 = 0$	0.99	0.01	0

It is obvious that such conclusions produced by the ER algorithm make sense in comparison to the irrational conclusions generated from the original D-S theory.

2.2.3 Utility based ranking for belief structure

Finally, utility measures have also been introduced to facilitate the ranking among alternatives.

Suppose the utility of an evaluation grade H_n is $u(H_n)$, then the expected utility of the aggregated assessment $S(y(a_l))$ is defined as follows:

$$u\left(S(y(a_l))\right) = \sum_{n=1}^{N} \beta_n(a_l) u(H_n). \tag{18}$$

The belief degree $\beta_n(a_l)$ stands for the lower bound of the likelihood that a_l is assessed to H_n, whilst the corresponding upper bound of the likelihood is given by $(\beta_n(a_l) + \beta_H(a_l))$, which leads to the establishment of a utility interval if the assessment is incomplete. Without loss of generality, suppose the least preferred assessment grade having the lowest utility is H_1 and the most preferred assessment grade having the highest utility is H_N. Then the maximum, minimum and average utilities of a_l can be calculated by

$$u_{max}(a_l) = \sum_{n=1}^{N-1} \beta_n(a_l) u(H_n) + \left(\beta_N(a_l) + \beta_H(a_l)\right) u(H_N), \tag{19}$$

$$u_{min}(a_l) = \left(\beta_1(a_l) + \beta_H(a_l)\right) u(H_1) + \sum_{n=2}^{N} \beta_n(a_l) u(H_n), \tag{20}$$

$$u_{avg}(a_l) = \frac{u_{max}(a_l)+u_{min}(a_l)}{2}. \quad (21)$$

It can be seen that if all original assessments $S(e_i(a_l))$ in the generalised decision matrix are complete, then $\beta_H(a_l) = 0$ and $u\left(S\left(y(a_l)\right)\right) = u_{max}(a_l) = u_{min}(a_l) = u_{avg}(a_l)$.

2.2.4 Rule and utility based transformation technique for evaluation grades

The basic framework of the ER approach was first proposed in (Yang and Singh 1994, Yang and Sen 1994), but this original approach is based on the assumption that all criteria are qualitative and employ the same set of evaluation grades. However, to facilitate data collection during the application, it is more natural and acceptable to acquire assessment information in a manner appropriate to a particular criterion. On the one hand, a basic criterion may be quantitative and can be measured using precise numbers or several values with different probabilities. On the other hand, a qualitative basic criterion may be assessed using a set of grades appropriate for this criterion but different from others. To facilitate the assessment aggregation, it is fundamental to transform the various sets of evaluation standards to a unified set so that all criteria can be assessed in a consistent and compatible manner. Hence, Yang (2001) developed the rule and utility based techniques to provide systematic yet rational ways for transforming information from one form to another.

The principle in conducting the transformation is to define a set of grades for a general criterion and then equivalently and rationally transform other sets of grades defined from basic criteria to this general set. The detailed transformation process could be summarized into four categories:

a) Rule based transformation of qualitative assessment:

If a grade H_n in a basic set H^i defined as $H^i = \{H_{n,i}, n = 1, \dots, N_i\}$ for criterion i, means a grade H_l in a general set H defined as $H = \{H_n, n = 1, \dots, N\}$ to a degree of $\alpha_{l,n}(l = 1, \dots, N)$ with $0 \leq \alpha_{l,n} \leq 1$ and $\sum_{l=1}^{N} \alpha_{l,n} = 1$, then it is said that

$H_{n,i}$ is equivalent to $\{(H_l, \alpha_{l,n}), l = 1, \dots, N\}$. (22)

Given Eq. (22), a basic assessment $S^i(e_i)$ defined as

$$S^i(e_i) = \{(H_{n,i}, \gamma_{n,i}), n = 1, \dots, N_i\}, \quad (23)$$

is said to be equivalent to a general assessment $S(e_i)$ defined as

$$S(e_i) = \{(H_{n,i}, \beta_{n,i}), n = 1, \dots, N\}, \quad (24)$$

if and only if

$$\beta_{l,i} = \sum_{n=1}^{N_i} a_{l,n}\gamma_{n,i},\ l = 1, \dots, N. \quad (25)$$

b) Rule based transformation of quantitative assessment in the format of precise number:

Suppose a value $h_{n,i}$ for an criterion e_i is judged to be equivalent to a grade H_n, or

$h_{n,i}$ means $H_n (n = 1, \dots, N)$. (26)

Without loss of generality, suppose e_i is a "profit" criterion, that is a large value $h_{n+1,i}$ is preferred to a smaller value $h_{n,i}$. Let $h_{N,i}$ be the largest feasible value and $h_{1,i}$ the smallest. Then a value h_j on e_i may be represented using the following equivalent expectation:

$$S^i(h_j) = \{(h_{n,i}, \gamma_{n,j}), n = 1, \dots, N\}, \quad (27)$$

Where

$$\gamma_{n,j} = \frac{h_{n+1,i} - h_j}{h_{n+1,i} - h_{n,i}}, \gamma_{n+1,j} = 1 - \gamma_{n,j}\ if\ h_{n,i} \le h_j \le h_{n+1,i}, \quad (28)$$

$$\gamma_{k,j} = 0\ for\ k = 1, \dots, N, k \ne n, n+1. \quad (29)$$

Given the equivalence rules described in Eq. (26), a value h_j can be represented by the following equivalent expectation:

$$\mathrm{S}(h_j) = \{(H_n, \beta_{n,j}), n = 1, \dots, N\}, \tag{30}$$

Where

$$\beta_{n,j} = \gamma_{n,j}, \mathrm{n} = 1, \dots, \mathrm{N}. \tag{31}$$

The above transformation is proved to be justified if the underlying utility function of the criterion e_i is assumed to be piecewise linear.

c) Rule based transformation of quantitative assessment in the format of random variable:

A random variable briefly is a collection of several values with different probabilities for describing all the possible value of a criterion. It could be formulated as the following distribution:

$$S^i(e_i) = \{(H_j, p_j), j = 1, \dots, M_i\}, \tag{32}$$

Where $h_j (j = 1, \dots, M_i)$ are possible values that e_i may take and p_j is the probability that e_i takes a value h_j, where $\sum_{j=1}^{M_i} p_j \leq 1$. The above distribution reads that an criterion e_i takes a value h_j with a probability $p_j (j = 1, \dots, M_i)$. Note that e_i taking a single value h_j for sure is a special case of Eq. (32) with $p_j = 1$ and $p_l = 0 (l = 1, \dots, M_i, l \neq j)$.

Assuming a piecewise linear utility function for e_i, the distribution $S^i(e_i)$ given by Eq. (32) can be equivalently represented using $h_{n,i}$ by

$$\tilde{\mathrm{S}}^i(h_j) = \{(h_{n,i}, \tilde{\gamma}_{n,j}), n = 1, \dots, N\}, \tag{33}$$

Where

$$\tilde{\gamma}_{n,j} = \begin{cases} \sum_{j \in \pi_n} p_j \gamma_{n,j}, \ for\ n = 1, \\ \sum_{j \in \pi_{n-1}} p_j (1 - \gamma_{n-1,j}) + \sum_{j \in \pi_n} p_j \gamma_{n,j}, \ for\ 2 \leq n \leq N - 1, \\ \sum_{j \in \pi_{n-1}} p_j (1 - \gamma_{n-1,j}), \ for\ n = N, \end{cases} \tag{34}$$

$$\pi_n = \begin{cases} \{j | h_{n,i} \le h_j < h_{n+1,i}, j = 1, \dots, M_i\}, n = 1, \dots, N-2, \\ \{j | h_{n,i} \le h_j \le h_{n+1,i}, j = 1, \dots, M_i\}, n = N-1, \end{cases} \quad (35)$$

and $\gamma_{n,j}$ is calculated by Eqs. (28) & (29). Note that $\pi_l \cap \pi_k = \emptyset (\mathrm{l}, \mathrm{k} = 1, \dots, \mathrm{N}-1; \mathrm{l} \neq \mathrm{k})$ and $\bigcup_{n=1}^{N-1} \pi_n = \{1,2, \dots, M_i\}$.

Given the equivalence rule represented by Eq. (26), $S^i(e_i)$ can be equivalently represented by the following exppectation suing the general grade set:

$$\mathrm{S}(e_i) = \{(H_n, \beta_{n,i}), n = 1, \dots, N\}, \text{ with } \beta_{n,j} = \tilde{\gamma}_{n,j}. \quad (36)$$

d) Utility based transformation:

It can be seen in the three categories above, rule based transformation, as its name, is conducted using the decision maker's knowledge and experience described as rules, such as equivalence rules described by Eq. (22) & (26).

Now consider the Eq. (23) for qualitative criteria and Eq. (27) for quantitative criteria, a utility based transformation technique could also be formulated, in which the decision maker's knowledge and experience is described as utility estimations.

Suppose the utilities of the evaluation grade $H_j (\mathrm{j} = 1, \dots, \mathrm{N})$ of a general criterion y have been estimated and denoted by $\mathrm{u}(H_j)(j = 1, \dots, N)$. Then, given $\mathrm{u}(H_j)(j = 1, \dots, N)$ and $\mathrm{u}(H_{n,i})(n = 1, \dots, N)$ (for qualitative criteria) or $\mathrm{u}(h_{n,i})(\mathrm{n} = 1, \dots, \mathrm{N})$ (for quantitative criteria), an original assessment $S^i(e_i)$ can be transformed to an equivalent expectation $\mathrm{S}(e_i)$ as follows:

$$\mathrm{S}(e_i) = \{(H_j, \beta_{j,i}), j = 1, \dots, N\}, \quad (37)$$

Where

$$\beta_{j,i} = \begin{cases} \sum_{j \in \pi_j} \gamma_{n,i} \tau_{j,n}, \ for\ j = 1, \\ \sum_{n \in \pi_{j-1}} \gamma_{n,i} (1 - \tau_{j-1,n}) + \sum_{n \in \pi_j} \gamma_{n,i} \tau_{j,n}, \ for\ 2 \le j \le N-1, \\ \sum_{n \in \pi_{j-1}} \gamma_{n,i} (1 - \tau_{j-1,n}), \ for\ j = N, \end{cases} \quad (38)$$

For a qualitative criterion

$$\tau_{j,n} = \frac{u(H_{j+1}) - u(H_{n,i})}{u(H_{j+1}) - u(H_j)} \text{ if } \mathrm{u}(H_j) \le u(H_{n,i}) \le u(H_{j+1}), \tag{39}$$

For a quantitative criterion

$$\tau_{j,n} = \frac{u(H_{j+1}) - u(h_{n,i})}{u(H_{j+1}) - u(H_j)} \text{ if } \mathrm{u}(H_j) \le u(h_{n,i}) \le u(H_{j+1}), \tag{40}$$

And

$$\pi_j = \begin{cases} \{n | u(H_j) \le u(H_{n,i}) < u(H_{j+1}), n = 1, \dots, N_i\}, j = 1, \dots, N-2, \\ \{n | u(H_j) \le u(H_{n,i}) \le u(H_{j+1}), n = 1, \dots, N_i\}, j = N-1. \end{cases} \tag{41}$$

Note that $\pi_l \cap \pi_k = \emptyset (\mathrm{l}, \mathrm{k} = 1, \dots, \mathrm{N}-1; \mathrm{l} \neq \mathrm{k})$ and $\bigcup_{j=1}^{N-1} \pi_j = \{1,2, \dots, N_i\}$.

e) Summary of transformation

It is summarized that information transformation could be conducted at three levels. If no preference information is available, it could be assumed that the utilities of evaluation grades for a qualitative criterion are equidistantly distributed in the normalised utility space and a linear utility function might be assumed for a quantitative criterion. At this basic level, there is no participation of the decision maker in information transformation. If the decision maker has sufficient expertise in analysing an assessment problem but is not confident in estimating utilities, the rule based technique could be used for information transformation. If the decision maker is capable of estimating utilities, information transformation could be conducted through utility estimation.

2.3 Naturalistic Decision Making (NDM)

The NDM movement is a collection of field study attempting to understand how people make decisions in real-world contexts that are meaningful and familiar to them. The

movement, as a community of practice, was initiated in 1989 in a conference in Dayton, Ohio, sponsored by the Army Research Institute. A group of 30 researchers who studied decision making in natural settings met for several days in an effort to find commonalities between the decision making processes of firefighters, Navy officers, Army commanders, nuclear power plant controllers, and other populations (Kahneman & Klein, 2009). Most of them were applied scientists who found the substantial existing literature on human judgment and decision making unsatisfactory for addressing their particular system design or training needs and went looking for a different approach (Howell, 1997). No less than nine NDM models were identified (Lipshitz 1993) specific to different field settings. Two of the most widely cited models were Rasmussen's (1983, 1986) skills/rules/knowledge account along with the 'decision ladder', and Klein's Recognition-Primed Decision (RPD) model. Working separately, they all reached similar conclusions. People were not generating and comparing option sets. People were using prior experience to rapidly categorize situations. People were relying on some kind of synthesis of their experience – call it a schema or a prototype or a category – to make these judgments (Klein, 2008). Eight factors were highlighted in characterizing the complexities of decision making in real-world settings, including: ill-structured problems; uncertain dynamic environments; shifting, ill-defined or competing goals; action/feedback loops; time stress; high stakes; multiple players, and organizational goals and norms (Orasanu & Connolly, 1993).

Since then, the NDM community has met every 2 or 3 years, alternating between North American and European venues. Each of the NDM meetings has generated a book describing the research and the ideas of the conference participants (Zsambok & Klein, 1997; Flin et al., 1998; Salas & Klein, 2001; Montgomery et al., 2004; Hoffman, 2006; Schraagen et al., 2008; Mosier et al., 2010; Wong et al., 2009). The movement has also received cooperation from the American Psychological Association (APA) and Human Factors and Ergonomics Society (HFES) by the late 1990s (Moon, 2002). In addition, the Cognitive Engineering and Decision making Technical Group, formed to provide an

outlet for NDM research, has become one of the largest and most active in the HFES (Klein, 2008).

As one of the most widely cited model in NDM literature, Klein (1993)'s Recognition-Primed Decision (RPD) model is briefed here, as the prototypical model (Lipshitz, Klein, Orasanu, & Salas, 2001a), to shed light on the advantage and limitations of methodologies and decision models developed in this literature.

The RPD model was originally developed on the basis of interviews and observations of fire ground commanders working in challenging circumstances (Klein et al., 1986). Semi-structured interviews with probe questions were conducted with 26 firefighters who had an average of 23 years of experience, to obtain retrospective data about 156 highly challenging decision points during critical incidents. The data of firefighters' decision making process were synthesized into three variations (Figure 2-1).

In the simplest variation, a DM sizes up a situation and responds with the initial option identified. In the second variation, more complex situation is encountered in which the DM performs some conscious evaluation of the option, typically using mental simulation, to uncover problems prior to carrying it out. In the most complex circumstance (the third variation), the evaluation reveals flaws and the option is modified accordingly, or the option is judged inadequate and rejected in favour of the next most typical option. The three variations explain how DMs can handle the constraints and stressors often found in tough field settings. Under extreme time pressure, the first variation will result in reasonable reactions without the need to perform any deliberations or analysis. Under dynamic shifting situations, the DM is prepared to react quickly without having to re-do analysis.

Since the initial proposal, the three variations of decision behaviour suggested by the RPD model have been replicated in other domains, like neonatal intensive care (Crandall & Calderwood, 1989), Chess playing (Klein, Wolf, Militello, & Zsambok, 1995), Platoon Commanding (Brezovic, Klein, & Thordsen, 1987), Army command and control (Pascual

& Henderson, 1997), all of which involves extreme time pressure and dynamic shifting situations.

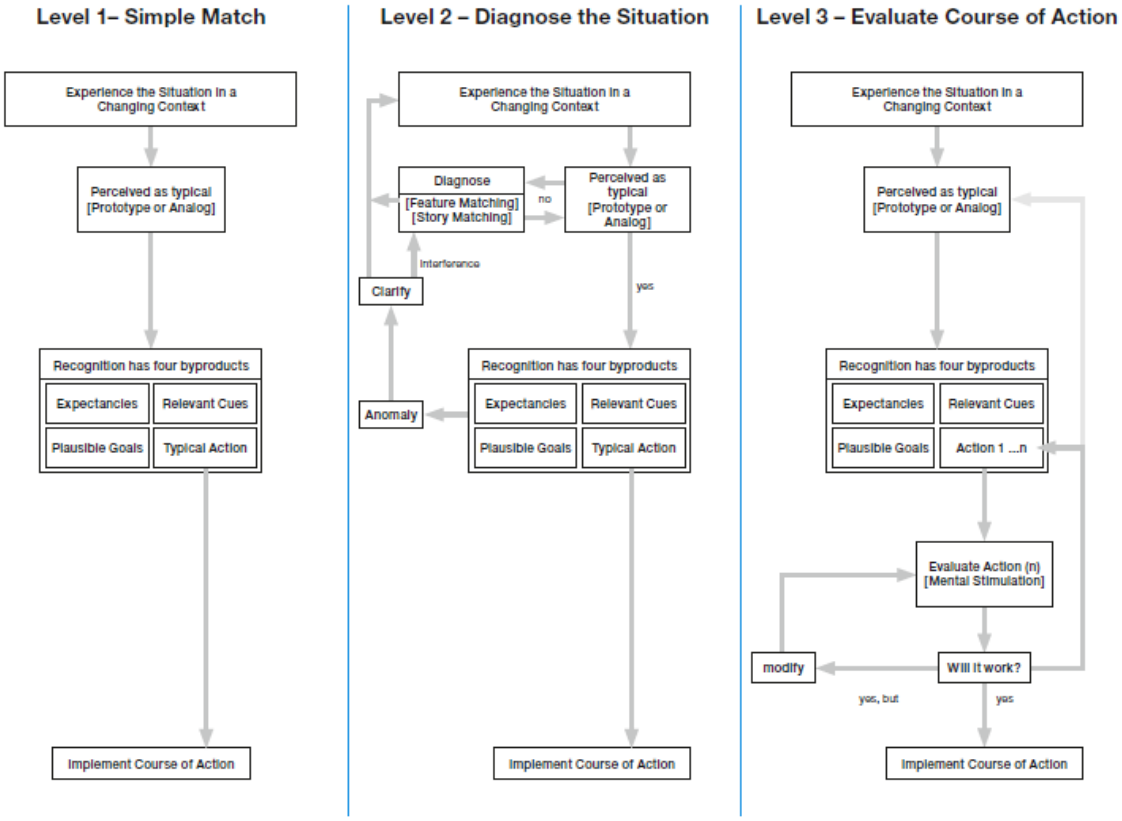


Figure 2-1, Recognition-Primed Decision (RPD) model (Klein, 1993)

2.3.1 Characteristics of NDM

As a collection of research effort, the NDM community has no general theory like other formal metaphors. But the research produced by the community could be marked by five essential characteristics: proficient decision makers, situation-action matching decision rules, context-bound informal modelling, process orientation, and empirical-based prescription (Lipshitz, Klein, Orasanu, & Salas, 2001a).

Proficient decision makers

It has been repeatedly acknowledged and highlighted that the primary factor defining NDM studies is expertise. NDM researchers do not see domain practitioners as infallible, but nevertheless respect their dedication, skills, and knowledge (Schraagen, Klein,

Hoffman, 2008). This attitude towards the proficient decision makers stems from their philosophy attitude towards rationality. Working in the field setting in the first place, they have a natural sympathy of the belief that the shaping features of the task ecology have a major effect on the decision performance. People trade decision making policy for the environment demands. They see their point of view as Simon's bounded organizational studies at the individual decision level (Beach & Lipshitz, 1993), a version of "bounded rationality" that rejects the omnipotent rational assumption due to the constraints of environment rather than blaming the incapability of human reasoning. In this sense, "People with greater expertise can see the world differently. They have a larger storehouse of procedures to apply. They notice problems more quickly. They have richer mental simulations to use in diagnosing problems and in evaluating courses of action. They have more analogies to draw upon." (Klein, 1998 pp.280). They are generally considered as the important sources of power other than rational analysis.

Process orientation

In contrast to input-output orientation of standard normative decision analysis, NDM researchers, through their observation of real world decisions, suggest the difference between the decision events and decision-making activities (Orasanu & Connolly, 1993). Research on decision events tends to focus on the ways in which decision makers pull together all available information into their choice of a best alternative. The decision-making activities suggested in real world however, offer few clean examples of decision events. Pre-decision processes appear to be critical to success (McLennan and Omodei, 1996). The emphasis in NDM is then expanded to those cognitive functionalities that are involved in the decision process (see Klein et al., 2003; Ross et al., 2006; Schraagen et al., 2008). From this perspective, the term NDM is just a historical accident stemming from the adoption of the traditional judgment and decision making literature as the reference point in the mid-1980s (Beach et al., 1997).

Situation-action matching decision rule

Lipshitz et al., (2001a) defines the "matching" here as the generic label for decisions with the basic structure of "Do A because it is appropriate for situation S". In defending their position towards the concurrent choice formalized in normative decision analysis, NDM researchers argue a paradox that: if the differences between choices are large and apparent, then the decision would be straightforward enough that the decision maker do not actually demand to deliberate choose at all; if the differences between the choices diminish, the decision maker would then wind up concentrating on trivial issues because of the "zone of indifference", which describes a phenomenon whereby the closer two options are to each other, the less important the consequences of selecting the best, but the more difficult the choice (Minsky, 1986). Empirical findings in NDM reflect the adaption of Subjective Matter Experts (SMEs) against this paradox in real world, especially with tough settings. It is suggested that SMEs do not wrestle with choice (e.g., Brezovic et al., 1987; Klein, Calderwood, & Clinton-Cirocco, 1986; Klein, Wolf, Militello, & Zsambok, 1995; Stokes et al., 1990; also see reviews in Klein 1998; Ross et al., 2006). Instead, they generate options sequentially along the time, and typically carry out the first course of action they identified. Their ability to handle decision points appeared to depend on their skill at recognizing situation as typical, as instances of general prototypes that they had developed through experience. In situation where is unusual or uncertain enough for evaluating, 'satisficing' is applied to adjust and diminish those difference from a standard prototype. Emphasis shifts from choosing between alternatives to situation assessment.

Context-bound informal modelling

As the expertise is certainly domain- and context-specific (Ericsson & Lehman, 1996), the context itself avoidably becomes the target of interventions, in looking at the ways that SMEs' reason within their own domains. In this sense, generic context-free method certainly could not fulfil this requirement, or performed at a higher level. In addition, NDM researchers urgently doubt if the complexity of environment could be reconstructed in simplified formal task, and if those reasoning processes emerged in

real world are still available for study in lab (Klein et al., 2003; Woods, 1988). Consequently, they draw a great deal of qualitative method to capture the decision making phenomena in an informal way, such as ethnography, cognitive task analysis (Hoffman & Militello, 2009). This certainly raises the critical debate regarding the study rigor, falsifiability, and generalizability (Lipshitz et al., 2001b).

Empirical-based prescription

In a quite pragmatic view, NDM researchers believe that "prescriptions which are optimal in some formal sense but which cannot be implemented are worthless" (Lipshitz et al., 2001a, p.335). This is quite different from the normative based prescription, like MCDM, in which formal proofs of optimization are believed to be independent of its descriptive validity. They define their empirical-based prescription as "deriving prescriptions from descriptive models of expert performance" (Lipshitz et al., 2001a, p.335). The application, then, is, in a qualitative manner, to improve feasible decision makers' characteristic modes of making decisions, rather than replacing them altogether.

2.4 Social Judgment Theory (SJT)

NDM's ecological and cognitive focus on decision making study is actually not new. There is a long history in psychology that works on the interaction between environmental and cognitive systems. One of the research metaphors that extend its investigation into human judgment is Kenneth Hammond's Social Judgment Theory (SJT) (Hammond, Stewart, Brehmer, & Steinmann, 1975).

Researchers in SJT share some essential theoretical arguments of NDM. On the one hand, it is acknowledged that the ambiguous environment must somehow be described. On the other hand, it highlights the cognitive processes, such as perception, learning, and thinking, to bear on the problem of reducing causal ambiguity of the environment. The emphasis, thus, placed on the more general process of knowing, in contrast to choice which was taken as a sub-problem. Human judgement, as an exercise, is a

cognitive activity of last resort. But differently, they attempts to represent such human judgment process in a formal integrated system.

The SJT, in general, is the result of a systematic application of Brunswik's probabilistic functionalism to the problem of human judgment in social situations. Therefore, to capture the theme of SJT, let's look at the Brunswik's probabilistic functionalism first.

2.4.1 Brunswik's probabilistic functionalism

Brunswik's probabilistic functionalism embodies several critical concepts and ideas, many of which were new to psychology at the time (Brunswik, 1943; 1952; 1955c; 1956; for reviews, see Hammond, 1966b; Hammond & Stewart, 2001; Postman & Tolman, 1959). They are illustrated in the following.

2.4.1.1 The role of ecology

From Brunswik's point of view, the environment in which an organism is embedded, and the organism itself should receive equal emphasis in the decision analysis. A proper decision analysis should begin by considering the distal states of affairs in the ecology with which the organism must cope (Hammond, 1966b). A distal variable presents itself to the perceiving organism by a set of **proximal cues** which are then processed centrally within the organism to yield some functional response (judgment). Thus, there exist two systems: the task system which is defined in terms of the relations between the cues and the distal variable of interest to the organism; and the cognitive system which is defined in terms of the relations between the cues and the central functional response. The fundamental decision analysis is then to be the simultaneous investigation of those two systems within the organism-ecology interface.

2.4.1.2 Principle of parallel concepts

Brunswik also recognize that both the cognitive system and the task system must be described in terms of the same kinds of concepts. He further claimed that these concepts should be statistical (Brehmer, 1988). As the relations between proximal cues and distal variables in perception, or relations between the proximal behaviours and goals in action, are at best probabilistic. He spoke such probabilistic relationships as

ecological validity specific to the task system and **functional or utilization validity** specific to the cognitive system, respectively. Ecological validity was defined as the correlation between a proximal cue and distal criterion and utilization validity was defined, in parallel fashion, as the correlation between a proximal cue and the organism's functional response (Brunswik, 1940; 1952). These two validities were then unified in the Lens Model Equation (see next section) for comprehensive analysis.

Meanwhile, the proximal cues are themselves interrelated, thus introducing redundancy (or intra-ecological correlations) into the environment. This led the phenomenon that successful achievement could occur even if the organism emphasized different cues on successive occasions as long as the cues themselves were intercorrelated. He labelled this process as **vicarious mediation**. The stronger the correlations, the greater the overlap in the information value among the cues. This redundancy permits the organism to make functional trade-offs in emphasis among the cues when drawing inferences about the distal criterion – the process of parallel concept, **vicarious functioning**. In other words, distal variables can be attained vicariously through proximal variables, and proximal variables can themselves be used vicariously through other proximal variables. It is implicated by Brunswik as a key reason why systematic experimental designs with their strong emphasis on controlling such correlations by forcing them to zero, failed to yield generalizable indications of behaviour (Cooksey, 1996). It was assumed that organisms learn the ecological validity of cues and their intercorrelations through experience.

2.4.1.3 The Lens Model

Brunswik created the Lens Model as a device for representing how the various concepts involved in probabilistic functionalism could be summarized.

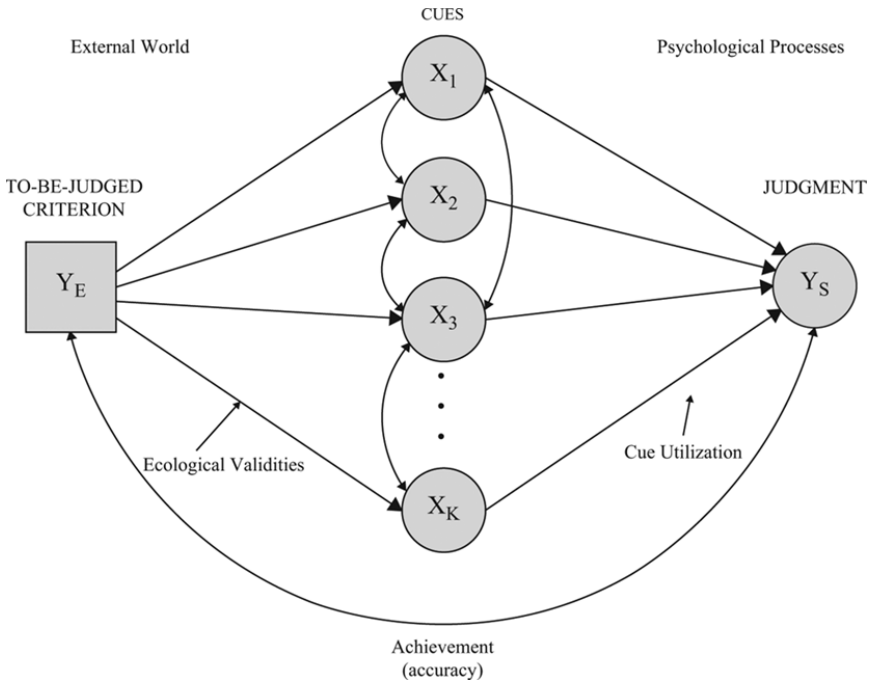


Figure 2-2, Adapted Lens Model. From "The conceptual framework of psychology" (p.20), E.Brunswik, 1952, Chicago, IL: University of Chicago Press. Copyright 1952 by University of Chicago Press.

2.4.1.4 Representative design of experiments

A natural methodological consequence of the above theory of probabilistic functionalism was the concept of representative design. It was argued that since the primary behavioural unit concerned distal-proximal-central parallel layers of reference, equal attention must be focused on obtaining representative samples of events or objects within the ecology, as well as on representative samples of organism. Any statistical analysis must then not only support inferences with respect to organisms, but also to situations and conditions within the ecology. In this view, any abstract, context-free systematic design, and policy of isolating and controlling selected variables, destroys the naturally existing causal texture of the environment to which an organism has adapted (Brunswik, 1944).

2.4.2 Social Judgment Theory: Hammond's further development of Probabilistic Functionalism

Brunswik's main concern in developing his system was with the problem of perception. SJT goes beyond Brunswik's original conception in that it focus on human judgment in his social environment. It could be characterized by four varieties of the lens model.

The single-system design considers only the judgment process itself. The criterion variable of the ecological side is either unavailable or is of no interest and thus the system represents only the subject side of lens model.

In the double-system design, the parameters of both the ecological side and subject side of the lens model are known. It has been used to study judgment accuracy (Hammond, 1955), multiple-cue probability learning (Hammond & Summers, 1965; Klayman, 1988), and cognitive feedback (Balzer et al., 1989; F.J. Todd & Hammond, 1965).

The triple-system is an expansion of the lens model. It involves a task situation and two DMs making use of the same probabilistic cues. This has enabled the study of interpersonal learning (Earle, 1973; Hammond, Wilkins, & F.J. Todd, 1966), and interpersonal conflict (B. Brehmer, 1976; Hammond, 1965; 1973; Hammond, F.J. Todd, Wilkins, & Mitchell, 1966).

Finally, the n-system involves more than two DMs and may or may not include an outcome criterion of task system. It therefore, enables the study of group judgment (Rohrbaugh, 1988).

For details of these designs and correspondent applications, see Cooskey (1996) and Hammond (2001).

2.4.2.1 Substantive situational sampling and formal situational sampling

To facilitate the investigation, SJT further goes beyond Brunswik's original proposal on representative design of experiments by differentiating the concepts between substantive situational sampling and formal situational sampling (Hammond, 1966b).

The former focuses on sampling the substance of an ecology so that it retains its realistic content and feel from the organism's point of view, and is analogous to Brunswik's original definition of representative design. The latter, on the other hand, focuses on the formal properties of the task (i.e., number of cues, their values, distributions, intercorrelations, and ecological validities), irrespective of its content. The formal properties define the universe of stimulus (or situation) populations, and would ideally be determined by an in-depth or understanding of the natural ecology so that non arbitrary values may be sampled. In this way, it permits the construction and presentation of stimuli that are formally representative of the natural stimulus population, and the practical difficulties of representative design (Dhami, Hertwig, & Hoffrage, 2004) could be overcome to some extent. Although the formal situational sampling still faced the difficulty in defining sampling frame (Brehmer, 1979), Hammond (1972) concluded that it is "clearly feasible", and he advocated that until technological advances allowed substantive situational sampling, researchers should use formal situational sampling (Hammond, 1966b; see also Hammond et al., 1987, as an illustrative example in practice).

2.4.2.2 Characteristics in measurement

Finally, to be compared with MCDM in measurement, two essential characteristics of SJT are worth to be addressed here.

Firstly, the probabilism that forms the major point of departure for SJT offers a conceptual link to the probabilism of MCDM in general and ER in particular. A major difference, however, is that probabilism of the approaches in MCDM demands explicit procedures for inquiring about a subject's probabilities regarding the occurrence of various conditions, or outcomes. SJT obtains a subject's probabilities through observing its judgments. In other words, while the MCDM approaches require that the subjects directly "measure" their own uncertainty, SJT measures the subjects' uncertainty in terms of their performance on judgment tasks, such as judgment consistency, interpersonal conflict (Hammond et al., 1980). Direct estimations from subjects

themselves were considered as problematic and unreliable in SJT. These subjective estimations were mostly compared with the probabilistic behaviour captured from observation, to reveal the subjects' self-insight into their own policies.

Secondly, subjects in MCDM are considered merely replicates of each other and hence, are evaluated on average levels of performance, consistent with a nomothetic orientation toward seeking generality and lawfulness. However, SJT argues the uniqueness of each organism as it engaged in functional behaviour within the context of a particular ecology. It is then that each subject is supposed to be individually examined and statistically tested before attempting to generalize behavioural trends, which is an idiographic orientated analysis method in contrast to compared the nomothetic analysis (Hammond, et al., 1980).

2.4.3. Judgment Analysis: a technique used by SJT researchers

Researchers in SJT employ the method of Judgment Analysis (JA) in their investigations. JA (Christal, 1963), also known as "policy capturing" (Dudycha, 1970), is a research method that using the lens model to capture and describe an individual's judgment policy. Generally speaking, the method involves asking individuals to make decisions on a set of cases that might be either real or hypothetical and which comprise a combination of cues (information). Each individual's judgment policy is then inferred from his or her behaviour, traditionally through the use of multiple linear regression analysis. This method attempts to avoid the pitfalls of verbal protocol methods in the context of hypothesis testing (Biggs et al., 1993; Nisbett & Wilson, 1977). There have been hundreds of studies which have employed JA of some variety in capturing judgment of professionals in a variety of domain. Some basic investigation guidelines of the method have been explicated in Cooksey (1996). The reporting of results of a JA study would generally involve the following 10 parts, which indicates the general procedures of the JA method (Stewart, 1988):

i. Descriptions of the DMs and how they were selected.
ii. Descriptions of the cues and how they were chosen.

iii. Definition of the judgment and the response scale.

iv. Distributions of the cue values and how they were generated.

v. Intercorrelations among the cues.

vi. Reliabilities of the DMs, computed by correlating their judgments over repeated cases.

vii. Multiple correlations for linear models and for any nonlinear models investigated.

viii. Standard errors of the regression coefficients.

ix. Relative weights and the method used to calculate them.

x. Function forms, if nonlinear function forms were used.

Representative experiment design in JA

The critical dimension of JA method which distinguishes it from nearly all other formal methods in the behavioural decision science is its insistence upon applying the principle of representative design to guide the structure of specific investigations. In addition to those formal statistical properties to be considered in the JA task design (e.g., cue distributions, intercorrelations in formal situational sampling), at least another two critical issues are addressed in the literature. Cooksey (1996) termed them as task familiarity and task congruence. Task familiarity reflects the extent to which a judge has prior experience, within a particular ecology, in making judgments of the type required by the judgment task. Task congruence, on the other hand, encompasses considerations of cue labelling, measurement and representation.

The term “policy capturing” implies that the subjects have some policy that could be captured. If a person has no experience with the kind of judgments required, there cannot very well be any policy to capture. JA, however, has often been used to study judgment involving graduate students on unfamiliar task. Such studies are typically designed to investigate individual learning by involving feedback in the experiment process. Brehmer (1987) has termed this type of application as “policy construction” rather than policy capturing. Statistically, the lack of an experiential basis for the

judgments can lead to unstable results that are highly sensitive to seemingly inconsequential aspects of the study design (Stewart & Ely, 1984).

In terms of task congruence, ideally, JA would be expected to present cue information with the measurement units which would be naturally observed within the ecology (for example, using actual kilogram to account for the weight of an object), a concept termed as concrete representation. In contrast however, historically, investigators employing regression technique tend to code the cues into scale for the benefit of data analysis. Such abstract representations of cue information rise critical issue of whether the coding scheme and the subsequent presentation format reflect the true perceptual process of the DMs. Even this agreement is hold, it still face the issues such that coding may move the judgment process towards the analytic end of the cognitive continuum so that artificially inflating consistency and insight (Hammond 1996a).

The two dimensions elaborated above depict the two broader aspects of the more general problem of representative design. They integrate much of the essential meaning of Hammond's formal and substantive task characteristics at a higher conceptual level. Cooksey (1996) broke down these two dimensions into four categories to indicate the level of representativeness of JA study. Generally, the task context described in Cell A of Table 2-5 (Familiar task / Concrete presentation) yields the greatest level of representative design opportunity and Cell D (Unfamiliar task / Abstract presentation) yields the lowest level of representative design opportunity. There can be strong differences in the statistical characteristics of judgment undertaken in different cell contexts (Gorman et al., 1978).

Table 2-5, Broad categorization of Judgment Analysis research contexts using the dimensions of task familiarity and task congruence (Cooksey, 1996, p.89)

		Task Congruence	
		Concrete	Abstract
Task Familiarity	Familiar	• Judge has made these sorts of judgments before in real life • Task information is represented and/or obtained in original units of measurement encountered in the ecology **A**	• Judge has made these sorts of judgments before in real life • Task information is represented and/or obtained using abstract conceptual variables **B**
	Unfamiliar	• Judge has seldom, if ever, made these sorts of judgments before in real life • Task information is represented and/or obtained in original units of measurement encountered in the ecology **C**	• Judge has seldom, if ever, made these sorts of judgments before in real life • Task information is represented and/or obtained using abstract conceptual variables **D**

To be noted at the end, close adherence to the principle of representative design is often difficult to achieve in practice. Actually, few studies reporting JA have employed the Cell A context (Cooksey, 1996), and orthogonal cues are applied in most of JA investigations (Brehmer & Brehmer, 1988). It is generally recommended that the investigators should consider utilizing those strategies and techniques which will afford the best chance of arguing for the representativeness of his or her research design within the particular context constraints. Trade-offs are inevitable; but they can be made sensibly in many cases, and may be circumvented in others (Cooksey, 1996).

2.5 Heuristic decision making: Adaptive Toolbox (AT) program

2.5.1 Origin

The Adaptive Toolbox (AT) program, which systematically investigates when, how, and why simple cognitive strategies (simple heuristics) can help people make good judgments and decisions (Marewski, Gaissmaier, & Gigerenzer, 2010), grew out of a number of interesting findings in literature of human judgment during early 1990s.

Firstly, according to Gigerenzer, Hoffrage, & Kleinbolting (1991)'s "probabilistic mental models" theory (PMM), the manner in which questions are sampled is relevant for the demonstration of the overconfidence effect – a prominent cognitive illusion catalogued in the literature of Heuristics & Biases program (Kahneman, Slovic, & Tversky, 1982). They suggested through empirical experiments that the overconfidence effect may stem from the fact that researchers failed to sample general-knowledge questions randomly, but tended to over-represent items in which cue-based inferences would lead to wrong choices. It is therefore, under specific designed conditions, the effect could appear, disappear, or invert. Such impact of representative design on overconfidence effect was further supported in a review of 130 overconfidence data sets (Juslin et al., 2000), where consistent findings were identified that the effect was on average pronounced with selected item samples and close to zero with representative item samples. In other words, the environment matters in the study design of human decision making.

Secondly, in their work on the "adaptive decision making", Payne, Bettman, and Johnson (1993) suggested that, as a result of prior experiences and training, a decision maker generally had more than one strategy available to solve a decision problem of any complexity, and chose among them depending on their costs and accuracy trade-off given constraints of the task environment such as time pressure. The strategy selection under this trade-off was generally adaptive and intelligent, if not optimal. The idea is that "human beings learn how to handle (and perhaps mishandle) risky decisions of various kinds not by introspecting about axioms but as a result of the piecemeal experience of having to deal with different types of problems encountered at different

times" (Payne et al., 1993, p.17). In other words, the idea of single unified model of individual decision making under risk and uncertainty (single general-purpose tool) was in doubt. Rather the same individual might use different models to deal with different problems.

Thirdly, Gigerenzer and Goldstein (1996), a further step forwardly, through computer simulation, demonstrated that their fast and frugal heuristics (mostly lexicographic rules) could match or outperform the comprehensive optimal way (e.g., weighted additive rule) in inferential accuracy. They further theorized it that as long as a decision rule exploited the structure of environment constraints, trade-off between the accuracy and cost was not necessary. Less could be more and better (less-is-more effect).

2.5.2 Main topics

It was therefore in the late 1990s, the AT, as a major research program on people's heuristic decision making, was launched, marked by the publications "simple heuristics that make us smart" (Gigerenzer et al., 1999) and "Bounded rationality: The adaptive toolbox" (Gigerenzer & Selten, 2001). Drawing on Simon's bounded rationality (1955; 1972), the program worked on "(1) proposing computational models of candidate heuristics that are realistically based on human competences, and testing whether they work via simulation; (2) mathematically analyzing when and how the heuristics work with particular environmental structures; and (3) experimentally testing when people use these heuristics" (Todd & Gigerenzer, 2003, p.148).

Computational models of heuristic

Heuristic, according to Wikipedia, refers to experience-based techniques for problem solving, learning, and discovery that give a solution which is not guaranteed to be optimal. Where the exhaustive search is impractical, heuristic methods are used to speed up the process of finding a satisfactory solution via mental shortcuts to ease the cognitive load of making a decision. Examples of this method include using a rule of thumb, an educated guess, an intuitive judgment, stereotyping, or common sense. It has been mostly studied by behavioural biologists (Hutchinson & Gigerenzer, 2005),

researchers in Artificial Intelligence (Newell & Simon, 1976) and cognitive psychology (Kahneman, 2003a; Kahneman & Tversky, 1982; Gigerenzer et al., 1999).

In study of human judgment and decision making, heuristics are most known as unreliable decision strategies that are often lead to errors (Kahneman & Tversky, 1982). The resulting Heuristics & Biases (HB) program has had immense influence, contributing to the emergence of behavioural economics and behavioural law and economics (Kahneman, 2003b). It is however, the heuristics generalized in HB are defined as one-word labels, and are criticized in their vagueness in explaining phenomena. It was argued by Gigerenzer (1996) that those one-word labels, like representativeness, availability and anchoring (Tversky & Kahneman, 1974), explain too little and too much: too little, as the underlying processes are left unspecified, and too much, because, with sufficient imagination, one of them could be fit to too much empirical result, even the opposite ones (e.g., Ayton & Fischer, 2004). Once the heuristic is formalized, conclusion could be changed (e.g., Sedlmeier et al., 1998). "The basic flaw in the original research was to introduce a label after data were observed, rather than formulate a heuristic as a computational model that allows predictions to be expressed in a testable way" (Gigerenzer & Brighton, 2009, p.127).

Accordingly, heuristics in AT are conceptualized in a step-by-step and computational manner through hypothesized building blocks to describe people's decision making process. It is therefore, highly transparent and open to scrutiny. Typical building blocks of a heuristic would be like the search rule for guiding a search for information, the stopping rule that terminates the search, and the decision rule that make inference based on the results of the search.

For example, search through information (termed as cue in the following) may be random, ordered according to their relative importance, or based on the memory of the last cue used. Search may be stopped as soon as the first cue that favours an alternative in a choice task is found, or be stopped after all the available cues have been considered.

Finally, a decision can be made on one cue only, or be made by weighting and combining the impact of all the cues available.

According to the impact of cues on a decision, heuristics in AT could be identified as compensatory or non-compensatory. While compensatory heuristics permit trade-off between cues, i.e. every cue is considered and may influence the final decision, non-compensatory heuristics works on an opposed way, i.e. an unfavourable value in one cue cannot be offset by a favourable value in other cues.

According to the number of cues being looked up in a search, the frugality of the heuristics could be classified. For example, a heuristic that only looks up one cue would be more frugal than the heuristics that look up all the cues available.

According to the complexity of the computation involved, the speed of heuristics could be classified. For example, a compensatory heuristic that integrates all the cues but ignore their relative importance would be faster than a compensatory heuristic that integrates the cue values according to the correspondent cue weights.

In other words, by involving a search and stopping process, the heuristics formalized in AT could address the process dynamics of people's decision making (e.g., information search, shift in information emphasis or judgment policy), and diverse decision making behaviour could be precisely tested through the approach.

In the following, the take-the-best heuristic is depicted as an example of heuristics that have been developed in the AT literature.

Take-the-best (Gigerenzer & Goldstein, 1996) is the first, and perhaps the most well-known and researched heuristic in AT. It embodies another simpler heuristic, namely the recognition heuristic (Goldstein & Gigerenzer, 2002). Both these heuristics were designed to make decision in a two-alternative choice task, where the alternatives and the cues (information) about the alternatives are binary. These two heuristic are illustrated in Figure 2-3.

The first step in the decision process involves the recognition heuristic. This is considered the simplest of all the heuristics developed to date. It is called ignorance based decision making as it relies on a lack of knowledge of the alternatives, and only on recognition memory. Therefore, the recognition heuristic would not work if every alternative was recognized. In such circumstance, the process continues to the second step, namely a search through the cue values. Search through cues is ordered. The cues are ordered according to their ecological validities. The ecological validity of a cue is defined in terms of the proportion of correct inferences for that cue alone, when one alternative has a positive value and the other does not. Cues with higher ecological validities are better able to correctly predict a choice. The take-the-best heuristic retrieves, from memory, the values of the first rank ordered cue for the two alternatives. If the cue discriminates (i.e., for a binary cue, it has a positive value for one alternative and a negative or unknown value for the other), then the search for values of other cues is stopped and alternative with the positive cue value is chosen. If not, the search process continues. If after all of the cues have been searched and no cue discriminates, an alternative is chosen randomly.

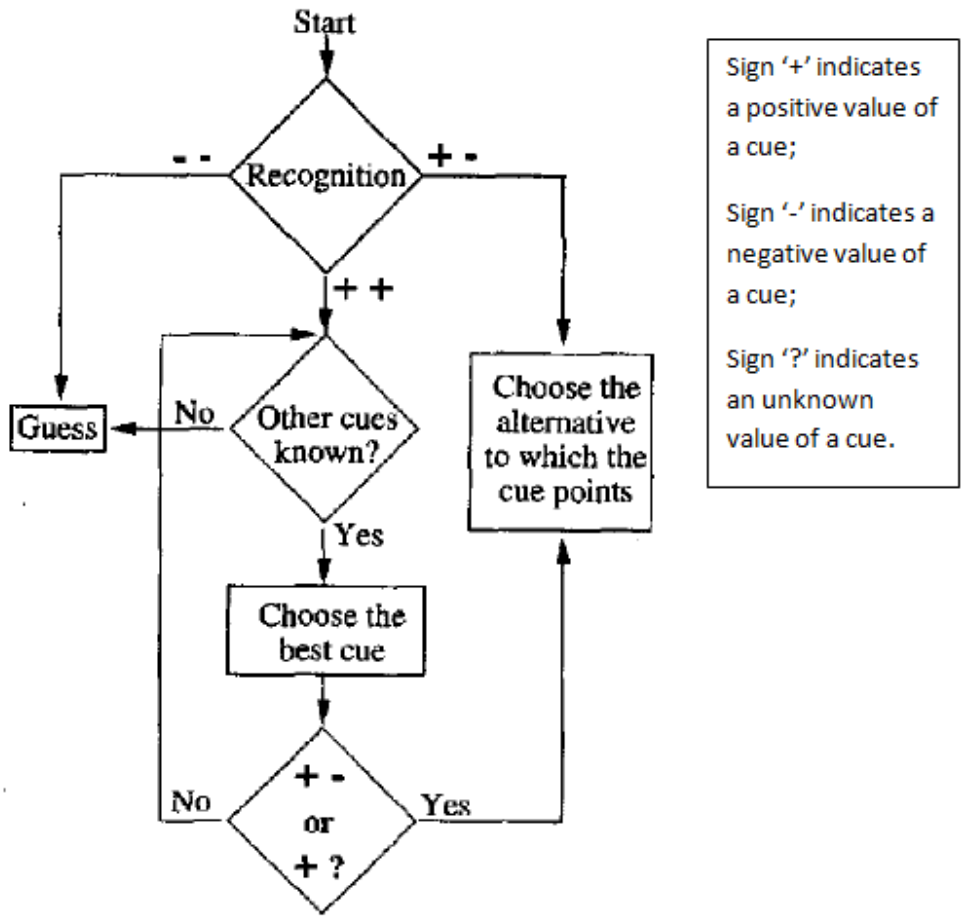


Figure 2-3, Flow diagram of take-the-best heuristic (Gigerenzer & Goldstein, 1996)

Environmental structure

The mathematical analysis of environmental structure is to figure out the rationality behind the performance of those computational models of heuristics. This is achieved by drawing on the assumption that evolution would seize upon informative environmental dependencies and exploit them with specific heuristics if they would give a decision-making organism an adaptive edge. From this standpoint, the repertoire of "adaptive toolbox" of heuristics (Gigerenzer & Selten, 2001) succeeds in its specificity, and the investigation focus was moved onto the fit between the simple heuristic and the particular environment it succeed. It is named as the new version of rationality: ecological rationality (Todd et al., 2011). A number of information structures have been

demonstrated of essential impact on the performance of particular heuristics, like the degree of uncertainty, the number of alternatives, and the size of the learning sample (Todd et al., 2011).

Empirical testing

In terms of the methodology to facilitate the identification of a most appropriate heuristic to describe a DM's decision making process, two positivistic approaches have been applied in literature.

(To be noted, research in this literature has normally used the term "decision making policy" to account for the decision making process applied by DMs. In the following, we will use these two terms interchangeably depending on the appropriateness of textual context.)

One is process-tracing approach (Juslin & Montgomery, 1999) to observe DMs' activities before they finally make their decisions. Here, the fundamental belief is that cognitive processes should be studied by collecting data during the decision process as often as possible (Svenson, 1983). The approach is therefore, designed to directly assess what information is accessed to form a decision and the order in which the information is accessed. Sequences of information acquisition are monitored through the information board technique in which hidden information is presented in the form of an alternative-by-attribute (cue) matrix, and has to be actively uncovered by turning cards or opening information boxes on the computer screen (e.g., Newell, Weston, Shanks, 2003; Rieskamp & Hoffrage, 1999; Rieskamp & Otto, 2006). The data obtained are then interpreted on the basis of the assumption that the order of information search depends on the DM's decision making policy. For example, a compensatory decision making policy is believed to induce an alternative-wise search pattern, whereas a non-compensatory decision making policy is believed to induce an attribute-wise search pattern. It is however, the approach has been criticized for involving artificial methods to assess what information is searched for, not precise enough to identify a particular

strategy the DM applied, and unjustified assumptions of information search patterns (e.g., Broder, 2000b; Rieskamp & Hoffrage, 1999).

Another approach is outcome oriented, usually named as "structural modelling" (Harte & Koele, 2001). The approach works on the structural relationships between the information (the input) and final decision (the output). One prominent representative of this approach is the Judgment Analysis (JA) method (see Section 2.4.3). Generally speaking, the method involves asking individuals to make decisions on a set of cases that might be either real or hypothetical and which comprise a combination of cues (information). Each individual's judgment policy is then inferred from his or her decisions on the cases, traditionally through the use of multiple linear regression analysis. Some variants of JA has been applied to empirically test people's use of heuristics in their working conditions (e.g., Dhami, 2003; Dhami & Ayton, 2001; Dhami & Harries, 2001; Garcia-Retamero & Dhami, 2009; Smith & Gilhooly, 2006). It is however, no investigation has been conducted in working conditions involving dynamic shifting situations. In fact, the issue of how the method is to best address dynamic decision making task ecologies has been identified as "the most pressing problem" of JA for future research (Cooksey, 1996, p.323). "This becomes a critical problem if Judgment Analysts hope to study decision making as it unfolds in naturalistic contexts" (p.323). The latest evidence of people's use of heuristics has been brought together in Gigerenzer, Hertwig, & Pachur (2011).

2.5.3 Characteristics of AT

The AT, as its origination suggested, holds a strong Brunswikian tone, such as the importance of task environment and representative design of experiment. At the same time, it shares a lot interesting arguments held in NDM. To enable parallel comparison, we characterize the program here through five dimensions suggested in section 2.3.

Proficient decision makers

Similar to the NDM, the AT shares the same contention of bounded rationality that addresses the constraints of environment rather than blaming the incapability of human

reasoning. A decision is ecologically rational to the degree that it is adapted to the structure of an environment. The coherence criteria are, therefore, argued to be replaced by the correspondence criteria which focus on the adaptive fit of the internal cognitive structure towards the external environmental structure, a notion of ecological rationality (Todd et al., 2011). Under this notion, human decisions are seen as inductive inference adapted to its past experience/environment (Cosmides & Tooby, 2006). In this sense, the notion of capable decision maker arises. We see a highlight of "expertise" like NDM, but in laypeople. However, methodologically, other than utilizing the expertise straightforwardly as did in NDM, the AT program like other laboratory-based cognitive psychology, sees task experience as a confounding factor to be avoided for the benefit of experimental control. Instead, studies in AT normally would plan a learning paradigm to approximate the learning process (domain experience) proposed by Gigerenzer, Hoffrage, & Kleinbolting (1991) in natural environments.

Process orientation

AT researchers hold the similar contention towards process orientation as NDM, but formalize the contention at a different level. They argue that the pure input-output orientation dispenses with the need to specify how, in mechanistic terms, data are processed to yield behavior, leaving the human decision as a black box (Brighton & Gigerenzer, 2008). In contrast, they suggest an algorithmic level model to provide a mechanistic account of how the human cognition processes data in order to address the problem. Such a model is algorithmic in the sense that it describes the steps required to transform inputs to outputs such that these steps could plausibly be implemented on some form of computing machinery. In their formalization of adaptive toolbox (Gigerenzer & Selten, 2001), a computational model of heuristic would at least involve search rules for guiding a search for alternative or information, stopping rules for terminating the search at some point, and decision rules for making inference based on the results of the search. The involving of the search rule and stopping rule address the essential characteristic of real world situations where information or alternatives must be actively sought. But to be noted, their decomposing or reducing the cognitive

phenomena into hypothetical building blocks is inherently different from the NDM's work on qualitative explaining (Klein et al., 2003), which will be further addressed in the following.

Situation-action matching decision rule

AT' notion of matching decision rules agree with the NDM on: (1) the active searching process usually leads to options being considered sequentially; (2) options are selected or rejected against a standard, or say an aspiration level, rather than their relative merits or complex optimal-stopping calculations (Todd & Gigerenzer, 2001). To demonstrate above parallel, a notion of nested structure is worth to mention here. In the light of sequential process as well as human's adaptive learning as biological entities (Pinker 1997; Wimsatt, 2007), heuristics in AT are assumed to be able to be combined by nesting one inside another. In this way, an alternative heuristic could be seen as the natural response to the additional circumstance, information, or knowledge, on top of the existing one. As an example, the recognition heuristic (Goldstein & Gigerenzer, 2002) works on the basis of recognition memory. But when the memory is incapable or improper to make inference and the subject is more knowledgeable and can recall other facts of the object, knowledge-based heuristics such as take-the-best (Czerlinski et al., 1999) would be applied. In this sense, the recognition heuristic serves as the first step of take-the-best. Interestingly, similar nested structure could be identified in NDM. In the prominent RPD model (Klein 1993, 1998), experts' decision making is characterized into three variations, among which the simplest variation – recognize the situation and implement the obvious response, serves as the first step of the more advanced and complicated level 2 and level 3 variations.

The disagreement on the situation-action matching decision rule, however, lies in the process of matching. In against to the NDM's proposal on subjective pattern matching and informal reasoning, AT calls for more specific formal models. Its initial attempt could be identified on two perspectives. From the perspective of organizational level, AT theorizing the collection of heuristics through building blocks which is common among

each other. This allows “reducing the larger number of heuristics to a smaller number of components, similar to how the number of chemical elements in the periodic table is built from a small number of particles” (Gigerenzer & Gaissmaier, 2011, p.456). From the perspective of rule level, AT has explored a number of possibilities in the context of searching for cues, including checking them in order of their past record of success, or past use, or randomly (e.g., Martignon & Hoffrage, 2002; Newell et al., 2004).

Context-bound informal modelling

Researchers in AT, by drawing the bias-variance dilemma (Geman et al., 1992), agree with the notion that the general-purpose inductive inference in natural environment is unachievable to learn or impossible to be actually applied (Brighton & Gigerenzer, 2008). Rather, people are relying on an adaptive toolbox of biased, specialized heuristics (Gigerenzer & Selten, 2001). However, they keep a reservation on this context specificity through introducing a notion of generality and specificity trade-off. The decision heuristics formalized in AT maintain their level of specificity at the moderate level (neither too specific nor too general) by their very simplicity (constrained building blocks and free parameters), which allows them to be robust in the face of environmental change and enables them to generalize well to new, but not every, situation, a bounded context specificity indeed.

In terms of formalization, while NDM researchers emphasize the difficulty and even inability of formalization through the present state of art, researchers in AT, like other members from mainstream cognitive psychology, take it as indispensable for scientific progress (Gigerenzer & Gaissmaier, 2011). Holding a strong positivist position, they emphasize the predictive understanding of phenomena, and highlight the importance of hypothesis testing. They see subjective criteria, vague terms, as well as after-the-fact explanation as an unacceptable step backwards of research (Gigerenzer & Gaissmaier, 2011), and a misleading opposition (Todd & Gigerenzer, 2001).

Empirical-based prescription

Researchers in AT share the same attitude towards prescription as the NDM researchers did. In addition, they take another step forward. They argued that norms of reasoning could be derived from descriptive statements of empirical science. And through the demonstration of their essential empirical evidence of less-is-more effect (Gigerenzer & Goldstein, 1996), the justification of general-purpose rule of good reasoning was questioned. In their view, optimization and heuristic strategies should be treated equally, as any of them can be justified only relative to the structure of the environment.

The difference of the two programs, however, lies in the consequent goal of empirical-based prescription. While the prescription in NDM, in a qualitative way, aims to improve feasible decision makers' characteristic modes of making decisions (Lipshitz, Klein, Orasanu, & Salas, 2001a); the prescription in AT works on quantitatively suggesting the match between different styles of decision making policy and their correspondent environment (Gigerenzer & Sturm, 2012).

2.6 Conclusion

This chapter has reviewed the literature of MCDM, ER approach, NDM movement, SJT, and AT program. The review indicates that MCDM, as a family of methods in operations research, provides a formal tool for supporting decision makers to tackle with decision making problem through multiple, and sometimes hierarchical, criteria. This has been further demonstrated in the review of ER approach. The approach, as one of the most recent development in MCDM literature, is capable of dealing with MCDM problems with uncertainties and hybrid nature of information.

The review also indicates that there has been a movement of decision making study (NDM) that attempts to understand how people make decisions in real-world contexts that are meaningful and familiar to them. The real-world contexts are mostly critical, characterized by the factors like ill-structured problems; uncertain dynamic

environments; shifting, ill-defined or competing goals; action/feedback loops; time stress; high stakes; multiple players, and organizational goals and norms. Accordingly, the methodology applied and research outputs produced by researchers in NDM are quite different from the normative decision analysis. These differences are characterized by proficient decision makers, situation-action matching decision rules, context-bound informal modelling, process orientation, and empirical-based prescription.

Following the contentions addressed in NDM, the review then explores another two literatures of decision making study, SJT and AT program. The SJT provides a theoretical framework and a methodology alternative of studying people's decision making under environmental constraints in a more formal and quantitative way. AT program inherits the theoretical and methodological contentions of SJT and further highlights the importance of formal models of heuristics in such study.

These literatures together build the theoretical and methodology basis of this thesis, and will be referenced and utilized in the elaboration of our methodologies development and case study applications.

Chapter 3 : Research Methodology

A research methodology is the general research strategy that outlines the way in which the research questions are to be answered and, among other things, identifies the methods to be used in it. These methods, described in the methodology, define the means or modes of data collection and how a specific result is to be produced.

As briefed in chapter one, multiple methods have been utilized complementarily to accomplish the investigation of the research in the thesis. They together facilitate two novel methods being developed and applied in our research: Evidence-based Trade-Off (EBTO) and Judgment Analysis with Heuristic Modelling (JA-HM).

This chapter provides a detailed account of these two methods regarding the literature gaps they have filled, their implementation procedures, and those techniques involved in their data collection and analysis process.

3.1 Evidence-based Trade-Off (EBTO)

In the real world, people are often engaged in decision making task under imprecise goal and uncertain information.

Here a simple example is given. Suppose you are a graduate student and you are deciding your graduate job. Obviously, you want a “good” job. This is your general goal. But what does a “good” job stand for? It is probably not so precise. You may think a “good” job must have a good monetary compensation. You may also want the job to involve a great training package for your future development and flexible working time for work-life balance. Then, you get at least two objectives or criteria that stand for a “good” job: maximizing the monetary compensation, and optimizing nature of work. It is however, that the two objectives are most of time conflicting between each other. Higher salary can often be associated with working at high pace and regularly overtime. The conflict indicates that at some point the decision maker will be faced with the proposition that further achievement on one objective can only be accomplished at the expense of achievement on the other. What make things more complex is that, as a

graduate student, you probably won't get complete and accurate information of any particular job. You may ask the recruiters about the training package, but you probably can only subjectively judge by yourself of how appropriate the training package is for you. You may have no or partial information about the work style of the job, which is common. Most of the time you would know what exact the salary is. But it is quantitative, and sometimes involves deterministic basic salary and probabilistic bonus. And how could you trade-off it with your subjective judgment of the quality of training package.

The above scenario illustrates the complexity of decision problem under the imprecise goal and uncertain information. The goal in scenario is imprecise and is mostly driven by multiple conflicted objectives. The information in scenario is uncertain, characterized by the subjective judgment, and missing or incomplete data. Furthermore, the information is hybrid in nature involving incommensurable units, mixture of qualitative and quantitative, deterministic and probabilistic data.

Our research proposes a method, on the basis of the ER algorithm (is described in Section 2.2.2), for human decision making task under imprecise goal and uncertainty. The method is named as Evidence-based Trade-Off (EBTO). The EBTO provides a novel framework to aid people's decision making under uncertainty and imprecise goal. Under the framework, the imprecise goal is modelled through an analytical structure, and is independent of the task requirement; the task requirement is specified by the trade-off strategy among criteria of the analytical structure through an importance weighting process, and is subject to the requirement change of a particular decision making task; the evidence available, that could contribute to the evaluation of general performance of the decision alternatives, is formulated with belief structures which are capable of capturing various format of uncertainties that arise from the absence of data, incomplete information and subjective judgments.

The framework of EBTO is illustrated in Figure 3-1 below.

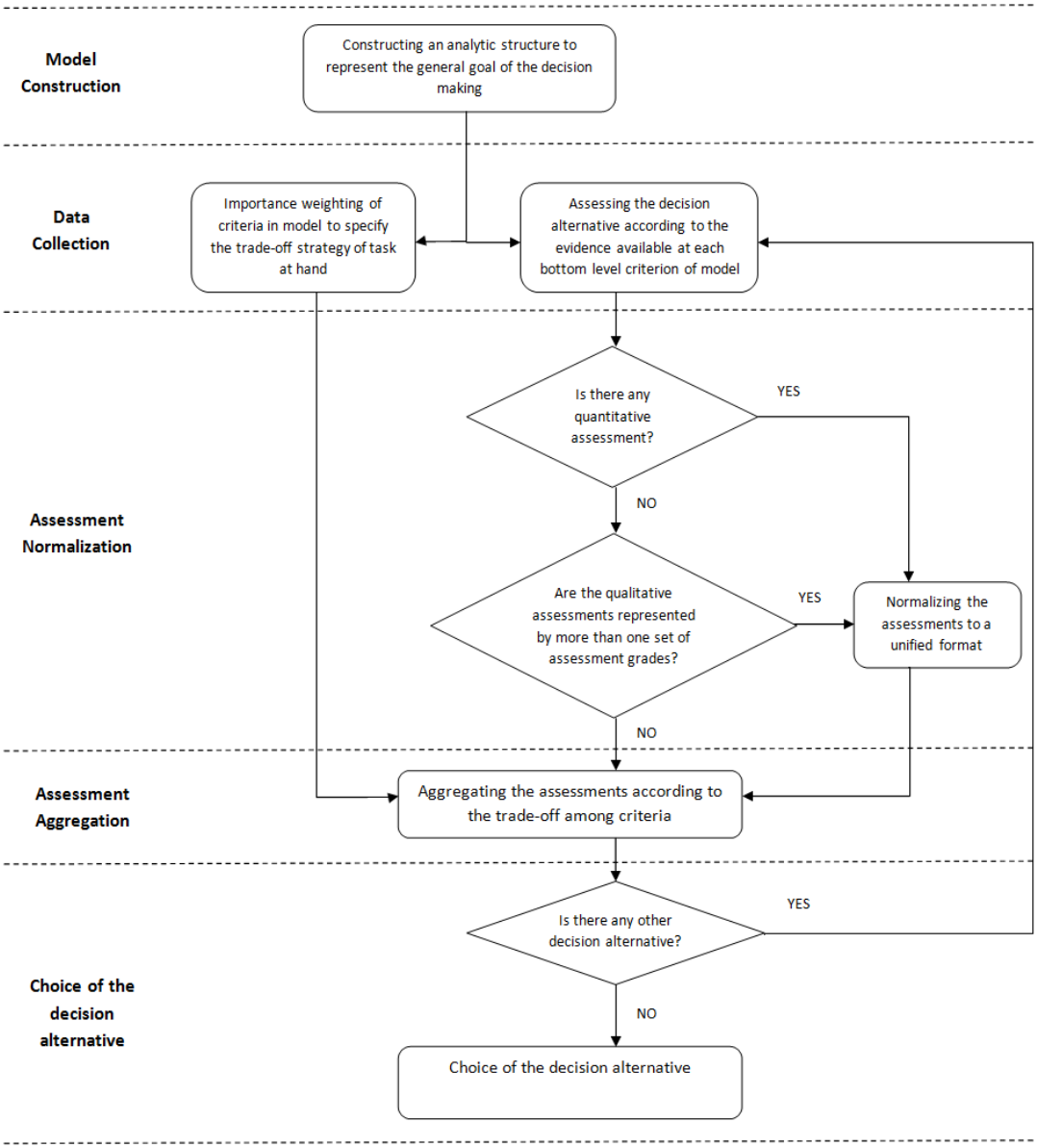


Figure 3-1, The framework of EBTO method

3.1.1 Model construction

The method starts from developing an overall analytic structure to represent the general goal of the decision making task. The analytic structure is generally termed as Model. The development of the model is accomplished by first identifying criteria that would contribute to the evaluation of a decision alternative. Each criterion would reflect one objective of the general goal and may be described by and evaluated through its associated sub- and sub-sub criteria. This leads to a hierarchical structure, in which each

of bottom level criteria is supposed to be directly measurable by means of either numerical values or subjective judgments. This modelling process could be conducted through multiple methods, including examination of the relevant literature, interviews with Subject Matter Experts (SMEs), and field observation.

There are five desirable properties for any set of criteria identified (Keeney & Raiffa, 1993).

Firstly, the set of criteria must be complete. This condition would be satisfied when the lowest-level criteria in a hierarchy include all areas of concern in the problem at hand.

Secondly, the set of criteria must be operational. This implies a number of ideas. Basically, since the modelling process itself is to help the decision maker assess and choose alternatives, the criteria must be meaningful to the decision maker so that he can understand the implication of the alternatives. In cases where the partial purpose of the assessment is to advocate a recommendation, the criteria should also facilitate explanations to others, like senior officers.

Thirdly, the set of criteria should be decomposable, so that aspects of the assessment process can be simplified by breaking it down into parts. By this we mean that any assessment task could be broken down into parts of smaller dimensionality. For instance, if a problem involves a set of seven criteria, it might be possible to break the assessment into two parts, one involving three criteria and one involving four.

Fourthly, the set of criteria should be defined to avoid double counting of any impact. That is non-redundancy. One common way that redundancies creep into a set of criteria is when criteria are associated with both means and ends objectives where means-ends relationships of the objectives are not clearly indicated. Another way in which redundancies enter a set of criteria is when some criteria represent variables that are inputs to a system and others represent variables that are outputs.

Finally, the set of criteria is desired to be minimal, so that the problem dimension is kept as small as possible. Each time a criterion is subdivided, possibilities of excluding

important concerns occur. In addition, the computational complexity increases greatly as the number of problem dimension increases.

Once the assessment model has been developed, the next step is to manipulate the model to identify an optimal decision alternative of the problem. This involves three sequential steps. They are data collection, data normalization, and data aggregation.

3.1.2 Data collection

The data collection involves two parts of work. One is to assign the relative importance to each criterion. The other is to assess the performance of decision alternative at each bottom level criterion. These two parts of data collection process are driven by decision making analysis at two dimensions: 1) the task requirement of decision making; 2) and the evidence available in evaluating the performance of decision alternatives.

3.1.2.1 Importance weighting of criteria

The task requirement of decision making, under the general goal, is not necessary constant and stable across different cases.

For example, in deciding a new product design for a customer, the general goal and analytical structure is objective and independent of the subjective preference. A good product design would always try to maximize the quality, reliability, safety, maintainability, serviceability, manufacturability, etc. and minimize the cost. It is however, different costumers may have different preference under the same budget. Some may want to guarantee the reliability and safety. Some may emphasize the manufacturability. The decision making from each costumer's perspective stands for an individual decision making task. Its task requirement could be specified by a trade-off strategy among criteria of the analytical structure. The importance weighting of criteria serves as a process of transforming such trade-off strategy into the model for holistic analysis under the general goal of decision making.

In assigning the weights, multiple of methods can be used, such as simple direct rating by SMEs, or more elaborate methods based on the pair-wise comparison technique (Saaty 1994).

3.1.2.2 Assessment at the bottom level criteria

The assessment at the bottom level criteria, on the other hand, concerns with the performance of each decision alternative. As explained in the model construction section, the bottom level criteria of the model are supposed to be directly measurable by means of either numerical values or subjective judgments. Through the assessment at the bottom level criteria, the detailed performance of a decision alternative could be specified.

The assessment is conducted on the basis of evidence available. The evidence in real world is normally presented in three natures: precise numbers, probabilistic occurrences (random variables), and qualitative accounts (subjective judgments from the DMs).

Precise numbers are single or exact values. For example, the “starting salary” of a job could be a fixed number, like 20,000 pounds per year.

The present method employs a belief structure to account for those random variables and subjective judgments. For example, the “bonus” of a job could be a random variable, like “50% chance of 2,000 pounds per year, 30% chance of 3,000 pounds per year, and 20% chance of 5000 pounds per year”. In this case, this assessment of “bonus” could be described with a belief structure as {(2000, 50%),(3000, 30%),(5000,20%)}.

Take another example, the “quality of training package” of a job could be subjectively judged to be “Excellent” with 30% of belief degree, “Good” with 70% of belief degree. This subjective judgment could be described with belief structure as {(Excellent,30%), (Good,70%)}. The terms “Excellent” and “Good” are assessment grades that indicate the different levels of qualitative preference.

In addition, the sum of the degree of beliefs assigned to each criterion could be between 0 and 1. If sum = 1, it indicates a complete assessment that the DM is 100%

sure about the judgment, or a random variable that is fully informed. If sum < 1, it is an incomplete assessment which reveal that the DM is not fully confident about a judgment or a random variable due to a lack of evidence available. The belief structure accepts the raw evidence as it is. It provides a meaningful tool for the qualitative and incomplete assessment. In addition, any absence of data could also be described through belief structure as {(unknown,100%)}. The assessment model is therefore, capable of dealing with uncertainties that arise from the absence of data, incomplete information and subjective judgments, which are critical to real world problem.

3.1.3 Assessment normalization

For the assessment at the bottom level criteria, it is a natural application that a qualitative criterion may be assessed using a set of grades appropriate for this criterion but different from others. For example, criterion “quality of training package” of a job choice problem may be assessed with a set of grades from 1 to 5, while the criterion “flexibility of working time” may be better assessed with a set of more grades from 1 to 9.

In addition, precise numbers are, formally, different from the qualitative belief structure.

Therefore, an assessment normalization process is required to transform the raw assessments to a unified format for the benefit of following aggregation process of data.

For those qualitative assessments (subjective judgments), the normalization process refers to first defining a general set of grades and then transforming the sets of grades, which are utilized in each criterion, into this general set of grades, so that the transformed assessments are equivalent with regard to underlying utility and rational in terms of preserving the features of original assessments.

For those quantitative assessments (single number or random variable), the normalization process refers to first defining a general set of grades and then transforming the quantitative assessments into this general set of grades, so that the

transformed assessments are equivalent with regard to underlying utility and rational in terms of preserving the features of original assessments.

The transformation can be conducted using the DMs' knowledge and experience described as rules. For instance, the "starting salary" of a graduate job at 30,000 pounds per year may mean that the general preference of the job is "Top" as far as "starting salary" is concerned. Then an equivalence rule could be set up as "If starting salary is 30,000 pounds per year, the preference of the job is top".

The same transformation process can also be conducted through utility estimation. For instance, a DM may estimate u(good) = 0.8 for criterion "quality of training package", and estimate u(very preferred) = 0.8 for general preference of a job; then an evaluation grade "good" in "quality of training package" assessment is said to be equivalent to a grade "very preferred" in general preference of a job.

The formalized normalization process could be found in Section 2.2.4, where we reviewed the Yang (2001)'s original proposal on this process.

3.1.4 Assessment aggregation & choice of decision alternative

Following the data collection, the assessments at the bottom level criteria could be aggregated through a data aggregation algorithm according to the trade-off strategy among criteria (importance weighting of criteria). The aggregated output reflects the general preference of a decision alternative. The ER algorithm is applied in our method to serve the assessment aggregation process.

The formulation of the ER algorithm has been reviewed in Chapter Two. Briefly, the aggregation is conducted through a bottom-up approach, in which, belief structures of the lowest level criteria are aggregated as input for the second lowest level criteria, which is, in turn, aggregated to produce belief structures of higher level criteria. This recursive process continues until the belief structure of the criterion at the top level of the model is generated. This general belief structure reflects the general assessment (preference) of a decision alternative. Depending on the number of decision alternatives

available, the aggregation process repeated until all the decision alternatives have been assessed. Consequently, the general preference of each decision alternative could be ranked and the most preferred decision alternative could be chosen.

For instance, if alternative one for a job choice problem has a general belief structure of {(Worst, 0), (Poor, 5%), (Average, 21%), (Good, 34%), (Excellent, 23%), (Top, 17%)} whereas alternative two for a job choice problem gets a belief structure of {(Worst, 3%), (Poor, 8%), (Average, 21%), (Good, 34%), (Excellent, 23%), (Top, 11%)}, by comparing the belief degree at each assessment grade, the alternative one could be identified to be preferred than alternative two, and it would be chosen as the alternative for the decision making task at hand.

Usually, the direct comparison between belief structures is not quite convenient. In such circumstance, utility measures could be introduced to transform the belief structure into numerical scores to help the ranking among decision alternatives. Take the above comparison as an example. The process starts from defining the utilities of each evaluation grades in the belief structure as follow:

$$u(Worst) = 0; u(Poor) = 0.1; u(Average) = 0.3;$$

$$u(Good) = 0.55; u(Excellent) = 0.8; u(Top) = 1.$$

Then the utilities of the two decision alternatives could be calculated as:

$$u(alternative\ one) = 0 + 0.1 \times 5\% + 0.3 \times 21\% + 0.55 \times 34\% + 0.8 \times 23\% + 1 \times 17\% = 0.609;$$

$$u(alternative\ two) = 0 + 0.1 \times 8\% + 0.3 \times 21\% + 0.55 \times 34\% + 0.8 \times 23\% + 1 \times 11\% = 0.552.$$

The alternative one has a higher utility which suggests it is a better choice in comparison to alternative two.

The formalized equations of this utility transformation process could be found in Section 2.2.3.

3.2 Judgment Analysis with Heuristic Modelling

The EBTO discussed above is a typical normative based method to overcome the human's cognitive limitations when they make decisions. It is useful in aiding people's decision making when time pressure is low and problem structure is stable over time. When applied in dynamic conditions, interactive mode and time dimension could be introduced to the aid (e.g., Hu, Si, & Yang, 2010; Si, Hu, & Zhou, 2010), where preferences and value judgments are restricted in some particular stage rather than applied globally.

But when the time pressure is extremely high and decision situation is dynamically changing that require the major shift in the way the decision makers understand the situation, the same type of application turned out not so satisfied. During 1970s and early 1980s, the U.S. army had spent millions of dollars to build very expensive decision aids for battle commanders in the field. However, they failed to be got adopted in critical conditions that were characterized by time pressure and dynamic shifting situations, and no commanders would actually use them (Klein, 1998). As in such conditions, the problem structure is not stable any further, where any dynamic change would not be enumerable in prior (e.g., causal structure, preference, value judgment). In addition, the decision makers, under extreme time pressure, hardly get the opportunity to deliberate on goals, information, and choice alternatives available.

It is becoming apparent that, in such conditions, any decision aids should not violate the DMs' deep cognitive concern (Klein, 1998). Research focus has changed to the understanding of how the DMs use their experience to make decisions, through qualitative models (see details in Section 2.3). As mentioned in chapter one, a literature gap is to be filled, that there is lack of formal quantitative model to describe the people's decision making process under critical conditions like extreme time pressure

and dynamic shifting situations. Correspondent with the quantitative modelling, a different methodology is demanded to facilitate the research investigation.

Our research addressed this literature gap by proposing a novel method, named as Judgment Analysis with Heuristic Modelling (JA-HM). The JA-HM further extends the traditional JA method to account for unique features of decision making tasks under extreme time pressure and dynamic shifting situations. The method would be elaborated step by step in the following. The novel aspects of JA-HM in against to the traditional JA would be summarized in Section 3.3.

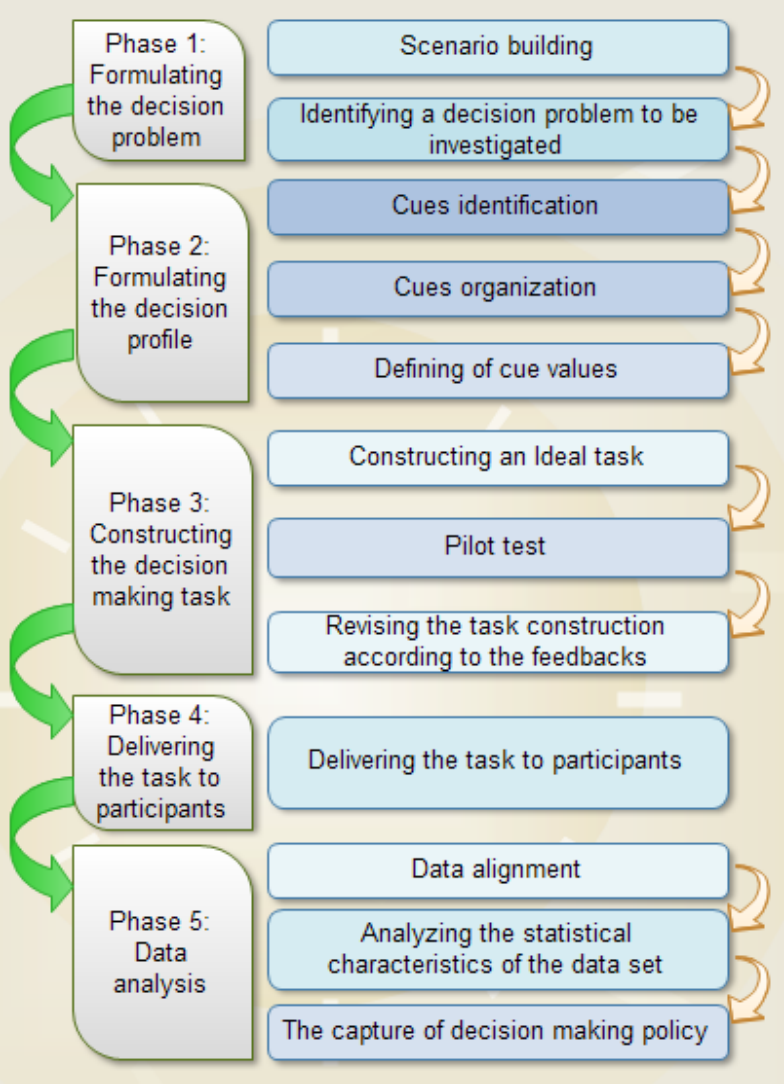


Figure 3-2, Work flow of JA-HM: A variant of JA method for decision making task under extreme time pressure and dynamic shifting situations

3.2.1 Phase 1: Formulating the decision problem

3.2.1.1 Scenario building

The methodology starts from the field investigation to identify the general scenario of the decision making. This involves the selection of a particular type of event, the definition of a decision point of the event to time stamp the scenario, and the identification of the DM.

Decision making in any domain is context specific in the first place. A Soldier mission could be a patrol task conducted by soldiers, or could be an assault task carried out together with air force. A fire incident could be an incident involving transport systems, or an incident involving biological hazards, or an incident in high rise building, etc. Different types of events get different characteristics and task demanding. To achieve the meaningful control of the study, only one type of event could be investigated at a time.

Decision making in a natural event is mostly dynamic as well. The dynamic change of situation tends to require the major shift in the way the DMs understand the situation (Klein, 1998). New information may be received, or old information invalidated, and the goals can become radically transformed. For example, as the fire develops, the commander's goal may shift from saving lives to protecting the fire spread, or as the battle develops, the Soldiers' goal may shift from destroying enemy to defending themselves. It is estimated that the situation changed an average of five times per incident during a fire operation (Klein, 1998). In this sense, we need define some time stamped decision points to deconstruct the event in a way that problem could be formulated. A decision point here refers to a critical time point when the DM experienced a major shift in his or her understanding of the situation or took some action that affected the events (Crandall, Klein, & Hoffman, 2006). It was originally proposed in the Critical Decision Method (CDM) (Klein, Calderwood, & Macgregor, 1989), and most widely applied in the NDM community. This is the first time that the decision point concept is introduced to the JA tradition to tackle the dynamic issue of the decision making task.

Decision making in a natural event, in the third place, is mostly a collaboration and co-ordination activity that involves multiple practitioners who hold different roles and are responsible for different job duties. For example, Fire and Rescue Service (FRS) in UK implements a role structure that denotes the hierarchy of its "operational management". The structure defines a taxonomy of command roles during the response

to an fire incident, ascending from front line Breathing Apparatus Wearer (BA Wearer), Breathing Apparatus Entry Control Officer (BAECO), Crew Commander (CC), to more senior Sector Commander (SC), Operations Commander (OC), Incident Commander (IC), as well as supporting positions such as Command Support (CS), Safety Officer (SO), Runner (HM Government, 2008a). It has been acknowledged that there are significant differences in the balance of cognitive skills required of command roles, across different levels (Home Office, 1997). It is therefore, essential to identify who is the DM that the scenario targets for.

3.2.1.2 Identify a decision problem to be investigated

Having built the scenario, the next step is to identify a decision problem to be investigated that is appropriate to the study aim. According to the representativeness principle held in JA tradition, the decision problem identified should be concrete problem that would be normally encountered in the real situation, rather than artificial abstract concept that are not familiar to the DMs. For example, risk awareness is essential to any fire practitioners during the fire emergency response. It is statutory requirement and routine responsibility of ICs to carry out suitable and sufficient assessments of the risks involved in responding to incidents (HM Government, 2008a). It is however, a decision problem of "what is the risk level of the incident" would be an artificial problem, as ICs have never explicitly made such general judgment during the fire emergency response.

The type of decision problems encountered by any DM during a task, characterized by extreme time pressure and dynamic shifting situations, could be classified into two categories (Klein, 1998).

One is routine problem, of which the solution is specialized and routinized in advance by department. The decision making of a DM, in this situation, is to identify the situation faced and retrieval the rule or taught method for dealing with this particular situation from memory. For example, the Standard Operation Procedures (SOPs) of responding a high rise building fire incident calls for 4 fire appliances carrying a minimum of 13 fire

crews. The first appliance in attendance would be responsible for information gathering, securing the water supply, taking control the lift. The second appliance in attendance would be responsible for setting up the bridgehead. The third and forth appliances in attendance would be responsible for other operations. The judgments involved in this SOPs sequence are 1) the control room operator identifies that it is a high rise building; 2) the commanders of the fire appliances identify how many appliances have arrived at the incident prior to their attendances. These judgments are straightforward in general.

Opposed to the routine problem, another category of decision problems cannot routinely specified, and require independent judgment of the DMs themselves. This type of decision problems are mostly at the management level, like determining priorities, calling for additional resources, diagnosing an unfamiliar situation, etc. They demand personal expertise, and the decisions made may differ among different DMs. This is the type of decision problem that could shed light on the DMs' decision making process, and is the type of decision problem to be addressed.

This categorization is another essential feature of decision making task under extreme time pressure and dynamic shifting situations, and is particularly formulated here. It is time and expertise demanding to deliberate everything instantly by the DMs themselves under these conditions. The standardized routine problems liberate the DMs from trivial details of operations so that they could focus on the management issues and response to those challenging situations that are critical to them. Without proper understanding and formulation of this problem categorization, investigations could be misleading and research findings could be difficult to compare with each other.

3.2.2 Phase 2: Formulating the decision profile

Having identified the decision problem, the information that could specify the purpose and context of the decision problem ought to be identified. The identified information could then be used to produce a variety of decision profiles to simulate different cases of decision problem. We use "cue" to indicate the information variable of the task ecology that allows for meaningful extrapolation. For example, the "time of incident" is

a cue, while the “the incident started from 9am” is a value of the cue “time of incident”. The formulation of the decision profile phase could be broke into three steps: cues identification, cues organization, and value definition of cues.

3.2.2.1 Cues identification

Ideally, the set of cues identified would be just the sort of information variables which a DM in the natural ecology would have access to, no more and no less. Several methods for cue identification have proven useful in JA literature, including survey/interview method, document analysis method, objective analysis of the ecology, and verbal protocol analysis (Cooksey, 1996).

Survey/interview method use interviews or surveys of a small sample of target DMs, or of acknowledged experts in the field, to generate a list of cues. It typically involves two distinct phases: a cue identification phase and a cue reduction phase. Cues generated in the identification phase tend to be a large number, many of which overlap in meaning and intent. The researcher would then in conjunction with the DMs or experts, refine and narrow the list down to the most appropriate set of cues (e.g., Burnside, 1994; Miesing & Dandridge, 1986; Schmitt, et al., 1987; Stewart, et al., 1992).

Document analysis method involves content analysis of document and records actually employed by target DMs in their normal practice (e.g., Maniscalco, et al., 1980; Zimmer, 1981). In general, it is somewhat more objective than the survey/interview method, but practically limited by the availability and quality of the document gathered.

Objective analysis of the ecology method usually involves the analysis of historical data where both cues and a decision are already known for a sample of potential cases (e.g., Cooksey et al., 1986), or a conceptual analysis of previously published literature targeting the decision domain (e.g., Cooksey et al., 1990; Dalgleish 1988). Depending on the availability, it is relatively the most objective method in generating the cues set. However, it is also encountered the speculation of whether or not objectively identified cues are necessarily the cues which DMs would choose to include in a cues set (Cooksey, 1996).

Verbal protocol analysis method involves applying think-aloud verbal protocol technique to generate verbal record from a small sample of target DMs, and then analysed using particular encoding and analysis scheme. To be noted, it is conceptually and methodologically different from the technique of generating verbal accounts which are generally retrospective in nature and mostly produce information similar to that recovered with survey/interview method. Indirect examples of how this method could be employed as a cue discovery technique can be found in studies conducted by Hamm (1988) and Hammond, Frederick, Robillard, & Victor (1989).

3.2.2.2 Cues organization

Two issues necessitate a cues organization process. Firstly, cues generated in the identification phase tend to be of a large number, of varying degree of significance to the DMs, and of varying degree of variability of the cue values. Secondly, many of the cues would only be comprehended by DMs in combination with others. For example, the cue "time of a high rise fire incident" itself makes no difference to fire commanders. It will make difference only when it links to some particular risks, like sleep risk. It is therefore, "time at 3am" could be a serious time for a high rise apartment fire, but a positive time for a high rise office fire. As there would be heavy sleep risk in apartment at 3am, but no sleep risk at all for office building.

It is therefore, cues organization process refers to two parts of work.

Firstly, to facilitate an appropriate control of the study, cues could be further categorized as constant cues or varied cues. Constant cues fix their values at levels that do not vary across cases. They together may specify the purpose of the decision, conditions leading up to the decision, or any other invariant characteristics of the problem that are to be decided. For those varied cues, there would be a number of possible values defined for each of them to enable the variations across the cases. They typically address the most essential characteristics of the problem, and are taken as the predictors of a DM's decision making policy.

Secondly, some of cues would be combined together as composite cues according to the domain experts' opinion.

To be noted, while the cues categorization process has been commonly applied in JA tradition, cues combination process is a novel process of our method to account for the dynamic feature of decision making task. Three collections of literature have contributed to the application of cues combination process in our method.

Firstly, literature in Situation Awareness (SA) has suggested an iterative process of DMs' perception of information and comprehension of their meanings in dynamic situations (Endsley, 1995a; 2004). In Endsley's 3 levels SA model, people are described to build their comprehension of situation through perceived elements of environment and utilize this comprehension together with some other perceived elements of environment to interpret other interest.

Secondly, JA literature has also mentioned this process but in a different way. Hammond et al. (1975) once proposed a hierarchical judgment model to account for the complex problems, in which task ecology is broken into smaller sub-problems (multiple levels), and judgments made at one level serve as the cue values for the next level. The difference, however, is that the hierarchical judgment model proposed remained a static linear process, where the first-order cues can only be utilized for first-order judgment, and those first-order judgments are always integrated with other first-order judgments. No dynamic iterative process was promoted. Another two literatures in JA that shed light on the cues combination process could be found in Phelps & Shanteau (1978)'s study on the breeding quality of gilts judged by experienced livestock DMs, and Ebbesen & Konenci (1975)'s study on the DMs' policies for setting bail. Both of studies suggested that the subjects might be following a sequential model, where they first used groups of cues to judge a limited number of more abstract qualities which they then integrated into an overall judgment. Unfortunately, neither of them has explicitly indicated whether such embedded structure was like the strict one way hierarchical process or the dynamic iterative process suggested in SA theory.

Thirdly, in addition to above applied domain, at the abstract level, the compound cue processing is not new to cognitive psychology, especially the work on knowledge acquisition and representation. It was suggested that people can and do process several cues as a configuration in certain environment (e.g., Edgell, 1993; Shanks, Charles, Darby, & Azmi, 1998; Williams & Braker, 1999). Garcia-Retamero, Hoffrage, & Dieckmann (2007) and Garcia-Retamero, Hoffrage, Dieckmann, & Ramos (2007) have made the excellent review on this. And their investigations have found people's use of compound cues in the decision-making process in a heuristic way.

3.2.2.3 Defining of cue values

Having re-organized the cues, the final step of formulating the problem phase is defining the possible values that can be taken on by each cue. For those constant cues, the process refers to defining one value for each cue that would constantly be exhibited across the cases of a decision making task. For varied cues, the process refers to defining a number of possible values for each cue that could be appeared across the cases of a decision making task.

Two issues are associated with this process. One is how the cue values should be presented in a decision profile.

In circumstance where the cues represent something that can be measured in concrete units, such as floor level of a building, those units should be used in presenting cue values.

In circumstance where cues represent something that has no natural unit of measurement, such as verbal description, traditional JA studies tended to code the information into numerical scales (for example, using simple 1 to 10 or Low to High scales to indicate the values of a cue) for ease of regression analysis and the so-called generalizability of the study. This coding process rested upon an assumption that the coding exactly reflected how subjects simplify their perceived information for judgment. We generally agreed that there was a coding process of subjects when they perceived the cue information from ecology. But presenting the coded numerical scales

straightforwardly to the DMs is fundamentally contradicting to the representativeness principle of the study design.

It is therefore, for those cues that are verbally encountered by DMs in the real ecology, concrete linguistic phrases would be utilized to represent cues in the present methodology. Then another coding process is applied to transform those linguistic phrases into numerical scales in the data analysis stage to facilitate a meaningful analysis.

We believe our strategy in presenting the cues is aligned with the task congruence requirement as summarized in Cooksey (1996), as well as the face validity demanding held in NDM community when applying simulation and laboratory techniques (Orasanu & Connolly, 1993; Zsambok & Klein, 1997).

The second issue of the defining cue values step is the degree of variability to be defined for those varied cues.

The determination of the degree of variability depends on the type of problem designed to be addressed. Take an example, the computational models of heuristics hypothesized in the AT tradition is principally designed for tasks in which all predictors are binary to simulate a binary coding process of how DMs simplify their perceived information (Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Goldstein, 2008). In such case, binary values would be defined for those varied cues. In practice, two possible presentation of each varied cue would be defined. The two presentations would be discriminated (coded) as negative (1) and positive (0) in the data analysis stage to simulate the DMs' simplification of the information they gathered. The negative presentation indicates a more serious situation to DM. The positive presentation suggests a less serious circumstance.

3.2.3 Phase 3: Constructing the decision making task

A decision making task is an experimental task that would enable the investigator to collect the data of ICs' decisions when they are facing different scenarios. Depending on

the number of varied cues formulated, there would be a population of decision profiles that could be produced through the manipulation of those varied cues.

For example, if there are 8 varied cues with binary values defined in a decision profile, there could produce 256 (2^8=256) different hypothetical decision profiles in total through manipulation of these 8 cues.

A decision making task would be constructed by involving a sequence of different decision profiles from the population through a sampling plan, to enable a statistical estimation of DMs' decision making policies.

The sampling plan should be carefully determined, so that it will: 1) be representative of the ecology as argued in SJT (see Section 2.4); 2) enable stable estimation of the parameters of decision making policy; 3) and be feasible for each participant to complete the task in reasonable time without fatigue or boredom.

It is however, these goals may conflict among each other (Stewart, 1988). On the one hand, statistical estimation is more precise and stable when cues are few and uncorrelated and a large volume of decision data is available. On the other hand, representative design often dictates correlated cues, and the participants' time and toleration on task is constrained. In addition, as reviewed in Section 2.4, the formal properties of task ecology are often unavailable which makes representative design inflexible. It is therefore, the final task designed typically emerges from compromises between the ideally desirable features of the design and the practically realizable features which various constraints impose.

In practice, the task construction phase of our method involves three critical determinations on the sampling plan of specific profiles. They are 1) determining whether a decision making task is completed by an individual participant or by a group of participants; 2) determining the number of profiles to be included in a decision making task; 3) determining the distribution and intercorrelations of cues in a decision making task.

These determinations would be achieved through a three-step cycle process.

a) Ideal design

An ideal task is designed according to the general recommendation in JA literature.

JA literature generally recommended that a decision making task would be designed for an individual participant to complete (Hammond et al., 1980).

In terms of the number of decision profiles (cases) to be included in a decision making task, it is generally agreed in the JA literature that a ratio of 5 cases to every varied cue would be the minimum ratio that one should proceed with in constructing a decision making task (Cooksey, 1996). To be noted, although it is not compulsive, JA tradition generally advises investigators to include a small number of replicated cases in a task to assess the decision inconsistency. They may either be randomly interspersed with the original cases or placed as a set at the end. In these cases where replications are introduced, the recommendation of cases cues ratio is still generally hold (Cooksey, 1996).

Regarding the distribution and intercorrelations of cues, it is recommended that the distribution and intercorrelations of cues should resemble those in the real environment. When these formal characteristics of environment are not available, most JA studies in literature, chose to randomly generate the cue profiles or deliberately construct them to satisfy an orthogonal design, with uniform distributions and zero intercorrelations of cues (Cooskey, 1996).

b) Pilot test

Having produced the ideal task, a pilot test on a small group of target DMs is ought to be conducted. The pilot test would enable the investigators to double check the appropriateness of cues organization, identify any potential conflict of cue presentations, and most importantly, determine the number of cases that individual participant would be willing to complete.

The double checking of the appropriateness of cues organization is to ensure that the cues, especially for those composite cues, are perceived and comprehended by participants as expected in the design of decision profile.

The identification of potential conflict of representations of cues is to ensure that there would be no unrealistic or atypical profiles produced.

The determination of number of profiles that individual participant would be willing to complete, provides essential knowledge of whether an ideal decision making task could be completed by an individual participant or not. If not, adjustment is demanded to revise the construction of decision making task.

Two types of adjustment could be applied. If the discrepancy between the ideal number of profiles in a task and the number of profiles that individual participant would be willing to complete is small, adjustment could be made by reducing the number of decision profiles included in a decision making task. In this case, the statistical power of estimation is sacrificed, but the task is able to be conducted at the individual level.

If the discrepancy is large, adjustment must be made to have the decision making task complete at the group level. In other words, a decision making task would be constructed to be completed by a group of participants in turn. Two references in JA literature shed lights on this adjustment plan. In Dhami (2003)'s investigation, DMs in London courts were observed to examine how punitive decisions were made. The observations were analyzed at the court level due to the fact that individual benches make too few decisions for meaningful analysis. A similar example could also be found in Ebbesen & Konenci (1975). Both of these two studies were widely acknowledged and cited, and yielded similar findings of those studies that conducted decision making task at individual level (e.g., Dhami & Ayton, 2001). In this sense, we see it as an acceptable trade-off between the theoretical proposal and the real situation constraint.

c) Revising the construction of decision making task

Accordingly, the construction of decision making task is revised based on the feedback collected from the pilot test. To be noted, if adjustment is made to have the decision making task complete at the group level, two additional issues needs to be addressed.

Firstly, a decision making task completed by a group of participants implies that those participants of the same group are assumed to hold the same decision making policy. This would demand a careful sampling plan of participants to ensure the expert level of participants within one group would be close. As the decision data of a decision making task would be idiographically analyzed in the end (Idiographic data analysis will be described in Section 3.2.5).

Secondly, having a decision making task completed by a group of participants, on the other hand, offers a higher flexibility in the number of decision profiles that could be included in the decision making task. As a group of participants together are more plausible to complete a large number of decision profiles. Accordingly, a decision making task could be constructed with more decision profiles for the advantage of statistical estimation.

3.2.4 Phase 4: Delivering the decision making task to participants

Having constructed the decision making task, the next phase is to have the participants complete the task. Methods for preparing and presenting decision profiles and for recording decision data are to be considered.

JA regularly employs paper-and-pencil implementations to deliver decision making tasks because of the greater degree of control over task aspects and flexibility in executing. Specialized computer systems have also come into place since the technology advances in computer, such as the POLICY PC (Executive Decision Services, Inc, 1991).

According to the representative principle, it is ideally that a decision making task appears in the same form in the study as when the DMs ordinarily perform the task. Actual experiments, however, could rarely fulfil this high standard requirement and would utilize the so called ‘paper people’ or ‘paper pig’ instead, in which a set of coded

cues are graphed or listed in paper, simulating the real person or real situation for deciding. Decades of studies in JA tradition generally agreed that "the paper format does not lead to any important distortions in the policies obtained … and that policies obtained with paper patients, pupils and pigs can be used to predict judgments about their real counterparts at least when the characteristics of the tasks are well known." (Brehmer & Brehmer, 1988, p.89).

Executing the decision making task is most frequently implemented through organizing one or multiple sittings to have DMs involving together at a time. Mail survey could also be found in the literature (e.g., Dhami & Ayton, 2001).

3.2.5 Phase 5: Data analysis

The final phase of the JA-HM method is analyzing the decision data through statistical tools and computational models of heuristics to indicate the decision making policies applied by the DMs.

Two general concepts need to be addressed here for understanding the following data analysis procedure employed in this thesis. They are "nomothetic" and "idiographic" (Hammond et al., 1980). When more than one DM make decisions about a set of cases, there are two possible ways for analyzing. The data could be averaged over to obtain a mean decision for each case and then those mean decisions can be analyzed. This way is termed "nomothetic". It assumes that, with respect to the given decision making task, all DMs are essentially replicates of one another. With such an assumption, large numbers of DMs are generally demanded to provide large number of decision records of the same case to increase statistical power.

Alternatively, data could be analyzed individually of each DM. This is called "idiographic". It assumes that there exist important and reliable individual differences between DMs with respect to the given case. With such an assumption, it is generally advantageous to observe many responses from the same DM. But such task burden of individual DM often serves to limit the number of DMs included in any one study. So, compared to a

typical nomothetic study, the idiographic method usually has smaller number of DMs and larger number of cases completed by the DMs (Hammond et al., 1980).

Following the JA tradition (Cooskey, 1996), the data analysis in our method is always idiographic. When a decision making task is completed by a group of participants in turn, the method assumes the participants replicate the decision making policy of one another within group, and the participants differ the decision making policy of one another across groups.

Three general procedures contribute to the data analysis phase. They are data alignment, analyzing the statistical characteristics of the data set, and the capture of decision making policy.

3.2.5.1 Data alignment

Data alignment refers to recoding the decision data into a number of fixed categories for consequent data analysis.

As mentioned in the problem formulation section, the identified decision problem which could shed light on the DMs' decision making policy is the type of problem that is mostly at the management level, demand personal expertise, and the decisions made may differ among different DMs. The decision variability expressed by different DMs needs to be organized in a way that could enable a meaningful analysis of decision data across DMs.

One example of such alignment process could be found in Lipshitz & Strauss (1997), where the DMs' qualitative decision data in coping with uncertainties are summarized into five categories (Reduction, Assumption-based reasoning, Weighing pros and cons, Forestalling, and Suppression). Through such alignment process, the decision data from different DMs could be meaningfully gathered together and compared among each other.

3.2.5.2 Analyzing the statistical characteristics of the data set

Having aligned the decision data, formal data analysis could be applied. Depending on the research objective, some statistical characteristics of the data set would normally be analyzed in the first place. Typical statistical characteristics include: intra-individual consistency in decisions, inter-individual agreement in decision making policies, self-insight of decision making policy, and post-decisional confidence.

Intra-individual consistency in decisions

Hammond et al. (1975) proposed that consistency should be measured in terms of the variance of decisions made in a test-retest situation. In practise, it is normally measured through the correlation between set of decisions on the test-retest profiles.

It is argued in the JA tradition that the ability of a model to predict an individual's decision data is limited by his or her consistency in making decisions. Because it is assumed that an inconsistent individual will be difficult to predict.

When the decision making task is planned at the group level, the consistency measure of test-retest profiles indicates the intra-group consistency in decisions.

People may be inconsistent in their decisions for a number of reasons, including fatigue, shifts in attention, boredom, and the fact that the task is highly unpredictable.

Inter-individual agreement in decision making policies

Agreement may also be considered at the level of the decision making policies, by comparing relative cue weights through computing correlations or cluster analysis (Hammond et al., 1975).

When the decision making task is planned at the group level, the mean of cue weights within the group is firstly calculated to account for the decision making policy for the group. And then the measure would indicate the inter-group agreement in decision making policies.

Self-insight of decision making policy

Researchers in JA tradition have also compared the subjective weights elicited from direct assignment with the statistical weights derived from the computational model (e.g., regression model). Taken the statistical weights as objective, such comparison has been supposed to indicate the DMs' self-insight of their own decision making policies.

The results of studies on this self-insight are mixed and no general conclusion has been drawn (Brehmer & Brehmer, 1988). On the one hand, the typical conclusion has been that subjective weights were generally inferior to statistical weights suggesting that DMs lacked sufficient insight into the relative importance of cues they applied in decision (Cooksey, 1996). On the other hand, some researchers have also showed that DMs can recognize their decision policy as long as the appropriate methodology being applied in obtaining the subjective weights (Reilly & Doherty, 1989; 1992), and there was no significant difference between subjective policies and statistical policies in terms of predictive accuracy (Cook & Stewart, 1975). Considering the potential exists for introducing additional method biases when gathering subjective weight information, the use of subjective weights has generally been less preferable to statistical weights for addressing the DMs' decision making policy in JA tradition (Cooksey, 1996).

Post-decisional confidence

Frequently, studies in JA tradition would design a separate continuous scale followed with each decision making profile to measure DMs' feeling of confidence in the decisions. As one of the goals of SJT research is to improve existing decision making policies by providing cognitive feedback and decision aids (Hammond et al., 1975), and the ongoing feelings of confidence is useful indicator of whether individuals will be amenable to such intervention (Zakay, 1997). The high confidence in a policy may imply an unwillingness to change it.

3.2.5.3 The capture of Decision making policy

The capture of decision making policy refers to identifying a computational model that best predicts the decisions made by the DMs. By capturing the decision making policy, the DMs' decision making processes could be specified.

Historically, the capturing procedure in JA has been intimately associated with the statistical process of multiple regression. This procedure yields a weighted linear model that describes a DM's decision making policy in terms of statistically significant cues in the model, the relative cue weights, the form of the function (e.g., linear, additive), and the predictability as measured by the model (e.g., R^2).

The pervasive use of regression models as the only model in describing DMs' decision making process is surprising, and critics have questioned the validity of this methodology principle in terms psychological plausibility, flexibility and adaptability.

Firstly, the decision making process characterised by the regression model is not easily reconciled with what has been known about human psychological capabilities. According to the regression model, multiple cues are differentially weighted and integrated in a compensatory way, to inform a decision. However, the human mind is characterised by limited attention, memory, and cognitive processing capability (e.g., Kahneman, 1973). In fact, people may choose strategies that reduce cognitive effort (e.g., Payne, Bettman, & Johnson, 1993).

Secondly, the regression models provide a static description of human decision making behaviour, where the same cues are used in the same way when deciding on each case. This way of modelling portrays humans as inflexible. However, according to Brunswik (1952)'s vicarious functioning, distal variables can be attained vicariously through proximal variables, and proximal variables can themselves be used vicariously through other proximal variables (see Section 2.4.1.2). In other words, DMs is capable to use different subsets of cues for making the same kinds of judgment. The DMs are supposed to be flexible on the information they used in a decision.

Thirdly, empirical research has reported that certain cognitive strategies are more descriptively and predictively valid under particular task structures (e.g., Czerlinski, Gigerenzer, & Goldstein, 1999; Einhorn & Hogarth, 1975; Martignon & Hoffrage, 1999). In other words, the decision making processes are adapted to the structure and demands of the task (Gigerenzer & Selton, 2001).

It is therefore, recent JA studies started to conduct performance competition among alternative models, and the model that best predicts a DM' decisions is identified as the plausible decision making policy the DM applied (e.g., Dhami, 2003; Dhami & Ayton, 2001; Garcia-Retamero, Dhami, 2009). In fact, Brunswik (1955b, 1956), in his initial proposal of probabilistic functionalism, did not rule out the use of other models, and Hammond (1996b) has also confessed that "a … sin of commission on my part was to overemphasize the role of the multiple regression technique as a model for organizing information from multiple fallible indicators into a judgment" (p.244).

Our method follows this principle of competitive test of models as the way of identifying the DM's decision making policy (Gigerenzer & Brighton, 2009; Gigerenzer & Gaissmaier, 2011).

Model selection criteria

The implementation of competitive test is, in essence, a process of deciding which model provides a best explanation for the data. This comparison of alternative models is called model selection. A number of model selection criteria are available in literature (see Jacobs & Grainger, 1994, for a detailed overview). It could be falsifiability, that is, whether the models can be proven wrong. It could be psychological plausibility, such as whether the computations postulated by a model are tractable in the world beyond the laboratory (e.g., Gigerenzer, Hoffrage, & Goldstein, 2008).

One most widely used model selection criterion is descriptive adequacy which is often evaluated in terms of "goodness of fit". That is, when two or more models are compared, the model that provides the smallest deviation from existing data (measured,

for example, as R^2) is favoured over a model that results in a larger deviation from the data. However, selecting model based exclusively on such measures of fit comes across one essential limitation: the problem of overfitting (Pitt, Myung, & Zhang, 2002).

In real world circumstance, noise-free data are practically impossible to obtain. Therefore, researchers are confronted with the problem of disentangling the variation in data caused by noise and the variation caused by psychological mechanisms (e.g., decision making policy). "Goodness of fit" measures alone cannot make this distinction. As a result, a model can end up overfitting the data. That is, it can capture not only the variance caused by the cognitive process under investigation but also that caused by random error.

One alternative criterion that could complement the limitation of "goodness of fit" criterion is the "generalizability". That is, the degree to which a model is capable of predicting all potential samples generated by the same cognitive process, rather than fitting only a particular sample of existing data.

Figure 3-3 illustrates an example that could shed light on the difference. In the figure, the model A overfits existing data by chasing after idiosyncrasies in that data. This model fits the existing data perfectly but does a poor job of predicting new data. Model B, albeit not as good at fitting the existing data, captures the main tendencies in that data and ignores the idiosyncrasies. This makes it better in predicting new data. In general, the more free parameters a model has, the better the fit, but this does not hold for predictions (Pitt, Myung, & Zhang, 2002).

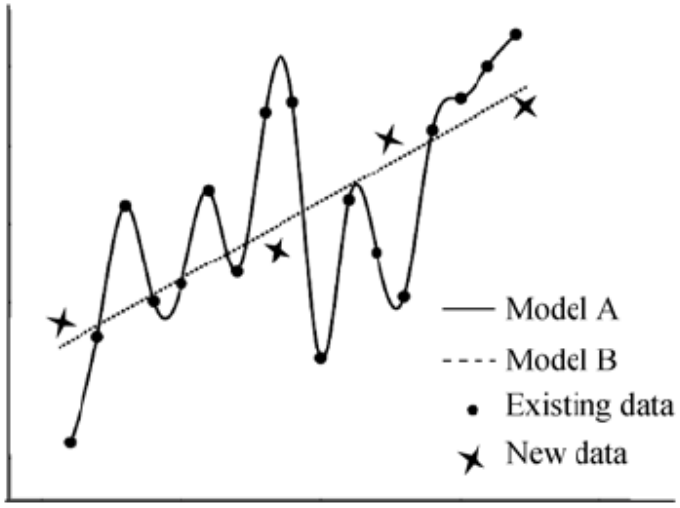


Figure 3-3, Schematic illustration of how two models fit existing data and how they predict new data. Model A overfits the existing data and is not as accurate as Model B in predicting new data (Pitt, Myung, & Zhang, 2002).

Cross-validation

In practice, the two model selection criteria illustrated above, could be calculated through a cross-validation procedure (Browne, 2000; Stone, 1974; 1977). The procedure generally involves the following steps:

1) Randomly split a sample of cases into two sub-groups identified as the derivation sample and the validation sample. Common schemes of such division have included 50%/50%, 60%/40%, or 75%/25% splits between derivation and validation sample sizes (Hair et al., 1992).

2) The parameters of each hypothesized model are estimated using the derivation sample. Each model is then used to generate predictions for all the cases in the same sample and the model fit on derivation sample is calculated, which represent the model's "goodness of fit".

3) Each model is then used to generate predictions for all the cases in the validation sample and the model fit of validation sample is computed, which represent the model's "generalizability".

Noted, we refer to the percentage of decisions correctly predicted by the models as “fit”. Often, to avoid any idiosyncratic effect caused by the sample division, the above three steps can be repeated N times to yield N-round cross-validations (e.g., Dhami, 2003). The computations of model fit can then be averaged for the final estimates.

Finally, the conducting of cross-validation would demand an additional number of cases to be assembled in the task construction phase. As both derivation and validation sample need to have appropriate sample size to permit efficient and stable statistical estimates to be computed.

3.3 Summary

This chapter has addressed the research methodology of our study. Two novel methods, EBTO and JA-HM were elaborated, each of which could accomplish the investigation of decision making tasks under specific conditions.

The EBTO provides a novel framework to aid people’s decision making under uncertainty and imprecise goal. Under the framework, the imprecise goal is objectively modelled through an analytical structure, and is independent of the task requirement; the task requirement is specified by the trade-off strategy among criteria of the analytical structure through an importance weighting process, and is subject to the requirement change of a particular decision making task; the evidence available, that could contribute to the evaluation of general performance of the decision alternatives, are formulated with belief structures which are capable of capturing various format of uncertainties that arise from the absence of data, incomplete information and subjective judgments.

The JA-HM further extend the traditional JA method, through a number of novel methodological procedures, to account for the unique features of decision making tasks under extreme time pressure and dynamic shifting situations. By elaborating the procedures of JA-HM step by step, we have discussed why and how these novel aspects

of JA-HM could accomplish the deficiencies of traditional JA. We summarize them here as the end of this chapter:

1) In dynamic shifting situations, the situation structure is not stable, and any dynamic change would not be enumerable in prior (e.g., causal structure, preference, value judgment). To deconstruct the dynamic shifting situations in a way that decision problem could be identified and formulated, a notion of decision point is introduced by drawing the literature from NDM. (Section 3.2.1.1)

2) In tasks that characterized by extreme time pressure and dynamic shifting situations, decision problems could be classified into two categories: routine problem and non-routine problem. Only the non-routine problem demands independent judgment of the DMs and could shed light on the DMs' decision making process. It is the type of decision problem to be addressed. However, due to the demand of personal expertise and complexity of the problem, the decisions made on a non-routine problem may differ among different DMs. Accordingly, a data alignment process need to be applied to enable meaningful gathering and analyzing of decision data from different DMs. We introduce these problem classification and data alignment processes based on the literature from the NDM. Without proper formulation of them, the applications of JA, in domains characterized by extreme time pressure and dynamic shifting situations, could be misleading, and their research findings would be difficult to compare with each other. (Section 3.2.1.2 & Section 3.2.5.1)

3) In a dynamic task environment, DMs tend to apply an iterative process to comprehend situation through the information they perceive. Under this process, an individual cue would not necessarily be meaningful in its own. Rather it would often be comprehended together with some other cues or together with the comprehension from other cues. Accordingly, a notion of composite cue is formulated in the decision profile development process of the method to account for such unique information processing behaviour in dynamic situations. We discuss this novel notion by drawing the literature from Situation Awareness (SA) and cognitive psychology (Section 3.2.2.2).

4) To account for the time constraints and process dynamics of DMs' decision making under extreme time pressure and dynamic shifting situations (e.g., information search, shift in information emphasis or decision making policy), computational models of heuristics are applied to replace the static multiple regression in modelling the DMs' decision making policies. By capturing the decision making policy, the DMs' decision making processes could be specified. (Section 3.2.5.3)

5) Associated with the methodological principle of competitive testing, two model selection criteria and a cross-validation process are formulated to facilitate the application of the method. (Section 3.2.5.3)

Chapter 4 : Case study of Soldier System decision making

This chapter applies the EBTO method to a case study of Soldier system decision making to demonstrate how the method is applicable to address the decision making task under imprecise goal and uncertain information. The background and problem will be firstly illustrated. According to the procedures suggested in EBTO, an analytical structure (assessment model) was then constructed to account for the general goal of Soldier system decision making. We then weighted the relative importance of the model criteria to capture the task requirements of two Soldier missions, respectively, and assessed the performance of two sets of Soldier system through the bottom criteria of the model. These assessments were finally aggregated according to the importance weighting of criteria, and the choice of decision alternative was prescribed for each Soldier mission. Through the application, the method, as a tool, is demonstrated to be able to provide the holistic analysis regarding the requirements of Soldier missions, the physical conditions of Soldiers, and the capability of their equipment and weapon systems, which is critical in domain.

4.1 Soldier system decision making

Soldiers are typically classified as Army service members. The Soldier System consists of the individual Soldier and those items and equipment that the Soldier wears, carries, or consumes. It includes all items in the Soldier's load and those items of equipment to accomplish unit missions that the Soldier must carry (TRADOC, 2006).

Soldier System decision making concerns with the deciding of the optimal match between a specific Soldier mission and an individual Soldier system.

The problem is raised from the concept of Soldier as a System (SaaS) which was proposed by U.S. Army during 1990s (U.S. Department of the Army, 1991).

In contrast to the traditional view of individual Soldier as a "replaceable item", the SaaS concept confirmed Soldier as the centrepiece of Army formations, and a valuable military asset with large impact on operational performance and mission success. The

individual Soldier, under this concept, was supposed to hold superior capabilities to accomplish assigned tasks across the spectrum of conflict, in any operational environment (Lockhart, 2006). The demanding of such fully integrated capability, sometimes, obscures the goal of individual Soldier in a mission, and resulted in the field equipping heavy, bulky and burdensome, degrading the Soldier's effectiveness and performance in an unexpected way. During operations in Iraq for example, soldiers involved in dismounted close combat were often carrying loads in excess of 60kg (Jackson 2004), which seriously impact their endurance, combat effectiveness and resulting in long-term health implications. According to a recent study by Danish Army Combat Centre, a range of physical problems closely associated with weight, with 10-15% of troops suffering long-term injuries and with significant proportions of patrols being on painkillers for parts of their mission (Baddeley 2010).

A capability gap is identified in the current force that there is lack of integrated methodology to provide a process and management structure with which to respond to the concerns on the requirements of mission task, the physical conditions of Soldiers, and the capability of their equipment and weapon system, in a holistic manner (TRADOC, 2006).

The effectiveness of any methodology to address the capability gap could be demonstrated by the following scenario.

Supposing an Army combat developer are deciding an appropriate Soldier system for assault mission and patrol mission, respectively.

There are two sets of Soldier system available. They are designed based on the recommendation of the Soldier Integrated Protective Ensemble (SIPE) program (Victor et al., 2000), in which a Soldier system was composed of a number of subsystems. Specifically, the Soldier system 1 covers the components that a sergeant commonly has, and is estimated to be weighted around 50kg. The Soldier system 2 includes the components with more advanced and comprehensive configuration, but is estimated to be weight around 65kg. They are depicted in Table 4-1 below.

The two missions, on the other hand, are recognized as the two most commonly implemented missions for soldiers. In assault mission, two armies would maneuver to contact, at which point they would form up their infantry and other units opposite each other. Then one or both would advance and attempt to defeat the enemy force. It is a mentally and physically demanding activity. While in patrol mission, the soldier has to be able to walk for a certain amount of distance for security or reconnaissance purpose, looking out for anything out of the ordinary – which if found will be reported for assistance or dealt with as appropriate.

Two questions are to be answered by the combat developer. Which Soldier system is more appropriate to the patrol mission? Which Soldier system is more appropriate to the assault mission?

Table 4-1, The Soldier systems to be considered

Subsystems	Soldier system 1	Soldier system 2
1. Integrated Headgear Subsystem (IHS)	*Soldier-to-soldier communications	*Weapons interface (M16A2-mounted thermal sight and laser aiming light)
2. Advanced Clothing Subsystem (ACS)	*Integrated body armor/ammunition carriage	*Integrated body armor/ammunition carriage
	*Handwear (combat, chemical/biological)	*Handwear (combat, chemical/biological)
	*Footwear (integrated combat boot, gaiter)	*Footwear (integrated combat boot, gaiter)
		*Uniform components (chemical vapor undergarment, advanced combat uniform, advanced shell garment)

		*Passive cooling T-shirt
3. Microclimate Conditioning/Power Subsystem (MC/PS)	*Blower	*Filter
4. Weapon Subsystem (WS)	*M16A2 (standard infantryman's rifle)	* M16A2 (standard infantryman's rifle)
		*Aim-1D laser aiming light
5. Individual Soldier Computer (ISC)	*Message management/reporting	*Global positioning system/digital mapping
6. Others	*Helmet	*Vest with dagger and 2 liter water pack
	*Protection vest with 1 liter water pack	*Bag with supplies for 48 hours

4.2 The assessment model for a Soldier system

Following the framework suggested in EBTO method, the first step to tackle the decision making task described above is to develop an assessment model to represent the general effectiveness of a Soldier system (general goal) in an objective, formal structure.

Previous research in literature has identified criteria for assessing the effectiveness of individual Soldier (e.g., O'Keefe IV, Mcintyre III, & Middleton, 1992; O'Keefe IV & Porter, 1992; Victor et al., 2000; CDE, 2009). The model proposed in this study is developed on the basis of these literatures, and is illustrated in Figure 4-1 below.

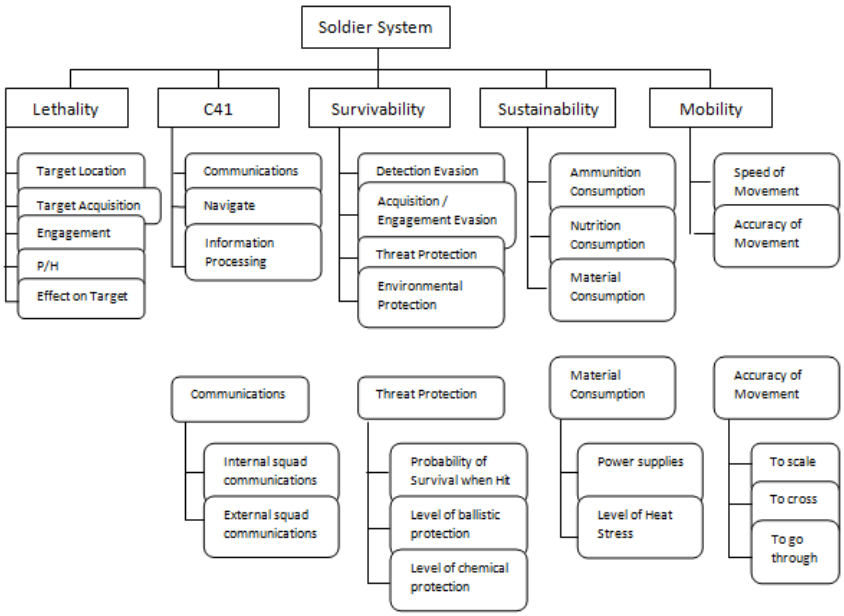


Figure 4-1, Soldier system hierarchical model

The lethality, as the name suggests, indicates the ability of the soldier system to destroy, neutralize, suppress, or bring desired effects on a threat, through the employment of organic and coordinated use of non-organic weapon systems. It further constitutes five sub-capabilities, including the ability to locate/position targets, to acquire enemy targets/info, to engage enemy, to incapacitate/destroy targets, and the hit chance of a shot (P/H).

The C41 (command, control, communications, computers and intelligence), refers to the soldier's ability to direct, coordinate and control personnel, weapons, equipment, information and procedures necessary to accomplish the mission. It can be further determined by the internal and external squad communication, capability of navigation and information processing.

The survivability describes the degree to which a soldier system is able to avoid or with stand a natural and manmade hostile environment, without suffering an abortive impairment to accomplish its designated mission, such as the ability to avoid detection, to disperse, and the protection accomplished.

The sustainability refers to the ability of the soldier system to sustain operations and be logistically supported in order to accomplish its assigned tasks. It includes areas such as power supplies with the system, availability and maintainability factors.

The mobility suggest the quality that permits soldier system to move from place to place or perform individual tasks in a timely fashion, while retaining the ability to fulfil their primary mission. It can be further broke down into four components: to scale (climbing over walls, ascending & descending cliffs), to cross (gaps, rivers, mine fields), to go through (walls, tunnels and doors etc.), and to move (speed on all surfaces, e.g., road, stones, sand, snow, marsh, slope).

4.3 Data collection

Once the assessment model has been constructed, the data regarding the decision making are collected at two dimensions: the task requirement of the specific Soldier mission to be carried out, and the detailed performance of Soldier system to be deployed.

4.3.1 Importance weighting of criteria

The task requirements of the patrol mission and assault mission are different and could be specified by different trade-off strategies among criteria of the assessment model. Such trade-off strategies were addressed through assigning the relative importance of criteria in the model, respectively. The assignments were collected through an interview with an ex-soldier. He was asked to assume himself as a combat developer, and judge the relative importance of each criterion specific to the task requirements of two missions, respectively.

The weighting assignments are depicted in Table 4-2.

Table 4-2, Relative importance of the criteria

Criterion (Importance weighting for assault mission/patrol mission)		
Level 1	Level 2	Level 3
1. Lethality (30%/30%)	1.1 Target Location (30%/10%)	
	1.2 Target Acquisition (10%/10%)	
	1.3 Engagement (30%/20%)	
	1.4 P/H (10%/20%)	
	1.5 Effect on Target (20%/40%)	
2. C41 (20%/30%)	2.1 Communications (30%/40%)	2.1.1 Internal Squad Communications (70%/40%)
		2.1.2 External Squad Communications (30%/60%)
	2.2 Navigate (20%/20%)	
	2.3 Information Processing (50%/40%)	
3. Survivability (20%/20%)	3.1 Detection Evasion (40%/40%)	
	3.2 Acquisition / Engagement Evasion (30%/30%)	
	3.3 Threat Protection (20%/10%)	3.3.1 Probability of Survival when Hit (30%/40%)
		3.3.2 Level of ballistic protection (40%/40%)
		3.3.3 Level of chemical protection (30%/20%)
	3.4 Environmental Protection (10%/20%)	

4. Sustainability (10%/10%)	4.1 Ammunition Consumption (60%/80%)	
	4.2 Nutrition Consumption (30%/10%)	
	4.3 Material Consumption (10%/10%)	4.3.1 Power Supplies (60%/60%)
		4.3.2 level of Heat Stress (40%/40%)
5. Mobility (20%/10%)	5.1 Speed of Movement (40%/60%)	
	5.2 Accuracy of Movement (60%/40%)	5.2.1 To scale (30%/40%)
		5.2.2 To cross (30%/30%)
		5.2.3 To go through (40%/30%)

The weighting assignments indicated that the task emphasis was not extremely differed between the two missions. While patrol mission may demand more efficient C41 through trading off the mobility in against to the assault mission, the two missions were assessed to hold similar demanding on lethality, survivability, and sustainability.

Further examining the detailed level 2 and 3 factors, greater difference is identified of the task requirement between two missions. In terms of “lethality”, the ability to position targets and engage enemy are more demanding in assault mission. This is probably due to the fact that the objective of assault is to defeat and dislodge the enemy force, thereby establishing control of the area. On the other hand, the emphasis of “lethality” in patrol mission is changed to the “P/H” and “effect on target”. The two criteria together describe the quality and accuracy of a shot. Because patrol mission is typically conducted by small unit of soldiers, their instant reaction of enemy activity is to report to camp and wait for support. They won’t start a fire by their own, unless it is necessary. And if they start a fire, they want to hit and destroy the enemy as quick and efficient as possible. This mission requirement difference is also reflected in the

consideration of “C41”, in which internal squad communication is more important for assault mission, while external squad communications is comparatively essential to patrol mission. Regarding the “survivability”, the difference is minimum, in which slight trade-off between “threat protection” and “environmental protection” is identified of the two missions. Regarding the consideration of “sustainability”, the relative importance of ammunition consumption is greater for patrol mission. As patrol mission is commonly conducted in turn, in which each unit is responsible for 3 hours’ duration. Nutrition and material consumption are not so critical in such short mission duration. Finally, looking at the “mobility”, the accuracy of movement is comparatively more important to assault mission due to the potential demanding of close combat.

4.3.2 Assessment at the bottom level criteria

Importance weighting of criteria specifies the task requirements of the missions, assessment at the bottom level criteria, on the other hand, indicates the performance of the Soldier system when conducting the mission. It depends on the set of equipment and weapon system that is assembled to the individual Soldier and psychical condition of himself. The assessment would often be characterized by uncertainty. As the performance of a Soldier system in a mission has to be estimated in prior, which are subjective in nature and sometimes incomplete due to the unpredictable nature of Soldier missions and mission contexts.

The assessment data of this study was collected through an interview with the same ex-soldier. He was asked to assume himself as a Soldier, and assess each bottom level criterion according to the performance he would achieve when he is assembled with the two Soldier systems, respectively. The interviewee consistently used a five grades set to assess the criteria, from “Very Good” (VG) to “Very Poor” (VP). The complete assessments are presented in Table 4-3 below.

Table 4-3, Assessments of bottom level criteria

Level 2	Level 3	Soldier system 1	Soldier system 2
1.1 Target Location		{(G, 0.3), (VG, 0.7)}	{(G, 0.6), (VG, 0.4)}
1.2 Target Acquisition		{VG, 1.0}	{G, 1.0}
1.3 Engagement		{G, 1.0}	{(N, 0.3), (G, 0.7)}
1.4 P/H		{G, 0.8}	{(N, 0.1), (VP, 0.9)}
1.5 Effect on Target		{*}	{(N, 0.1), (VP, 0.9)}
2.1 Communications	2.1.1 Internal Squad Communications	{(G, 0.6), (VG, 0.4)}	{(N, 0.1), (VG, 0.9)}
	2.1.2 External Squad Communications	{*}	{(N, 0.1), (VP, 0.9)}
2.2 Navigate		{VG, 1.0}	{(G, 0.7), (VG, 0.3)}
2.3 Information Processing		{G, 1.0}	{G, 1.0}
3.1 Detection Evasion		{(N, 0.4), (VG, 0.6)}	{N, 1.0}
3.2 Acquisition / Engagement Evasion		{N, 1.0}	{G, 1.0}
3.3 Threat Protection	3.3.1 Probability of Survival when Hit	{P, 1.0}	{N, 1.0}
	3.3.2 Level of ballistic protection	{P, 1.0}	{N, 1.0}
	3.3.3 Level of chemical protection	{N, 1.0}	{N, 1.0}
3.4 Environmental Protection		{N, 1.0}	{N, 1.0}
4.1 Ammunition Consumption		{*}	{VG, 1.0}
4.2 Nutrition Consumption		{G, 1.0}	{VG, 1.0}
4.3 Material Consumption	4.3.1 Power Supplies	{*}	{(G, 0.4), (VG, 0.6)}
	4.3.2 level of Heat Stress	{*}	{*}
5.1 Speed of Movement		{VG, 1.0}	{(G, 0.2), (VG, 0.8)}

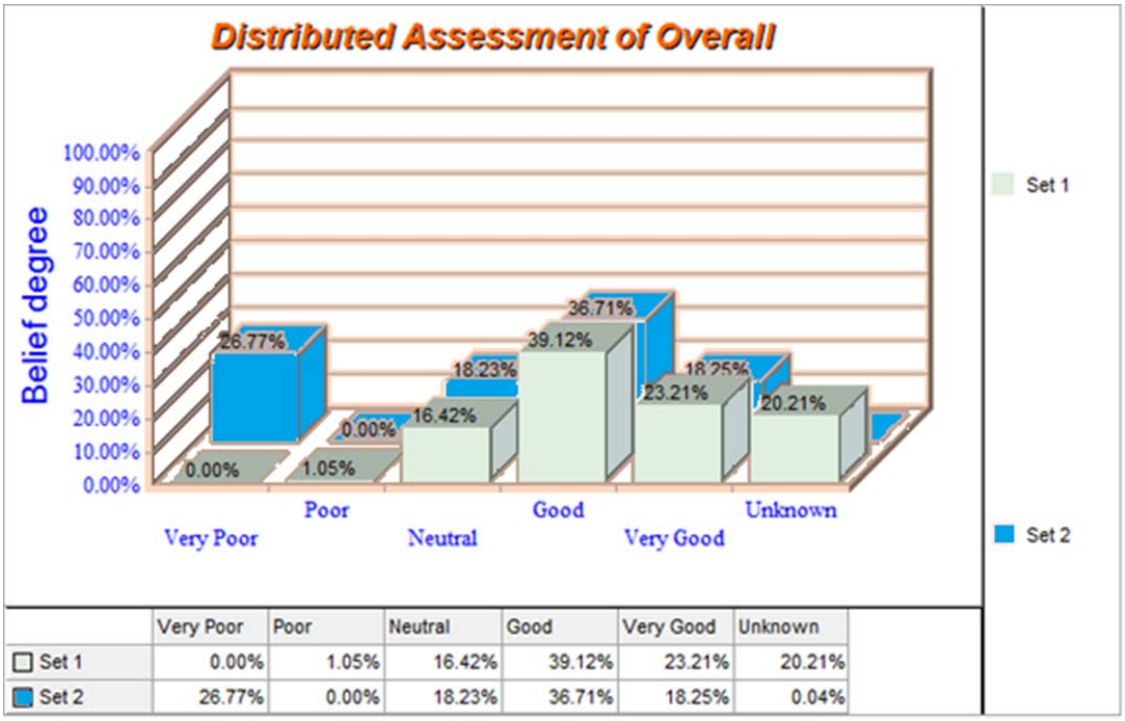
5.2 Accuracy of Movement	5.2.1 To scale	{G, 1.0}	{G, 1.0}
	5.2.2 To cross	{(G, 0.4), (VG, 0.6)}	{G, 1.0}
	5.2.3 To go through	{{N, 0.4}, (G, 0.6)}	{G, 1.0}

(The assessment grades are defined as VP – very poor, P – poor, N – neutral, G – good and VG – very good. The number behind the assessment grade is its associated belief degree. Missing judgment is represented with symbol *, in which our interviewee has no idea of its assessment based on the information he obtained.)

4.4 Assessment aggregation & Choice of Soldier system

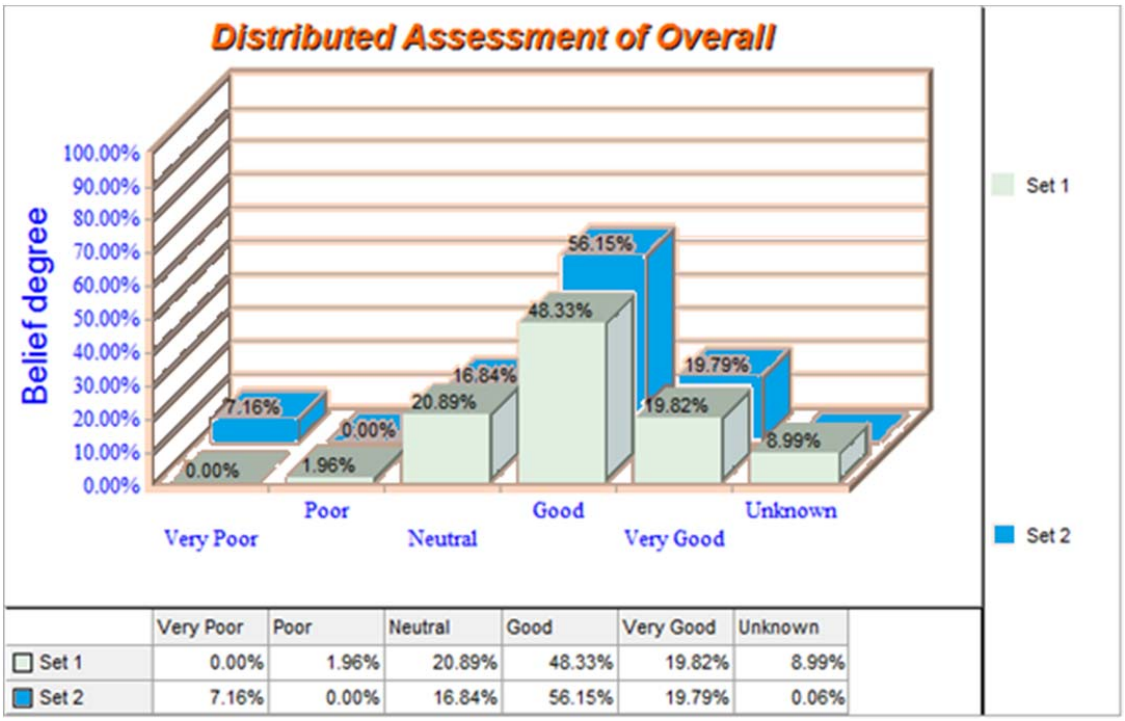
According to the procedures suggested in EBTO method, an assessment normalization process would often be applied to the raw assessments of bottom level criteria as a pre-step of further aggregation. In this study, as each of the bottom level criteria were consistently assessed through a five grades set, there was no demand of any normalization process. Accordingly, we aggregated the assessments according to the importance weighting of criteria, directly. The assessment aggregation was conducted through a window based software package named Intelligent Decision System (IDS). The IDS was developed by Yang & and Xu (2000) on the basis of the ER algorithm. It transforms the lengthy and tedious aggregation process into a program that could be conducted by the computer.

The result of the assessment aggregation process is illustrated in Figure 4-2 & 4-3. It indicates the general preference of each Soldier system for the two missions.



	Very Poor	Poor	Neutral	Good	Very Good	Unknown
Set 1	0.00%	1.05%	16.42%	39.12%	23.21%	20.21%
Set 2	26.77%	0.00%	18.23%	36.71%	18.25%	0.04%

Figure 4-2, Overall assessment of Soldier systems for patrol mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2)



	Very Poor	Poor	Neutral	Good	Very Good	Unknown
Set 1	0.00%	1.96%	20.89%	48.33%	19.82%	8.99%
Set 2	7.16%	0.00%	16.84%	56.15%	19.79%	0.06%

Figure 4-3, Overall assessment of Soldier systems for assault mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2)

The belief structure illustrated in Figure 4-2 & 4-3 indicates raw output of the aggregation process. The preference order, however, is not easy to be observed. In this case, the utility analysis procedure, as suggested in section 2.2.2, is applied to transform the distributed belief structures to numerical scores. We suppose the utility scales of the assessment grades are given as follow:

$$u(Very\ Poor) = 0$$

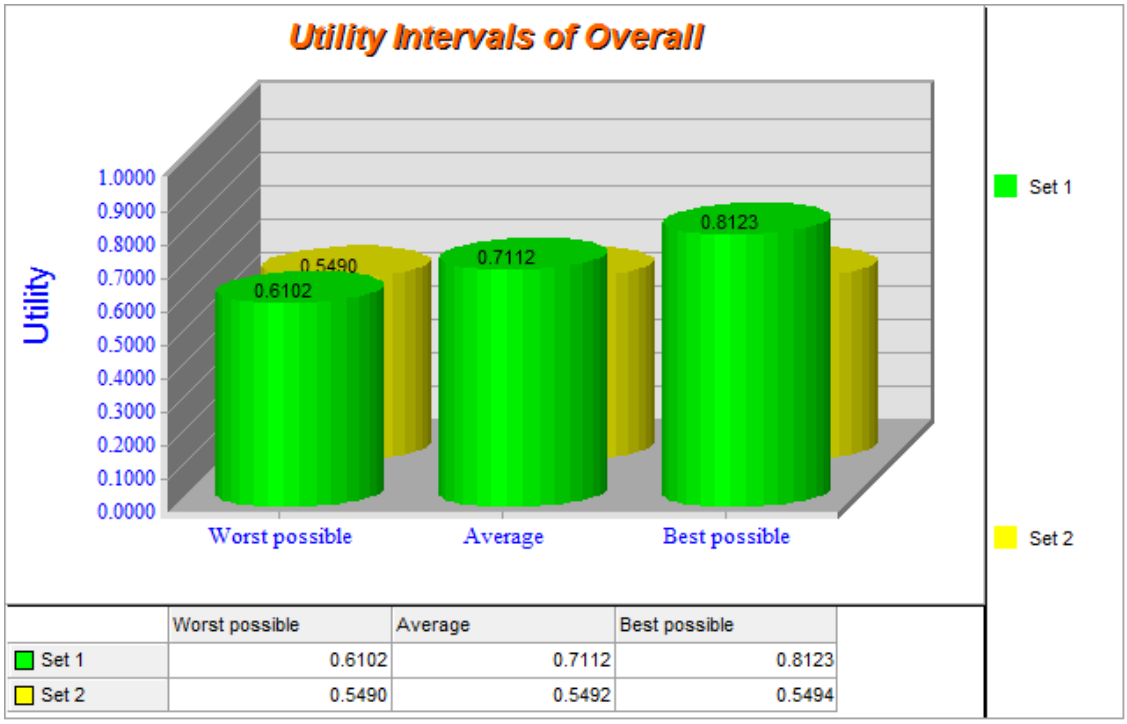
$$u(Poor) = 0.25$$

$$u(Neutral) = 0.5$$

$$u(Good) = 0.75$$

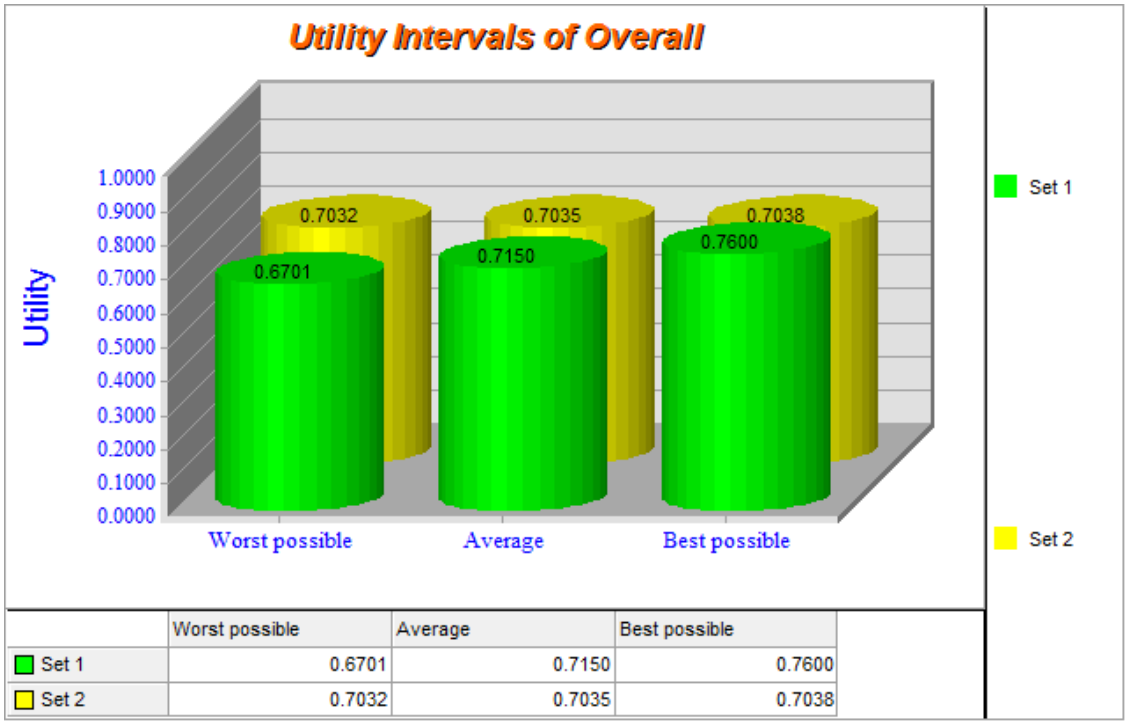
$$u(Very\ Good) = 1$$

where VP has the lowest utility value of zero and VG has the highest utility value of one. Then the utility value of the belief structures could be calculated with the formula (18)-(21). In particular, as there exists belief degree of unknown portion due to the incomplete judgment collected in the assessment process, an interval utility was obtained for each belief structure as depicted in Figure 4-4 & 4-5 below. In the worst case, the full belief degree of unknown portion could be assigned to the worst grade VP and got the lowest utility value. It stands for the lower bound of utility that a Soldier system could be performed. In the best case, the full belief degree of unknown portion could be assigned to the best grade VG and got the highest utility value. It stands for the upper bound of utility that a Soldier system could be performed.



	Worst possible	Average	Best possible
Set 1	0.6102	0.7112	0.8123
Set 2	0.5490	0.5492	0.5494

Figure 4-4, Utility interval of Soldier systems for patrol mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2)



	Worst possible	Average	Best possible
Set 1	0.6701	0.7150	0.7600
Set 2	0.7032	0.7035	0.7038

Figure 4-5, Utility interval of Soldier systems for assault mission (Set 1 indicates the Soldier system 1; Set 2 indicates the Soldier system 2)

It could be found that Solder system 1 performs dominantly better than Soldier system 2 in patrol mission, as

$u_{min}(Soldier\ system\ 1\ for\ patrol) = 0.6102 >$
$u_{max}(Soldier\ system\ 2\ for\ patrol) = 0.5494.$

But there is no strict preference between two Soldier systems in assault mission, as

$$u_{min}(Soldier\ system\ 1\ for\ assault) = 0.6701 < u_{min}(Soldier\ system\ 2\ for\ assault) = 0.7032$$

$$u_{max}(Soldier\ system\ 1\ for\ assault) = 0.7600 > u_{max}(Soldier\ system\ 2\ for\ assault) = 0.7038$$

$u_{avg}(Soldier\ system\ 1\ for\ assault) = 0.7150 \approx$
$u_{avg}(Soldier\ system\ 2\ for\ assault) = 0.7035.$

The above results provide the decision maker a valuable support to make the choice between two systems for the two missions. Regarding patrol mission, the assessment surely indicates that the Soldier system 1 is a better choice in comparison to the Soldier system 2. In terms of assault mission, however, no strict prescription could be made in determining which one is a better choice. This is due to the uncertainties of initial assessment data. According to the average utilities, Soldier system 1 holds a slight higher possibility that it is the better choice than Soldier system 2. If this possibility is not large enough to be accepted by the decision maker, the initial assessments at the bottom level criteria should be re-evaluated by the decision maker, in which more precise assessments are encouraged.

4.5 Summary

This chapter presents a case study of applying EBTO in the Soldier system decision making. The study is motivated by the concept of “soldier as a system” and the individual burden issue encountered by the operations of modern soldiers. A soldier system assessment model which constitutes five categories of criteria, namely lethality,

C41, survivability, sustainability, and mobility, is used. The model consists of 23 measurable criteria at two levels. Through the application, the methodology is demonstrated to be able to provide the holistic analysis regarding the requirements of Soldier missions, the physical conditions of Soldiers, and the capability of their equipment and weapon systems, which is critical in domain. (TRADOC, 2006).

Chapter 5 : Case study of fire emergency decision making

One typical example of decision making under extreme time pressure and dynamic shifting conditions is the decision making during fire emergency response. In this chapter, a case study of Incident Commanders' (ICs') decision making during high rise apartment fire will be reported. According to the procedures suggested in Judgment Analysis with Heuristic Modelling (JA-HM) method, we first determined the high rise apartment as the type of fire incident to be investigated, and identified the ICs' resource demand decision at the on-arrival stage of incident as the decision problem to be studied (Section 5.1, 5.2). We then formulated a collection of cues that could specify the context of the decision problem (Section 5.3). These cues were then manipulated to construct an experimental decision making task to be completed by participants (Section 5.4). Data were then collected by involving ICs completing the decision making task (Section 5.5). The collected data were finally analyzed through inferential statistics and heuristic modelling (Section 5.6).

5.1 The problem

For firefighting purposes, a high rise building is considered to be one containing floors at such a height or position, or design that external firefighting and rescue operations may not be feasible or practicable (HM Government 2008b). A more straightforward definition of a high rise building is a building with 8 floors or more. During a high rise fire, vehicular access is often difficult; travel distances, communications, and operations are extended; logistical supports are often challenging; firefighting may affect means of escape. It has been identified as one of the essential high-risk built environments of fire emergency response (Yang et al., 2009), and has been an important area of research.

Operational response to a high rise building fire has been organized into five temporal dimensions in UK FRS. They are: 1) operational pre-planning and information gathering; 2) on arrival; 3) proceeding to and establishing the bridgehead; 4) Firefighting, search and rescue; 5) Post firefighting operations.

Operational pre-planning and information gathering is carried out during normal time through inspection and site visits of FRS. The intelligence gathered highlights things that are tactically advantageous to fire crews when attending an incident (water supplies, fixed installations, ventilation systems, etc.), as well as those inherent hazards and risks such as hazardous materials, compromised settings. These pre-planned intelligence reports go to form "Site Specific Risk Information / Premise Risk Management" (SSRI/PRM) documents, and are required to be available for first attending crews. In addition, pre-planning would also make provision for a pre-determined response/attendance (PDR/PDA) that reflects the access and facilities provided for the FRS and the type of incident likely to be encountered. Normal practice across UK FRS would consider offensive internal attack on a fire to require a minimum of 2 appliances. In high rise scenarios, the standard expectation of PDA would be 4 appliances, and 1 aerial ladder platform (ALP).

Once the actual incident occurs, the primary function of the IC at the on arrival stage is to ascertain as much information as available. This would start from the en-route to the incident, when SSRI/PRM is checked through the Mobile Data Terminal (MDT) installed on fire appliance. This is followed by a visual check of the structure, as well as gathering information from building occupants and any fire control systems they may be present when arriving at the incident. During this time stage, the ICs are always engaged in a decision of

"Re-assess the requirement for further resources – being cognisant of the running time of reinforcing appliances in relation to likely escalation" (HM Government 2008b, p.19).

The decision of the amount of additional resources calling for (mostly additional fire appliance) reflects the IC's overall judgment of the scale and extent of the incident and its likely escalation. It is the type of decision that demand personal expertise and could shed light on their decision making policy in terms of the information they collected at the time. There is no departmental SOPs on this decision problem.

It is therefore, the present study determined, the ICs' resource demand decision making at the on-arrival stage of the high rise building fire incident, as the problem to be investigated. Among the four main property classes that account for high rise fires (apartments, hotels, facilities that care for the sick, and offices) (John, 2013), the high rise apartment fire is selected as the type of high rise incident to be investigated, as it is considered as the most commonly happened class of high rise fire (John, 2013).

5.2 Background

5.2.1 Incident Command System (ICS) in UK FRS

Fire and Rescue Service (FRS) in the UK has the primary role of preventing further escalation of an incident by controlling or extinguishing fires, rescuing people, and undertaking other protective measures. They may also be demanded to deal with released chemicals or other contaminants in order to render the incident site safe; or assist other agencies in the removal of large quantities of flood water; or assist ambulance services with casualty-handling; or assist police with the recovery of bodies (HM Government 2013).

A total of 48 fire and rescue services are established, currently, to cover the countrywide operations in the UK. Any individual fire and rescue service response to a fire emergency is managed by a nationwide safe and effective system, named Incident Command System (ICS). It constitutes the doctrine of the FRS in the context of operational incident management, command competence, and the functional command and control processes that flow from it.

At the organizational level, the ICS defines a taxonomy of command roles on the incident ground, ascending from front line Breathing Apparatus Wearer (BA Wearer), Breathing Apparatus Entry Control Officer (BAECO), Crew Commander (CC), to more senior Sector Commander (SC), Operations Commander (OC), Incident Commander (IC), as well as supporting positions such as Command Support (CS), Safety Officer (SO),

Runner (HM Government, 2008a). They are managed with two essential concepts, namely 'Span of Control' and 'Sectorisation'.

The concept of Span of Control indicates that the direct lines of communication and areas of involvement of any officer are limited in number to enable the individual to deal effectively with those areas. No individual should be responsible for a number of activities or operations that are difficult or impossible to give sufficient attention. Hence, once the demands placed upon an IC beyond his capability, the need of sectorisation arises. There are two main categories of sectorisation, geographical based (i.e., fire sector, search sector, lobby sector for high-rise fire) or functional based (i.e., water sector, logistics sector for the same fire).

It is the IC to exercise authority over fire service resources on the incident ground, maintaining an appropriate span of control by introducing additional roles into the incident command structure when the demands on any individual's attention become excessive. In circumstance where multi-agency Bronze, Silver, Gold system is activated, the tactical officer may be introduced to take over the incident, or allow the existing IC to remain in charge. Strategic commander, however, would not take charge of operations on the actual incident ground.

At the operational level, the ICS acknowledges that the central guidance on FRS operations is based on the 'Safe Person Concept' (Home office, 1997), which recommended that all FRS operations should be based on assessments of risk to personnel as the central factor. Operational procedures and practices are designed to promote safe systems of work. The ICs are taught to operate a process of Dynamic Risk Assessment (DRA) to take into account the continually and sometimes rapidly evolving nature of an incident, and to ensure the safe systems of work being implemented and maintained throughout an incident. The DRA process involves steps including: (1) evaluate the situation, task and persons at task; (2) select systems of work; (3) assess the chosen systems of work; (4) introduce additional controls if necessary; (5) re-assess systems of work and additional control measures.

At the command competence level, the ICS together with the National Occupational Standards (NOS) outlines the skills knowledge and understanding required by an IC and the importance of maintaining such competencies. It is the IC's leadership role to maintain shared situational awareness by effective communications, to plan and set operational priorities clear, to direct and focus activity in pursuit of objectives, and to ensure subordinates have freedom and resources to carry out their role safely within the plan.

5.2.2 Rank structure of FRS in UK

In addition to the ICS, each FRS in UK implements a rank structure that denotes the hierarchy of its "operational management", ascending from the most junior firefighter to senior managers including Crew Manager (CM), Watch Manager (WM), Station Manager (SM), Group Manager (GM), Area Manager (AM) and Brigade Manager (BM). Among them, A Crew Manager (CM) is the leading firefighter that could take charge of the watch at smaller fire stations in the Retained Service. During the incident, a CM would take charge of the crew of its fire appliance, and could take command of small scale incidents. A Watch Manager (WM) is a sub officer senior to the CM, but junior to the Station Manager (SM). A WM could take charge of the watch at larger fire stations in the Retained Service. Once an incident occurs, conveniently, the Officer-in-charge (OIC) of the first attending appliance will be the initial IC. With an attendance of only a single appliance, status quo will apply. But as further appliances attend, the function of IC may be passed on. CM and WM are therefore, the primary characters that would take command at the on arrival stage of a fire incident.

5.3 Formulation of decision profile

As the JA-HM suggested, a decision profile would be formulated by a number of cues. In the present study, the cues were determined on the basis of: (a) a review of national fire safety regulations on high rise building; (b) a review of literature on fireground decision making; (c) an analysis of the FRS departmental training documents; and (d) seven semi-

structured individual interviews with five watch managers and two crew managers of Leicestershire FRS (LFRS).

The review of national fire safety regulations on high rise building helped the investigator develop the essential background knowledge of the concepts that is unique to the high rise building (e.g., Communities and Local Government, 2007; Local Government Group, 2011). For example, it is the building regulations that a riser must be installed for buildings that exceed 11 metres. The riser is a supply system intended to distribute water to multiple levels or compartments of a building, as a component of its firefighting systems. The water supply of a high rise incident would not only depend on the source water the FRS get, but also depend on the condition of the riser of the building. A defective or otherwise inoperative dry riser will seriously effect fire service operations.

The review of literature on fireground decision making offered some reference on the type of cues could be defined. An example could be found in Klein, Calderwood & Clinton-Cirocco (1988; 2010)'s study on fire commanders' decision making, in which 9 dimensions of cues are summarized for fire incident in general.

The analysis of the FRS departmental training documents facilitates the initial checklist of the possible cues would be considered. For example, generic risk assessment for high rise firefighting has been formally documented in HM Government (2008b), in which essential hazards of high rise firefighting are identified, such as height of the building, falling objects and burning debris, wind patterns and velocity, etc. Previous incident reports would also be an important source file, like Shirley Towers incident report (Hampshire Fire and Rescue Service, 2010).

The semi-structured interviews help check the relevance of cues identified in the documents analysis, further complement potential cues, produce the presentation of the cues, and help organize the cues in the way that the ICs would perceive and comprehend.

For example, the complexity of internal layout is an essential hazard identified in the FRS training document (HM Government, 2008b). Large or complex floor layouts and a lack of information on the internal layout of the building would challenge fire crews seeking safe access and egress routes to and from the scene of the fire and may increase the risk of crews becoming disorientated or lost. It is however, one of our interviewees commented on this that "*it's rare that you got a complex layout in the high rise unless you got a big building where it's like an old factory. If it is just a square block straight up, a tower, there is no complex layout. The only time it will be slightly more complex if it was a converted factory that being made into flat*". Accordingly, the internal layout of building was determined as a constant cue to provide the decision context of the scenario.

Take another example, the cue "time of incident" has been identified as important information to be considered by ICs. It is however, in addressing the possible impact of the "time of the incident", one interviewee commented: "*let me elaborate more about the time, my experience would suggest that if the incident happened around or at the midnight, my expectation is people would be inside sleeping, if it happened in mid day, there is greater chance of being out of the occupants whether the university college or work, so the occupancy matters … the only difference between time will relate to occupancy …"* Another interviewee also commented on this: *"… so the time itself make no difference, it will make difference only when it links some particular risks, like sleep risk etc*." Accordingly, the cue "time of incident" was combined with "occupancy of the building", "self-evacuation status" and "smoke spread" as a composite cue "evacuation consideration" in formulating the decision profile.

The experience of seven interviewees ranged from years to 24 years in FRS, and from 4 years to 15 years as IC. The interviews lasted from 1 hour to 3 and half hours. Interviewees were told that the information provided would be used to develop a set of hypothetical cases that may be encountered in a high rise incident, and which require a resource demand decision to be made. Confidentiality was assured. Audio records of

the interviews were taken, and were later transcribed for the purpose of confirming the investigator's understanding developed in the interviews. Interviewees were first presented with a hypothetical case and asked to comment on its content validity, and how it could be improved. A list of cues was then presented to have their opinions on the impact of each cue. The hypothetical case and the list of the cues utilized in the interviews are kept on being revised according to the feedback obtained in the interviews which have been conducted. But cautions are also paid on this process. A cue would not be eliminated from the checklist unless the investigator is fully sure about it. All of the interviewees said that the case was realistic and was quite similar with the simulation package they utilized in training.

Finally, 19 cues were identified to formulate the decision profile. They were presented in Table 5-1 & Table 5-2. Thirteen of them are fixed constant of their values to provide the decision context of the scenario. Other six cues are manipulated with binary values as the predictors of the ICs' decision making policies. According to the argument of "composite cue" in JA-HM (described in Section 3.2.2.2), four of these six cues were composite cues that are compounded through seven more basic information variables.

In addition, as the computational models of heuristics formalized in the AT tradition is principally designed for tasks in which all predictors are binary to simulate a binary coding process of how DMs simplify their perceived information, the six varied cues were defined with binary values here. The representation of cue values is through linguistic phrases and visual image, which are aligned with the real circumstance.

An example decision profile that was produced from these cues was presented in Figure 5-1 below. The complete explanation of each cue and their impact on the ICs' resource demand decision are elaborated in Appendix A. The interview guide could be found in Appendix B.

Table 5-1, Six varied cues

Cues	**Positive Features (coded as value 0)**	**Negative Features (coded as value 1)**
Horizontal fire spread	black smoke rolling out of the left balcony	black smoke rolling out of the left balcony and its two adjacent windows
Vertical fire spread	Fire is visible behind the balcony	Fire is visible on the balcony with Large flames flickering up to the upper balcony level
Floor level	On the first floor	On the eighth floor
Wind effect	Wind is light, and heavy smoke is spread upwards and to the east.	Wind is strong blowing heavy smoke upwards and to the east
Evacuation consideration	• Spring afternoon; • Normal residential with elderly people reported; • The building has a 'stay put' policy. However, residents around the fire flat seems well self-evacuated, crowded on the ground; • AFD suggests a number of floors above are affected by the spread smoke.	• Summer daybreak; • Normal residential with elderly people reported; • The building has a 'stay put' policy. And due to the daybreak time as well, the self-evacuation seems quite limited at the moment. Elderly people reported on the floors above; • AFD suggests a number of floors above are affected by the spread smoke.
Search & Rescue	All persons accounted for	One family is reported, one couple, three children.

Table 5-2, Thirteen constant cues

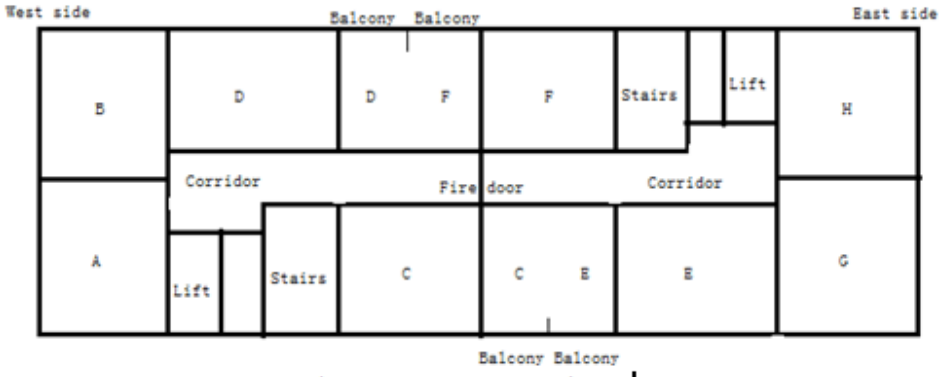
Cues	**Features**
Temperature	Spring afternoon and summer daybreak, both of which reflect a normal temperature
Location	Near city center
General construction	Built in 1960s, 11 floors apartment building, pre-fabricated concrete construction
Floor layout	Identical and available
Lift	Two lifts, both are not firefighting proved, but has a fireman's switch
Staircase	Two protected staircases
Ventilation system	Mechanical ventilation could be found in corridors and staircases
Automatic Fire Detection (AFD) installed in the building	Conventional, zoning the building by floors
Water supply	Sufficient
Internal maintenance	Good
Pre-determined Attendance (PDA) of the building	4 appliances, 1 Aerial Ladder Platform (ALP)
Cause of fire	Unknown
Falling object	Burning debris is observed falling out.

Your pre-knowledge of the scenario cases followed:

It is a high rise apartment near city centre, normal residential with elderly people reported. It was built in 1960s, 11 floors, pre-fabricated concrete construction (figure 1). Floor compartmentation plan is illustrated in figure 2. It is symmetrical layout with one protected staircase and one lift in both sides of the building. The lift is not fire fighting proved, but has a fireman's switch. Mechanical ventilation could be found in corridors and staircases. Eight flats are in each floor, labelled from 'A' to 'H' in figure 2. The AFD of the building is conventional, zoning the building by floors. The water supply of the building area is sufficient. The maintenance inside is good. The PDA for this building is 4 appliances, 1 ALP.



Building profile



Floor compartmentation layout

Trial No.5

It is the spring afternoon of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony and its two adjacent windows, on the eighth floor (as circled in picture below). According to the floor layout, they belong to the same flat but cross two compartmentations. Fire is visible from the balcony. Burning debris is observed falling out. Wind is light, and heavy smoke is spread upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground. The cause of fire remains unknown. Evidence suggests all persons accounted for.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
... ...
10, if you are absolutely confident about your assessment.)



Figure 5-1, An example of decision profile

5.4 Construction of decision making task

As elaborated in JA-HM method (Section 3.2.3), a decision making task is an experimental task that would enable the investigator to collect the data of ICs' decisions when they are facing different scenarios. Here, through manipulating the 6 varied cues with binary values identified in the last section, there could be 64 different hypothetical decision profiles produced (i.e. 2^6=64, the complete population of decision profiles). A decision making task would be constructed by involving a sequence of different decision profiles from the complete population through a sampling plan. The sampling plan specifies 1) the number of decision profiles to be included in the decision making task; and 2) the distributions and intercorrelations of cue values of the decision profiles. By involving individual participant or a group of participants completing a decision making task, their decision making policies could be revealed through heuristic modelling and statistical analysis of the decision data.

According to the procedures suggested in JA-HM, a pilot test on 3 ICs from LFRS revealed that participants would be willing to engage in a task that demands less than 15 minutes. The time would only be able to enable each participant to complete around 10 cases, which is far from the ideal number of cases for any meaningful statistical estimations of decision making policy of the individual.

Accordingly, the present decision making task was constructed as follows. Firstly, the complete population of decision profiles was included in the decision making task (Hogarth & Karelaia, 2005). This yielded 64 different cases of decision profile. Secondly, 16 cases were randomly selected from the decision profile population and included in the decision making task to permit the assessment of decision inconsistency (Cooskey, 1996). Altogether this yielded 80 cases in a decision making task. The 80 cases were then randomly ordered and divided into 8 sub-tasks in turn. Each sub-task involved 10 of the cases and was to be completed by an individual participant. It is therefore, a decision making task constructed in the present study would be completed by a group

of 8 participants. The 16 replicated cases would be randomly completed by the 8 participants of the group to indicate the intra-group consistency of their decisions.

Consequently, the consideration of cue distributions and intercorrelations in the present decision making task were simplified (uniform distribution and zero intercorrelations). The fact that little is known about the cues distributions and intercorrelations in the real incidents means that for now, little can be said about how this may affect the representativeness of the cases (Brunswik, 1956). Indeed, a similar sampling plan in terms of cues distribution and intercorrelations, orthogonal design, is common in research using JA (Dhami, Hertwig, & Hoffrage, 2004). Brehmer & Brehmer (1988) suggested a possible remediation for orthogonal set by carefully eliminated those unrealistic or atypical cases. The present study followed this suggestion by carefully defining the representation of those six varied cues to ensure any combination of them is not impossible.

The cases of the decision making task and the division of the sub-tasks are presented in Appendix C, with the cue values as coded in Table 5-1.

Weighting task

In addition to the decision making task, a direct weighting task was designed to capture ICs' explicit decision making policy (Dhami & Ayton, 2001; Cook & Stewart, 1975). Participants were asked to scale the six varied cues according to the relative importance that they attached to them during their judging of the cases. A scale of 10 indicated that the cue was so important that it completely dominated their resource demand decision. A scale of 0 indicated the cue had absolutely no impact on their resource demand decision.

5.5 Participants and Data collection process

Each sub-task generated from the decision making task was followed by a weighting task in a 15-page booklet format (see a sample booklet in Appendix D). The booklet started

from one page cover letter that introduce the study, guaranteed respondents anonymity, and requested participation.

In the sub-task of resource demand decision, participants were instructed to respond to the hypothetical cases by firstly determine the number of additional fire appliances they would like to request. The more fire appliance being requested, the more critical of the situation they judged. Participants were then asked to indicate how certain they were that they had made the appropriate decision, based on the information provided, on an 11-point scale. Zero on the scale represented “absolutely no confidence” and ten on the scale indicated “absolutely confident”. The rating here were requested for each case individually rather than for the set of 10 cases overall, as profiles differed in terms of the information provided (i.e., on the cue features) and ICs may feel more confident assessing on one case than on another.

Participants were instructed to complete the cases individually, to complete at their own pace as they would be in normal practice, and not to return to cases which had been completed. Participants were also asked to specify what further information, if any, they would have liked in the set of cases which would help them to make decision. Participants’ demographic details, namely their current role, number of years in the FRS, number of years as IC, and IC qualification level held were also requested.

The distribution of the booklets was conducted by drop-in visits of fire station in Leicestershire FRS (LFRS). As introduced in Section 5.2, CM and WM are the primary characters that would take command at the initial stage of a fire incident. They would consist of two groups of participants of the present study. To ensure each group completing one decision making task (80 cases), 8 WMs and 8 CMs were required to be involved, respectively. Each of them completed one booklet. This was achieved by distributing 26 booklets in total, with 17 booklets fully completed and returned (One of the returned booklets was excluded because the participant did not answer the questions as the instruction required).

The participants of WM group, on average, held 18.3 years of experience in the FRS (M=18.3, SD=4.3) and 8 years of experience as IC (M=8.0, SD=1.7). The participants of CM group, on average, held 12.9 years of experience in the FRS (M=12.9, SD=4.4) and 4.4 years of experience as IC (M=4.4, SD=2.6).

5.6 Analysis and results

5.6.1 Raw data and data alignment

The raw data collected were integral scores from 0 to 10 (M=2.69, SD=2.427 for WMs group; M=1.66, SD=1.242 for CMs group), indicating the number of additional fire appliances that participant ICs would like to request, in response to the scenarios described in the hypothetical cases (Appendix C & D). However, according to our field investigation, ICs could have their different personal preference towards the amount of resources in responding to the same incident, which indicated that, the extent of difference between integral scores among different ICs, were different. Therefore, the raw data were first recoded into ordinal scales to specify the decision variability expressed by the participant ICs (Section 3.2.5.1 Data alignment). The decision variability here is defined by the number of different integral scores utilized by an individual participant IC in the cases he judged.

Among 16 participants, 6 of them (4 WMs, 2 CMs) judged the cases with solely two integral scores (2-point scale), 4 of them (1 WMs, 3 CMs) judged the cases with three different integral scores (3-point scale), 4 of them (3 WMs, 1CMs) judged the cases with four different integral scores (4-point scale), and another 2 CMs judged the cases with solely one integral scores (1-point scale), in which no variety (discrimination) was expressed.

The data alignment was conducted through two recoding processes, yielding two sets of data.

According to the raw data, the minimum number of fire appliances requested by each participant indicated their decision baseline. Accordingly, the first recoding process

applied was dichotomizing each participant' decisions of resource demand according to its correspondent decision baseline. The recoding would depict the ICs' binary classification process of decision making. For those participants deciding the cases with 2-point scale, the process is straightforward. The lower point was treated as baseline and recoded as 0. And the higher point was recoded as 1. For those participants deciding the cases with 3-point or 4-point scale, their baseline point were recoded as 0, and their non-baseline points were recoded as 1. This first recoding process yielded a binary dataset, in which 0 indicates the judgment of less resources demanding scenario and 1 indicates the judgment of severe resources demanding scenario.

However, for those participants deciding the cases with 3-point or 4-point scale, we acknowledged that the dichotomizing process would lead to the loss of information about variance among those non-baseline points (MacCallum et al., 2002). It was therefore, the second recoding process was applied to capture the lost variations. The recoding depicted the variations of those non-baseline points by aligning them into ordinal categories. For those participants deciding the cases with 3-point scale, their data were recoded into 0, 1, and 2 in order, where 0 indicates the judgment of the least resources demanding scenario, and 2 indicates the judgment of the most severe resources demanding scenario. For those participants deciding the cases with 4-point scale, their data were recoded into 0, 1, 2, and 3 in order, where 0 indicates the judgment of the least resources demanding scenario, and 3 indicates the judgment of the most severe resources demanding scenario. The recoding depicted the ICs' ranking process of decision making.

To be noted, for those participants deciding the cases with one point scale, their decision data were eliminated from our analysis, as no decision variation was expressed by them and their data was unable to indicate their decision making policies.

5.6.2 Statistical characteristics of the data

The following statistical characteristics of the data were analyzed through IBM SPSS Statistics 21.

Intra-group consistency

Following with the data alignment process, two sets of consistency measures were calculated here.

The intra-group consistency on the set of replicated cases was measured through the Spearman's rank correlation. For the replicated cases applying the first recoding process, a non-significant correlation of 0.313 was found for WM group (2-tailed $p>0.05$, $n=16$). A non-significant correlation of 0.258 was found for CM group (2-tailed $p>0.05$, $n=7$). For the replicated cases applying the second recoding process, a non-significant correlation of 0.253 was found for WM group (2-tailed $p>0.05$, $n=16$). A non-significant correlation of 0.54 was found for CM group (2-tailed $p>0.05$, $n=7$).

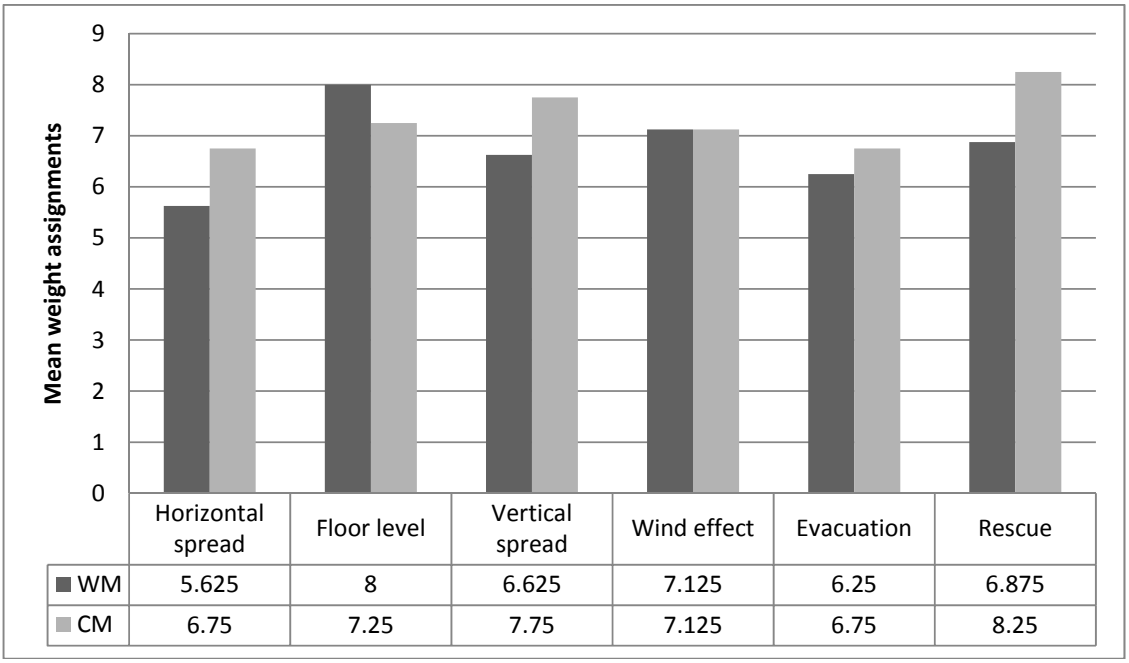
ICs' post-assessment confidence

Mean post-assessment confidence ratings made over the cases were calculated for each IC. For the whole sample, the ratings ranged from 6 to 10 ($M = 8.19$, $SD = 1.00$). For WM group only, mean ratings ranged from 6 to 9.10 ($M = 8.09$, $SD = 1.02$). For CM group only, mean ratings ranged from 7 to 10 ($M = 8.29$, $SD = 1.03$). In order to examine the relationship between ICs' mean post-decisional confidence and their experience as IC, a Kendall's tau-b correlation was computed. A non-significant correlation of -0.11 was found (two-tailed $p>0.05$, $n=15$).

For each group data, a Spearman's rank correlation was computed to examine the relationship between the decisions made on the cases and the correspondent post-decisional confidence ratings. For the WM data applying the first recoding process, a non-significant correlation of -0.10 was found (two-tailed $p>0.05$, $n=80$). For the CM data applying the first recoding process, a significant correlation of -0.35 was found (two-tailed $p<0.01$, $n=60$). In terms of decision data applying the second recoding process, a non-significant correlation of -0.12 was found for WM group (two-tailed $p>0.05$, $n=80$), and a significant correlation of -0.36 was found for CM group (two-tailed $p<0.01$, $n=60$).

ICs' subjective weights of cues

The mean subjective weights of cues assigned were compared between the WM group and CM group (Figure 5-2). There was no significant difference found according to the Wilcoxon signed ranks test (two-tailed $p>0.05$).



	Horizontal spread	Floor level	Vertical spread	Wind effect	Evacuation	Rescue
WM	5.625	8	6.625	7.125	6.25	6.875
CM	6.75	7.25	7.75	7.125	6.75	8.25

Figure 5-2, Comparison between mean subjective weights of cues assigned by WMs and CMs

5.6.3 ICs' decision making policies

Ideally, fire ICs are expected to use all of the relevant information, and weight and combine it appropriately. As their decisions have significant consequences in terms of property and life loss. An example of such expectation could be identified in the FRS operational risk philosophy following an unacceptable level of firefighter deaths occurred during the late 1980s and early 1990s. It is highlighted departmentally that the benefits of proceeding with a task must be weighed carefully against the risks. "In a highly calculated way, firefighters will take some risk to save saveable lives; may take some risk to save saveable property; will not take any risk at all to save lives or properties that are already lost" (HM Government, 2008a, p.65).

On the other hand, however, decision making in fire emergency is characterized by extreme time pressure. It is unreasonable to expect fire ICs behave in an optimal way. Numerical literatures have suggested that people tend to use fewer information and simple non-compensatory strategies under conditions of time pressure (e.g., Payne, Bettman, & Johnson, 1993; Rieskamp & Hoffrage, 1999; 2008).

Accordingly, to capture the decision making policies applied by the DMs, seven computational models of heuristics are hypothesized and competitively tested in the following. The seven models are Franklin's rule, Dawes' rule, Matching Heuristic (MH), Matching Heuristic for 3-point ranking (MH-3R), Weighted Additive linear model for 3-point ranking (WADD-3R), Matching Heuristic for 4-point ranking (MH-4R), and Weighted Additive linear model for 4-point ranking (WADD-4R). All of them are applying the cue validity as the index in their searching of information.

5.6.3.1 Information search rule

As elaborated in literature and methodology chapters (Section 2.5 & Section 3.2), heuristic modelling applied in this study involves a search and stopping process to address the process dynamics of people's decision making (e.g., information search, shift in information emphasis or judgment policy). It has been acknowledged that the search of information has two main categories: random search and ordered search (Gigerenzer, Dieckmann, & Gaissmaier, 2011). Random search is ecologically rational when individual has little knowledge about relative importance of cues. In contrast, ordered search is mostly happened when individual has some knowledge about the cues. Accordingly, ordered search of ICs is assumed in this study. In particular, we applied cue validity as the index of the ICs' ordered search.

The search by cue validity assumes that a DM goes through cues in the descending order of their validities. The validity of cue i is calculated as:

$$v_i = \frac{R_i}{R_i + W_i}$$

Where R_i is the number of cases where the value of cue i is 1 and the value of decision are 1 or 1+, and W_i is the number of cases where the value of cue i is 1 but the value of decision is 0. The validity, in binary classification situation, indicates the predictive power of a cue. The validity, in ranking situation, indicates the capability of a cue in narrowing down the possible categories (Berretty et al., 1999). A calculation example in binary classification situation is demonstrated in Figure 5-3 below.

No. of the cases / Cue Value / Decision	0	1
0	7	**8**
1	5	**12**
Validity of the cue:	12/(8+12)=0.6	

Figure 5-3, Calculation of the cue validity: supposing a binary cue (0-1) and a dataset containing binary decision data of 32 cases in total

The cue validity was first proposed in Gigerenzer, Hoffrage, Kleinbolting (1991)'s PMM model, arguing that people constructed a hierarchy of cue validities in their reference class through the operation of some frequency-encoding mechanism. Given the large literature on people's learning and storing frequency information of event co-occurrences in the environment (e.g., Gigerenzer, 1984; Hasher & Zacks, 1979; 1984; Hintzman, Nozawa, & Irmscher, 1982), this argument seems reasonable, and the searching cues by their validities has been the most widely acknowledged and utilized search rule in AT literature. In a sense, it is a frugal ordering, as it does not attempt to grasp the dependencies between cues. The frugality makes it be able to be estimated from a small sample of objects and cues (Czerlinski et al., 1999), and be more robust when generalizing to new sample, in comparison to optimal index (Martignon & Hoffrage, 2002). Its prediction power has also been demonstrated in Brighton (2006).

5.6.3.2 Seven computational models of heuristics

a) Franklin's rule

This is a model for binary decision data. The model multiplies each cue value by its weight and sums the total. Then, a decision would be made according to the comparison between the sum and a threshold value. If the sum is equal or greater than the threshold value, then a decision value of 1 is predicted. If not, then a decision value of 0 is predicted.

The cue weight here refers to cue validity defined above.

The threshold value is calculated through following steps (Dhami & Ayton, 2001; Dhami, 2003; Broder, 2002; MacCallum et al., 2002):

1) Calculating the sum value of each case in a dataset through the model, and sorting the sum values in the descending order;

2) Calculating the base rate between the value 1 and value 0 of the decision data of the dataset (i.e., if there are 20 decision data of 1 and 12 decision data of 0 in a dataset containing 32 cases in total, then the base rate is 20:12, see Figure 5-4);

3) Splitting the sum values order produced in step 1 into two portions, according to the base rate produced in step 2 (i.e., one portion contains 20 higher sum values, another portion contains 12 lower sum values, see Figure 5-4);

4) The value on the splitting boundary is set as the threshold value of the model for the dataset (i.e., the 20^{th} sum value in the descending order produced in step 1, see Figure 5-4).

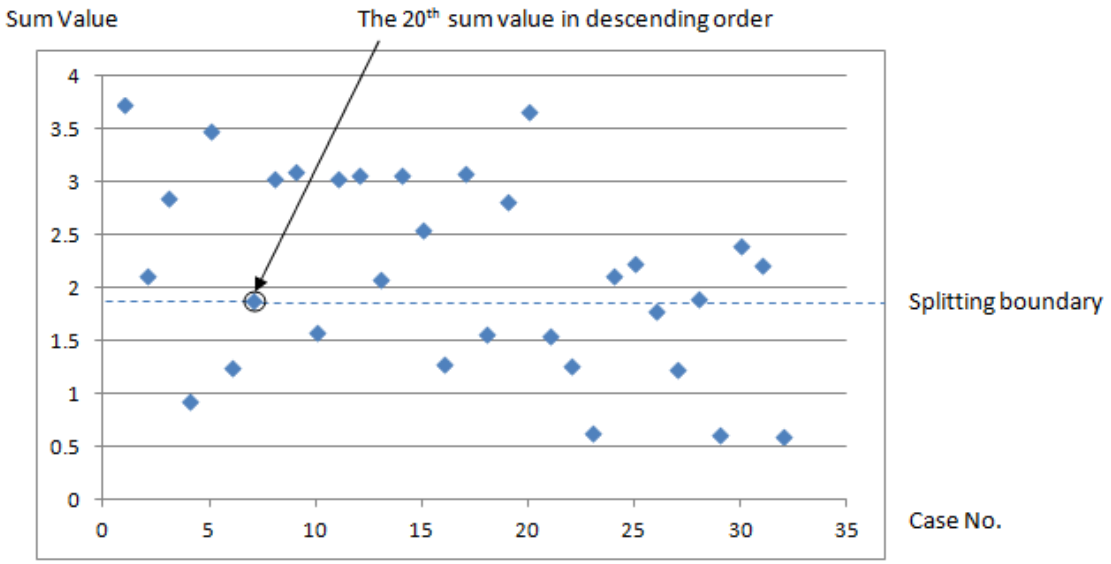


Figure 5-4, An example of defining the threshold value of the Franklin's rule for a dataset with 32 cases

The Franklin's rule is drawn from the Franklin's moral algebra (Franklin 1772/1987). Compared to the traditional regression model, It ignore the inter-dependence of cues, and is therefore faster and more psychological plausible than the regression model. It has been commonly used as a prototypical model in the AT tradition to reflect the classical viewpoint of rational inference that involves comprehensive search and compensatory integration, especially for those empirical studies where case-to-cue ratio was not sufficient for regression analysis (e.g., Rieskamp & Hoffrage, 2008; Rieskamp & Otto, 2006; Gigerenzer et al., 1999). An example of the implementation of Franklin's rule is demonstrated in Figure 5-5 below.

b) Dawes' rule

This is a model for binary decision data. The model ignores the weights of cues and simply adds up the cue values. Then, a decision would be made according to the comparison between the sum and a threshold value. If the sum is equal or greater than the threshold value, then a decision value of 1 is predicted. If not, then a decision value of 0 is predicted.

The cue weight here refers to cue validity defined above.

The definition of threshold value follows the same procedure of the calculation in Franklin's rule.

The Dawes' rule is named after Robyn Dawes, a judgment and decision making researcher, who conducted precursory work on demonstrating that the less computational demanding model could be excellent approximations to regression models in terms of descriptive and predictive validity (e.g., Dawes 1979; Dawes & Corrigan, 1974). The model is fast (it does not involve much computation), but not frugal (it looks up all cues). It is also another prototypical model utilized in AT literature (e.g., Gigerenzer & Goldstein, 1996). An example of the implementation of Dawes' rule is demonstrated in Figure 5-5 below.

Cues	Cue features	Cue value	Cue validity	Decision made by IC
Horizontal fire spread	black smoke rolling out of the left balcony and its two adjacent windows	1	0.64	1
Floor level	Eighth floor	1	0.95	
Vertical fire spread	Fire is visible from the balcony	0	0.85	
Wind effect	Wind is light	0	0.63	
Evacuation consideration	The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground.	0	0.61	
Search & Rescue	Evidence suggests all persons accounted for	0	0.67	

Decision models		Prediction
Franklin's rule	1*0.64+1*0.95+0+0+0+0=1.59 < 1.88 (threshold value)	0
Dawes' rule	1+1+0+0+0+0=2 < 3 (threshold value)	0
MH (K=2)	Search the 1st rank ordered cue, floor level. Value 1 is identified. Then stop search and make prediction 1.	1

Figure 5-5, Calculations of Franklin's rule, Dawes' rule, and MH where K=2, on an example case

c) Matching Heuristic (MH)

This is a model for binary decision data. For each case, the model searches through K of the available cues in rank order of their weights, looking for a value 1 on each cue. If a value 1 is found, the model stops searching and a decision value of 1 is predicted. Otherwise, the model searches through the value of the next rank ordered cue. The model continues this procedure until K cues being searched. If by this time no value 1 on each cue has been found, a decision value of 0 is predicted. For illustrative purpose, Figure 5-6 shows the process of MH where K=2.

The cue weight here refers to cue validity defined above.

To determine the K, the fit of the model with all possible K value (1-6 in the present study) is to be systematically tested on the modelling set. The K value with the best fit in terms of percentage of correct predictions is chosen. In case that two or more K value has the same fit, the smaller value is chosen. As the smaller of the K value, the more fast and frugal of the model is.

The Matching Heuristic was first proposed by Mandeep K. Dhami in her empirical studies of professional magistrates' decision making (Dhami 2003; Dhami & Ayton, 2001). It is fast, as the model predicts decision without complex computation. It is frugal, as it searches through a small subset of the cues. It is non-compensatory, as a decision is based on the value of one cue, and would not be altered by values of other cues. It is also highly flexible and context-sensitive, as different cues can be used to make decisions on different cases. It belongs to the fast-and-frugal tree (Gigerenzer & Gaissmaier, 2011), and specifically designed for binary classification task (Dhami & Ayton, 2001). An example of the implementation of MH (where K=2) is demonstrated in Figure 5-5 above.

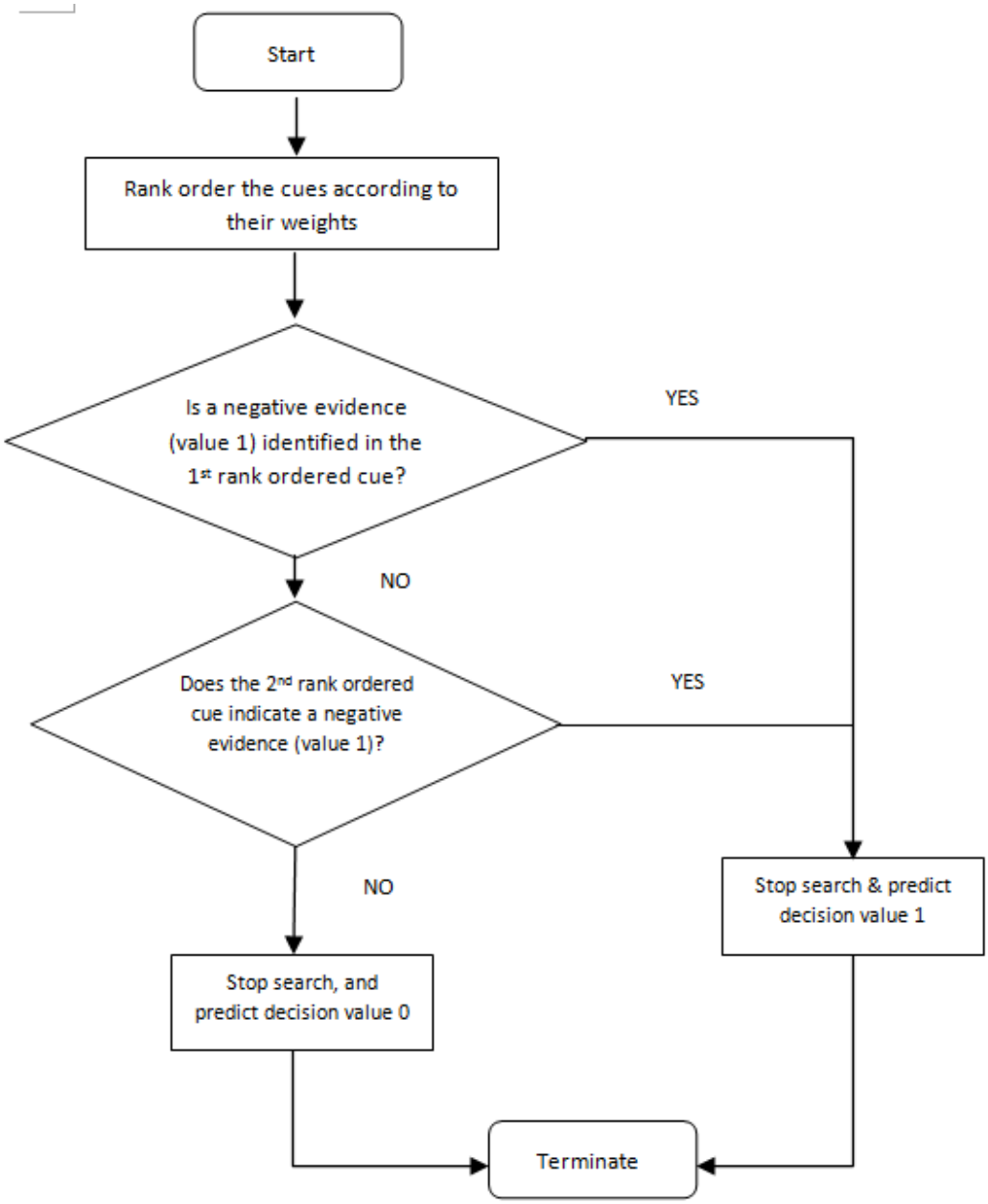


Figure 5-6, Flowchart of Matching Heuristic (K=2)

d) Weighted Additive linear model for 3-point ranking (WADD-3R)

This is a model for 3-point ordinal decision data. The model multiplies each cue value by its weight and sums the total. Then, a decision would be made according to the comparison between the sum and two threshold values. If the sum is equal or greater than the higher threshold value, then a decision value of 2 is predicted. If the sum is

equal or greater than the lower threshold value but smaller than the higher threshold value, then a decision value of 1 is predicted. If the sum is smaller than the lower threshold value, then a decision value of 0 is predicted.

The cue weight here refers to cue validity defined above.

The threshold value is calculated through a similar procedure of calculation in Franklin's rule:

1) Calculating the sum value of each case in a dataset through the model, and sorting the sum values in the descending order;

2) Calculating the base rate between the value 2, 1 and 0 of the decision data of the dataset;

3) Splitting the sum values order produced in step 1 into three portions, according to the base rate produced in step 2;

4) The values on two splitting boundaries are set as the threshold values of the model for the dataset.

e) Weighted Additive linear model for 4-point ranking (WADD-4R)

This is a model for 4-point ordinal decision data. The model multiplies each cue value by its weight and sums the total. Then, a decision would be made according to the comparison between the sum and three threshold values. If the sum is equal or greater than the highest threshold value, then a decision value of 3 is predicted. If the sum is equal or greater than the moderate threshold value but smaller than the highest threshold value, then a decision value of 2 is predicted. If the sum is equal or greater than the lowest threshold value but smaller than the moderate threshold value, then a decision value of 1 is predicted. If the sum is smaller than the lowest threshold value, then a decision value of 0 is predicted.

The cue weight here refers to cue validity defined above.

The threshold value is calculated through a similar procedure of calculation in WADD-3R.

The models WADD-3R and WADD-4R developed here, drew the classical weighted linear model for preferential choice (e.g, Keeney & Raiffa, 1993; Payne et al., 1993; Zurada, 1992) to multiple points ranking data. It reflects the behavior of comprehensive search and compensatory integration in the multiple alternatives ranking task. This is similar to the studies conducted in Berretty, Todd, & Martignon (1999) and Blythe, Todd, & Miller (1999), where weighted additive linear model is utilized in multiple categorization problems. The reason that we don't use the regression model here, such as logistic regression, is the same as we mentioned in the Franklin's rule section. But we don't consider the present model to be especially limited in reflecting the behavior we want to investigate (Dhami & Ayton, 2001; Gigerenzer & Goldstein, 1996).

f) Matching Heuristic for 3-point ranking (MH-3R)

This is a model for 3-point ordinal decision data. For each case, the model searches through K of the available cues in rank order of their weights, looking for a value 1 on each cue. If a value 1 is found, the model would keep on searching the next rank ordered cue as a confirmation process. If another value 1 is identified of the confirmation process, the model stops searching and a decision value of 2 is predicted. If no value 1 is identified of the confirmation process, the model stops searching and a decision value of 1 is predicted. If by the time that K cues have been searched and no value 1 on each cue has been found, a decision value of 0 is predicted. For illustrative purpose, Figure 5-7 shows the process of MH-3R where K=2.

The cue weight here refers to cue validity defined above.

The MH-3R is built on the basis of MH. Its K value is therefore, equal to the K value of the correspondent MH.

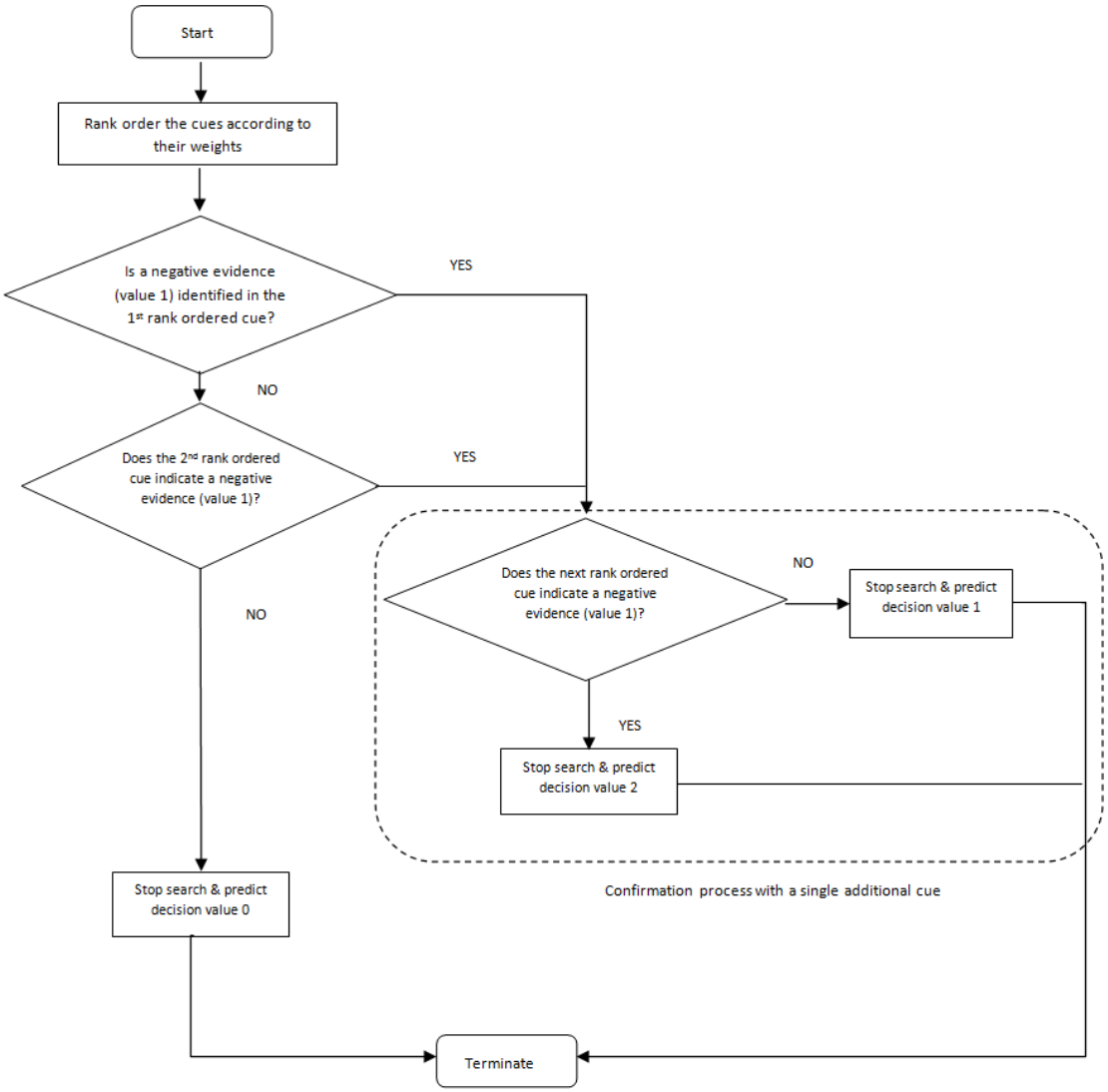


Figure 5-7, Flowchart of MH-3R (K=2)

g) Matching Heuristic for 4-point ranking (MH-4R)

This is a model for 4-point ordinal decision data. Just like the MH-3R that is built on the basis of MH, the MH-4R is further built on the basis of MH-3R. It involves a two cues confirmation process after the first value 1 of cue is identified.

It predicts a decision value of 3, if both of the two additional searched cues are identified to be of value 1. It predicts a decision value of 2, if only the first additional searched cue is identified to be of value 1. It stops searching and predicts a decision

value of 1, if no value 1 is identified in the first additional searched cue. And If by the time that K cues have been searched and no value 1 of cue being found, the model predicted a decision value of 0. For illustrative purpose, Figure 5-8 shows the process of MH-4R where K=2.

The cue weight here refers to cue validity defined above.

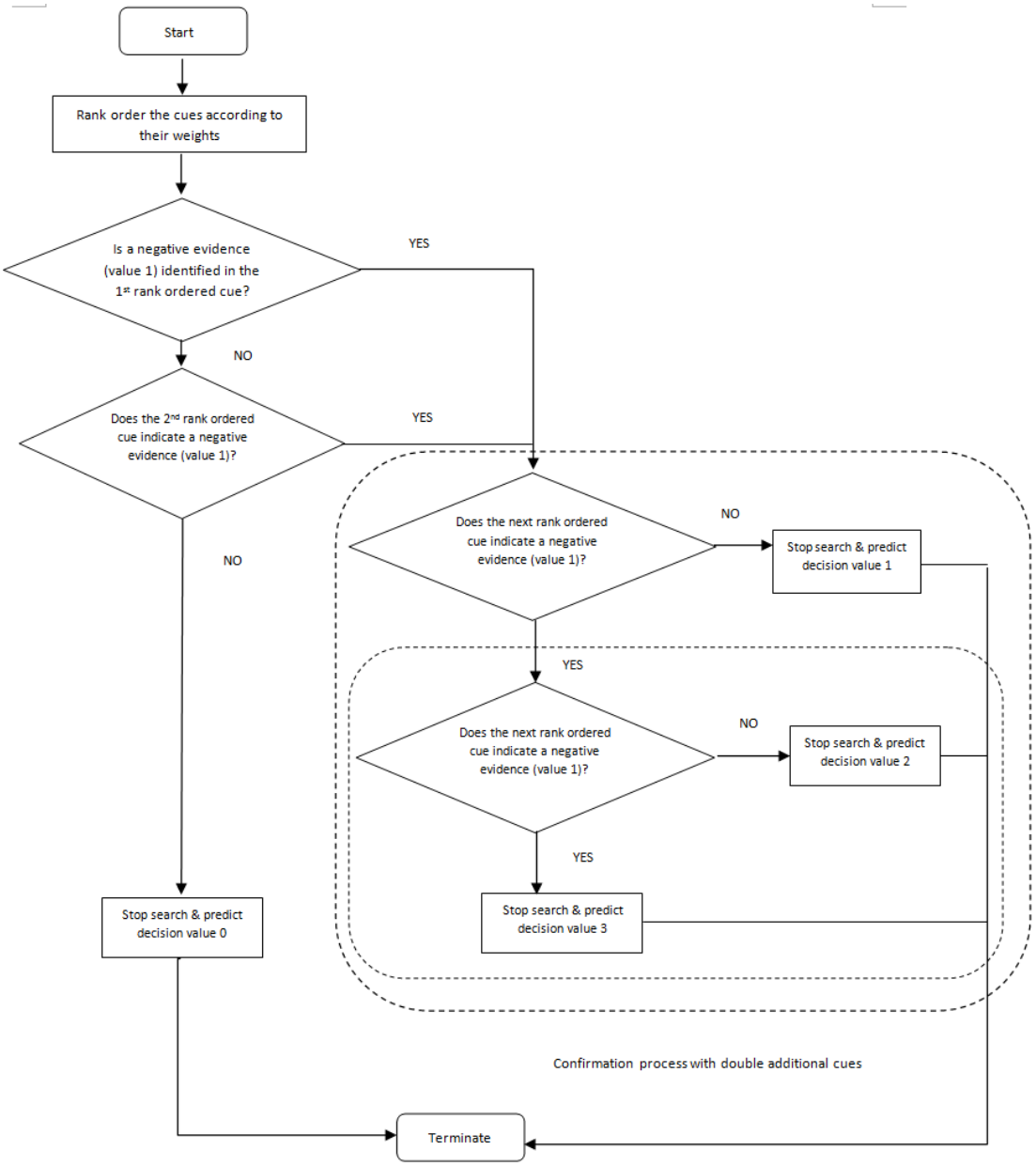


Figure 5-8, Flowchart of MH-4R (K=2)

The MH-3R and MH-4R were the further developments of MH drawing on three essential literatures of heuristic study.

The first literature is Tversky's (1972) classic elimination-by-aspects (EBA) model. The original model follows a multi-stage process that pares down a large set of alternatives by eliminating them on the basis of probabilistically selected criteria. It has been adapted in the AT literature to resolve multi-alternative choice tasks, such as Hogarth & Karelaia (2005)'s Deterministic EBA (DEBA), Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer (2010)'s Elimination by Recognition, as well as categorization task involving multiple categories of cue values (Berretty et al., 1999). It is however, that the ranking task being investigated in the present study remains essential difference from the problems formalized in above literature. To develop an appropriate model for the present data, further adaption of EBA is required.

The second literature that sheds light on our new developments of MH came from the most recent development of AT. In their investigation of stopping rules of heuristic searching process, a notion of confirmation process has been proposed (Gigerenzer, Dieckmann, Gaissmaier, 2011). It has been acknowledged that people sometimes look for confirming evidence so that their conclusion is doubly sure (Bruner et al., 1956). Such process is remarkably robust and insensitive to cue ordering, and therefore, would work well in situations in which the decision maker knows little about the validity of the cues and the costs of cues are rather low (Karelaia 2006). The number of cues to be confirmed in such confirmation process is, in principle, open to all possibilities. For instance, Meyers-Levy's (1989) work on advertisement and consumer behaviour suggests that men stop earlier than women do when looking for cues regarding potential purchases. Rieskamp & Dieckmann (2011) has also investigated a take-two heuristic that looking for one cue confirmation.

The third literature is the principle held in AT that any new heuristics should be built from the parts of the old ones rather than from scratch, a notion of nested construction of heuristic (Gigerenzer & Todd, 1999). It is a reasonable principle if we believe the

heuristics conceptualized in AT describe the human mental process at the algorithm level (Brighton & Gigerenzer, 2008). In this light, an alternative heuristic could be seen as the natural response to the additional circumstance, information, or knowledge, on top of the existing one. Examples could be found in the constructions of take-the-best heuristic and Elimination-by-Recognition heuristic, in which the human recognition was taken as the first processing step of the new heuristics (Gigerenzer & Goldstein, 1999; Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010).

5.6.3.3 Competitive testing of models of heuristics

The following calculations were conducted through Matlab programming. The scripts are attached in Appendix G.

a) Prediction accuracy of models on the binary decision data

Franklin's rule, Dawes' rule, and Matching Heuristic were used to predict the binary decision data.

As suggested in JA-HM (Section 3.2), 10-round cross-validation were implemented (Dhami, 2003). The process yielded 10 randomly divided (50%/50% split) derivation and validation sample of WMs data and CMs data, respectively. In each round, the cue validities, K value of Matching Heuristic, and threshold values of Franklin's rule & Dawes' rule were first estimated using the derivation sample, and then the model fit of each heuristic was calculated with the derivation sample and validation sample, respectively. As stated in the methodology chapter, we refer to the percentage of decisions correctly predicted by the models as "fit". The binary nature of the decision data implies that any valid model should be expected to perform better than chance (i.e. predict more than 50% of the decision data).

In practice, the data of replicated cases need to be eliminated before each round of cross-validation process. This would finally yield 64 records of WMs data, and 53 of CMs data to be modeled from the original 80 records of WMs data and 60 records of CMs data (see Section 5.6.2). Due to the low intra-group decision consistency (see Section

5.6.2), and the fact that the replicated cases were completed by different participants, any single elimination process in terms of those data of replicated cases would yield an idiosyncratic effect on the following data analysis. It was therefore, the elimination of replications process in our study was executed 10 rounds. Each round of elimination process was bounded together with one round of cross-validation process. We believe through this plan, any idiosyncratic effect of replication elimination would be avoided.

Table 5-3 below show the results of the each model in predicting the decision data on derivation sample and cross-validation sample, respectively.

Table 5-3, Mean percentage of binary decision data predicted correctly by the models

	Models					
	Matching Heuristic		Franklin's rule		Dawes' rule	
	M	SD	M	SD	M	SD
	Watch Managers					
Model fit of derivation sample	82.8	5.2	78.1	4.4	73.4	7.3
Model fit of cross-validation sample	81.6	5.8	77.5	5.3	74.4	6.4
	Crew Managers					
Model fit of derivation sample	75.6	5.0	69.6	8.2	64.1	3.5
Model fit of cross-validation sample	68.5	8.5	64.6	7.2	62.3	4.4
Note. Means (M) and standard deviations (SD) are calculated over 10 test rounds.						

Whereas Franklin's rule and Dawes' rule searched through all 6 cues, the maximum number of cues searched (K) by the Matching Heuristic was on average 1.6 (SD = 0.5) for WMs group and 1.5 (SD = 0.5) for CMs group. The K value of Matching Heuristic differed slightly across the 10 tests, because the elimination of replications process and properties of the derivation sample changed in each test round, which in turn produced

different rank order of cues in each test round. As the results showed, despite the large difference in cue use among the models, the Matching Heuristic outperformed Franklin's rule and Dawes' rule on both WMs data and CMs data.

b) Prediction accuracy of models on the 3 & 4-point ordinal decision data

In this competitive test, we examined the performance of MH-3R, MH-4R, WADD-3R, and WADD-4R in predicting the 3 & 4-point ordinal decision data.

The cue validities and K value of Matching Heuristic produced in the 10 round cross-validations of binary decision data, were directly utilized as the cue validities and K value of the models MH-3R and MH-4R, because they were developed as the natural extension of Matching Heuristic. The same cue validities were utilized in WADD-3R and WADD-4R. It is therefore, yielding 10 rounds test of performance. The 3 & 4-point ordinal nature of the decision data implies that any valid model should preferably predict more than 33% of decision data. Table 5-4 shows the results.

Table 5-4, Mean percentage of 3 & 4-point ordinal decision data predicted correctly by the models

	Models			
	MH-3R & MH-4R		WADD-3R & WADD-4R	
	M	SD	M	SD
	Watch Managers			
Model fit of 3 & 4-point ordinal data	40.3	4.4	46.9	4.2
	Crew Managers			
Model fit of 3 & 4-point ordinal data	50.5	8.1	60.0	5.7
Note. Means (M) and standard deviations (SD) are calculated over 10 test rounds.				

Whereas the MH-3R/MH-4R searched one/two additional cues on the basis of the Matching Heuristic, the WADD-3R & WADD-4R utilizing all the cues like the Franklin's rule. As the results showed, all of the models yielded a moderate performance in predicting the data. In particular, the compensatory models (WADD-3R & WADD-4R) outperformed the non-compensatory models (MH-3R & MH-4R) both on WMs data and CWs data.

5.6.3.4 Self-insight of decision making policy

As discussed in the methodology chapter, researchers in JA tradition have also compared the subjective weights elicited from direct assignment with the statistical weights derived from the computational model. Taken the statistical weights as objective, such comparison has been supposed to indicate the DMs' self-insight of their own decision making policies.

Following this tradition, we were interested in assessing the difference between ICs' explicit estimation of cue importance (subjective cue weights) and their implicit consideration of cue importance (cue validity). Therefore, the rank order of the mean subjective weights of cues assigned by the ICs themselves and the mean rank order of cue validities derived from the 10 round cross-validation were compared here (Figure 5-9 & Figure 5-10).

According to the comparison depicted in Figure 5-9, the two rank orders of cues were differed between each other for WMs. While both of the cue validities and subjective cue weights suggested that the "floor level" was the most important cue, they disagreed with the rank order of other cues. In particular, cue validities took the "wind effect" as the least important, while the mean subjective cue weights indicated the "wind effect" to be second most important.

According to the comparison depicted in Figure 5-10, the two rank orders of cues were differed between each other for CMs as well. If CMs search cues by their validities, "Floor level" would be the first cue to be looked up; "Rescue", "Wind effect", and "Vertical spread" would be then searched in order. If CMs search cues by their

subjective weights, “Rescue” would be the first cue to be looked up in general; “Vertical spread”, “Floor level”, and “Wind effect” would be then searched in order.

These differences indicate that ICs differ in what they publicly state they do and what they actually do.

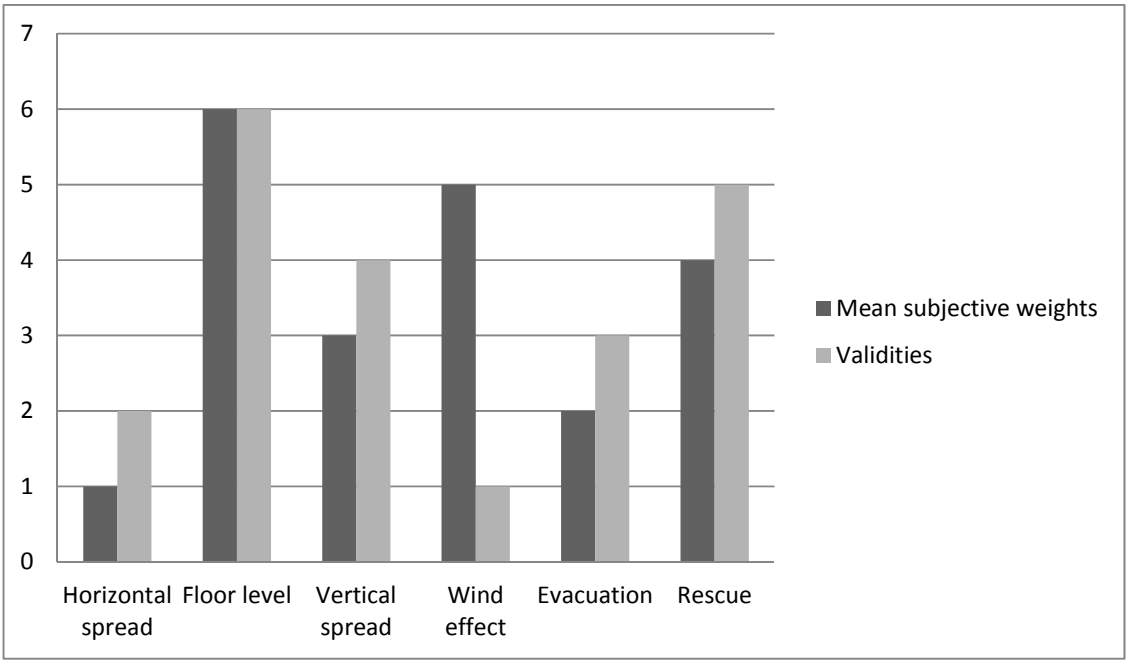


Figure 5-9, Comparison of cue rank orders applied by WMs

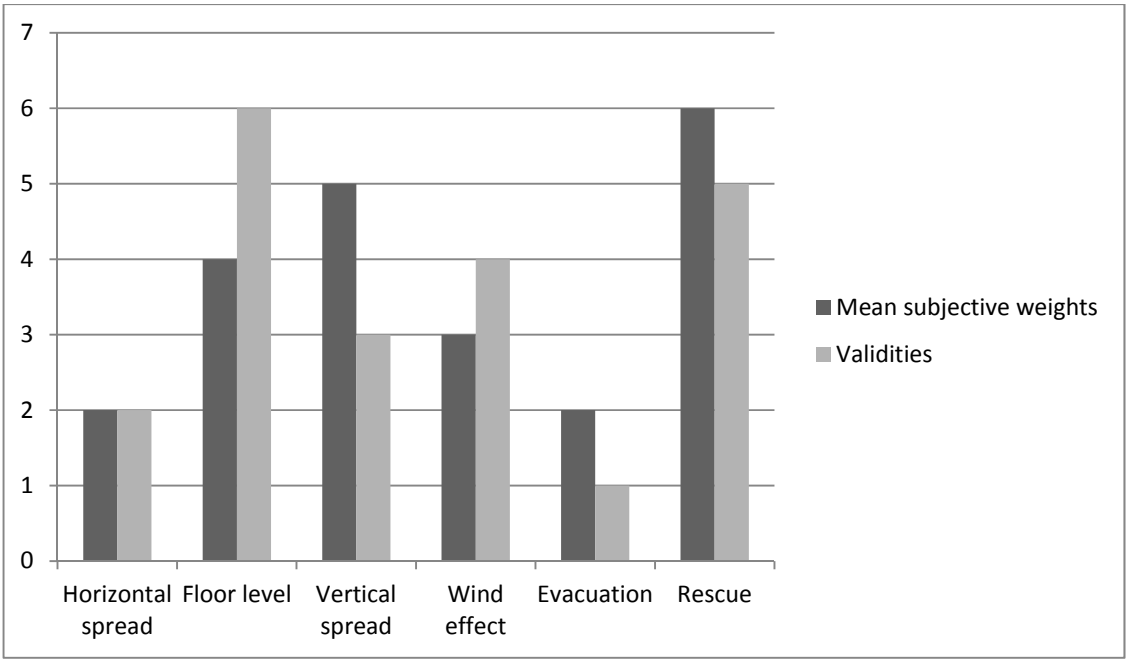


Figure 5-10, Comparison of cue rank orders applied by CMs

To be noted in the figures, the rank order of importance was reversed for ease of illustration, so a rank order of six represents the most important cue.

5.7 Summary

The case study presented in this chapter investigated the ICs' on-site decision making process through our JA-HM method. To our best knowledge, it was the first behavioural test of the validity of the computational models of heuristics, in predicting the DMs' decision making during fire emergency response. It was also the first behavioural test of the validity of the non-compensatory heuristics in predicting the DMs' decisions on ranking task.

We identified a panel of information requirement of ICs' resource demand decision at the on-arrival stage of high rise fire incident. We use this information panel to construct experimental decision making task to collect ICs' decision data. Two groups of ICs (WMs & CMs) from Leicestershire FRS have participated in the study.

The raw data suggest that WMs tended to request more resources in comparison to the CMs. The raw data were further aligned to facilitate the investigation of their decision making policies. The alignment is accomplished through two recoding process, yielding three categories of decision data: binary data, 3-point ordinal data, and 4-point ordinal data.

The binary decision data assumes a binary classification process of decision making of ICs. The 3 & 4-point ordinal decision data, however, indicates the ICs are applying a ranking process of decision making with 3 & 4 ordinal categories.

Our statistical analysis of these aligned data finds that ICs exhibits high levels of post-decisional confidence in their decisions; the CMs are significantly less confident when they requesting more resources; both WMs and CMs have demonstrated high level of intra-group inconsistency; both WMs and CMs differed in the implicit and explicit rank ordering of cues, which suggested their lack of self-insight.

Regarding the ICs' decision making policies, we competitively tested seven computational models of heuristics in predicting the aligned decision data.

In terms of decision making policy of binary classifying the resource demand of the incidents, we found that the both WMs' and CMs' decisions were best predicted by a non-compensatory heuristic, called Matching Heuristic. According to the Matching Heuristic, WMs and CMs only used one or two cues (out of a possible six) in making decisions.

For those WMs and CMs who have ranked (3 & 4-point) the resource demand of the incidents, we found that their decisions were better predicted by the compensatory models (WADD-3R & WADD-4R) in against to the non-compensatory competitors (MH-3R & MH-4R). According to the models, these ICs make decisions by taking into account of all the cues and their relative importance.

These findings contributed to the field of fire emergency response to enable better understanding of the ICs' decision making process. They will be further discussed in the following chapter.

Chapter 6 : Discussions, Conclusions and Future Research

This chapter summarizes and discusses the work conducted in our research. The chapter begins by the summary of research findings and the discussion of their implications in the context of the existing knowledge. This is then followed by a highlight of research contributions. The chapter finally explains the research limitations and indicates the possible future work that could address those limitations.

6.1 Summary and discussion of research findings

Our study has addressed three research questions regarding the methods and applications of decision making under different conditions, like imprecise goal, uncertainty, extreme time pressure, and dynamic shifting situations. To provide answers to the research questions, five research objectives have been enumerated to be achieved. Accordingly, the summary and discussion of research findings are organized through these five research objectives.

1) Carry out literature review on the theory and methodology of the decision making study.

We review the literature of Multiple Criteria Decision Making (MCDM), Evidential Reasoning (ER) approach, Naturalistic Decision Making (NDM) movement, Social Judgment Theory (SJT), and Adaptive Toolbox (AT) program.

The review indicates that MCDM, as a family of methods in operations research, provides a formal tool for supporting decision makers to tackle decision making problems through multiple, and sometimes hierarchical, criteria. In addition, the ER approach, as one of the most recent development in MCDM literature, is capable of dealing with MCDM problems with uncertainties and hybrid nature of information.

The review also indicates that there has been a movement of decision making study (NDM) that attempts to understand how people make decisions in real-world contexts that are meaningful and familiar to them. The real-world contexts are mostly critical,

characterized by the factors like ill-structured problems; uncertain dynamic environments; shifting, ill-defined or competing goals; action/feedback loops; time stress; high stakes; multiple players, and organizational goals and norms (Orasanu & Connolly, 1993). Accordingly, the methodology applied and research outputs produced by researchers in NDM are quite different from the normative decision analysis. These differences are characterized by proficient decision makers, situation-action matching decision rules, context-bound informal modelling, process orientation, and empirical-based prescription.

Following the contentions addressed in NDM, the review then explores another two literature approaches of decision making study, SJT and AT program. The SJT provides a theoretical framework and a methodology alternative of studying people's decision making under environmental constraints in a more formal and quantitative way. AT program inherits the theoretical and methodological contentions of SJT and further highlights the importance of formal models of heuristics in such study.

2) Investigate the theory and methodology used for different conditions in decision making.

We investigated the theory and methodology of decision making under conditions like imprecise goal, uncertain information, extreme time pressure, and dynamic shifting situations.

The investigation indicates that decision making under imprecise goal is normally resolved through MCDM method, to decompose the imprecise goal, in a hierarchical fashion, into a set of more precise sub- and sub-sub criteria which could be incompatible or conflicting, but precise by themselves. Each bottom level criterion is then supposed to be able to be measured in some numerical way. Through a certain mathematical algorithm, these numerical assessments at each bottom level criterion could be aggregated into a general score. Accordingly, comparison among decision alternatives becomes possible and the optimal choice could be identified.

When the information is uncertain, the criterion is difficult and sometimes impossible to be simply assessed through a single number. Major solutions towards this uncertainty include Probability Theory (PT), Dempster-Shafer theory of evidence (D-S theory), and fuzzy set theory. PT represents uncertainty through a probability distribution on a proposition space, and new knowledge is supposed to learn by conditionalization. However, it is unable to represent the ignorance properly. And it assumes that the reinforcement of belief in one state would be associated with a decrease of belief in other states, which is not necessarily the case in real life problem. D-S theory is proposed to overcome the two limitations of PT. But when it is applied in MCDM problem for aggregating conflict evidence, irrational results may be produced. Fuzzy set theory represents uncertain information with fuzzy membership functions. However, the main concern of the fuzzy set theory is the meaning of information rather than its measurement. When applied in the measurement oriented problem, the final fuzzy set, which is aggregated from individual assessments, is difficult to generate an accurate prescription.

It is concluded that common difficulties in making decisions under imprecise goal and uncertain information are: the human cognitive limitations of dealing with multiple factors, the lack of universally agreed scale to represent the uncertainty knowledge in a numerical way, and the need to combine different type of scales on the factors in a consistent manner. The correspondent research is then focused on developing formal procedures to aid the DMs tackle these difficulties so as to support their decision making.

The investigation also indicates that when studying decision making under extreme time pressure and dynamic shifting situations, research focus has changed to the understanding of how decisions are really made by DMs. It is believed that any decision aids should not violate the DMs' deep cognitive concern under such critical conditions. Otherwise, the decision aids developed would hardly get adopted in field. Accordingly, researchers turn to utilize interview and observation methods to develop qualitative

models to account for the DMs' decision making behaviour, such as how do the DMs think when they deciding strategy, how is the strategy generated by them, etc. It is however, the description solely with qualitative account failed to provide any clear and precise predictions of decision behaviour. In addition, such qualitative account has another shortcoming. It explains too little and too much: too little, as the underlying processes are left unspecified, and too much, because, with sufficient imagination, one of them could be fit to too much empirical result.

It is concluded that a literature gap is to be filled, that there is lack of formal quantitative model to account for the people's decision making process under critical conditions like extreme time pressure and dynamic shifting situations. Through formal quantitative model, the prediction of DMs' decisions would become possible. Correspondent with the quantitative modelling, a different methodology is demanded to facilitate the research investigation.

3) Choose proper cases for the applications of the research

We choose two cases for the applications of the research.

The first case is Soldier System decision making, that concerns with the deciding of the optimal matching between a specific Soldier mission and an individual Soldier system. This decision making task is characterized by imprecise goal. Because the individual Soldier, under the concept of Soldier as a System (SaaS), is supposed to hold superior capabilities to accomplish assigned tasks across the spectrum of conflicts, in any operational environment. The demanding of such fully integrated capabilities, obscure the goal of individual Soldier in a mission, and results in the field equipping heavy, bulky and burdensome, which in turn degrades the Soldier's effectiveness and performance in an unexpected way. The decision making task is also characterized by uncertainty. As the performance of a Soldier system in a mission has to be estimated in prior, which are subjective in nature and sometimes incomplete due to the unpredictable nature of Soldier missions and mission contexts.

The second case is fire emergency decision making, which concerns with the decisions made by Incident Commanders (ICs) during their fire emergency response. This decision making task is characterized by extreme time pressure. As it has been identified that ICs make around 80 percent of their decisions in less than one minute (Klein, Calderwood, & Clinton-Cirocco, 1988). The decision making task is also characterized by dynamic shifting situations. As it has been estimated that the situation changed an average of five times per incident (Klein, 1998), which demand the correspondent shift in the way the ICs understand the incident.

4) Discuss and propose an alternative framework to aid people's decision making under uncertainty and imprecise goal, and apply it to a case such as Soldier system decision making.

By applying the ER algorithm, we propose a novel framework to aid people's decision making under uncertainty and imprecise goal. Under the framework, the imprecise goal is objectively modelled through an analytical structure, and is independent of the task requirement; the task requirement is specified by the trade-off strategy among criteria of the analytical structure through an importance weighting process, and is subject to the requirement change of a particular decision making task; the evidence available, that could contribute to the evaluation of general performance of the decision alternatives, are formulated with belief structures which are capable of capturing various format of uncertainties that arise from the absence of data, incomplete information and subjective judgments.

By applying the framework to a case of Soldier system decision making, the method Evidence-based Trade-Off (EBTO), as a tool, is demonstrated to be able to provide the holistic analysis regarding the requirements of Soldier missions, the physical conditions of Soldiers, and the capability of their equipment and weapon systems, which is critical in domain.

5) Discuss and propose a methodology to investigate DMs' decision making process when they are under extreme time pressure and dynamic shifting situations, and apply it to a case such as fire emergency decision making.

We develop a novel variant of JA method, JA-HM, to accomplish the study of decision making problem in extreme time pressure and dynamic shifting situations. The JA-HM further extend the traditional JA method, through a number of novel methodological procedures, to account for the unique features of decision making tasks under extreme time pressure and dynamic shifting situations. By elaborating the procedures of JA-HM step by step, we have discussed why and how these novel aspects of JA-HM could accomplish the deficiencies of traditional JA. These novel aspects have been summarized in Section 3.3. A briefing is restated in the following:

- To deconstruct the dynamic shifting situations in a way that decision problem could be identified and formulated, a notion of decision point is introduced by drawing the literature from NDM.
- To avoid misleading investigation and to enable meaningful comparison of research findings, a classification of decision problems that would be encountered in the dynamic shifting situations is formulated and an associated data alignment process is proposed by drawing the literature from the NDM.
- To account for the DMs' iterative process of information perception and comprehension in dynamic task environment, a notion of composite cue is formulated in the decision profile development process by drawing the literature from SA and cognitive psychology.
- To account for the time constraints and process dynamics of DMs' decision making under extreme time pressure and dynamic shifting situations (e.g., information search, shift in information emphasis or decision making policy), computational models of heuristics are applied to replace the static multiple regression in modelling the DMs' decision making policies. By capturing the decision making policies, the DMs' decision making processes could be specified.

- Associated with the methodological principle of competitive testing, two model selection criteria and a cross-validation process are formulated to facilitate the application of the method.

By applying the JA-HM in the case study of fire emergency decision making, we have identified a panel of information requirement of ICs' resource demand decision at the on-arrival stage of high rise apartment fire incident. We use this information panel to construct experimental decision making task to collect ICs' decision data. Two groups of ICs (WMs & CMs) from Leicestershire FRS have participant in the study.

The raw data suggest that WMs tend to request more resources in comparison to the CMs. This is compatible with the department policy applied in FRS. It is regulated in the UK FRS department that the maximum number of fire appliances that could be managed by an IC is constrained by his rank order and level of qualification held. Typically, CMs qualified at Supervisor level 1 (Sup 1) could only command simple 1 or 2 appliances incidents; whereas WMs qualified at Supervisor level 2 (Sup 2) can command incidents with a life risk or up to 6 appliances. It is therefore, CMs would mostly get used to manage and request small amount of resources as they believe any further demanding is out of their responsibility.

The raw data were further aligned to facilitate the investigation of ICs' decision making policies. The alignment is accomplished through two recoding process, yielding three categories of decision data: binary data, 3-point ordinal data, and 4-point ordinal data.

The binary decision data assumes a binary classification process of decision making of ICs. The 3 & 4-point ordinal decision data, however, indicates the ICs are applying a ranking process of decision making with 3 & 4 ordinal categories.

Our statistical analysis of these aligned data finds that ICs exhibits high levels of post-decisional confidence in their decisions; the CMs are significantly less confident when they request more resources; both WMs and CMs have demonstrated high level of

intra-group inconsistency; both WMs and CMs differed in the implicit and explicit rank ordering of cues, which suggested their lack of self-insight.

The high level of post-decisional confidence of ICs implies that any decision aids that violate their current decision behaviour of ICs should be carefully introduced. As they might be unwilling to make change (Zakay, 1997).

The finding that the CMs were significantly less confident when they were requesting more resources indicate that CMs were less confident to deal with more severe incidents. In comparison, WMs' confidence level would not be significantly affected by the severity of the incidents. This difference is in line with their different rank levels. CMs, as the junior ICs, have less chance to command the more severe incidents according to the department policy described above. They are therefore, less experienced in commanding severe incident and, in turn, less confident when requesting more resources.

The high level of intra-group inconsistency, on the other hand, indicated the high degree of subjective uncertainty of ICs in making decisions of high rise fire incident during the on-arrival stage (Hammond et al., 1980).

The lack of self-insight of ICs has significant implication of FRS. As departmental training and knowledge exchange among colleagues, is supposed to play an important role in the skills development of ICs. If ICs differ in what they publicly state they do and what they actually do, the knowledge development of the department would be negatively affected. Future research could investigate ICs' self-insight by using policy recognition methods like those advocated by Reilly & Doherty (1992).

Regarding the ICs' decision making policies, we competitively tested seven computational models of heuristics in predicting the aligned decision data.

In terms of decision making policy of binary classifying the resource demand of the incidents, we find that both WMs' and CMs' decisions is best predicted by a non-

compensatory heuristic, called Matching Heuristic. According to the Matching Heuristic, WMs and CMs only use one or two cues (out of a possible six) in making decisions.

The best competence of Matching Heuristic in predicting both WMs' and CMs' binary decision data is compatible with the empirical evidence identified in the AT literature that decisions of experienced professionals are more likely to be predicted by non-compensatory models (e.g., Dhami 2003; Dhami & Ayton, 2001).

For those WMs and CMs who have ranked (3 & 4 ordinal points) the resource demand of the incidents, we found that their decisions are better predicted by the compensatory models (WADD-3R & WADD-4R) in against to the non-compensatory competitors (MH-3R & MH-4R). According to the models, these ICs make decisions by taking into account of all the cues and their relative importance.

This finding is in contrast to the findings identified from the binary decision data, where both WMs and CMs are believed to apply a non-compensatory model, Matching Heuristic, to binary classifying the resource demand of the incidents. According to the Matching Heuristic, WMs only used one or two cues (out of a possible six) in making decisions. This contradiction implies that decision making policies concluded from the binary classification task would not necessary be justified in more complicated multiple-category ranking task.

Task formalized in AT literature has been concentrated on the two-alternative choice task, where the alternatives and the cues (information) about the alternatives are binary. This type of task, being a staple of experimental psychology, holds an idealization that problems of greater complexity are reducible through this elementary case (Goldstein & Gigerenzer, 1999). The contradicting findings concluded here, regarding ICs' decision making policies in binary classification task and multiple-category ranking task, arouse the caution that any findings in two-alternative choice task is not necessary generalizable to the more complicated task. Very little investigations in AT literature have been conducted on this more complicated dimension. To our best knowledge, the only empirical study was conducted by Frosch, Beaman, & McCloy (2007), in which

some evidence of people's use of heuristic, when the choice was among up to four alternatives, was suggested. But they have not applied any competitive testing among different models.

Looking at the implications other than the AT literature, as argued in the NDM literature, decision making during fire emergency response is highly associated with the DM's situation awareness. The SA serves as a major input of consequent decision making and may impact the process of decision making itself (Endsley, 1997; Klein, 2000). Our quantitative modelling from the higher order information processing perspective provides a useful alternative to complement and validate the qualitative findings argued in the literature regarding those lower order mental mechanisms in decision making. According to the discrimination method applied in Jones et al. (2003), the critical cues should be less of a factor for mental simulation as opposed to pattern matching, since numerous cues will be considered and employed in the mental simulation process. In this regard, the prevailing of non-compensatory policy of WMs and CMs in binary classifying the resource demand of the incidents suggested that they were more plausible applying pattern matching for decision making in this simplest occasion, while the prevailing of compensatory policy of those WMs and CMs in multiple-category ranking the resource demand of the incidents suggested that they were more plausible turning to apply mental simulation for decision making if they assessed that the situations were more complex. This implication is compatible with the qualitative account suggested in the classical RPD model (Klein, 1993), in which DMs were described to use pattern matching to quickly respond to the situation in the simplest occasions, and to further use mental simulation to perform some conscious evaluation of decisions in more complex situations.

Finally, the finding that WMs and CMs tended to apply the same decision making policies is in contrast to the previous literature of professional decision making in terms of expert-novice difference (e.g., Garcia-Retamero & Dhami, 2009; Shanteau, 1992). It was argued that less experienced professionals are more likely to use more information

in making decisions. This could be partly due to the different definition of novice in different studies. For instance, in Garcia-Retamero & Dhami (2009)'s study of decision policies for residential burglary, novice was defined as naïve participants (i.e. college students), who have absolute no experience on the problem. In contrast, CMs studied in our research, as the group of most novice ICs in the department, still have an average of 12.9 years of experience in the FRS and 4.4 years of experience as IC. In addition, the context analysis of FRS suggested that social learning (Gigerenzer, Hoffrage, & Goldstein, 2008), like departmental training and knowledge exchange among colleagues, is supposed to play an important role in the skills development of ICs. This would stimulate the ICs share the same decision making policies for the same type of problem. Such organizational context could provide some additional reasons of why WMs and CMs were using the same decision making policies. Further evidence of this social influence could be found in the comparison of mean subjective cue weights between WMs group and CMs group. Despite of the difference in rank level and personal experience, there is no significant difference of their mean subjective cue weights identified.

6.2 Research contributions

By identifying the research findings illustrated above, our research contributions could be summarized as follows:

- Through the employment of ER algorithm, we proposed a novel framework to aid DMs making decisions under imprecise goal and uncertain information.
- Through applying the framework to the case study of Soldier system decision making, our method EBTO is demonstrated to be able to provide the holistic analysis regarding the requirements of Soldier missions, the physical conditions of Soldiers, and the capability of their equipment and weapon systems, which is critical to the domain.

- The same application has also been the pioneering employment of ER algorithm in Soldier system decision making, which contribute to the empirical study of ER literature.
- The JA-HM developed in our research provided a novel variant of JA method for the investigation of people's decision making process under extreme time pressure and dynamic shifting conditions, which contribute to the JA literature.
- The application of JA-HM to the case of fire emergency decision making produced a panel of information requirement of ICs' resource demand decision at the on-arrival stage of high rise apartment fire incident. The application also identified a number of characteristics of ICs' resource demand decisions, like their decision consistency, post-decisional confidence, self-insight of decision making policy, etc. These findings contributed to the field to better understand the ICs' resource decision making of high rise apartment incident.
- The case study of fire emergency decision making was the first behavioural test of the validity of the computational models of heuristics, in predicting the DMs' decision making during fire emergency response. The evidence of ICs' use of matching heuristic in binary classification process of decision making extend the literature of people's use of non-compensatory strategy in professional decision making.
- The case study of fire emergency decision making was also the first behavioural test of the validity of the non-compensatory heuristics in predicting the DMs' decisions on ranking tasks. Our study developed two novel non-compensatory heuristics for this type of task (MH-3R & MH-4R). The two heuristics yielded a moderate performance in predicting ICs' decisions. They together extended the candidate heuristics in AT literature.
- Finally, by comparing our findings of ICs' decision making policies with the qualitative account suggested in the classical RPD model, this study has demonstrated how the decision making studies from different perspective (i.e., higher order information processing perspective VS. lower order mental

mechanisms perspective) could complement and validate each other. The study has therefore, contributed to the cross-disciplinary research of AT and NDM.

6.3 Limitations and future research

Our research, through the development of two methodologies and the application of two case studies, presents encouraging findings that could contribute to a diverse literature of decision making study. However, it is acknowledged that our research is also subject to the following limitations, which could address the possible direction of future research.

Firstly, due to the lack of engaging partners, the Soldier system assessment model utilized in the case study of Soldier system decision making was based on the previous research in literature. The assessment model could be improved with more sophisticated and up-to-date interviews and field observations if a stronger field access is available.

Secondly, the decision making task in the case study of fire emergency decision making designed a complete cases sample that could be hypothesized with the six varied cues. The consideration of cue distributions and intercorrelations were therefore simplified (uniform distribution and zero intercorrelations). This is due to the fact that the complete ecological information of cue distributions and intercorrelations were just not available, which was a common practice in JA literature. The case study was therefore limited in terms of the fact that we failed to represent the ecological properties toward which generalizations are intended. It is however, empirical comparisons of the policies captured with representative and unrepresentative designs suggest a controversial picture (e.g., Phelps & Shanteau, 1978; Braspenning & Sergeant, 1994), and any definite conclusions regarding the effects of representative sampling on research findings have not be drawn in literature (Dhami, Hertwig, & Hoffrage, 2004). Brehmer & Brehmer (1988) suggested a possible remediation for orthogonal set by carefully eliminating those unrealistic or atypical cases. The present study followed this suggestion by

carefully defining the representations of those six varied cues to ensure any combination of them is not impossible.

Thirdly, Recent development in AT literature has demonstrated the difficulty of people's learning of cue validities (Broder 2003; Newell & Shanks, 2003; Newell, Weston, & Shanks, 2003; Rakow, Hinvest, Jackson, & Palmer, 2004; Todd & Dieckmann, 2005) and a number of additional cue orders have been proposed and experimentally tested like "discrimination rate", "usefulness" and "success" (e.g., Gigerenzer, Dieckmann, & Gaissmaier, 2011; Newell, Rakow, Weston, & Shanks, 2004; Rakow, Newell, Fayers, & Hersby, 2005). However, these orders only differ from cue validities when the discrimination rates vary. As our experiment plan were designed with the unique discrimination rate for all cues, our study failed to be able to compete the cue validities with these orders. As Gigerenzer, Dieckmann, & Gaissmaier, (2011) most recently proposed, future study could be improved by involving competition among search rules, in addition to computational models of heuristics, to facilitate richer findings.

Finally, the pilot test of our case study of fire emergency decision making suggested that ICs were unwilling to engage in task with too many cases. Consequently, we design the decision making task at the group level. One of our reflections on the reason of the unwillingness was that the cases just simulated a one-shot situation of the firefighting activity. This is different from the studies conducted in traditional JA. Task applied in traditional JA studies are typically static, in which the judgment itself is complete and regularly encountered by the DMs. Most of the time, judging a set of profiles itself is part of the DMs' regular work (e.g., weather forecasting). In contrast, fire operations during the incident involve dynamic activities following the timeline. Any one-shot situation, although may be critical at the time, do not dominant the whole operation, and is thus less of interest to the fire officers. Consequently, insisting too many similar cases in an experimental decision making task for an individual participant, as the most JA studies did, is not representative for decision making in dynamic fire emergency response.

This partly explains our literature confusion that although Kenneth R. Hammond, the prominent leader of JA, actually participated in the first Naturalistic Decision Making (NDM) conference (Klein et al., 1993), the impact of his important conceptualization on the NDM community has been minimal. Howell (1997) suspected this phenomenon as the NDM community's "definition-by-contrast thinking" and "intentional overlook in their zeal to distance NDM from 'traditional decision theory'" (p.39). Here, we might encounter an alternative explanation that study design of JA, which traditionally involved large numbers of similar cases for judging, was already conflicting to the representative principle (Cooskey, 1996), where dynamic story-like response to changing environment was demanded. Accordingly, an improved version of our JA-HM would be formulating a number of decision points following the timeline of the incident, and analyzing decision data of each decision point at the group level, where larger data set could be obtained and richer findings of decision making policy could be revealed.

References:

Allais, M. (1953). Le comportement de l'homme rationel devant le risque: Critique des postulats et axioms de l'Ecole Americaine. *Econometrica, 21: 503-46.*

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness?. *Memory & cognition*, *32*(8), 1369-1378.

Baddeley, A. (2010). "Lightening The Soldier Platform", *Defence Director*, Jan 2010, pp. 21-22.

Balzer, W.K., Doherty, M.E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin, 106,* 410-433.

Beach, L.R., Chi, M., Klein, G.A., Smith, P., Vicente, K. (1997). Naturalistic decision making and related research lines. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 29-36). Mahwah, NJ: Erlbaum.

Beach, L.R., & Lipshitz, R. (1993). Why classical decision theory is an in appropriate standard for evaluating and aiding most human decision making? In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods (pp. 21-35).* Norwood, NJ: Ablex.

Bedny, G., & Meister, D. (1999). Theory of activity and situation awareness. *International Journal of Cognitive Ergonomics, 3(1),* 63-72.

Bernoulli, D., 1738. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperiales Petropolitanae, 5, 175-192*. (Trans. by L. Sommer in *Econometrica*, 1954, 22, 23-36).

Berretty, P.M., Todd, P.M., & Martignon, L. (1999). Categorization by Elimination: Using Few Cues to Choose. In G. Gigerenzer, P.M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 235-256). New York: Oxford University Press.

Bouyssou, D. (1990). Building criteria: A prerequisite for MCDA. In C.A. Bana e Costa (Eds.), *Readings in Multiple Criteria Decision Aid* (pp. 58-80). Berlin: Springer.

Blythe, P.W., Todd, P.M., & Miller, G.F. (1999). How Motion Reveals Intention: Categorizing Social Interactions. In G. Gigerenzer, P.M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 257-286). New York: Oxford University Press.

Biggs, S. F., Rosman, A. J., & Sergenian, G. K. (1993). Methodological issues in judgment and decision-making research: Concurrent verbal protocol validity and simultaneous traces of process. *Journal of Behavioral Decision Making*, *6*(3), 187-206.

Brans, J.P., & Vincke, P. (1985). A preference ranking organisation method: The PROMETHEE method for MCDM. *Management Science, 31(6),* 647-656.

Braspenning, J., & Sergeant, J. (1994). General practitioners' decision making for mental health problems: Outcomes and ecological validity. *Journal of clinical epidemiology*, *47*(12), 1365-1372.

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment The SJT view* (pp. 41-74). New York: North-Holland.

Brehmer, B. (1971). Effects of communication and feedback on cognitive conflict. *Scandinavian Journal of Psychology, 12,* 205-216.

Brehmer, B. (1976). Social judgment theory and the analysis of interpersonal conflict. *Psychological Bulletin, 83*, 985-1003.

Brehmer, B. (1979). Preliminaries to a psychogy of inference. *Scandinavian Journal of Psychology, 20,* 193-210.

Brehmer, B. (1987). *Policy construction. Transforming declarative knowledge into judgments.* Uppsala: University of Uppsala, Department of Psychology, Unpublished manuscript.

Brehmer, B. (1988). The development of social judgment theory. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment The SJT view* (pp. 13-40). New York: North-Holland.

Brezovic, C.P., Klein, G.A., & Thordsen, M. (1987). *Decision making in armoured platoon command (*Contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA). Fairborn, OH: Klein Associates Inc.

Brighton, H. (2006). Robust Inference with Simple Cognitive Models. In AAAI Spring Symposium: Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems (pp. 17-22).

Brighton, H., & Gigerenzer, G. (2008). Bayesian brains and cognitive mechanisms: harmony or dissonance? In N. Chater, M. Oaksford, *The Probabilistic Mind: Prospects for Bayesian Cognitive Science (eds.), pp. 189-208.* New York: Oxford Univ. Press.

Broder, A. (2000a). Assessing the empirical validity of the “take-the-best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1332-1346.

Broder, A. (2000b). A methodological comment on behavioural decision research. *Psychologische Beitrage, 42*, 645-662.

Broder, A. (2002). Bailing and jailing the unknown way: A critical examination of a study reported by Dhami and Ayton (2001). *Berichteaus dem Psychologischen Institut der Universitat Bonn, 28(1).*

Bröder, A. (2003). Decision making with the" adaptive toolbox": Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 611.

Broder, A. & Schiffer, S. (2003). “Take the best” versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General, 132*, 277-293.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108-132.

Bruner, J.S., Goodnow, J.J., & Austin, A.A. (1956). *A study of thinking.* New York: Wiley.

Brunswik, E. (1934). *Wahrnehmung und Gegenstandswelt: Grundlegung einer Psychologie vom Gegenstand her* [Perception and the world of objects: The foundations of a psychology in terms of objects]. Leipzig, Germany: Deuticke.

Brunswik, E. (1940). Thing constancy as measured by correlation coefficients. *Psychological Review, 47*, 69-78.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review, 50*, 255- 272.

Brunswik, E. (1944). Distal focusing of perception: Size constancy in a representative sample of situations. *Psychological Monographs, 56, 1-49.*

Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, No. 10, pp. 656-760). Chicago: University of Chicago Press.

Brunswik, E. (1955a). In defense of probabilistic functionalism: A reply. *Psychological Review, 62, 236-242*.

Brunswik, E. (1955b). “Ratiomorphic” models of perception and thinking. *Acta Psychologica, 11, 108-109*.

Brunswik, E. (1955c). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62, 193-217.*

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Buchanan, B.G., & Shortliffe, E.H. (1984). *Rule based expert systems*. MA, Reading: Addison-Wesley.

Burnside, D.G. (1994). *A study of land management judgments in the rangelands of Western Australia and South Australia.* Unpublished Ph.D. thesis, University of Queensland, Brisbane, Queensland, Australia.

Beynon, M., Curry, B., & Morgan, P. (2000). The Dempster–Shafer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega*, *28*(1), 37-50.

Centre for Defense Enterprise (CDE) Seminar, (2009). *A Strategy for Soldier System Integration, Integrated Soldier System Executive (ISSE).* Harwell: Ministry of Defence (MOD).

Chin, K.S., Wang, Y.M., Yang, J.B., Poon, K. (2009). An evidential reasoning based approach for quality function deployment under uncertainty. *Expert Systems with Applications, 36, 5684-5694.*

Chin, K.S., Yang, J.B., Guo, M., Lam, J. (2009). An evidential reasoning-interval based method for new product design assessment.. *IEEE transactions on Engineering Management, 56(1), 142-156.*

Christal, R.E. (1963). *JAN: A technique for analyzing individual and group judgment.* Lackland Air Force Base, Texas: Personnel Research Laboratory, Areospace medical Division.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation for the behavioural sciences.* Hillsdale, NJ: Erlbaum.

Communities and Local Government (2007). *The Building Regulations Approved Document B Fire Safety, Volume 2 – Building Other Than Dwellinghouses*. London: NBS Publications.

Communities and Local Government (2009). *Generic Risk Assessment 5.8, Flashover, backdraught and fire gas ignitions*. London: TSO Publications.

Communities and Local Government (2012). *Fire and Rescue Service Operational guidance, Operational Risk Information*. CLG, London.

Cook, R. L., & Stewart, T. R. (1975). A comparison of seven methods for obtaining subjective descriptions of judgmental policy. *Organizational Behavior and Human Performance*, *13*(1), 31-45.

Cooksey, R.W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.

Cooksey, R. W., Freebody, P., & Bennett, A. J. (1990). The ecology of spelling: A lens model analysis of spelling errors and student judgments of spelling difficulty. *Reading Psychology: An International Quarterly*, *11*(4), 293-322.

Cooksey, R. W., Freebody, P., & Davidson, G. R. (1986). Teachers' predictions of children's early reading achievement: An application of social judgment theory. *American Educational Research Journal*, *23*(1), 41-64.

Cosmides, L., & Tooby, J. (2006). Evolutionary psychology, moral heuristics, and the law. In G. Gigerenzer, C. Engel (Eds.), *Heuristics and the Law,* (pp. 175-205). MA: MIT Press.

Crandall, B., & Calderwood, R. (1989). *Clinical assessment skills of experienced neonatal intensive care nurses.* (Contract 1 R43 NR0191101 for The National Center for Nursing, NIH). Fairborn, OH: Klein Associates Inc.

Crandall, B., Klein, G.A., & Hoffman, R.R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* Cambridge, MA: MIT Press.

Czerlinski, J., Gigerenzer, G., Goldstein, D.G. (1999). How good are simple heuristics? In: G. Gigerenzer, P.M. Todd, the ABC research Group (eds.), *Simple heuristics that make us smart.* Oxford University Press, New York, pp. 97-118.

Dalgleish, J. (1988). Decision making in child abuse cases: Applications of social judgment theory and signal detection theory. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 317-360). New York: North-Holland.

Dawes R.M. (1979). The robust beauty of improperlinear models in decision making. *Am. Psychol. 34: 571-582.*

Dawes R.M., Gorrigan B. (1974). Linear models in decision making. *Psychol. Bull. 81: 95-106.*

Dempster, A.P. (1967). Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics 38: 325-339.*

Dhami, M.K. (2003). Psychological models of professional decision making. *Psychological Science, 14,* 175-180*.*

Dhami, M.K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14,* 141-168*.*

Dhami, M.K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin, 130*, 959-988.

Dhami, M.K., & Harries, C. (2001). Fast and frugal versus regression models of human judgment. *Thinking and Reasoning, 7*, 5-27.

Dieckmann, A., & Todd, P.M. (2011). Simple Rules for Ordering Cues in One-Reason Decision Making. In P.M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological Rationality: Intelligence in the world* (pp. 274-308). New York: Oxford University Press.

Dreyfus, H.L. (1997). Intuitive, Deliberative, and Calculative models of expert performance. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 17-28). Mahwah, NJ: Erlbaum.

Dreyfus, H.L., & Dreyfus, S.E. (1986). *Mind over machine: the power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Dreyfus, S.E. (1981). Formal models versus human situational understanding: Inherent limitations on the modelling of business expertise. Technical report No. ORC 81-3. Operations Research Center, University of California, Berkley.

Dudycha, A.L. (1970). A monte carlo evaluation of JAN: A technique for capturing and clustering raters' policies. *Organizational Behavior and Human Performance*, 5, 501-516.

Earle, T.C. (1973). Interpersonal learning. In L. Rappoport & D.A. Summers (Eds.), *Human judgment and social interaction* (pp. 240-266). New York: Holt, Rinehart & Winston.

Ebbesen, E.B. & Konenci, V.J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32*, 805-821.

Edgell, S.E. (1993). Using configural and dimensional information. In N.J. Castellan (Ed.), *Individual and Group Decision Making* (pp. 43-64). Hillsdale, NJ: Erlbaum.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin, 51: 380-417.*

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics, 75: 643-69.*

Endsley, M.R. (1987). *SAGAT: A methodology for the measurement of situation awareness (NOR DOC 87-83).* Hawthorne, CA: Northrop Corp.

Endsley, M. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting*, pp. 97-101. Santa Monica, CA: Human Factors Society.

Endsley, M.R. (1995). Towards a theory of situation awareness. *Human Factors, 37(1),* 32-64.

Endsley, M.R. (1997). The role of situation awareness in naturalistic decision making. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 269-284). Mahwah, NJ: Erlbaum.

Endsley, M.R. (2000). Theoretical underpinnings of situation awareness: a critical review. In: M.R. Endsley, D.J. Garland, eds. *Situation Awareness Analysis and Measurement.* Mahwah, NJ: Laurence Erlbaum Associates.

Endsley, M.R. (2004). Situation awareness: Progress and Directions. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 317-341). Aldershot, UK: Ashgate.

Endsley, M.R., Bolte, B., & Jones, D.G. (2003). *Designing for situation awareness: an approach to human-centered design*. London: Taylor & Francis.

Endsley, M.R., & Garland, D.J. (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum.

Ericsson, K.A., & Leman, A.C. (1996). Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Executive Decision Services, Inc. (1991). *POLICY PC Judgment analysis software reference manual* (3rd Ed.) [Computer software manual]. Albany, NY: Executive Decision Services, Inc.

Flin, R., Salas, E., Strub, M., & Martin, L. (1998). *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot, UK: Ashgate.

Figuera, J., Greco, S., & Ehrgott, M. (Eds) (2005). *Multiple Criteria Decision Analysis, State of the Art Surveys,* New York: Springer.

Fishlock, M., (2013). High-rise fire fighting: Developing better responses to fires in tower blocks and high-rise buildings. [online] Available at: < http://www.highrisefirefighting.co.uk/> [Accessed 17 January 2014].

Franklin, B. (1987). Writings. New York: The Library of America. (Original letter written Sept. 19, 1772).

Fredholm, L. (1997). Decision making patterns in major fire-fighting and rescue operations. In R. Flin, E. Salas, M. Strub & L. Martin (Eds), *Decision making under stress: emerging themes and applications* (pp. 107-115). Aldershot, UK: Ashgate.

Friedman, M. (1953). Essays in Positive Economics. Chicago: University of Chicago Press.

Frosch, C. A., Beaman, C., & McCloy, R. (2007). A little learning is a dangerous thing: An experimental demonstration of ignorance-driven inference. *The Quarterly Journal of Experimental Psychology*, *60*(10), 1329-1336.

Garcia-Retamero, R., & Dhami, M.K. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychon. Bull. Rev.* 16: 163-169.

Garcia-Retamero, R., Hoffrage, U., & Dieckmann, A. (2007). When one cue is not enough: Combining fast and frugal heuristics with compound cue processing. *Quarterly Journal of Experimental Psychology, 60*, 1197-1215.

Garcia-Retamero, R., Hoffrage, U., Dieckmann, A., & Ramos, M. (2007). Compound cue processing within the fast and frugal heuristic approach in non-linearly separable environments. *Learning & Motivation, 38*, 16-34.

Garcia-Retamero, R., Takezawa, M., & Gigerenzer, G. (2006). How to learn good cue orders: When social learning benefits simple heuristics. In *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1352-1358).

Garcia-Retamero, R., Takezawa, M., & Gigerenzer, G. (2008). Group communication and decision-making strategies. *Psicothema, 20*, 753-759.

Garcia-Retamero, R., Wallin, A., & Dieckmann, A. (2007). Dose causal knowledge help us be faster and more frugal in our decisions? Memory & Cognition, 35, 1399-1409.

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural computation*, *4*(1), 1-58.

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *The American journal of psychology*, 185-195.

Gigerenzer, G. (1987). Survival of the fittest probabilist: Brunswik, Thurstone, and the two disciplines of psychology. In L. Kruger, G. Gigerenzer, & M.S. Morgan (Eds.), *The probabilistic revolution. Volume 2: Ideas in the sciences.* Cambridge, MA: The MIT Press, 49-72.

Gigerenzer, G. (1991a). How to make cognitive illusions disappear: beyond "heuristics and biases". In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology (vol. 2, pp. 83-115).* Chichester, UK: Wiley.

Gigerenzer, G. (1991b). On cognitive illusions and rationality. In E. Eells, & T. Maruszewski (Eds.), *Probability and rationality: studies on L. Jonathan Cohen's philosophy of science (Poznan studies in the philosophy of the sciences and the humanities) (vol. 21, pp. 225-249).* Amsterdam: Rodopi.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky. *Psychological Review,* 103, 592-596.

Gigerenzer, G. (2004). Striking a blow for sanity in theories of rationality. In M. Augier & J.G. March (Eds.), *Models of a man: Essays in memory of Herbert A. Simon* (pp. 389-409). Cambridge, MA: MIT Press.

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

Gigerenzer, G., Dieckmann, A., & Gaissmaier, W. (2011). Efficient Cognition Through Limited Search. In P.M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological Rationality: Intelligence in the world* (pp. 241-273). New York: Oxford University Press.

Gigerenzer, G., Fiedler, K., Olsson, H. (2011). Rethinking Cognitive Biases as Environmental Consequences. In: P.M. Todd, G. Gigerenzer, & The ABC Research Group (eds.), *Ecological Rationality: Intelligence in the World.* Oxford, UK, pp. 80-110.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annu. Rev. Psychol.,* 62, 451-482.

Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*, 650-669.

Gigerenzer, G., & Goldstein, D.G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer, P.M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 75-95). New York: Oxford University Press.

Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behaviour*. New York: Oxford University Press.

Gigerenzer, G., Hoffrage, U., & Goldstein, D.G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review, 115*, 230-239.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review, 98*, 506-528.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: The MIT Press.

Gigerenzer, G., & Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese, 187*, 243-268.

Gigerenzer, G., & Todd, P.M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G.Gigerenzer, P.M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 3-34). New York: Oxford University Press.

Gigerenzer, G., Todd, P.M., & The ABC Research Group (1999). *Simple heuristics that make us smart.* Oxford, UK: Oxford University Press.

Goldstein, D.G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*, 75-90.

Gorman, C.D., Clover, W.H., & Doherty, M.E. (1978). Can we learn anything about interviewing real people from "interviews" of paper people? Two studies of the external validity of a paradigm. *Organizational Behavior and Human Performance, 22*, 165-192.

Grimwood, P. (2014). Take the fire first. [online] Available at: < http://www.firetactics.com/TACTICAL-OBJECTIVES.htm> [Accessed 22 January 2014].

Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1992). *Multivariate data analysis with readings* (3rd Ed.). New York: Macmillan Publishing Company.

Halpern, J. (2003). *Reasoning about Uncertainty*. The MIT Press.

Hamm, R. M. (1988). Moment-by-moment variation in experts' analytic and intuitive cognitive activity. *Systems, Man and Cybernetics, IEEE Transactions on*, *18*(5), 757-776.

Hammond, K.R. (1955). Probabilistic functioning and the clinical method. *Psychology Review, 62, 255-262*.

Hammond, K.R. (1965). New directions in research in conflict resolution. *Journal of Social Issues, 21*, 255-262.

Hammond, K.R. (1966a). *The psychology of Egon Brunswik.* New York: Holt, Rinehart & Winston.

Hammond, K.R. (1966b). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K.R. Hammond (Ed.), *The psychology of Egon Brunswik* (pp. 15-80). New York: Holt, Rinehart and Winston.

Hammond, K.R. (1972). Inductive knowing. In J.R. Royce & W.W. Rozeboom (Eds.), *The psychology of knowing.* New York: Gordon & Breach, pp. 285-320.

Hammond, K.R. (1973). The cognitive conflict paradigm. In L. Rappoport & D.A. Summers (Eds.), *Human judgment and social interaction* (pp. 188-205). New York: Holt, Rinehart & Winston.

Hammond K.R. (2001). Expansion of Egon Brunswik's psychology, 1955-1995. In K.R. Hammond & T.R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 464-478). New York: Oxford University Press.

Hammond K.R. (1996a). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavailable injustice.* New York: Oxford University Press.

Hammond, K.R. (1996b). Upon reflection. *Thinking & Reasoning, 2,* 239-248.

Hammond, K.R. (1996c). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K.R. Hammond (Ed.), *The psychology of Egon Brunswik (pp. 15-80)*. New York: Holt, Rinehart and Winston.

Hammond, K.R., Frederick, E., Robillard, N., & Victor, D. (1989). Application of cognitive theory to the student-teacher dialog. In D. Evans & V. Patel (Eds.), *Cognitive science in medicine: Biomedical modelling* (pp. 173-210). Cambridge, MA: The MIT Press.

Hammond, K.R., Hamm, R.M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transations on Systems, Man, and Cybernetics, SMC-17(5), 753-770.*

Hammond, K.R., Hursch, C.J., & Todd, F.J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71, 438-456*.

Hammond, K.R., McClelland, G., & Mumpower, J. (1980). *Human judgment and decision making: Theories, methods, and procedures*. New York: Praeger Scientific.

Hammond, K.R., Mumpower, J., & Smith, T.H. (1977). Linking environmental models with models of human judgment: A symmetrical decision aid. *IEEE Transactions on Systems, Man, and Cybernetics, 7*, 358-367.

Hammond, K.R., Wilkins, M.M., & Todd, F.J. (1966). A research paradigm for the study of interpersonal learning. *Psychological Bulletin, 65, 221-232.*

Hammond, K.R., Stewart, T.R. (2001). *The essential Brunswik: Beginnings, explications, applications (Eds.).* Oxford, England: Oxford University Press.

Hammond, K.R. Stewart, T.R., Brehmer, B., & Steinmann, D.O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision making.* New York: Academic Press.

Hammond, K.R., & Summer, D.A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review, 72*, 215-224.

Hammond, K. R., Todd, F. J., Wilkins, M., & Mitchell, T. O. (1966). Cognitive conflict between persons: Application of the “lens model” paradigm. *Journal of Experimental Social Psychology*, *2*(4), 343-360.

Hampshire Fire and Rescue Service (2010). *Fatal fire investigation – Hampshire Fire and Rescue Service*. HFRS, Southampton.

Harries, C., & Harvey, N. (2000). Taking advice, using information and knowing what you are doing. *Acta Psychologica*, *104*(3), 399-416.

Harte, J. M., & Koele, P. (2001). Modelling and describing human judgement processes: The multiattribute evaluation case. *Thinking & reasoning*, *7*(1), 29-49.

Hartin, E. (2009). *Fire Development and Fire Behavior Indicators.* [PDF] Available at: < http://www.cfbt-us.com/pdfs/FBIandFireDevelopment.pdf> [Assessed 28 August 2013].

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of experimental psychology: General*, *108*(3), 356.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist*, *39*(12), 1372.

Hintzman, D. L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 127-141.

HM Government (2008a). *Fire Service Manual Volume 2, Fire Service Operation: Incident Command*. London: TSO Publications.

HM Government (2008b). *Generic Risk Assessment 3.2, High rise firefighting*. London: TSO Publications.

HM Government (2013). *Emergency Response and Recovery, Non statutory guidance accompanying the Civil Contingencies Act 2004.* London: Cabinet Office.

Hoffman, R.R. (2006). *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hoffman, R. R., & Militello, L. G. (2009). Perspectives on cognitive task analysis: Historical origins and modern communities of practice. Psychology Press.

Hogarth, R.M., & Karelaia, N. (2005). Simple models for multi-attribute choice with many alternatives: When it does and does not pay to face tradeoffs with binary attributes. *Management Science, 51,* 1860-1872.

Home Office (1997). *Dealing with Disaster (3rd edition)*. London: TSO.

Horswill, M.S., & McKenna, F.P. (2004). Drivers' Hazard Perception Ability: Situation Awareness on the Road. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 155-175). Aldershot, UK: Ashgate.

Howell, W.C. (1997). Progress, prospects, and problems in NDM: A global view. In G.E. Zsambok & G.A. Klein (Eds.), *Naturalistic decision making* (pp. 37-48). Mahwah, NJ: Erlbaum.

Hu, C. H., Si, X. S., & Yang, J. B. (2010). Dynamic evidential reasoning algorithm for systems reliability prediction. *International Journal of Systems Science*, *41*(7), 783-796.

Hutchinson, J., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural processes*, *69*(2), 97-124.

Jackson, M. (2004), "Soldier Systems - The British Future Integrated Soldier System (FIST)", *Military technology.,* vol. 28, no. 11, pp. 71.

John, R.H.J. (2013). *High-Rise Building Fires.* MA: NFPA (National Fire Protection Association).

Jones, D.G., Quoetone, E.M., Ferree, J.T., Magsig, M.A., & Bunting, W.F. (2003). An initial investigation into the cognitive processes underlying mental projection. In Proceedings of the human factors and ergonomics society 47th annual meeting (pp. 596-600). Santa Monica, CA.

Juslin, P., & Montgomery, H. (Eds.). (1999). Judgment and decision making: Neo-Brunswikian and process-tracing approaches. Psychology Press.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological review*, *107*(2), 384.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D. (2003a). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kahneman, D. (2003b). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 1449-1475.

Kahneman, D., & Klein, G.A. (2009). Conditions for intuitive expertise: A failure to disagree. American *Psychological Association, Vol. 64, No. 6,* 515-526.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, MA: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47(2),* 263-291.

Kahneman, D., & Tversky, A., 1982. On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases (pp. 493-508).* New York: Cambridge University Press.

Kangari, R., & Riggs, L. S. (1989). Construction risk assessment by linguistics. *Engineering Management, IEEE Transactions on*, *36*(2), 126-131.

Karelaia, N. (2006). Thirst for confirmation in multi-attribute choice: Does search for consistency impair decision performance?. *Organizational Behavior and Human Decision Processes*, *100*(1), 128-143.

Keeley, S. M., & Doherty, M. E. (1972). Bayesian and regression modeling of graduate admission policy. *Organizational Behavior and Human Performance*, *8*(2), 297-323.

Keeney, R.L., Raiffa, H. (1993). Decisions with multiple objectives, second ed. *Cambridge University Press.*

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 115-162). New York: North-Holland.

Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In: G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds), *Decision making in action: Models and methods*, pp. 138-147. Norwood, NJ: Ablex.

Klein, G.A. (1997a). An overview of naturalistic decision making applications. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 49-60). Mahwah, NJ: Erlbaum.

Klein, G.A. (1997b). The current status of the naturalistic decision making framework. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 11-28). Aldershot, UK: Ashgate.

Klein, G.A. (1998). *Sources of Power: How people make decisions.* Cambridge, MA: The MIT Press.

Klein, G.A. (2000). Analysis of Situation Awareness from Critical Incient Reports. In: Endsley M, Garland DJ, eds. *Situation Awareness Analysis and Measurement.* Mahwah, NJ: Laurence Erlbaum Associates.

Klein, G.A. (2001). The fiction of optimization. In G. Gigerenzer, R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 103-122). Cambridge, MA: MIT Press.

Klein, G.A. (2008). Naturalistic Decision Making. *Human Factors. Vol. 50, No. 3, pp.* 456-460.

Klein, G.A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. In *Proceedings of the Human Factors Society 30th Annual Meeting*. San Diego: HFS.

Klein, G.A., Calderwood, R., & Clinton-Cirocco, A. (1988). *Rapid decision making on the fireground (Technical Report 796).* Alexandria: U.S. Army Research Institute for the Behavioral and Social Sciences.

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, *4*(3), 186-209.

Klein, G.A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19(3),* 462-472.

Klein, G.A., & Crandall, B.W. (1995). The role of mental simulation in naturalistic decision-making. In: P. Hancock, J. Flach, J. Caird, & K. Vincente (Eds), *Local applications of the ecological approach to human-machine systems, Vol. 2*, pp. 324-358. Hillsdale, NJ: Erlbaum.

Klein, G., & Hoffman, R.R. (2008). Macrocognition, mental models, and cognitive task analysis methodology. In J.M. Schraagen, L.G. Militello, T. Ormerod, & R. Lipshitz (Eds.), *Naturalistic Decision Making and Macrocognition* (pp. 57-80). Aldershot, UK: Ashgate.

Klein, G.A., Orasanu, J., Calderwood, R., and Zsambok, C.E. (1993). *Decision Making in Action: Models and Methods.* Norwood, NJ: Ablex Publishing Corporation.

Klein, G.A., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18(3),* 81-85.

Klein, G.A., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes*, 61(3).

Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods (pp. 103-137).* Norwood, NJ: Ablex

Lipshitz, R. (1994). Decision making in three modes. *Journal for the Theory of Social Behavior, 24,* 47-66.

Lipshitz, R. (2000). Two cheers for bounded rationality [Commentary]. *Behavioral and Brain Sciences, 23,* 756.

Lipshitz, R. (2010). Rigor and relevance in NDM: How to study decision making rigorously with small Ns and without controls and (inferential) statistics. *Journal of Cognitive Engineering and Decision Making, Vol. 4, No. 2,* pp. 99-112.

Lipshitz, R., Klein, G., Orasanu, J., Salas, E. (2001a). Taking stock of Naturalistic Decision Making. *Journal of Behavioral Decision Making, 14:* 331-352.

Lipshitz, R., Klein, G., Orasanu, J., Salas, E. (2001b). Rejoinder: a welcome dialogue – and the need to continue. *Journal of Behavioral Decision Making, 14:* 385-389.

Lipshitz, R., Omodei, M., McClellan, J., & Wearing, A. (2007). What's Burning? The RAWFS Heuristic on the Fire Ground. In R.R. Hoffman (Eds.) *Expertise out of context* (pp. 97-112). New York: Lawrence Erlbaum Associate.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: a naturalistic decision-making analysis. *Organizational Behaviour and Human Decision Processes, 2 (69),* 152-154.

Liu, J., Yang, J., Wang, J. & Sii, H.S. (2002). Review of uncertainty reasoning approaches as guidance for maritime and offshore safety-based assessment. *Journal of UK Safety and Reliability Society*, 23(1), 63-80.

Liu, J., Yang, J.B., Wang, J., Sii, H.S. (2005). Engineering system safety analysis and synthesis using fuzzy rule-based evidential reasoning approach. *Qual. Reliab. Eng. Int., vol. 21, no. 1, pp. 7-21*.

Lockhart, C.E. (2006). *Soldier as a system*. PA: U.S. Army War College.

Local Government Group (2011). *Fire safety in purpose-built blocks of flats*. London: LGA Publications.

MacCallum, R.C., Zhang, S.B., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, Vol. 7, No. 1,* 19-40.

Maniscalco, C. I., Doherty, M. E., & Ullman, D. G. (1980). Assessing discrimination: An application of social judgment technology. *Journal of Applied Psychology*, *65*(3), 284.

Marewski, J.N., Gaissmaier, W., Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing, 11,* 103-121.

Marewski, J.N., Gaissmaier, W., Schooler, L.J., Goldstein, D.G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multi-alternative inference. *Psychonomic Bulletin & Review, 17*, 287-309.

Marewski, J.N., Schooler, L.J. (2011). Cognitive niches: An ecological model of strategy selection. Psychological Review, 118, 393-437.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision, 52*, 29-71.

McCloy, R., Beaman, C. P., & Smith, P. T. (2008). The Relative Success of Recognition-Based Inference in Multichoice Decisions. *Cognitive Science*, *32*(6), 1037-1048.

McLennan, J., & Omodei, M.M. (1996). The role of prepriming in recognition-primed decision making. *Perceptual and Motor Skills, 82*, 1059-1069.

Meyers-Levy, J. (1989). The influence of a brand name's association set size and word frequency on brand memory. *Journal of Consumer Research*, 197-207.

Miesing, P., & Dandridge, T. C. (1986). JUDGMENT POLICIES USED IN ASSESSING ENTERPRISE-ZONE ECONOMIC SUCCESS CRITERIA*. *Decision Sciences*, *17*(1), 50-64.

Militello, L.G., & Hutton, R.J.B. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, Special Issue: Task Analysis, 41(11),* 1618-1641.

Minsky, M. (1986). The Society of Mind. New York, NY: Simon and Schuster.

Montgomery, H., Lipshitz, R., & Brehmer, B. (2004). *How Professional Make Decisions*. Mahwah, NJ: Lawrence Erlbaum Associate.

Moon, B. (2002). Naturalistic Decision Making: Establishing a naturalistic perspective in Judgment and Decision Making research. *Advanced Decision Architectures Collaborative Technology Alliance cooperative agreement DAAD19-01-2-0009,* US Army Research Laboratory.

Mosier, K.L., & Fischer, U.M., (2010). *Informed by Knowledge: Expert Performance in Complex Situations.* Sussex, UK: Taylor & Francis.

Murphy, C.K. (2000). Combining belief functions when evidence conflicts. *Decision Support Systems 29, 1-9.*

Negoita, C. V., Zadeh, L. A., & Zimmermann, H. J. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, *1*, 3-28.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113-126.

Newell, B.R., Rakow, T., Weston, N.J., & Shanks, D.R. (2004). Search strategies in decision making: The success of "success". *Journal of Behavioral Decision Making, 17*, 117-137.

Newell, B.R., & Shanks, D.R., (2003). Take the best or look at the rest? Factors influencing "one-reason" decision-making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 53-65.

Newell, B.R., & Shanks, D.R., (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 923-935.

Newell, B.R., Weston, N.J., & Shanks, D.R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "take-the-best". *Organizational Behavior and Human Decision Processes, 91*, 82-96.

Ng, K. C., & Abramson, B. (1990). Uncertainty management in expert systems. *IEEE Intelligent Systems*, *5*(2), 29-48.

Nisbett, Richard E., and Timothy D. Wilson. "Telling more than we can know: Verbal reports on mental processes." *Psychological review* 84.3 (1977): 231.

O'Keefe IV, J.A., Mcintyre III, R.T., & Middleton, V.E. (1992). Modelling and Simulation of the Soldier as a System. *Technical Paper NATICK/TP-10/005.* Massachusetts: U.S. Army Natick Soldier Research, Development and Engineering Center Natick.

O'Keefe IV, J.A., & Porter, C.D. (1992). Soldier System Modelling and Integrated Unit Simulation System (IUSS). *Technical Paper NATICK/TP-10/002*. Massachusetts: U.S. Army Natick Soldier Research, Development and Engineering Center Natick.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods (pp. 3-20).* Norwood, NJ: Ablex.

Pascual, R., & Henderson, S. (1997). Evidence of naturalistic decision making in military command and control. *Naturalistic decision making*, 217-226.

Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety*, *54*(2), 95-111.

Payne, J.W., Bettman, J.R., & Johnson, E.J. (1993). *The adaptive decision maker*. Cambridge: Cambridge University Press.

Pearl, J. (1983). *Heuristics: Intelligent Search Strategies for Computer Problem Solving.* New York, Addison-Wesley.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, Inc..

Phelps, R.M., & Shanteau, J. (1978). Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance, 21,* 209-219.

Pinker, S. (1997). How the mind works. 1997. *NY: Norton*.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological review*, *109*(3), 472.

Postman, L., & Tolman, E.C. (1959). Brunswik's probabilistic functionalism. In S. Koch (Ed.), *Psychology: A study of science* (pp. 502-564). New York: McGraw-Hill.

Pruitt, J.S., Cannon-Bowers, J.A., & Salas, E. (1997). In search of naturalistic decisions. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 29-42). Aldershot, UK: Ashgate.

Rakow, T., Hinvest, N., Jackson, E., & Palmer, M. (2004). Simple heuristics from the adaptive toolbox: Can we perform the requisite learning? *Thinking and Reasoning, 10*, 1-29.

Rakow, T., Newell, B. R., Fayers, K., & Hersby, M. (2005). Evaluating three criteria for establishing cue-search hierarchies in inferential judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1088.

Rasmussen, J. (1983). Skill, rules and knowledge: Signals, signs and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13(3),* 257-266.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering.* Amsterdam: North-Holland.

Reilly, B. A., & Doherty, M. E. (1989). A note on the assessment of self-insight in judgment research. *Organizational Behavior and Human Decision Processes*, *44*(1), 123-131.

Reilly, B. A., & Doherty, M. E. (1992). The assessment of self-insight in judgment policies. *Organizational Behavior and Human Decision Processes*, *53*(3), 285-309.

Reimer, T., & Hoffrage, U. (2011). Ecological Rationality for Teams and Committees: Heuristics in Group Decision Making. In P.M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological Rationality: Intelligence in the world* (pp. 335-359). New York: Oxford University Press.

Rieskamp, J., & Dieckmann, A. (2011). Redundancy: Environment Structure That Simple Heuristics Can Exploit. In P.M. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological Rationality: Intelligence in the world* (pp. 187-215). New York: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P.M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 141-167). New York: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica, 127*, 258-276.

Rieskamp, J., & Otto, P. (2006). SSL: a theory of how people learn to select strategies. *J. Exp. Psychol.: Gen. 135*: 207-236.

Rohrbaugh, J. (1988). Cognitive conflict tasks and small group processes. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 199-226). New York: North-Holland.

Romero, C. (1986). A survey of generalized goal programming (1970-1982). *European Journal of Operational Research, 25*, 183-191.

Ross. K.G., Shafer, J.L., & Klein, G.A. (2006). Professional Judgments and "Naturalistic Decision Making". In K.A. Ericsson, N. Charness, P.J. Feltovich, R.R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 403-420). New York: Cambridge University Press.

Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: Prospects and limits on the search for mental models. *Psychological Bulletin, 100,* 349-363.

Roy, B. (1990). The outranking approach and the foundations of ELECTRE methods. In C.A. Bana e Costa (Eds.), *Readings in Multiple Criteria Decision Aid* (pp. 155-183). Berlin: Springer.

Saaty, T.L., (1980). *The Analytic Hierarchy Process. New York:* McGraw-Hill.

Saaty, T.L., (1994). *Fundamentals of decision making with the analytic hierarchy process*. USA:RWS Publications.

Salas, E. and Klein, G. (2001). *Linking Expertise and Naturalistic Decision Making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Samurcay, R. & Rogalski, J. (1988). Analysis of operators cognitive activities in learning and using a method for decision making in public safety. In J. Patrick & K.D. Duncan (Eds.), *Training Human Decision Making and Control*. Amsterdam: Elsevier Science.

Savage, L.J. (1954). *The Foundations of Statistics.* New York: Wiley.

Schmitt, N., Gleason, S. E., Pigozzi, B., & Marcus, P. M. (1987). Business climate attitudes and company relocation decisions. *Journal of applied psychology*, *72*(4), 622.

Schraagen, J.M., Klein, G., Hoffman, R.R. (2008). The macrocognition framework of naturalistic decision making. In: J.M. Schraagen, L. Militello, T. Ormerod, & R. Lipshitz (eds.)*, Naturalistic Decision Making and Macrocognition* (pp. 3-27). London: Ashgate.

Schraagen, J.M., Militello, L., Ormerod, T., & Lipshitz, R. (2008). *Naturalistic Decision Making and Macrocognition.* London: Ashgate.

Sedlmeier, P., Hertwig, R., & Gigerenzer, G. (1998). Are judgments of the positional frequencies of letters systematically biased due to availability?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 754.

Sentz, K., & Ferson, S. (2002). *Combination of evidence in Dempster-Shafer theory* (Vol. 4015). Albuquerque, New Mexico: Sandia National Laboratories.

Shafer, G. (1976). A mathematical theory of evidence. *Princeton: Princeton University Press.*

Shafer, G. (1990). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, *4*(5), 323-362.

Shafer, G. (1992). The Dempster-Shafer theory. *Encyclopedia of artificial intelligence*, 330-331.

Shanks, D.R., Charles, D., Darby, R.J., & Azmi, A. (1998). Configural processes in human associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1353-1378.

Shanteau, J. (1992). How much information does an expert use? Is it relevant?. *Acta Psychologica*, *81*(1), 75-86.

Si, X.S., Hu, C.H., Zhou, Z.J. (2010). Fault prediction model based on evidential reasoning approach. *Science China: Information Sciences, 53(10), 2032-2046.*

Simon, H.A. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics, Vol. 69, No. 1,* pp. 99-118.

Simon, H.A. (1990). Invariants of human behaviour. Annu. *Rev. Psychol. 41,* 1-19.

Smith, K. & Hancock, P.A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, 37, 137-148.

Smith, L., & Gilhooly, K. (2006). Regression versus fast and frugal models of decision-making: The case of prescribing for depression. *Applied Cognitive Psychology, 20*, 265-274.

Smith, T.H. (1975). Social judgment theory and methods applied to university planning (Report No. 190). Center for Research on Judgment and Policy, University of Colorado, Boulder, CO.

Stewart, T.J. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *OMEGA International Journal of Management Science, 20*, 569-586.

Stewart, T.R. (1988). Judgment analysis: Procedures. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment The SJT view* (pp. 41-74). New York: North-Holland.

Stewart, T.R., & Ely, D.W. (1984). *Range sensitivity: A necessary condition and a test for the validity of weights* (Report No. 270). Center for Research on Judgment and Policy, University of Colorado, Boulder, CO.

Stewart, T. R., Heideman, K. F., Moninger, W. R., & Reagan-Cirincione, P. (1992). Effects of improved information on the components of skill in weather forecasting. *Organizational behavior and human decision processes*, *53*(2), 107-134.

Stokes, A., Belger, A., and Zhang, K. (1990). Investigation of factors comprising a model of pilot decision making: Part 2. Anxiety and cognitive strategies in expert and novice aviators. ARL-90-8/SCEEE-90-2. Urbana-Champaign, IL: Institute of Aviation.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 111-147.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44-47.

Sverson, O. (1983). Decision rules and information processing in decision making. In L. Sjoberg, T. Tyszka & J. Wise (Eds.), *Human decision making* (pp. 131-162). Bodafors, S: Doxa.

Taylor, R.M. (1990). Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In: *Situational Awareness in Aerospace Operations (AGARD-CP-478)* (pp. 3/1-3/17). Neuilly Sur Seine, France: NATO-AGARD.

Todd, F.J., & Hammond, K.R. (1965). Differential feedback in two multiple-cue probability learning tasks. Behavioral Science, 10, 429-435.

Todd, R.M., & Dieckmann, A. (2005). Heuristics for ordering cue search in decision making. In L.K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in neural information processing systems* (Vol. 17, pp. 1393-1400). Cambridge, MA: MIT Press.

Todd, P.M., & Gigerenzer, G. (2001). Putting naturalistic decision making into the adaptive toolbox. *Journal of Behavioral Decision Making, 14(5),* 381-383.

Todd, P. M., & Gigerenzer, G. (2003). Bounding rationality to the world. *Journal of Economic Psychology*, *24*(2), 143-165.

Todd, P.M., & Gigerenzer, G. (2011). What is Ecological Rationality? In: P.M. Todd, G. Gigerenzer, & The ABC Research Group (eds.), *Ecological Rationality: Intelligence in the World.* Oxford, UK, pp. 3-30.

Todd, P.M., Gigerenzer, G., & The ABC Research Group (2011). *Ecological Rationality: Intelligence in the World.* Oxford, UK: Oxford University Press.

TRADOC Pamphlet 525-97, (2006). *Military Opereations Soldier as a System*. Virginia: Department of the Army Headquarters, United States Army Traing and Doctrine Command Fort Monroe.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, *79*(4), 281.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

U.S. Department of the Army, (1991). *Army Science Board Summer Study, Final Report, Soldier as a System*. Washington: OASA.

Verhagen, T.L.A., Woering, A.A., Ubink, E.M., Knijnenburg, S.G., de Jong, K.Y. & van Helden, H.P.M. (2008). "Technical Optimisation of the ``Soldier System'' The Physical

Human Centric Modelling Approach", *MILITARY TECHNOLOGY,* vol. 32, no. 6, pp. 144-151.

Victor, M., Ken, S., Bob, M., John, O.I. (2000). Soldier Integrated Protective Ensemble (SIPE) Advanced Technology Demonstration (ATD). *Technical Report NATICK/TR-00/019.*

Von Neumann, J., Morgenstern, O. (1944). *Theory of Games and Economic Behaviour.* Princeton, N.J.: Princeton University Press.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge University Press.

Wallasch, K., & Stock, B. (2009). Single Living [pdf]. FRM Journal. Available at: < https://www.frmjournal.com/MainWebSite/Resources/Document/Single%20living.pdf> [Accessed 17 January 2014].

Wang, Y.M., Yang, J.B., Xu, D.L. (2006). Environmental impact assessment using the evidential reasoning approach. *European Journal of Operational Research 174: 1885-1913.*

Wang, Y.M., Yang, J.B., Xu, D.L., Chin, K.S. (2006). The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees. *European Journal of Operational Research 175(1): 35-66.*

Wierzbicki, A.P. (1980). The use of reference objectives in multiobjective optimization. In G. Fandel & T. Gal (Eds.), *Multiple Criteria Decision Making Theory and Application (pp. 468-486)*. Berlin: Springer.

Williams, D.A., & Braker, D.S. (1999). Influence of past experience on the coding of compound stimuli. Journal of Experimental Psychology: Animal Behaior Processes, 25, 461-474.

Wimsatt, W. C. (2007). Re-engineering philosophy for limited beings: Piecewise approximations to reality. Harvard University Press.

Wong, B.L.W., & Stanton, N.A. (2009). *Naturalistic Decision Making and Computers: Proceedings of the 9th Bi-annual International conference on Naturalistic Decision Making*. London: British Computer Society.

Woods, D.D. (1988). Coping with complexity: The psychology of human behaviour in complex systems. In L.P. Goodstein, H.B. Andersen, & S.E. Olsen (Eds.), *Tasks, errors, and mental models*. London: Taylor & Francis.

Xu, L., & Yang, J.B. (2001). Introduction to multi-criteria decision making and the evidential reasoning approach. *Working Paper No. 0105, Manchester School of Management, ISBN 1 86115111 X.*

Xu, D.L., Yang, J.B., Liu, J.P., Wang, J., Jenkinson, I. (2007). Inference and learning methodology of belief-rule-based expert system for pipeline leak detection. *Expert Systems with Applications, 32(1), 103-113.*

Yang, J.B. (2001). Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. *European Journal of Operational Research 131 (1), 31-61.*

Yang, J.B., Wang, J., Bonsall, S., Fang, Q.G. (2009). Use of fuzzy evidential reasoning in maritime security assessment. *Risk Analysis, vol. 29, no. 1.*

Yang, J.B., Sen, P. (1994). A general multi-level evaluation process for hybrid MADM with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics 24 (10), 1458-1473.*

Yang, J.B., Singh, M.G. (1994). An evidential reasoning approach for multiple attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 32(3), 289-304.*

Yang, J.B., Xu, D.L. (2000). *Intelligent Decision System via Evidential Reasoning, Version 1.1*, Cheshire, England.

Yang, J.B., Xu, D.L. (2002a). On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 32(3): 289-304.*

Yang, J.B., Xu, D.L. (2002b). Nonlinear information aggregation via evidential reasoning in multiattribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans 32(3): 376-393.*

Yang, L., Prasanna, R., & King, M. (2009). On-Site Information Systems Design for Emergency First Responders. *Journal of Information Technology Theory and Application, 10(1),* pp. 5-27.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338-353.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions on*, (1), 28-44.

Zakay, D. (1997). Post-decisional confidence – can it be trusted? In R. Ranyard, W.R. Crozier, & O. Svenson (Eds.). *Decision making: Cognitive models and explanations* (pp. 233-246). London: Routledge.

Zimmer, I. (1981). A comparison of the predictive accuracy of loan officers and their linear-additive models. *Organizational Behavior and Human Performance, 27*, 69-74.

Zimmermann, H.J. (2000). An application-oriented view of modelling uncertainty. *European Journal of Operational Research, 122*, 190-199.

Zsambok, C. (1997). Naturalistic decision making: where are we now? In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 3-16). Mahwah, NJ: Erlbaum.

Zsambok, C. & Klein, G. (1997). *Naturalistic Decision Making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Zurada, J. M. (1992). *Introduction to artificial neural systems* (Vol. 8). St. Paul: West publishing company.

Appendix A: Field context of high rise apartment fire

For firefighting purposes, a high rise building is considered to be one containing floors at such a height or position, or design that external firefighting and rescue operations may not be feasible or practicable (HM Government 2008b). A more straightforward definition is a building with 8 floors or more. The following illustrates those cues we identified that are essential in the response of fire incident.

Staircases

The staircases within high-rise buildings are required to be enclosed in fire-resisting material and access to them should be through self-closing fire-resisting doors. As well as providing a means of escape for residents, staircases also provide vital access for the fire crews to reach affected floors at an incident. It is, theoretically, best to have more than one staircase to provide dual-pathway of fire and rescue operations. Although there is no requirement of the number of staircases to be provided within residential flats has been established in UK (Wallasch & Stock, 2009), most of the high rise apartment are built with at least two staircases. In addition to the number of staircase, another issue in terms of staircase is its working conditions. During an incident, a compromised stair enclosure could be identified due to unclosed fire doors, lack of maintenance or malicious damage. Such conditions would require additional resources to deal with the circumstance. One interviewee commented on this that "… *definitely require more resources, if you get structured staircases that compromised by fire, then it is going to be require more resources from that point of view because your bridge has to be confirm away from fire floor, depends on what the phase you get there, you may already lost, it is probably more aloud experience in the developing stage*".

In terms of most commonly encountered status of staircase, further comments from interviewees would be like "*you could have broken doors, fire doors which mean it fire could penetrate very quickly from one floor to another, or into the protecting staircase if the self-closing mechanism on the door doesn't work, so generally you will found well maintained building and nothing will be problem, all the doors will be fine, if you find it's a badly maintained building, you will find debris on the floor and doors are probably*

broken and the dry riser are probably damaged"; "*it is always protected staircase in high rise building if maintenance is good*".

Dry Risers

A dry riser is a supply system intended to distribute water to multiple levels or compartments of a building, as a component of its firefighting systems. It is a building regulations requirement in residential buildings exceeding 11 metres in height but not exceeding 50 metres (superior protection of wet risers in buildings over 50 metres). It typically consists of three components: external inlets at the ground of the building for connection of Fire service water supplies, pipe work usually enclosed within fire resisting enclosures or shafts, as well as outlets in floors to enable the fire service to attach and advance its hose lines within a building. A defective or otherwise inoperative dry riser will seriously effect fire service operations. Riser failure can be brought about through many factors, such as an open outlet, restrictive blockage, mechanical failure of the pipe work, vandalism and more recently theft of the inlet/outlet fittings (Fishlock, 2013). We excluded this essential information in the present task for two reasons. Firstly, the decision point we set is the time point that normally the acknowledged dry riser condition is not available. According to the SOPs in Leicester FRS, at the time point, there would be two fire crews sent into the building to particularly make sure every riser outlet in building is closed. However, the full conditions like restrictive blockage, vandalism, and their effect on the water flow will only be confirmed once the bridgehead is set up and attacking is starting. Secondly, the dry riser is so essential to the high rise building that it is always well inspected during the FRS's regular visit of the site. It is therefore, for those well maintained building, dry rise is mostly in working order. And it is default condition that would be assumed by the ICs when they are making decisions at the initial stage of the fire incident.

Lifts

Lifts can be an invaluable resource to fire operations, as firefighters working in full Personal Protective Equipment (PPE) and carrying heavy equipment up stairwells will

suffer from heavy degrees of exertion and exhaustion. However, the lift would distinct between the normal passenger lift and the fire fighting lift. It is not always that a fire fighting lift is installed in any high rise building, especially for those tower blocks pre-date any nationally approved standard. In such case, the lift would normally be required to involve a Fireman's Switch (lift over-ride switch) for firefighters fully control the lift during a fire incident. There are also a variety of safety installations available to a lift system, including overload detectors, door close protection sensors, emergency lighting, emergency communication equipment / alarm button, floor level detection, smoke detection and buffer stops. Theoretically, a 'balancing pros and cons' analysis is generally recommended for those normal lift that ICs would use their acknowledged ability to risk assess if the lift can or should be used during an incident. In practice, the ICs would not be involved in such detailed trade-off, and would normally choose to use the lift as long as they could fully control it. "*It's rare we lose the lift, we will still use the lift with light smoke unless dense smoke in it*" (a comment from one interviewee). In this sense, although theoretically is critical, the issue that whether the building is installed with firefighting lift or normal lift makes little difference to the ICs, especially in the assessment of the resources demand. One interviewee commented on this: "*… so a lot safer if you got firefighting proved lift, but little difference about the resources demand …*"

Ventilation systems / Smoke control system

Ventilation systems are designed to remove unpleasant smells and excessive moisture, to introduce outside air, to keep interior building air circulating, and to prevent stagnation of the interior air. It is a requirement that ventilation systems are provided to common corridors and stairs in apartment buildings to protect from smoke and to help maintain tenable conditions for means of escape in the event of fire. Methods of ventilation are typically divided into two types: natural ventilation, and mechanical ventilation.

Natural ventilation harnessing naturally available forces to supply and remove air in the building, including pressure-driven which relies upon the buoyancy of heated or rising air ('stack effect'), as well as wind driven which relies upon the force of the prevailing wind to pull and push air through the enclosed space.

Mechanical ventilation circulates fresh air using ducts and fans, rather than airflow through windows. The most commonly applied mechanical ventilation systems include mechanical extraction, staircase pressurisation, as well as automatically operating vent system normally activated by the automatic fire alarm system in the building.

Firefighters are trained to initiate and control ventilation when they feel appropriate for the benefit of their fire operations. It is generally agreed that mechanical ventilation systems are to be relatively a positive feature. But in terms of the resources demand, the impact difference is minimum. "*Big difference for the operation, but no difference for resources demand, mechanical could be a lot easy for operation*" (a comment from one interviewee).

Water supply

Under building regulations in the UK (Communities and Local Government, 2007) there would be one hydrant within 18M of any riser inlet. Secondary hydrants, ideally from separate mains, are also recommended to be identified during FRS Site Specific Risk Information (SSRI) investigation. The realistic of water supply being poor is extremely small. Therefore the most significant issue of water during a fire incident is to actually the condition of riser. "*… water source is always sufficient. if it is not, we are likely to be told and record in our MDT, as sometimes you got road works anything like that, we should get notified if water supply get effected by any those work, and we will amend up the PDA"* (a comment from one interviewee).

Wind effect

Fire growth and spread may be exacerbated by strong winds. Wind speed generally increases with height. Nearby buildings will also have a marked effect on local wind

currents. High rise buildings magnify wind speed and can create complicated and unpredictable wind patterns, causing hazard of 'blow torch', increasing the likelihood of flashover, 'plane' the falling objects and burning debris, etc. (HM Government, 2008b). Any additional hazard, as well as rapid fire development driven by wind will demand altered operational plan and require more resources to take control of. Therefore, the strong wind is considered as a strong indicator of more serious situation. The present descriptions of wind effect indicate the general strength and direction of wind. One interviewee commented on this:

"*That is very important, if we got high wind and fire is high, then we know that should that become a ventilated fire is really going to cause huge problems, likewise if it is unventilated fire and then we ventilate it, it was going to cause huge problems as well. So high wind dictated that we've got big problems. That would possibly change the strategy for your firefighting as well. The description gives enough to be able to work, more information is always better, but this would be sufficient information for me there to be able to make assessment … basically you should put there whether the building is blowing into, so it was going to directly affect our fire, … so just knowing which direction is going with regards to the building, such as if it from left to right, from back to front, and then the location of fire in comparison to where the wind is important as well*."

Building evacuation plan

There are basically 3 pre-planned evacuation policies that are likely to be encountered in high rise building: 'stay put', simultaneous evacuation, and phased (partial) evacuation (Local Government Group, 2011). The type of policies encountered is highly associated with the fire detection and alarm system installed in high rise building.

'Stay put' policy is a relatively recent concept aligned with the technology advances of compartmentation. It is used to minimise the number of people evacuating. Those that deem themselves not at direct risk from the fire or smoke will stay in their compartment and use the compartmental passive fire protection of the building to keep them safe

until the FRS extinguish the fire. Those that deem they are at risk area should evacuate. It is the most widely used policy in high rise apartment.

The alternative to a 'stay put' policy is one involving simultaneous evacuation. It requires a means to alert all of those residents to the need to evacuate, i.e. a communal fire detection and alarm system. It is sometimes applied to buildings converted into blocks of flats, but usually only where it has not been possible to achieve the level of compartmentation required for a 'stay put' policy. Purpose-built blocks of flats are not normally applied with such policy (Local Government Group, 2011).

A hybrid and preferred policy by FRS is the phased (partial) evacuation where an intelligent, pre-programmed alarm system is able to notify those at risk within certain areas of the building to evacuate, but does not order evacuation of those deemed as currently 'not at risk' with just a pre-warning alarm. However, it is rarely used in UK due to the requirement of new intelligent fire alarm system. It is most commonly encountered as the subsequent policy of 'stay put', when the development of the fire necessitates the FRS to organize the area based evacuation for operational reasons. In such cases, the IC must determine and control the most appropriate way to manage the residents of the building.

To be noted, actual evacuation behaviour of residents during an incident would not strictly follow the pre-planned policy. It would be common circumstance that residents self-evacuated from 'stay-put' building before FRS on arrival, or residents stay-put in simultaneous evacuation building. One interviewee commented on this: "… *anything is possibly, even on the stay put policy you may get where the floor of the origin has to evacuate but the one above and below don't …*" However, it is always a positive feature that residents could have self-evacuated themselves. In addition, the most recent incident (Hampshire Fire and Rescue Service, 2010) indicated that the 'stay-put' might not work.

Fire detection and alarm system

An automatic fire detection and alarm system (AFD) would typically include the following components: fire detectors (e.g., smoke and heat); manual call points (break glass call points) next to exits with at least 1 call point on each floor; electronic sirens or bells; and a control and indicator panel. In addition to the effect on the building evacuation policy, the control panel of the AFD is generally considered as an essential information source for ICs to identify the source of fire and building areas affected at the initial stage of any incident. Its usefulness and efficacy, again, depends on the type of system installed. Generally, there are two system categories being encountered in high rise apartment: conventional and addressable.

In a conventional AFD, a number of call points and detectors are wired to the control panel in zones (pre-divided building sections). Once any detector trigger, the corresponding zone will display on the panel. That is, you cannot pinpoint exactly which device has been activated. An addressable AFD, however, provide more advanced programs to identify individual alarms, and potentially displaying initial detection times, development rate, areas affected.

One interviewee commented on the most common situation of AFD that "*it will give you the floor and possibly the flat number, but it won't exactly say which bedroom is in or anything like that. Solely AFD is absolutely not enough, I will get as much information as possible … AFD is very good guide, but I am not taking it as per se, not 100%.*"

Compartmentation & Floor layout

Normally, a building layout map would be associated with the AFD panel. Such layout map indicates the zoning of the building and the compartmentation of each floor.

Compartmentation is known as the high degree of fire separation between flats and the common parts achieved by making each flat a fire-resisting enclosure. A compartment simply a part of building bounded by walls and floors that will resist the passage of fire for a specified period of time (Local Government Group, 2011). Theoretically, the fire resistance of this construction is such that, normally, a fire will burn itself out before

spreading to other parts of the building. It is mostly a horizontal area, typically a room, flat or corridor, but may also be a vertical area, namely a staircase or lift shaft. It is the basic design principle and the most useful design aspect of high rise building in relation to fire.

To be noted, in addition to the compartmenatation layout, there is another version of layout that indicates the detailed floor and flat plan. It is typically available in the "Premises Information Box" in the lobby/reception area of the building, which is designed to store essential building information for use by fire crews so that immediate decisions can be made in the absence of the owner/occupier. Large or complex floor layouts and a lack of information on the internal layout of the building would challenge fire crews seeking safe access and egress routes to and from the scene of the fire and may increase the risk of crews becoming disorientated or lost (HM Government, 2008b; example could see Hampshire Fire and Rescue Service, 2010). However, "*it's rare that you got a complex layout in the high rise unless you got a big building where it's like an old factory. If it is just a square block straight up, a tower, there is no complex layout. The only time it will be slightly more complex if it was a converted factory that being made into flat*" (a comment from one of our interviewees).

General construction

The majority of the UK residential high rise buildings are built of reinforced concrete. Theoretically, many factors will affect how concrete will behave under fire conditions. These may include: quantity and type of aggregate used in the mix, thickness (and thus protection of reinforcement), type of cement used, water content of the concrete, loading bearing, fire exposure time, temperature, application of water (Firefighting Jets), cladding or covering, and age. It is however, in fire operations, ICs would never think of these detailed factors. They generally trust the reinforced concrete and compartmentation as mention above. Collapsing would not be a consideration unless the incident, like '911'. It is generally accepted the concrete is able to provide up to 4 hours protection, and is considered as the stoutest of all construction types when

exposed to fire. However, at temperatures of between 400-600C, explosive surface spalling can start to occur within 30-60mins of most concrete, causing the hazard of concrete falling on the firefighters post entry (Fishlock 2013).

Occupancy profile

Occupancy profile indicates the number and type of the occupants of the building. The level and nature of occupation and socio-economic factors may adversely impact on the ability of the firefighters to carry out operations. One example is illustrated in HM Government (2008b) that anti-social behaviour ranging from verbal attacks through to physical assaults has been reported in areas of social deprivation. However, "*it's not common, possibly London Manchester Liverpool I think, but Leicester we don't attend a great deal, we have some but not a great deal. So mainly would be physical ability of individual to self-evacuate*" (a comment from one of our interviewees). The physical ability of individual to self-evacuate here could refer to those sleep risk in the night, or those elderly and disabled residents.

Maintenance status

Maintenance status is essential. As the fire safety facilities in building (e.g, compartmentation, fire door) are not under the direct control of the FRS. Their effectiveness and availability cannot be taken for granted but are dependent upon the vigilance of the responsible person (including building owner/occupier) and premises management. Compromised facilities, such as the fire door not shutting right to hold the floor, stair enclosure compromised for malicious damage, would serious effect the fire operations. It normally highly related to the occupancy profile of the building. For example, in areas of social deprivation, vandalism could be commonly identified.

Sprinkler system

A fire sprinkler system is an active fire protection measure, consisting of a water supply system, to detect and suppress a fire at the earliest stage possible. It is one significant safety measure that contributes to the fire emergency response. One of our interviewee

commented on this: "*It could be the sprinkler system isn't adequate to put the fire out, but are likely hold it less, the chance of developed fire is less, but is absolutely possible*". However, it is mainly used in factories and large commercial buildings. Systems for apartment building are only available in modern building. The building utilized in our present study is too old to have such system. We therefore excluded this information from the present study.

Height of the fire (floor level)

The height of the incident floor(s) will have a profound effect of the tactics used. The fire below 3rd storeys could be approached in a similar way to other domestic compartment fires. While for upper floors, key items of FRS equipment such as appliances, ladders, will be of limited use. The physiological effect of deploying crews would impact upon the speed of deployment and effectiveness of crews. External ventilation (wind effect) will be influential. The use of dry riser become to be a necessary, and flow rate will be reduced and may affect branch performance.

Falling objects and burning debris

Falling objects is an essential hazard specific to high rise building, and would always be looked for by the ICs. Failed windows or balcony doors can pre-warn of an extensive high energy / long burning fire and possible vertical fire spread. The falling debris can be ejected explosively from the building or in the case of glass and curtain walling can 'plane' and travel over a considerable distance. It is hazardous to personnel working at ground level and may cause secondary fires. However, in terms of resource demand, the influence is minimum. One interviewee commented on this: "*… again it is very important, we set our safety code that it is important we don't put people into a risk area if there is debris falling down inside of building, then that is a risk area when you need move thing back all the way. so it is important information. in terms of resource, if there are items, you got to need more people to deal with it, as you need create safety cordon, again we would try to put hands to police, but initially we would have to deal with it, that may be a draw on your initial resources, it may certainly indicate a*

ventilated fire and something is failed and stuffs are falling out, that gives more information with regards the ventilation of the fire as well."

Weather temperature

Normally, except for the wind effect illustrated above, weather conditions like temperature, rainy or dry, would not be a great issue to fire operations. One exception is the weather with extreme hot or cold temperature. In extreme cold day, the "stack effect" may become significant, in which the temperature difference between inside and outside a building creates constant airflow either open windows, ventilation openings, or other forms of leakage. In case of a fire, such effect increase the risk of smoke and fire spread, especially in those vertical shafts like staircase, refuse chute, and need to be controlled to maintain tenable conditions for victims and firefighters. In extreme hot weather, another issue arise that firefighters may more prone to heat stress. One interviewee commented on this: "*if it is extremely hot, you want more resources there, because it is very difficult to deal with the task if you are contained within your fire kit, you extremely hot are going to become tired, have difficulty performing, but probably not immediately, but within the first 10 to 15 minutes, I wouldn't suggest calls due to that straight away, not until the incident start to develop, you get more information towards*".

Incident Location

Incident location has two main effects. One is in terms of traffic. It may affect which way the fire appliances will approach incident from. And the traffic condition may stop them getting to the incident as quick as they want to. Another effect is in terms of resources. As one of the interviewees commented: "*again if it is city centre, we know we have a large amount of resources based on the city centre, more fire engine, if it is in suburban or village, you know the resources are going to take much longer to get to it. Yes I would suggest would probably a consider, make it extra resource at an early stage*".

Time of the incident

Incident time is essential when it relates to hazard. The most regularly encountered example is the sleep hazard, in which a fire happened in midnight, and most people would be expected to be inside sleeping. And when there is an escalating situation, FRS may require more resources to wake them from sleep. One interviewee commented: "*let me elaborate more about the time, my experience would suggest that if the incident happened around or at the midnight, my expectation is people would be inside sleeping, if it happened in mid-day, there is greater chance of being out of the occupants whether the university college or work, so the occupancy matters … the only difference between time will relate to occupancy …"* Another interviewee also commented on this: *"… so the time itself make no difference, it will make difference only when it links some particular risks, like sleep risk etc*."

Search & Rescue

Search & rescue could refer to two dimensions. At the wider dimension, the FRS sectorizes any high rise incident with a three sectors structure as illustrated in Figure A-1. Search and rescue here refers to the work at the search sector. It is part of the evacuation process. At the narrow dimension, search & rescue refers to the task of rescuing any person trapped in the fire. UK FRS, in practice, uses the term "persons reported" to indicate that there are believed to be persons in a building that are trapped and require rescuing; and uses the term "all persons accounted for" to indicate no persons to be rescued. Philosophically, FRS will take some risk to save saveable lives, and will not take any risk at all to try to save lives or properties that are already lost (HM Government, 2008a). A saveable life situation here refers to one in which the FRS is assured that a person remains trapped and an immediate entry into the compartment will result in rescue. Such scenarios may include: persons trapped still on phone to FRS control; persons trapped recently observed from outside building; persons heard in incident apartment; information from local resident confirming other survivors trapped. To determine whether a rescue is appropriate is critical to ICs, and demanding serious trade-off between the benefit and risk. There is some debate over priority of fire control

and rescue in domain (Grimwood, 2014). As a consideration, the circumstance of more than one person reported would generally alter the resources demand. One interviewee commented on this: "*… depending the circumstance, if one person there, the resource would be the same, if we got two or three person reported, I would make call to have another two or one fire engine there, that definitely alter with the amount of people from that assessment …*"

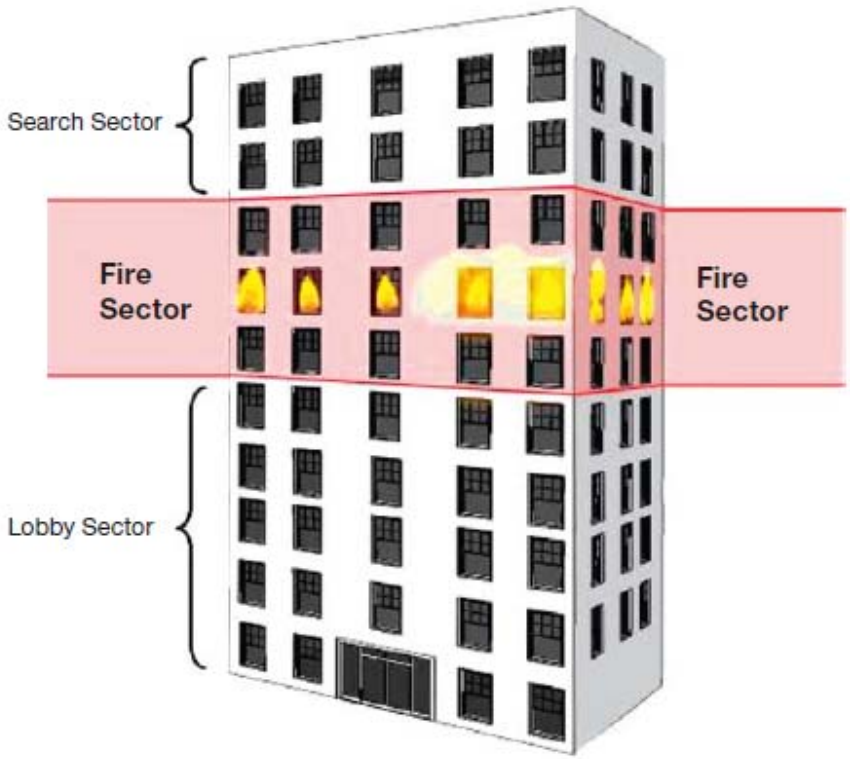


Figure A-1, Sector structure of high rise fire (HM Government, 2008a)

Fire flame & smoke

Fire flame and smoke are the most straightforward indicators of the fire status. To make clear of this information, we would like to address three concepts first. The first concept is the stages of fire development. Compartment fire development could be described as being comprised of four stages: incipient (or initial), growth (or developing), fully developed, and decay (see Figure A-2). The second concept is the hazards of flashover and backdraught. The third concept is the discrimination between ventilated fire and

under/un-ventilated fire. Recognizing these concepts and correspondent fire behaviour marks a skilled firefighter who is able to “read a fire” and make decision accordingly.

Briefly, the hazards of flashover and backdraught are the two significant risks that fire operations are always trying to avoid. Firefighters are trained specifically to identify the signs and counter act the effect (see details in Communities and Local Government, 2009). One simplified and generally agreed process is to recognize the ventilation status of the fire and its development stage.

According to the FRS department definition (Communities and Local Government, 2009), a flashover indicates the sudden and sustained transition of a developing fire to a fully developed fire (as see in Figure A-1). The transition make the compartment absolutely unsustainable for any people where the total thermal radiation from the fire plume, hot gases and hot compartment boundaries (ceilings and walls) causes the radiative ignition of all exposed combustible surfaces within the compartment. It would be mostly happened in ventilated situation, where fresh air (sufficient oxygen) is constantly feeding the burning so that speeding up the heating of the compartment. When flashover occurs, burning gases will push out openings in the compartment (such as a door leading to another room) at a substantial velocity.

A backdraught, on the other hand, is a sudden deflagration with smoke (rather than burning gases), where limited ventilation lead to a fire in a compartment producing significant partial combustion products and unburnt pyrolysis products (seen as black and dense smoke). If these smokes accumulate, the admission of air to the compartment, when an opening is made, would lead to this serious and sudden deflagration out of the opening. Different from the flashover, the backdraught could happen in any stage from growth to decay (there might be no fully developed stage due to limited ventilation and fuel supply).

Through the ventilation status and some other signs like colour and pressure of the smoke, visibility of fire flame, the fire status could be accurately estimated by most of experienced firefighters.

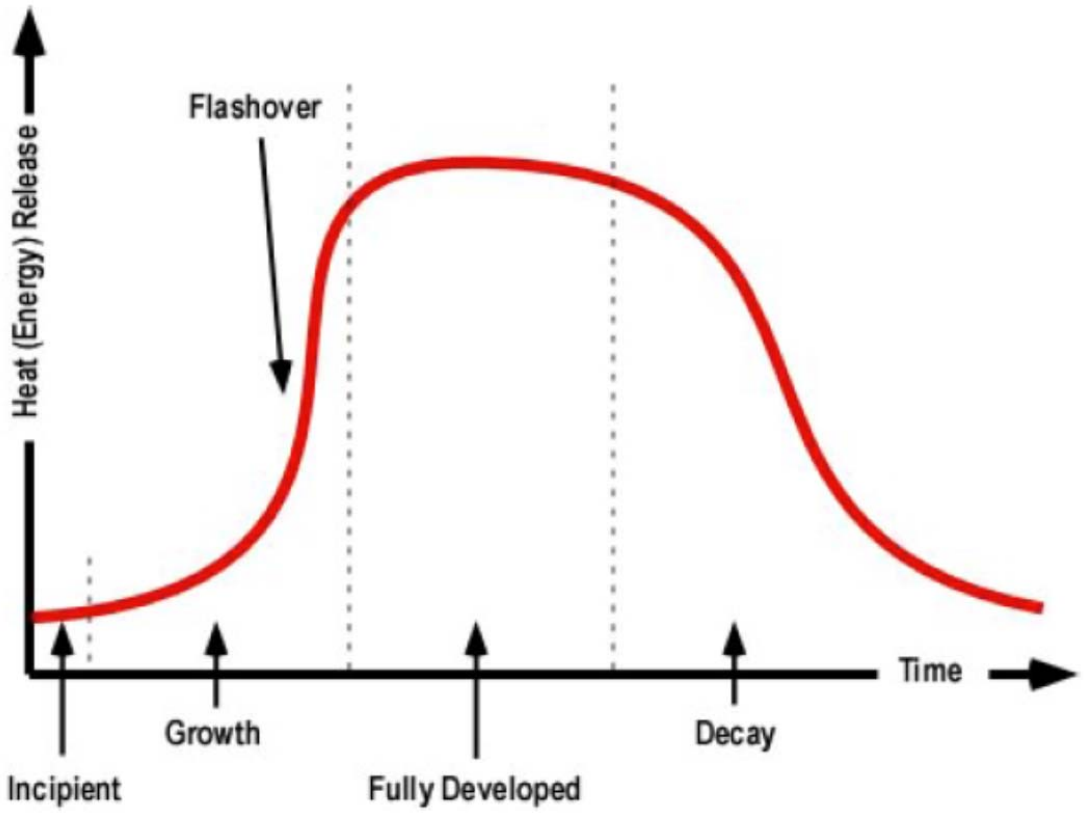


Figure A-2, Fire development in a compartment (Hartin, 2009)

Specific to the high rise situation, a confident fire reading of ICs demands observation from both external (window) and internal (front door testing in the corridor). In addition, due to the compartmentation of the high rise, there could be a developing fire in a compartment with little or even no smoke seen from outside. It is therefore, a perceived, relatively less ambiguous, fire status of ICs would not be possible until the fire crews actually start attacking the fire. In other words, solely external observation of visible flame and heavy smoke might not be a worse situation to ICs, in contrast to the little smoke situation. It is therefore, our initial attempt (pilot questionnaires) of modelling fire status difference with text description came across real difficulty if we wanted to ensure unambiguous and consistent interpretation among different ICs. As the result, we gave up the subtle discrimination of fire status, and concentrated on discriminating the fire spread instead. In terms of the resources demand, whether the fire is contained

within the original compartment or not is critical to ICs. And this is the information that could be identified and relied through external observation.

Accordingly, we defined two cues, horizontal fire spread and vertical fire spread, to reflect the fire smoke profile of the incident. Three issues would be addressed here. Firstly, we define fire and smoke together as they are highly related and always interpreted together from ICs' point of view. Secondly, we differ between horizontal dimension and vertical dimension as they, though both are indications of fire spread, are different to ICs. Due to the compartmentation design, the number of possible fire spread situations that could be assumed in high rise apartment are quite constrained. Fire would either be spread horizontally from the original room to the other rooms of the same flat, or be spread vertically through balcony. Any other serious spread is rarely happened. The vertical spread indicates the broken of the compartmentation, while the horizontal one indicates a larger size of fire but still contained within the same compartment. The vertical spread is therefore a more serious circumstance. Thirdly, to enable the vertical fire spread possible, the choice of fire status is restricted. Only those ventilated fire with visible flame would possibly spread to its upper floors. It is therefore, all the variations of horizontal spread and vertical spread are defined on the basis of ventilated circumstance.

Appendix B: Interview guide

1. Establishing rapport:
Tell the interviewee a little about yourself. Give the interviewee a chance to talk a little bit about himself.

2. Briefly introduce the research, explain the purpose and procedure of the interview, ask for the permission of audio recording.

3. Record the general information regarding the interview, including:
Interview time & location, interviewee's name, age, current role, No. of years in FRS, No. of years as Incident Commander.

4. Present the hypothesized case, and ask for the comments.
Use the interview prompts like:
Does it contain enough information for you to have an initial assessment of the incident?
What additional information would you like to better assess the situation?
What is your interpretation of the scenario suggested in the case?

5. Explain the development of the case hypothesized, and start to check individual cues one by one. Use interview prompts like:
When and how would you observe/get this cue?
Does the cue relevant to your resource demand decision?
What are the features of the cue that would make a difference of your assessment?
What are your interpretations of these different features?
What is the typical feature (most commonly happened)?
What is the terminologies that are used in your normal practice to describe the cue features?

6. Depending on the availability of time, further things could be checked regarding:
presentation of the case
temporal order of cues
inter-dependence of cues

7. Interview should be closed by thanking the interviewee for the participation, and request possible support for the future phase of study.

Appendix C: Hypothetical cases in a decision making task

Index of sub-task	Index of cases	Horizontal fire spread	Floor level	Vertical fire spread	Wind effect	Evacuation status	Search & Rescue
1	27	1	1	0	1	1	0
1	37	1	0	1	0	0	1
1	13	1	0	1	1	0	0
1	8	0	0	0	1	0	0
1	3	1	1	0	0	0	0
1	38	0	1	1	0	0	1
1	60	0	0	1	1	1	1
1	19	1	1	0	0	1	0
1	56	0	0	0	1	1	1
1	62	0	1	1	1	1	1
2	11	1	1	0	1	0	0
2	32	0	0	0	0	0	1
2	1	1	0	0	0	0	0
2	43	1	1	0	1	0	1
2	15	1	1	1	1	0	0
2	61	1	0	1	1	1	1
2	10	0	1	0	1	0	0
2	49	1	0	0	0	1	1
2	53	1	0	1	0	1	1
2	34	0	1	0	0	0	1
3	63	1	1	1	1	1	1
3	3	1	1	0	0	0	0
3	55	1	1	1	0	1	1
3	16	0	0	0	0	1	0
3	36	0	0	1	0	0	1
3	22	0	1	1	0	1	0
3	50	0	1	0	0	1	1
3	57	1	0	0	1	1	1
3	14	0	1	1	1	0	0
3	39	1	1	1	0	0	1
4	5	1	0	1	0	0	0
4	51	1	1	0	0	1	1
4	30	0	1	1	1	1	0
4	44	0	0	1	1	0	1
4	41	1	0	0	1	0	1
4	47	1	1	1	1	0	1
4	28	0	0	1	1	1	0
4	24	0	0	0	1	1	0
4	6	0	1	1	0	0	0
4	36	0	0	1	0	0	1
5	22	0	1	1	0	1	0
5	46	0	1	1	1	0	1

5	27	1	1	0	1	1	0
5	29	1	0	1	1	1	0
5	4	0	0	1	0	0	0
5	17	1	0	0	0	1	0
5	52	0	0	1	0	1	1
5	54	0	1	1	0	1	1
5	21	1	0	1	0	1	0
5	0	0	0	0	0	0	0
6	12	0	0	1	1	0	0
6	48	0	0	0	0	1	1
6	24	0	0	0	1	1	0
6	42	0	1	0	1	0	1
6	26	0	1	0	1	1	0
6	45	1	0	1	1	0	1
6	13	1	0	1	1	0	0
6	9	1	0	0	1	0	0
6	2	0	1	0	0	0	0
6	49	1	0	0	0	1	1
7	41	1	0	0	1	0	1
7	14	0	1	1	1	0	0
7	33	1	0	0	0	0	1
7	20	0	0	1	0	1	0
7	60	0	0	1	1	1	1
7	40	0	0	0	1	0	1
7	50	0	1	0	0	1	1
7	42	0	1	0	1	0	1
7	39	1	1	1	0	0	1
7	23	1	1	1	0	1	0
8	58	0	1	0	1	1	1
8	35	1	1	0	0	0	1
8	31	1	1	1	1	1	0
8	18	0	1	0	0	1	0
8	21	1	0	1	0	1	0
8	7	1	1	1	0	0	0
8	0	0	0	0	0	0	0
8	63	1	1	1	1	1	1
8	25	1	0	0	1	1	0
8	59	1	1	0	1	1	1

Note: Cue values are coded according to Table 5-1. Hypothetical cases are indexed from 1 to 64, indicating 64 different combinations of six cues (2^6=64). They together with another 16 replicated cases are further ordered (Section 5.4.5) to enable 8 sub-tasks, each of which would be completed by one participant.

Appendix D: Data collected from the questionnaires

1. Decision records of the cases

Index of sub-task	Index of cases	No. of fire appliances requested (WMs)	Confidence level (WMs)	No. of fire appliances requested (CMs)	Confidence level (CMs)
1	27	2	8	0	8
1	37	0	8	0	8
1	13	0	8	0	9
1	8	0	8	0	10
1	3	2	8	0	8
1	38	2	8	1	8
1	60	2	8	0	8
1	19	2	8	0	8
1	56	0	8	0	9
1	62	2	8	1	9
2	11	4	9	2	7.5
2	32	2	10	0	8
2	1	0	8	2	8
2	43	4	10	4	9
2	15	4	10	4	8
2	61	2	9	2	8
2	10	2	8	2	7
2	49	2	9	2	8
2	53	2	10	2	8
2	34	4	8	4	8
3	63	4	6	4	8
3	3	4	6	4	8
3	55	4	6	4	8
3	16	2	6	2	8
3	36	2	6	2	8
3	22	4	6	4	8
3	50	4	6	4	8
3	57	2	6	2	8
3	14	4	6	4	8
3	39	4	6	4	8
4	5	2	8	2	7
4	51	4	8	2	7
4	30	2	8	2	7
4	44	4	8	2	7
4	41	1	8	2	7
4	47	4	8	3	7
4	28	2	8	2	7
4	24	0	8	2	7
4	6	0	8	0	7

4	36	2	8	1	7
5	22	0	9	0	10
5	46	2	9	2	8
5	27	2	9	2	8
5	29	2	9	0	9
5	4	0	9	0	9
5	17	0	9	0	9
5	52	2	9	0	9
5	54	2	9	1	9
5	21	0	9	0	9
5	0	0	9	0	9
6	12	2	9	2	7
6	48	2	9	2	7
6	24	2	9	2	7
6	42	4	9	2	7
6	26	4	9	2	7
6	45	2	9	2	7
6	13	2	9	2	7
6	9	2	9	2	7
6	2	4	9	2	7
6	49	2	9	2	7
7	41	0	7	2	9
7	14	2	7	2	9
7	33	0	8	2	9
7	20	0	8	2	9
7	60	2	8	2	8
7	40	0	8	2	9
7	50	3	7	2	9
7	42	3	7	2	9
7	39	4	8	2	9
7	23	4	8	2	9
8	58	8	8	2	10
8	35	8	8	2	10
8	31	10	8	1	10
8	18	8	8	1	10
8	21	6	8	0	10
8	7	10	8	0	10
8	0	0	8	0	10
8	63	10	8	2	10
8	25	0	8	0	10
8	59	8	8	2	10

2. Subjective cue weights assigned by each IC

IC reference	Horizontal fire spread	Floor level	Vertical fire spread	Wind effect	Evacuation status	Search & Rescue
WM1	6	8	6	8	6	8
WM2	5	8	5	7	7	8
WM3	6	8	6	6	9	10
WM4	6	6	4	6	8	8
WM5	5	7	8	8	6	6
WM6	7	9	7	7	0	0
WM7	5	10	9	8	7	6
WM8	5	8	8	7	7	9
CM1	6	8	8	7	6	8
CM2	7	7	7	6	7	9
CM3	8	9	9	8	9	9
CM4	3	4	6	8	7	7
CM5	7	9	7	5	5	8
CM6	5	6	8	8	7	8
CM7	8	10	8	7	7	10
CM8	10	5	9	8	6	7

Appendix E: A sample questionnaire

Your participation counts!

Dear Fire Incident Commanders,

The Loughborough University is conducting a study about your situation assessment behaviour (assessing what is going on) during the fire incident. Your valuable support on this survey will feed us your personal opinion as an incident commander. The knowledge generated in this study would contribute to your department information support system design and training program.

It would be really appreciated if you could take your precious 10 minutes time to complete this questionnaire.

Information in this questionnaire will remain completely confidential and be used solely for the research purposes.

Yours Sincerely,

Emergency Management research team

Loughborough University

Background Information:

(Information in this questionnaire will remain completely confidential and be used solely for the research purposes).

Your Name:

Age:

Gender:

Current Role:

No. of years in the FRS:

No. of years as Incident Commander:

Incident Command Qualification (level 1/2/3):

Instructions:

The pages followed contain 10 scenario cases that simulate the initial stage of the high rise apartment fire. It is the time stage that you are on arrival, conducting your initial DRA on the ground. Each of the case contains the information you would probably gathered at this time stage, including your pre-knowledge of the building, on-scene observation, as well as information from the local.

Followed with each case, there are two fixed questions. **Response to them at your own pace as you would be in normal practice**. So do not return to cases which had already been completed.

To save your time and effort, your pre-knowledge of the incident building is separately presented in the first page followed. It is the same across all the cases. You would take it as your pre-knowledge from the MDT on your way to the incident.

Your pre-knowledge of the scenario cases followed:

It is a high rise apartment near city centre, normal residential with elderly people reported. It was built in 1960s, 11 floors, pre-fabricated concrete construction (figure 1). Floor compartmentation plan is illustrated in figure 2. It is symmetrical layout with one protected staircase and one lift in both sides of the building. The lift is not fire fighting proved, but has a fireman's switch. Mechanical ventilation could be found in corridors and staircases. Eight flats are in each floor, labelled from 'A' to 'H' in figure 2. The AFD of the building is conventional, zoning the building by floors. The water supply of the building area is sufficient. The maintenance inside is good. The PDA for this building is 4 appliances, 1 ALP.



Figure 1 Building profile

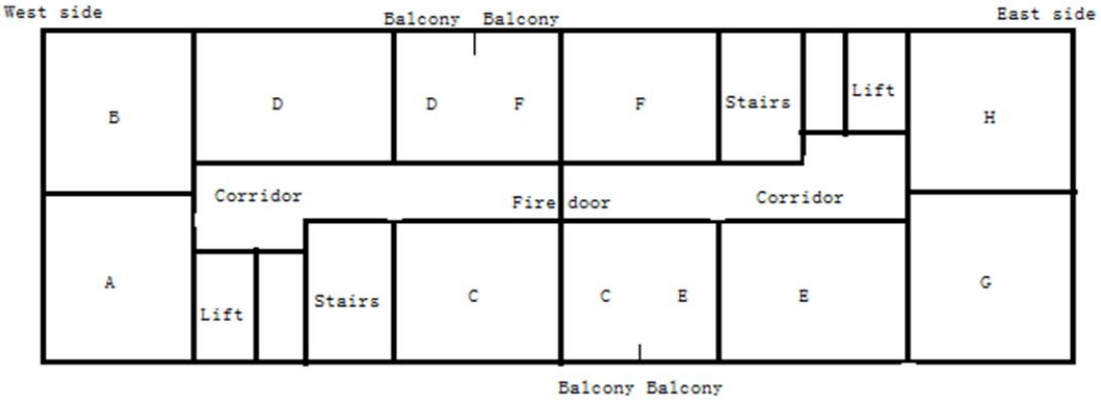


Figure 2. Floor compartmentation layout

Trial No.1

It is the summer daybreak of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony and its two adjacent windows, on the eighth floor (as circled in picture below). According to the floor layout, they belong to the same flat but cross two compartmentations. Fire is visible from the balcony. Burning debris is observed falling out. Wind is strong, blowing heavy smoke upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. And due to the daybreak time as well, the self-evacuation seems quite limited at the moment. The cause of fire remains unknown. Evidence suggests all persons accounted for.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.2

It is the spring afternoon of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony and its two adjacent windows, on the first floor (as circled in picture below). According to the floor layout, they belong to the same flat but cross two compartmentations. Fire is involving the balcony with large flames flickering up to the upper balcony level. Burning debris is observed falling out. Wind is light, and heavy smoke is spread upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground. The cause of fire remains unknown. One family is reported in the flat, one couple, three children.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.3

It is the spring afternoon of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony and its two adjacent windows, on the first floor (as circled in picture below). According to the floor layout, they belong to the same flat but cross two compartmentations. Fire is involving the balcony with large flames flickering up to the upper balcony level. Burning debris is observed falling out. Wind is strong, blowing heavy smoke upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground. The cause of fire remains unknown. Evidence suggests all persons accounted for.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.4

It is the spring afternoon of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony on the first floor (as circled in picture below). Fire is visible from the balcony. Burning debris is observed falling out. Wind is strong, blowing heavy smoke upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground. The cause of fire remains unknown. Evidence suggests all persons accounted for.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.5

It is the spring afternoon of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony and its two adjacent windows, on the eighth floor (as circled in picture below). According to the floor layout, they belong to the same flat but cross two compartmentations. Fire is visible from the balcony. Burning debris is observed falling out. Wind is light, and heavy smoke is spread upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground. The cause of fire remains unknown. Evidence suggests all persons accounted for.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.6

It is the spring afternoon of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony on the eighth floor (as circled in picture below). Fire is involving the balcony with large flames flickering up to the upper balcony level. Burning debris is observed falling out. Wind is light, and heavy smoke is spread upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. However, residents around the fire flat seem well self-evacuated, crowded on the ground. The cause of fire remains unknown. One family is reported in the flat, one couple, three children.

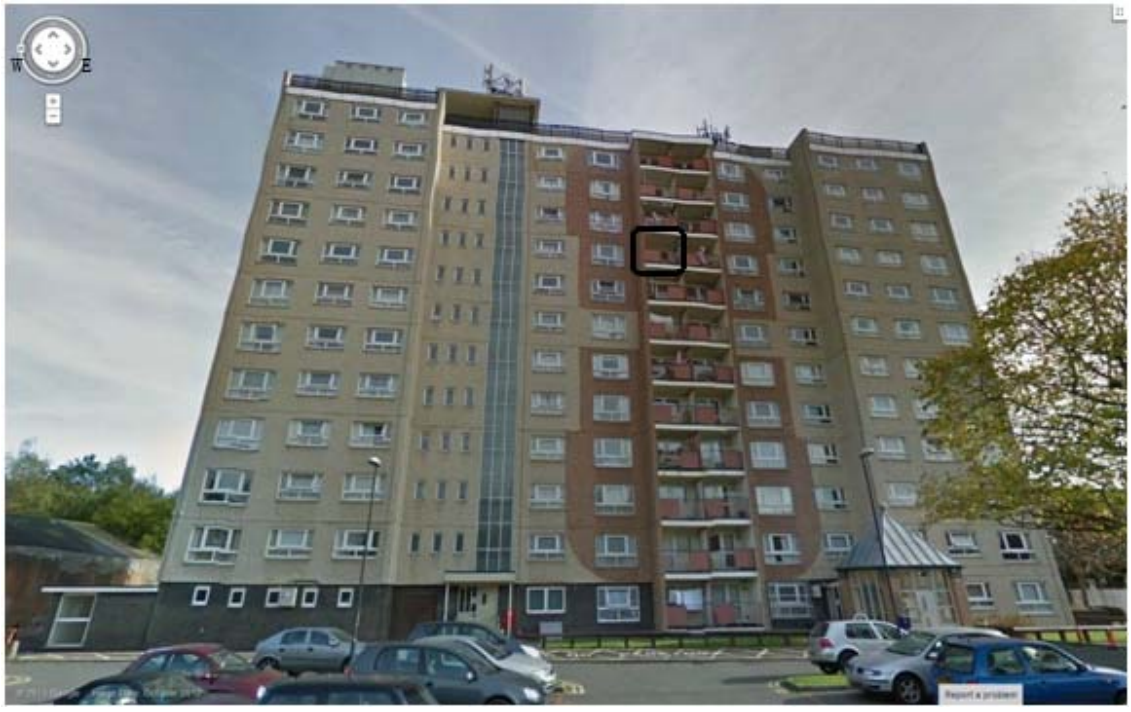
Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.7

It is the summer daybreak of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony on the first floor (as circled in picture below). Fire is involving the balcony with large flames flickering up to the upper balcony level. Burning debris is observed falling out. Wind is strong, blowing heavy smoke upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. And due to the daybreak time as well, the self-evacuation seems quite limited at the moment. The cause of fire remains unknown. One family is reported in the flat, one couple, three children.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.8

It is the summer daybreak of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony and its two adjacent windows, on the eighth floor (as circled in picture below). According to the floor layout, they belong to the same flat but cross two compartmentations. Fire is visible from the balcony. Burning debris is observed falling out. Wind is light, and heavy smoke is spread upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. And due to the daybreak time as well, the self-evacuation seems quite limited at the moment. The cause of fire remains unknown. Evidence suggests all persons accounted for.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.9

It is the summer daybreak of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony on the first floor (as circled in picture below). Fire is visible from the balcony. Burning debris is observed falling out. Wind is strong, blowing heavy smoke upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. And due to the daybreak time as well, the self-evacuation seems quite limited at the moment. The cause of fire remains unknown. One family is reported in the flat, one couple, three children.

Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



Trial No.10

It is the summer daybreak of a weekday. You are deployed as the first OIC to reach the incident. Upon arrival, you see black smoke rolling out of the left balcony on the eighth floor (as circled in picture below). Fire is involving the balcony with large flames flickering up to the upper balcony level. Burning debris is observed falling out. Wind is strong, blowing heavy smoke upwards and to the east. The AFD suggests a number of floors above are affected by the spread smoke. The building has a 'stay put' policy. And due to the daybreak time as well, the self-evacuation seems quite limited at the moment. The cause of fire remains unknown. One family is reported in the flat, one couple, three children.

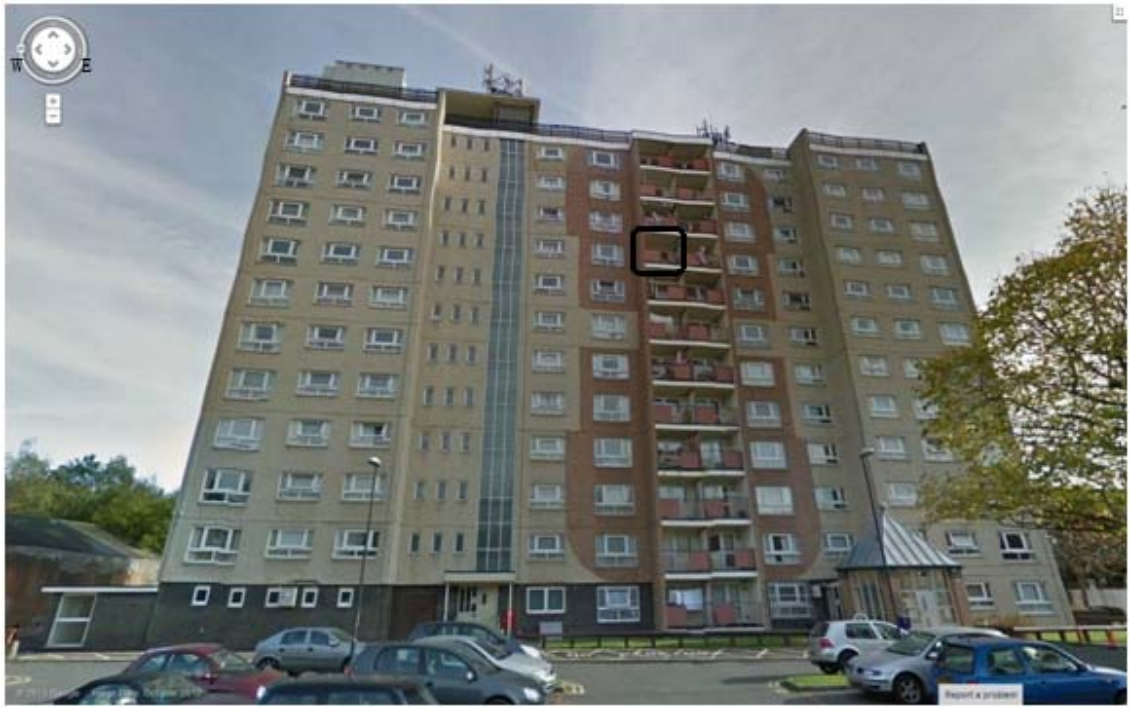
Question 1:
At this moment, you are re-assessing the resources demand. Would you request further resources in addition to the PDA? If Yes, how many additional pumps would you request?

_____

Question 2:
On a scale of 0 to 10, how certain are you feel about your assessment in question 1?

_____

(0, if you have absolutely no confidence about your assessment.
… …
10, if you are absolutely confident about your assessment.)



You are almost finished. One last task!

Ranking task:

Please scale the six issues below, according to the relative importance that you attached to them **during you assessing the above scenario cases**.

(10, if the issue is so important that completely dominate your resource demand assessment;
… …
0, if the issue is absolutely no impact on your resource demand assessment.)

Horizontal fire spread within flat
Fire floor level
Potential of vertical fire spread
Wind condition
Necessity and extent of evacuation
Demand of search & rescue

Additional comments:

Is there any further information you would have liked in the set of scenario cases at this time stage, which would help you to make the assessment? (if yes, please specify)

_____

_____

_____

_____

_____

Thank you so much for your precious time and effort!

If you have any doubt regarding this questionnaire, please feel free to contact with us at: bsys10@lboro.ac.uk

Appendix F: List of participants of interviews and questionnaires

List of interviewees:

Reference	Fire Station	Designation	Experience
Interviewee 1	Central	Crew Manager	16 years in FRS, 4 years as IC
Interviewee 2	Western	Watch Manager	19 years in FRS, 9 years as IC
Interviewee 3	Eastern	Crew Manager	9 years in FRS, 5 years as IC
Interviewee 4	Wigston	Watch Manager	16 years in FRS, 4 years as IC
Interviewee 5	Central	Watch Manager	20 years in FRS, 11 years as IC
Interviewee 6	Central	Watch Manager	19 years in FRS, 14 years as IC
Interviewee 7	Coalville	Watch Manager	24 years in FRS, 15 years as IC

List of participants of pilot study:

Reference	Fire Station	Designation	Age	Gender	No. of years in FRS	No. of years as IC	Incident Command level
Pilot1	Loughborough	WM	33	M	13	9	N/A
Pilot2	Loughborough	CM	46	M	23	14	1
Pilot3	Loughborough	WM	43	M	16	6	2

List of participants of questionnaires:

Reference	Fire Station	Designation	Age	Gender	No. of years in FRS	No. of years as IC	Incident Command level
WM1	Loughborough	WM	51	M	20	7	N/A
WM2	Loughborough	WM	41	M	22	8	2
WM3	Eastern	WM	42	M	15	10	2
WM4	Eastern	WM	38	M	18	9	2
WM5	Eastern	WM	N/A	M	N/A	N/A	N/A
WM6	Southern	WM	37	M	15	10	3
WM7	Eastern	WM	N/A	M	13	6	2
WM8	Southern	WM	49	M	25	6	N/A
CM1	Loughborough	CM	48	M	7	1	N/A
CM2	Loughborough	CM	39	M	14.5	5	1
CM3	Eastern	CM	46	M	14	9	2
CM4	Eastern	CM	28	M	6	2	1
CM5	Eastern	CM	33	M	13	4	1
CM6	Wigston	CM	43	M	19	7	1
CM7	Central	CM	42	M	16	3	1
CM8	Southern	CM	36	M	14	4	1

Appendix G: Matlab scripts used for data analysis

1. Main function

```
function
[ Val,ModMH,ModFK,ModDR,PreMH,PreFK,PreDR,T2,T3,T4,M,MHval3,MHval4,WAva
l3,WAval4 ] = croVal( N )
% N round cross-validation
% Val, 2X6XN matrix of cues validaties of each round
% Mod-MH/FK/DR, matrix of model fit of dichotomized data with validity
rule in training
% Pre-MH/FK/DR, matrix of prediction accuracy of dichotomized data with
validity rule in testing
% T2/T3/T4: matrix containing threshold value of FK and DR on 2/3/4
scaled data with validity rule
% M: NX1 matrix containing k value of MH in each round
% MHval3/MHval4,WAval3/WAval4

sourceData = load('FullDis.mat');
%
%For model comparison on dichotomized data:
Env=sourceData.EnvForCM;%You need to decide here, whether CM or WM.
[row_env,col_env] = size(Env);
Jug=sourceData.JudgmentCM;%Deciding whether CM or WM
%
%For model comparison on 3/4-point ordinal data
Env64=sourceData.EnvForWM;%Typical complete environment
Jug3S=sourceData.JugCM3Cat;%Deciding whether CM or WM
Jug4S=sourceData.JugCM4Cat;%Deciding whether CM or WM
%
%Dealing with replications
JugReplicate=sourceData.JugCMReplicates;%Deciding whether CM or WM
JugNRound=selection(Jug,JugReplicate,N);
rng('shuffle'); %seeds the random number generator based on the current
time
randNround = rand(row_env,N);%53XN for CM data, 64XN for WM data
tempForSort = [1:row_env];
tempForSort = tempForSort';
randNround = [randNround,tempForSort]; %reference matrix for the order
of cross-validation

%Initialization of the matrices
Val = zeros(2,col_env,N);

ModMH = zeros(col_env,4,N);
ModFK = zeros(1,3,N);
ModDR = zeros(1,3,N);

PreMH=zeros(1,3,N);
PreFK=zeros(1,3,N);
PreDR=zeros(1,3,N);

MHval3 = zeros(1,4,N);
MHval4 = zeros(1,5,N);
WAval3 = zeros(1,4,N);
WAval4 = zeros(1,5,N);
```

```
M=zeros(N,1);%k value in matching heuristic
T2=zeros(N,2);%threshold value utilized in FK and DR
T3=zeros(N,2);%threshold value for 3-point ordinal data in WADD3val
T4=zeros(N,3);%threshold value for 4-point ordinal data in WADD4val

for n = 1:N
    orderMatrixTemp = sortrows(randNround,n);
    orderofNround = orderMatrixTemp(:,N+1);
    % the N+1 column of randNround, A matrix of just one column for
indexing
    %
    %Computes the parameters relevant to validity rule
    [Val(:,:,n),T2(n,:),T3(n,:),T4(n,:)] =
validity(Env,JugNRound(:,:,n),orderofNround);

    %For model comparison on dichotomized data with validity rule:
    [ModMH(:,:,n),M(n)] =
modelingMH(Env,JugNRound(:,:,n),orderofNround,Val(:,:,n));
    PreMH(:,:,n) =
predictingMH(Env,JugNRound(:,:,n),orderofNround,Val(:,:,n),M(n));

    ModFK(:,:,n) =
modelingFK(Env,JugNRound(:,:,n),orderofNround,Val(:,:,n),T2(n,1));
    PreFK(:,:,n) =
predictingFK(Env,JugNRound(:,:,n),orderofNround,Val(:,:,n),T2(n,1));

    ModDR(:,:,n) =
modelingDR(Env,JugNRound(:,:,n),orderofNround,T2(n,2));
    PreDR(:,:,n) =
predictingDR(Env,JugNRound(:,:,n),orderofNround,T2(n,2));
    %
    %For model comparison on 3/4 scaled data
    MHval3(:,:,n) = MH3val(Env64,Jug3S,Val(:,:,n),M(n));
    MHval4(:,:,n) = MH4val(Env64,Jug4S,Val(:,:,n),M(n));
    WAval3(:,:,n) = WADD3val(Env64,Jug3S,Val(:,:,n),T3(n,:));
    WAval4(:,:,n) = WADD4val(Env64,Jug4S,Val(:,:,n),T4(n,:));

end
```

2. Function of calculating cue validities and threshold values

```
function [ Val, threshold2, threshold3, threshold4 ] = validity( Env,
Jug, Ord )
%compute the cue validities of this round
%input:Env matrix (64X6 for WM data, 53X6 for CM data), Judgment matrix,
Ordering vector
%output:validities matrix
%Val, 2X6 matrix of cue validity of this round
[row_env,col_env]=size(Env);
count=zeros(2,col_env);
% first row save the No. of cases with correct positive prediction
% second row save the No. of environment cases with positive prediction
Val=zeros(2,col_env);
modelProfileNo = ceil(row_env/2);
for r=1:modelProfileNo
    for c=1:col_env
        count(1,c)=count(1,c)+(Env(Ord(r),c)&&Jug(Ord(r),1));
        count(2,c)=count(2,c)+Env(Ord(r),c);
    end
end
for c=1:col_env
    Val(1,c)=count(1,c)/count(2,c);
    Val(2,c)=c; %second row as index
end
%calculating threshold value
elementFK=zeros(modelProfileNo,1);
elementDR=zeros(modelProfileNo,1);
rate=0;
threshold2=zeros(1,2);%threshold for dichotomized data
for m=1:modelProfileNo
    for n=1:col_env
        elementFK(m)=elementFK(m)+Env(Ord(m),n)*Val(1,n);
        elementDR(m)=elementDR(m)+Env(Ord(m),n);
    end
    if Jug(Ord(m),1)==1
        rate=rate+1;
    end
end
elementFK=sortrows(elementFK,-1);
elementDR=sortrows(elementDR,-1);
threshold2(1)=elementFK(rate);
threshold2(2)=elementDR(rate);

threshold3=zeros(1,2);%threshold for 3-point ordinal data
threshold4=zeros(1,3);%threshold for 4-point ordinal data
if row_env==64%Indicating WM data
    threshold3(1)=elementFK(13);
    threshold3(2)=elementFK(29);
    threshold4(1)=elementFK(8);
    threshold4(2)=elementFK(20);
    threshold4(3)=elementFK(24);
end
if row_env==53%Indicating CM data
    threshold3(1)=elementFK(8);%first column for threshold separating
scale 2 and 1
```

```
    threshold3(2)=elementFK(16);%second column for separating scale 1
and 0
    threshold4(1)=elementFK(3);%first column for separating scale 3 and
2
    threshold4(2)=elementFK(21);%second column for separating scale 2
and 1
    threshold4(3)=elementFK(24);%third column for separating scale 1
and 0
end

end
```

3. Predicting data on derivation sample with Matching Heuristic

```
function [ ModMH, K ] = modelingMH( Env, Jug, Ord, Val )
%Caculating the fit of modeling set with matching heuristic so that
determining the K
%Input:Env matrix, Judgment matrix, Ordering vector, Validity matrix
%Output:Fitting result matrix, value of K
%ModM, 6X4 matrix of modeling fit results (6 possibilities) with
matching heuristic
[row_env,col_env]=size(Env);
modelProfileNo = ceil(row_env/2);
ModMH=zeros(col_env,4);
rateM=zeros(col_env,row_env,3);% score of Matching rule during modeling
request=0;
temp=zeros(1,3);
tempVal=Val';
tempVal=sortrows(tempVal,[-1 2]);%sort rows first by column 1, and then
for any rows with equal value
                        %in column 1, to sort by column 2
for k=1:col_env
    request=0;
    temp=[0 0 0];
    for r=1:modelProfileNo
        if Jug(Ord(r),1)==1
            request=request+1;
        end
        for m=1:k %to identify a critical feature within the k cues,
                  %if identified, then break out from this process.
            if rateM(k,r,1)==1
                break;
            end
            rateM(k,r,1)=Env(Ord(r),tempVal(m,2));
            % column 1 for the score from profile
        end
        if rateM(k,r,1)==1&&Jug(Ord(r),1)==1
            rateM(k,r,2)=1; %column 2 for fitting of scale 1
        end
        if rateM(k,r,1)==0&&Jug(Ord(r),1)==0
            rateM(k,r,3)=1; %column 3 for fitting of scale 0
        end
        temp(1)=temp(1)+rateM(k,r,2)+rateM(k,r,3);
        temp(2)=temp(2)+rateM(k,r,2);
        temp(3)=temp(3)+rateM(k,r,3);
    end
    ModMH(k,1)=temp(1)/modelProfileNo;%percentage of correct prediction
    ModMH(k,2)=temp(2)/request;%accuracy of prediction on scale 1
    ModMH(k,3)=temp(3)/(modelProfileNo-request);%accuracy of prediction
on scale 0
    ModMH(k,4)=k;%remember the k
end
%remember producing K ...
ModMH=sortrows(ModMH,[1,-4]);
%first sorts the row in ascending order according to column 1, then
%descending order according to column 4.
K=ModMH(col_env,4);
end
```

4. Predicting data on cross-validation sample with Matching Heuristic

```
function [ PreMH ] = predictingMH( Env, Jug, Ord, Val, K )
%compute the prediction accuracy through matching heuristic
%PreMH:1X3 prediction result when just looking up K cues
[row_env,col_env]=size(Env);
modelProfileNo = ceil(row_env/2);
predictProfileNo = row_env-modelProfileNo;
PreMH=zeros(1,3);
rateM=zeros(row_env,3);% scoring of Matching rule for each item
request=0;
temp=zeros(1,3);
tempVal=Val';
tempVal=sortrows(tempVal,[-1 2]);
%for k=1:col_env
    for r=(modelProfileNo+1):row_env
        if Jug(Ord(r),1)==1
            request=request+1;
        end
        for m=1:K %to identify a critical feature within the K cues,
                  %if identified, then break out from this process.
            if rateM(r,1)==1
                break;
            end
            rateM(r,1)=Env(Ord(r),tempVal(m,2));
            % column 1 for the score from profile
        end
        if rateM(r,1)==1&&Jug(Ord(r),1)==1
            rateM(r,2)=1; %column 2 for fitting of request No.
        end
        if rateM(r,1)==0&&Jug(Ord(r),1)==0
            rateM(r,3)=1; %column 3 for fitting of non-request No.
        end
        temp(1)=temp(1)+rateM(r,2)+rateM(r,3);
        temp(2)=temp(2)+rateM(r,2);
        temp(3)=temp(3)+rateM(r,3);
    end
    PreMH(1)=temp(1)/predictProfileNo;%percentage of correct prediction
    PreMH(2)=temp(2)/request;%accuracy of prediction on scale 1
    PreMH(3)=temp(3)/(predictProfileNo-request);%accuracy of prediction
on scale 0
%end

end
```

5. Predicting data on derivation sample with Franklin's rule

```
function [ ModFK ] = modelingFK( Env, Jug, Ord, Val, threshold )
%Caculating the fit of modeling set with Franklin rule
[row_env,col_env]=size(Env);
modelProfileNo = ceil(row_env/2);
rateFK=zeros(row_env,3);% score of Franklin during modeling
ModFK=zeros(1,3);
temp=zeros(1,3);
request=0;
for r=1:modelProfileNo
    for c=1:col_env
        rateFK(r,1)=rateFK(r,1)+Env(Ord(r),c)*Val(1,c);%column 1 for
the score
    end
    if Jug(Ord(r),1)==1
        request=request+1;
    end
    if rateFK(r,1)/threshold>=1&&Jug(Ord(r),1)==1
        rateFK(r,2)=1; %column 2 for fitting of request No.
    end
    if rateFK(r,1)/threshold<1&&Jug(Ord(r),1)==0
        rateFK(r,3)=1; %column 3 for fitting of non-request No.
    end
    temp(1)=temp(1)+rateFK(r,2)+rateFK(r,3);
    temp(2)=temp(2)+rateFK(r,2);
    temp(3)=temp(3)+rateFK(r,3);
end
ModFK(1)=temp(1)/modelProfileNo;
ModFK(2)=temp(2)/request;
ModFK(3)=temp(3)/(modelProfileNo-request);


end
```

6. Predicting data on cross-validation sample with Franklin's rule

```
function [ PreFK ] = predictingFK( Env, Jug, Ord, Val, threshold )
%Caculating the prediction accuracy on cross-validation set with
Franklin rule
[row_env,col_env]=size(Env);
modelProfileNo = ceil(row_env/2);
predictProfileNo = row_env-modelProfileNo;
rateFK=zeros(row_env,3);% score of Franklin during modeling
PreFK=zeros(1,3);
temp=zeros(1,3);
request=0;
for r=(modelProfileNo+1):row_env
    for c=1:col_env
        rateFK(r,1)=rateFK(r,1)+Env(Ord(r),c)*Val(1,c);%column 1 for
the score
    end
    if Jug(Ord(r),1)==1
        request=request+1;
    end
    if rateFK(r,1)/threshold>=1&&Jug(Ord(r),1)==1
        rateFK(r,2)=1; %column 2 for fitting of request No.
    end
    if rateFK(r,1)/threshold<1&&Jug(Ord(r),1)==0
        rateFK(r,3)=1; %column 3 for fitting of non-request No.
    end
    temp(1)=temp(1)+rateFK(r,2)+rateFK(r,3);
    temp(2)=temp(2)+rateFK(r,2);
    temp(3)=temp(3)+rateFK(r,3);
end
PreFK(1)=temp(1)/predictProfileNo;
PreFK(2)=temp(2)/request;
PreFK(3)=temp(3)/(predictProfileNo-request);


end
```

7. Predicting data on derivation sample with Dawes' rule

```
function [ ModDR rateDR ] = modelingDR( Env, Jug, Ord, threshold )
%Modeling with Dawes' rule
[row_env,col_env]=size(Env);
modelProfileNo = ceil(row_env/2);
rateDR=zeros(row_env,3);% score of Dawes during modeling
ModDR=zeros(1,3);
temp=zeros(1,3);
request=0;
for r=1:modelProfileNo
    for c=1:col_env
        rateDR(r,1)=rateDR(r,1)+Env(Ord(r),c);%column 1 for the score,
the only difference with FK
    end
    if Jug(Ord(r),1)==1
        request=request+1;
    end
    if rateDR(r,1)/threshold>=1&&Jug(Ord(r),1)==1
        rateDR(r,2)=1; %column 2 for fitting of request No.
    end
    if rateDR(r,1)/threshold<1&&Jug(Ord(r),1)==0
        rateDR(r,3)=1; %column 3 for fitting of non-request No.
    end
    temp(1)=temp(1)+rateDR(r,2)+rateDR(r,3);
    temp(2)=temp(2)+rateDR(r,2);
    temp(3)=temp(3)+rateDR(r,3);
end
ModDR(1)=temp(1)/modelProfileNo;
ModDR(2)=temp(2)/request;
ModDR(3)=temp(3)/(modelProfileNo-request);


end
```

8. Predicting data on cross-validation sample with Dawes' rule

```
function [ PreDR ] = predictingDR( Env, Jug, Ord, threshold )
%Caculating the prediction accuracy on cross-validation set with Dawes'
rule
[row_env,col_env]=size(Env);
modelProfileNo = ceil(row_env/2);
predictProfileNo = row_env-modelProfileNo;
rateDR=zeros(row_env,3);% score of Dawes during modeling
PreDR=zeros(1,3);
temp=zeros(1,3);
request=0;
for r=(modelProfileNo+1):row_env
    for c=1:col_env
        rateDR(r,1)=rateDR(r,1)+Env(Ord(r),c);%column 1 for the score,
the only difference from FK
    end
    if Jug(Ord(r),1)==1
        request=request+1;
    end
    if rateDR(r,1)/threshold>=1&&Jug(Ord(r),1)==1
        rateDR(r,2)=1; %column 2 for fitting of request No.
    end
    if rateDR(r,1)/threshold<1&&Jug(Ord(r),1)==0
        rateDR(r,3)=1; %column 3 for fitting of non-request No.
    end
    temp(1)=temp(1)+rateDR(r,2)+rateDR(r,3);
    temp(2)=temp(2)+rateDR(r,2);
    temp(3)=temp(3)+rateDR(r,3);
end
PreDR(1)=temp(1)/predictProfileNo;
PreDR(2)=temp(2)/request;
PreDR(3)=temp(3)/(predictProfileNo-request);


end
```

9. Predicting 3-point ordinal data with MH-3R

```
function [ MHval3 ] = MH3val( Env, Jug, Val, K )
%computing the prediction accuracy of MH-3R model with validity rule
%Env is the standard 64X6 matrix
%Jug would be 30X3 for CM data and  10X3 matrix for WM data
%Val here contains the 2X6 matrix produced by validity function
%K is the k value produced from the modelingMH function

[row_Jug,col_Jug]=size(Jug);
MHval3=zeros(1,4);
rateMH3=zeros(row_Jug,4);
request=zeros(1,2);
temp=zeros(1,4);
tempVal=Val';
tempVal=sortrows(tempVal,[-1 2]);%sort rows first by column 1, and then
for any rows with equal value
                        %in column 1, to sort by column 2
if row_Jug==30%one replication in this data set
    rng('shuffle');
    randNo=rand(2);
    count=row_Jug-1;
else
    count=row_Jug;%then row_Jug=10
end

    for r=1:row_Jug
        if row_Jug==30
            if (randNo(1)>=randNo(2))&&(r==27)%choose record 20;
                continue;
            end
            if (randNo(1)<randNo(2))&&(r==20)%choose record 27;
                continue;
            end
        end
        if Jug(r,1)==1
            request(1)=request(1)+1;%request at scale 1
        end
        if Jug(r,1)==2
            request(2)=request(2)+1;%request at scale 2
        end
        for m=1:K %the search of critical value, and consequent
confirmation process
            rateMH3(r,1)=Env(Jug(r,2),tempVal(m,2));
            if rateMH3(r,1)==1%One additional cue confirmation process

rateMH3(r,1)=rateMH3(r,1)+Env(Jug(r,2),tempVal((m+1),2));
                break;
            end
        end
        if rateMH3(r,1)==2&&Jug(r,1)==2
            rateMH3(r,2)=1; %column 2 for fitting of scale 2
        end
        if rateMH3(r,1)==1&&Jug(r,1)==1
            rateMH3(r,3)=1; %column 3 for fitting of scale 1
        end
```

```
        if rateMH3(r,1)==0&&Jug(r,1)==0
            rateMH3(r,4)=1; %column 4 for fitting of scale 0
        end
        temp(1)=temp(1)+rateMH3(r,2)+rateMH3(r,3)+rateMH3(r,4);
        temp(2)=temp(2)+rateMH3(r,2);
        temp(3)=temp(3)+rateMH3(r,3);
        temp(4)=temp(4)+rateMH3(r,4);
    end
    MHval3(1)=temp(1)/count;%percentage of correct prediction
    MHval3(2)=temp(2)/request(2);%percentage prediction at scale 2
    MHval3(3)=temp(3)/request(1);%percentage prediction at scale 1
    MHval3(4)=temp(4)/(count-request(2)-request(1));% ... scale 0

end
```

10. Predicting 4-point ordinal data with MH-4R

```
function [ PreMH4S ] = MH4val( Env, Jug, Val, K )
%computing the prediction accuracy of MH-4R model with validity rule
%Env is the standard 64X6 matrix
%Jug would be 30X3 WM data and 10X3 matrix for CM data
%Val here contains the 2X6 matrix produced by validity function
%K is the k value produced from the modelingMH function

[row_Jug,col_Jug]=size(Jug);
PreMH4S=zeros(1,5);
rateM=zeros(row_Jug,5);
request=zeros(1,3);
temp=zeros(1,5);
tempVal=Val';
tempVal=sortrows(tempVal,[-1 2]);%sort rows first by column 1, and then
for any rows with equal value
                        %in column 1, to sort by column 2
if row_Jug==30%one replication in this data set
    rng('shuffle');
    randNo=rand(2);
    count=row_Jug-1;
else
    count=row_Jug;%then row_Jug=10
end



    for r=1:row_Jug
        if row_Jug==30
            if (randNo(1)>=randNo(2))&&(r==11)%choose record 5;
                continue;
            end
            if (randNo(1)<randNo(2))&&(r==5)%choose record 11;
                continue;
            end
        end
        if Jug(r,1)==1
            request(1)=request(1)+1;%request at scale 1
        end
        if Jug(r,1)==2
            request(2)=request(2)+1;%request at scale 2
        end
        if Jug(r,1)==3
            request(3)=request(3)+1;%request at scale 3
        end
        for m=1:K %the search of critical value, and consequent
confirmation process
            rateM(r,1)=Env(Jug(r,2),tempVal(m,2));
            if rateM(r,1)==1%Two additional cues confirmation process
                rateM(r,1)=rateM(r,1)+Env(Jug(r,2),tempVal((m+1),2));
                if rateM(r,1)==2

rateM(r,1)=rateM(r,1)+Env(Jug(r,2),tempVal((m+2),2));
                    break;
                end
                break;
```

```
            end
        end
        if rateM(r,1)==3&&Jug(r,1)==3
            rateM(r,2)=1; %column 2 for fitting of scale 3
        end
        if rateM(r,1)==2&&Jug(r,1)==2
            rateM(r,3)=1; %column 3 for fitting of scale 2
        end
        if rateM(r,1)==1&&Jug(r,1)==1
            rateM(r,4)=1; %column 4 for fitting of scale 1
        end
        if rateM(r,1)==0&&Jug(r,1)==0
            rateM(r,5)=1; %column 5 for fitting of scale 0
        end
        temp(1)=temp(1)+rateM(r,2)+rateM(r,3)+rateM(r,4)+rateM(r,5);
        temp(2)=temp(2)+rateM(r,2);
        temp(3)=temp(3)+rateM(r,3);
        temp(4)=temp(4)+rateM(r,4);
        temp(5)=temp(5)+rateM(r,5);
    end
    PreMH4S(1)=temp(1)/count;%percentage of correct prediction
    PreMH4S(2)=temp(2)/request(3);%accurate prediction at scale 3
    PreMH4S(3)=temp(3)/request(2);%accurate prediction at scale 2
    PreMH4S(4)=temp(4)/request(1);%accurate prediction at scale 1
    PreMH4S(5)=temp(5)/(count-request(1)-request(2)-
request(3));%...scale 0

end
```

11. Predicting 3-point ordinal data with WADD-3R

```
function [ WAval3 ] = WADD3val( Env, Jug, Val, threshold3 )
%computing the prediction accuracy of WADD-3R model with validity rule
%Env is the standard 64X6 matrix
%Jug would be 30X3 for CM data and 10X3 matrix for WM data
%Val here contains the 2X6 matrix produced by validity function
%threshold3 is a 1X2 vector containing 2 cut-off values of the data

[row_env,col_env]=size(Env);
[row_Jug,col_Jug]=size(Jug);
WAval3=zeros(1,4);
rateWA3=zeros(row_Jug,4);
request=zeros(1,2);
temp=zeros(1,4);
if row_Jug==30%one replication in this data set
    rng('shuffle');
    randNo=rand(2);
    count=row_Jug-1;
    %indicating CM data
else
    count=row_Jug;%then row_Jug=10
    %indicating WM data
end

    for r=1:row_Jug
        if row_Jug==30
            if (randNo(1)>=randNo(2))&&(r==27)%choose record 20;
                continue;
            end
            if (randNo(1)<randNo(2))&&(r==20)%choose record 27;
                continue;
            end
        end
        if Jug(r,1)==1
            request(1)=request(1)+1;%request at scale 1
        end
        if Jug(r,1)==2
            request(2)=request(2)+1;%request at scale 2
        end
        for c=1:col_env
            rateWA3(r,1)=rateWA3(r,1)+Env(Jug(r,2),c)*Val(1,c);%column
1 for the score
        end
        if rateWA3(r,1)/threshold3(1)>=1&&Jug(r,1)==2 %remember to
define threshold value
            rateWA3(r,2)=1; %column 2 for fitting of scale 2
        end
        if
rateWA3(r,1)/threshold3(1)<1&&rateWA3(r,1)/threshold3(2)>=1&&Jug(r,1)==
1
            rateWA3(r,3)=1; %column 3 for fitting of scale 1
        end
        if rateWA3(r,1)/threshold3(2)<1&&Jug(r,1)==0
            rateWA3(r,4)=1; %column 4 for fitting of scale 0
        end
```

```
        temp(1)=temp(1)+rateWA3(r,2)+rateWA3(r,3)+rateWA3(r,4);
        temp(2)=temp(2)+rateWA3(r,2);
        temp(3)=temp(3)+rateWA3(r,3);
        temp(4)=temp(4)+rateWA3(r,4);
    end
WAval3(1)=temp(1)/count;%percentage of correct prediction
WAval3(2)=temp(2)/request(2);%accurate prediction at scale 2
WAval3(3)=temp(3)/request(1);%accurate prediction at scale 1
WAval3(4)=temp(4)/(count-request(1)-request(2));%...scale 0

end
```

12. Predicting 4-point ordinal data with WADD-4R

```
function [ WAval4 ] = WADD4val( Env, Jug, Val, threshold4 )
%computing the prediction accuracy of WADD-4R model with validity rule
%Env is the standard 64X6 matrix
%Jug would be 30X3 for WM data and 10X3 matrix for CM data
%Val here contains the 2X6 matrix produced by validity function
%threshold4 is a 1X3 vector containing 3 cut-off values of the data

[row_env,col_env]=size(Env);
[row_Jug,col_Jug]=size(Jug);
WAval4=zeros(1,5);
rateWA4=zeros(row_Jug,5);
request=zeros(1,3);
temp=zeros(1,5);

if row_Jug==30%one replication in this data set
    rng('shuffle');
    randNo=rand(2);
    count=row_Jug-1;
    %indicating WM data
else
    count=row_Jug;%then row_Jug=10
    %indicating CM data
end

    for r=1:row_Jug
        if row_Jug==30
            if (randNo(1)>=randNo(2))&&(r==11)%choose record 5;
                continue;
            end
            if (randNo(1)<randNo(2))&&(r==5)%choose record 11;
                continue;
            end
        end
        if Jug(r,1)==1
            request(1)=request(1)+1;%request at scale 1
        end
        if Jug(r,1)==2
            request(2)=request(2)+1;%request at scale 2
        end
        if Jug(r,1)==3
            request(3)=request(3)+1;%request at scale 3
        end
        for c=1:col_env
            rateWA4(r,1)=rateWA4(r,1)+Env(Jug(r,2),c)*Val(1,c);%column
1 for the score
        end
        if rateWA4(r,1)/threshold4(1)>=1&&Jug(r,1)==3
            rateWA4(r,2)=1; %column 2 for fitting of scale 3
        end
        if
rateWA4(r,1)/threshold4(1)<1&&rateWA4(r,1)/threshold4(2)>=1&&Jug(r,1)==
2
            rateWA4(r,3)=1; %column 3 for fitting of scale 2
        end
```

```
        if
rateWA4(r,1)/threshold4(2)<1&&rateWA4(r,1)/threshold4(3)>=1&&Jug(r,1)==
1
            rateWA4(r,4)=1; %column 4 for fitting of scale 1
        end
        if rateWA4(r,1)/threshold4(3)<1&&Jug(r,1)==0
            rateWA4(r,5)=1; %column 5 for fitting of scale 0
        end


temp(1)=temp(1)+rateWA4(r,2)+rateWA4(r,3)+rateWA4(r,4)+rateWA4(r,5);
        temp(2)=temp(2)+rateWA4(r,2);
        temp(3)=temp(3)+rateWA4(r,3);
        temp(4)=temp(4)+rateWA4(r,4);
        temp(5)=temp(5)+rateWA4(r,5);
    end
WAval4(1)=temp(1)/count;%percentage of correct prediction
WAval4(2)=temp(2)/request(3);%accurate prediction at scale 3
WAval4(3)=temp(3)/request(2);%accurate prediction at scale 2
WAval4(4)=temp(4)/request(1);%accurate prediction at scale 1
WAval4(5)=temp(5)/(count-request(1)-request(2)-request(3));%...scale 0

end
```