Specimen Theoriae Novae de Mensura Complexūs

Mesures Objectives de la Complexité
Objective Measures of Complexity

Thèse

Sylvain Pronovost

Doctorat en Psychologie
Philosophiæ doctor (Ph.D.)

Québec, Canada

Specimen Theoriae Novae de Mensura Complexūs

Mesures Objectives de la Complexité
Objective Measures of Complexity

Thèse

Sylvain Pronovost

Sous la direction de:

Sébastien Tremblay, directeur de recherche

Résumé

Mesures Objectives de la Complexité pour la Prise de Décision Dynamique

La gestion efficace de systèmes sociotechniques complexes dépend d'une compréhension des interrelations dynamiques entre les composantes de ces systèmes, de leur évolution à travers le temps, ainsi que du degré d'incertitude auquel les décideurs sont exposés. Quelles sont les caractéristiques de la prise de décision complexe qui ont un impact sur la performance humaine dans l'environnement moderne du travail, constamment en fluctuation et sous la pression du temps, exerçant de lourdes demandes sur la cognition ? La prise de décision complexe est un concept issu de la macrocognition, impliquant des processus et des fonctions de bas et haut niveaux de description tels que la métacognition, soit pour un individu de penser à propos de son propre processus de pensées. Dans le cas particulier de la prise de décision complexe, ce phénomène est nommé la pensée systémique. L'étude de la prise de décision complexe en-dehors de l'environnement traditionnel du laboratoire, permettant un haut niveau de contrôle mais un faible degré de réalisme, est malheureusement difficile et presque impossible. Une méthode de recherche plus appropriée pour la macrocognition est l'expérimentation basée sur la simulation, à l'aide de micromondes numérisés sous la forme de jeux sérieux. Ce paradigme de recherche est nommé la prise de décision dynamique (PDD), en ce qu'il tient compte des caractéristiques de problèmes de prise de décision complexe telles que des séquences complexes de décisions et de changements d'états d'un problème interdépendants, qui peuvent changer de façon spontanée ou comme conséquence de décisions préalables, et pour lesquels la connaissance et la compréhension du décideur peut n'être que partielle ou incertaine.

Malgré la quantité de recherche concernant la PDD à propos des difficultés encourues pour la performance humaine face à des problèmes de prise de décision complexe, l'acquisition de connaissances à propos de systèmes complexes, et à savoir si le transfert de l'apprentissage est possible, il n'existe pas de mesure quantitative de ce en quoi un problème de décision est considéré comme étant complexe. La littérature scientifique mentionne des éléments qualitatifs concernant les systèmes complexes (tels que des interrelations dynamiques, une évolution non-linéaire d'un système à travers le temps, et l'incertitude à propos des états d'un système et des issues des décisions), mais des mesures quantitatives et objectives exprimant la complexité de problèmes de décision n'ont pas été développées. Cette dissertation doctorale présente les

concepts, la méthodologie et les résultats impliqués dans un projet de recherche visant à développer des mesures objectives de la complexité basées sur les caractéristiques de problèmes de prise de décision dynamique pouvant expliquer et prédire la performance humaine. En s'inspirant de divers domaines d'application de la théorie de la complexité tels que la complexité computationnelle, la complexité systémique, et l'informatique cognitive, un modèle formel des paramètre de la complexité pour des tâches de prise de décision dynamique a été élaboré. Un ensemble de dix mesures objectives de la complexité a été développé, consistant en des mesures de la complexité structurelle, des mesures de la complexité informationnelle, la complexité de la charge cognitive, et des mesures de la difficulté d'un problème, de la non-linéarité des relations, de l'incertitude concernant l'information et les décisions, ainsi qu'une mesure de l'instabilité d'un système dynamique sous des conditions d'inertie.

Une analyse des résultats expérimentaux colligés à partir de cinq scénarios de PDD révèle qu'un nombre restreint de candidats parmi des modèles de régression linéaires multiple permet d'expliquer et de prédire les résultats de performance humaine, mais au prix de certaines violations des postulats de l'approche classique de la régression linéaire. De plus, ces mesures objectives de la complexité présentent un degré élevé de multicolinéarité, causée d'une part par l'inclusion de caractéristiques redondantes dans les calculs, et d'autre part par une colinéarité accidentelle imputable à la conception des scénarios de PDD. En tenant compte de ces deux considérations ainsi que de la variance élevée observée dans les processus macrocognitifs impliqués dans la prise de décision complexe, ces modèles présentent des valeurs élevées pour le terme d'erreur exprimant l'écart entre les observations et les prédictions des modèles.

Une analyse additionnelle explore l'utilisation de méthodes alternatives de modélisation par régression afin de mieux comprendre la relation entre les paramètres de la complexité et les données portant sur performance humaine. Nous avons d'abord opté pour une approche de régression robuste afin d'augmenter l'efficience de l'analyse de régression en utilisant une méthode réduisant la sensibilité des modèles de régression aux observations influentes. Une seconde analyse élimine la source de variance imputable aux différences individuelles en focalisant exclusivement sur les effets imputables aux conditions expérimentales. Une dernière analyse utilise des modèles non-linéaires et non-paramétriques afin de pallier les postulats de la modélisation par régression, à l'aide de méthodes d'apprentissage automatique (*machine learning*). Les résultats suggèrent que l'approche de régression robuste produit des termes d'erreur substantiellement plus faibles, en combinaison avec des valeurs élevées pour les mesures de variance expliquée dans les données de la performance humaine. Bien que les méthodes non-linéaires et non-paramétriques produisent des modèles marginalement plus efficients en comparaison aux modèles de régression linéaire, la combinaison de ces modèles issus du domaine de l'apprentissage automatique avec les données restreintes aux effets imputables aux conditions expérimentales produit les meilleurs résultats relativement à l'ensemble de l'effort de modélisation et d'analyse de régression.

Une dernière section présente un programme de recherche conçu pour explorer l'espace des paramètres pour les mesures objectives de la complexité avec plus d'ampleur et de profondeur, afin d'appréhender les combinaisons des caractéristiques des problèmes de prise de décision complexe qui sont des facteurs déterminants de la performance humaine. Les discussions concernant l'approche expérimentale pour la PDD, les résultats de l'expérimentation relativement aux modèles de régression, ainsi qu'à propos de l'investigation de méthodes alternatives visant à réduire la composante de variance menant à la disparité entre les observations et les prédictions des modèles suggèrent toutes que le développement de mesures objectives de la complexité pour la performance humaine dans des scénarios de prise de décision dynamique est une approche viable à l'approfondissement de nos connaissances concernant la compréhension et le contrôle exercés par un être humain face à des problèmes de décision complexe.

Abstract

Objective Measures of Complexity for Dynamic Decision-Making

Managing complex sociotechnical systems depends on an understanding of the dynamic interrelations of such systems' components, their evolution over time, and the degree of uncertainty to which decision makers are exposed. What features of complex decision-making impact human performance in the cognitively demanding, ever-changing and time pressured modern workplaces? Complex decision-making is a macrocognitive construct, involving low to high cognitive processes and functions, such as metacognition, or thinking about one's own thought processes. In the particular case of complex decision-making, this is called systems thinking. The study of complex decision-making outside of the controlled, albeit lacking in realism, traditional laboratory environment is difficult if not impossible. Macrocognition is best studied through simulation-based experimentation, using computerized microworlds in the form of serious games. That research paradigm is called dynamic decision-making (DDM), as it takes into account the features of complex decision problems, such as complex sequences of interdependent decisions and changes in problem states, which may change spontaneously or as a consequence of earlier decisions, and for which the knowledge and understanding may be only partial or uncertain.

For all the research in DDM concerning the pitfalls of human performance in complex decision problems, the acquisition of knowledge about complex systems, and whether a learning transfer is possible, there is no quantitative measure of what constitutes a complex decision problem. The research literature mentions the qualities of complex systems (a system's dynamical relationships, the nonlinear evolution of the system over time, and the uncertainty about the system states and decision outcomes), but objective quantitative measures to express the complexity of decision problems have not been developed. This dissertation presents the concepts, methodology, and results involved in a research endeavor to develop objective measures of complexity based on characteristics of dynamic decision-making problems which can explain and predict human performance. Drawing on the diverse fields of application of complexity theory such as computational complexity, systemic complexity, and cognitive informatics, a formal model of the parameters of complexity for dynamic decision-making tasks has been elaborated. A set of ten objective measures of complexity were developed, ranging

from structural complexity measures, measures of information complexity, the cognitive weight complexity, and measures of problem difficulty, nonlinearity among relationships, information and decision uncertainty, as well as a measure of the dynamical system's instability under inertial conditions.

An analysis of the experimental results gathered using five DDM scenarios revealed that a small set of candidate models of multiple linear regression could explain and predict human performance scores, but at the cost of some violations of the assumptions of classical linear regression. Additionally, the objective measures of complexity exhibited a high level of multicollinearity, some of which were caused by redundant feature computation while others were accidentally collinear due to the design of the DDM scenarios. Based on the aforementioned constraints, and due to the high variance observed in the macrocognitive processes of complex decision-making, the models exhibited high values of error in the discrepancy between the observations and the model predictions.

Another exploratory analysis focused on the use of alternative means of regression modeling to better understand the relationship between the parameters of complexity and the human performance data. We first opted for a robust regression analysis to increase the efficiency of the regression models, using a method to reduce the sensitivity of candidate regression models to influential observations. A second analysis eliminated the within-treatment source of variance in order to focus exclusively on between-treatment effects. A final analysis used nonlinear and non-parametric models to relax the regression modeling assumptions, using machine learning methods. It was found that the robust regression approach produced substantially lower error values, combined with high measures of the variance explained for the human performance data. While the machine learning methods produced marginally more efficient models of regression for the same candidate models of objective measures of complexity, the combination of the nonlinear and non-parametric methods with the restricted between-treatment dataset yielded the best results of all of the modeling and analyses endeavors.

A final section presents a research program designed to explore the parameter space of objective measures of complexity in more breadth and depth, so as to weight which combinations of the characteristics of complex decision problems are determinant factors on human performance. The discussions about the experimental approach to DDM, the experimental results relative to the regression models, and the investigation of further means to reduce the variance component underlying the discrepancy between the observations and the model predictions all suggest that establishing objective measures of complexity for human performance in dynamic decision-making scenarios is a viable approach to furthering our understanding of a decision maker's comprehension and control of complex decision problems.

Contents

Résumé iii

Abstract vi

Contents viii

List of Tables xii

List of Figures xv

List of Acronyms xx

Acknowledgements xxiii

Prologue xxiv

Introduction 1
From simple decision-making to complex decision-making . . . 2
The psychology of decision-making . . . 3
Dynamic decision-making . . . 7
Objective of this dissertation: tackling the problem of quantifying complexity for dynamic decision-making problems . . . 13
Modeling complex systems . . . 15
Models of complexity from cognitive informatics . . . 20
Task complexity and the complexity of dynamic decision-making problems . 22

1 Experimental Approach and Methodology 25
1.1 Introduction . . . 25
1.2 Microcognition and macrocognition . . . 25
1.3 Simulations, microworlds, and serious games . . . 31
1.3.1 Simulations . . . 32
1.3.2 Microworlds . . . 33
1.3.3 Serious games . . . 34
1.3.4 The CODEM simulation environment . . . 35
1.3.5 A complete example: the Arctic 3 DDM scenario in CODEM . . . 39
1.4 Objective parameters of complexity . . . 53
1.4.1 Structural complexity . . . 55
1.4.2 Information complexity . . . 56
1.4.3 Cognitive weight . . . 56

1.4.4 Difficulty . . . 57
1.4.5 Nonlinearity . . . 59
1.4.6 Uncertainty . . . 59
1.4.7 Instability . . . 60
1.5 Discussion . . . 61

2 Modeling and Validation of Objective Measures of Complexity for Dynamic Decision-Making 63
2.1 Introduction . . . 63
2.2 Method . . . 64
2.2.1 Participants . . . 64
2.2.2 Task: dynamic decision-making scenarios . . . 64
2.2.3 Experimental design and procedure . . . 68
2.2.4 Apparatus . . . 69
2.2.5 Response metric: measures of performance in dynamic decision-making 70
2.2.6 Control measures: objective parameters of complexity . . . 71
2.3 Results . . . 73
2.3.1 Data analysis and variable selection . . . 73
2.3.2 Multiple linear regression model selection for the dynamic decision-making scenarios . . . 80
2.3.3 Regression model validation . . . 91
2.4 Discussion . . . 96
2.4.1 Relationships between the objective parameters of complexity . . . 96
2.4.2 Relationships between DDM performance and the objective measures of complexity . . . 98
2.4.3 Limitations and way ahead . . . 100

3 Nonlinear, Non-Parametric, and Robust Methods to Model Objective Measures of Complexity for Dynamic Decision-Making 102
3.1 Introduction . . . 102
3.1.1 Coping with the bias-variance trade-off: differences for micro- and macrocognitive phenomena . . . 103
3.1.2 Nonlinear and non-parametric methods for classification and regression in computational learning theory . . . 104
3.2 Method . . . 105
3.2.1 Robust regression models . . . 105
3.2.2 Regression analysis using group means and group medians . . . 108
3.2.3 Nonlinear and non-parametric methods: the machine learning approach . . . 109
3.3 Results . . . 118
3.3.1 Robust regression models . . . 118
3.3.2 Regression analysis using group means and group medians . . . 121
3.3.3 Nonlinear and non-parametric regression models . . . 123
3.4 Discussion . . . 128
3.4.1 Comparing the linear regression models with robust regression, regression using group means and medians, and nonlinear methods . . . 129
3.4.2 To underfit or to overfit? Compromising between bias and variance in modeling objective measures of complexity . . . 131

3.4.3 Limitations and way ahead 132

Conclusion **135**
Summary of the objectives 135
Summary of the results 136
Research hypotheses and answers 137
Tackling dynamic decision-making with simulation-based experimentation 138
Objective measures of complexity for dynamic decision-making problems 140
Working around the assumptions of linear regression analysis while dealing with the bias-variance trade-off 144
Contributions 146
Theoretical contribution 146
Methodological contribution 147
Practical contribution 149
Limitations and way ahead 151
A research program for objective measures of complexity 157
Candidate parameters of complexity and models of human performance in dynamic decision-making 157
Feature space and dynamic decision-making models 159

Epilogue **161**

References **163**

A Dynamic Decision-Making Scenarios **187**
A.1 Arctic 1 187
A.1.1 Variables 187
A.1.2 Relations between variables 188
A.1.3 Interventions and influences from interventions 188
A.1.4 Miscellaneous 188
A.2 Arctic 2 189
A.2.1 Variables 189
A.2.2 Relations between variables 190
A.2.3 Interventions and influences from interventions 191
A.2.4 Miscellaneous 191
A.3 Arctic 3 192
A.3.1 Variables 192
A.3.2 Relations between variables 193
A.3.3 Interventions and influences from interventions 194
A.3.4 Miscellaneous 194
A.4 COIN 1 195
A.4.1 Variables 195
A.4.2 Relations between variables 196
A.4.3 Interventions and influences from interventions 197
A.4.4 Contribution of interventions to action points 198
A.4.5 Miscellaneous 198
A.5 COIN 2 198
A.5.1 Variables 198

A.5.2 Relations between variables . . . 199
A.5.3 Interventions and influences from interventions . . . 200
A.5.4 Contribution of interventions to action points . . . 201
A.5.5 Miscellaneous . . . 201
A.6 *Cybernetia* in Ecopolicy . . . 201
A.6.1 Variables . . . 201
A.6.2 Relations between variables . . . 202
A.6.3 Interventions and influences from interventions . . . 202
A.6.4 Contribution of interventions to action points . . . 203
A.6.5 Miscellaneous . . . 203

B Objective Measures of Complexity Calculations **204**
B.1 Structural complexity C_{DDM} . . . 205
B.1.1 Cyclomatic complexity CC_{DDM} . . . 205
B.1.2 Coefficient of network complexity CNC_{DDM} . . . 205
B.1.3 System dynamics complexity SDC_{DDM} . . . 206
B.2 Information complexity I_{DDM} . . . 206
B.2.1 Information complexity variant I_{DDM} . . . 206
B.2.2 Action complexity variant A_{DDM} . . . 207
B.3 Cognitive weight W_{DDM} . . . 207
B.4 Difficulty D_{DDM} . . . 208
B.5 Distance to linearity L_{DDM} . . . 209
B.6 Uncertainty U_{DDM} . . . 210
B.7 Distance to stability S_{DDM} . . . 210
B.8 Objective parameters of complexity for the DDM scenarios in CODEM . . . 211
B.9 Objective parameters of complexity for the DDM scenarios including Ecopolicy 216
B.10 Objective parameters of complexity for DDM scenarios suitable for a future research program . . . 222

C Details for the Regression Models created through Machine Learning **225**
C.1 Exhaustive search with `leaps` using the adjusted R^2 selection criterion . . . 226
C.2 Exhaustive search with `glmulti` using the $AICc$ selection criterion . . . 229
C.3 Model-averaged importance of the parameters of complexity selected through the exhaustive search . . . 232
C.4 QQ plots of the residuals against the fitted values for the candidate MLR models used for regression diagnostics . . . 235
C.5 Graphs of Cook's distance for the candidate MLR models used for regression diagnostics . . . 238
C.6 Relative importance of the regression model variables for the selected candidate models using 1000-sample bootstrap validation . . . 241

List of Tables

1 Characterization of the seven objective measures of complexity[1] based on features included and type of calculation. . . . 22
2 Task complexity matrix linking a task model with a task complexity model (reproduced from Stouten & Größler, 2017). . . . 24

3 Variables in Arctic 3. . . . 50
4 Relations between variables in Arctic 3. . . . 51
5 Interventions and influences from interventions in Arctic 3. . . . 52
6 Summary of the objective measures of complexity. . . . 54
7 Cognitive weights of basic control structures (BCS), from Wang (2009). . . . 57

8 Summary of the objective measures of complexity. . . . 72
9 Descriptive statistics for the Performance metric used as a response variable in the five DDM scenarios. . . . 74
10 Values for the objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2). . . . 75
11 Candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2). . . . 88
12 Multicollinearity factors for the candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2). . . . 88
13 Relationship between candidate MLR model coefficients for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) and the performance measures. . . . 90
14 Descriptive statistics for the Performance metric used as a response variable in the DDM scenarios including Ecopolicy. . . . 92
15 Values for the objective measures of complexity in the DDM scenarios, including Ecopolicy. . . . 93
16 Cross-validation of candidate MLR models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) with the addition of Ecopolicy, using the stratified *1,000*-repeated *10*-fold method. . . . 95

17 $RMSE$ for robust regression of candidate MLR models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) without Ecopolicy, using the iterated re-weighted least squares (IRLS) method. 120

18 $RMSE$ for robust regression of candidate MLR models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) with Ecopolicy, using the iterated re-weighted least squares (IRLS) method. 120
19 Robustness weights for the $SDC_{DDM} + I_{DDM} + D_{DDM}$ model of objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) without Ecopolicy, using the iterated re-weighted least squares (IRLS) method. Convergence in 14 IRLS iterations. 21 observations are outliers with $|\text{weight}| = 0$ ($< .0005$). 24 weights are ~ 1. The remaining 154 weights are summarized in the table below. 121
20 Candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using group mean performances only. 122
21 Candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using group median performances only. 123
22 $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2). 125
23 $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) with Ecopolicy. 126
24 $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using mean performances. 127
25 $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using median performances. 127
26 $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2), including Ecopolicy, using mean performances. 128
27 $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) including Ecopolicy, using median performances. 128
28 Values for the objective measures of complexity in DDM scenarios for a future experimental design. 160

29 Variables in Arctic 1. 187
30 Relations between variables in Arctic 1. 188
31 Interventions and influences from interventions in Arctic 1. 188
32 Variables in Arctic 2. 189
33 Relations between variables in Arctic 2. 190
34 Interventions and influences from interventions in Arctic 2. 191
35 Variables in Arctic 3. 192
36 Relations between variables in Arctic 3. 193
37 Interventions and influences from interventions in Arctic 3. 194
38 Variables in COIN 1. 195
39 Relations between variables in COIN 1. 196
40 Interventions and influences from interventions in COIN 1. 197

41 Contribution of interventions to action points in COIN 1. 198
42 Variables in COIN 2. 198
43 Relations between variables in COIN 2. 199
44 Interventions and influences from interventions in COIN 2. 200
45 Contribution of interventions to action points in COIN 2. 201
46 Variables in Ecopolicy. 201
47 Relations between variables in Ecopolicy. 202
48 Interventions and influences from interventions in Ecopolicy. 202
49 Contribution of interventions to action points in Ecopolicy. 203

50 Values of the parameters of complexity OMC_{DDM} for the DDM scenarios. . . . 204
51 Cyclomatic complexity CC_{DDM} for the DDM scenarios. 205
52 Coefficient of network complexity CNC_{DDM} for the DDM scenarios. 205
53 System dynamics complexity SDC_{DDM} for the DDM scenarios. 206
54 Information complexity I_{DDM} for the DDM scenarios. 206
55 Action complexity A_{DDM} for the DDM scenarios. 207
56 Cognitive weights of basic control structures (BCS), from Shao and Wang (2003). 207
57 Cognitive weight W_{DDM} for the DDM scenarios (columns feature the total cognitive weight for each characteristic) *. 208
58 Difficulty D_{DDM} for the DDM scenarios. 209
59 Distance to linearity L_{DDM} for the DDM scenarios (excluding relationships and interventions where y is constant). 209
60 Uncertainty U_{DDM} for the DDM scenarios. 210
61 Distance to stability S_{DDM} for the DDM scenarios (the turns handicap to inertial failure is determined by the ratio of the maximum number of turns minus the number of turns to fail in an idle state, to the maximum number of turns). 211

List of Figures

1 Examples of two independent manipulations: (i) number of exogenous and endogenous variables, and (ii) number of relations between variables (Funke, 2014, personal communication). . . . 12
2 A taxonomy of models of complexity in various research domains (Kinsner, 2010). 16
3 Decision graph for a minimal complex system. . . . 17
4 Decision graphs for the Tribes scenario in CODEM. . . . 17
5 Examples of reinforcing (positive) feedback effects (Vester, 1999). . . . 19
6 Examples of balancing (negative) feedback effects (Vester, 1999). . . . 19
7 Example of interlocking feedback. . . . 20
8 Mapping of the original interlocking feedback and corresponding functional form (Vester, 1999). . . . 20
9 Example of a delayed effect with feedback (Vester, 1999). . . . 20
10 A generic task model for DDM simulation problems (Stouten & Größler, 2017). 23

11 Newell's levels of analysis in cognitive science (Gray, 2012). . . . 29
12 A dynamic overview of a DDM scenario in the Ecopolicy serious game. . . . 35
13 The CODEM in-game interface for the Arctic 3 DDM scenario. . . . 37
14 The CODEM scenario editor interface. . . . 38
15 Introduction of the Arctic 3 scenario in CODEM, presenting the context. . . . 40
16 Introduction of the Arctic 3 scenario in CODEM, presenting the goals. . . . 40
17 The *situation* tab of the Arctic 3 scenario in CODEM, presenting the variables. 41
18 The *relations* tab of Arctic 3, presenting the relationships between variables. . . 41
19 The details of a relationship between two variables in the Arctic 3 scenario. . . 42
20 The *decision* tab of the Arctic 3 scenario, presenting the available interventions. 43
21 The *decision* tab of the Arctic 3 scenario, where interventions have multiple influences. . . . 44
22 The details of the effects of an intervention on a variable in the Arctic 3 scenario. 44
23 The *situation* tab of Arctic 3 in the aftermath of a completed turn of play. . . . 45
24 The *feedback* window detailing the effects of a particular intervention in Arctic 3. 45
25 An events message box during the Arctic 3 scenario. . . . 46
26 The messages section of the *information* tab in Arctic 3. . . . 46
27 The history section of the *information* tab in Arctic 3. . . . 47
28 The feedback section of the *information* tab in Arctic 3. . . . 47
29 The *notes* tab in Arctic 3. . . . 48
30 Ending the Arctic 3 scenario with a failure. . . . 49

31 Decision graph for the SpaceLab scenario. . . . 64
32 Decision graph for the Arctic 1 scenario. . . . 65

33 Decision graph for the Arctic 2 scenario. . . . 65
34 Decision graph for the Arctic 3 scenario. . . . 66
35 Decision graph for the COIN 1 and 2 scenarios. . . . 67
36 Decision graph for the Cybernetia scenario in Ecopolicy. . . . 68
37 Performance scores for the five DDM scenarios. . . . 73
38 Correlogram representing the Pearson (r) correlations for the mean and median performances in DDM scenarios with the objective parameters of complexity. . 77
39 Correlogram representing the Spearman (ρ) correlations for the mean and median performances in DDM scenarios with the objective parameters of complexity. 77
40 Parallel coordinates chart presenting the relationships between the parameters of complexity for the five DDM scenarios. The parameter values are centered and scaled. . . . 78
41 A bivariate plot representing the principal component analysis loadings of the ten parameters of complexity for the five DDM scenarios. . . . 80
42 Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. . . . 83
43 Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. . . . 83
44 Model-averaged variable importance using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. . . . 84
45 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters. . . . 86
46 Cook's distance D plot for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters. . . . 87
47 Relative importance of the model variables for SDC_{DDM}, I_{DDM}, D_{DDM}. . . . 90
48 Performance scores for the DDM scenarios including Ecopolicy. . . . 92
49 A bivariate plot representing the principal component analysis loadings of the ten parameters of complexity for the DDM scenarios including Ecopolicy. . . . 93

50 An example of piecewise linear modeling in the MARS algorithm (from Briand, Freimut, & Vollei, 2004). . . . 110
51 An example of SVM classification. Mapping a 2-dimensional nonlinear categorization function to a linearly separable categorization in 3-dimensional space. . 112
52 An example of SVM regression (Kuhn & Johnson, 2013). The top graph presents the contrast between a standard regression line and the SVM in the presence of a highly influential outlier. The bottom graph shows the observations used as support vectors (grey circles) maximizing the margin in the fit of the predictions for other samples (red crosses). . . . 112
53 Relationship between a model residual and its contribution to the regression line for several techniques. From Kuhn and Johnson, 2013. . . . 113
54 Tuning the k parameter to improve a k-NN classification algorithm. . . . 114
55 The large margin nearest neighbor (LMNN) algorithm for optimization in k-NN algorithms. . . . 114
56 Schematics of a multilayer perceptron, a simple feedforward artificial neural network. . . . 116
57 Sigmoid function of the linear combination of input predictors. . . . 116
58 Linear function of the hidden units. . . . 116

59 Example of LOESS regression lines for three different span values. The black line is the actual data. . . . 118
60 A bivariate plot representing the principal component analysis loadings of the ten parameters of complexity for the DDM scenarios used in chapters 2 and 3. . 158
61 A bivariate plot representing the principal component analysis loadings of the candidate parameters of complexity for the prospective DDM scenarios. 160

62 Measures of structural complexity based on the CC calculation for the five DDM scenarios. . . . 211
63 Measures of structural complexity based on the CNC calculation for the five DDM scenarios. . . . 212
64 Measures of structural complexity based on the SDC calculation for the five DDM scenarios. . . . 212
65 Measures of information complexity (I_{DDM}) for the five DDM scenarios. 213
66 Measures of action complexity (A_{DDM}) for the five DDM scenarios. 213
67 Measures of cognitive weight (W_{DDM}) for the five DDM scenarios. 214
68 Measures of difficulty (D_{DDM}) for the five DDM scenarios. 214
69 Measures of nonlinearity (L_{DDM}) for the five DDM scenarios. 215
70 Measures of uncertainty (U_{DDM}) for the five DDM scenarios. 215
71 Measures of system stability (S_{DDM}) for the five DDM scenarios. 216
72 Measures of structural complexity based on the CC calculation for the DDM scenarios including Ecopolicy. . . . 216
73 Measures of structural complexity based on the CNC calculation for the DDM scenarios including Ecopolicy. . . . 217
74 Measures of structural complexity based on the SDC calculation for the DDM scenarios including Ecopolicy. . . . 217
75 Measures of information complexity (I_{DDM}) for the DDM scenarios including Ecopolicy. . . . 218
76 Measures of action complexity (A_{DDM}) for the DDM scenarios including Ecopolicy. . . . 218
77 Measures of cognitive weight (W_{DDM}) for the DDM scenarios including Ecopolicy. 219
78 Measures of difficulty (D_{DDM}) for the DDM scenarios including Ecopolicy. . . 219
79 Measures of nonlinearity (L_{DDM}) for the DDM scenarios including Ecopolicy. . 220
80 Measures of uncertainty (U_{DDM}) for the DDM scenarios including Ecopolicy. . 220
81 Measures of system stability (S_{DDM}) for the DDM scenarios including Ecopolicy. 221
82 Structural complexity SDC_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance. . . . 222
83 Information complexity I_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance. . . . 223
84 Difficulty D_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance. 223
85 Uncertainty U_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance. 224
86 Instability S_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance. 224

87 Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 226
88 Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 226
89 Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 227
90 Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 227
91 Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 228
92 Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 228
93 Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 229
94 Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 229
95 Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 230
96 Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 230
97 Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 231
98 Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 231
99 Model-averaged variable importance using CC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 232
100 Model-averaged variable importance using CC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 232
101 Model-averaged variable importance using CNC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 233
102 Model-averaged variable importance using CNC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 233
103 Model-averaged variable importance using SDC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}. 234
104 Model-averaged variable importance using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}. 234
105 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CC_{DDM}, I_{DDM}, D_{DDM} parameters. 235
106 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CNC_{DDM}, I_{DDM}, D_{DDM} parameters. 235
107 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters. 236
108 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CC_{DDM}, U_{DDM}, S_{DDM} parameters. 236
109 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CNC_{DDM}, I_{DDM}, S_{DDM} parameters. 237
110 A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the SDC_{DDM}, U_{DDM}, S_{DDM} parameters. 237

111 Cook's distance D plot for the MLR model using the CC_{DDM}, I_{DDM}, D_{DDM} parameters. . . . 238
112 Cook's distance D plot for the MLR model using the CNC_{DDM}, I_{DDM}, D_{DDM} parameters. . . . 238
113 Cook's distance D plot for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters. . . . 239
114 Cook's distance D plot for the MLR model using the CC_{DDM}, U_{DDM}, S_{DDM} parameters. . . . 239
115 Cook's distance D plot for the MLR model using the CNC_{DDM}, U_{DDM}, S_{DDM} parameters. . . . 240
116 Cook's distance D plot for the MLR model using the SDC_{DDM}, U_{DDM}, S_{DDM} parameters. . . . 240
117 Relative importance of the model variables for CC_{DDM}, I_{DDM}, D_{DDM}. . . . 241
118 Relative importance of the model variables for CNC_{DDM}, I_{DDM}, D_{DDM}. . . . 241
119 Relative importance of the model variables for SDC_{DDM}, I_{DDM}, D_{DDM}. . . . 242
120 Relative importance of the model variables for CC_{DDM}, U_{DDM}, S_{DDM}. . . . 242
121 Relative importance of the model variables for CNC_{DDM}, U_{DDM}, S_{DDM}. . . . 243
122 Relative importance of the model variables for SDC_{DDM}, U_{DDM}, S_{DDM}. . . . 243

List of Acronyms

A $_{\mathbf{DDM}}$	measure of action complexity
adj. R^2	coefficient of determination
AICc	Akaike information criterion
aNN	artificial neural networks
C $_{\mathbf{DDM}}$	measure of structural complexity
CC $_{\mathbf{DDM}}$	cyclomatic complexity
CNC $_{\mathbf{DDM}}$	coefficient of network complexity
D $_{\mathbf{DDM}}$	measure of difficulty
DDM	dynamic decision-making
I $_{\mathbf{DDM}}$	measure of information complexity
***k*-NN**	k-nearest neighbors
L $_{\mathbf{DDM}}$	measure of nonlinearity
LOESS	locally weighted scatterplot smoothing
MARS	multivariate adaptive regression splines
MLR	multiple linear regression
OMC $_{\mathbf{DDM}}$	objective measure of complexity
RMSE	root-mean-square error (standard error of the regression)
S $_{\mathbf{DDM}}$	measure of instability
SVMr	support vector machine regression
U $_{\mathbf{DDM}}$	measure of uncertainty
W $_{\mathbf{DDM}}$	cognitive weight complexity

In honor of Daniel Bernoulli's ***Specimen Theoriae Novae de Mensura Sortis*** *(1738), where he formulated the expected utility hypothesis, in order to model choice under uncertainty, thereby laying the foundations for modern decision theory.*

In dedication to my wife, Svetlana, and my son, Felix.

You keep my life always interesting and challenging, at the narrow frontier between order and chaos.

I think the next century will be the century of complexity.

Across the frontiers of science, this new more complete, whole systems approach is replacing the old reductionist paradigm, where scientists traditionally tried to understand and describe the dynamics of systems by studying and describing their component parts. Complexity science is moving us away from a linear, mechanistic view of the world, to one based on nonlinear dynamics, evolutionary development and systems thinking. It represents a dramatic new way of looking at things; not just looking at more things at once. Insights from complex systems research provide a new *theory-driven framework* for thinking about, understanding and influencing the dynamics of complex systems, issues and emerging situations.

- Stephen Hawking, 2003

Acknowledgements

I would like to thank Dr Sébastien Tremblay, my dissertation advisor, for having given me the opportunity of pursuing a doctoral degree in research psychology in a field which never stops to fascinate me. Dr Tremblay provided an unwavering motivation to pursue every milestone along the way of the doctoral requirements, no matter how big or small, and was always there to provide precious feedback in a swift manner. His enthusiasm and energy at the laboratory, in meetings, and at conferences will always surprise me, he is a force to be reckoned with and an inspiration for hard work where no compromises are acceptable.

I would also like to thank Dr Daniel Lafond and Dr François Vachon, members of my dissertation committee who are as friendly as they are admirable, and provided me with insights, inspiration, and motivation along those three years in the laboratory. Special thanks to Dr Jean-François Gagnon, who was very involved in the research project on dynamic decision-making from a time well-before my arrival at the lab, and who also provided me with considerable insights and feedback during my tenure at Université Laval. This work would not have been possible without the help of fellow grad students and undergraduates from the Co-DOT lab who ran participants, collected the experimental data, and provided support for the analyses and presentations related to the project, namely Jean-Denis Latulippe-Thériault, Marie-Ève St-Louis, Philippe Carpentier, Roxanne Poulin, Julien Lambert-Song, Louis-Charles Reny et Laurence Caron.

Last but not least, I would like to thank my parents, Lina and Pierre, for encouraging me throughout my long and turbulent academic life. Your support will always be remembered. To my wife, Svetlana, and to my very young son, Felix, I thank you dearly for your patience, for your kindness, and for all the little things that you populated my daily life with. You provided me with a constant distraction from the seriousness of academic and professional life, through the best of times, and through the worst of times. I love you with every decision I make, no matter how good or bad, no matter how simple or complex, and no objective measure of complexity will ever capture that.

Prologue

Cognitive psychology is interested in understanding the elementary information processes on which are organized more complex cognitive functions, through operations on mental representations. Those low-level cognitive processes (such as attention, memory, and perception) are elementary units of cognitive functions, grounded in the neural activity, and operate on small time scales. The higher-level cognitive functions (such as decision-making, language comprehension, numeracy, etc.) rely on the interactions of those elementary information processes, and the internal codes on which they operate are also progressively more complex, from low-level encoding of percepts to the intricacies of semantic representations for language comprehension, to the mental models involved in reasoning and inference.

What features of complex decision-making impact human performance in the cognitively demanding, ever-changing and time pressured, complex sociotechnical systems that are our modern workplaces? Complex decision-making is a macrocognitive construct involving metacognition, or thinking about one's own thought processes. In the particular case of complex decision-making, it involves systems thinking, which is best studied through simulation-based experimentation, using computerized microworlds in the form of serious games. That research paradigm is called dynamic decision-making (DDM), as it takes into account the features of complex decision problems, such as complex sequences of interdependent decisions and changes in problem states, which may change spontaneously or as a consequence of earlier decisions, and for which the knowledge and understanding may be only partial or uncertain.

For all the research in DDM concerning the pitfalls of human performance in complex decision problems, the acquisition of knowledge about complex systems, and whether a learning transfer is possible, there is no quantitative measure of what constitutes a complex decision problem. The research literature mentions the qualities of complex systems, such as a system's dynamical relationships, the nonlinear evolution of the system over time, and the uncertainty about the system states and decision outcomes. This dissertation presents the concepts, methodology, and results involved in a research endeavor to develop objective measures of complexity based on characteristics of dynamic decision-making problems which can explain and predict human performance, drawing on the diverse fields of application of complexity theory such as computational complexity, systemic complexity, and cognitive informatics.

Introduction

> Choose a life. Choose a job. Choose a career. Choose a family [...] But why would I want to do a thing like that? — *Irvine Welsh, Trainspotting (1993).*

What are the characteristics of complex decision problems that affect our understanding of the choices we face in everyday situations, from choosing an academic program, choosing a career, choosing a mate, *et caetera*? How do we exert our influence and control over complex and dynamically evolving situations, in the modern sociotechnical systems that are the work environments of many professionals? Complex decision problems are typically characterized by risk and uncertainty in the choices the decision maker is facing. They involve sequences of choices and outcomes which transform the original decision problem over time, and the relationships between those choices and outcomes are sometimes nonlinear in nature. They can also yield delayed effects, aggravating the comprehension of the decision maker in his or her choices.

The fundamental, experimental, and applied streams of cognitive psychology focused on the interactions between people and technology in workplaces are interested in questions concerning the comprehension and control of humans in the face of complex decision-making problems. Yet psychology alone, for all its concepts, methods, and results, does not possess all the answers to tackle the issue of decision complexity, and a broader theoretical approach is required, drawing on concepts and methods from complexity theory, modeling and simulation engineering, and computational learning theory.

This dissertation presents the concepts, methodology, and results involved in a research endeavor to model the objective characteristics of complex decision problems, with regards to their capacity to explain the variance in human performance for such decisions. The present chapter introduces the themes underlying this research project, starting with a discussion covering the psychology of decision-making, from simple decision-making to complex problem solving. A detailed breakdown of the research objectives follows, along with the hypotheses and the proposed structure of the dissertation intended to support the investigation of the impact of the complexity of decision tasks on human performance. A final section presents a summary of the concepts and methods borrowed outside of the field of cognitive psychology, with a particular focus on complex systems theory, computational complexity theory, and cognitive systems engineering.

Chapter 1 will detail the experimental approach to modeling and testing human performance in complex decisions problems, from a discussion about macrocognition to the use of microworlds, to the creation of the objective measures of complexity themselves. Chapter 2 covers the analysis of the experimental results, where a number of candidate models for the objective measures of complexity are evaluated against the human performance data, including a validation analysis to assess their generalizability. Chapter 3 presents an alternative analysis of the same experimental data from the point of view of more flexible methods of regression modeling, based on issues encountered in the analysis phase of the former chapter, with the aim of improving model accuracy without compromises. Chapter 3.4.3, the discussion section, synthesizes the findings in light of the objectives of the project and presents a research program which could further the investigation of the impact of the features of complexity of dynamic decision-making on human performance.

From simple decision-making to complex decision-making

Historically, there has been two main trends in decision theory: a first, *normative approach*, is based on rational choice theory and classical economics, and construes decision-making on *a priori* assumptions, or formal requirements based on mathematical equations concerning utility functions. It features rational agent, which have well-defined preferences, and an extensive knowledge of the options and choice features involved in such a process.

A second, *descriptive approach*, is based on decision behavior theory (behavioral economics) and the cognitive science of decision-making, which is primarily concerned with the consistency between models of decision and the empirical evidence for decision behavior. It features agents with bounded rationality, who use heuristics (rules of thumb in the decision process), are subject to cognitive biases, and have limited knowledge of the possible choices and outcomes of decisions, as well as limited knowledge of their own preferences.

A third, more recent approach is the *prescriptive approach*, an instrumental view which aims to provide the means by which rationally-bounded agents can improve their decision-making, i.e., to get as close as possible to the normative requirements of rational choice theory. The prescriptive approach is thus an application of the theoretical findings made available through both the normative and descriptive approaches. The following section details the characteristics of decision-making theory, from research interested in simple decision problems, to the more challenging domain of complex and dynamic decision-making.

The psychology of decision-making

What is decision-making?

Decision-making is the process by which we choose an option or a course of action among alternatives, based on a variety of features, such as preferences and expected outcomes (Tversky & Kahneman, 1974, Kahneman, Slovic, & Tversky, 1982, Tversky & Shafir, 1992, Shafir & Tversky, 1995, Shafir, 1999). A variety of features and contexts influence the process of decision-making, such as the complexity of the options and features at hand, contexts of conflicting preferences, and decisions involving uncertainty. Decision-making is studied by an amalgam of research areas grouped under the umbrella of decision theory, but the disciplines concerned with decision-making have traditionally little to do with each other, such as the research on individual factors as studied by decision behaviorists and cognitive scientists, and the group dynamics and relations studied by economics and game theory.

Decision-making is commonly analyzed as an incremental process involving steps or stages, such as a coarse-grained primary distinction between a prior information gathering stage, a midway stage involving deliberation and evaluation, and finally, making a choice among the available options. Such fragmentations of the decision process are more specific when applied to contextualized decision tasks, such as the rather simple evaluation of the probability of a wager, or a more complex task involving conflicting preferences, uncertain outcomes, and multiple features to be weighted, as in the case of establishing a diagnostic, a prognostic, and prescriptive measures for the treatment of a malignant disease for a physician.

Normative and descriptive decision-making

A first, normative account of decision-making comes from economics and philosophy, owing to utilitarianism's conception of rationality as the maximization of utility, i.e., the measurement and representation of preferences. This account, named *rational choice theory*, claimed to be neutral of any psychological features, and embodied an ideal representation of instrumental rationality. Indeed, rational choice theory shares many features with logic, such as being essentially syntactic, axiomatic, and formal, emphasizing consistency above all (Bernoulli, 1738, Von Neumann & Morgenstern, 1944, Savage, 1954). Rational decision makers purport to maximize their decisions based on preference ranking, a process called *expected utility*, involving the assignment of subjective probabilities on outcomes and choice preferences. The process of calculating the expected utility for rational decision makers involves the consideration of only a restricted set of abstract alternatives, emphasizing only the essential features, and differences, of the choice options and their features. This restriction is essential even for rational choice theorists, since even they concede that the range of possible choice options and preferences cannot be represented and processed, both in principle and in practice. Rational choice theory and rational decision-making achieve consistency in expected utility calculations based on a number of 'favorable' conditions, such as the complete pre-ordering of alternatives in the decision process, and the aforementioned restriction of alternatives to a finite set. Given those two requirements, finding maximally desirable, or preferred, alternatives is then made possible.

The depiction of the decision process provided by rational choice theory is appealing, and would be useful if most decision tasks would happen to be solved as such. But the rationality assumption of rational decision-making and expected utility theory is unrealistic, and a normative account of decision processes fails to capture the actual performance of decision makers. A descriptive account of decision-making, as provided by decision behavior theory and cognitive science, shows that the reality of human decision is far from the idealized, formal, and consistent depiction of rational agency. The psychological neutrality assumption of instrumental rationality does not provide an accurate model of real-world decision-making, having sacrificed the correspondence between theory and phenomena for the consistency of an abstract, axiomatic theory of choice and preferences.

A significant number of violations of the principles of rational choice theory have been investigated, and the descriptive approach to decision-making has by now a well-established corpus of knowledge on the psychological factors involved in decision processes. Decision makers commonly violate reflexivity, transitivity, and completeness in framing and evaluating choices and preferences. The fluctuation of preferences over time, caused by additional deliberation, experiences, and actions, have little to do with the preference weighting of expected utility theory, and decision makers are rarely found to optimize or maximize the utility of their choices, but rather opt for satisfactory alternatives, as Herbert Simon coined by the term of '*satisficing*' (Simon, 1956). The departure from rational choice theory to the *bounded rationality* assumption (Simon, 1957) of the descriptive approach to decision-making has been observed in all of human activities, from everyday laymen decisions to expert decision-making (Hutton & Klein, 1999, Hastie & Dawes, 2001, Gilovich, Griffin, & Kahneman, 2002). The idealized tenets of expected utility and rational choice theory are too strong to ever be met, implying logical omniscience, and an unbounded access and comprehension of the decision maker's own choices and preferences.

A foundational research program in the descriptive decision-making literature is *prospect theory*, developed by Tversky and Kahneman (1974. See also Shafir & Tversky, 1995, and Gilovich, Griffin, & Kahneman, 2002). Contrary to what is postulated by expected utility theory, empirical studies on decision-making revealed some factors that have been found to influence the decision-making process and contradict normative assumptions, such as framing effects, which occur when different descriptions of an identical problem task cause the decision makers to choose different alternatives. Describing a decision task in terms of gains or losses influence the evaluation process, and can give rise to contradictory subjective utility values in decisions involving risk. This phenomenon thus violates the requirement of description invariance, as postulated by rational choice theory and expected utility theory. Another factor that has been found to conflict with normative assumptions and requirements is the way that preferences are elicited, i.e., the way that people evaluate and weight their choices. According to the requirement of procedure invariance, as postulated by rational choice theory, logically equivalent ways of weighting preferences should give rise to identical preference orderings. Prospect theory provides evidence that the use of different means of preference elicitation, such as by relative valuation or independent evaluation according to a metric, often violates procedure invariance. All such characterizations violate the normative intuitions of rational choice theory, but are replicable through simple decision tasks involving explicit and complete information to the decision makers, in contexts where transparency is not an issue (Shafir, 1999).

Heuristics and biases

The choices that people make are influenced by a great number of factors, which generally escape the breadth of rational choice theory and expected utility (Fischhoff, 2002). To name a few: *procedural factors* (such as the characterization of conflicting and complex alternatives in a decision, inabilities regarding future predictions and hindsight judgment, and the influence of resources already involved in decisions), *emotional factors* (such as feelings of regret or satisfaction anticipated if other alternatives would have proven to be better choices, and the attachment to alternatives already taken), and justification or *consistency factors* (such as the role of reasons in adjudicating the decision of one alternative over others, and the effects of distance in time concerning past choices, and future decisions). Thus, a number of psychological factors influence the decision-making process, and as mentioned earlier, the psychological neutrality assumption of the conception of instrumental rationality involved in the normative approach to decision-making is untenable.

Research conducted through behavioral economics theory and cognitive science has provided a wealth of insights into the psychological factors that influence decision-making. One central domain of investigation of descriptive decision theory concerns *heuristics and biases* (Tversky & Kahneman, 1974). Decision makers, as mentioned above, use heuristics, or rules of thumb, in order to deal with a wide variety of problems. Those heuristics are surprisingly efficient in practical terms, but can lead to errors, from trivial to systematic ones. The use of heuristics is most prevalent in conditions of uncertainty (which is explored in the next section). Reliance on such judgment heuristics causes decision makers to be biased in the way they deal with problems and decisions (Kahneman, Slovic, & Tversky, 1982).

The original contribution of the pioneers of research on judgment heuristics and cognitive biases, Amos Tversky and Daniel Kahneman, established three general categories of heuristics that can potentially bias decision makers towards erroneous choices. Those three coarse-grained classes are *availability heuristics* (reliance on the ease of retrieval of instances of a category to assess the size of that category), *representativeness heuristics* (reliance on salient properties of events to assess the likelihood of such events), and *anchoring-and-adjustment* *heuristics* (the estimation of a quantity or sample size by relying on some initial value, and then trying to assess how it might be larger or smaller than the chosen base value. Gilovich, Griffin, & Kahneman, 2002).

Uncertainty

Decisions often involve a degree of uncertainty, whether in the form of inaccurate or incomplete information, and decision makers can even be uncertain about their own degree of uncertainty (Kahneman, Slovic, & Tvserky, 1982, Howard, 1988, Henrion, Breese, & Horvitz, 1991). Uncertainty is traditionally represented, and dealt with, through the mathematics of probability. But probability and uncertainty are themselves controversial concepts, since they are hard to define. In order to relate them to decision-making, they must first be clarified. One concern in the characterization of uncertainty is epistemic: what type of uncertainty are we concerned with in probabilistic models, fuzzy logic, and non-monotonic logics? Probability is usually defined as the degree of belief in a proposition being true, concerning the likelihood of an event (Shafir & Tversky, 1995). Two opposite views have put forward an interpretation of uncertainty with regards to probability: the frequentist view and the personalist view. The former approach characterizes probability and uncertainty from an objective perspective, holding that probability is the frequency of some event occurring over a large number of trials, whereas the latter holds a subjective, and more pragmatic account of probability as a person's degree of belief in a proposition being true, in light of all currently known information.

Uncertainty must be carefully modeled and represented in any account of the processes underlying decision-making, as uncertain information and uncertain inferences are essential to any situation involving risk, complex and multiple variables (such as highly complex choice alternatives and choice features), multiple decision makers who might not be collaborating (in competitive and strategic situations), and a dynamic environment in which the validity and value of probability assessments and degrees of confidence is fluctuating. As supported by both the normative and descriptive approaches to decision-making, subjective probability is essential in describing attitudes towards risk, and uncertainty biases decision makers towards risk aversion, or risk-seeking, depending on the point of reference, i.e., the domain being one of gains or losses.

Multi-attribute choices and multidimensional features weighting

Uncertainty is not the sole source of complexity in the decision-making process. Real-world decisions, quite unlike monetary gambles and simple laboratory tasks, involve multiple, sometimes highly complex features and requirements that do not lend themselves to any easy tractable way of figuring out the best alternative. People have been found to be generally uncertain about which features of the choice alternatives to focus on, and the proper way of weighting such features (Shafir & Tversky, 1995). The attribution of weight and preference to features and alternatives has been observed to be often contingent on the description and nature of the alternatives in a decision task, and on changes of the decision task itself that do not lead to significant alterations in the expected outcomes. Additionally, prospect theory has revealed that conflicting alternatives in a decision task often lead decision makers to search for further alternatives, when better ones are available but the decision problem is hard, than when inferior alternatives are available, but the decision task is easier.

The descriptive approach to decision-making also provides evidence against traditional assumptions in economics, such as the evaluation of options being invariant regardless of the method used for such evaluations (Shafir, 1999). Multi-attribute choice options are evaluated according to the individual attributes' contribution to the desirability of the option, but evidence shows that the weighting process is influenced by its compatibility with required or expected responses. A decision maker can for example give more weight to the potential payoff, or financial outcome, in pricing a gamble, rather than in choice between gambling alternatives. The compatibility effect can also lead to other interesting biases, such as preference reversal, a situation in which the desirability of alternatives yields preference based on one attribute or feature of the alternatives, but to the opposite result when the emphasis is shifted to another feature.

Dynamic decision-making

Managing complex sociotechnical systems in various domains such as healthcare, defense and security, transportation and critical infrastructures, all depend on an understanding of the dynamic interrelations of such systems' components, their evolution over time, and the degree of uncertainty to which decision makers are exposed. Beyond the necessity of understanding the problem spaces they have to deal with, decision makers must also prove skillful and adaptive so as to determine how to successfully influence a complex situation, anticipate the consequences, react to surprises, and meet durable objectives. Shortcomings in the comprehension of the impacts of minute interventions, or even of long term strategies, may lead to disastrous consequences. It is thus essential to insure that decision makers are made aware of the challenges inherent to dynamic decision-making, through appropriate training and supported by adequate technologies and organizational strategies.

Interactive learning environments (ILE) based on simulations of complex systems and framed as serious game can be valuable tools to facilitate performance and learning in dynamic decision-making (Karakul & Qudrat-Ullah, 2008), and to conduct experiments focused on the relationship between decision problem complexity and human performance (Pronovost, Gagnon, Lafond, & Tremblay, 2014). Those simulations allow the compression of time and space in which complex decision problems unfold, and thus provide an opportunity for participants to learn about the intrinsic dynamic properties of complex systems from both a piecemeal and a holist points of view, by feedback on common non-adaptive heuristics provided throughout a simulation run. The breadth and depth of feedback from such simulated environments may support a metacognitive function, i.e., support thinking about one's own decisions, and thus provide solid foundations for the development of *systems thinking*, i.e., understanding how variables interrelate when dealing with complex systems (Dörner, 1989, Sterman, 1989, Funke, 1995, Vester, 1999, Bakken, 2008, Gonzalez, 2012, Lafond, DuCharme, Gagnon & Tremblay, 2012). Metacognition, in a broad sense, is the thoughts that a person has about their own thoughts, which is further charcaterized as to include how effective a person is at monitoring their own performance on a given task (self-regulation), a person's understanding of their capabilities on particular mental tasks, and the ability to apply cognitive strategies (Flavell, 1979, Metcalfe & Shimamura, 1994, Schraw, 1998).

Characterizing dynamic decision-making

Dynamic decision-making is a research paradigm interested in determining the factors underlying the strategic processes involved in complex problem solving, with a particular focus on the decision-making process itself, beyond the more traditional interest for decision results (Dörner, 1986, Brehmer, 1992, Diehl & Sterman, 1993, Funke, 1991, Gonzalez, Vanyukov & Martin, 2005). It draws on normative decision-making theory's idealization of concepts such as utility, rational agency, and optimality (the rational choice theory of Bernoulli, 1738, Von Neumann & Morgenstern, 1944, Savage, 1954) while nevertheless considering the constraints of descriptive decision-making concerning human decision makers' bounded rationality, the use of heuristics and the vulnerability to biases, and the impact of uncertainty on cognition (Simon, 1957, Ellsberg, 1961,Tversky & Kahneman, 1974). DDM has a direct inheritance from the idea of operative intelligence (Dörner, 1986), where decision makers gather information, elaborate goals, plan courses of actions and make decision, and finally, must manage its own resources in order to cope with decision constraints (Funke, 2010). The earliest references to dynamic decision-making are imputable to Edwards (1954, 1961, 1962) who defines dynamic decision-making as the category of complex decision tasks which have three common features (referenced in Karakul & Qudrat-Ullah, 2008, and Hotaling, Fakhari, & Busemeyer, 2015):

— a series of actions must be taken over time to achieve some overall goal,

— the actions are interdependent so that later decisions depend on earlier actions, and

— the environment changes both spontaneously and as a consequence of earlier actions.

Based on observations concerning the misperception of feedback (Sterman, 1989) in complex and dynamic decision problems, as well as observations of different stages of information processing such as the distinction between knowledge acquisition and knowledge application (Funke, 2001), researchers interested in DDM developed a number of approaches to tackle the cognitive and behavioral constraints involved in complex problem solving, such as the use of modeling and simulation technologies like serious games and microworlds (Brehmer & Dörner, 1993). The microworlds approach exhibit more ecological validity (Neisser, 1976) than traditionally artificial and contrived laboratory experiments of experimental psychology, aiming for the generalizability of experimental results in more naturalistic environments where macrocognition can be studied (Cacciabue & Hollnagel, 1995, Klein et al, 2003, Klein, 2008), ranging from the psychology of everyday decision heuristics (Gigerenzer & Selten, 2002), to the study of sociotechnical systems (Trist & Bamforth, 1951, Card, Moran, & Newell, 1983, Vicente, 1999). The concepts of macrocognition, simulation-based experimentation, and microworlds are explored in more detail in the following chapter.

Formally, dynamic decision-making is a form of *stochastic optimal control theory* (Hotaling, Fakhari, & Busemeyer, 2015), where the optimal performance of the decision maker is the solution to an objective function (a cost minimization function, or any other form of optimization function). A DDM tasks can thus be modeled as a stochastic linear optimal control problem (Bellman, 1957, Bertsekas, 1976, Rouse, 1980), through linear structural equations, dynamic programming methods, or nonlinear control models (Rapoport, 1967, Zionts & Wallenius, 1976, Cormen, Leiserson, & Rivest, 1990, Funke, 1991, Holland, 1994). Note that while the dynamic decision-making scenarios in this research project contain complex, nonlinear relationships and outcomes, as well as events unanticipated by the decision makers, those DDM problems are not stochastic. They are deterministic in nature, i.e., rule-based and event-driven scenarios, where the initial conditions and the conditional events are all scripted in advance, and never change from one simulation to another. This is done in order to control as much variability in the human performance as possible, relative to the complexity of the DDM problems.

Karakul and Qudrat-Ullah (2008) suggest that while knowledge about the characteristics of dynamic decision-making proper are well-known by now (citing Sterman, 1989, Bakken, 1993, and Funke, 1995, to name a few), we lack a certain knowledge concerning the interactions between decisions, the decision makers themselves, and the decision-making context. Karakul and Qudrat-Ullah distinguish between three different, yet complementary research targets for DDM, namely *task performance* (an operationalized concept aiming to explain how decision makers achieve a certain degree of success in controlling a dynamical system, understood as a DDM problem), *task knowledge* (the extent to which a decision maker exhibits an understanding of the dynamical system, i.e., the accuracy of his or her mental model of the DDM problem), and *transfer learning* (the notion of whether learning about, and performing in, DDM problems, dynamical systems, and systems thinking, actually generalizes this knowledge, competence, and performance to other DDM problems). The authors established a short taxonomy of the characteristics of the research on dynamic decision-making, segregated in the three broader categories of (i) the *characteristics of the learner*, which are concerned with research about a decision maker's prior knowledge, task experience, motivation, cognitive styles, computing skills, and decision heuristics; (ii) the *characteristics of the decision task*, which are interested in features such as task transparency, task complexity, and semantic embedding; and (iii) the *characteristics of the decision-making environment*, concerned with the task type, the time pressure, and information feedback. The present research project focuses exclusively on *task performance* as a response variable expressing human capability in coping with dynamic decision-making problems. From the taxonomy of characteristics of interest in DDM research, we will discuss features pertaining to *task complexity*, our primary topic of interest of course, but features concerning *task transparency*, the *task type*, the *time pressure*, and the *information feedback* are inexorably linked to our main subject of the impact of the complexity of dynamic decision-making tasks on human performance.

Human performance in dynamic decision-making

There are diverging views concerning the limits of human performance in dynamic decision-making scenarios. All such observations and explanations are still competing theories in a rather young research program, by science's standards, even by psychology's history and upheavals. A first view comes from Rapoport (1975), who observed that human performance in DDM could be modeled as an optimal model to which information processing constraints are added (i.e., limiting the number of steps a decision maker can plan ahead, such as a maximum of three steps in the future), and by modeling the decision choices with the inclusion of subjective utility calculations in the objective model (i.e., weighting the options with additional subjective criteria). Another view of human performance is that of Brehmer (1992) and Sterman (1989, 1994, Diehl & Sterman, 1993) on suboptimal DDM performances being caused by a mismatch between the model of the dynamical system and the decision maker's mental model. This misconception is particular to the lack of comprehension or control of feedback loops and nonlinear relationships in the stock and flow models of DDM problems, particularly in the presence of delayed feedback loops. Bußwolder (2015) observed that a generic, structured method designed for system thinking in DDM problems increased performance, as did mental model accuracy.

Another subjective approach to DDM performance models decision choices as a multiple linear regression analysis where the decision maker's subjective probabilistic judgment are coefficients in a linear control model (Jagacinski & Miller, 1978, Jagacinski & Hah, 1988), similar to a lens model approach, or probabilistic functionalism, to decision-making (Slovic & Lichtenstein, 1971, Kirlik, Miller, & Jagacinski, 1993, Kleinmuntz, 1993). Some researchers have attempted to model performance in DDM with regards to the use of decision heuristics (Kleinmuntz, 1985, Kleinmuntz, D., & Thomas, 1987, Kerstholt, 1996), where decision makers were observed to overuse information seeking strategies, relative to a strategy favoring an immediate decision in DDM scenarios revolving around health care management. There was no particular gain in performance following the information seeking strategy, which runs contrary to what was observed by Payne, Bettman, and Johnson (1993), Cronin, Gonzalez, and Sterman (2009), and Pronovost, St-Louis, Lafond, Gagnon, DuCharme, and Tremblay (2015).

Payne and colleagues found that decision makers tend to minimize effort while attempting to maximize performance, so the information collection strategy of decision makers in Kleinmuntz' observations is counterintuitive, as the pursuit of information did not follow up with an improvement in performance, and constituted a poor trade-off. Cronin and colleagues found that the suboptimal performance in DDM scenarios was linked in part to the use of inappropriate heuristics. While those heuristics could sometimes produce appropriate responses, they unfortunately misrepresented the relationships and effects from interventions in the DDM scenarios. Pronovost et al (2015) found that information-seeking behavior and strategies tended to improve DDM performance (supported by Greiff, Niepel, Scherer, & Martin, 2016), but the use of decision heuristics traditionally associated to micromanagement or simple decision problems, did not improve the performance of decision makers. Those simple decision heuristics, such as the static allocation of action points (representing the decision maker's pool of resources in a DDM problem) from one simulation turn to another, the flat allocation of those points split evenly through all possible interventions, and the allocation of resources towards the reduction of the distance to particular goals from one turn to another, were not found to be correlated with lower or higher performance scores in a conclusive way.

Another topic of interest in dynamic decision-making research is the research on the determinants of individual differences. Dörner (1980) split his participants in groups based on prior performance in DDM (low and high performers), and observed that high performers were more prone to integrative goal-setting (as opposed to focused goals), they were more intensely involved in information-seeking behaviors about relationships in the dynamical system and feedback from interventions, and were more capable of assessing whether they were progressing towards a DDM problem's objectives (see also Funke, 1991, Lafond et al, 2012, 2016, and Pronovost et al, 2015). Similarly, Bisantz, Kirlik, Gay, Phipps, Walker, and Fisk (2000) found that top performers in DDM problems exhibited a higher degree of consistency in their subjective probabilistic judgments, and not relative to different judgment policies than in low and average performers. Individual differences in situation awareness ability were found to predict the transfer of DDM performance in load and high cognitive workload conditions (Nicholson & O'Hare, 2014). Some researchers could not find significant differences in DDM performance between participants involved in non-conscious (intuitive, implicit) vs. deliberate decision-making (Größler, Rouwette, & Vennix, 2016).

Supporting metacognition in dynamic decision-making

Prescriptive approaches have been suggested, following on the idea of promoting systems thinking, also known as causal cognition, as a key strategy to help decision makers cope with complex decision problems (Senge, 1990). Prescriptive decision-making frameworks targeting different levels of description and explanation in how cognitive systems deal with complexity have been suggested, in line with the multidisciplinary paradigm of the cognitive sciences

(Bechtel, 1994, Newell, 1990, Dawson, 1998). Those prescriptive approaches range from a minimal complex systems framework (Funke, 2014, see Figure 1), differentiating between a "*vary one thing at a time*" (VOTAT) approach where DDM problems may yield analytical solutions, to the complex problem solving (CPS, see Greiff, Fischer, Wüstenberg, et al, 2013) approach of full-fledged, nigh-intractable DDM problems in ILEs featuring intelligent tutoring systems (ITS) to support metacognitive strategies (Diehl & Sterman, 1993, Gonzalez, 2012, Lafond, et al, 2012). In particular, Cronin, Gonzalez, and Sterman (2009) have found that a verbalization protocol (a think-aloud protocol) enhance DDM performance in an accumulation task. Sterman (2010) and his colleagues observed that training graduate students on system dynamics improved their performance for DDM problems in general, while Gonzalez (2005), Kopainsky and Sawicka (2011), and Qudrat-Ullah (2015) have reported that the use of decision support tools in DDM scenarios facilitated the understanding and performance of decision makers. Lafond et al (2016) have observed some modest improvements in DDM performance for decision makers using an intellligent tutoring system, but only in comparisons involving low performers across different experimental conditions, when the bimodal distribution of performance scores are split between low and high performers in highly complex and difficult DDM problems. The overall performance in dynamic decision-making scenarios tends to improve with rehearsal (Mackinnon & Wearing, 1985, Brehmer, 1992, Gonzalez & Quesada, 2003, Lafond et al, 2016, Pronovost et al, 2015), but to a certain threshold of complexity (Pronovost et al, 2015).

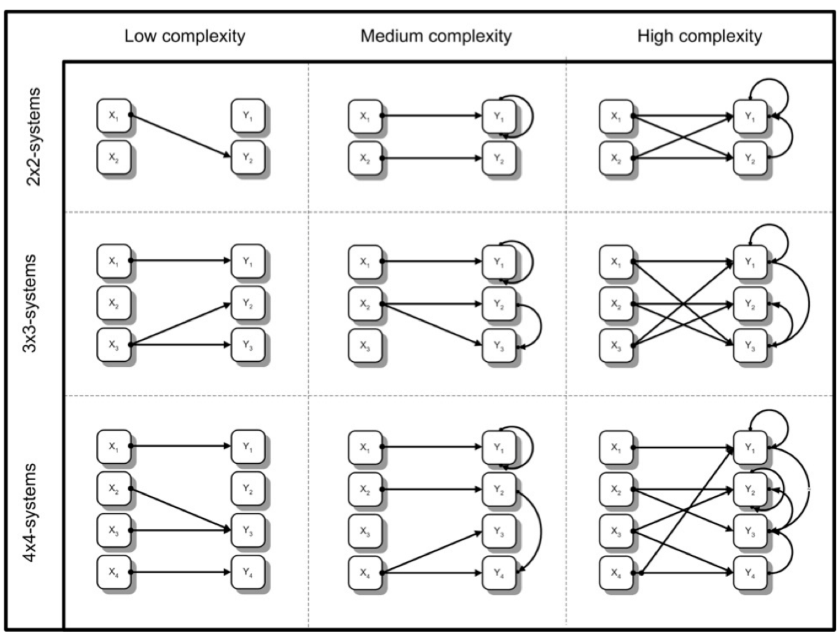

Figure 1 – Examples of two independent manipulations: (i) number of exogenous and endogenous variables, and (ii) number of relations between variables (Funke, 2014, personal communication).

Objective of this research: tackling the problem of quantifying complexity for dynamic decision-making problems

What features of complex decision-making impact human performance? There are many characteristics of decision problems which can be envisioned as parameters of complex systems, from a formal perspective. Dynamic decision-making problems can thus be understood as complex, dynamic systems, which generally exhibit uncertainty, in terms of information opacity or completeness, influencing and limiting the decision maker in its choices and in its comprehension of the system it is attempting to influence; they involve non-linearity and delays in the relationships between parameters and variables, impairing the understanding of complex systems; and they exhibit considerable variability and evolution over time, i.e., the dynamics of a system make it so that its parameters and variables may change states over time, both independently and as a result of human intervention.

In order to understand the numerous issues related to the comprehension and control of complex, dynamic systems, an experimental testbed must represent all such features in a well-parameterized design and produce robust measures yielding some insights into participants' situational awareness and performance in attempting to stabilize a complex decision problem space. This dissertation presents a methodology, as well as some empirical results, to support the analysis of human performance in dynamic decision-making, and provide some insights in the issue of the comprehension of complex systems, by means of objective models of complexity explaining the variance of human performance in DDM scenarios.

The research objectives can be summarized as follows:

The primary objective is to develop objective measures of complexity based on characteristics of dynamic decision-making problems which can explain the variance in human performance scores,

objective 1.1: use DDM scenarios of varying complexity, in terms of structure, information, difficulty, nonlinearity, and uncertainty values, in order to measure the performance of participants with the help of the CODEM microworld, a simulation-based experimentation approach to macrocognition,

objective 1.2: explore the parameter space for separate subsets of the objective measures of complexity (as some of the parameters exhibit redundant characteristics), using exhaustive searches and optimization algorithms from machine learning methods, in order to obtain a few candidate multiple linear regression models explaining the maximum variance in DDM scenario performance,

objective 1.3: assess whether the MLR approach to model the objective characteristics of complexity is a better fit to explain the variance in DDM scenario performance, in comparison with the the same candidate models using alternative statistical regression analyses (MLR using only group means and medians, nonlinear and non-parametric methods, as well as robust regression methods).

There are two hypotheses associated with those goals:

hypothesis 1: the objective measures of complexity will function as parameters in a MLR model explaining the variance in the human performance data, whereby higher net complexity for a DDM scenario over a given subset of complexity metrics will yield lower performance scores, and lower complexity ratings will yield higher performances,

hypothesis 2: the functional form of the relationship between the objective parameters of complexity and the performance scores in DDM scenarios may not be linear, and a nonlinear and/or non-parametric regression model may fit the multivariate data better than the ordinary least squares (OLS) method to calculate a linear regression model's residuals.

The structure of this document organizes the content in order to achieve those goals and to test those hypotheses:

The introduction detailed so far the concept of dynamic decision-making, and further elaborates on the concepts of complex and dynamical systems, the literature on models and metrics of complexity, as well as the challenge of creating complexity measures for dynamic decision-making problems based on research ideas about the complexity of a task,

chapter 1 presents an experimental approach to modeling dynamic decision-making tasks using a simulation-based methodology, after having discussed the appropriateness of this methodology for complex cognitive functions involved at the macrocognitive level of analysis. The objective measures of complexity retained for DDM problems are detailed, including the equations for their calculations,

chapter 2 presents the empirical data for human performance in five DDM scenarios, and where those scenarios are parameterized in accordance with the objective measures of complexity presented in chapter 1. A number of candidate multiple linear regression models are created through an exhaustive search in the parameter space, and a regression model validation is presented, using cross-validation with the inclusion of another DDM scenario (a standard holdout sample validation is explored in chapter 3, using a different MLR approach),

chapter 3 discusses the validity of the candidate MLR models created through the traditional OLS approach and attempts to find nonlinear and non-parametric models which could reduce the variance in the results, to provide better measures of goodness of fit. In order to mitigate the departures from the basic assumptions of linear regression modeling, such as the non-normality of multivariate residuals, the non-constant error variance, and the presence of influential observations, three types of analyses are explored: (i) a robust regression method, (ii) a standard MLR approach using only group mean and median performances, and (iii) statistical machine learning algorithms.

the conclusion synthesizes the results of the previous chapters in light of the research objectives and to assess the plausibility of the hypotheses. A discussion of the contribution of the research project is presented, with regards to the theory, the methodology, and the practical aspects of research on dynamic decision-making. A final section presents a research program designed to explore the parameter space of objective measures of complexity for DDM scenarios in more breadth and depth, so as to weight which combinations of the characteristics of complex decision problems are determinant for the human comprehension and control of complex and dynamical systems.

Modeling complex systems

The complexity of a decision problem from the point of view of human cognition is not the focal point of research in the scientific literature on complexity theory and complex systems dynamics (Diehl & Sterman, 1993, Bar-Yam, 1997, Kinsner, 2010). According to Bar-Yam (1997),

> Loosely speaking, the complexity of a system is the amount of information needed in order to describe it. The complexity depends on the level of detail required in the description.

Models and measures of complexity can emphasize differences such as local vs. global complexity; simple vs. multi-scale models; deterministic vs. probabilistic systems; absolute, differential, or relative calculations; static vs. dynamic measures of complexity; average tendencies vs. asymptotic measures; and arithmetic vs. logical models (Edmonds, 1999, Couture, 2007, Lloyd, 2008). In cognitive psychology, the notion of *relational complexity* has been developed and tested to assess the limits of working memory and information processing of humans relative to the mental representations of arguments and variations (relations) in sentence comprehension involved in problem-solving (Halford, Wilson, & Phillips, 1998, 2010, Halford, Cowan, & Andrews, 2007). A literature review of models and measures of complexity in various fields of research suggests that *algorithmic computational complexity*, as well as *structural systemic complexity*, are more appropriate for the type of problems involved in DDM, given that they are sensitive to information about both quantities of information, as well as the interrelatedness of components (Sipser, 1996, Cardoso, Mendling, Neumann & Reijers, 2006, as well as Kinsner's taxonomy, 2010, seen in Figure 2 below).

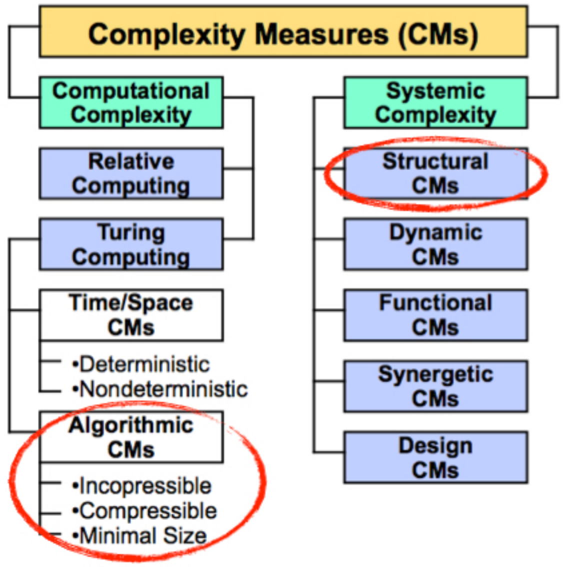

Figure 2 – A taxonomy of models of complexity in various research domains (Kinsner, 2010).

Senge (1990) makes a distinction between the *detail complexity* (structural and informational characteristics) versus the *dynamic complexity* (characteristics pertaining to change, feedback, and the complex interplay between cause and effect) in systems of human activity. The systems metaphor applied to complex decision problems is particularly appropriate when we attempt to model a DDM problem and formalize its states and behaviors. Drawing on concepts from graph theory such as connectivity, graph models, and decision graphs (Sylvester, 1878, König, 1936, Harary, 1969), as well as from concepts of systems dynamics such as stocks and flow models (Forrester, 1961, Dörner, 1989, Sterman, 2000), DDM problems can be studied both formally and empirically so as to ascertain the impact of complexity on human cognition.

DDM problems are *labeled cyclic multidigraphs*: directed graphs with multiple connections, where the connections express transition functions to altered states in the variables (Figures 1, 2, and 3). Input (exogenous) and output (endogenous) nodes are therefore connected by edges in causal loop diagrams, where relationships represent reinforcement (positive) and balancing (negative) loops. For the purposes of discrete time dynamic systems such as the ones represented in the turn-based DDM scenarios used in psychology experiments, the transition functions are recurrences relations (difference equations), instead of the continuous relationships (differential equations) found in stocks and flow models. The systems dynamics of a DDM problem represented as a decision graph involves both external influences (manipulations from a decision maker) and internal influences, or *Eigendynamik*, where the states of certain variables change over time without interventions, and thus exhibit feedback in a complex system. The circuit form (or "oriented cycle") of a decision graph for CPS scenarios allow any variable to be endogenous, exogenous, and/or "eigendynamic" (compare Figures 3 and 4).

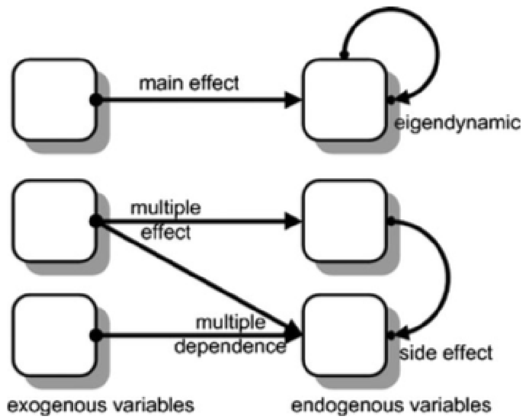

Figure 3 – Decision graph for a minimal complex system.

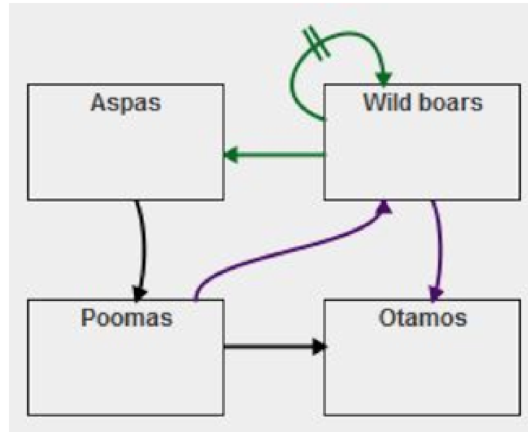

Figure 4 – Decision graphs for the Tribes scenario in CODEM.

From the point of view of the semantics of this formal approach to DDM, the behavior of complex systems is said to exhibit nonlinearity in the relationships between variables, where actions can incur delays in their effects, and the system may be fraught with uncertainty, or intransparency in its components, adding to the burden of comprehension of a DDM problem. The result is an exponential growth of the possible future states of a dynamic system, where the parameter state of the system exhibits a *combinatorial explosion* of anticipated consequences in attempting to influence its behavior. The interconnectivity of complex and dynamic systems thus causes DDM problems to exhibit *polytely*, or a multiplicity of oftentimes conflicting objectives; *polytropy*, a multiplicity of courses of action undertaken to meet such objectives; and *polyplasy*, whereby the uniqueness of a system at time t is caused by the interplay between polytely and polytropy, creating a novel problem space after each choice of a decision maker facing a DDM problem in the CPS paradigm.

The number of functional, i.e., relevant, parts in a complex system varies over time (Brehmer & Allard, 1991), and the intrinsic dynamics of a system may impact its own complexity (Kerstholt & Raaijmakers, 1997). Even an apparently modest DDM problem can thus yield intractable state (problem) spaces, as the formal complexity class of DDM scenarios is connected to *constraints satisfaction problems* for which optimal solutions may not be computable [2]. Quantitative approaches such as multiple-criteria decision analysis (MCDA) may leverage some of those problems by looking for a compromise in efficiency in the form of non-dominated solutions, which mirror Simon's (1956) notion of *satisficing* in human problem solving.

2. There are many formal problems for which analytical solutions may not be knowable, due to reasons such as the limitations of the conceptual framework in which the problems are represented (e.g., *Entscheidungsproblem*, or the decidability problem of Hilbert & Ackermann, 1928, Church, 1936, Turing, 1937), limitations on available computational resources (P *vs.* NP complexity class problems, Cook, 1971), or because of logical and epistemic constraints on knowledge (e.g., Fitch's paradox of knowability, Fitch, 1963, Williamson, 2000).

The biocybernetician, ecologist, and systems theorist F. Vester (1983, 1985, 1999) presented a number of possible relationships in complex and dynamical systems featured in DDM problems. The simplest relationship between variables in a stock and flow model is the *linear relationship*, where the affected stock, or endogenous variable, changes in proportionality to the exogenous variable's input. A second type of relationship is the category of *simple nonlinear relationships*, which covers monotonic functions, such as logarithmic functions, exponential functions, sigmoid functions, and the inverse of linear functions. Those relationships preserve or reverse the order of a set of input values, and are either increasing or decreasing over the domain values. Simple nonlinear relationships may be asymptotic over their range, such as in the particular case of saturation relationships, i.e., relationships for which there is a progressively diminishing effect for additional values over the domain of the function.

Higher-order relationships are polynomials where the coefficients of the variables are transformed through simple operations such as addition, subtraction, multiplication, and non-negative integer exponents. Adding exponents yields polynomial functions where values over the range are reversed for certain values over the domain, such as quadratic and cubic functions common in linear regression models. Some relationships exhibit *effects with limit or threshold values*, where the values of the domain of a function are limited to a particular interval, beyond which the function cannot yield a value over the range . For example, a bow only has a restricted interval of tension values which can be applied. Below a certain threshold value, the lack of tension on the bow will not impact the flying range of an arrow, and above a certain tension limit, the bow will break, reducing the flying range potential to zero. A limiting value can represent a catastrophic point in the relationship, beyond which the functional relationship no longer holds.

Another important type of relationships in systems dynamics and in dynamic decision-making problems are the relationships that exhibit *effects with feedback*, such as reinforcing and balancing effects. *Reinforcing relationships* can have positive upward effects (such as in a scenario where two stocks affect each other through a relationship of accumulation), or positive downward effects (where two stocks affect each other through a relationship of depletion). Reinforcing relationships are thus relationships where two variables affect each other in a way that intensifies their magnitudes (Figure 5). *Negative or balancing relationships* are relationships where the effects act in opposite directions for two stocks. An increase in one variable causes a decrease in another variable, or vice-versa (Figure 6). Effects with feedback can involve more than two variables, where chains of relationships exhibit reinforcing or balancing effects.

When stocks, or variables, are involved in more than one feedback relationship at a time, such complex feedback loops are called *interlocking feedback effects* (Figures 7 and 8). The complex, nonlinear dynamics caused by interlocking feedback effects are very hard to anticipate for decision makers, as they can exhibit oscillations, unstable and aperiodic changes sensitive to initial conditions, and even spontaneous order of radically different patterns. A final category of relationships for complex and dynamical systems concerns *effects with time lags*, where simple to complex relationships only exhibit an effect after a change was made in the stock of origin, or exogenous variable. Applying an intervention with a delayed effect may cause the system to oscillate in an undesirable outcome condition, and decision makers are reported to attempt corrective countermeasures that are suboptimal (Sterman, 1989, 1994, Brehmer, 1992, Diehl & Sterman, 1993) as in the example in Figure 9, where corrective behavior only exacerbates the deviation from the optimal state due to improper understading and control over delayed effects from interventions. In this example, the intended negative feedback effect was turned into a positive feedback loop because of the delay in the effect.

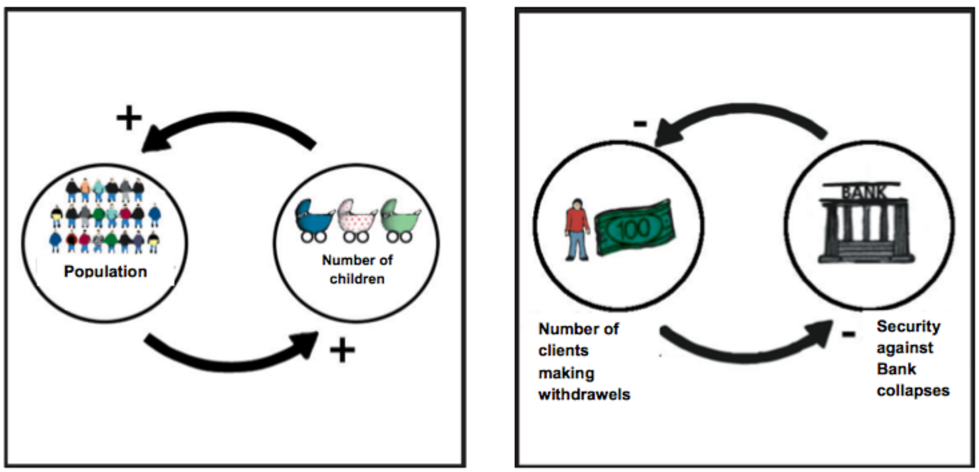

Figure 5 – Examples of reinforcing (positive) feedback effects (Vester, 1999).

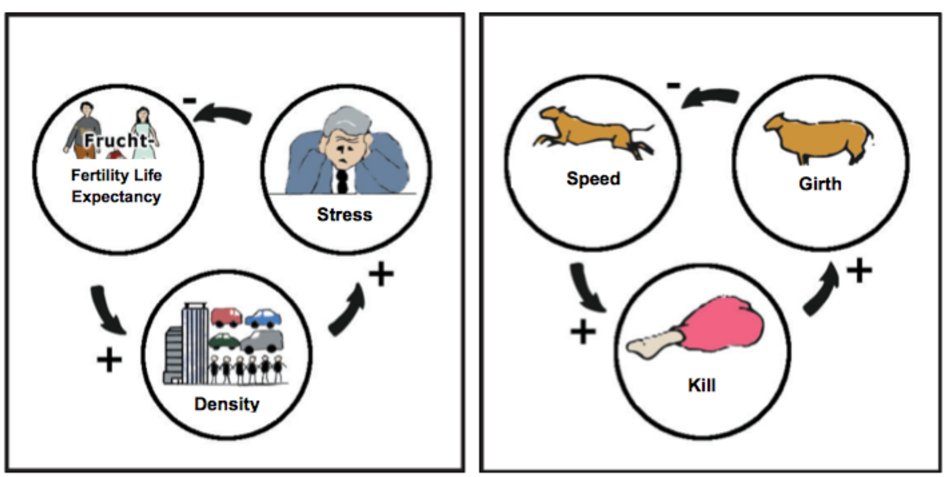

Figure 6 – Examples of balancing (negative) feedback effects (Vester, 1999).

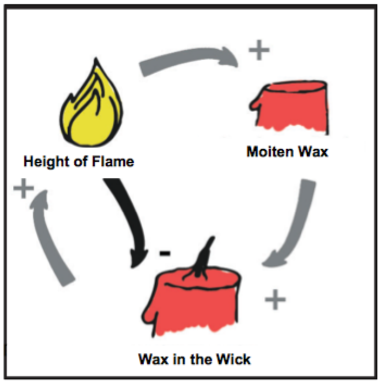

Figure 7 – Example of interlocking feedback.

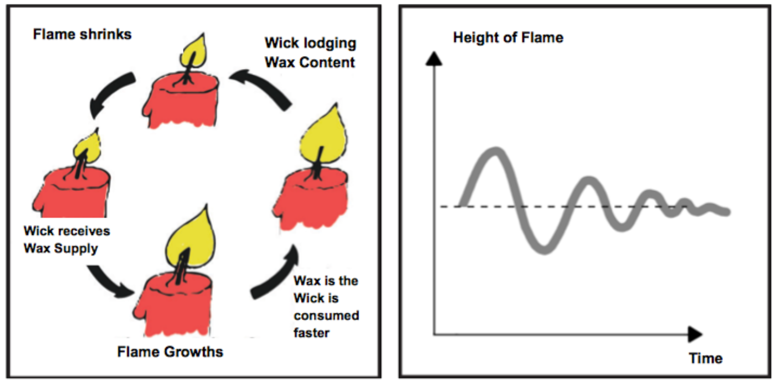

Figure 8 – Mapping of the original interlocking feedback and corresponding functional form (Vester, 1999).

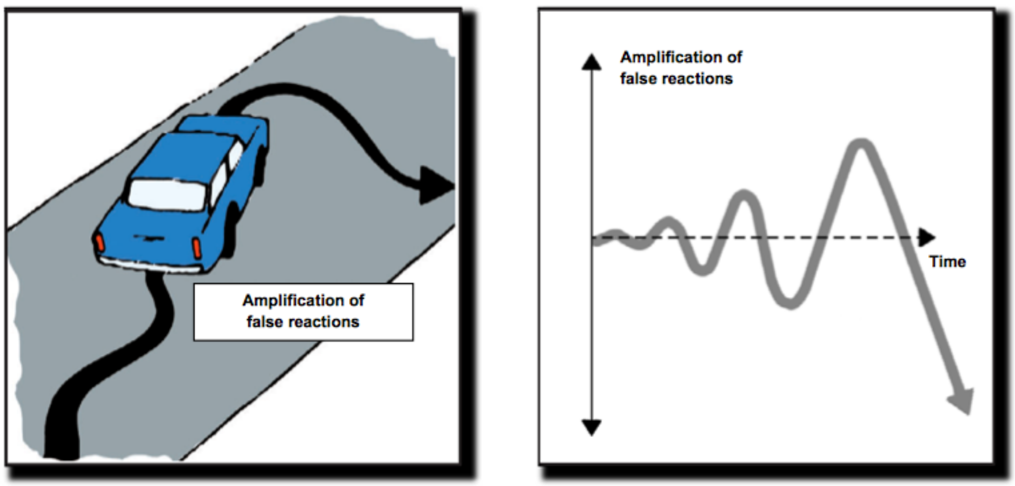

Figure 9 – Example of a delayed effect with feedback (Vester, 1999).

Models of complexity from cognitive informatics

Previous models of complexity used to assess the impact of the complexity of DDM problems on human performance involved the quantification of structure, information, and cognitive load via models suggested in the literature on various areas of science such as graph theory, computability theory, information theory, as well as more applied areas such as business process management and software engineering. Seven measures of complexity were originally retained to assess their relevance and their efficacy in the context of behavioral and cognitive science research, following on the positive empirical results of a number of researchers with regards to the possibility and practicality of deploying objective measures of complexity in the abovementioned various domains (De Silva, Kodagoda, & Perera, 2012, De Silva & Kodagoda, 2013, De Silva, Weerawarna, Kuruppu, et al, 2013, Kinsner, 2010, Pronovost et al, 2014). Some of those objective measures of complexity were composite models, that is, they merged together a number of parameters which blended structure, information, etc. in order to compare models between themselves. The models are presented summarily below. Table 1 presents a synthesized view of the features and means of computation for the various models of complexity (i.e., whether they involved scalars expressing magnitudes or ratios representing relationships).

The foundational work of Wang (Shao & Wang, 2003, Wang, 2007, 2009, Wang, Kinsner, Anderson, et al, 2009) in the innovative domain of *cognitive informatics* provides some insights into more rigorous measures of complexity that include a cognitive component. By including cognitive weights in complexity measures, Wang et al. model the additional *semantic* properties of the comprehension of complex systems, beyond mere structure and information flow. Cognitive complexity is measured as the product of architectural complexity and operational complexity, where architecture refers to structural factors and operations refer to operator-system interactions (Wang, 2007, 2009).

McCabe's *Cyclomatic Complexity* (1976), is a measure inspired by algorithmic graph theory, which determines the number of independent paths in a control flow graph. It is computed as a simple subtraction of the number of graph nodes from graph edges, plus the number of connected components. We use a modified version of *Halstead's Software Metrics* (1977), an implementation-independent complexity measure of algorithms. Halstead's metric depends on the number of unique and total operators and operands, and is thus a measure of algorithmic size, a strictly informational measure of complexity. Cardoso's *Interface Complexity* (Cardoso et al, 2006), adapted from Henry and Kafura's Information Flow (1981), determines the total complexity of an algorithm as the product of structural complexity and information flow, i.e., inputs and outputs. Wang's *Cognitive Functional Size* (Shao and Wang, 2003) computes the complexity of an algorithm as a function of the product of summed inputs and outputs by the "total cognitive weight", this latter construct being itself an additive function of "basic control structures" which bear various cognitive weights determined via empirical studies. Misra's *Cognitive Weight Complexity Measure* (2006), based on Wang's CFS, is a metric considering exclusively the total of cognitive weights, as a function of executed instructions in an algorithm. Kushwaha and Misra's *Cognitive Information Complexity Measure* (2006), also based on Wang's CFS, is a metric combining a weighted information score multiplied by the total cognitive weight. Finally, Wang's *Cognitive Systems Complexity* (Wang, 2007, 2009) is a more rigorous measure of a system's complexity and size, because it represents its real *semantic complexity* (as opposed to mere symbolic quantification) by integrating both the *operational complexity* and the *architectural complexity* of a system in a coherent measure.

These objective measures of complexity were tested against human behavioural data obtained through CODEM experimentations. The goal was to evaluate whether these measures could account for human behavior, comprehension and performance in scenarios of varied complexity. This foundational experiment, presented in Pronovost, Gagnon, Lafond, and Tremblay (2014), guided the elaboration of the objective parameters of complexity presented in the following chapter, where the compound models of complexity measures were broken down into smaller sets of quasi-independent features.

Table 1 – Characterization of the seven objective measures of complexity [3] based on features included and type of calculation.

Models	Features			Calculation		
	structure	information	cognition	absolute	weighted	scaled
CC	✔			✔		
HM		✔		✔		
CWCM			✔		✔	
IC	✔	✔		✔		
CFS	✔	✔	✔	✔	✔	
CICM	✔	✔	✔	✔	✔	
CSC	✔	✔	✔	✔	✔	✔

Task complexity and the complexity of dynamic decision-making problems

An interesting and recent review of what constitutes task complexity features for DDM was produced by Stouten and Größler (2017). Drawing on Liu and Li's (2011, 2012, 2014) and Park's (2014) interest in establishing objective measures of task complexity for systems reliability engineering, Stouten and Größler provided an informal taxonomy of features for a standard task model, as well as features, or dimensions, of a task complexity model for dynamic stock control tasks. The authors suggest that only through a formalized definition of what constitutes a *task* (drawing on Hackman, 1969, Rouse & Rouse, 1979, Kieras & Polson, 1985, Wood, 1986, and Fischer Greiff, & Funke, 2012) can we claim to develop a model of *task complexity* (see also Bedny, Karwowski, & Bedny, 2012). A *task model* is suggested to have objective features such as objectives, inputs, processes, outputs, simulation time, a presentation mode, and a simulation model (Figure 10).

3. CC: Cyclomatic Complexity; HM: Halstead's Software Metrics, CWCM: Cognitive Weight Complexity Measure; IC: Interface Complexity; CFS: Cognitive Functional Size; CICM: Cognitive Information Complexity Measure; CSC: Cognitive Systems Complexity.

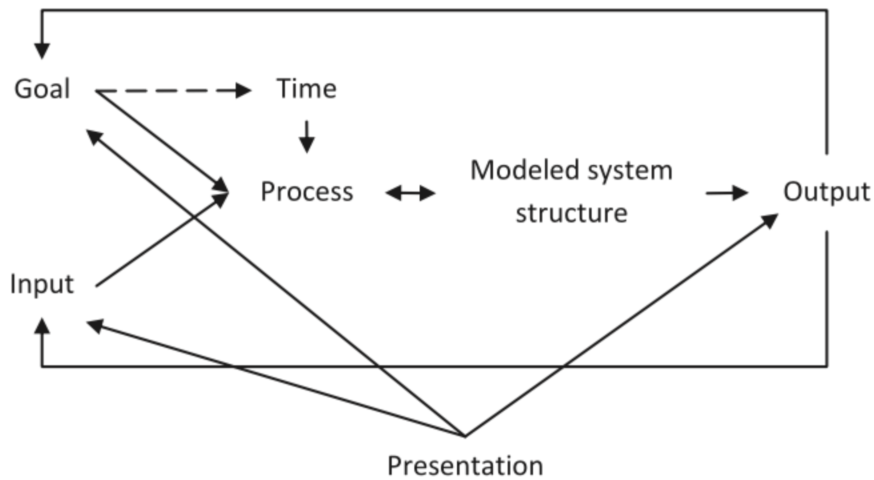

Figure 10 – A generic task model for DDM simulation problems (Stouten & Größler, 2017).

The relationship between the objective features of a task model can be mapped to complexity dimensions relevant to the task complexity model of a dynamic stock control task, yielding ten parameters, namely: the task size, variety, redundancy, ambiguity, variability, unreliability, novelty, incongruity, connectivity, and temporal demand. The resulting matrix (Table 2) constitutes an informal collection of objective features of complexity in the particular context of stock and flow tasks. The three parameters of complexity suggested through the composite models of cognitive informatics (structure, information, and cognitive load) map to some of the features of Stouten and Größler's (2017) task complexity model, with the various metrics of cognitive weight finding an analogue in Liu and Li's (2012) concept of *action complexity*. Yet the complexity of algorithms is not identical to the complexity of dynamic stock control tasks, nor is the task complexity model of a DDM problem used through the CODEM environment identical to the continuous dynamic stocks and flow model underlying Stouten and Größler's (2017) examples. For this reason, complexity metrics for individual parameters of complexity which map to the task complexity model of DDM scenarios were used in the present experiment, and are described in the methodology section below. Those parameters of complexity (namely structure, information, cognitive weight, difficulty, nonlinearity, uncertainty, and instability) are at a sufficiently high level of abstraction to represent dynamic decision-making problems, and should be generalizable to other experiments (examples of earlier attempts to use similar subsets to those parameters for DDM scenarios can be found in Rouwette, Größler, & Vennix, 2004, Osman & Speekenbrink, 2011, and Özgün & Barlas, 2015).

Table 2 – Task complexity matrix linking a task model with a task complexity model (reproduced from Stouten & Größler, 2017).

		Task model						
	Task Component class → **Complexity dimensions** ↓	**Goal**	**Input**	**Process**	**Modelled system structure**	**Output**	**Presentation**	**Time**
Task complexity model	**Size**	• Quantity	• Quantity	• Quantity (of actions / steps / paths / decisions)	• Quantity (of variables / relationships / equations) / and potential states or potential behaviours	• Quantity		
	Variety	• Diversity	• Diversity	• Diversity (of actions / steps / paths / decisions)	• Diversity (of variables / relationships / equations) / and potential states or potential behaviours	• Diversity	• Heterogeneity in presentation format	
	Redundancy	• Redundancy	• Redundancy	• Redundancy	• Redundancy	• Redundancy		
	Ambiguity	• Clarity • Structure	• Clarity • Structure			• Clarity • Structure	• Clarity • Structure	
	Variability	• Rate of change	•Rate of change • Random events		• Robustness • Stochasticity	• Rate of change		
	Inaccuracy	• Inaccuracy	• Inaccuracy		• Technical correctness	• Inaccuracy		
	Novelty		• Non-routine events	• Repetitiveness			• Non-routine presentation formats	
	Incongruity	• Conflict • Mismatch	• Conflict • Mismatch	• Conflict • Mismatch	• Conflict • Mismatch	• Conflict • Mismatch	• Compatibility	
	Connectivity				• Interrelatedness of variables			
	Temporal demand							• Time availability • Concurrency

Chapter 1

Experimental Approach and Methodology

1.1 Introduction

This chapter presents the experimental approach used to explore the relationships between the objective parameters of complexity of dynamic decision-making and human performance. We will first explore the concepts of micro- and macrocognition, and highlight the features of DDM which necessitates a different approach to research concerning the high-level cognitive functions involved in complex task environments. Secondly, the notions of simulation-based experimentation will be presented, where a variety of research tools and techniques are discussed, such as microworlds and serious games. The CODEM simulation environment used in the remainder of the dissertation is presented, along with a detailed example of play in one of the DDM scenarios used for the experimentation. A third and final section presents the detailed breakdown of the objective measures of complexity discussed in the introduction, featuring the rationale for their inclusion in this research project as well as the equations used to extract the characteristics of complexity from the DDM scenarios. The experimental approach presented herein is expanded in chapters 2 and 3, where the details concerning the DDM scenarios used to gather the empirical data are presented. A detailed breakdown describing the implementation of each DDM scenario is presented in appendix A, and B presents the calculations of the objective parameters of complexity for those DDM scenarios.

1.2 Microcognition and macrocognition

Cognitive psychology is interested in understanding the elementary information processes on which are organized more complex cognitive functions, through operations on mental representations (Neisser, 1967, Baddeley, & Bernses, 1989, Eysenck, 1990, Mandler, 2002, Anderson, 2010). Those low-level cognitive processes (such as attention, memory, and perception) are elementary units of cognitive functions, grounded in the neural activity, and operate on small time scales. The higher-level cognitive functions (such as decision-making, language comprehension, numeracy, etc.) rely on the interactions of those elementary information processes, and the internal codes on which they operate are also progressively more complex, from low-level encoding of percepts to the intricacies of semantic representations for language comprehension, to the mental models involved in reasoning and inference (Craik, 1943, Forrester, 1968, 1985, Johnson-Laird, 1983, Gentner & Stevens, 1983, Evans, Newstead, & Byrne, 1993, Ford & Sterman, 1998, Schaffernicht & Groesser, 2009). Cognitive processes are understood to operate in stages, or sequences, and the overall cognitive system of a human agent is thus understood as a system that processes inputs, performing some operations, and producing outputs (be they novel mental representations, outward information-bearing behavior such as communication through language, or simply psychomotor behaviors).

The traditional approach to observing behaviors, manipulating experimental conditions, and producing causal inferences about the underlying cognitive processes, higher-level functions, or the type of mental representations, has been to isolate the cognitive processes in terms of those stages of information processing, which is what the study of microcognition is concerned with. Isolating microcognitive processes has been a dominant and fruitful method in experimental cognitive psychology through the recuperation of methods from psychophysics, behaviorism, and psychophysiology (such as the mental chronometry concerning reaction times, objective measures of response accuracy in perceptual or behavioral outcomes, and the manipulations invovled in the design of experiments). The microcognitive approach to studying the information processes involved in low-level cognition is said to be a *reductionist* theory, insofar as the explanation of the higher level cognitive functions rely on observations and inferences concerning those lower level cognitive processes, on the one hand, and that such cognitive processes could in turn be explained by inferences and observations concerning the even lower level of the underlying neural and biological processes.

The term macrocognition was created by Cacciabue and Hollnagel (1995) to distinguish the systemic approach of modeling cognitive systems from the traditional microcognitive approach of cognitive science in general, and cognitive psychology in particular. The intuition is that, much like it was argued by proponents of ecological validity (Neisser, 1979), interactive behavior (Gibson, 1977, Card, Moran, & Newell, 1983, Norman, 1988), and naturalistic decision-making (Klein, 2008), the kinds of cognitive functions involved in natural environments are more complex, exposed to richer sources of information, and involve time scales well beyond what the contrived and artificial experiments found in the experimental psychology research focused on microcognition. Naturalistic decision-making, in particular, occurs on different time scales than in experimental research concerned with decision processes and outcomes in the controlled environment of the laboratory (Klein, Klein & Klein, 2000, Klein, Ross, Moon, Klein, Hoffman, & Hollnagel, 2003, Klein, 2008). A focus on cognition involving complex decision-making still requires the useful and indispensable information processing metaphor, but its conceptual units and the causal inferences which are made about it involve careful considerations to the *characteristics of the learner* (prior knowledge, task experience, motivation, cognitive styles, computing skills, and decision heuristics), the *characteristics of the decision task* (task transparency, task complexity, semantic embedding), and the *characteristics of the decision-making environment* (task type, time pressure, information feedback). A macrocognitive framework of research involving those three factors could explain the performance, knowledge acquisition, and transfer learning taking place in complex decision-making (Karakul & Qudrat-Ullah, 2008, Qudrat-Ullah, 2014, 2015).

Macrocognition in general, just like the more specific field of naturalistic decision-making, is thus distinguished from microcognition by its reliance on cognitively complex functions in demanding, real-world situations. Those real-world situations are characterized by situational uncertainty, time pressure and time limits, marred by multiple sources of risk of failure, ambiguities in choice outcomes and even in the goals or objectives of the multiple stakeholders, in unstable conditions and constrained by the task environment, organizational constraints, etc. Macrocognition involving complex decision-making often involves the study of experts working in complex sociotechnical systems (Baxter & Sommerville, 2011), as opposed to novices in artificial tasks such as the undergraduates recruited in universities for academic research (Hollnagel, 2001). In a way, macrocognition pursues the opposite trend to scientific reductionism typical in the experimental research on microcognition, in that it embraces multiple levels of description for phenomena and processes pitched at different time scales (see Vespignani, 2012 for a discussion of the challenges of considering different time scales in the modelling of dynamical processes for complex sociotechnical systems). Such is the prerogative of cybernetics, of systems theory, and of contemporary cognitive science. This *holistic* approach favors a hierarchical view of cognitive systems, where higher level functions and complex behaviors are to be explained by reference to lower level entities, properties, and relationships, and can be in turn used to explain even more complex phenomena, such as the individual development, education and learning, as well as organizational and social behavior.

Previous research on dynamic decision-making (Pronovost et al, 2014, Pronovost et al, 2015, Lafond et al, 2016) discussed the the possibility that the combined effect of parameters of complexity and difficulty on DDM performance follows a relationship curve mirroring a *just-noticeable difference* threshold as formulated by the Weber–Fechner law (Fechner, 1860). The hypothesis is that beyond a particular degree of complexity and/or difficulty (the differential *limen*, or threshold, as formulated in psychophysics), decision-makers would no longer be able to achieve any gains in performance, in the likeness of a monotonic curve, such as an inverse scale or a model similar to an exponential decay function. As we have discussed above, cognitive psychology is typically interested at the level of description of microcognition, where scientist generally endeavor to produce simple models of regression analysis to model universal processes pitched at a low level of information processing. The cognitive functions which depends on those processes, such as complex decision-making, exhibit much higher degrees of variance within and between individuals (Tyler, 1947, Carroll, 1993, Gruszka, Matthews, & Szymura, 2010). Complex, possibly nonlinear relationships exhibiting high variance in behavioral measures such as DDM performance are thus likely to require different methods for quantifying, analyzing, and interpreting phenomena which departs from the microcognitive level of description.

This suggests a departure from the type of methodology pursued at the level of cognitive processes, such as the relationship between external stimuli and their perception and sensation in psychophysics. A perceptual limen is a more universal characteristic of human cognition than a macrocognitive process, given its more fundamental information processing function, in reference to Newell's (1990) "bands" of cognition. That is, the minimum discernible difference in intensity used in psychophysics to determine perceptual thresholds for stimuli of various sources are processed at a fundamentally lower "band" of human information processing. Perceptual limen as formulated by Weber (1851), Fechner (1860), and Stevens (1957) occur at the "cognitive" scale (Figure 11) of approximately 100 milliseconds, whereas the DDM tasks involve time scales situated at the "rational" level (minutes to hours, Anderson 1990, 2002, 2007, also called the computational level by Marr, 1982). Learning about systems thinking, i.e., the development of metacognition, occurs at the level of sociotechnical systems (weeks to years).

Insofar as cognitive processes exhibit intra- and inter-individual differences at multiple levels of description in proportionality with the complexity of the information processing strata as illustrated by Newell's levels of analysis, we expect that more fundamental processes such as found in the biological band (neural signaling in the cerebral cortex) and the cognitive band (perception, memory, motor outputs) exhibit less variance in their execution than the cognitive activities described at higher levels of analyses such as the rational and the social-organizational bands (Schoelles, Neth, Myers, & Gray, 2006, Myers, Gluck, Gunzelmann, & Kruskmark, 2010, Gray, 2012). A similar and parsimonious assumption could be construed with regards to the accuracy of information processing responses, beyond the consideration of information processing latencies: the accuracy of information processing responses relative to the levels of analyses of cognition exhibit increasingly higher intra- and inter-individual differences than their lower information processing stages.

NEWELL'S LEVELS OF ANALYSIS

Scale (sec)	Time Units	System	Analysis	World (theory)
1000000000	decades	Technology	Culture	Social & Organizational
100000000	years	System	Development	
10000000	months	Design	Education	
1000000	week	Task		
100000	days	Task	Traditional Task Analysis	Bounded Rationality
10000	hours			
1000	10 min			
100	min	Subtask	Strategies & Procedures	
10	10 sec	Unit task	Procedures & Methods	Cognitive Band (symbolic)
1	1 sec	MicroProcedures	Embodiment Level (1/3 to 3 s)	
0.1	100 ms	Activities	Internal operations of semi-independent modules such as memory, perception, motor	
0.01	10 ms	Atomic Components	Architectural	Biological Band (subsymbolic)
0.001	1 ms	Parameters		

Figure 11 – Newell's levels of analysis in cognitive science (Gray, 2012).

To emphasize the relationship between time scales, the levels of analysis, and the type of experimental approach one might use in tackling cognition processes or functions of various degrees of complexity, consider the problem of choosing a system level (as per Newell's taxonomy) in the pursuit of knowledge about human cognition. Gray (2012) discusses the notion of the 'specious present', inspired by William James' (1893) conception of one's intimate perceptions as they occur over brief periods of time, bearing a sense of what the 'present' is. This time frame of a few seconds is where "immediate interactive behavior" occurs, or as he labels it, the time span of natural interaction with simple artifacts, technologies, and in everyday communications. Gray claims that the specious present is a very different phenomenon from other time spans of human experience, and purports to show that this so-called *exceptionalism* of the specious present is necessary but insufficient in explaining behavior and cognition at higher levels of description, as he favors research at the functional level of cognition, ot the *unit task* level of analysis. This level of analysis is favored by researchers in specialized fields of applied cognitive psychology and multidisciplinary research interested in complex, interactive behavior and cognition, such as human factors and ergonomics (Wickens, Hollands, Banbury, & Parasuraman, 2013), human-computer interaction (Card, Moran, & Newell, 1983), and cognitive systems engineering (Hollnagel & Woods, 1983). The field of cognitive science interested in computational cognitive modeling is particularly interested in the relationships between microcognition and macrocognition, insofar as the creation of a theory of the architecture of the mind rests upon the creation and the validation of a global computational system (literally called a cognitive architecture) capable of producing human-like cognition and behavior, based on assumptions about the low-level cognitive processes above-mentioned (Laird, Newell, & Rosenbloom, 1987, Kieras & Meyer, 1997, Anderson, 2007, Pronovost, 2009, 2010, 2012, Pronovost & West, 2007, 2008, West & Pronovost, 2009).

Anderson (2002) claims that

> Much of cognitive psychology focuses on effects measured in tens of milliseconds while significant educational outcomes take tens of hours to achieve. The task of bridging this gap is analyzed in terms of Newell's (1990) bands of cognition—the biological, cognitive, rational, and social bands. The 10 millisecond effects reside in his biological band while the significant learning outcomes reside in his social band.

Anderson holds three particular views concerning the relationship between microcognitive and macrocognitive aspects of human cognition, namely a thesis concerning **decomposition** (*the claim that learning occurring at the social band can be reduced to learning occurring at lower bands*), another thesis about **relevance** (*the claim that instructional outcomes at the social band can be improved by paying attention to cognition at the lower bands*), and finally, a **modeling** thesis stating that *cognitive modeling provides a basis for bridging between events on the small scale and desired outcomes on the large scale.* He further argues that the unit task situated in the cognitive band is particularly well-suited to test the three above-stated claims, as it is at the boundary of the rational band, and bears a very close relationship to the kinds of experiments which may validate some claims about macrocognition from a microcognitive perspective, and vice-versa. In his words:

> The unit-task level, at the boundary of the cognitive and rational bands, is useful for assessing these theses. There is good evidence for all three theses in efforts that bridge from the unit-task level to educational applications. While there is evidence for the decomposition thesis all the way down to the 10 millisecond level, more work needs to be done to establish the relevance thesis and particularly the modeling thesis at the lower levels.

In summary, complex decision-making, systems thinking, and metacognitive abilities all justify a departure from the traditional experimental framework of cognitive psychology. Brehmer and Dörner (1993) championed experimentation through the use of microworlds to expand the validity of laboratory-based research to natural work environments, a mandatory step in their opinion in generalizing research results to complex sociotechnical systems where DDM is commonplace. The internal validity (the validity of causal inferences in experimental research, Brewer, 2000) and the external validity (the extent to which such causal inferences can be generalized beyond the research) in traditional laboratory based research may be threatened by *conceptual* (e.g., abstracting away from characteristics of the decision-maker, the decision task, and the task environment), *methodological* (e.g., operating on different time scales), and *practical* limitations (e.g., focusing on non-experts), which can mitigated by the conduct of experimental research through complex and dynamic tasks in simulation environments. Such simulations offer a compromise between experimental control and rigor in observing the impact of the complexity of DDM scenarios on human peformance, on the one hand, and insuring the validity of the observed and measured results in terms of the realism of DDM tasks and the generalizability of such results to real-world settings, on the other hand (Gray, 2002, Gonzalez, Vanyukov, & Martin, 2005, Lebiere & Best, 2009). The following section presents in detail the type of simulation-based experimentation platform used in for the elaboration of objective measure of complexity for DDM problems.

1.3 Simulations, microworlds, and serious games

The complexity and dynamic features of a work environment where complex decision-making is used are difficult to replicate and operationalize in a more traditional laboratory environment (Klein et al, 2003). The characteristics of a dynamic task, contrary to the static tasks commonly involved in experimental psychology (such as assessing the outcome of a wager, locating an item on a display, or recalling a sequence of symbols after a presentation), is that dynamic tasks involve multiple decisions made over time, and the outcome of those decisions actually changes the state of the problem at subsequent times for new decisions. We can thus say that a task is dynamic by virtue of (i) involving multiple decisions occurring over time, (ii) where decision choices are a product of the state of a problem and the interaction with a decision maker, and (iii) meaning that the state of the problem is ever-changing, i.e., it is sensible to the decision maker's choices in a way that can lead to outcomes ranging from benign to catastrophic changes in the problem's state (Forrester, 1961, Dörner, 1980, Brehmer, 1990, Funke, 1991, Sterman, 1994). What experimental method is best suited to capture such feature-rich task, on the one hand, and the intricacies of macrocognitive functions, on the other hand? Brehmer and Dörner (1993) referred to this problem as the researcher's dilemma:

> Psychology lives with many tensions. One is that between research in the laboratory and research in the field. Field researchers criticize laboratory research for lack of relevance, or "ecological validity" as it is now often called. Laboratory researchers, on the other hand, criticize field researchers for lack of control with attendant problems in making causal interpretations of their findings. Both are right, of course, especially in the eyes of the applied psychologists who try to use the knowledge produced by psychological research. The root of these problems lies in the inability to handle *complexity*. In field research, there is often too much of it to allow for any more *definite* conclusions, and in laboratory research, there is usually too little complexity to allow for any *interesting* conclusions.

1.3.1 Simulations

Simulationism is the practice of designing models for real or imagined systems, upon which observations and experiments are made (Banks & Carson, 1984). Simulations are a way to gain insights about a system, in order to train to control of it or to learn how to make better decisions based on such knowledge. It is a very old concept, insofar as imitation of real-world systems in order to gain a deeper understanding of their inner workings or their outcomes predates computer simulations (e.g., turn-based strategy games such as chess and other abstract strategy games, such as miniature wargaming).

Simulation-based experimentation in cognitive psychology can model the impact of three fundamental features of dynamic decision-making discussed in the introduction, namely the complexity, dynamics, and uncertainty involved in complex problem solving (Gonzalez, 2005). Various degrees of complexity in the decision task can be manipulated in order to observe the impact on the decision maker's workload and performance. The complex decision problem, understood as a dynamical system composed of stocks and flows, evolves over time due to the decision maker's choices, and based on its own intrinsic dynamics, all of which can be captured in the execution of a simulation, and the perceived time pressure on the decision maker can be measured against the states of the DDM scenarios as they unravel over time. This is done via process tracing, or the collection of multiple data points over the entire simulation. Finally, the opacity, or uncertainty in both the quality and quantity of information available to the decision maker can also be modeled and manipulated through simulation-based experimentation, reflecting the environment-rich nature of decision problems occurring in complex, real-world sociotechnical systems (Funke, 2010).

Gray (2002) has established a flexible taxonomy of simulated task environments with non-exclusive categories encapsulating some features of simulation-based research methods. *High-fidelity simulations of complex systems* are full-fledged simulation environments used to train or test personnel on highly critical tasks in a secure setting, such as an aircraft simulator. *High-fidelity simulations of simple systems* focus on less complex task environments which would involve narrower interaction between the agent and the system, such as a single agent training on a particular piece of technology. *Scaled worlds* are focused on particular configurations of complex task environment, abstracting away from information and tasks extraneous to the functional relationships of interest. Researchers use scaled worlds to generalize some experimental results in a different task, or using a similar task with different parameters.

Synthetic environments and microworlds are theory-driven simulated task environments from which the functional relationships between a cognitive system and the task environment are studied in order to stress the details of how such functional relationships unravel in various tasks. Those synthetic environments are thus an opportunity to exploit regularities among tasks which elicit the functional relationships of interest, and to observe behaviors without wasting time on a wealth of information and details which do not directly pertain to the tasks (and thus constitute lower-fidelity simulations). The fifth and final category is the *simple laboratory environment*, included as a baseline method of research in experimental psychology, with the lowest fidelity (and external validity for research at the macrocognitive level).

Those simulated task environments differ across the three dimensions of *tractability* (whether the simulated task is easily controlled and manageable, or the experimental results are hardly tractable), *correspondence* (whether the simulation models multiple aspects of one system, or one aspect of multiple systems), and *engagement* (whether the simulation is an immersive and involving experience or a rather technical and demotivating experience). Those three dimensions are characteristics affecting the researcher's perspective, the tasks itself, and the participant's experience, respectively. We will focus on microworlds in particular, as they are the simulated task environment of choice for experimental research on dynamic decision-making (Sterman, 1989, Brehmer & Dörner, 1993, Gonzalez, Vanyukov, & Martin, 2005), with desirable properties such as the tractability of a controlled environment, generalizability of a few functional relationships (i.e., the investigation of cognitive functions involved) over multiple tasks, and a game-like engagement for participants (Gray, 2002).

1.3.2 Microworlds

A microworld offers a robust compromise in balancing internal validity and ecological validity in that it is used to create scenarios reflecting the complex and dynamic properties of human cognition by the simulation of information-rich task environments which afford realistic actions and produces plausible outcomes. Using a microworld also favors learning and decision support, insofar as the interaction with a simulated task environment mirroring real-world settings may be repeated and rehearsed for an indefinite number of times, with or without minute variations, in a way to explore the problem space of complex decision tasks. The possibility of running "what-if" scenarios over and over by interacting with a microworld is a considerable advantage of simulation-based experimentation. It far outweighs the trade-off of the benefits vs. the limitations of both the traditional, laboratory-based experimentation, on the one hand, and of the natural, field-based observation studies, on the other hand (Rassmussen, Pejtersen, & Goodstein, 1994).

A microworld recreates the functional relations of a task, providing a high degree of external realism compared to laboratory studies, yet retaining a level of control over the manipulation and measurement of task variables that is not possible in the field (Brehmer & Dörner, 1993; Gonzalez, Vanyukov & Martin, 2005; Gray, 2002). Confounding and extraneous variables may be isolated and controlled, like in the experimental designs interested in microcognition. The compression of time and space for a task in a microworld suits the complex nature not only of the task environment at hand, but mostly to elicit the complexity of the low- to high-levels of cognitive processes and functions involved in seeking information about a decision problem, assessing the available courses of action, and evaluating the outcome of choices made by the agent. Those computer-based simulations may be geared towards pedagogy and training on particular skills, such as in an interactive learning environment, or ILE, (Bakken, 1989, Issacs & Senge, 1994, Karakul & Qudrat-Ullah, 2008).

The microworld methodology is therefore well-suited for DDM research, as microworlds are environments characterized by the need for people to make multiple, interdependent, real-time decisions in reaction to both external changes and to the effects of their past decisions (Brehmer, 1992). The semantically-rich microworlds create an immersive experience, and share certain key characteristics with their real-world counterparts insofar as they involved complex problem solving in that they are dynamic (evolving autonomously in real time), complex (involve interacting components and conflicting goals), and opaque (relationships between variables must be inferred). Microworlds offer the same kinds of experimental control and measurements available to the laboratory environment, with a potential for scalability. Process and performance measures are captured in real-time experimental conditions involving different types of individual and intra-organizational interactions in cognitively demanding tasks. Microworlds are thus a perfect match for DDM research, as they favor task repeatability, rehearsal for metacognitive skills such as systems thinking, learning involved in the development of expert performance such as managerial skills, constraints satisfaction and optimization abilities, etc. The microworld fosters the exploration of *what-if* scenarios in a non-threatening exercise environment. Feedback is immediate, and can be used to learn from one's choices (Diehl & Sterman, 1993).

1.3.3 Serious games

Microworlds are particularly appropriate for educational purposes, having emerged from considerations about pedagogical constructivism (the idea that human learning is most effective through making things, "learning by doing", also known as experiential learning, Piaget, 1957, Clements, 1989, Ackermann, Wilson, 1995, Rieber, 1996). As a small-scale representation of real-world interactions, a microworld is a simulation task environment where a learner may explore alternatives, test hypotheses, and discover novel facts about the problems being simulated. Educational microworlds have been around for decades, such as Papert's *Logo* (1980), Forrester's *Stella* (1989), and White's *ThinkerTools* (1993). Another feature of microworlds that is important to control is participant engagement, as seen in Gray's taxonomy above (Gray, 2002), in order to keep participants motivated in the complex tasks at hand.

A particular type of microworld geared towards both learning and entertainment is the category of simulations named *serious games* (Sterman, 1989). While serious games are microworlds geared towards education, they are designed to challenge and entertain the decision maker in order to immerse him or her, so as to keep a higher level of motivation throughout the challenging (and sometimes stressful or frustrating)simulation experience. Such microworlds emphasize an added pedagogical value of entertainment through rewards, incremental challenges, and competition, providing a richer engagement and a more immersive experience. Exemples of microworlds with a ludic flavor specifically geared towards the training and the evaluation of performance for dynamic decision-making are Sterman's *Beer Game* (1989), Omodei and Wearing's *Fire Chief* (1995), *Ecopolicy* ™ (sold by *Ravensburger*, 1985, 1997, see Vester, 1999), Gonzalez, Lerch, and Lebiere's *Water Production Plant* (2003, also Gonzalez, Thomas, & Vanyukov, 2005), and *Democracy* ™ (sold by *Positech Games*, 2005, 2007, 2013). The following section presents the two simulated tasks environments we have used for the current research project, which share the aforementioned characteristics of synthetic environments, microworlds, and serious games.

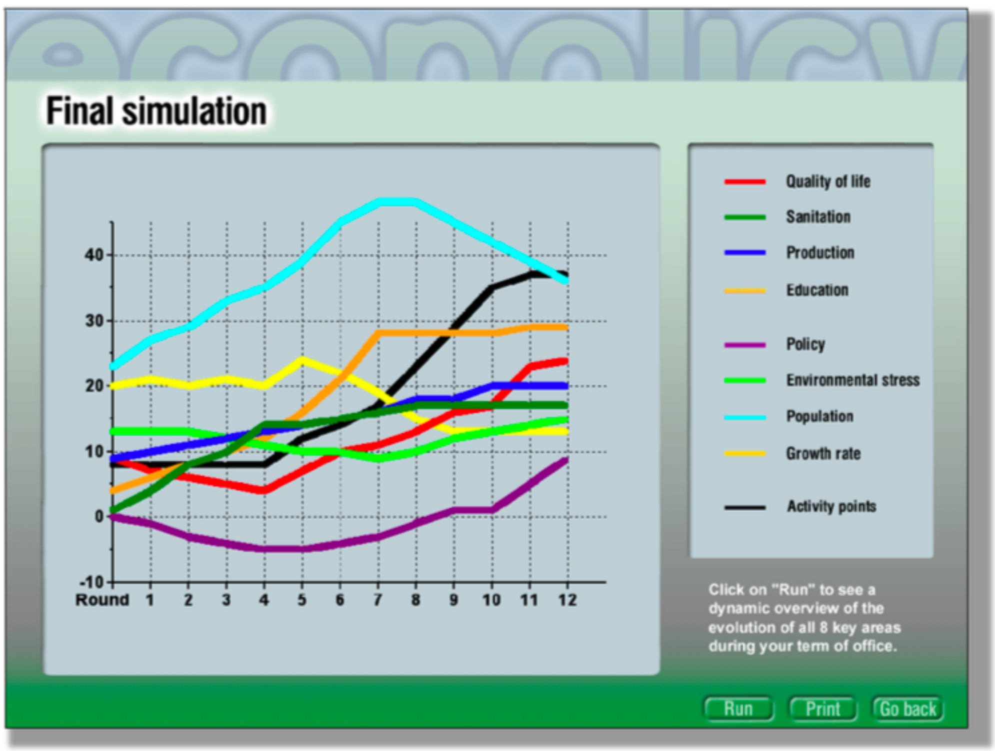

Figure 12 – A dynamic overview of a DDM scenario in the Ecopolicy serious game.

1.3.4 The CODEM simulation environment

The main DDM scenarios used for experimentation were designed with the **Co**mplex **De**cision **M**aking Experimental Platform (CODEM) developed by Defence R&D Canada, and run on standard personal computers. CODEM is a simulator and testbed used to study the impact of interactive learning environments on the comprehension of individuals involved in complex decisions problems. CODEM contains a scenario editor allowing the complete oversight and total control on scenario design down to every single DDM entity, parameter, and relationship, allowing the creation of experimental content through a GUI. It can then be deployed as scenarios through which participants experience dynamic decision-making problems, and finally, CODEM generates a plethora of raw data outputs from which statistical analyses may be undertaken.

The scenarios are contextualized implementations of dynamic decision-making problems, where a participant may observe and influence the changes in variable values of key parameters in a complex system (Figure 13). Some of the variables must be minimized or maximized, or attain a particular threshold value, in order to achieve success (or merely avoid catastrophic failure). Figure 14 presents a snapshot of the CODEM scenario editor, where the scenarios are designed as mathematical abstractions of DDM problems following a rule-based, discrete-event system. In the scenario editor one can create the DDM variables and the relationships between them, the interventions from participants used to influence the system's state, scenario events to surprise and alter the flow of the system's dynamics, etc.

CODEM is a turn-based computer simulation, and require that participants expend some "action points" in order to effect some changes in the system variables. The more complex scenarios used in the present experiment, the third version of the Arctic development scenario (presented below) and the two versions of COIN (which stands for Counter-Insurgency), were developed with the support of subject matter experts in domains such as defence and aerospace, and governance and policy-making.

Those DDM scenarios are therefore considered to be realistic complex problem solving tasks exhibiting a high degree of external validity, with meaningful entities and relationships for expert decision-makers in complex sociotechnical systems. This stands in contrast with contrived scenarios of a more artificial nature, such as in the case of the Tribes and SpaceLab scenarios, designed with only playfulness in mind. The latter two scenarios are only used for the familiarization of participants with the user interface of CODEM, and to give them a quick tutorial concerning the essential concepts involved in dynamic decision-making and systems thinking.

The following sections present brief summaries for the dynamic decision-making scenarios, as well as a visual representation of their structural complexity in the form of decision graphs. Another DDM scenario is used to provide data in a regression model validation phase, as presented at the end of chapter 2. This scenario, called *Cybernetia*, was used in the Ecopolicy ™ serious game environment, developed by MCB Publishing House. Ecopolicy is advertised as an 'edutaintment' product, meaning that it aims to provide an entertaining experience while allowing the user to educate itself in the process of playing.

The game can be understood as an abstract model of policy-making and resources management, and the level of realism and detail of that model supports a user experience which carries a certain degree of reusable and repurposeable knowledge in the context of real-world problems. Short summaries for each of the DDM scenarios used in the various experimental conditions are presented in the next chapter, while the next section presents a complete example of complex decision-making using the Arctic 3 DDM scenario. Appendix A contains a detailed breakdown of all the DDM scenarios used in this document.

CODEM and Ecopolicy are discrete-event simulation environments which can be executed either as completely or partially deterministic simulations, i.e., they can include some randomness if so desired in order to observe the impact of stochasticity on scenario performance. The feature of random events has been disabled due to our interest of observing the relationships between the parameters of complexity and human performance in DDM problems, affording a more controlled design of the variance in both experimental and response variables.

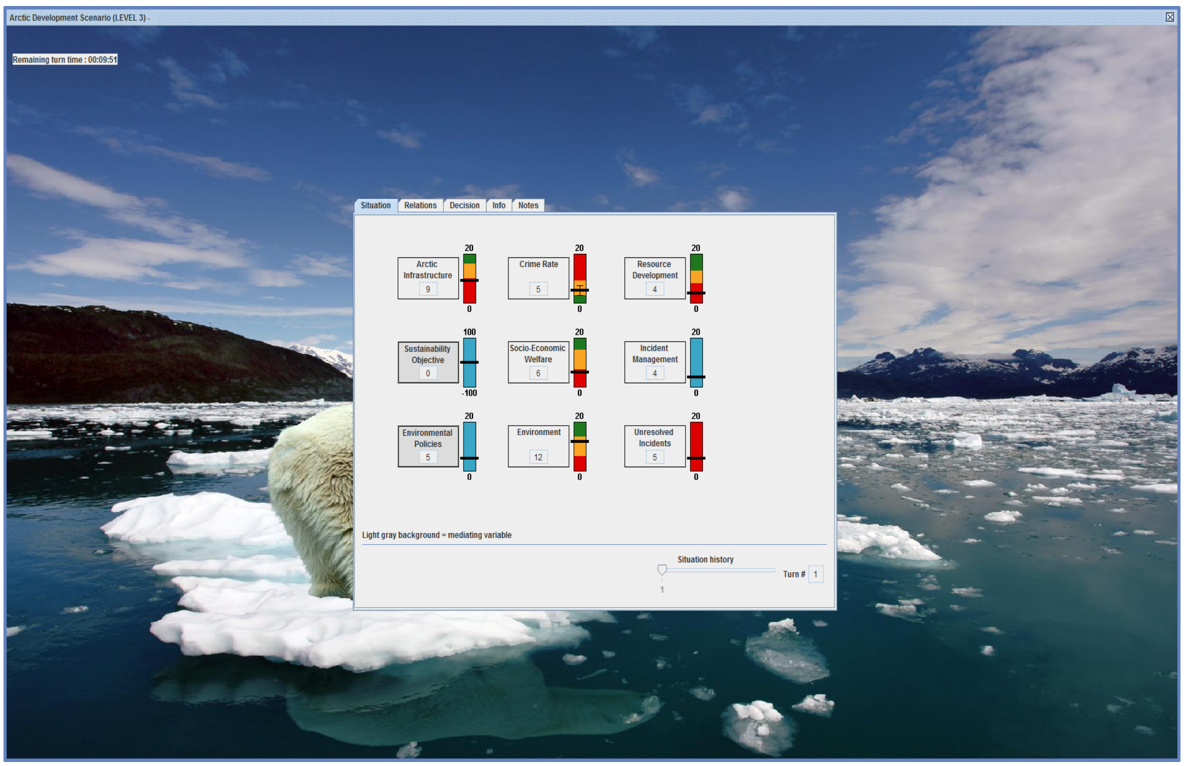

Figure 13 – The CODEM in-game interface for the Arctic 3 DDM scenario.

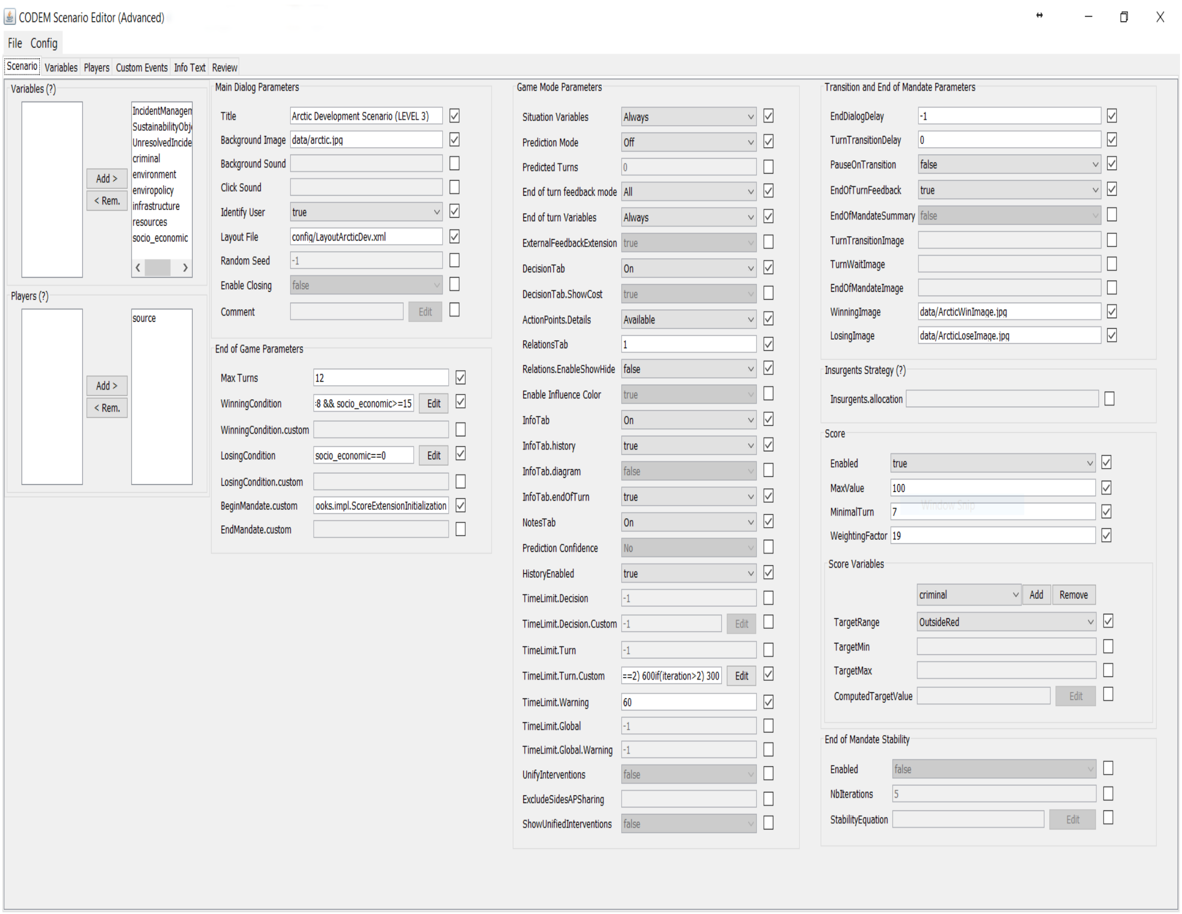

Figure 14 – The CODEM scenario editor interface.

1.3.5 A complete example: the Arctic 3 DDM scenario in CODEM

No amount of structured text can convey the experience of playing through a dynamic decision-making scenario in a simulation-based experiment. In the absence of more appropriate media to capture and illustrate the process of interacting with a serious game to evaluate human performance in DDM, we will rely on the moderately more appropriate medium of using pictures of the microworlds as they appear in the CODEM simulation environment. This section presents a summary of the multiple informations sources and available interactions during the simulation of the Arctic 3 scenario in the CODEM environment. Also included are a breakdown of the scenario variables, relationships between variables, the interventions and influences (outcomes) from such interventions, as well as miscellaneous information relative to the initial and end conditions of the scenario. Data concerning the other DDM scenarios discussed in this document are available in appendix A.

1.3.5.1 Playing through a dynamic decision-making scenario as a serious game

The Arctic development scenario is a fictional microworld set in the Arctic Archipelago in a hypothetical near-future geo-political context (Figure 15). The player is a leader of a strategic advisory group who has a mandate to improve the socio-economic status of his or her state by establishing a presence in the area and achieving the sustainability of their operations in the long term (Figure 16). Table 3 presents the 8 variables of interest for the scenario, plus a ninth variable labeled *sustainability objective*, which only stands as a gauge for the progression of a player's efforts (Figure 17). Also included are the variables values at the onset of a simulation, as well as the target values for success (which must *all* be met) and the conditions for failure. Table 4 presents the relationships between the aforementioned variables (also seen in 18), with the array of values from one variable bearing an impact on the values of target variables. The linear, partially linear, polynomial, and otherwise nonlinear relationships for such variables are unfortunately not presented herein, since it would requires a considerable amount of space for the purpose of illustration (see Figure 19). Appendix B nevertheless presents the calculations for the objective measure of nonlinearity, or L_{DDM}, which represents the average of the regression slopes for all relations in a scenario.

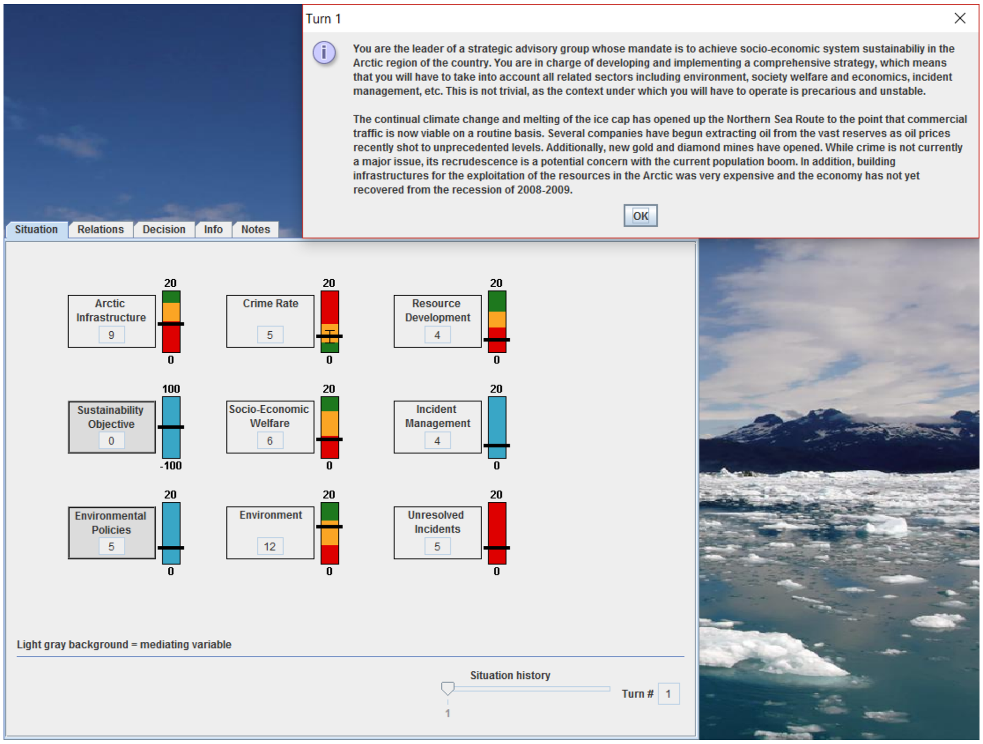

Figure 15 – Introduction of the Arctic 3 scenario in CODEM, presenting the context.

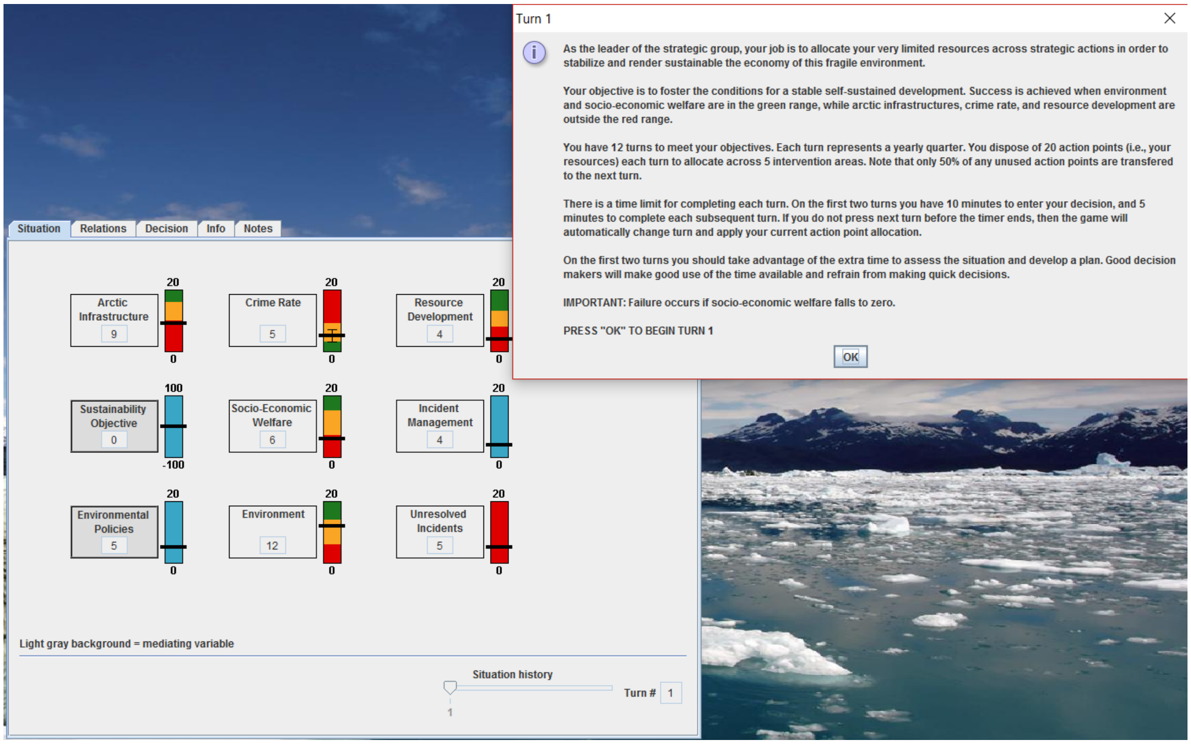

Figure 16 – Introduction of the Arctic 3 scenario in CODEM, presenting the goals.

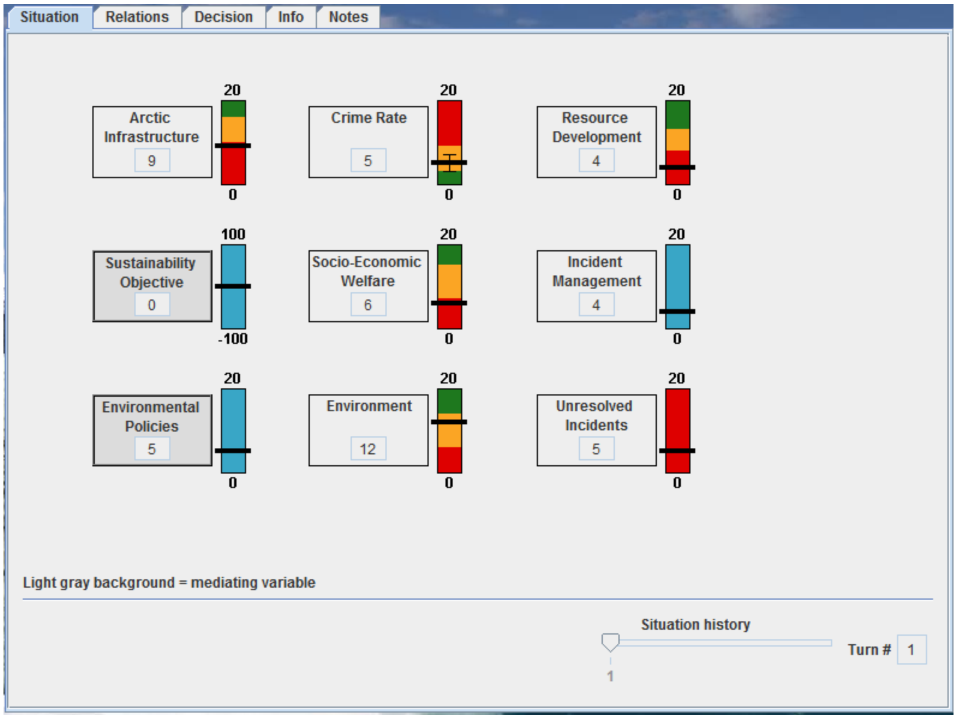

Figure 17 – The *situation* tab of the Arctic 3 scenario in CODEM, presenting the variables.

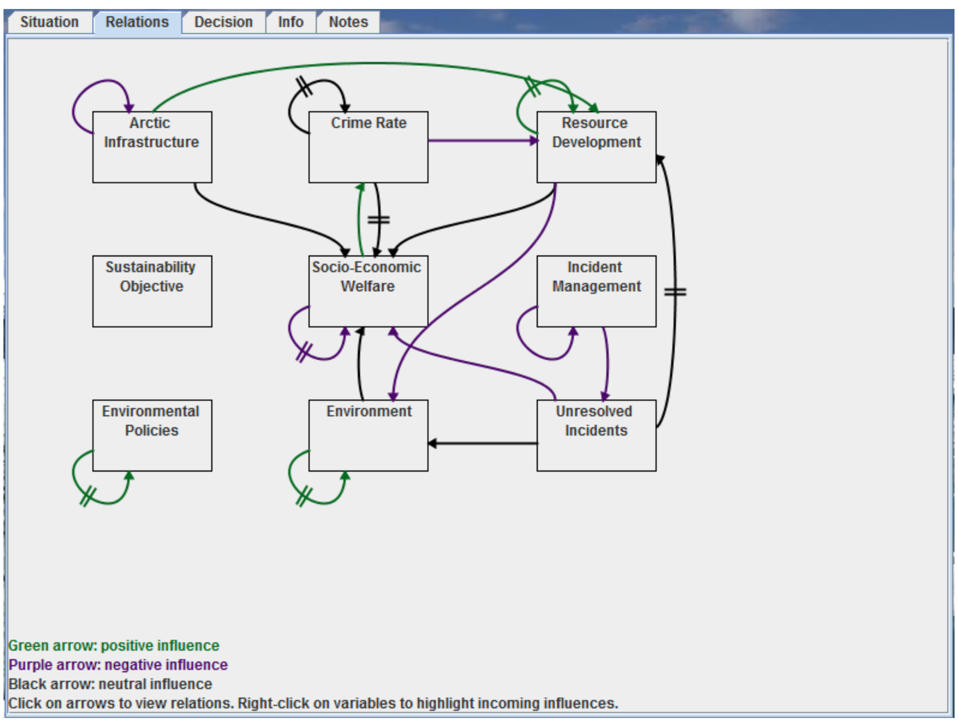

Figure 18 – The *relations* tab of Arctic 3, presenting the relationships between variables.

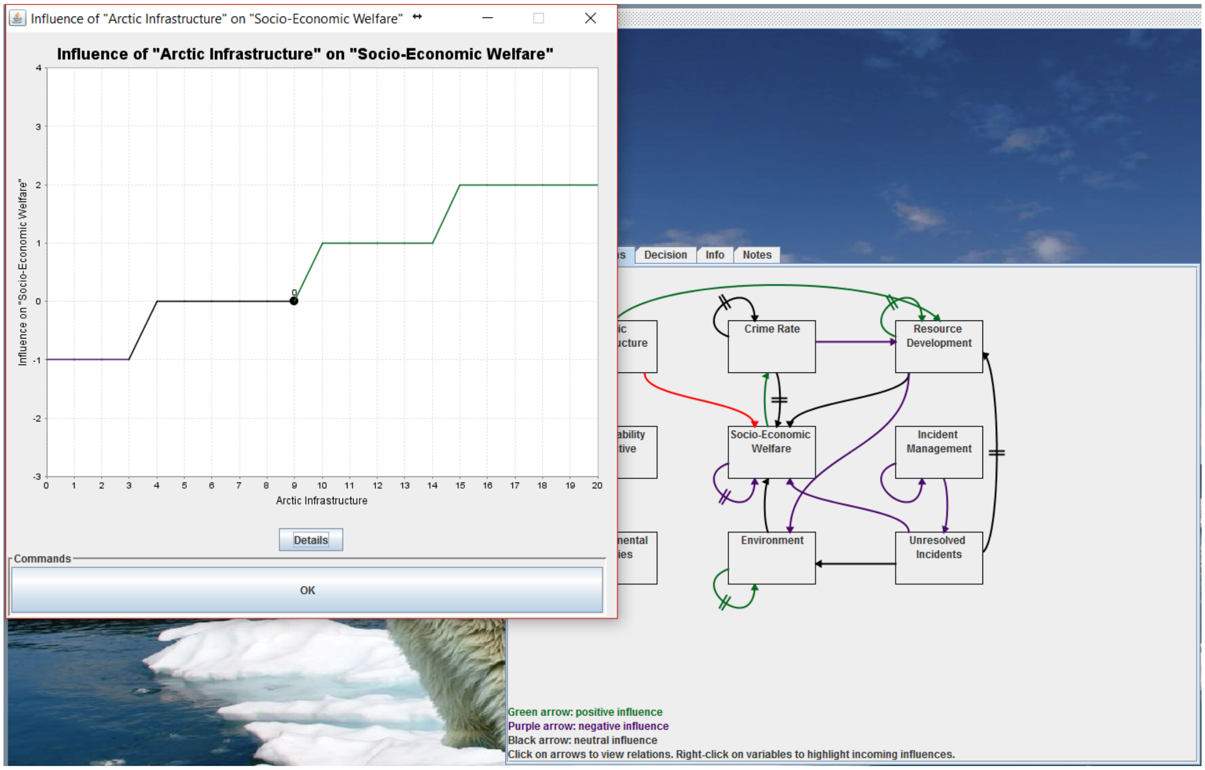

Figure 19 – The details of a relationship between two variables in the Arctic 3 scenario.

The player has up to 12 turns, which may be interpreted as arbitrarily representing a month or a yearly quarter, in order to reach the winning conditions. A number of action points, representing resources such as financial capital, human capital, etc. can be allocated at each turn of play, starting with 20 action points.

The player has 5 interventions through which to use such actions points (Figure 20), as presented in Table 5. Such interventions yield a varying number of impacts on the system's variables, with possible delays, differential impacts based on interactions with other variables (referred to as "conditional influences"), and even some contributions to the pool of action points (though it is not the case in the Arctic scenarios).

The effect of interventions on other variables may be positive or negative, or a combination of both, so a player is advised to weigh all such possibilities prior to committing to a particular allocation of action points (Figures 21 and 22). Once a player has committed to a decision, the DDM scenario is updated to new variable state values, as the simulation advances to another turn of play (compare Figures 17 and 23).

The participant is then presented with an interactive feedback window displaying the various effects from interventions, from which he or she can consult individual effects (Figure 24). Scripted scenario events and delayed effects are also displayed (Figure 25), adding to the complexity of both the intelligibility of, and the influence of a player on, the underlying dynamic system of the DDM scenario.

Players may consult the information tab of the CODEM interface in order to consult general information concerning the scenario goals and the events which have occurred (the messages sub-tab, Figure 26), keep track of variable states and previous action point allocation (the history sub-tab, Figure 27), and revisit the effects of past decisions on future system states (the feedback sub-tab, Figure 28). The player can also take notes directly into the CODEM interface in order to organize his or her ideas about the DDM problem (the notes tab, seen in Figure 29).

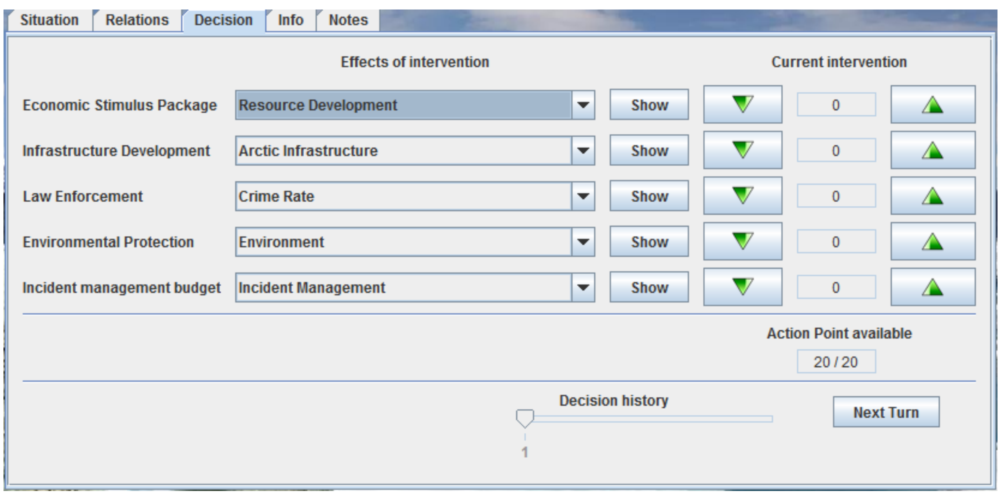

Figure 20 – The *decision* tab of the Arctic 3 scenario, presenting the available interventions.

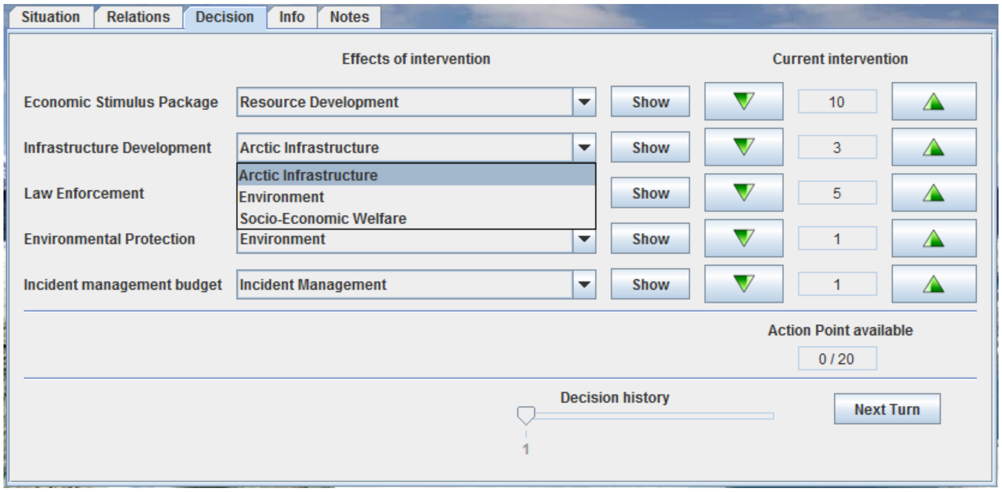

Figure 21 – The *decision* tab of the Arctic 3 scenario, where interventions have multiple influences.

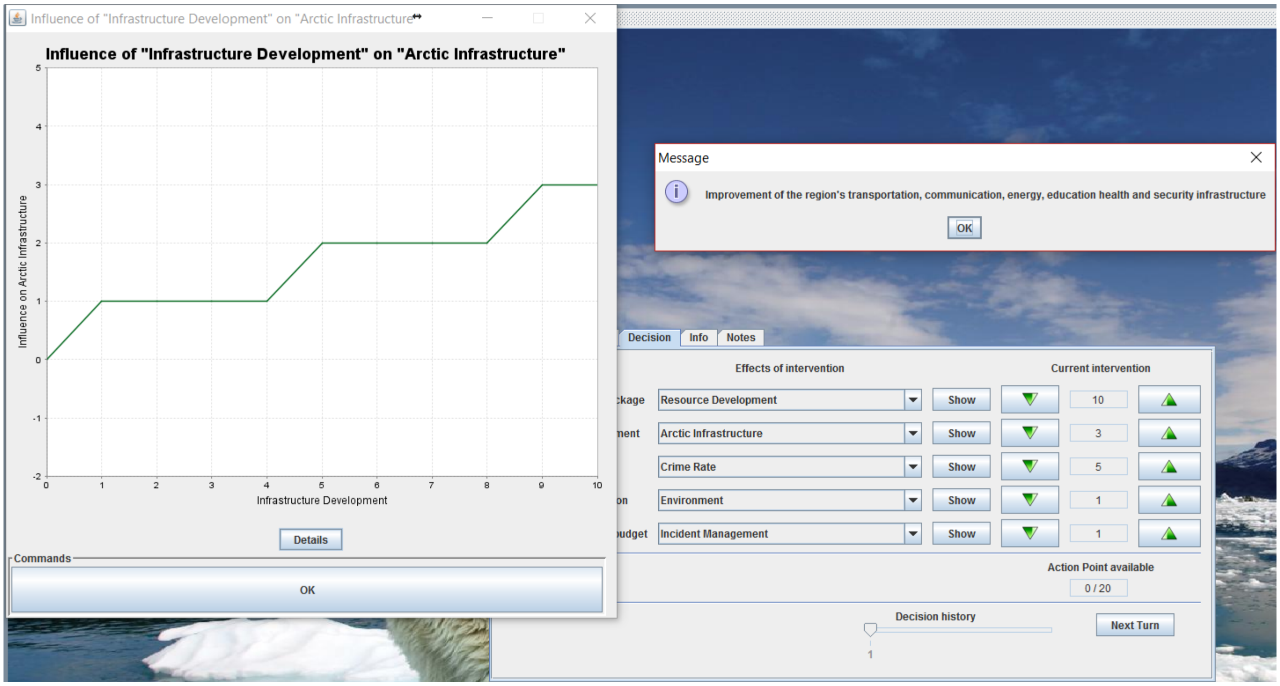

Figure 22 – The details of the effects of an intervention on a variable in the Arctic 3 scenario.

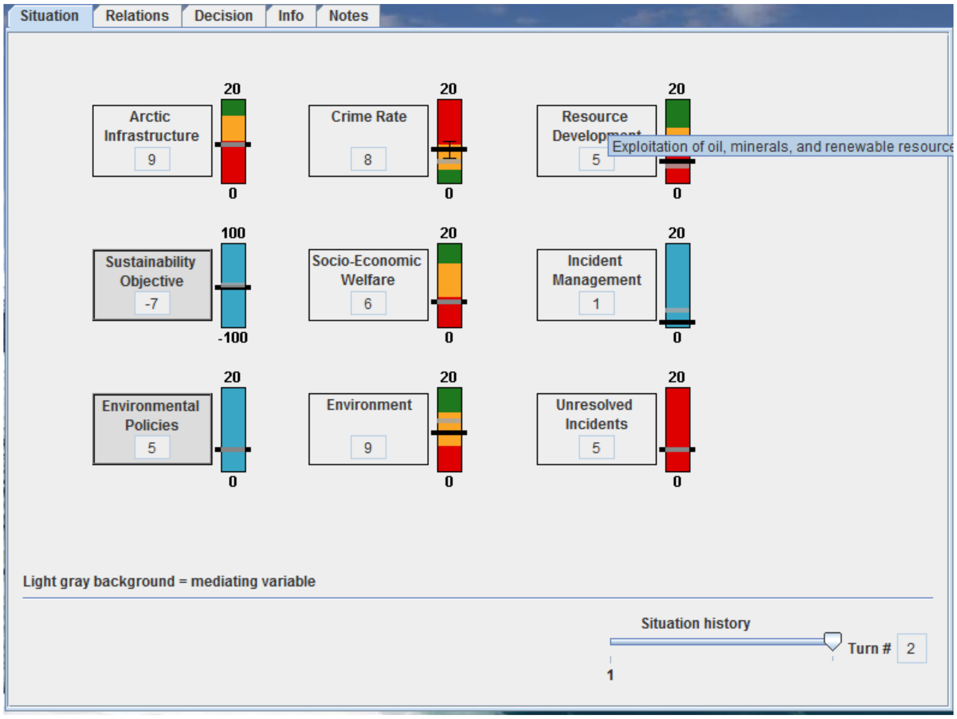

Figure 23 – The *situation* tab of Arctic 3 in the aftermath of a completed turn of play.

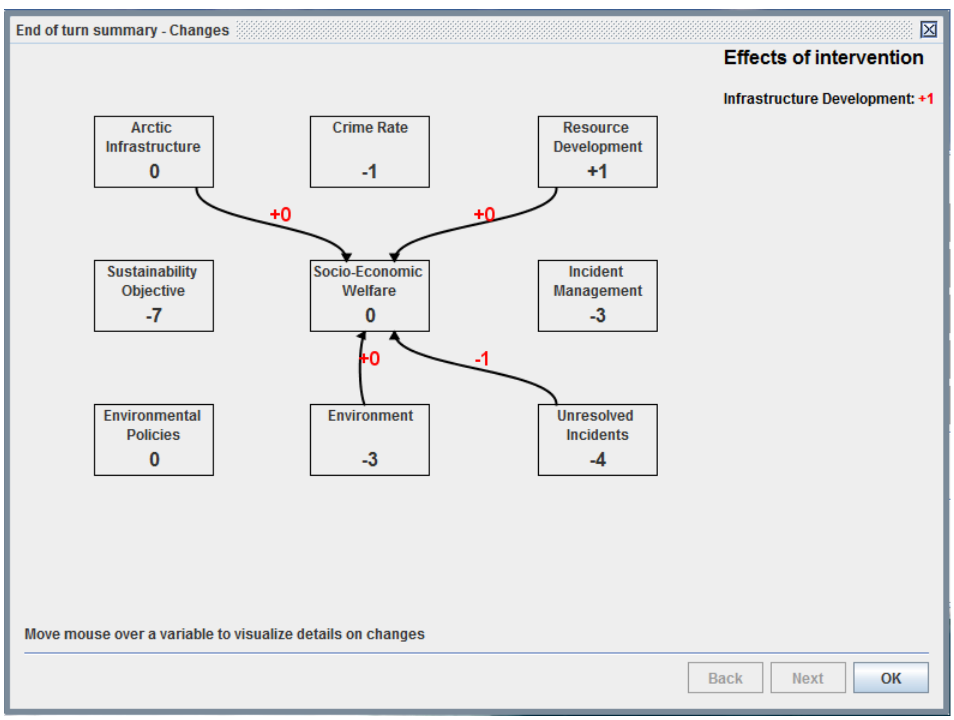

Figure 24 – The *feedback* window detailing the effects of a particular intervention in Arctic 3.

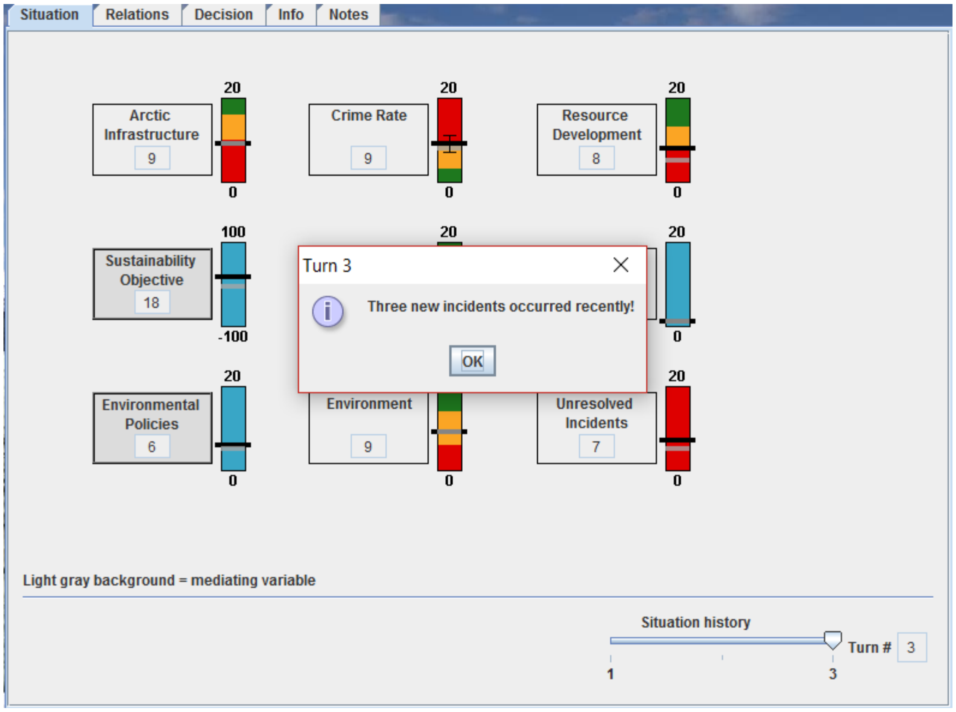

Figure 25 – An events message box during the Arctic 3 scenario.

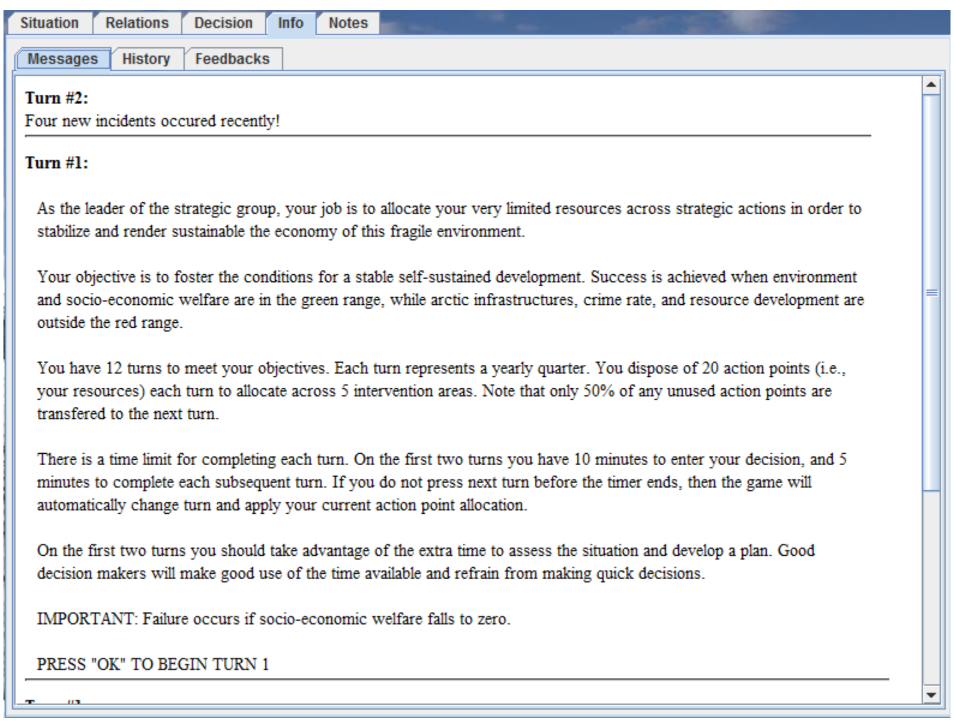

Figure 26 – The messages section of the *information* tab in Arctic 3.

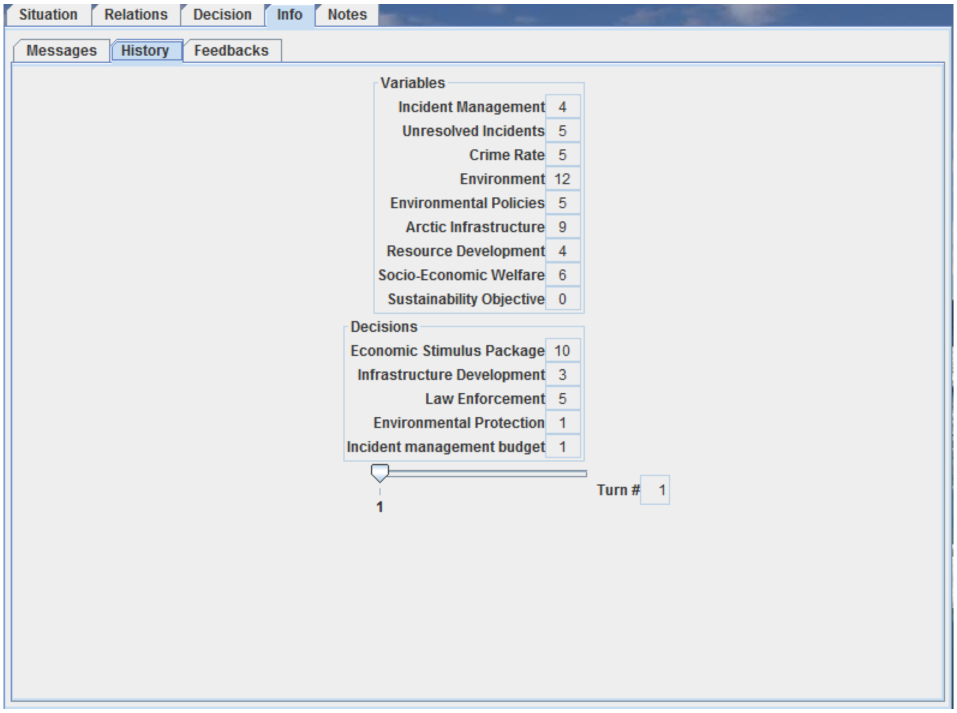

Figure 27 – The history section of the *information* tab in Arctic 3.

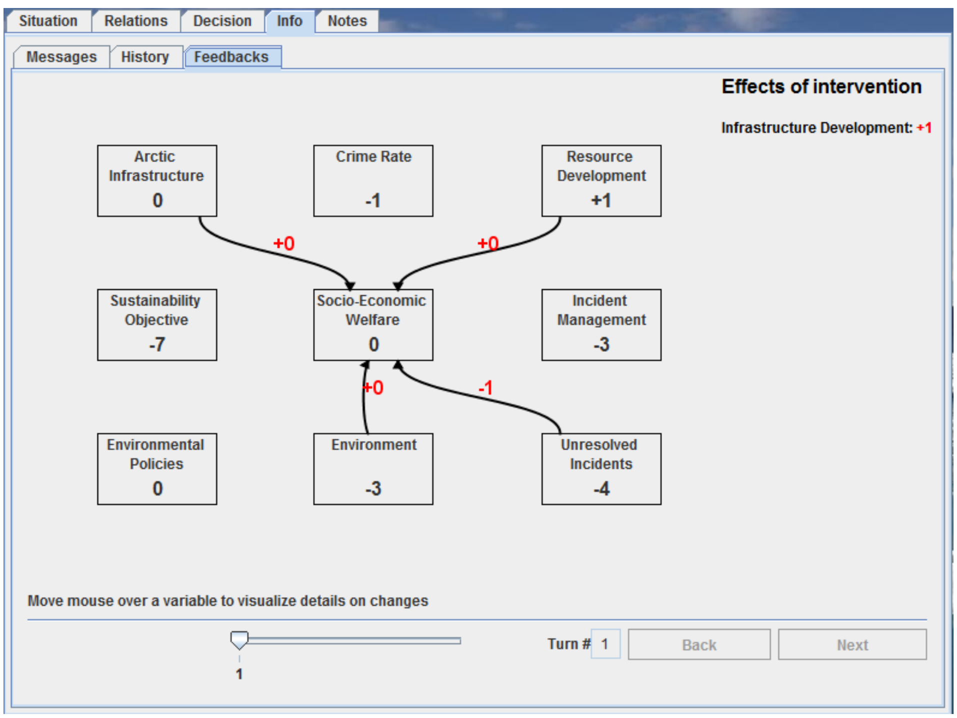

Figure 28 – The feedback section of the *information* tab in Arctic 3.

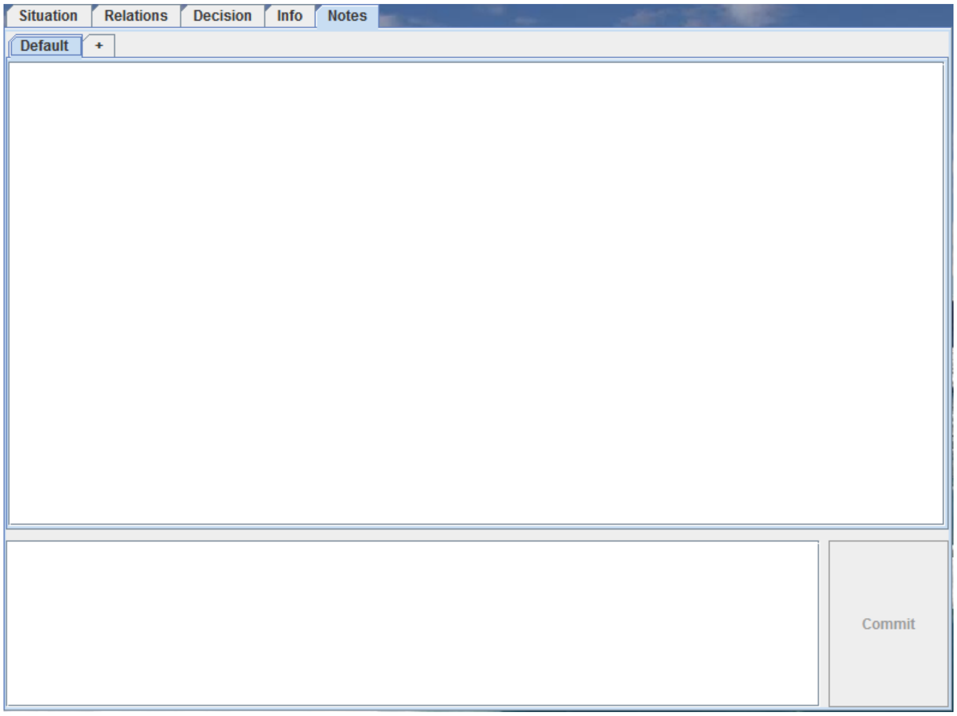

Figure 29 – The *notes* tab in Arctic 3.

The Arctic 3 scenario ends with either success or failure as soon as the winning or loosing conditions are met. Since the success outcome is dependent on *all* of the variable states having reached a certain threshold, the conjunctive nature of the winning condition makes it only possible to win if the 5 variables of interest have been achieved. The opposite would be true in most DDM scenarios with regards to failure, i.e., the loosing conditions are usually disjunctive, as the failure condition is reached as soon as one of the variable states is reached. In the context of Arctic 3, there is only one variable state which qualifies for a failure condition, the *socio-economic welfare*. It should be noted that the net balance of many system variables are either desirable or undesirable, so it could be said that the state of the socio-economic welfare variable is highly dependent on other variable values, given the amount of interconnectivity in the system. Whether the player is successful or not, a vignette displaying a message at the end of the simulation displays a performance score as a percentage, ranging from 0 to 100 (an example of which can be seen in Figure 30). The performance score is a gradient score, and participants can thus aim to improve their own performances by benchmarking their own various degree of successes and failures, whether or not the simulation ended on a positive or a negative note.

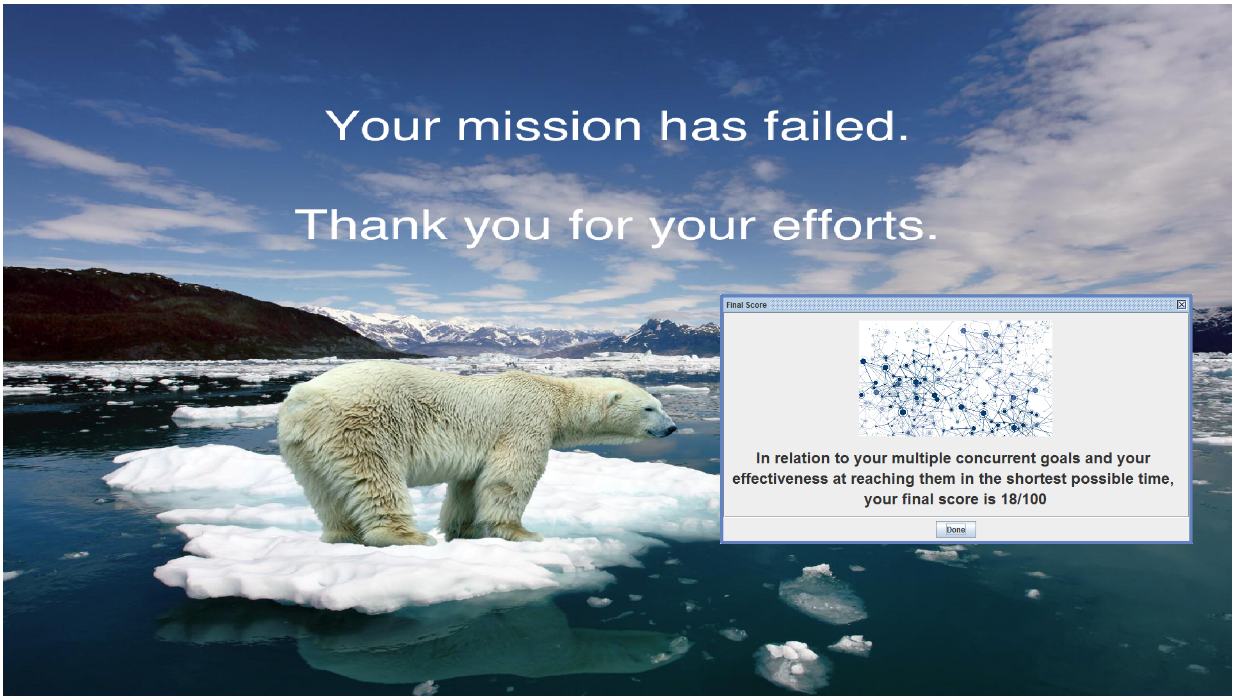

Figure 30 – Ending the Arctic 3 scenario with a failure.

1.3.5.2 Variables

The variables in a DDM scenario designed through the CODEM simulation environment may be endogenous and/or exogenous, they can be hidden from the player or even fuzzy (ambiguous to a given number of units). Variables are either colored in a red-yellow-green scheme, indicating the intervals of values which are undesirable, approaching an undesirable state, or favorable, or they can be uniformly blue, indicating a neutral variable (such as a mediating variable or a simple gauge for scoring) covering an arbitrary range (such variables should be scrutinized through the interface as their role is not particularly straightforward).

Table 3 – Variables in Arctic 3.

variable	**details**	**initial value**	**winning value**	**failure value**
incident management	no color (mediating)	4		
sustainability objective	no color (mediating)	0		
unresolved incidents		5		
crime rate	uncertainty $\pm$ 2	7	<= 9	
environment		12	>= 14	
environmental policies	no color (mediating)	5		
Arctic infrastructure		9	>= 10	
resource development		4	>= 13	
socio-economic welfare		6	>= 15	0

1.3.5.3 Relations between variables

The relationships between variables in a DDM scenario designed through the CODEM simulation environment may be positive or negative, and they can also be hidden from the player. Some relationships exhibit two- or three-fold conditional statements which alter the shape and values of the functional relationship between variables based on the state of one or more other variables in the system. Variables can point back to themselves in an *eigendynamic* relationship, such as the *environmental policies* relationship in Arctic 3, as seen in Figure 19. Furthermore, relationships may yield effects at a later time, occurring after a delay of one or more turns, as seen in the case of the *crime rate* variable in the same Figure. Relationships may have various functional forms, such as purely linear relationships, or they may be simple yet nonlinear (monotonic, logarithmic), higher-order (polynomial or exponential) relationships, and they can have effects with limits or threshold values.

Table 4 – Relations between variables in Arctic 3.

source	values	target	values	details
incident management	0 to 20	incident management	0 to -20	
		unresolved incidents	0 to -20	
unresolved incidents	0 to 20	resources development	1 to -6	delay 1
		environment	0 to -4	
		socio-economic welfare	0 to -6	
crime rate	0 to 20	crime rate	0 to -2	delay 1
		socio-economic welfare	0 to -3	delay 1
		resources development	0 to -3	
environment	0 to 20	environment	-3 to 2	delay 1
		socio-economic welfare	-3 to 1	
		environmental policies	2 to -2	hidden
environmental policies	0 to 20	environmental policies	-1 to -1	conditional env. policies > 10), delay 1
		environmental policies	0 to 0	conditional env. policies = 10), delay 1
		environmental policies	1 to 1	conditional env. policies < 10), delay 1
Arctic infrastructure	0 to 20	socio-economic welfare	-1 to 2	
		resources development	0 to 2	
		Arctic infrastructure	0 to -3	
resource development	0 to 20	socio-economic welfare	0 to 2	
		environment	0 to -4	conditional env. policies <= 6)
		environment	0 to -2	conditional env. policies > 6 and <= 13)
		environment	0 to 3	conditional env. policies > 13)
		resources development	0 to 3	conditional env. policies < 11, delay 1)
		resources development	0 to 2	conditional env. policies > 10, delay 1
socio-economic welfare	0 to 20	socio-economic welfare	-1 to 1	delay 1
		crime rate	3 to 2	

1.3.5.4 Interventions and influences from interventions

Action points obtained at the beginning of each turn can be used towards any or all of the 5 interventions available in Arctic 3, which in turn yield influences on the scenario variables, up to a number of 11 influences. A number of secondary effects from interventions (listed in drop-down menus such as in Figure 21) may not be desired, yet co-occur with a more desirable impact, so the balance of all influences should be carefully assessed prior to committing to a course of action. Some of the effects of an intervention can also be concealed from the player, adding to the difficulty of keeping a cogent mental model of the underlying system dynamics of the DDM scenario. The Arctic scenarios do not feature contributions of interventions towards additional action points, something that occurs only in the COIN scenarios. Unused action points may be transfered to the next turn of play, up to a certain threshold (half of the unused points are transferable in the context of Arctic 3).

Table 5 – Interventions and influences from interventions in Arctic 3.

source	**values**	**target**	**values**	**details**
law enforcement	0 to 10	resources development	-1 to 2	
		environment	0 to -2	
		crime rate	1 to -4	
incident management budget (resupply)	0 to 10	incident management	0 to 10	
infrastructure development	0 to 10	environment	0 to -3	
		socio economic welfare	0 to 4	
		Arctic infrastructure	0 to 3	
environmental protection (request support)	0 to 10	environmental policies	0 to 3	
		environment	0 to 3	delay 1
economic stimulus package	0 to 10	resources development	-1 to 3	delay 1
		socio economic welfare	0 to 4	delay 1

1.3.5.5 Miscellaneous

A simple list of essential features to know about the Arctic 3 scenario in CODEM. The other DDM scenarios are described in the same way in appendix A. Any other parameter of the DDM scenarios which may appear in the calculations of the objective measures of complexity is featured in appendix B.

action points begin with 20
carryover half of the unused actions points to the next turn

total number of turns maximum of 12 turns

turns required to win minimum of 6 turns

scenario events 13 events

1.4 Objective parameters of complexity

We have so far discussed the methodological aspects of conducting research on dynamic decision-making using simulation-based experimentation, using microworlds in the form of serious games to elicit complex decision-making performances through playing in DDM scenarios. But once we have implemented stock and flow models in DDM scenarios, and collected data regarding human performance in those scenarios, how do we assess the relationship between a DDM scenario's complexity and its impact of the participants' performance? This section presents the complexity metrics derived from a literature review on complexity theory, the sources of which have been discussed in the introduction. The equations used for the calculations of those complexity metrics are presented herein, whereas the detailed breakdown of the values of those complexity metrics for each DDM scenario used in the next two chapters are presented in appendix B.

Previous models of complexity for DDM involved the quantification of structure, information, and cognitive load in a compound fashion (as seen in Table 1 in the introduction), which yielded modest results in order to tackle the variance explained in the response variable. The compound models were decomposed into smaller sets of parameters, following the suggestions of Liu and Li (2011, 2012, 2014), Pronovost et al (2014), and Stouten and Größler (2017). The intuitive idea is that separating the composite models into a subset of maximally orthogonal parameters would likely avoid strong associations between the independent variables, or multicollinearity, thereby also maximizing the association between the individual objective parameters of complexity with the response variable. From the compound models presented earlier were retained the commonly observed notions of structure and information, as well as the concept of cognitive weight from the field of cognitive computing.

A particular attention is given to a variety of ways of calculating the structural complexity of decision problems, as the structural and systemic complexity is a favorable way to frame the complexity of dynamic decision-making problems. Another non-trivial issue is the means to calculate the information complexity of such problems, as discussed in Pronovost et al (2014), whereby the inclusion or exclusion of action points, or then 'action potential' afforded by the design of a particular DDM scenario may significantly and substantially influence the human performance results. Finally, an additional set of four parameters, namely difficulty, nonlinearity, uncertainty, and instability, was combined to the above. A summary table of the objective measures of complexity is presented in Table 6, detailing the formal type and associated quantification for each parameter, as well as a short description of their purpose and their means of computation. Structure, information, and cognitive weight are described as magnitudes, that is, they are unbounded scalars with no ceiling value.

Table 6 – Summary of the objective measures of complexity.

parameter	type	quantification	what it is	how it's calculated
structure CC_{DDM} (Cyclomatic complexity)	magnitude	scalar	arcs and nodes in graph representation	difference between arcs and nodes
structure CNC_{DDM} (Coefficient of Network Complexity)	magnitude	scalar	arcs and nodes in graph representation	ratio of squared arcs to nodes
structure SDC_{DDM} (System Dynamics Complexity)	magnitude	scalar	arcs and nodes in graph representation, incl. functional role of elements	product of endogenous variables with exogenous variables and relations
information complexity I_{DDM}	magnitude	scalar	inputs, interventions, and outcomes	product of inputs and outputs
action complexity A_{DDM}	magnitude	scalar	inputs, interventions, and outcomes	product of action points, inputs, and outputs
cognitive weight W_{DDM}	magnitude	scalar	workload (time) to recognize operations	sum of products of basic control structures
difficulty D_{DDM}	relationship	ratio	distance to goals (variable values)	root-mean-square of distances to win and fail
nonlinearity L_{DDM}	relationship	ratio	distance to linearity for relationships	average of regression slopes for all relations
uncertainty U_{DDM}	relationship & magnitude	ratio incl. scalar	hidden and fuzzy variables, relations	averaged hidden / fuzzy elements + events
instability S_{DDM}	relationship	ratio	distance to end state under idleness	rms of dist. to end values (incl. # of turns to fail)

In principle, there could be an indefinite number of variables, and relationships between variables, for a DDM scenario. Difficulty, nonlinearity, and instability express relationships, whereby a number of elements from which the first three parameters are derived could be found to exhibit such properties, out of a bounded maximum. Finally, the parameter of uncertainty is a mixed case, since it comprises both a bounded maximum of fuzzy and/or hidden elements drawn from structure or information, but also involves the scenarios events — or surprise changes affecting the state of a DDM problem along a simulation. There is no theoretical or technical upper limit to the number of events that can be featured in a DDM problem, although generally there are less events than there are simulation turns.

In a formal notation, the parameter space for the objective measures of complexity can be concatenated in an equation similar to a regression model. The objective measure of complexity OMC_{DDM} can be written as follows:

$$OMC_{DDM} = C_{DDM} + I_{DDM} + W_{DDM} + D_{DDM} + L_{DDM} + U_{DDM} + S_{DDM}$$

where $C_{DDM}$is the structural complexity proper, I_{DDM} is the information flow, W_{DDM} is the cognitive weight complexity factor, D_{DDM} is the measure of difficulty, L_{DDM} is the measure of nonlinearity, U_{DDM} is the degree of uncertainty, and S_{DDM} is a measure of instability for a dynamic system.

1.4.1 Structural complexity

The **structural complexity** C_{DDM} is considered paramount to the computation of complex decision problems, as suggested in the introduction. DDM problems can be expressed through graphs representing variables and relations with edges and nodes. A first calculation for the structural complexity of graphs is McCabe's cyclomatic complexity (1976), inspired by algorithmic graph theory. The CC_{DDM} model of **cyclomatic complexity** therefore refers to the proportion of edges to nodes, plus connected components, in a decision graph. The structural complexity C_{DDM} parameter is computed as the magnitude $V(G)$ of a graph G with n vertices, e edges, and p connected components, including delayed, conditional, uncertain, and hidden relations in the following formula:

$$CC_{DDM} = v(G) = e - n + 2p$$

Another version of the **structural complexity** C_{DDM} is the CNC_{DDM} model based on the **coefficient of network complexity**. The CNC metric was developed by Kaimann (1974) for the purpose of network analysis in management science. It was used to assess the complexity of products derived from the *program (or project) evaluation and review technique* (PERT), a statistical tool used in project management. In the particular context of network analysis in management science, it is defined as the quotient of activities squared divided by events, or by preceding work items squared divided by work items. Its network graph representation and accompanying means of computation are similar to the cyclomatic compelxity above:

$$CNC_{DDM} = \frac{e^2}{n}$$

A third and final version of the **structural complexity** C_{DDM} is the SDC_{DDM} model of **system dynamics complexity**, inspired by the concepts of Henry and Kafura's information flow (1981), Cardoso, Mendling, Neumann, and Reijers' interface complexity(2006), and Funke's differentiation between minimal complex systems and complex problem solving (2014). Cardoso's interface complexity (IC) determines the total complexity of an algorithm as the product of structural complexity and information flow, i.e., inputs and outputs. This was used as a differentiation criterion between endogenous and exogenous variables, as a system's many variables may have more than one role within a DDM problem. SDC_{DDM} may be computed as:

$$SDC_{DDM} = \textit{endogenous variables} \text{ x } \textit{exogenous variables} \text{ x } \textit{relations}$$

1.4.2 Information complexity

The **informational complexity** I_{DDM} is the information flow containing the product of inputs and possible outcomes, where the inputs contain both the action points and the number of possible interventions, and the outputs are the sum of the number of possible influences from the interventions, the number of possible contributions from variables to action points, all of which may contain delayed, conditional, uncertain, and hidden influences and contributions. The measure was inspired from the interface complexity model of Cardoso et al (2006) and the relational complexity of Wang (2009):

$$I_{DDM} = \textit{interventions} \text{ x } \textit{(influences from interventions + contributions to action points)}$$

The **action complexity** A_{DDM} is merely a variant metric based on the the information flow metric I_{DDM} presented immediately above, motivated by previous observations to the effect that small variations in the pool of action points for a DDM scenario can greatly affect the performance results. It is also inspired by Liu & Li's (2012) notion of *action complexity* in the context of elaborating measures of complexity for a task model. It is a quasi-linear transformation of I_{DDM} and will be treated as an alternative to that parameter in the remainder of this document. The equation representing the action complexity A_{DDM} can thus be summarized as follows:

$$A_{DDM} = \textit{action points} \text{ x } \textit{interventions} \text{ x } \textit{(influences from interventions + contributions to action points)}$$

1.4.3 Cognitive weight

The **cognitive weight** complexity W_{DDM} is similar to the CWCM composite model in that it focuses exclusively on the total of cognitive weights of a system as a function of the relations and influences from interventions in a DDM problem, as if it was the set of executed instructions in an algorithm. This metric corresponds to the operational complexity component of the CSC composite model (Wang, 2009). The cognitive weight complexity measure is the weighted sum of basic control structures by their cognitive weight type. A detailed breakdown of those basic control structures can be found in Shao and Wang (2003), Misra (2006), Wang (2007, 2009), and Pronovost (2014). A summery is presented in Table 56.

$$W_{DDM} = (relations \text{ x } W_i) + (influences\ from\ interventions \text{ x } W_i) + (contributions\ to\ action\ points \text{ x } W_i)$$

where the relations between variables, the influences from interventions on variables and action points include basic relationships, conditional relationships, and delays, and where W_i is determined by the theory of cognitive weights as featured in Wang (2009). Hidden relations and influences from interventions are excluded from the calculations.

Table 7 – Cognitive weights of basic control structures (BCS), from Wang (2009).

BCS	**notation**	**calibrated cognitive weight** w_i
sequence	$\rightarrow$	1
branch	$\mid$	3
switch	$\mid \ldots \mid \ldots$	4
for-loop	$R^{\,i}$	7
repeat-loop	$R^{\,*}$	7
while-loop	$R^{\,*}$	8
function call	$\rightarrowtail$	7
recursion	$\circlearrowleft$	11
parallel	$\parallel$ or $\square$	15
interrupt	↯	22

1.4.4 Difficulty

The objective measure of **difficulty** D_{DDM} was introduced as a means to differentiate between parameters of complexity and difficulty, in that the goal attainment of a decision problem of any kind can in principle be extremely easy or extremely hard to achieve, no matter the underlying complexity of the decision problem (Bedny, Karwowski, & Bedny, 2012). Indeed, the features of a complex system, such as structural intricacies or information flow volume, neither translate directly, nor in isolation from one another, into some measure of cognitive demand on a decision maker (Gonzalez, Vanyukov, & Martin, 2005). A complex system with twice as many interrelated components is not twice as difficult to understand or influence (Funke, 1988). Gonzalez et al (2005) also note that

> Decision makers may be able to achieve the goals of some dynamic decision-making systems by considering only a subset of system variables.

This is consistent with Mackinnon and Wearing's (1980, also mentioned in Karakul & Qudrat-Ullah, 2008) finding that interactions between subsystems in DDM facilitate overall performance. That is, by capitalizing on certain structural and informational features in a DDM problem which facilitate the achievement of goals, a decision maker may find itself facing a less challenging problem than the apparent complexity would suggest. In order to insure that a complex system does not yield to such narrower strategies, the experimenters must increase the interrelatedness of components so as to eliminate such subsystems. As the authors note,

> In such tightly coupled systems, complexity remains relatively constant because each dynamic decision requires consideration of most or all system variables.

Conversely, local processing (focusing on specific elements) in a sufficiently complex stock and flow system, as opposed to a global processing approach (focusing on a system's overall structure), is a recipe for failure (Fischer & Gonzalez, 2015).

It is with the above-mentioned concerns in mind that an objective measure of difficulty $D_{DDM}$was designed. D_{DDM} is a measure of the distance to goal attainment for the overall DDM scenario, and is calculated on the basis of the relative distance between the initial variable values and the values required to succeed (as detailed in appendix A). The goal distance metric also accounts for the distance between the initial values and the values of variables which warrant a failure, thus constituting a metric of "distance to win" with a "distance to fail" component. The computation of D_{DDM} is based on the *root-mean-square* (RMS) of distances to win and fail, as presented in the following equation:

$$D_{DDM} = \textit{(distance to win + (1 - distance to fail)) / 2}$$

where

$$\textit{distance to win} = \left\{ \sqrt{1/n \times \sum_{i=v_1}^{v_n} \left(\frac{G_{success_i} - G_{initial_i}}{G_{optimal_i}} \right)^2} \right\}$$

and

$$\textit{distance to fail} = \left\{ \sqrt{1/n \times \sum_{i=v_1}^{v_n} \left(\frac{G_{initial_i} - G_{failure_i}}{G_{range_i}} \right)^2} \right\}$$

Note that unlike the objective measure of Stability S_{DDM} presented below, there is no turn handicap to weight the number of turns used in a simulation relative to the minimal required number of turns to achieve success. This is due to the fact that such a turn handicap is already computed in the performance variable. Its presence in the response variable would therefore make it redundant to include it in the computation of the measure of difficulty D_{DDM}.

1.4.5 Nonlinearity

A measure of distance to linearity (or **nonlinearity**) L_{DDM} represents the weighted quantification of the degree of distance to a purely linear relation (of the form $y = ax + b$) for each of the relations between variables and the influence from interventions on both variables and action points. This measure was inspired from research by Brehmer (1974), Sterman (1989), Cronin, Gonzalez, and Sterman (2009), Soyer and Hogarth (2015), and Özgün and Barlas (2015). The metric is based on the average of the slopes a for each linear equation representing a relationship between variables or an influence from an intervention. The absolute values of the slopes a are then subtracted from a value of 1, to represent to opposite of a measure of linearity. The relations and influences from interventions where y is constant are excluded.

$$L_{DDM} = \textit{(distance to linearity for relations + distance to linearity for influences from interventions) / 2}$$

where

$$distance\ to\ linearity\ for\ relations\ = \left\{ 1/n * \sum_{i=rel_1}^{rel_n} (1 - |L_{a_i}|) \right\}$$

and

$$distance\ to\ linearity\ for\ influences\ from\ interventions\ = \left\{ 1/n * \sum_{i=int_1}^{int_n} (1 - |L_{a_i}|) \right\}$$

1.4.6 Uncertainty

Decision-making uncertainty and a lack of task transparency have been found to negatively impact human performance in DDM scenarios (Kopainsky & Alessi, 2015, Osman, Glass, & Hola, 2015, and Qudrat-Ullah, 2015 for a general overview). The objective measure of **uncertainty** U_{DDM} is a rather simple parameter expressing the quantity of information which is ambiguous (based on a fuzzy range of values), hidden from the decision maker in the DDM scenarios, as well as a quantification of the number of unforeseen events occurring in the scenarios. The net parameter value for the calculation of U_{DDM} is therefore based on the sum of the ratios of uncertain elements to their respective category, plus the ratio of the number of scenario events to the number of simulation turns. Note that the latter element of the computation of the U_{DDM}, i.e., the ratio of events to the number of turns, can in principle be larger than 1, and therefore, the overall value of U_{DDM} could be larger than 1 itself, hence the idea that this parameter of complexity is both a scalar and a ratio in its composition, as mentioned above in Table 6. Elements which may exhibit fuzziness, such as variables, relations, as well as influences from interventions on variables and on action points, may be uncertain for a number of units higher than 1 (e.g., the crime rate variable in COIN is uncertain at ± 2 units), and each point of fuzziness thereby counts as an additional unit in the calculations for the U_{DDM} parameter.

$$U_{DDM} =$$

$$\left\{ \frac{\frac{fuzzy_{var}+hidden_{var}}{variables} + \frac{fuzzy\ _{rel}+hidden_{rel}}{relations} + \frac{fuzzy_{int}+hidden_{int}}{infl.\ on\ variables} + \frac{fuzzy_{ap}+hidden_{ap}}{infl.\ on\ action\ points} + \frac{\#\ events}{\#\ turns}}{\#\ of\ categories\ above\ with\ non-zero\ values} \right\}$$

1.4.7 Instability

The measure of distance to stability (or **instability**) S_{DDM} is a measure of the degree of inertia of the DDM scenario from the point of its intrinsic dynamics. It expresses the DDM problem's overall tendency to remain stable, on a spectrum ranging from extreme inertia to extreme volatility, the latter being a sign of a tendency to catastrophic consequences when the system is left unchecked and unbalanced. The measure was inspired by Kerstholt (1996), Hagmayer, Meder, Osman, et al (2010), and Osman and Speekenbrink (2011). As an objective parameter of complexity, S_{DDM} measures the weighted distances between initial variable values and end game inertial values, using the RMS computation. This weighted distance is combined with a handicap for the proportion expressed between the number of turns required to fail in inertial conditions relative to the maximum number of turns, in order to emphasize the expediency of inertial failures in a system, beyond the overall end state upon failure.

$$S_{DDM} = \textit{(distance to end game inertial values + turn handicap to inertial failure) / 2}$$

where

$$distance\ to\ end\ game\ inertial\ values\ = \left\{ \sqrt{1/n \times \sum_{i=v_1}^{v_n} \left(\frac{S_{inertial_i} - S_{initial_i}}{S_{range_i}} \right)^2} \right\}$$

and

$$\textit{turn handicap to inertial failure} = \textit{(maximum \# turns - \# turns to fail from inertia) / maximum \# turns}$$

1.5 Discussion

The experimental method proposed herein aims to reproduce the conditions involved in the everyday dynamic decision-making problems to which experts are confronted in complex sociotechnical systems. While we are abstracting such DDM problems in a way that university undergraduates can participate in our experimental design, we do not want to sacrifice the external validity concerning the realism of those tasks. The DDM scenarios we have retained were developed with subject-matter experts, and implement mathematical abstractions of stock and flow models which mirror real-world complex decision problems, such as resource managements tasks, policy making tasks, and conflict resolution tasks.

The traditional experimental approach in cognitive psychology draws its methods from reductionist paradigms such as psychophysics, applied behavior analysis, and psychophysiology, with the essential addition of using concepts from information theory in order to explain the role of low-level informational processes, using internal codes (mental representations) as building blocks for higher-level cognitive functions. This microcognitive approach is indispensable to research interested in attention, memory, and perception, and plays a good part in the explanations concerning cognitive functions such as (simple) decision-making, language comprehension, and numeracy, but scales poorly to more complex, interactive behavior and cognition. Moving from microcognition to a macrocognitive framework is critical to explain the information-rich, interaction-involving, and hierarchically layered cognitive functions involved in the complex task environments outside the laboratory.

Dynamic decision-making involves cognitive processes, mental representations, and cognitive functions occurring at different time scales and best described at multiple levels of analysis, on the one hand. It also requires a wealth of particular task characteristics found in naturalistic settings which are typically not elicited nor reproduced in experimental designs concerned strictly with the microcognitive level of description. Capturing and studying DDM in a realistic experimental design requires a shift towards simulation-based experimentations, whereby the functional relationships between an agent and its environment (situation assessment, weighting choice utilities, learning from outcomes), the wealth of task characteristics (such as time pressure, dynamic problem states, and information opacity), and a proper context implementing a complex problem solving scenario (a resource management scenario, for example) are all accounted for.

Synthetic environments such as microworlds are interactive simulation environments suitable for experimentation in DDM, with an emphasis on high-fidelity task realism in low-fidelity laboratory constructs, featuring both experimental and instructional capabilities, and enhanced engagement through the gamification of the experimental design. The use of an experimental suite featuring a synthetic environment, data collection algorithms, and analytical tools such as the CODEM simulation environment is a "one stop shop" to experiment on DDM, from (i) the design of scenarios implementing stock and flow dynamical models into engaging stories with various constraints and objectives, (ii) running participants who interact with such scenarios as turn-based strategy games, and (iii) collecting meaningful data about the evolution of the DDM problem over time, including systems states, as well as the information-seeking behaviors of participants, their decisions, and overall performance.

Combined with the simulation-based experimentation presented in this chapter, the objective measures of complexity described in the previous section will allow us to analyze the results obtained through the empirical investigation reported in the next chapter. Ten objective measures of complexity were elaborated to capture the characteristics of DDM scenarios which confer those tasks a quantitative representation of the challenging nature of systems thinking. Those objective measures of complexity are the three variants of the structural complexity C_{DDM}, the two alternative measures of information complexity I_{DDM}, the cognitive weight complexity W_{DDM}, the measure of difficulty D_{DDM}, the measure of nonlinearity L_{DDM}, the degree of uncertainty U_{DDM}, and a measure of system instability S_{DDM}. As will be seen in the following chapters, some of those metrics of complexity exhibit redundant characterization of DDM problems, which translates into the statistical issue of multicollinearity.

Chapter 2 explores in depth the possibility of modeling the linear relationship between subsets of the metrics of complexity as parameters of a multiple linear regression model (MLR), with the human performance scores in five DDM scenarios. Machine learning methods of MLR are used in order to explore the parameter space of the multivariate data produced by comparing subsets of those MLR models between themselves, using goodness of fit criteria such as the coefficient of determination (a relative measure of the variance explained in the response variable), the standard error of the regression (an absolute measure of the amount by which the values predicted by a model differ from the quantities being observed), and the corrected Akaike information criterion (a measure of the relative quality of a model in terms of the entropy, or information loss, generated in fitting a model to the observations, relative to the information-theoretic complexity of that model).

Chapter 3 goes further in the modeling, analysis, and validation endeavor with the objective measures of complexity by exploring the bias–variance trade-off in the statistical models of dynamic decision-making through the use of nonlinear and non-parametric methods from the machine learning literature, using between-condition sources of variance exclusively, as well as through the use of robust regression models. A small number of MLR models which are found to be the best candidates to explain the variance in the human performance data are compared with their equivalent models in those nonlinear and non-parametric analyses, to assess whether the functional form of the relationship between the parameters of complexity and the performance scores is nonlinear in nature.

Chapter 2

Modeling and Validation of Objective Measures of Complexity for Dynamic Decision-Making

2.1 Introduction

This chapter presents the empirical groundwork used to establish objective models of complexity for dynamic decision-making scenarios, from a preliminary analysis of the parameterization of the characteristics of complex problem solving tasks discussed in the preceding chapter on the global experimental approach and the particular methodology advocated for the study of DDM, to the validation of the objective models of complexity. We present a summary of the human performance data and the results of the parameter selection for objective features of complexity. Then follows a presentation of the multiple linear regression models built from such parameters as well as a comparison between competing candidate models. Finally, a predictive analysis is presented in order to validate the candidate regression models of complexity for human performance in dynamic decision-making problems, using a cross-validation technique on an out-of-sample method involving an additional test scenario. The discussion section covers issues such as the multicollinearity of the parameters of complexity, the sensitivity of the regression models to small nuances in some of the calculations used for the parameterization and issues regarding the regression coefficients, as well as potential ways to improve the goodness of fit and the predictive power of the objective measures of complexity in order to account for the human performance data.

2.2 Method

2.2.1 Participants

215 university students (mean age = 22.25, *SD* = 4.29, 43.18% males, 56.82% females) from various domains of study (mainly psychology), received a financial compensation to undertake experimentation sessions of approximately 3 hours. The 215 students were spread over five particular groups of interest which were part of a larger, 3-year research project, which explains why the groups are unequal in sizes. The results of one participant (in the experiment using the COIN 2 DDM scenario) were improperly recorded due to a technical issue, and therefore the sample has been corrected as $n = 59$ in the following analyses.

2.2.2 Task: dynamic decision-making scenarios

2.2.2.1 SpaceLab

The first DDM scenario is SpaceLab, used for rehearsal in the CODEM environment after exposure to a brief tutorial (in the form of a PowerPoint presentation), and a very short exposure to CODEM itself through a minimally complex scenario named *Tribes* in order for the participants to familiarize themselves with the interface. SpaceLab is considered to be far less complex than the other DDM scenarios used in the experiment.

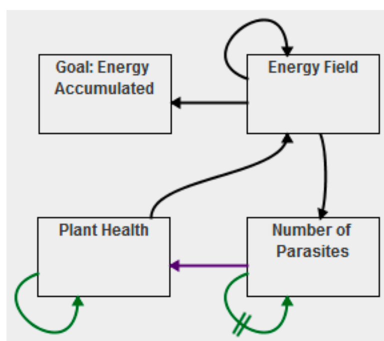

Figure 31 – Decision graph for the SpaceLab scenario.

2.2.2.2 Arctic, versions 1, 2 and 3

The Arctic development scenario is a complex DDM problem used for multiple purposes beyond the current research project, including learning stability and transfer studies involving intelligent tutoring systems. There are three versions of the DDM scenario of low to very complexity, as will be presented in the chapter concerning the experimental results and the analyses. Progressively more complex than the SpaceLab training scenario, they are nevertheless still less complex than the two COIN scenarios, from the point of view of structure, information, and cognitive weight. However, as it will also be presented in the next chapter, the Arctic scenarios present significantly more difficult goal achievement target values (encapsulated by the notions of difficulty and system instability), relative to the COIN scenarios.

Arctic 1 has 6 variables, it does not feature the environmental policies mediating variable, nor the incident management and the unresolved incidents. There are no feedback loops nor any delays in relationship effects, so the scenario only has 8 relationships. Arctic 2 adds a mediating variable, the environmental policies, as well as the conditional relationships involving this variable. It adds feedback loops but does not feature delays, relative to Arctic 1 and Arctic 3. Arctic 2 has a total of 6 variables and 15 relations. The complete and final version of the Arctic development scenario, Arctic 3, features a total of 8 variables and 20 relations. It adds incident management and unresolved incidents as variables, adding another layer of supply chain management complexity to the stakes. This is also the only version of Arctic which features delayed effects between variables and delays in the influences from interventions.

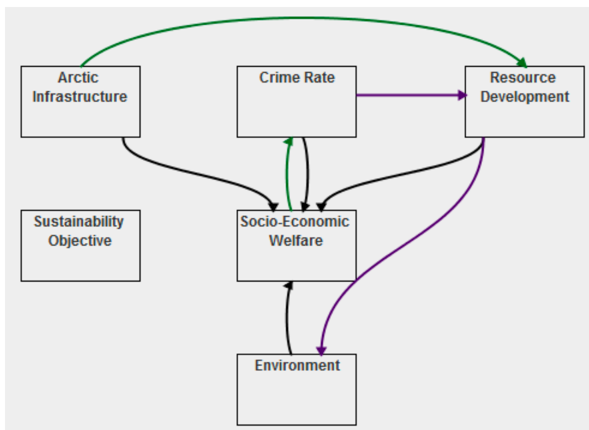

Figure 32 – Decision graph for the Arctic 1 scenario.

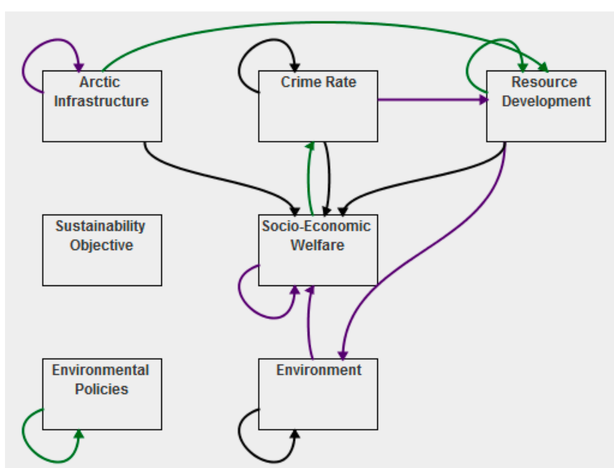

Figure 33 – Decision graph for the Arctic 2 scenario.

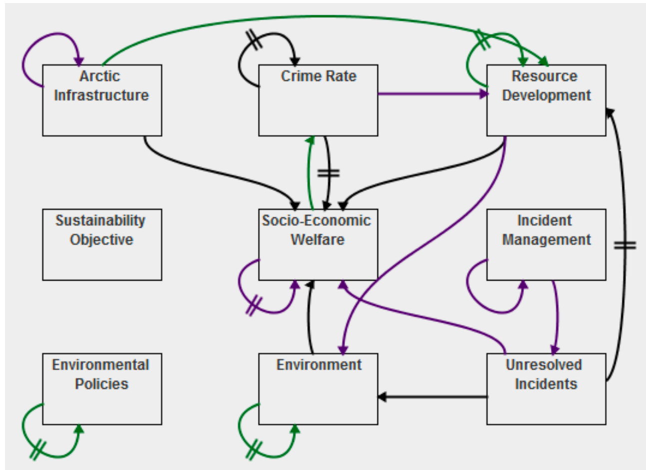

Figure 34 – Decision graph for the Arctic 3 scenario.

2.2.2.3 COIN 1 and COIN 2 (Counter-Insurgency)

The **co**unter-**in**surgency scenario is a highly complex DDM problem developed in a joint effort with the defence and aerospace community in order to teach notions concerning systems thinking to decision makers. It is framed as a geo-political scenario where the player is coordinating a multinational effort in order to stabilize a failing state, which lies in the grips of a rising insurgency. The player's goal is to return the host nation to a stable state and in a self-sustaining condition. The COIN scenario presents a considerable challenge to participants, with many uncertain (hidden, fuzzy) elements, and rather unstable internal dynamics rapidly leading decision makers to failure. Both COIN 1 and 2 have 9 variables, connected by a total of 34 relationships (7 of which are hidden to the player), and the player has 7 possible interventions, from which occur 17 influences on variables states, as well as 4 possible contributions to action points in future turns of play.

Winning at COIN is only possible if *all* of the situation indicators (with the exception of the mediating variable of *cultural understanding*) are outside the unfavorable red zone. Failure is insured if the *population allegiance* falls to zero. Adding to the difficulty of an already complex DDM scenario, the insurgent forces in COIN possess an adaptive quality which may yield adverse impacts of the system variable states, as the antagonists adjust their behavior in order to prevent the player's success. Such adaptive measures are presented as conditional events which depend on the system state and the player's decisions.

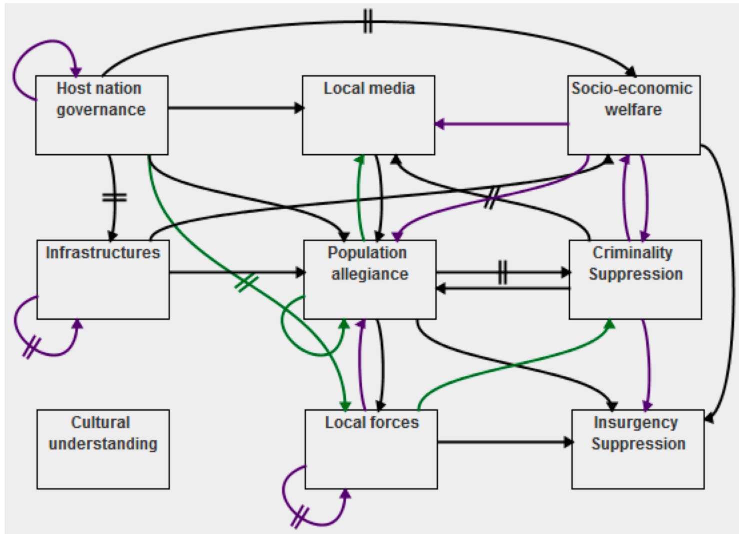

Figure 35 – Decision graph for the COIN 1 and 2 scenarios.

While the COIN scenarios are highly complex DDM problems in terms of structure, information, and cognitive load, we have used two slightly different versions of COIN that yielded considerably different distributions of performance scores, as will be seen in the next chapter. Version 1 of COIN is designed in a way that is slightly more manageable, with different initial conditions such as an increase in the pool of action points and a more favorable state of the important variable *population allegiance*, as well as a reduced number of turns of play. This affects not only the performance distributions, but also the calculations of the objective parameters of complexity, in particular the potential of action complexity A_{DDM}, the measure of relative difficulty D_{DDM}, and the measure of system instability S_{DDM}, which are presented in detail in the following section.

2.2.2.4 Cybernetia scenario in the Ecopolicy serious game

The Ecopolicy serious game was used as a test environment for the purposes of regression model validation. Ecopolicy offers three different DDM scenarios of varying complexity by defaults, and we have used the popular *Cybernetia* scenario, which employs a fictional setting resembling the political, economic, and cultural characteristics of a contemporary western nation. As the head of the state, the player must manage the nation with regards to resources, economic development, and quality of life, with a particular emphasis on the dangers of affecting the ecosystem if the adopted policies are too harmful to the environment.

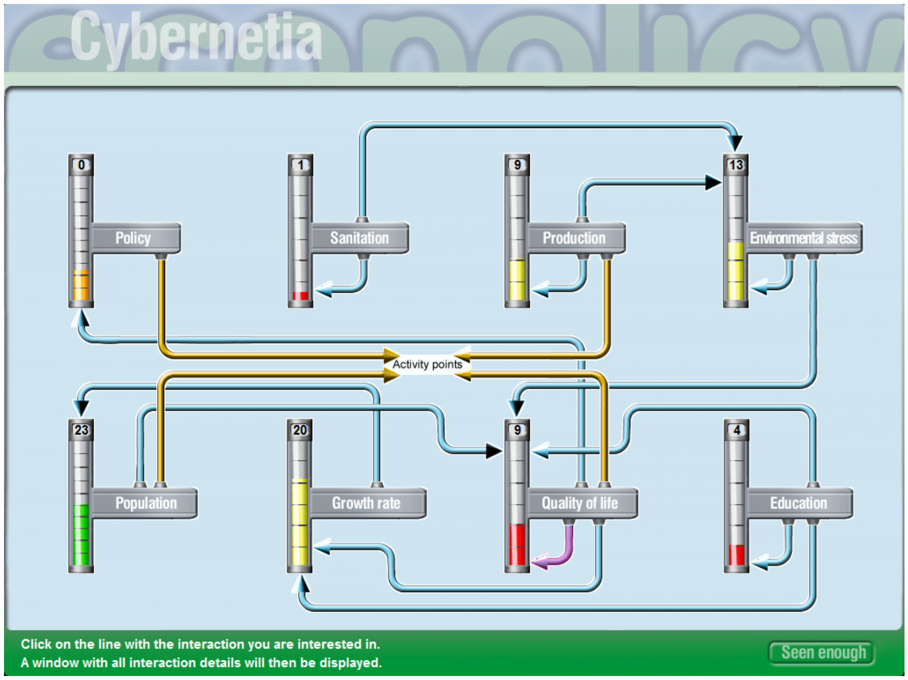

Figure 36 – Decision graph for the Cybernetia scenario in Ecopolicy.

Ecopolicy only has 4 interventions available to the player, but the degree of interconnectivity between interventions and variables is quite high. There are 8 variables and 14 relationships between them, and there are 4 possible influences from variables states and direct interventions on the pool of action points. Ecopolicy appears to be moderately complex relative to the DDM scenarios developed in CODEM, yet its measure of relative difficulty D_{DDM} is quite high in comparison to COIN and its measure of nonlinearity in relationships and interventions L_{DDM} is much higher than in all other scenarios, as will be presented in the next chapter. It can be run with or without randomly occurring events (positive and negative changes happening throughout the simulation), but this feature has been turned off in the current experimentation. An unfortunate consequence is the complete absence of uncertainty for this particular scenario, relative to the ones from the CODEM environment, as Ecopolicy does not feature fuzzy or hidden variables, relationships, and influences from interventions.

2.2.3 Experimental design and procedure

The overall experiment is a factorial design involving between-groups fixed and random effects factors, based on one response variable (performance), with the control variables consisting of six DDM scenarios (fixed effects factor), and ten objective parameters of complexity (random effects factor). Five experimental conditions featuring of groups of n = 20, 40, 60, 59, and 15 respectively were used to conduct the analyses. Within each of those experimental conditions, participants played through the a maximum of three DDM scenarios, which are presented below. The *null* hypothesis of the entire experiment is that a combination of the objective parameters of complexity, as instantiated in the various DDM scenarios, would not explain the variance observed in the participants' performance. The participants play through the DDM scenarios of varying degrees of difficulty and complexity in a strict linear process, from simpler to more complex. They could not be counterbalanced, given the nature of the tasks, as progressively more complex DDM scenarios require some degree of familiarization and instruction prior to moving up to more challenging decision problems. The participants in the five experimental conditions played through the DDM scenarios in the following sequences (not in chronological order for the purpose of simplification):

— SpaceLab, Arctic 1, Arctic 2

— SpaceLab and Arctic 3

— SpaceLab and COIN 1

— SpaceLab and COIN 2

— SpaceLab and Ecopolicy

2.2.4 Apparatus

The five main DDM scenarios used for experimentation were designed with the **Co**mplex **De**cision **M**aking Experimental Platform (CODEM) developed by Defence R&D Canada, and run on standard personal computers. CODEM is a simulator and testbed used to study the impact of interactive learning environments on the comprehension of individuals involved in complex decisions problems. CODEM contains a scenario editor allowing the complete oversight and total control on scenario design down to every single DDM entity, parameter, and relationship, allowing the creation of experimental content through a GUI. It can then be deployed as scenarios through which participants experience dynamic decision-making problems, and finally, CODEM generates a plethora of raw data outputs from which statistical analyses may be undertaken.

The scenarios are contextualized implementations of dynamic decision-making problems, where a participant may observe and influence the changes in variable values of key parameters in a complex system. Some of the variables must be minimized or maximized, or attain a particular threshold value, in order to achieve success (or merely avoid catastrophic failure). The stock and flow models underlying the dynamical systems implemented through the scenarios are mathematical abstractions of DDM problems following a rule-based, discrete-event system. In the scenario editor one can create the DDM variables and the relationships between them, the interventions from participants used to influence the system's state, scenario events to surprise and alter the flow of the system's dynamics, etc.

CODEM is a turn-based computer simulation, and require that participants expend some "action points" in order to effect some changes in the system variables. The more complex scenarios used in the present experiment, the third version of the Arctic development scenario (presented below) and the two versions of COIN (which stands for Counter-Insurgency), were developed with the support of subject matter experts in domains such as defence and aerospace, and governance and policy-making. Those DDM scenarios are therefore considered to be realistic complex problem solving tasks exhibiting a high degree of external validity, with meaningful entities and relationships for expert decision-makers in complex sociotechnical systems. This stands in contrast with contrived scenarios of a more artificial nature, such as in the case of the Tribes and SpaceLab scenarios, designed with only playfulness in mind. The latter two scenarios are only used for the familiarization of participants with the user interface of CODEM, and to give them a quick tutorial concerning the essential concepts involved in dynamic decision-making and systems thinking.

The following sections present brief summaries for the dynamic decision-making scenarios, as well as a visual representation of their structural complexity in the form of decision graphs. A sixth scenario was used to provide data in a regression model validation phase, as presented at the end of chapter 2. This scenario, called *Cybernetia*, was used in the Ecopolicy $^{\text{TM}}$ serious game environment, developed by MCB Publishing House. Ecopolicy is an 'edutaintment' product, meaning that it aims to provide an entertaining experience while allowing the user to educate itself in the process of playing. Games can be understood as abstract models, and the level of realism and detail of such models can determine whether the user experience will carry a certain degree of reusable and repurposeable knowledge in the context of real-world problems.

CODEM and Ecopolicy are discrete-event simulation environments which can be executed either as completely or partially deterministic simulations, i.e., they can include some randomness if so desired in order to observe the impact of stochasticity on scenario performance. The feature of random events has been disabled due to our interest of observing the relationships between the parameters of complexity and human performance in DDM problems, affording a more controlled design of the variance in both experimental and response variables.

2.2.5 Response metric: measures of performance in dynamic decision-making

The scenario performance assesses goal attainment for each scenario, suggesting a measure of the degree of the decision makers' comprehension and success at controlling the underlying dynamic system in a DDM problem. The performance measures for the five DDM scenarios used in the CODEM environment, in terms of goal attainment following a computation of the goal distance between initial values and values which warrant the successful completion of a scenario, were weighted over the number of turns required to complete a given scenario, as the scenarios did not share identical characteristics in order to establish comparable scores. Both the weighted performance scores and the turn handicap for having completed the scenarios in less turns than the maximum number of turns allowed were computed using the distributions of scores in comparison with the results of a data mining algorithm. All five scenario performance distributions were thus analyzed with the support of 300,000 simulations for each DDM scenario in CODEM. For example, the minimal number of turns in order to achieve successful goal states in Arctic 3 and COIN 2 are 6 and 5 turns respectively.

2.2.6 Control measures: objective parameters of complexity

Previous models of complexity for DDM involved the quantification of structure, information, and cognitive load in a compound fashion (as seen in Table 1 in the introduction), which yielded modest results in order to tackle the variance explained in the response variable. The compound models were decomposed into smaller sets of parameters, following the suggestions of Liu and Li (2011, 2012, 2014), Pronovost et al (2014), and Stouten and Größler (2017). The intuitive idea is that separating the composite models into a subset of maximally orthogonal parameters would likely avoid strong associations between the independent variables, or multicollinearity, thereby also maximizing the association between the individual objective parameters of complexity with the response variable.

From the compound models presented earlier were retained the commonly observed notions of structure and information, as well as the concept of cognitive weight from the field of cognitive computing. A particular attention is given to a variety of ways of calculating the structural complexity of decision problems, as the structural and systemic complexity is a favorable way to frame the complexity of dynamic decision-making problems. Another non-trivial issue is the means to calculate the information complexity of such problems, as discussed in Pronovost et al (2014), whereby the inclusion or exclusion of action points, or then 'action potential' afforded by the design of a particular DDM scenario may significantly and substantially influence the human performance results.

Finally, an additional set of four parameters, namely difficulty, nonlinearity, uncertainty, and instability, was combined to the above. A summary table of the objective measures of complexity is presented in Table 8, detailing the formal type and associated quantification for each parameter, as well as a short description of their purpose and their means of computation. Structure, information, and cognitive weight are described as magnitudes, that is, they are unbounded scalars with no ceiling value.

Table 8 – Summary of the objective measures of complexity.

parameter	type	quantification	what it is	how it's calculated
structure CC_{DDM} (Cyclomatic complexity)	magnitude	scalar	arcs and nodes in graph representation	difference between arcs and nodes
structure CNC_{DDM} (Coefficient of Network Complexity)	magnitude	scalar	arcs and nodes in graph representation	ratio of squared arcs to nodes
structure SDC_{DDM} (System Dynamics Complexity)	magnitude	scalar	arcs and nodes in graph representation, incl. functional role of elements	product of endogenous variables with exogenous variables and relations
information complexity I_{DDM}	magnitude	scalar	inputs, interventions, and outcomes	product of inputs and outputs
action complexity A_{DDM}	magnitude	scalar	inputs, interventions, and outcomes	product of action points, inputs, and outputs
cognitive weight W_{DDM}	magnitude	scalar	workload (time) to recognize operations	sum of products of basic control structures
difficulty D_{DDM}	relationship	ratio	distance to goals (variable values)	root-mean-square of distances to win and fail
nonlinearity L_{DDM}	relationship	ratio	distance to linearity for relationships	average of regression slopes for all relations
uncertainty U_{DDM}	relationship & magnitude	ratio incl. scalar	hidden and fuzzy variables, relations	averaged hidden / fuzzy elements + events
instability S_{DDM}	relationship	ratio	distance to end state under idleness	rms of dist. to end values (incl. # of turns to fail)

In principle, there could be an indefinite number of variables, and relationships between variables, for a DDM scenario. Difficulty, nonlinearity, and instability express relationships, whereby a number of elements from which the first three parameters are derived could be found to exhibit such properties, out of a bounded maximum. Finally, the parameter of uncertainty is a mixed case, since it comprises both a bounded maximum of fuzzy and/or hidden elements drawn from structure or information, but also involves the scenarios events — or surprise changes affecting the state of a DDM problem along a simulation. There is no theoretical or technical upper limit to the number of events that can be featured in a DDM problem, although generally there are less events than there are simulation turns. A detailed breakdown of the characteristics and computations for each objective parameter of complexity is presented in section 1.4 of the previous chapter. Appendix A presents the characteristics of the DDM scenarios which are used to quantify those measures of complexity, while Appendix B presents the calculations specific to the DDM scenarios for each parameter of complexity.

2.3 Results

2.3.1 Data analysis and variable selection

Performance results for the five DDM scenarios using the CODEM simulation environement varied between 5% and 95%, with average scores ranging from 60.30% ($sd = 16.64$) in the simpler scenario Arctic 1, to 14.71% ($sd = 5.02$) in the more complex and difficult scenario COIN 2. A box plot graph presents the central tendencies and variances of performance scores for each DDM scenario in Figure 37, while a summary of the descriptive statistics is presented in Table 9. A one-way analysis of variance reveals that performances vary significantly across scenarios with $F(4,\ 194) = 43.75$, $p < .001$. Scenario performance medians and means are presented together in order to illustrate the discrepancies in distributions for the response variable. The importance of the differences between group means and medians will be discussed at length following the MLR analyses in the present chapter, as well as for the purposes of the alternative regression analyses in the following chapter.

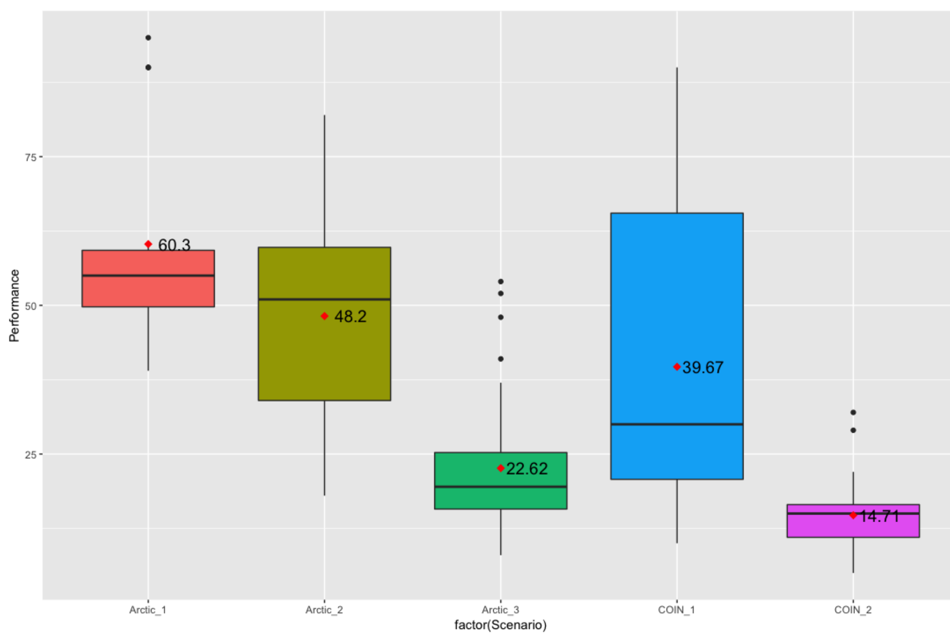

Figure 37 – Performance scores for the five DDM scenarios.

Table 9 – Descriptive statistics for the Performance metric used as a response variable in the five DDM scenarios.

		Performance				
Scenario	n	mean	median	sd	min	max
Arctic_1	20	60.30	55.00	16.64	39.00	95.00
Arctic_2	20	48.20	51.00	18.24	18.00	82.00
Arctic_3	40	22.62	19.50	10.70	8.00	54.00
COIN_1	60	39.67	30.00	23.79	10.00	90.00
COIN_2	59	14.71	15.00	5.02	5.00	32.00
All	199	31.77	22.00	22.12	5.00	95.00

2.3.1.1 Parameters of complexity for the dynamic decision-making scenarios

Based on the complexity metrics for individual parameters of complexity listed in the methodology section, the values for the five DDM scenarios, namely Arctic 1, Arctic 2, Arctic 3, COIN 1, and COIN 2, were computed, and are presented in Table 10 below. Appendix B presents Figures that illustrate the relationship between all five scenarios for each objective parameter of complexity (Figures 62 to 71). It can observed that while the overall complexity of DDM scenarios in CODEM increase across Arctic versions 1 to 3, and similarly from COIN 1 to 2, such scenarios can differ greatly across some parameters of complexity while remaining similar or event constant across a number of other parameters.

For example, Arctic 1 and 2 differ in terms of structural and informational complexity, while retaining identical values in terms of distance from initial conditions to both success and failures, thereby yielding identical scores on the difficulty (D_{DDM}) parameter. They are also constant in terms of uncertainty (U_{DDM}). The Arctic 3 scenario is quite different on all parameters from both its Arctic 1 and 2 siblings and the COIN scenarios, yet it also yields an identical difficulty (D_{DDM}) rating as Arctic 1 and 2. COIN 1 and 2, on the other hand, vary in terms of action complexity (A_{DDM}), difficulty (D_{DDM}), they vary slightly in terms of uncertainty (U_{DDM}), and more substantially in terms of system stability (S_{DDM}). COIN 2 affords less manipulations (less action points than COIN 1), yields slightly higher difficulty in terms of distance to success and failure, exhibits a bit more uncertainty, and is more prone to failure under inertia.

COIN 1 and 2 both differ greatly from the Arctic scenarios on most parameters, with higher scores for structural and informational complexity, as well as action complexity and cognitive weight. Their difficulty ratings (D_{DDM}) are substantially lower than the single rating shared across the Arctic scenarios, which may seem counterintuitive at first given the lower performances summarized above for the COIN scenarios, particularly the very low performance scores in COIN 2. But this phenomenon must not be interpreted in isolation from the other objective parameters of complexity, as the multiple linear regression models will reveal in the following sections.

Table 10 – Values for the objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2).

parameter	**Arctic 1**	**Arctic 2**	**Arctic 3**	**COIN 1**	**COIN 2**
CC_{DDM}	8	20	32	50	50
CNC_{DDM}	24.20	88.17	171.13	277.78	277.78
SDC_{DDM}	200	756	1600	2128	2128
I_{DDM}	36	40	70	378	378
A_{DDM}	576	640	1400	8316	7560
W_{DDM}	57	88	142	245	245
D_{DDM}	.5016	.5016	.5016	.1834	.2072
L_{DDM}	.0795	.1076	.0913	.0882	.0882
U_{DDM}	.0944	.0944	.2306	.1957	.2028
S_{DDM}	.5016	.5443	.5228	.5217	.5657

2.3.1.2 Relationships between the parameters of complexity

We have observed the relationships between the DDM scenarios across parameters of complexity in the previous section, but what of the relationships between the objective parameters of complexity between themselves? The correlograms in Figures 38 and 39 present the linear relationships between the parameters, and a parallel coordinates chart in Figure 40 illustrates such relationships through the use of standardized scores. It can be inferred that many of the parameter values for the DDM scenarios are highly collinear, a consequence of their constant values among some scenarios, and given the small set of scenarios.

It should be noted that the parameter values implemented in the five DDM scenarios of the present research project are assumed to be random samples in a larger pool of unrestricted domains, i.e., they represent random effects for the purposes of statistical inference tests such as analyses of variance and measures of association such as general linear models. All of the analyses performed through the exploration of the parameter space of the objective measures of complexity use variance components models, that is, they are random effects models. Since the highly collinear nature of some of the parameters of complexity will likely cause them to inadequately explain the variance in the performance variable given the phenomenon of multicollinearity, the next section presents a more rigorous method of parameter selection through the principal component analysis method.

Before we proceed with a principal component analysis and the regression modeling endeavor, we must distinguish between *accidental* and *semantic* collinearity between the objective parameters of complexity. The presence of collinearity between parameters is said to be accidental if the parameters yield high variance-covariance scores due to design choices in the creation of DDM scenarios, and not because the parameters express characteristics of complexity which are redundant. That is, the accidental collinearity is so by virtue of arbitrary design choices in the features expressed by the DDM scenarios. For example, the values of difficulty, nonlinearity, uncertainty, and instability can take arbitrary values which could be made independent between one another, yet they depend in turn on a certain measure of structural complexity and/or information complexity.

On the other hand, the presence of collinearity between parameters is said to be caused by the semantic properties of such parameters, insofar as the parameters yield high variance-covariance scores due to the actual characteristics of complex decision-making captured through the parameters. The alternate measures of structural complexity C_{DDM} are obviously dependent on one another due to their representation of similar, intrinsic properties of complex decision problems, namely the graph-based computation of variables and relations as graph edges and nodes. Likewise, the cognitive weight is closely related to structural and information complexity. Lastly, the action complexity is also a quasi-linear transformation of the information complexity parameter. It is expected that some parameters will be dropped in favor of more successful ones in the MRL analyses if they are semantically related, while the parameters which turn out to be accidentally collinear will be retained in the analyses if they afford higher goodness of fit scores.

Lastly, the correlograms also present the linear and monotonic relationships for the individual parameters of complexity with the mean and median performances for all DDM scenarios. It can be observed that the difficulty D_{DDM} and the nonlinearity L_{DDM} have peculiar relationships with the performance data, the first being positively correlated with performance, while the second is barely positively or negatively correlated at all, depending on the use of linear or monotonic correlation scores. Combined with the variable selection analysis discussed in the following section, this will be factored into the multiple linear regression model selection hereafter.

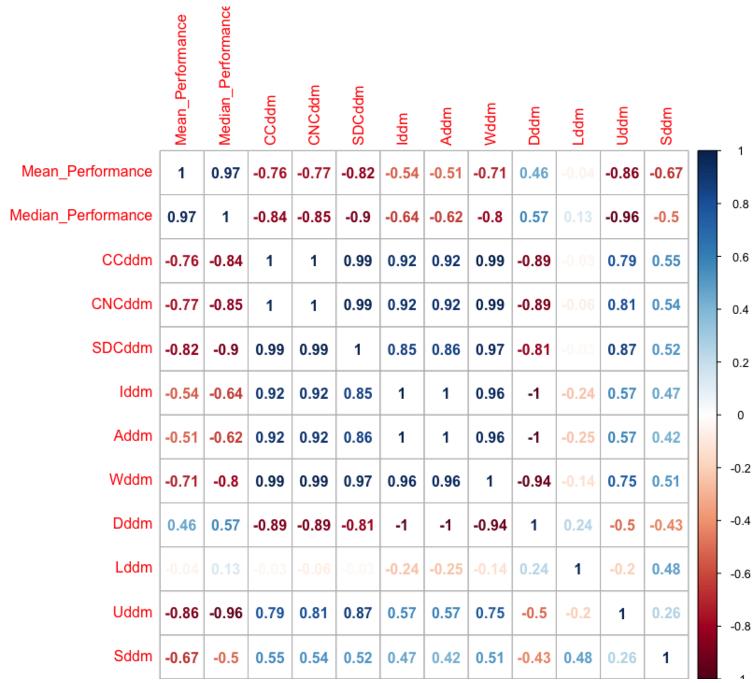

Figure 38 – Correlogram representing the Pearson (r) correlations for the mean and median performances in DDM scenarios with the objective parameters of complexity.

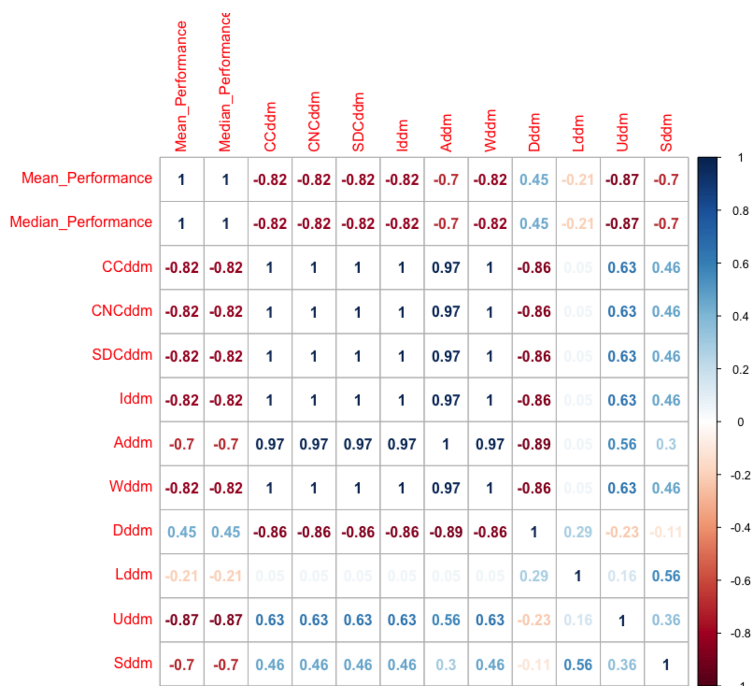

Figure 39 – Correlogram representing the Spearman (ρ) correlations for the mean and median performances in DDM scenarios with the objective parameters of complexity.

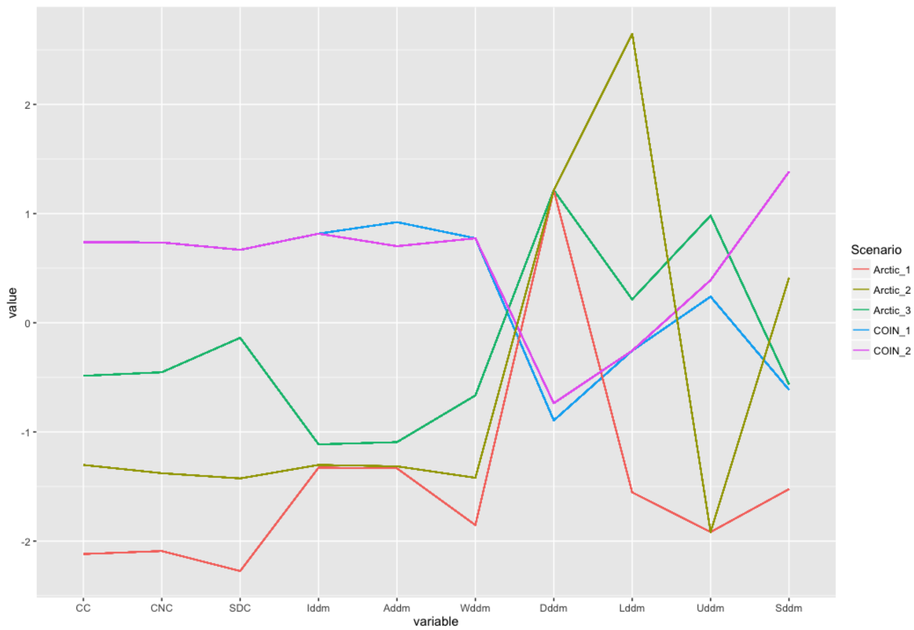

Figure 40 – Parallel coordinates chart presenting the relationships between the parameters of complexity for the five DDM scenarios. The parameter values are centered and scaled.

2.3.1.3 Variable selection through principal components analysis

A principal component analysis (PCA) may be helpful in order to reduce some of the highly collinear - and therefore redundant - information captured through the objective parameterization of complexity of DDM problems. A PCA is a way to analyze the structure of a correlation matrix whereby a few key explanatory variables are created, thus reducing the dimensionality of a multivariate dataset. Those new explanatory variables, called principal components, maximize the quantity of variance explained in the multivariate data in an ordinal fashion, that is, the first principal component captures as much of the explained variance, then the second attempts to capture as much of the remaining variance, etc.

The principal components capture the variance in an uncorrelated fashion, i.e., they are independent and orthogonal between themselves. The original variables map to the principal components through *loadings* (standardized and weighted). Assessing their relative importance in the components thus supports the analysis of whether their explanatory power, in terms of relevance and magnitude, makes such variables independent or redundant, and to what extent.

Figure 41 shows a bivariate representation of the topology linking the original 10-parameter space of the objective measures of complexity as *eigenvectors* (or characteristic vector of a linear transformation, yielding *eigenvalues*, or characteristic values representing transformation weights) [1] to the first two principal components of the PCA model. A "correlation circle" marks the value range of correlated eigenvectors with those principal components using a standardized range of value over the parameter space, and the DDM scenarios are included in this space in order to capture their relationship to both the eigenvectors of the parameter space and the principal components themselves.

The first principal component hints that all three structural complexity parameters CC_{DDM}, CNC_{DDM}, and SDC_{DDM} are unsurprisingly highly related to one another, and to information complexity I_{DDM}, action complexity A_{DDM}, as well as the cognitive weight complexity W_{DDM}. This is in turn related to the particularly high values of the DDM scenarios of tremendous structural and informational complexity, namely COIN 1 and COIN 2, and to Arctic 3. The difficulty parameter D_{DDM} also holds a strong influence relative to the first principal component, spread apart by the high discrepancies between the Arctic scenarios and the COIN scenarios.

The second principal component yields strong loadings for the nonlinearity L_{DDM} parameter and the system instability S_{DDM}, differentiated through the discrepancies between the Arctic 1 and Arctic 2 scenarios, with the remaining three scenarios (Arctic 3, COIN 1, and COIN 2) holding a middle ground. A third principal component (not displayed in the bivariate plot), weighting at less than 10% of the variance explained in the 10-parameter space, appears to correlate highly with the uncertainty parameter U_{DDM} (approximately 85%) and somewhat with difficulty D_{DDM} (above 25%), so those two parameters will be under scrutiny in the MLR model building effort presented in the next section.

1. An eigenvalue yielding scores larger than 1 is an indication that the principal component explains more variance than that accounted by one of the original variables in the standardized data. This is a traditional cutoff point to retain principal components in a PCA analysis.

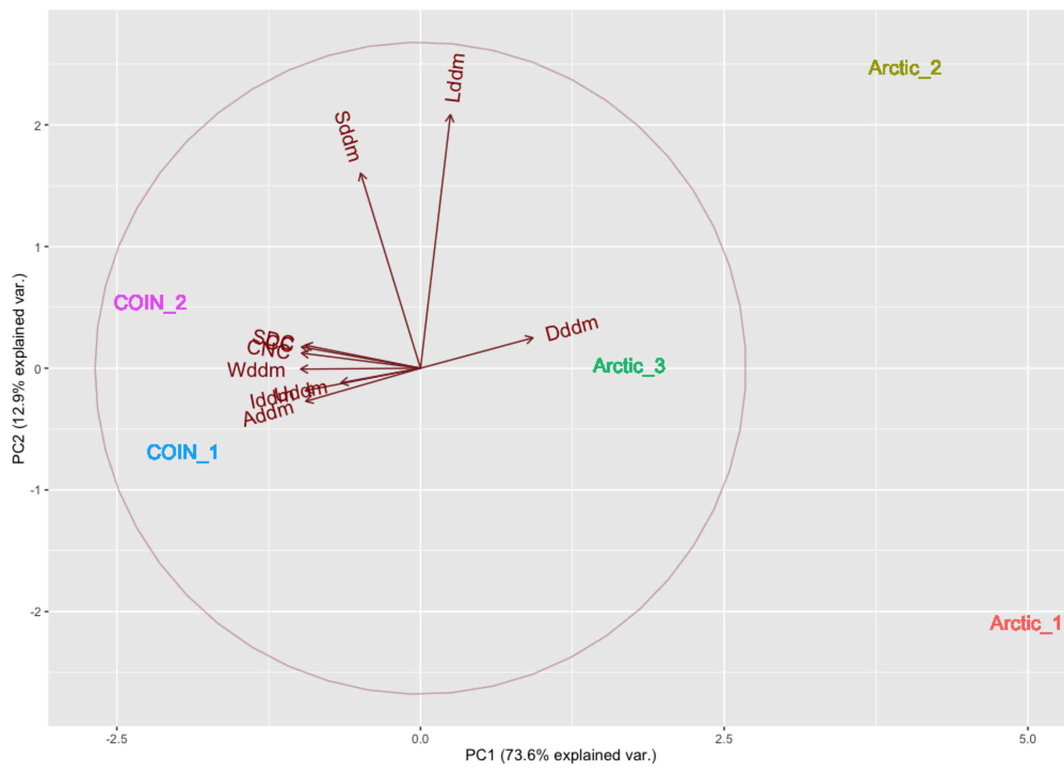

Figure 41 – A bivariate plot representing the principal component analysis loadings of the ten parameters of complexity for the five DDM scenarios.

2.3.2 Multiple linear regression model selection for the dynamic decision-making scenarios

The initial parameter space for the selection of candidate models would be a considerably large, 10-dimensional space, with an upper bound of $10! = 3{,}628{,}800$ combinations. We have seen in the previous section that the ten parameters are not only redundant in terms of collinearity, they are so because of semantic or accidental properties of the features of complexity and the design of the DDM scenarios. In order to segregate between the highly collinear parameters by virtue of linear transformation, a few subsets of parameters have been segmented. Three main subsets are thus segregated by the three parameters representing the structural complexity C_{DDM}, as they are semantically related, and for each of the three subsets, an additional subset of two parameter sets are segregated by the inclusion of information complexity I_{DDM} or action complexity A_{DDM}, as those parameters are not only semantically related but they are also a linear transformation of one another. A number of considerations are taken into account, in light of the observations concerning the relationships between the parameters of complexity as they are implemented through the five DDM scenarios:

— the cognitive weight complexity W_{DDM} has been dropped altogether, as it is invariably a linear transformation of the structural complexity C_{DDM} and information complexity

I_{DDM}. The cognitive weight metric has been successfully correlated with subjective measures of complexity (De Silva, 2012), and may play a role in future research,

— the difficulty D_{DDM} parameter will be preserved for the analyses, even though it appears positively correlated with performance, because of its semantic and methodological significance in previous experiments (Pronovost et al, 2014), as well as being substantially different in ratings across two clusters of the DDM scenarios in CODEM, i.e., the Arctic vs. COIN scenario variations,

— the L_{DDM} parameter has been dropped for being either mildly positively (Pearson r) or negatively (Spearman ρ) correlated with the performance data, while also being almost identical among the DDM scenarios, unable to tease apart any reasonable differences in performance. The L_{DDM} scores for the five scenarios in CODEM differ over less than 3% in values across the entire metric (on a percentage scale itself).

The complete set of parameters used to build multiple linear regression models of objective measures of complexity in order to explain performance in DDM scenarios is thus a matrix of 3 x 2 subsets of 5 dimensions, or six sets of 5 parameters where each parameter subset has $5! = 120$ possible combinations. The factorial design reflects an interest in conducting an exhaustive search of the parameter space over all possible sequences of parameters in a regression analysis, but a priority is given to the structural and information complexity measures C_{DDM} and I_{DDM} respectively, based on prior literature on dynamic decision-making (Senge, 1990, Brehmer & Allard, 1991, Kinsner, 2010, Funker, 2014, Pronovost et al, 2014).

CC-based structural complexity component C_{DDM} based on the CC calculation

- **I_{DDM}-based** information complexity component based on interventions, influences from interventions, and contributions from variables to action points
- **A_{DDM}-based** information complexity component based on the pool of action points, interventions, influences from interventions, and contributions from variables to action points

CNC-based structural complexity component C_{DDM} based on the CNC calculation

- **I_{DDM}-based** information complexity component based on interventions, influences from interventions, and contributions from variables to action points
- **A_{DDM}-based** information complexity component based on the pool of action points, interventions, influences from interventions, and contributions from variables to action points

SDC-based structural complexity component C_{DDM} based on the SDC calculation

- **I_{DDM}-based** information complexity component based on interventions, influences from interventions, and contributions from variables to action points
- **A_{DDM}-based** information complexity component based on the pool of action points, interventions, influences from interventions, and contributions from variables to action points

2.3.2.1 Stepwise linear regression modeling and automatic model selection by exhaustive search

The six sets of 5-parameter models of multiple linear regression where processed through three different methods, namely blocked, stepwise, and factorial (exhaustive search) designs, and assessed using three goodness of fit measures: the relative measure expressed by the coefficient of determination (adjusted R^2), the absolute measure of the standard error of the regression, or root-mean-square error ($RMSE$), and the corrected Akaike information criterion ($AICc$), which gives a relative estimate of the information loss when a model is used to represent data (a trade-off between a goodness of fit indicator and an assessment of model complexity).

Appendix C presents the exhaustive search for the best fit multiple linear regression models based on the six subsets of parameters of complexity, conducted with the help of the `leaps` and `glmulti` packages in `R`. An example of the exhaustive search with `leaps` using the adjusted R^2 selection criterion is presented in 42, while an example of the exhaustive search with `glmulti` using the $AICc$ selection criterion is presented in 43. Both examples use the 5-parameter subset including the structural complexity component C_{DDM} based on the SDC calculation, and the action complexity A_{DDM} variant of the information complexity component.

The figure featuring the exhaustive search using the adjusted R^2 criterion (Figure 42) illustrates how a few candidate models featuring 3- and 4-parameters converge towards a ceiling of approximately 46% of explained variance in the human performance scores. The figure featuring the exhaustive search using the $AICc$ information criterion (Figure 43) shows the actual spread of the progression in narrowing down which MLR models explain as much of the variance in the data relative to its *maximum likelihood* as calculated from the entropy of the unexplained variance (here, the total information loss yielded by fitting a given model to the data). Smaller values indicate a better fit, so models figuring under the horizontal line of two IC (information criteria) units above the best model are considered good candidates.

Appendix C also presents an exploration of the model-averaged variable importance for the parameters of complexity in the exhaustive search phase. Figure 44 shows an example, using the same sample parameter set as above (with SDC_{DDM} and A_{DDM}). This graph displays each parameter of complexity featured in the 5-parameter MLR model subsets as its estimated importance (or relative evidence weight), computed as the sum of the relative evidence weights of all models in which the term appears (through the simulation of $5! = 120$ possible combinations for each subset). The vertical red line is only displayed as a conventional cutoff indicator of extreme importance (at the 80% mark) for a predictor in a given model.

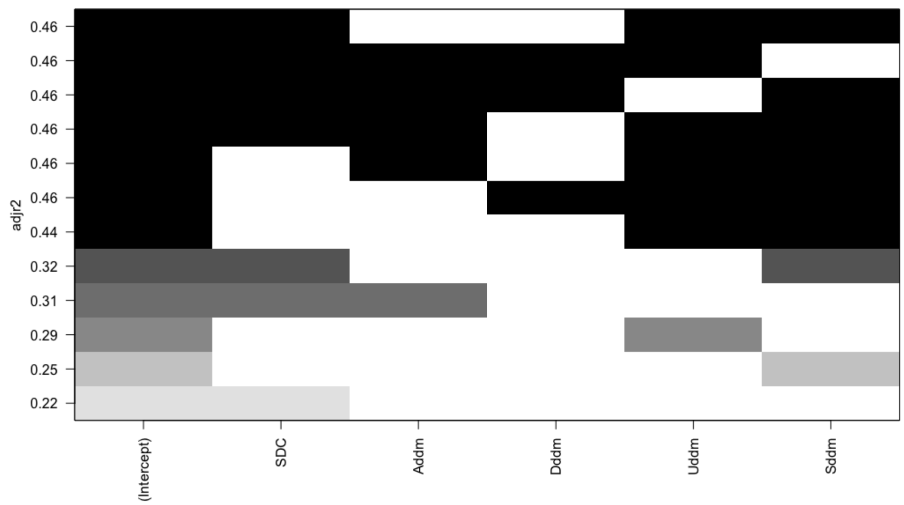

Figure 42 – Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

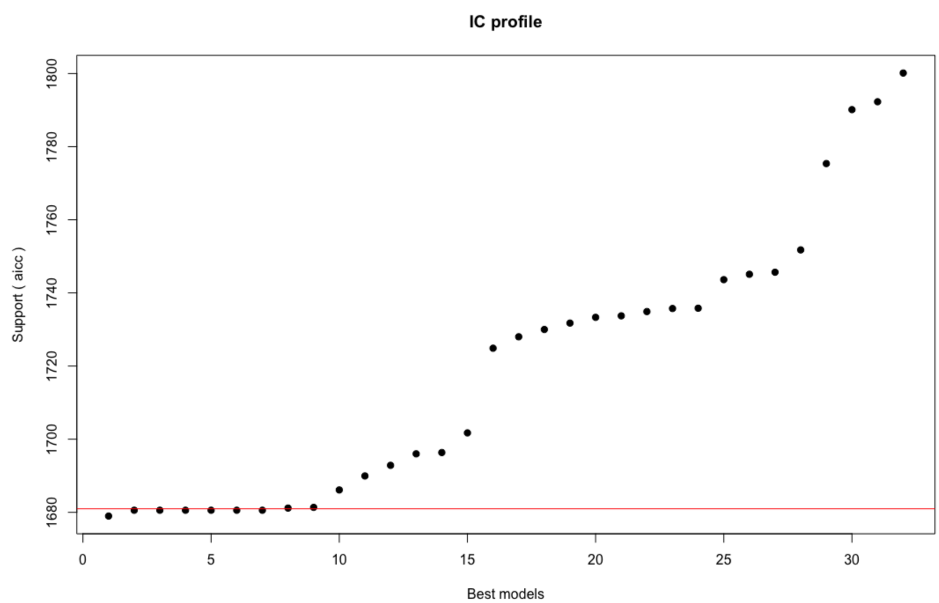

Figure 43 – Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

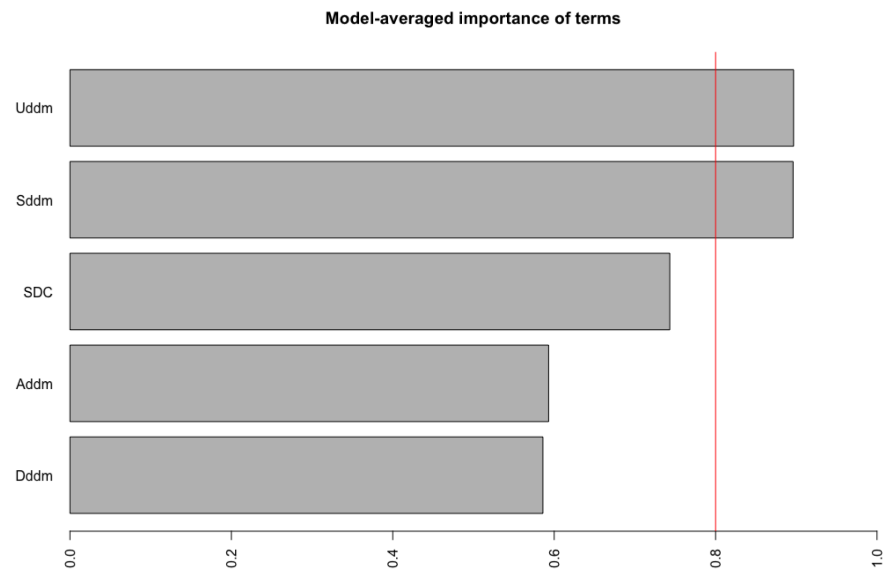

Figure 44 – Model-averaged variable importance using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

2.3.2.2 Regression diagnostics

Various sample distribution and model assumptions have been tested in order to validate the regression analysis effort, and to apply corrective measures where and when it is warranted. Those assumptions concern the sample size for regression analysis, the normality of variables and model residuals, multicollinearity amongst independent variables, homoscedasticity, as well as testing for the presence of multivariate outliers.

Sample size Tabachnick and Fidell (2012) advise a minimum of 50 + 8 (k) observations for an overall MLR model, or 104 + k observations for single predictor tests, where k is the number of predictors. Our dataset varies between 199 and 214 observations, and our predictors are limited to a few subsets in 2- to 4-predictor models of multiple linear regression.

Multicollinearity As mentioned in the section on data analysis and variable selection, the predictors of the MLR models representing our parameters of complexity are either intrinsically collinear (if they redundantly feature structural or information characteristics, for example) or accidentally collinear (if they exhibit high correlation values by virtue of an accidental linear relationship in the design of the DDM scenarios, i.e., a relationship that is not necessary and could be orthogonal). The MLR model comparison presented below uses multicollinearity diagnostic statistics, namely the tolerance indicator and the related variance inflation factor (VIF), but the remaing collinearity scores should be taken with a grain of salt, given the effort undertaken above in segregating the parameter space into smaller subsets of 5-parameter models.

Normality of residuals The distribution of studentized residuals, as observed on quantile-quantile plots (see an example in Figure 45, all other plots are presented in appendix C), indicates that we have heavy-tailed residuals. There are too many extreme positive and negative residuals, and the assumption that the error terms are normally distributed is therefore not met. Normally distributed multivariate data will enhance the MLR solution to a regression equation, in terms of statistical power and accuracy of goodness of fit indicators. The corollary assumption of linearity makes us expect to see residuals that are not too far away from the standardized 0 value. Values less than -2 or greater than 2 are considered problematic. For the purpose of assessing our MLR models using the standard, ordinary least squares (OLS) approach to linear regression modeling, the following chapter will include model comparisons with non-parametric and nonlinear approaches which may befit our multivariate data in a more rigorous matter.

Homoscedasticity The homoscedasticity assumption is assessed via the observation of the spread, regularity of variations, and possible presence of patterns in the residuals for univariate and multivariate data. All three tests used to observe sample non-constant error variance (Bartlett's test, Levene's test, and the Fligner-Killeen test) returned negative results for the homoscedasticity of error variance in the performance scores relative to the DDM scenarios. This is not surprising, given the spread of variance illlustrated in the box plot graph at the beginning of this chapter (Figure 37). Additionally, using the Breusch-Pagan test (1979) and the Goldfeld–Quandt test (1965) of non-constant error variance for the complete MLR models also fails to support assumption of homoscedasticity, which can also be diagnosed from the above-mentioned QQ plot (Figure 45). As with the irregularities amongst residuals mentioned above, model comparisons with non-parametric and nonlinear approaches will likely mitigate the impact of heteroscedasticity.

Multivariate outliers Figure 46 illustrates the influence of individual observations relative to one another in computing the candidate MLR models presented in the following section, using the SDC_{DDM}, I_{DDM}, and D_{DDM} parameters. It uses Cook's distance D as an indicator of the effect of removing a data point on all the parameters combined. Observations with high D values hint that they exert a leverage effect on the MLR model, and constitute influential observations which bias the regression coefficients. The cutoff value for Cook's D is the $4/(n-k-1)$ method, a more conservative approach the the traditional cutoff value of $D =$ 1.0 or greater, but deemed appropriate for small datasets (Hair, Black, Babin & Anderson, 2010). The cutoff value is thus .04 in the plot presented in Figure 46 (the remaining graphs are presented in appendix C).

Looking at all Cook's distance D plots, it appears that a substantial number of observations exert a high influence on the regression coefficients representing the objective parameters of complexity, if we include the distance values along the cutoff line. Since the variance of error is non-constant across groups and there is a number of influential observations, corrective measures and alternative means of goodness of fit will be assessed in the next chapter, particularly through the use of weighted least squares and other non-parametric methods, as well as nonlinear approaches implemented in various machine learning algorithms. A particular attention will be focused towards robust and resistant estimators, and the absolute measure of the standard error of the regression, or root-mean-square error ($RMSE$), will be adopted as goodness of fit indicators, in contrast to the less efficient measure of adjusted R^2 in the context of non-normal, heteroscedastic, and influential multivariate data.

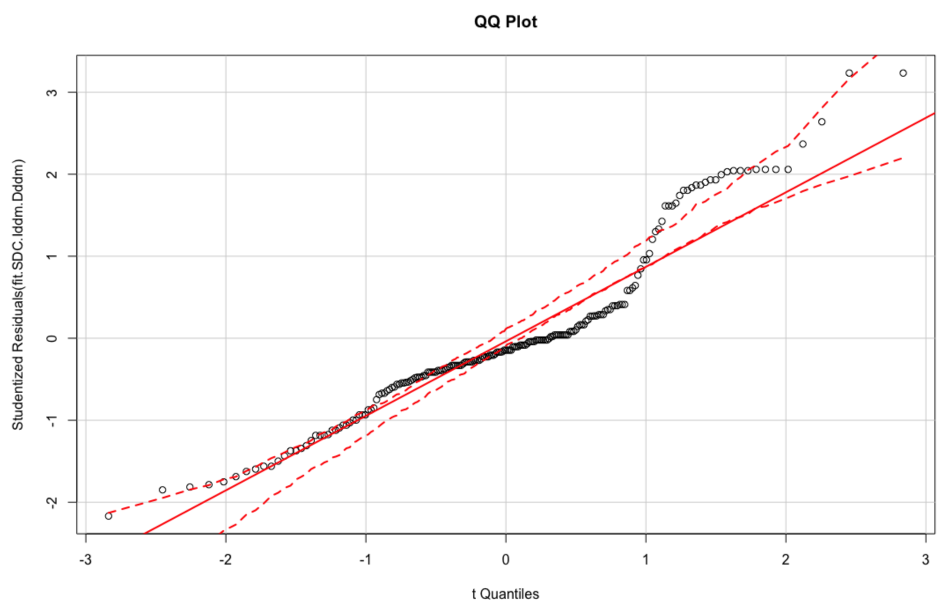

Figure 45 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters.

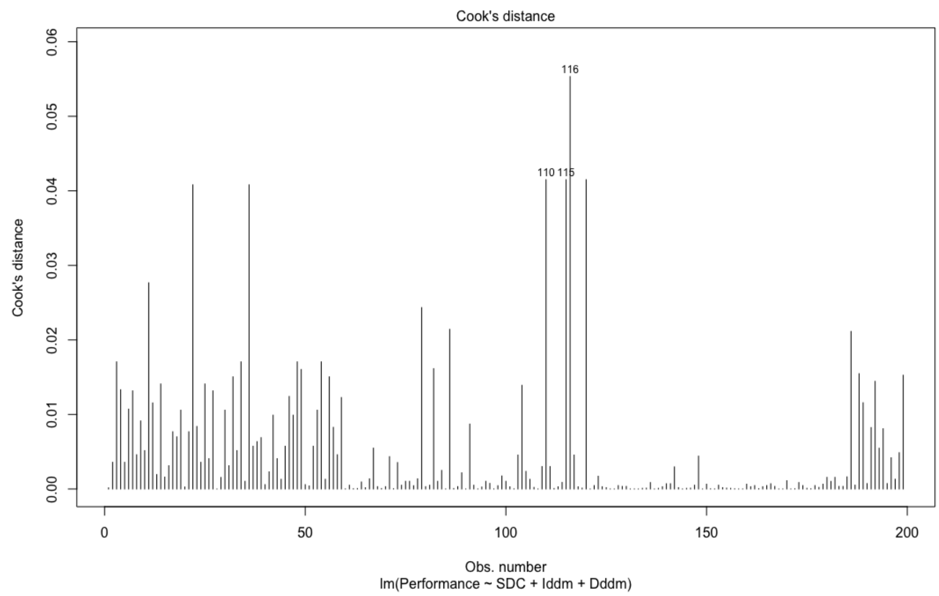

Figure 46 – Cook's distance D plot for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters.

2.3.2.3 Regression model comparison using goodness of fit criteria and multi-model inference using information criteria

Following the model building effort and the exhaustive search through the parameter space, six candidates models for multiple linear regression were retained, as presented in Table 11. The six models contain 3-parameter sets of predictors, and also include a reference model including four parameters[2] for benchmarking and comparison. The MLR models were assessed using three goodness of fit measures: the coefficient of determination (adjusted R^2), the standard error of the regression ($RMSE$), and the Akaike information criterion corrected for small samples ($AICc$) (higher adjusted R^2, and lower $RMSE$ and $AICc$ are better). Table 12 presents the multicollinearity factors for the same candidate models, discussing the variance inflation factor and model tolerance index (lower VIF and higher tolerance are preferrable).

2. Any 4-parameter combination among structural complexity (CC_{DDM}, CNC_{DDM}, or SDC_{DDM}), information complexity (I_{DDM} or A_{DDM}), and D_{DDM}, U_{DDM} or S_{DDM}.

Table 11 – Candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2).

model	**adj.** R^2	**RMSE**	**AICc**
4-parameter models	max. .4634	min. 16.20	min. 1680.56
$SDC + U_{DDM} + S_{DDM}$	.4647	16.18	1678.97
$CC + U_{DDM} + S_{DDM}$	.4635	16.20	1679.45
$CNC + U_{DDM} + S_{DDM}$	.4630	16.21	1679.63
$CC + I_{DDM} + D_{DDM}$	.4607	16.24	1680.48
$CNC + I_{DDM} + D_{DDM}$	.4602	16.25	1680.66
$SDC + I_{DDM} + D_{DDM}$	.4598	16.25	1680.78

Table 12 – Multicollinearity factors for the candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2).

model	**VIF**	**tolerance**
$SDC + U_{DDM} + S_{DDM}$	moderately correlated	high
$CC + U_{DDM} + S_{DDM}$	moderately correlated	high
$CNC + U_{DDM} + S_{DDM}$	moderately correlated	high
$CC + I_{DDM} + D_{DDM}$	highly correlated (I_{DDM}, D_{DDM})	very low
$CNC + I_{DDM} + D_{DDM}$	highly correlated (I_{DDM}, D_{DDM})	very low
$SDC + I_{DDM} + D_{DDM}$	highly correlated (I_{DDM}, D_{DDM})	very low

As previously mentioned, the low tolerance and high variance inflation factors which determine the multicollinearity of the predictors used in the MLR models are only worrisome when such multicollinearity is the result of non-accidental collinearity between independent variables. As such, the intrinsic relationships between SDC_{DDM}, CC_{DDM}, CNC_{DDM}, and W_{DDM} can be said to be non-accidental, the cause of their highly collinear nature, while the relationship between SDC_{DDM} and D_{DDM}, L_{DDM}, U_{DDM}, or S_{DDM} are contingent on the design of the DDM scenarios. Ideally, one should aim to obtain orthogonal predictors, that is, parameters of complexity which are completely non-redundant with one another, but for practical purposes as well as for increased intelligibility, models including structure and information are favored.

Using the adjusted R^2 measure of goodness of fit, the objective models of complexity explain between **45.98%** and **46.47%** of the variance in the performance scores for the five DDM scenarios, for the top six MLR candidate models. This is a positive outcome, given the low measures of variance explained in previous research projects (Pronovost et al, 2014, Pronovost et al, 2015, Lafond et al, 2016), as DDM problems tend to involve substantial proportions of intra- and inter-individual differences, reflected in the statistical analyses as unexplained variance.

The best models use combinations of one of the three methods of computation for the structural complexity C_{DDM} term, paired with either I_{DDM} and D_{DDM} or with U_{DDM} and S_{DDM}. Other combinations of three parameters exhibit a significant drop in goodness of fit, with a ceiling under 42%. All 2-parameter model combinations were tested, but they only produced adjusted R^2 values between 27% and 33%. No gains in goodness of fit above four-parameter models were observed in any model design combinations. That is, no four- or more parameter combinations increased the variance explained in the response variable in a significant way (ceiling at 46%), relative to the three-parameter models presented in Table 11.

Another way to compare the influence of predictors both within- and between-models of MLR is the analysis of the relative importance of regressors in such models. The `relaimpo` package provides measures of relative importance for each of the predictors in the model, expressing the predictors' weight in the balance of variance explained by a complete MLR model. Figure 47 presents the predictors' relative importance for the MLR model featuring the SDC_{DDM}, I_{DDM}, D_{DDM} parameters, in continuity with the graphs presented above. This analysis is generated via a bootstrap validation based on 1,000 samples of the candidate MLR models, using the LMG method, where the R^2 contribution is averaged over orderings among regressors (Lindeman, Merenda & Gold, 1980, Chevan & Sutherland, 1991).

Looking at the complete set of graphs for the relative importance of predictors for the total variance explained for the performance scores in appendix C), the predictors appear to explain anywhere between 15% and 45% of the total R^2 on average. Additional insights are provided through those graphs. For instance, candidate MLR models which feature the U_{DDM} and S_{DDM} parameters minimize the impact of the structural complexity component C_{DDM}, with U_{DDM} looming larger than any other predictor, while models featuring I_{DDM} and D_{DDM} appear to weight their three parameters in a similar proportion between themselves.

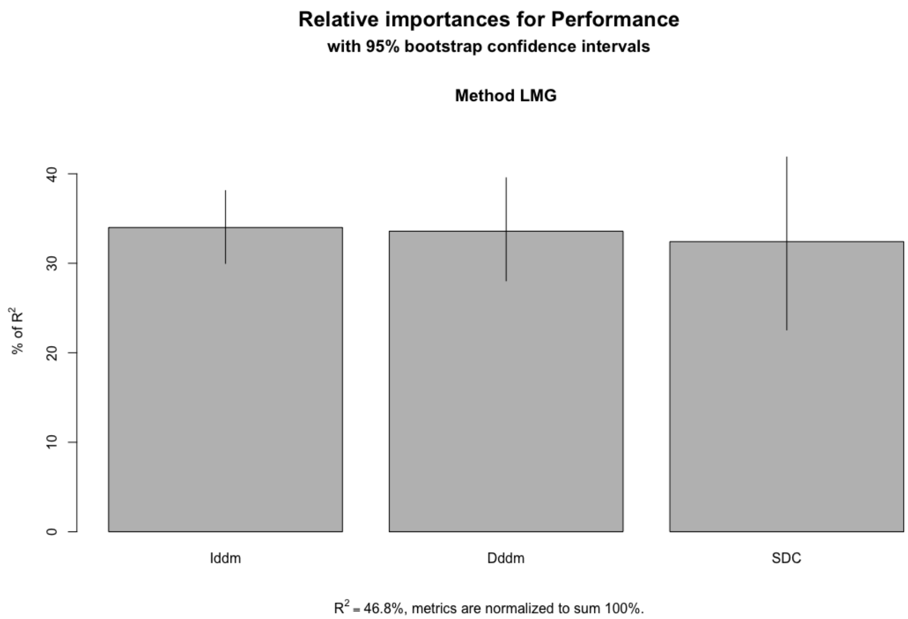

Figure 47 – Relative importance of the model variables for SDC_{DDM}, I_{DDM}, D_{DDM}.

While all six models more or less exhibit similar performances through their goodness of fit indicators, the regression coefficients of the objective models of complexity do not necessarily map well with the intuitions underlying this research project. As can be seen in Table 13, the structural complexity component C_{DDM} yields positive coefficients when combined with U_{DDM} and S_{DDM}, while it is negative when combined with I_{DDM} and D_{DDM}. The latter is far more consistent for our research purposes, as one of the primary hypotheses of this experiment is that an increase in structural complexity would negatively impact human performance in dynamic decision-making problems. The MLR models involving structural complexity, information complexity, and the measure of difficulty will therefore be favored in the remainder of the analyses, as will be discussed at the end of this chapter.

Table 13 – Relationship between candidate MLR model coefficients for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) and the performance measures.

model	**coeff. of 1st parameter**	**coeff. of 2nd parameter**	**coeff. of 3rd parameter**
$SDC + U_{DDM} + S_{DDM}$	positive	negative	negative
$CC + U_{DDM} + S_{DDM}$	positive	negative	negative
$CNC + U_{DDM} + S_{DDM}$	positive	negative	negative
$CC + I_{DDM} + D_{DDM}$	negative	negative	negative
$CNC + I_{DDM} + D_{DDM}$	negative	negative	negative
$SDC + I_{DDM} + D_{DDM}$	negative	negative	negative

2.3.3 Regression model validation

This section presents a validation effort to test the multiple linear regression candidate models for the objective measures of complexity through predictive analytics. A cross-validation is performed using the stratified n-repeated k-fold method, over the five DDM scenarios and a test scenario using the *Cybernetia* scenario of the Ecopolicy serious game.

2.3.3.1 Adding a test scenario outside CODEM: the Ecopolicy serious game

Results from an earlier experiment using the *Cybernetia* scenario of the Ecopolicy serious game are used in this section as a target sample to validate the previous MLR models. Figure 48 presents a box plot graph of the performances for the DDM scenarios including Ecopolicy. We can see that the scores exhibit high variance around the median score of 23.00 (m = 40.69, sd = 29.73, min = 6.67, max = 78.00, compare with the results of Table 9), probably exacerbated by the small sample size (n = 15). A one-way analysis of variance including the Ecopolicy scenario reveals that performances vary significantly across scenarios with $F(5, 208)$ = 30.92, $p < .001$. Scenario performance medians and means are presented together in order to illustrate the discrepancies in distributions for the response variable.

Table 15 presents the objective parameters of complexity updated with the values for the Ecopolicy scenario, and Figure 49 shows the impact of the added parameter values from Ecopolicy on the principal component analysis. In the bivariate plot, the first component differentiates scenarios over the structural complexity and the informational complexity, on the one hand, and difficulty, on the othe hand, in a similar fashion to the PCA with the five DDM scenarios in CODEM. The second component differentiates scenarios according to uncertainty (the Ecopolicy scenario does not have any uncertainty component at all), as well as over nonlinearity and instability scores. Ecopolicy has very low structural complexity like Arctic 1, while exhibiting the lowest action complexity score of all scenarios yet its information complexity ranges between the Arctic 2 and Arctic 3 scenarios. Figures 72 to 81 in appendix B present each of the objective parameters of complexity relative to the DDM scenarios, updated with the inclusion of the Ecopolicy scenario.

While the structural complexity C_{DDM} based on the SDC computation is nearly identical to the other two C_{DDM} metrics in the original PCA, the addition of Ecopolicy exacerbates the differences over SDC between the scenarios, due to the high number of variables that are both endogenous and exogenous in Ecopolicy, relative to the total number of variables factored in the other C_{DDM} metrics. The projection of PCA loadings over the L_{DDM} and S_{DDM} on one vector, and of U_{DDM} on another vector are now segregating between the three Arctic scenarios differently, as Arctic 1 and 3 are now closer in the PCA space relative to two out of those three parameters (nonlinearity and instability). Ecopolicy is far away from all the other scenarios, a phenomenon exacerbated by the null value of Ecopolicy in the uncertainty parameter, and high values of nonlinearity and instability.

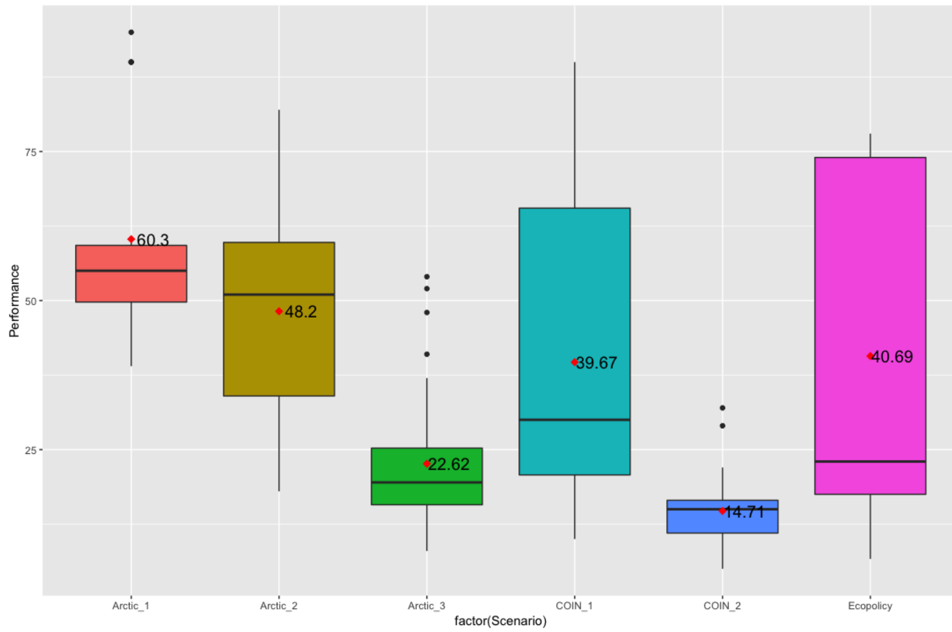

Figure 48 – Performance scores for the DDM scenarios including Ecopolicy.

Table 14 – Descriptive statistics for the Performance metric used as a response variable in the DDM scenarios including Ecopolicy.

		Performance				
Scenario	n	mean	median	sd	min	max
Arctic_1	20	60.30	55.00	16.64	39.00	95.00
Arctic_2	20	48.20	51.00	18.24	18.00	82.00
Arctic_3	40	22.62	19.50	10.70	8.00	54.00
COIN_1	60	39.67	30.00	23.79	10.00	90.00
COIN_2	59	14.71	15.00	5.02	5.00	32.00
Ecopolicy	15	40.69	23.00	29.73	6.67	78.00
All	214	32.40	22.00	22.76	5.00	95.00

Table 15 – Values for the objective measures of complexity in the DDM scenarios, including Ecopolicy.

parameter	Arctic 1	Arctic 2	Arctic 3	COIN 1	COIN 2	Ecopolicy
CC_{DDM}	8	20	32	50	50	9
CNC_{DDM}	24.20	88.17	171.13	277.78	277.78	28.13
SDC_{DDM}	200	756	1600	2128	2128	784
I_{DDM}	36	40	70	378	378	60
A_{DDM}	576	640	1400	8316	7560	480
W_{DDM}	57	88	142	245	245	90
D_{DDM}	.5016	.5016	.5016	.1834	.2072	.5272
L_{DDM}	.0795	.1076	.0913	.0882	.0882	.2007
U_{DDM}	.0944	.0944	.2306	.1957	.2028	.0000
S_{DDM}	.5016	.5443	.5228	.5217	.5657	.5476

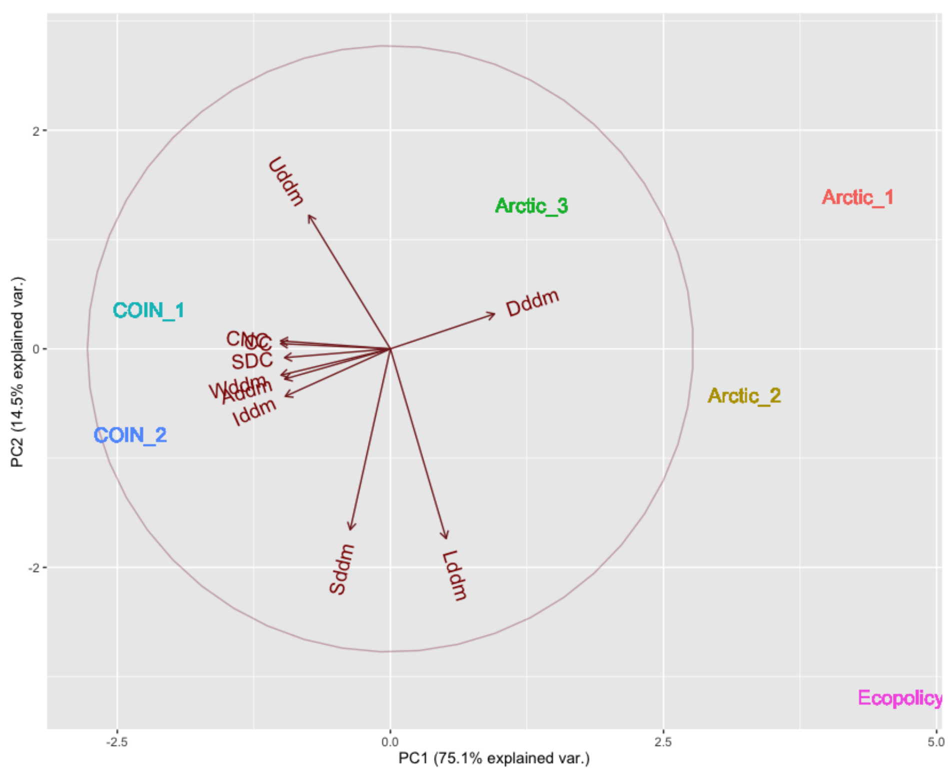

Figure 49 – A bivariate plot representing the principal component analysis loadings of the ten parameters of complexity for the DDM scenarios including Ecopolicy.

2.3.3.2 Cross-validation using the stratified n-repeated k-fold method

A stratified *1,000*-repeated *10*-fold cross-validation (Allen, 1974, Stone, 1974, Geisser, 1975) was conducted in order to observe the performance of the candidate MLR models generated with the five DDM scenarios in CODEM in the context of model predictions, using the Ecopolicy scenario as a novel, in-sample subset of performance data. The stratified n-repeated k-fold method is by definition a non-exhaustive cross-validation approach which does not sample all possible susbets of the observations, yet with a simulation design using 1,000 random repetitions of 10 randomly partitioned subsets of the complete dataset, the results will warrant a rather exhaustive qualification. For each of the 1,000 simulations, one of the k subsets is used as the test set and the other *k-1* subsets are used together to form a training set. The average error is computed across the k trials for each simulation, then for the overall 1,000 repetitions.

Likewise, the variance of the resulting $RMSE$ estimate is reduced as the number of folds k and repetitions n are increased. How the data gets divided through the folds is not particularly important, as each observation is represented in a test set exactly once, and gets to be in a training set *k-1* times. This particular type of n-repeated k-fold cross-validation is said to be stratified by virtue of the folds being selected so that each fold contains roughly the same proportions of observations from the different strata. Here, the strata correspond to the DDM scenario themselves, so that each simulation contains observations from each of the scenario parameter values. This is important given that the set of values of the parameters of complexity for a given scenario are constants, even if they are presumed to be random effects models.

This particular approach was used because the structure and the values of the dataset prohibited a simple out-of-sample approach. Indeed, the dataset prevented any attempt at hold-out sampling, bootstrap, or leave-one-out ($LOOCV$) validation, because of the *rank-deficient fit* of the variance-covariance matrix for the independent variables (Breiman & Spector, 1992, Schaffer, 1993, Kohavi, 1995, Diamantidis, Karlis, & Giakoumakis, 2000, Krstajic, Buturovic, Leahy, & Thomas, 2014). That is, the parameters of complexity for each and all scenarios are a small set of parameters which are held at constant values, relative to the variance in the dependent variable. Moreover, the highly multicollinear nature of the parameters further exacerbates the rank deficiency phenomenon. Adding another scenario with a few constant parameter values does not allow the above-mentioned validation procedures to fit a MLR model to the data, as there is insufficient information contained in the data to estimate such models.

The stratified *1,000*-repeated *10*-fold cross-validation was thus used to hold out random samples belonging to different scenarios, including the Ecopolicy data. As can be seen in Table 16, the objective models of complexity can predict between approximately **37.24%** and **40.49%** of the variance in the performance scores when Ecopolicy is included, for models ranging from

3- to 4-parameter sets. If we focus strictly on the range of variance predicted in performance scores for 3-parameter models, the results vary from **37.24%** to **39.66%**, a decent range of predictive power given the considerable intra- and inter-group variance in performances for DDM scenarios. Chapter 3 presents an alternative regression model validation analysis using the group means and group medians in the human performance data in order to conduct an out-of-sample validation analysis, as a way to mitigate the impact of intra-group variance on the standard error of the regression analyses.

Table 16 – Cross-validation of candidate MLR models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) with the addition of Ecopolicy, using the stratified *1,000*-repeated *10*-fold method.

model	**adj.** R^2	**RMSE**
4-parameter models	~.4049	~17.86
$SDC + U_{DDM} + S_{DDM}$	~.3771	~18.33
$CC + U_{DDM} + S_{DDM}$	~.3916	~18.10
$CNC + U_{DDM} + S_{DDM}$	~.3900	~18.12
$CC + I_{DDM} + D_{DDM}$	~.3966	~17.99
$CNC + I_{DDM} + D_{DDM}$	~.3958	~18.00
$SDC + I_{DDM} + D_{DDM}$	~.3724	~18.33

An analysis of the regression coefficients produced through the stratified *1,000*-repeated *10*-fold validation models reveals that the regression signs remain the same as in the case of the five DDM scenarios from CODEM (as seen in Table 13. The structural complexity component C_{DDM} yields positive coefficients when combined with U_{DDM} and S_{DDM}, while it is negative when combined with I_{DDM} and D_{DDM}. It is therefore corroborating the choice of the MLR models involving structural complexity, information complexity, and the measure of difficulty as more realistic choices in the search for objective models of complexity for dynamic decision-making, as conjectured in the following discussion section. Chapter 3 will present a more nuanced analysis of the regression coefficients though, depending on the type of regression analysis in use, as well as on some transformations performed on the original data.

2.4 Discussion

This chapter presented an effort to quantify objective features of complexity involved in dynamic decision-making scenarios, by sampling the parameter space of complex decision problems as it relates to human performance, in order to retain a minimum of predictors while maximizing the model fit with the variance explained in the response variable. A subset of candidate models were proposed, in the form of multiple linear regression models, selected with the additional constraint of having as little multicollinearity as possible. The candidate models were then tested for their predictive power using a stratified n-repeated k-fold cross-validation method, through the use of an additional dynamic decision-making scenario using the Ecopolicy serious game.

The use of five DDM scenarios in the CODEM simulation environment were found to be implementations of ten parameters of complexity of various magnitudes and relationships, with some of the scenarios scoring high on a few parameters but not on others, as was seen in the table presenting the results of the parameter calculations for each scenario (Table 10), the parallel coordinates chart (Figure 40), and in the bivariate plot featuring the principal component analysis (Figure 41). Six MLR models were retained for their relative index of goodness of fit in explaining the variance in human performance (using the adjusted R^2), the absolute measure of the standard error of the regression ($RMSE$) indicating the magnitude of variance unexplained by each model, as well as the $AICc$ information criterion used to benchmark the loss function of the entropy of MLR models relative to one another. Collinearity indicators were included to assess whether the control variables exhibited a lot of redundancy in their respective models, and it was found that most models did in fact exhibit high levels of multicollinearity.

A regression model validation phase used a stratified *1,000*-repeated *10*-fold cross-validation method with and without an additional DDM scenario in the Ecopolicy serious game, which yielded an overall drop of 6% to 9% in predictive efficiency, based on the new coefficients of determination for the same objective models of complexity. Given the high variance previously observed and expected in DDM performances, due to intra- and inter-individual differences in performance for such complex information processing tasks, it is nevertheless with a positive outlook that such models can be appraised.

2.4.1 Relationships between the objective parameters of complexity

The complexity of DDM problems can be parameterized in many ways. What started as an investigation of the impact of the features of complexity on participant performance in DDM scenarios through the use of compound models (such as the cyclomatic number, the information flow, and the cognitive functional size) led to the adoption of a set of parameters, considered relevant in the scientific literature about dynamic decision-making and task complexity, that are as orthogonal as possible between themselves. The compound models from complexity theory, cognitive informatics, and knowledge engineering involved in business process modeling were decomposed into three common parameters (structure, information, and cognitive weight), to which an additional set of four parameters (difficulty, nonlinearity, uncertainty, and instability) was combined based on more recent literature (Liu & Li, 2011, 2012, 2014, Pronovost et al, 2014, Stouten & Größler, 2017).

That the inclusion of many or all parameters of complexity did not increase the explanatory power of the multiple linear regression models is understandable in light of the information concerning the high degrees of multicollinearity between some of the parameters. This is the case for two different yet complementary reasons, one being theoretical and the other being more practical. Firstly, from a theoretical point of view, some of the parameters may be seen as quasi-linear transformations of other parameters. Such is the case with the cognitive weight parameter, which can be seen as being commensurate with both structural complexity variants (the part concerning the relationships) and the information complexity variants (the part concerning the possible interventions). Those parameters of complexity which exhibit redundant characteristics of DDM scenarios bear an *instrinsic* collinearity and were therefore tested separately from one another, namely the three measures of structural complexity C_{DDM}, and the two measure of information complexity I_{DDM} and A_{DDM}. The cognitive weight complexity W_{DDM} must certainly be appropriate in a different context such as the evalution of the complexity of algorithms (as seen in De Silva & Kodagoda, 2013, De Silva, Kodagoda, & Perera, 2012, De Silva et al, 2013, and Kinsner, 2010), but in the current research project, this metric was removed from the final pool of parameters among the six subsets which were fed to the exhaustive search algorithms. The measure of nonlinearity L_{DDM} among variable relationships and for the effects of interventions was simply unusable in the context of this particular dataset, as the highest difference between the scenarios for that parameter was merely 3% in magnitude, and was very weakly correlated with the response variable.

The characterization of the various parameters of complexity as seen in the methodology section emphasized that some of the parameters were magnitudes (scalars) while other were relationships (ratios). The particular case of the uncertainty parameter, being both a scalar- and a ratio-based metric, also features a similar issue in that the uncertainty is always about the other magnitude metrics. That is, uncertainty about variables, relationships, or interventions requires having a certain number of variables, relationships, and interventions. This is related to the second phenomenon concerning the practical aspect of building DDM scenarios of various degrees of complexity, which generally involves scaling (i) the possible interventions and their related effects, (ii) the presence of nonlinear relationships, (iii) a number of uncertain features including scenario events, (iv) more difficult goal attainment thresholds, as well as (v) varying degrees of system stability, in a certain proportion with the structural complexity of the scenario. This tendency for *accidental* collinearity can be seen in the contrast between the Arctic 3 and the COIN 2 scenarios, as many of the parameters for those DDM scenarios exhibit high multicollinearity. Those scenarios are intended to be particularly challenging for the participants, after all. Where Arctic 3 and COIN 2 mostly differ is captured in their ratings on information complexity I_{DDM} and the degree of difficulty D_{DDM}. In summary, this means that the parameter space for the DDM scenarios exhibit high collinearity among their respective parameter scores, but in principle, such collinearity could be much lower in future DDM scenario designs. Nothing prevents controlling for such shortcomings in practice, as scenarios could have different ratios to compare among themselves for the parameters expressing relationships. A scenario with high information complexity (similar to COIN 1 and 2) could also exhibit high difficulty (similar to Arctic 1, 2, and 3) while having a very small score on structural complexity, for example.

2.4.2 Relationships between DDM performance and the objective measures of complexity

The original composite models involved too many parameters which exhibited high degrees of dependence, and for this reason, they were broken down into a smaller feature set of more orthogonal parameters of complex decision problems. The model selection to explain as much variance in the response variable was conducted through goodness of fit indicators and multicollinearity indicators, with the aim of maximizing the former and minimizing the latter. Based on the insights from the results section, it became more apparent that differentiating between similar scores in Arctic 2, COIN 1 and Ecopolicy (scenarios of considerable differences in most parameters of complexity) on the one hand, and different scores in Arctic 3 and COIN 1 (scenarios of substantial differences in scores which are nevertheless 'closer' in the parameter space of complexity, relative to simpler scenarios such as SpaceLab and Arctic 1), on the other hand, required a complex interaction of parameters which were not governed strictly by the magnitude of parameters such as the variants of structural and information complexity metrics.

In light of the information contained in the parameter values matrix (Table 15), the parallel coordinates chart for the ten parameters of complexity (Figure 40), and the bivariate plot illustrating the relationship between the PCA loadings and the DDM scenarios (Figure 49), it becomes clearer that goal attainment (the objective measure of difficulty D_{DDM}) determines the distribution of performance scores just as much as are the parameters of complexity reflecting the underlying structure and information contained in the DDM scenarios. The first component of the PCA explained 75% of the variance in the parameter space based in great part on contrasts between structural and information components on the one hand, and based on difficulty on the other hand. COIN scenarios yield higher structure and information ratings than the Arctic scenarios and Ecopolicy, while the opposite is true with regards to difficulty. The second component of the PCA captured 15% of the remaining differences between scenario parameters, differentiating between COIN 1, COIN 2, and Ecopolicy based on nonlinearity, uncertainty, and instability. A score of 50% in COIN 1 may appear easier to obtain than in Arctic 3, even with significantly higher scores in structural and information complexity, but the measures of difficulty and uncertainty make the Arctic 3 scenario fare more difficult to tackle. Similarly, while a score of 50% in Ecopolicy may appear to be as hard to get as it is for COIN 1 even if the latter yields substantially larger scores of structure and information complexity, it can be observed that Ecopolicy has the highest goal attainment (difficulty) score of all DDM scenarios, as well as the largest measure of nonlinearity in relationships and for influences from interventions.

The complex relationships of the response variable with some of the objective parameters of complexity require some work of interpretation in order to provide insights as to what is going on in the black box that is the mental model of a human decision maker who has to identify (*knowledge acquisition*) and control (*knowledge application*) a complex and dynamic system (Funke, 2001). For example, the relationship between the variant of the information complexity named the *action complexity*, or A_{DDM}, and the response variable is quite peculiar in that higher action complexity A_{DDM} tends to increase DDM performance if we compare the COIN 1 and COIN 2 scenarios. The relationship between the action complexity and human performance is counterintuitive at first glance because it would suggest that more inputs and outputs in the information flow of a complex and dynamic decision problem would facilitate performance. Our immediate intuitions about complexity would urge us to see some inverse proportionality between additional information to process and a measure of success in understanding and controlling a system. But when we break down the details of the action complexity parameter, we can observe that action complexity only differs from the information complexity metric insofar as it includes the pool of action points in its computation.

The relationship between action complexity and performance could therefore be construed in a different way, if we interpret the components of A_{DDM} as the gross potential pool of inputs and outcomes which may help a decision maker to understand and influence the current and future states of a discrete, turn-based complex decision problem such as the ones embodied in dynamic decision-making scenarios. That is, rather than envisioning the action complexity as a burden on the mental model of decision makers, it may be the case that, up to a certain threshold (the relationship is not expected to be entirely linear), the inputs-and-outcomes component of DDM scenarios supports the goals of the participants, enabling them to achieve their ends. All other things being unequal (and hard to interpret), it may be the case that having the possibility of influencing a complex and dynamic system through more interventions and with a larger 'budget' of action points may support the decision makers, instead of hindering them.

Significant differences in human performance scores between COIN 1 and COIN 2 are mainly explained by the increase in action complexity in COIN 1, while the measures of difficulty, uncertainty, and instability are higher in COIN 2. The final MLR model selection completely ignores the action complexity in favor of the original information complexity variant I_{DDM}, so action complexity A_{DDM} will not be revisited in the remainder of the next chapter. An analysis of the regression coefficients for the the candidate MLR models featuring a combination of structure, uncertainty, and instability revealed that those models bear a positive coefficient for the structural complexity (either CC_{DDM}, CNC_{DDM}, or SDC_{DDM}). This counterintuitive factor will be explored in more detail in chapter 3, but suggest that the MLR models combining structure, information, and difficulty are more likely to represent the best candidates of objective measures of complexity for DDM performance. Ideally, in a future research program, all parameters would be expected to impact performance, given an arbitrarily large number of scenarios with different parameter values and a sufficiently large sample of participants. Those parameters would probably do so with a differential impact on DDM performance, as the influence of some parameters may loom larger than others, as seen in our results. Therefore, whether some of the parameters were retained and excluded in our candidate MLR models, they should still be accounted for in future DDM experimentation.

2.4.3 Limitations and way ahead

A number of theoretical and practical limitations need to be accounted for in the pursuit of an objective model of complexity for dynamic decision-making. The high inter- and intra-individual variability associated with the many effects of low- (perception, working memory) to high-level cognitive processes (mental models, analytical skills, numeracy, etc.) pose a considerable challenge in the manipulation of the impact of control variables on our response variable. Inter-individual differences in performance have been observed to produce non-normal probability distributions for DDM scenario performances when the overall complexity of scenarios is high, such as in Arctic 3 and the COIN scenarios (Pronovost et al, 2014), and the best predictor of human performance in the context of tutoring trainees about systems thinking is a behavioral measure, namely the time spent seeking information about relationships and about feedback on past decisions (Karakul & Qudrat-Ullah, 2008, Lafond et al, 2012, Pronovost et al, 2015).

The above-mentioned constraints support the idea that the parameter space for the five DDM scenarios in CODEM exhibit high degrees of multicollinearity among their respective parameter scores, but that in principle, such scores could be much lower in future DDM scenario designs, where multiple scenarios would be available, providing additional data points for those random effects factors to compare with human performance. The non-normality and heteroscedasticity of residuals, combined with the presence of multivariate outliers or influential observations, suggest that other modeling approaches besides multiple linear regression using the ordinary least squares paradigm could potentially yield lower parameter bias as well as lower variance in the residuals. Pronovost et al (2014) report the possibility that the combined effect of parameters of complexity and difficulty on DDM performance follows a relationship curve mirroring a *just-noticeable difference* threshold as formulated by the Weber–Fechner law (Fechner, 1860). The hypothesis is that beyond a particular degree of complexity and/or difficulty (the differential *limen*, or threshold, as formulated in psychophysics), decision-makers would no longer be able to achieve any gains in performance, in the likeness of a monotonic curve, such as an inverse scale or a model similar to an exponential decay function.

In closing remarks, with peak coefficient of determination values around approximately 46% for the five DDM scenarios in CODEM, the subset of 3-parameter candidate models of complexity for DDM appear to be decent candidates to explain a substantial proportion of the variance in human performance. The cross-validation endeavor, although marred by small variations in the set of random effects models, also suggested that the 3-parameter MLR models were able to generalize to subsets of random data when some observations from a test scenario were withheld, with coefficients of determination ranging between 37% and 40%. Two major questions remain, in our humble view:

Firstly, are those models accurate enough to account for human performance in DDM, and could they be undermined by the violated assumptions of the ordinary least squares method to multiple linear regression analysis? Are there potential ways to improve the goodness of fit and the predictive power of the objective measures of complexity in order to account for the human performance data? The next chapter will address this question by comparing our MLR models with results generated through robust methods of non-parametric regression, with MLR models using mean and median group performances, and with nonlinear approaches using machine learning algorithms.

Secondly, which model is preferable? Is there a "right" one, above others? Since many candidate MLR models account for a decent proportion of variance explained, in part due to high collinearity between parameters on the one hand, yet they are also sensitive to minute differences in such parameters across scenarios such as COIN 1 and COIN 2 on the other hand, it would be disingenuous to favor one set of parameters from our six candidate models, and exclude others as inconsequential. There can be no definitive answer to this question at this stage, although the models involving structural complexity (either CC_{DDM}, CNC_{DDM}, or SDC_{DDM}), information complexity (I_{DDM}), and the measure of difficulty (D_{DDM}) are promising, given the high goodness of fit values and compelling regression coefficients, relative to the models involving uncertainty and instability among the six candidate MLR models. A research program involving a succession of experimental conditions, whereby an array of DDM scenarios would implement various configurations of the feature space of the objective parameters of complexity could answer this question. The final chapter addresses this issue by laying out the details of such a research program.

Chapter 3

Nonlinear, Non-Parametric, and Robust Methods to Model Objective Measures of Complexity for Dynamic Decision-Making

3.1 Introduction

This chapter presents an evaluation of the objective models of complexity for dynamic decision-making scenarios produced through the traditional multiple linear regression approach. We present a preliminary discussion on the bias-variance trade-off in experimental cognitive psychology and the cognitive sciences in general, whereby we argue that the ordinary least squares paradigm in linear regression is not the most appropriate measure of relationships between the predictors of complexity for dynamic decision-making and the human performance data. We then introduce three different analyses in order to compare the results of the original MLR analysis from the previous chapter with alternative methods of computation: (i) the statistical methods of robust and resistant regression are considered, in light of the violated assumption of standard linear regression with regards to the non-normality of multivariate residuals, the non-constant error variance, and the presence of influential observations affecting the candidate regression models for our objective measures of complexity; (ii) a standard MLR approach using only group means and medians is assessed, whereby the intra-group variance is completely eliminated from the regression analysis; and finally (iii) an exploration of alternative methods for regression modeling, analysis, and prediction drawing on advances in computational learning theory with regards to machine learning algorithms used for classification and regression. The discussion section compares the results from the MLR models with the analyses of the same models using the alternative methods, from the point of view of the bias-variance trade-off. Issues such as model complexity, interpretability, and utility are discussed, as well as concerns about underfitting and overfitting models based on the sources of error and on the assumptions underlying the model building endeavor.

3.1.1 Coping with the bias-variance trade-off: differences for micro- and macrocognitive phenomena

The previous section discussed the interest in finding a regression model using the traditional linear modeling approach to regression analysis, and the results produced a small set of candidate MLR models composed of three parameters, to explain human performance in dynamic decision-making scenarios. Those DDM scenarios are instances of a particular combination of values in the parameter space of each respective objective characteristic of complexity of the DDM task model, sampled among a much larger combinatorial range of possible scenarios. Therefore, the DDM scenarios themselves are merely a scarce number of points for random effects factors, which must be sampled in a larger research program, in order to properly assess the impact of objective parameters of complexity on human performance.

As mentioned in chapter 1, cognitive psychology is interested at the level of description of microcognition, i.e, that of attention, memory, and perception, and typically endeavors to produce simple models of regression analysis to model universal processes pitched at a low level of information processing, while the cognitive functions which depends on those processes, such as decision-making, problem-solving, the acquisition and comprehension of language, etc., exhibit much higher degrees of variance within and between individuals (Tyler, 1947, Carroll, 1993, Gruszka, Matthews, & Szymura, 2010). Complex, possibly nonlinear relationships exhibiting high variance in behavioral measures such as DDM performance are thus likely to require different methods for quantifying, analyzing, and interpreting phenomena which departs from the microcognitive level of description.

Those conjectures have been supported by previous research on the relationship between dynamic decision-making scenarios of varying degrees of complexity and difficulty on human performance (Pronovost et al, 2014, Pronovost et al, 2015, Lafond et al, 2016), with a working hypothesis that beyond a particular threshold of complexity and/or difficulty, the decision-maker would no longer be able to achieve success in DDM problems. A perceptual threshold is a more universal characteristic of human cognition than a macrocognitive process involving systems thinking. More fundamental information processing functions involved in perception, memory, and attention are processed at a fundamentally lower "band" of human information processing, in reference to Newell's (1990) "bands" of cognition (refer to Figure 11 in the introduction). Those microcognitive processes occur at the cognitive scale (hundreds of milliseconds), whereas the metacognitive skills involved in DDM tasks operate at different time scales situated at the rational level (minutes to hours, Anderson 1990, 2002, 2007).

Cognition, from the low-level processes of perception, memory, and attention to the high-level functions of decision-making, language comprehension, and numeracy, therefore exhibits intra- and inter-individual differences at multiple levels of description, in proportionality with the complexity of the information processing strata as illustrated by Newell's levels of analysis. The lower-level, more universal cognitive processes are expected to exhibit less variability between individuals, while the higher-level cognitive functions studied in the macrocognitive paradigm are expected to exhibit much higher degrees of variability in the population. Macrocognition, naturalistic decision-making, and simulation-based experimentation aim to find patterns and regularities at the rational and the social-organizational bands of cognition, patterns which are hard to reduce to microcognitive theories due in part to the complexities involved in quantifying and measuring cognitive activity over those time scales (Schoelles, Neth, Myers, & Gray, 2006, Myers, Gluck, Gunzelmann, & Kruskmark, 2010, Gray, 2012).

We conjecture that a similar assumption could be construed with regards to the accuracy of information processing responses, beyond the consideration of information processing latencies. That is, the accuracy of information processing responses (here, human performance in dynamic decision-making problems) relative to the levels of analysis of cognition exhibit increasingly higher intra- and inter-individual differences than their lower information processing stages. The following section presents alternative quantitative methods to tackle statistical distributions and estimators which may enhance our understanding of the relationship between the objective parameters of complexity and human performance in dynamic decision-making problems.

3.1.2 Nonlinear and non-parametric methods for classification and regression in computational learning theory

The high inter- and intra-individual variability associated with the many effects of low- (perception, working memory) to high-level cognitive processes (mental models, analytical skills, numeracy, etc.) pose a considerable challenge in the manipulation of the impact of control variables on our response variable. How could we assess the accuracy of the MLR models to account for human performance in DDM, and could they be undermined by the violated assumptions of the OLS method to multiple linear regression analysis? Are there potential ways to improve the goodness of fit and the predictive power of the objective measures of complexity in order to account for the human performance data? The non-normality and heteroscedasticity of residuals, combined with the presence of multivariate outliers or influential observations, suggest that other modeling approaches besides multiple linear regression using the ordinary least squares paradigm could potentially yield lower parameter bias as well as lower variance in the residuals. We propose to explore the bias–variance trade-off in statistical models of DDM through the robust methods of non-parametric regression, through the use of error-reduction strategies focusing on the inter-group differences in performance, as well as through the use of nonlinear and non-parametric methods from the literature on statistical machine learning.

The branch of artificial intelligence named *computational learning theory* is interested in the design and analysis of machine learning algorithms which, combined with pattern recognition algorithms, yield supervised methods (models trained over test sets) of recognizing relationships among inputs and outputs (Russell & Norvig, 2010). In the context of our research, the parameters of complexity composing a MLR model are fed to various algorithms which have different constraints in the way they model variable relationships and statistical error. They do so by virtue of flexibility in the postulates or assumptions they necessitate to produce an objective function, such as relaxing the assumptions concerning the shape of distributions and residuals (e.g., normality, homoscedasticity, influence of observations), assumptions about the functional form of the relationship between variables, or even specifying different boundaries and criteria in the error function to be minimized.

In the present chapter, mathematical optimization techniques are thus explored in order to minimize the bias-variance dilemma imposed by the highly sensitive performance measures of DDM. The enterprise is similar in spirit to the adoption of generalized linear model (GLM) methods such as the iteratively reweighted least squares (IRLS), but uses instead the sophisticated computational toolbox of the machine learning approach. We will compare the OLS approach with five nonlinear and non-parametric methods, ranging from the use of regression splines (partial/piecewise linear functions), local regression using smoothing (weighted interpolation) functions, metric-dependent similarity functions based on explicit feature space computations (such as the k-nearest neighbor approach), kernel methods (implicit mapping of feature spaces to higher-order dimensions to facilitate linear separation), and nonlinear parameter transformation through intermediate functions.

3.2 Method

3.2.1 Robust regression models

The ordinary least squares (OLS) method to calculate a model's residuals is the standard algorithm to fit a multiple linear regression model, which can exhibit the desirable property of having *uniformly minimum-variance unbiased estimators*, that is, a linear regression model that has lower variance than any other unbiased estimator for all possible values of the parameters (Rao, 1965). Yet this is only achievable *if and only if* the predictors are not collinear, and the residuals exhibit homoscedasticity. Pending the additional condition that those residuals are normally distributed, the ordinary least squares method for MLR also yields the maximum likelihood estimator for a response variable, that is, if all of the above-mentioned conditions are met, the OLS computation for a MLR model maximizes the likelihood of making those observations, given the parameters.

The original data concerning our candidate MLR models for objective measures of complexity explaining human performance in DDM scenarios unfortunately features neither homoscedascticity nor normality of residuals, and many of the parameters of complexity exhibit high degrees of multicollinearity, whether it is accidental (due to practical issues such as the design of the scenarios) or intrinsic (due to quasi-linear transformations of some of the features of complexity) in the parameters' calculation. The nonlinear and non-parametric approaches presented in the machine learning section use a number of strategies to cope with the possibility that the functional form of the regression between the objective parameters of complexity and the performance may not be linear (Kuhn & Johnson, 2013). The results section below presents the impact of exploiting such nonlinear approaches to assess whether the candidate MLR models can be improved on, through the adoption of methods ranging from segmented regression to kernel smoothing for function approximation, to feature mapping kernels in order to solve regression problems in high-dimensional spaces. But as will be discussed throughout this chapter, more complex models are not to be preferred to simpler, more parsimonious models, especially if the simpler model's fit is a reasonable approximation given an arbitrary criterion such as the adjusted R^2 or the $RMSE$.

What if we employed a regression method which could model linear relationships between predictors and the response variable without sacrificing all the assumptions of the OLS approach to regression analysis? In the presence of influential observations in the data, a metric less sensitive to such outliers may be used in combination with the least squares approach. So-called *robust regression* methods employ weighted least squares to minimize the impact of the influential outliers on the linear trend of a MLR model. Similar to the support vector regression method presented in the nonlinear and non-parametric section below, there are ways to calculate a model's residuals that are more or less sensitive to observations which stand at various distances from the regression line (Figure 53). The first method of robust regression was introduced by Huber (1973), and the family of robust methods for regression analysis all share the common feature of dampening the effect of non-constant error variance (heteroscedasticity) and multivariate outliers (influential observations) in the calculation of the residual sum of squares used for the linear regression analysis.

3.2.1.1 M-estimation, S-estimation, and MM-estimation

The intuition behind the robust regression approach is based on weighting the observations relative to their contribution to the sum of squared residuals and to the unexplained variance in a regression model. The most elementary way of decreasing sensitivity to observations with larger contributions to the variance is to use the least absolute errors (LAR, also known as least absolute deviations) for all observations, instead of the squares. The main issue of the LAR method is that there is no analytical solution for its calculation, so robust regression methods favor a hybrid computation of residuals whereby least absolute deviations are used for distant observations, while keeping the OLS computation for observations closer to the regression line. An iterative approach is required to calculate the optimal contribution of residuals relative to their distance and weight their impact according to threshold values. There are three common methods of robust regression involving this so-called *iteratively reweighted least squares* (IRLS) instead of the OLS computation (Draper & Smith, 1998, Susanti & Pratiwi, 2014). Huber's (1973) method, called **M-estimation**, uses a maximum likelihood-type function in order to find the minimum of the residuals in the error function of a regression analysis. Calculated as the zero of the derivative of a likelihood function for a given set of parameters in a MLR model, Huber's M-estimation yields an unbiased estimator which also exhibits minimum variance. However, while Huber's M-estimation is resistant to outliers in the response variable, it is susceptible to leverage points in the explanatory variables.

Another group of robust regression methods deals with multivaritate outliers in a different way, by trimming the dataset of influential observations in the distribution. The least trimmed squares (LTS) minimizes the sum of squared residuals over a subset of points, based on a set of ordered absolute values of the residuals (Rousseeuw, 1984, Rousseeuw & Leroy, 1987). The Kendall-Theil–Sen estimator, also named the single median method (Theil, 1950, Sen, 1968), selects the median slope among all lines through pairs of the two-dimensional space of the regression between predictors and the response variable. It is particularly effective for data featuring high degrees of skewness and heteroscedasticity. The most popular parameter estimation method in this family is the **S-estimation** (Rousseeuw & Yohai, 1984, Rousseeuw & Leroy, 1987, Salibian & Yohai, 2006), which uses a hyperplane to reduce the robust estimate of the scaled residuals. Whereas the M-estimator approach is not resistant to outliers in the explanatory variables, the S-estimator uses the residual standard deviation to overcome such a weakness. By determining a minimum robust scale estimator, the S-estimation is said to be a *resistant regression* method, as it is both resistant to leverage points in the explanatory variables and robust to multivariate outliers involving the response variable. Resistant regression methods are efficient in dealing with datasets where a number of undesirable observations are expected, as they give no weight to such observations in the regression analysis.

The third method, used herein, is the **MM-estimation**, a combining the benefits of both M-estimation and S-estimation (Yohai, 1987, Croux, Dhaene, & Hoorelbeke, 2003, Koller & Stahel, 2011). This algorithm first minimizes the scale of the residuals from the M-estimation using the resistant regression method of the S-estimation, and then proceeds in a second moment with the robust regression using the iteratively reweighted least squares computation of the M-estimator. MM-estimation is said to combine statistical *efficiency* (the measure of quality of an estimator, here using the minimization of the $RMSE$ as an indicator) with a *high breakdown point* (i.e., the proportion of outliers an estimator can handle before these influential observations affect the model. Maronna, R. A., Martin, R. D., & Yohai, 2006).

3.2.2 Regression analysis using group means and group medians

In the presence of non-normal and heteroscedastic data in the response variable, as well as of non-normality, multicollinearity, and non-constant error variance in the residuals of the candidate MLR models, one strategy that could boost the goodness of fit indicators is to perform multiple linear regression analyses and validation over the measures of central tendency for each group involved in the response variable. This has the immediate advantage of focusing strictly on inter-group differences (or between-group effects), by effectively eliminating the source of variance imputable to intra-group differences (or within-group effects). This type of calculation will favor a regression analysis based on distances to the regression line for the combined effect of the independent variables on the group performances, represented by as many (as few) points as there are dynamic decision-making scenarios. In the current context, this means that the group mean performances are a set of five points only! This has the immediate consequence of diminishing the ratio of cases-to-independent variables, another important requirement in the conduct of multiple linear regression analysis in order to obtain reliable goodness of fit indicators (Green, 1991, Khamis & Kepler, 2010, Tabachnick & Fidell, 2012).

Running a multiple linear regression analysis using group mean performances thus uses less data points, which may lead to the under-determination and misrepresentation of the original statistical distributions, in order to focus on the central tendencies. Using a few independent parameters of DDM complexity as regression parameters for such a limited subset of observations may lead to overfitting (Tabachnick & Fidell, 2012). The suggested analyses presented herein will therefore also use the group median performances, in order to observe model fit under less stringent assumptions. We have evoked on a few occasions in the previous chapter that the discrepancies between group means and medians were considerable (refer to Figures 37 and 48). In light of the regression diagnostics in chapter 2 as well as the discussion on the benefits of using non-parametric estimators with higher breakdown points in the section concerning robust regression analyses above, the group median performances will therefore be compared to the group mean performances, in order to assess which of the MLR models produce better goodness of fit indicators.

Furthermore, this type of analysis favors a holdout sample validation approach, as it was found that the rank-deficient fit of the variance-covariance matrix for the independent variables no longer affects the regression analyses. The results section will therefore detail the use the group mean and median performances from the scenarios in CODEM as a sample to validate the regression models, using the Ecopolicy data as an out-of-sample test set.

3.2.3 Nonlinear and non-parametric methods: the machine learning approach

Choosing between models and methods suitable for regression analysis among more sophisticated and flexible algorithms requires thoughtful consideration of the benefits and drawbacks of each computational approach to linking predictors to the response variable. Kuhn and Johnson (2013) suggest a three-stage process in order to sample the recipes available in the machine learning literature, based on the characteristics of the dataset and the type of research question(s) guiding our methodology:

> — Start with several models that are the least interpretable and most flexible, such as boosted trees or support vector machines. Across many problem domains, these models have a high likelihood of producing the empirically optimum results (i.e., most accurate).
>
> — Investigate simpler models that are less opaque (e.g., not complete black boxes), such as multivariate adaptive regression splines (MARS), partial least squares, generalized additive models, or naive Bayes models.
>
> — Consider using the simplest model that reasonably approximates the performance of the more complex methods.

As it will be presented in the results section below, opting for multiple models and comparing them on accuracy measures such as the proportion of explained variance (the adjusted R^2) or the standard error of the regression (the $RMSE$), the models reach an area which could be described as a performance ceiling, in an asymptotic fashion. For regression models produced by two arbitrarily complex algorithms, it is thus preferable to choose a simpler model which yields reasonably similar results to a more complex model, for the sake of interpretability and reproducibility. Some of the machine learning algorithms presented below exhibit considerable computational complexity, and the regression models they produce would be hard to interpret, even if they could potentially yield the highest model performances. For all of the abovementioned reasons, it is therefore not the strict criterion of model accuracy in explaining the variance of the human performance data that has to be considered in order to commit to a regression model over another.

3.2.3.1 MARS: multivariate adaptive regression splines

The multivariate adaptive regression splines, or MARS method (Friedman, 1991, Hastie, Tibshirani, & Friedman, 2009, Zhang & Singer, 2010) is a non-parametric type of regression where the structure of the observations from predictors and the response variable is modeled according to their nonlinear relationships. Such nonlinearities are captured by the algorithm as hinge functions, or "kinks" in a linear relationship. MARS is a segmented regression approach, or piecewise linear modeling method, as it attempts to retain linear relationships between predictors and the response variable up to a certain threshold determined by tuning parameters (see Figure 50 for a low-dimensional example of segmented regression using MARS).

MARS uses transformed predictors, or "surrogate features" instead of the original predictors to be entered in a stepwise regression, which involves both a forward and a backward pass, in order to reduce its residual error in an incremental and exhaustive process. So-called hinge functions are evaluated with each additional observation being considered as a potential cutting point, and the choice of which combination of surrogate features (the transformed predictors) with the potential cut point is more likely to minimize the model error is made. If the utility of a hinge function is lower than the standard linear regression model (featuring no hinge), then the linear model is preferred for this subset of the data points. After the complete set of features has been generated, the stepwise function eliminates individual features which do not significantly improve the model's error reduction.

There are two tuning parameters: the degree of added features (here, a single predictor at a time in the stepwise regression), and the number of retained features (automated through the pruning procedure which compares models for the error reduction function). The MARS approach offers considerable advantages over the traditional MLR approach if the modeler suspects that there might be some nonlinear relationships among the multivariate data: MARS is flexible, simple to interpret, and requires few transformations on the original dataset. This segmented, recursive partitioning regression approach also operates as a variable selection algorithm, which is not particularly desirable in the given context, as the variable selection was already done through careful considerations concerning accidental and intrinsic collinearity amongst the objective parameters of complexity. MARS tends to offer a lower variance with few additional costs in terms of bias, as the method considers the differences between models with and without hinge functions, assessed through the cost function used for error reduction.

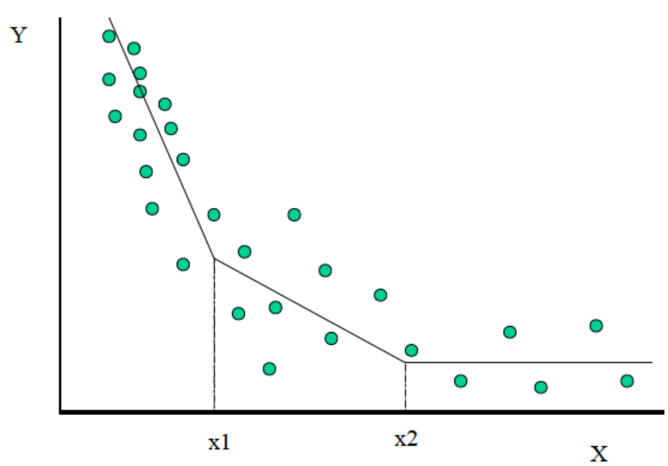

Figure 50 – An example of piecewise linear modeling in the MARS algorithm (from Briand, Freimut, & Vollei, 2004).

3.2.3.2 SVMr: support vector machine regression

Support vector machines are a useful technique to commonly employed for data classification, which comes from statistical learning theory (Boser, Guyon, & Vapnik, 1992, Cortes & Vapnik, 1995, Vapnik, 1998). SVMs use a mapping of the complex relationships between observations to a higher-dimensional *feature space* using hyperplanes (see Figure 51), where further analyses can be conducted in order to solve classification or regression problems. Such mappings are done through a transformation by means of *kernel functions*. The kernel function maps the input space by calculating the inner product (or dot product, also scalar product) of the feature space into a higher-dimensional space by means of a linear kernel, but there are also polynomial, radial basis (RBF), and sigmoid types of kernel functions.

For SVM applied to regression functions, the algorithm aims to find the support vectors in the observations which best model $y = f(x) + error$ where $f(x)$ is the regression function, while simultaneously sampling the space of the error function by sequential optimization. SVM regression is similar to the non-parametric approach of robust regression (discussed in detail in the previous section), where the regression algorithm aims to suppress the effect of highly influential outliers on the regression model (see Figure 52). Minimizing the sum of squared errors (or residual sum of squares) is affected by the distance to the regression line, using the ordinary least squares (OLS) approach. The SVM method uses a residual function that defines a threshold beyond which only the (non-squared) values of observations which exhibit very large residual error contribute to the linear regression equation (see Figure 53). The effect is twofold: firstly, the non-squared residuals yield less influence on the regression equation for outliers overall, and secondly, the model is actually built based on the residuals of those observations which yield higher residual values, not on the observation which fit the model well (i.e., those with small residuals). This threshold value defines the ϵ-insensitive SVM method for regression, while another tuning parameter (called a *cost* parameter) penalizes obervations yielding large residuals.

Parameter estimates for the SVMr model are a function of a set of unknown parameters α_i, where there are as many α parameters as there are observations. In effect, *all* the observations *are* the support vectors for subsets of data points, which use the error reduction function and the cost function in order to reduce an over-parameterized model to a small set of parameters which constitute the final support vectors. Observations yielding residual values within the $\pm\,\epsilon$ range of the error function possess a α_i value of 0 as parameters, and are excluded from the regression model. Consequently, only observations outside the range of ϵ in the error reduction function which have also not been discounted by the cost parameter are included in the final model, and those observations are the parameters called the support vectors of the regression. The cost parameter governs the model building complexity, as large costs yield more flexible models with higher variance (prone to overfitting), while small costs yield stricter models with higher bias (prone to underfitting).

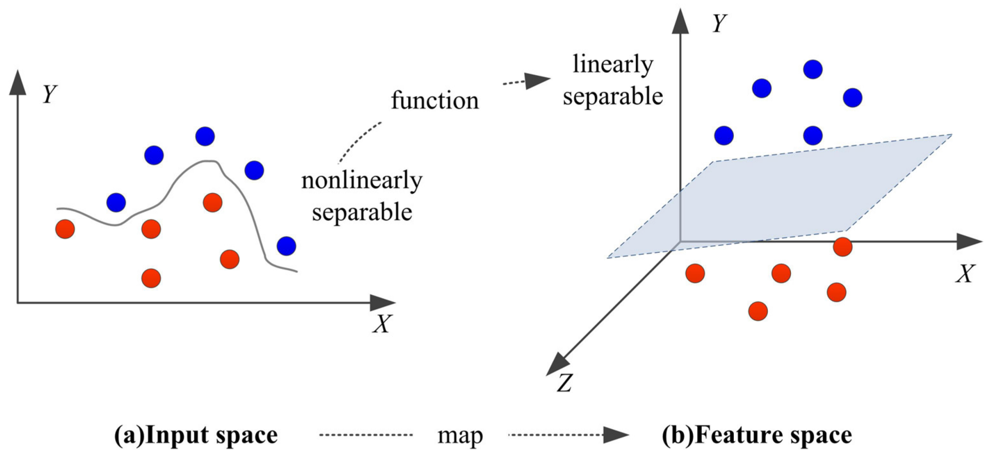

Figure 51 – An example of SVM classification. Mapping a 2-dimensional nonlinear categorization function to a linearly separable categorization in 3-dimensional space.

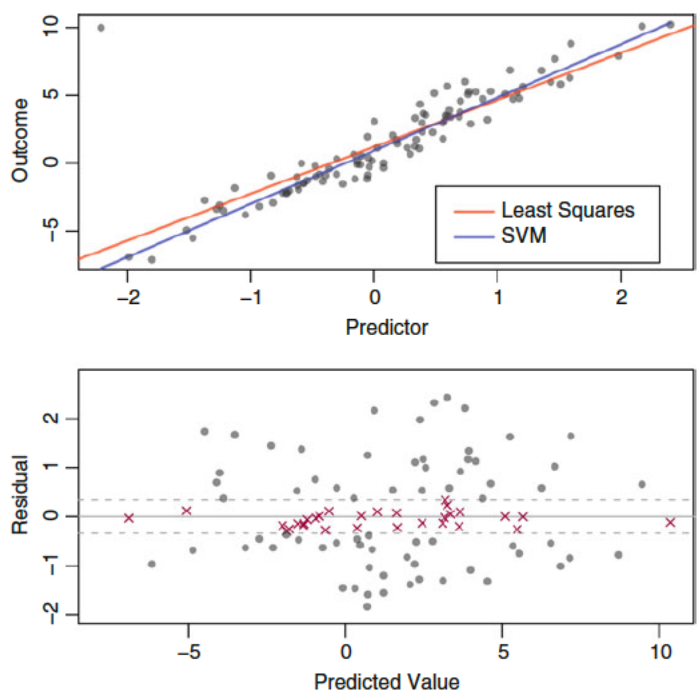

Figure 52 – An example of SVM regression (Kuhn & Johnson, 2013). The top graph presents the contrast between a standard regression line and the SVM in the presence of a highly influential outlier. The bottom graph shows the observations used as support vectors (grey circles) maximizing the margin in the fit of the predictions for other samples (red crosses).

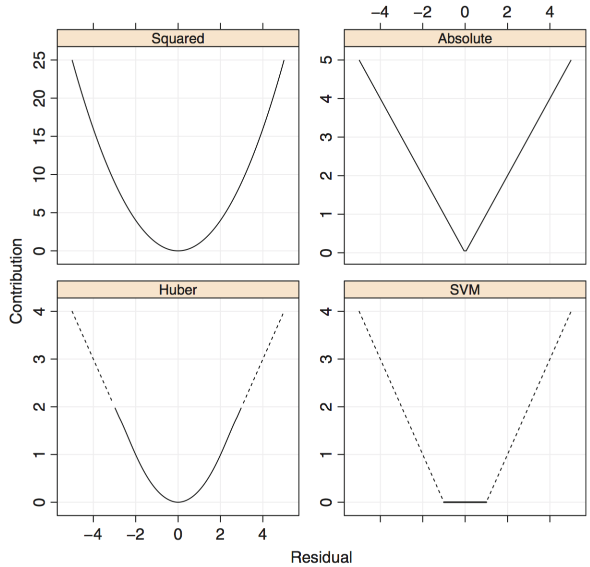

Figure 53 – Relationship between a model residual and its contribution to the regression line for several techniques. From Kuhn and Johnson, 2013.

3.2.3.3 *k*-NN: *k*-nearest neighbors

The k-nearest neighbors algorithm (Fix & Hodges, 1951), Cover & Hart, 1967, Altman, 1992) is a non-parametric and nonlinear approach to classification and regression which does not create a model through function approximation. It instead generates predictions in the response variable based on a similarity function in the set of observations, where this similarity is generally computed as the proximity (typically, the Euclidean distance). The prediction of a response value is thus based on its likelihood of being in proximity to an arbitrary number of k-closest samples in the predictor space of the training set. In the case of continuous valued-functions such as for k-NN regression, the output value is the *average* of the response values of its k-nearest neighbors, calculated though the *inverse distance weighted average* with the k-nearest multivariate neighbors. The k-nearest neighbors method is particularly useful when little to no information is available about the structure of the data, such as assumptions about distribution and possible functional forms linking response to control variables. It is considered one of the simplest types of algorithms from the point of view of computational learning theory, belonging to the broad category of instance-based learning.

Since the Euclidean distance is used as the similarity function to predict the mean response relative to the k-closest samples, it is critical to the analysis that the predictors be centered and scaled, in order to avoid undue influence from predictors featuring larger scales on the distance between samples. The k-NN algorithm is also very susceptible to missing data in the predictors, as it generates its classes or mean responses based on sample similarity. Imputation of missing data is required if any of the information concerning a sample or a predictor is to be preserved for the analysis.

Finding the optimal number of "neighbors" is done through tuning parameters such as resampling the data over a range of candidate values for the k parameter (Figure 54, Weinberger & Saul, 2009). Small values tend to over-fit, while large values of k tend to under-fit the observations. A common method to improve the performance of a k-NN is through *supervised metric learning* algorithms, such as *neighborhood components analysis* and *large margin nearest neighbor*. By transforming the predictor space of the sample data, the classification or regression performance of k-NN can be improved, such as by employing the Mahalanobis distance instead of the Euclidean distance for the cluster analysis (Figure 55, Weinberger & Saul, 2009). A major limitation of the k-NN algorithm is its sensitivity to the local structure in the dataset. A k-NN model's predictive ability is not a function of the local structure of the data, so the predictions are only as good as the relationship of the response variable being dependent on the local *predictor* structure. If this local predictor space is not relevant to the response variable, the set of predictors must be inspected in order to eliminate irrelevant, noisy, or redundant parameters. Such undersirable parameters undermine the similarity function based on sample distance over the global predictor space.

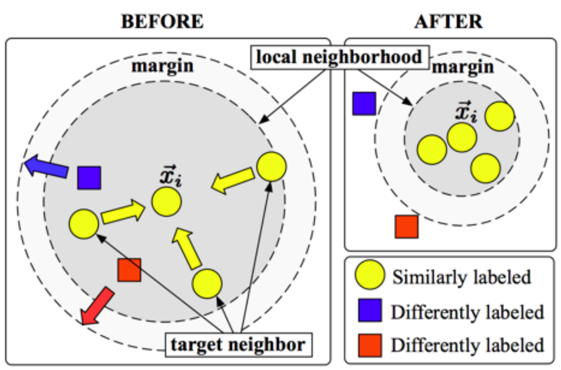

Figure 54 – Tuning the k parameter to improve a k-NN classification algorithm.

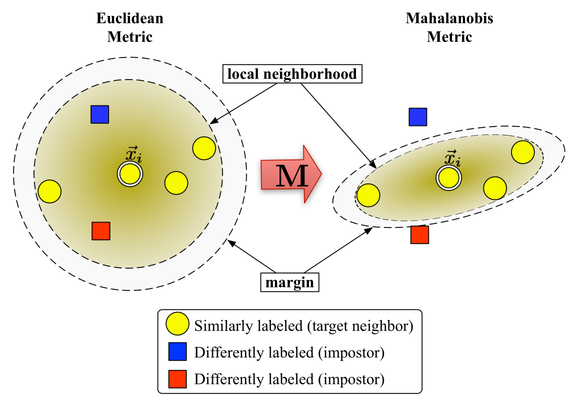

Figure 55 – The large margin nearest neighbor (LMNN) algorithm for optimization in k-NN algorithms.

3.2.3.4 aNN: artificial neural networks

Artificial neural networks (Rumelhart & McClelland 1986, Bishop, 1995, Ripley, 1996) are graph-based representations of computational units which yield a considerable nonlinear modeling capability, inspired by the architecture of the brain. A (typically) supervised learning approach in artificial intelligence in general, and machine learning in particular, aNNs can be used for pattern recognition (the classification of objects) and function approximation (such as regression analysis). The simplest neural network is a directed acyclic graph comprising three layers of nodes, or *neurons*, which receive vector values as inputs, connected to units in a further layer which may or may not be activated based on weights representing a sigmoid function, with this second layer potentially activating a third, connected output layer, producing outputs for either categorical labeling or in the form of vector values. When output values are continuous, such as in the current purposes for multiple linear regression, the output node yields the values of a function, instead of classes. The supervised learning technique used to minimize the error function is called the *backpropagation* algorithm (backward propagation of errors, Rumelhart, Hinton, & Williams, 1986), in combination with an optimization method of gradient descent, in order to find local minima for the connection weights, using a set of training data. This is done recursively until the model's residual can no longer be minimized, but it is not guaranteed to produce a global minimum in the optimization process.

The three-layer artificial neural network model is a preferred way to achieve optimal function approximation for most nonlinear relationships in regression analysis. Each input node is connected to an arbitrary number of nodes in the hidden layer, which are in turn all connected to the output layer. The input value to a node in the hidden layer is a cumulative function of the products of the input values by their respective connection weight, and the result is passed to an activation function, which may or may not "fire" based on the threshold value of the sigmoid shape of the function. Each of the hidden units is a function equivalent to the linear combination of the original predictors of a regression model, mapped to a logisitic (sigmoidal) function feeding the output layer. The output layer, however, is a linear combination of the predictors of the hidden layer. The units in the intermediate (hidden) layer are essentially a set of unobserved variables which act together in an unconstrained linear combination of the β coefficients, similar to MLR coefficients. Since the β coefficients in the hidden layer are not defined by the linear combination of predictors, but through initial random values updated through a supervised learning algorithm, they do not represent coherent pieces of information in themselves, and are treated as a "black box". That is, studying the structure of any sufficiently complex ANN will not provide insights on the structure of the functional form being approximated for the regression analysis. This is a disadvantage if a model's interpretability is important to the modeler, beyond the results. Rule-extraction algorithms have been developed to enhance the understanding of the inner workings of ANNs, based on mathematical rules, symbolic (propositional) logic, fuzzy logic, decision trees, and information visualization techniques (Hinton, McClelland, & Rumelhart, 1986, Duch, Setiono, & Zurada, 2004, Nordlie, & Plesser, 2010).

Training an ANN requires two phases: a feedforward stage and a backward phase. In the feedforward phase, the inputs are propagated to the three layers sequentially, usually with an initial set of random values for the connection weights. The weights and associated threshold functions are modified in the backward phase, based on the delta (the gradient of the error function) between the output results and the desired values, hence the name of backpropagation. Neural networks can thus achieve considerable complexity in nonlinear modeling without any particular knowledge about the structure of a model, converging on a near-optimal functional form for regression analysis. This is achieved by the constant re-engineering of network structure itself. Since there are no constraints on the parameters of this complex nonlinear modeling approach, the exploration of the parameter space which minimizes the squared sum of residuals is not guaranteed to be uniformly more efficient than other sets, and a global minimum of the error function may not be found out even through a considerable search in terms of computational time, especially for complex, high-dimensional datasets. A useful strategy in order to cope with having different locally optimal solutions in the error reduction function is to produce a stable model through *model averaging*, i.e., running the ANN with random initial values multiple times and averaging the model results.

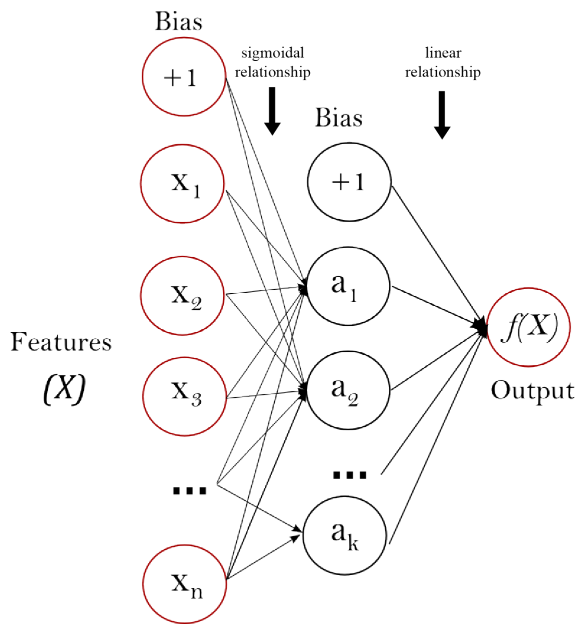

Figure 56 – Schematics of a multilayer perceptron, a simple feedforward artificial neural network.

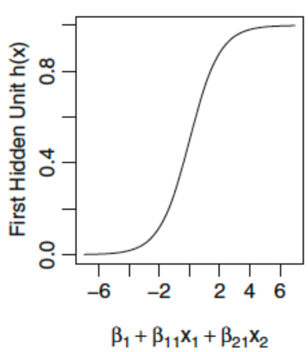

Figure 57 – Sigmoid function of the linear combination of input predictors.

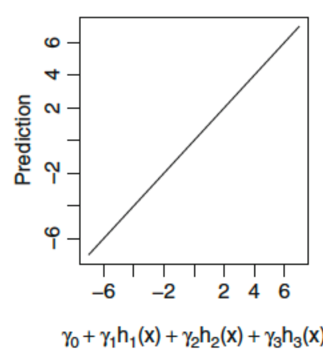

Figure 58 – Linear function of the hidden units.

3.2.3.5 LOESS: locally weighted scatterplot smoothing

The locally weighted scatterplot smoothing (LOESS, Cleveland, 1979, 1981, Cleveland & Devlin, 1988), also known as local regression, is a non-parametric method that uses the least squares regression method on localized subsets of the data. It is a type of *kernel smoother* (Hastie, Tibshirani, & Friedman, 2009), as it is used on a numerical vector to smoothen it and to predict the response variable locally, by using the noisy observations, when no parametric model for a function is known. LOESS is particularly effective at modeling nonlinearities between predictors and response variables, as a generalization of the spline, or piecewise regression (such as the MARS method above), bu using interpolation between data points. It is also similar to computing a moving average in time series analysis and signal processing.

The local linear regression is computed by solving a weighted least squares problem for a "neighborhood", or small region of the multivarate data. The range of the neighborhood can be controlled using a tuning parameter named the "span", which ranges between $\lambda + 1$ and 1 (where λ is the degree of the local polynomial). The span controls the degree of the smoothing kernel. Formally, each of the points along the smoothing curve is obtained by a *weighted quadratic least squares regression* over the local neighborhood (determined by the span parameter) of values of the y-axis scatter plot. In other terms, a low-degree polynomial is fitted to the local subset of data for each point in the range, giving more weight to points near the point whose response is being estimated. The span value, the degree of the polynomial function, as well as the type of weight function can be determined by the modeler.

At greater span values, the fitted curve is smoother. This can de described for technical purposes as a combination of multiple quasi-linear (polynomial) regression models driven by a k-nearest neighbors "meta-model". Finding the optimal smoothing span is a challenge, as the smoothing span changes, the accuracy of the fitted curve also changes. Error reduction through an optimization function is available in LOESS, based on comparisons of a range of span values relative to the sum of squared residuals. Although the LOESS algorithm can technically model high-dimensional predictor spaces, it uses up to four parameters in its current implementation, as the determination of the locality (the so-called neighborhoods) of subsets of data becomes less computationally tractable as the number of dimensions increases. Having too many parameters also increases the overall model variance, as well as the likelihood of overfitting a model. The LOESS method combines the simplicity of interpreting the least squares calculation of linear regression with the tremendous capability of nonlinear modeling, without having to commit to a particular functional form in fitting a model to the data. The major drawback of not having to specify a functional form for the regression analysis is that a technique such as LOESS may produce a model which is not easily interpretable as a mathematical function. In the particular context of finding an objective model of complexity to explain and predict human performance in DDM scenarios, this could be a problem.

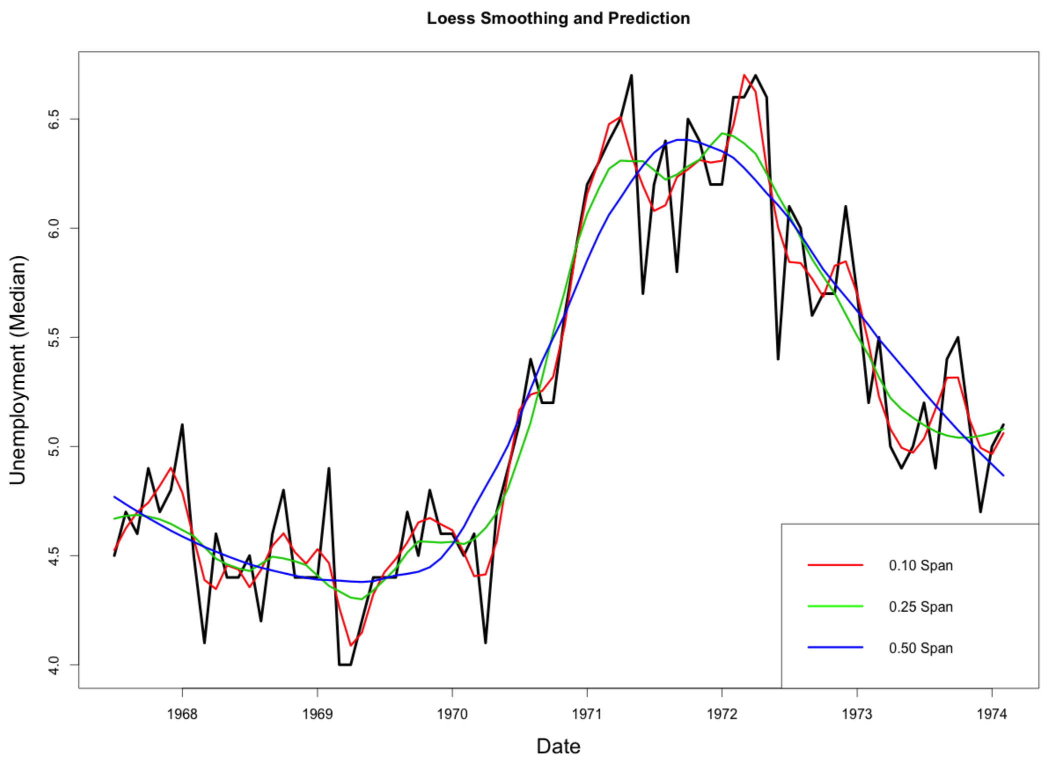

Figure 59 – Example of LOESS regression lines for three different span values. The black line is the actual data.

3.3 Results

The following subsections present the results of the alternative regression analyses (robust regression, MLR using group means and group medians, nonlinear and non-parametric regressions) used to compare with our candidate MLR models of objective measures of complexity as predictors of human performance in dynamic decision-making scenarios.

3.3.1 Robust regression models

The `R` package `robustbase` (Rousseeuw et al, 2016) was used in order to implement the MM-estimation method, using the iteratively reweighted least squares computation of the regression analysis. The `lmrob` function of the `robustbase` used a robust and efficient estimator with a 50% breakdown point (similar to using the median as an estimator) and 95% asymptotic efficiency for normal errors (Maronna & Yohai, 2000). The `lmrob` function not only yields $RMSE$ indicators to compare the goodness of fit of the regression models, but it also produces a *robust coefficient of determination for regression* (Renaud & Victoria-Feser, 2010), or robust adjusted R^2, as a relative measure of explained variance. This goodness of fit measure indicates, much like the adjusted R^2 of the linear regression approach, the suitability of the chosen explanatory variables in predicting the response variable. The coefficient is robust to influential observations on the regression line while not making any assumption on the distribution of the explanatory variables. This robust coefficient of determination for regression has been shown to exhibit consistency and unbiasedness, beyond the requirements of efficiency and robustness for an estimator, even in small samples. The tuning parameters in the `robustbase` package are passed to the `lmrob.control` function, and the only parameter tweak that was undertaken in the default settings was the maximum number of iterations (100) of the IRLS method to converge to an optimal set of weights maximizing the likelihood function for the robust regression models. The IRLS method of calculation for the sum of squared residuals of the candidate models converged to optimal solutions between 11 to 26 iterations.

The results of the robust regression analysis are detailed below in Tables 17 and 18 for the five scenarios in CODEM, and for an analysis including the Ecopolicy scenario directly in the dataset, respectively. The robust regressions modeling the relationship between the objective parameters of complexity and the performance metric yielded $RMSE$ values from 8.11 to 8.19, for the analysis involving the five DDM scenarios in CODEM (using the same 3-parameter candidate models). With the inclusion of the Ecopolicy scenario, the $RMSE$ scores ranged from 8.80 to 9.25 for the candidate 3-parameter models. The results for both analyses are considerably superior to the standard MLR approach, across all candidate models. The robust and resistant regression models for the objective measures of complexity appear to achieve a reduction of 50% of the standard error of the regression indicator, relative to the other methods of regression analysis. Looking at the robust adjusted R^2, the candidate models are able to explain anywhere between 70.33% and 71.79% for the analysis involving the five DDM scenarios in CODEM, and between 64.65% and 78.53% for the analysis including the sixth scenario in the Ecopolicy serious game.

The impressive results produced by the robust regression analysis come at a particular cost, relative to the ordinary least squares method of the standard MLR approach. The iteratively reweighted least squares computation used to minimize the error function of the robust regression models assigns weights ranging between 0 and 1 to every single observation in the sample data, in order to maximize the likelihood of observing this particular data given a model's set of parameters. This is done in order to minimize the impact of influential observations on the linear trend linking the predictors to the response variable. In the present research project, it has already been assessed that a number of influential observations were a nuisance to the standard linear regression modeling endeavor, as were the non-constant error variance and the non-normality of the multivariate residuals. The robust and resistant regression models presented in this section were able to achieve such low ratings for the absolute measure of the standard error of the regression ($RMSE$) and high scores for the relative measure of the robust adjusted R^2 by weighting down some observations all the way to a weight value of nearly 0, effectively eliminating those influential observations altogether. Tables 17 and 18 present a column featuring the number of observations which were weighted to a score of near-zero by the respective candidate model of robust regression. Such dropped observations range between 20 and 21 for the five scenarios in CODEM, and between 17 and 27 for the analysis adding the Ecopolicy scenario, for 3-parameter models. The remainder of the sample data is of course subject to strong variations in the distribution of weights for the calculation of the least squares. For example, Table 19 presents the weight distribution of the observations for the $SDC_{DDM} + I_{DDM} + D_{DDM}$ candidate model of robust regression, the most promising MLR model from the analyses produced in chapter 2, in the particular context of the analysis involving only the five DDM scenarios in CODEM.

Table 17 – $RMSE$ for robust regression of candidate MLR models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) without Ecopolicy, using the iterated re-weighted least squares (IRLS) method.

model	**robust adj. R^2**	**robust RMSE**	**# obs. dropped**
$CNC + U_{DDM} + S_{DDM}$	.7179	8.19	20
$CC + U_{DDM} + S_{DDM}$	.7172	8.18	20
$SDC + U_{DDM} + S_{DDM}$	.7154	8.16	20
$SDC + I_{DDM} + D_{DDM}$	.7131	8.11	21
$CNC + I_{DDM} + D_{DDM}$	.7089	8.11	21
$CC + I_{DDM} + D_{DDM}$	.7033	8.11	21

Table 18 – $RMSE$ for robust regression of candidate MLR models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) with Ecopolicy, using the iterated re-weighted least squares (IRLS) method.

model	**robust adj. R^2**	**robust RMSE**	**# obs. dropped**
$CNC + U_{DDM} + S_{DDM}$	.7853	8.80	27
$CC + U_{DDM} + S_{DDM}$	.7846	8.80	27
$SDC + U_{DDM} + S_{DDM}$	.7664	9.20	22
$SDC + I_{DDM} + D_{DDM}$	.6838	8.73	22
$CNC + I_{DDM} + D_{DDM}$	.6511	9.19	17
$CC + I_{DDM} + D_{DDM}$	.6465	9.25	17

Table 19 – Robustness weights for the $SDC_{DDM} + I_{DDM} + D_{DDM}$ model of objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) without Ecopolicy, using the iterated re-weighted least squares (IRLS) method. Convergence in 14 IRLS iterations. 21 observations are outliers with $|\text{weight}| = 0$ ($< .0005$). 24 weights are ~ 1. The remaining 154 weights are summarized in the table below.

minimum	1st quantile	median	mean	3rd quantile	maximum
0.0079	0.8638	0.9599	0.8619	0.9857	0.9987

Worth noting is a difference in the relationship between the regression coefficients for the objective parameters of complexity involved in the candidate MLR models and the response variable. The structural complexity component C_{DDM} in the candidate models involved in the robust regression analysis now yield negative coefficients when combined with uncertainty U_{DDM} and instability S_{DDM}, similar to the case of C_{DDM} combined with information complexity I_{DDM} and the measure of difficulty D_{DDM}. This reversal concerning the undesirable positive coefficients for the C_{DDM} metrics in the former set of candidate models provides additional credibility for the three candidate models involving U_{DDM} and S_{DDM}.

3.3.2 Regression analysis using group means and group medians

Table 20 presents the results of the MLR analysis using group mean performances in the DDM scenarios as a response variable, and features predictions for the Ecopolicy mean performance using the holdout validation method. As a reference, the mean performance in Ecopolicy is 40.69. The 3-predictor models are likely overfitting on the goodness of fit for the regression analysis (with adj. R^2 score between 92.74% and 98.48%), and offer very low scores on the standard error of the regression (ranging from 5.00 to 2.29). Yet all 3-predictor MLR models also offer very poor predictions of the Ecopolicy mean performance, failing to generalize to an additional scenario. The models involving one of the C_{DDM} metrics combined with the uncertainty U_{DDM} and instability S_{DDM} parameters were slightly better performers on the regression fit but underperformed even more than the models involving one of the C_{DDM} metrics with the information I_{DDM} and the difficulty D_{DDM} parameters on the prediction test.

Additional analyses involving a number of 2-parameter models were found to provide better model fit than in the traditional MLR analysis, and improved predictions using the Ecopolicy mean performance. The 2-parameter models involving A_{DDM} ($CNC + A_{DDM}$ and $SDC + A_{DDM}$) have positive coefficients for the action complexity parameter, which is very likely to capture the difference between COIN 1 and COIN 2 relative to the action points, as discussed in chapter 2. The $SDC + A_{DDM}$ offered the best prediction of all candidate MLR models in this particular analysis, with a RMSE score of 3.38.

The coefficients of the candidate regression models are similar to the original MLR analysis in chapter 2, where the structural complexity component C_{DDM} for 3-parameter models yields positive coefficients when combined with U_{DDM} and S_{DDM}, while it is negative when combined with I_{DDM} and D_{DDM}. This again suggests that the three $C_{DDM} + I_{DDM} + D_{DDM}$ models are more plausible fits to explain the variance in the human performance data for dynamic decision-making problems.

Table 20 – Candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using group mean performances only.

model	adj. R^2	RMSE	mean prediction (Ecopolicy)	mean prediction RMSE
$SDC + U_{DDM} + S_{DDM}$	.9848	2.29	76.31	35.62
$CC + U_{DDM} + S_{DDM}$	.9716	3.13	67.90	27.21
$CNC + U_{DDM} + S_{DDM}$	.9666	3.39	68.41	27.73
$CC + I_{DDM} + D_{DDM}$	.9452	4.35	14.84	25.85
$CNC + I_{DDM} + D_{DDM}$	.9351	4.73	14.15	26.54
$SDC + I_{DDM} + D_{DDM}$	.9274	5.00	10.34	30.35
$CNC + A_{DDM}$	.6527	10.94	61.71	21.02
$CNC + D_{DDM}$	.6176	11.48	58.70	18.02
$SDC + A_{DDM}$	.6171	11.49	44.06	3.38

Table 21 presents the results of the MLR analysis using group median performances in the DDM scenarios as a response variable, and features predictions for the Ecopolicy median performance using the holdout validation method. As a reference, the median performance in Ecopolicy is 23.00. The 3-predictor models are again likely overfitting on the goodness of fit for the regression analysis (with adj. R^2 between 94.17% and 96.87%), and offer even lower scores on the standard error of the regression (ranging from 4.38 to 3.21) than in the group mean performance analyses above. When models using one of the C_{DDM} metrics combined with the uncertainty U_{DDM} and instability S_{DDM} parameters are involved, they offer even poorer predictions of the Ecopolicy mean performance. On the other hand, the models involving one of the C_{DDM} metrics with the information I_{DDM} and the difficulty D_{DDM} parameters are now both very good fit over the group median performances and provide excellent predictions using Ecopolicy as an out-of-sample test set. Indeed, the $CNC_{DDM} + I_{DDM} + D_{DDM}$, $CC_{DDM} + I_{DDM} + D_{DDM}$, and $SDC_{DDM} + I_{DDM} + D_{DDM}$ now offer predictions of the median performances in Ecopolicy with mean prediction $RMSE$ scores of 2.55, 4.41, and 6.53 respectively.

In a similar fashion to the MLR analyses using group mean performances above, additional analyses involving 2-parameter models were found to provide better fitness than in the original MLR analysis, but this time offered very poor predictions using the Ecopolicy median performance akin to the 3-parameter models involving the uncertainty U_{DDM} metric. The 2-parameter models involving combinations of structural complexity U_{DDM} terms with uncertainty U_{DDM} yield negative regression coefficients. The coefficients of the candidate regression models are again similar to the original MLR analysis, whereby the structural complexity component C_{DDM} for 3-parameter models yields positive coefficients when combined with U_{DDM} and S_{DDM}, while it is negative when combined with I_{DDM} and D_{DDM}.

The results of the MLR analysis using group mean performances and group median performances support our hypothesis from chapter 2 and the introduction of the present chapter regarding our concerns for the irregularities in distributions for the response variable, affected by small sample sizes, non-normality, and non-constant error variance. Using the median instead of mean performances not only insured additional robustness against violation of the assumptions of parametric estimators, it exhibited considerable improvements in the indicators of fitness for both the regression analysis and regression validation phases. Additionally, more support is given to the three $C_{DDM} + I_{DDM} + D_{DDM}$ model variants as the best candidate regression models to both explain the variance in performance in the DDM scenarios in CODEM, and to generalize to a test scenario outside of CODEM as indicated through a regression validation analysis.

Table 21 – Candidate multiple linear regression models for objective measures of complexity in five dynamic decision-making scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using group median performances only.

model	adj. R^2	RMSE	median prediction (Ecopolicy)	median prediction RMSE
$SDC + I_{DDM} + D_{DDM}$	.9687	3.21	16.47	6.53
$CNC + I_{DDM} + D_{DDM}$	.9595	3.65	20.45	2.55
$CC + I_{DDM} + D_{DDM}$	.9478	4.15	18.59	4.41
$SDC + U_{DDM} + S_{DDM}$	.9477	4.15	74.06	51.06
$CC + U_{DDM} + S_{DDM}$	.9429	4.34	71.83	48.83
$CNC + U_{DDM} + S_{DDM}$	.9417	4.38	71.88	48.88
$CC + U_{DDM}$	.8612	6.76	75.16	52.16
$CNC + U_{DDM}$	.8603	6.78	74.92	51.92
$SDC + U_{DDM}$	.8597	6.80	70.90	47.90

3.3.3 Nonlinear and non-parametric regression models

All of the nonlinear and the non-parametric analyses were conducted through the `R` programming language. The MARS method used the `earth` package (Milborrow, 2016), including only a first-degree model of regression (i.e., it did not model interactions between predictors). As its default parameters of computation for the splines are used to conduct variable selection, we disabled the addition or removal of the regression parameters to ensure that only the parameters of interest are included in the analysis. A parameter named `thresh` prevented the removal of predictors in the forward pass, while the parameter `penalty` was tested at values ranging between [-1, 0, 1, 2] in order to prevent the removal of predictors based on cross-validation penalties incurred in the computation of the residual sum of squares. There was therefore a total number of 4 simulations per model of regression.

For the SVMr, the `ksvm` function of the `kernlab` package (kernel-based machine learning methods, including support vector machines, Karatzoglou, Smola, & Hornik, 2016) was used. The default `eps-svr` epsilon regression algorithm was used, with an array of [.01, .05, .1, .5, 1] for the `epsilon` parameter (the default parameter value is $\epsilon = 0.1$). 1,000 simulations were conducted for each parameter value, for a total of 5,000 for each regression model, as the SVMr method is a complex algorithm which may fail to converge to a global maximum at any given simulation execution, similar to the aNN method.

The k-NN method was used through the `caret` package (Kuhn, 2016), using the `knnreg` function, with the `k` parameter being tested over values of the array [1, 2, 3, 4, 5]. For $k = 1$, i.e., with *1*-nearest neighbor (also known simply as the "nearest neighbor" algorithm), the models tend to overfit in some of the following analyses, so this particular value was eliminated from the pool of best results when this was the case, as indicated in the text. There was a total number of 5 simulations per regression model.

The aNN model used the `nnet` from the package of the same name (Ripley, 2016), with the tuning of the initial random weights for a maximum of 1,000 iterations, a number of units in the hidden layer in the interval of [1, 2, 3] for the `size` parameter, and a weight `decay` value (a regularization method to avoid overfitting) in the interval of [.01, .05, .1]. Models with 2 or 3 units in the hidden layer tend to overfit in some of the results presented below, where it will be indicated that 1 hidden unit was used instead. 5,000 simulations were executed for each value of the two main parameters per regression model, not counting the required computations for the iterations over the tuning of the random weights. The total number of simulations for each regression model is therefore 45,000.

The LOESS algorithm used Ripley's implementation of Cleveland's (1979, 1981, Cleveland & Devlin, 1988) of the `loess` function in the default `stats` package in `R` (`R` Core Team et al, 2016). The parameters of the LOESS method were a first-degree model of regression, and `span` values in the interval [.75, .80, .85, .90, .95] were used, as the suggested lower bandwidth values ranging from .25 to .75 either returned error messages due to non-invertible matrices (caused by multicollinearity among the predictors), or did not successfully reduce the $RMSE$ at the levels of higher span values. Five simulations per regression model were executed.

The results of the nonlinear and non-parametric analyses are detailed below in Tables 22 and 23 for the five scenarios in CODEM, and for an analysis including the Ecopolicy scenario dataset, respectively. The nonlinear and non-parametric approach to modeling the relationship between the objective parameters of complexity and the performance metric yielded $RMSE$ values ranging from 16.00 to 16.53 for our candidate 3-parameter models, for the analysis involving the five DDM scenarios in CODEM. With the inclusion of the Ecopolicy scenario, the $RMSE$ scores ranged from 17.20 to 17.81 for the same 3-parameter models. The results for the SVMr and the aNN methods are approximate in nature, given that both of those algorithms attempt to minimize the error function of a regression model based on partially stochastic processes, i.e. through optimization algorithms which operate as a supervised search for a lowest minimum (which may or may not be the global minimum of the cost function). However, given the high number of simulations iterated over each of the tuning parameter values, the likelihood of obtaining a global minimum was greatly increased.

The results presented in Tables 22 and 23 below are not a substantial improvement over the multiple linear regression approach using the ordinary least squares method of calculation for bias and variance. Some of the nonlinear and non-parametric algorithms have indeed found lower model residuals relative to the MLR approach, such as the MARS models, the k-NN models, and the artificial neural networks, with an asymptotic floor value approaching 16.00 units of the performance metric, using the $RMSE$ fitness indicator. The support vector regression fared the worst (with higher $RMSE$ scores than those of the original MLR candidate models in the case of the analysis using the five DDM scenarios in CODEM), while the LOESS method yielded almost identical values to the the OLS method. This is most probably due to the issue of non-invertible matrices in computing the LOESS regression with low bandwidth values, as we could only manage to obtain results minimizing the error function at high span values close or equal to .95, which is barely any different from the OLS computation to begin with. Worth noting is that for other 3-parameter model combinations which are not presented here, the results of some of the nonlinear and non-parametric analyses are similar.

Table 22 – $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2).

model	**MARS**	**SVMr**	***k*-NN**	**aNN**	**LOESS**
$CC + I_{DDM} + D_{DDM}$	16.00	16.53	16.00	16.08	16.27
$CNC + I_{DDM} + D_{DDM}$	16.00	16.53	16.00	16.09	16.24
$SDC + I_{DDM} + D_{DDM}$	16.00	16.53	16.00	16.10	16.21
$CC + U_{DDM} + S_{DDM}$	16.00	16.46	16.00	16.00	16.30
$CNC + U_{DDM} + S_{DDM}$	16.00	16.46	16.00	16.00	16.31
$SDC + U_{DDM} + S_{DDM}$	16.00	16.46	16.00	16.00	16.32

Table 23 – $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) with Ecopolicy.

model	**MARS**	**SVMr**	***k*-NN**	**aNN**	**LOESS**
$CC + I_{DDM} + D_{DDM}$	17.20	17.73	17.20	17.29	17.46
$CNC + I_{DDM} + D_{DDM}$	17.20	17.73	17.20	17.31	17.43
$SDC + I_{DDM} + D_{DDM}$	17.20	17.81	17.20	17.36	17.47
$CC + U_{DDM} + S_{DDM}$	17.20	17.70	17.20	17.64	17.78
$CNC + U_{DDM} + S_{DDM}$	17.20	17.70	17.20	17.64	17.62
$SDC + U_{DDM} + S_{DDM}$	17.20	17.69	17.20	17.64	17.55

What happens if we run the same nonlinear and non-parametric analyses using only the central tendency estimators, as was the case in the previous section, namely the group mean performances and the group median performances? Tables 24 and 25 below present the results of the identical nonlinear and non-parametric analyses (i.e., using the same tuning parameters) over the dataset strictly involving the five DDM scenarios in CODEM, while Tables 26 and 27 present the same analyses for the dataset including the Ecopolicy scenario.

The first, rather obvious observation is a tendency to overfit for a number of the nonlinear and non-parametric methods. The multivariate adaptive regression splines, the support vector machine regressions, and the artificial neural networks appear to reduce the standard error of the regression scores close to null values, very likely due to the ratio of the number of parameters to the number of cases involved in using group mean and median performances. The k-nearest neighbors algorithm produces higher $RMSE$ scores, but this is a selection issue, as the use of the k-NN models with $k = 1$ produced overfit models with $RMSE = 0.00$, and were excluded from the analyses. Likewise, using aNN models with 2 or 3 units in the hidden layer produced null values for the standard error of the regression scores. The locally weighted scatterplot smoothing algorithm failed to produce any results for all of the candidate regression models, as LOESS does not work for small sets of data points (an error message stated that there were "fewer data values than degrees of freedom"), notwithstanding invertibility and near-singularity issues in the LOESS computations.

Some patterns of interest are observed in the results below: firstly, the standard error of the regression scores tend to be lower in the analyses conducted strictly on the five DDM scenarios in CODEM, relative to the analyses conducted in the dataset including the Ecopolicy scenario. This is consistent with the original MLR analyses, the robust regression models, as well as with the nonlinear and non-parametric analyses performed over the entire dataset. However, there are two exceptions, the first case being that of the artificial neural network models involving the $C_{DDM} + I_{DDM} + D_{DDM}$ models as predictor of the group median performances, which yield equal or lower $RMSE$ scores when Ecopolicy is included. The other case concerns the MARS models featuring the U_{DDM} and S_{DDM} parameters, which also exhibit lower $RMSE$ scores when Ecopolicy is included. Secondly, the greatest benefits of using group median performances as a response variable are noticeable in the particular case of the the models involving one of the C_{DDM} metrics with the information I_{DDM} and the difficulty D_{DDM} parameters. For those models, the standard error of the regression scores are consistently lower than in the analyses using the group mean performances (with two minor exceptions involving the k-NN models in the analyses using the Ecopolicy data). Thirdly, from the more extreme case of including Ecopolicy directly in the analyses, and the more robust choice of estimators for central tendency represented by group median performances, we obtain yet again more convincing results across all of the nonlinear and non-parametric methods when the $C_{DDM} + I_{DDM} + D_{DDM}$ model variants are involved.

Table 24 – $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using mean performances.

model	**MARS**	**SVMr**	***k*-NN**	**aNN**	**LOESS**
$CC + I_{DDM} + D_{DDM}$	1.02	3.88	10.47	2.55	fail
$CNC + I_{DDM} + D_{DDM}$	1.02	3.88	10.47	2.72	fail
$SDC + I_{DDM} + D_{DDM}$	1.02	3.88	8.88	2.82	fail
$CC + U_{DDM} + S_{DDM}$	0.38	2.63	10.47	0.59	fail
$CNC + U_{DDM} + S_{DDM}$	0.38	2.61	10.47	0.59	fail
$SDC + U_{DDM} + S_{DDM}$	0.38	2.68	8.88	0.59	fail

Table 25 – $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) using median performances.

model	**MARS**	**SVMr**	***k*-NN**	**aNN**	**LOESS**
$CC + I_{DDM} + D_{DDM}$	0.57	1.38	8.59	1.07	fail
$CNC + I_{DDM} + D_{DDM}$	0.56	1.38	8.59	1.07	fail
$SDC + I_{DDM} + D_{DDM}$	0.51	1.38	4.99	1.03	fail
$CC + U_{DDM} + S_{DDM}$	0.70	0.93	8.59	0.52	fail
$CNC + U_{DDM} + S_{DDM}$	0.70	0.92	8.59	0.52	fail
$SDC + U_{DDM} + S_{DDM}$	0.70	1.16	4.99	0.51	fail

Table 26 – $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2), including Ecopolicy, using mean performances.

model	**MARS**	**SVMr**	***k*-NN**	**aNN**	**LOESS**
$CC + I_{DDM} + D_{DDM}$	1.78	4.45	8.95	2.76	fail
$CNC + I_{DDM} + D_{DDM}$	1.93	4.45	10.66	2.87	fail
$SDC + I_{DDM} + D_{DDM}$	2.04	4.46	8.02	3.53	fail
$CC + U_{DDM} + S_{DDM}$	1.22	3.66	10.66	5.78	fail
$CNC + U_{DDM} + S_{DDM}$	1.11	3.63	10.66	5.79	fail
$SDC + U_{DDM} + S_{DDM}$	1.49	3.63	8.02	5.80	fail

Table 27 – $RMSE$ of nonlinear and non-parametric candidate models for objective measures of complexity in DDM scenarios (Arctic 1, Arctic 2, Arctic 3, COIN 1, COIN 2) including Ecopolicy, using median performances.

model	**MARS**	**SVMr**	***k*-NN**	**aNN**	**LOESS**
$CC + I_{DDM} + D_{DDM}$	1.69	3.36	7.30	1.09	fail
$CNC + I_{DDM} + D_{DDM}$	1.49	3.41	11.96	0.93	fail
$SDC + I_{DDM} + D_{DDM}$	1.31	3.37	9.24	0.72	fail
$CC + U_{DDM} + S_{DDM}$	0.51	3.98	11.62	10.13	fail
$CNC + U_{DDM} + S_{DDM}$	0.53	3.92	11.62	9.88	fail
$SDC + U_{DDM} + S_{DDM}$	0.45	3.45	9.24	4.46	fail

3.4 Discussion

This chapter aimed to explore alternative methods and models of the relationships between objective measures of complexity and performance in dynamic decision-making problems. Uncertainties about the accuracy and precision of the data and resulting analyses in the previous chapter led us to address the bias-variance trade-off as it differentially affects microcognitive and macrocognitive phenomena. We compared the psychophysics approach to modeling microcognitive phenomena, which inspired our linear regression modeling approach to DDM, with more complex approaches which might be more suitable for macrocognitive phenomena involved in metacognitive processes such as systems thinking. The linear modeling approach of MLR is based not only on observations in a dataset, but also on prior assumptions about underlying distributions and relationships between variables. In light of the non-normality of our candidate model residuals, in the presence of heteroscedasticity as well as highly influential observations, there was a requirement to move beyond the ordinary least squares method of the GLM approach to methods of regression analysis which have different constraints in the way they model variable relationships and statistical error. We have explored the use of nonlinear and non-parametric methods from the machine learning approach in computational learning theory, and we have also endeavored to compare the MLR models with the robust and resistant regression approach of MM-estimation, which uses the flexible iteratively reweighted least squares computation, combining statistical efficiency with the high breakdown point in the face of influential outliers (Yohai, 1987,Maronna et al, 2006, Rousseeuw et al, 2016). Another modeling approach used group mean and median performances as response variables to observe the effect of the elimination of intra-group variance attributed to inter-individual differences. This approach was combined with the nonlinear and non-parametric methods and produced considerable improvements for the goodness of fit indicators of our candidate models, although questions were raised as to the possibility of overfitting for such complex modeling approaches to regression.

3.4.1 Comparing the linear regression models with robust regression, regression using group means and medians, and nonlinear methods

Multiple linear regression models are a useful tool to model and interpret linear relationships when the postulates of independence of observations, the normality of the variables and of the multivariate residuals, the constancy of error variance, the lack of high collinearity among predictors, and the absence of influential observations are respected. Yet linear analysis is prone to high sensitivity in the calculation of its regression slope in the presence of observations which lie distant to the major trend in the data. Distant outliers can yield unduly large residuals as they are squared in the OLS computation, as the linear regression seeks to account for every data point without any consideration for its relevance or influence on the overall model. One way to adapt nonlinear trends in the multivariate data to linear regression analysis is to transform the predictors (by subjecting the coefficients to exponential, inverse, or logarithmic functions, etc.) This presumes that the particular functional form of the multivariate equation is already known to the analyst.

In order to cope with the large model residuals produced by the OLS method for our candidate MLR models of objective measures of complexity, we exploited the relationship between a model's residuals and their contribution to the regression line by minimizing the influence of influential outliers. By using the MM-estimation approach to robust regression, observations which had large residuals and/or high leverage values were weighted down in the computation of the regression line, in such a way that only observations which were clustered close to the regression line (and thus produced smaller residuals) contributed significantly to the determination of the linear regression slope. The robust regression method produced substantially higher model accuracy compared to the MLR candidate models, with nearly halved scores for the standard error of the regression indicators. The robust regression models could also be assessed for the proportion of explained variance in the human performance response variable using a robust adjusted R^2 indicator, whereby it was shown that 3-parameter models could explain anywhere between 70% to 72% of the performance for the five DDM scenarios in the CODEM simulation environment, and between 65% to 79% when an additional test scenario was included. An analysis using the group mean and median performances as the response variables for the MLR candidate models from chapter 2 revealed even lower $RMSE$ scores and inflated coefficients of determination for both analyses (using group means and group medians), at the expense of predictive power as revealed through an out-of-sample validation test using the Ecopolicy data. However, MLR analyses on the candidate models using the more robust estimator of group medians favored the models featuring C_{DDM} metrics with the information I_{DDM} and the difficulty D_{DDM} parameters, which could not only explain the variance in median performances with adjusted R^2 scores ranging between 95% and 97%, the same candidate models also managed to predict the Ecopolicy median score within 3% to 7% of the performance scale, as demonstrated through the median prediction $RMSE$ scores.

The final strategy to capture departures from linear regression due to violations of some of its underlying assumptions was through the use of nonlinear and non-parametric models which can exploit mathematical tricks through sophisticated algorithms. Here, we have used piecewise regressions using splines, interpolation and weighting through smoothing kernel functions, mapping multivariate data combinations into linearizable feature spaces, using intermediate transition functions to map nonlinear inputs to linearized output functions, and the classification of multivariate data based on similarity functions such as the Euclidean distance. It was found that the machine learning approaches yielded a performance ceiling which was either comparable to the candidate MLR models (for the SVMr and the LOESS methods), or only marginally better in accuracy (for the MARS method, the k-NN algorithm, and the artificial neural networks). However, the combination of the aforementioned multiple linear regression method using group mean and median performances as response variables with the nonlinear and non-parametric models yielded similar or even lower scores of standard error of the regression than in the previous regression analyses, suggesting that nonlinear models may be more appropriate than the candidate linear regression models. Combined with observations concerning the negative relationship between the regression coefficients for the parameters of complexity and the response variable, the final analysis also substantiated our conjecture developed throughout chapters 2 and 3 to the effect that MLR models involving structural complexity, information complexity, and a measure of difficulty are candidates of choice in order to explain the impact of complex and dynamic decision-making problems on human understanding and performance.

3.4.2 To underfit or to overfit? Compromising between bias and variance in modeling objective measures of complexity

The standard error of the regression is a measure of the differences between the observed and predicted values, which are then squared and summed for all the observations. The square root of those mean squared error represents the average departure from an observation in assessing a prediction for a given value of the response variable. This indicator of goodness of fit is useful as it is computed in the same unit as the original data, so that when we see the candidate model $SDC_{DDM} + I_{DDM} + D_{DDM}$ bearing a $RMSE$ score of 16.25 for the original set of models in the MLR analysis, we can straightforwardly assess that the nonlinear and non-parametric methods of MARS, k-NN, and the artificial neural network only marginally increase the models' accuracy with a $RMSE$ score of 16.00. Are those models more desirable than the original multiple linear regression analysis? Most probably not. Following Kuhn and Johnson's (2013) criteria for model comparison and selection, the aNN approach is a more opaque and complex algorithm relative to MARS, k-NN, and the MLR approach. In fact, the most interpretable and flexible machine learning algorithm among the above-mentioned three is the multivariate adaptive regression splines, as it merely segments the regression line based on kinks found in local subsets of the regression analysis. There are a few drawbacks: the MARS method does use transformed subsets of the model predictors, called surrogate features, in order to reduce the error function used to fit the predictions to observations, and the recursive partitioning nature of the stepwise process used to reduce this error function also acts as a variable selection procedure, which was not the intended function of the method for our current objectives (although we prevented the latter using tuning parameters).

The robust regression method we have used managed to reduce how far, on average, the residuals are from a null, value, i.e., to minimize the average distance between the observed values and the model predictions, as expressed by the standard error of the regression. Points which lie farther from the regression line and yield larger residual values, or influential observations, have been undervalued in the computation of the regression line. The candidate model $SDC_{DDM} + I_{DDM} + D_{DDM}$, for example, bears a $RMSE$ score of 8.11, or less than half of its corresponding value in both the original MLR analysis and in the best nonlinear and non-parametric approaches Is this a desirable outcome? The model accuracy is considerably enhanced by the adoption of the MM-estimation computation of the model residuals, which nevertheless comes at the price of selective insensitivity to data points in our pool of observations. In order to achieve lower heteroscedasticity in the multivariate residuals and cope with influential outliers, the robust regression eliminates the relative contribution of approximately 20 observations out of close to 200 data points. This constitutes just above 10% of the overall dataset. Is this a reasonable sacrifice? There were no principled way to pick and choose observations which qualified as undesirable outliers in the original analysis, given the considerable range in variance exhibited for the DDM scenarios, and there doesn't seem to be a non-statistical reason, i.e., outside of the IRLS procedure involved in the robust regression, to favor some observations instead of others. The proportion of down-weighted observations could still be appraised as a modest sacrifice in order to reduce the source of error attributable to a model's sensitivity to fluctuations in the data. We have deliberately chosen to reduce error attributable to variance, which tends to overfit by modeling the error term of a regression model, in favor of models with higher accuracy (lower overall residuals), at the cost of possibly having models with higher bias (which may fail to model some of the underlying structure linking the predictors to the response variable). However, we do not think that the models underfit the maximum likelihood estimator explaining the fit between observations and predictions. The robust regression method of MM-estimation is designed to reduce a model's variance imputable to influential outliers, and in the same process, it reduces the non-constant error variance term. The use of a robust regression analysis over the candidate models of the prior MLR analysis constitutes, in our opinion, a confirmatory analysis of the appropriateness of the overall regression endeavor, in using the objective parameters of complexity to explain and predict human performance in DDM.

Insofar as the MLR analyses using group mean performances are concerned, the majority of the candidate models exhibited very high coefficients of determination but failed to generalize to the test scenario, which hints at the likelihood of overfitting on the original dataset. On the other hand, the MLR models involving structural complexity, information complexity, and the measure of difficulty as predictors of group median performances exhibited both excellent explanatory and predictive indicators of goodness of fit, which constitutes a decent statistical support to our hypotheses concerning the capability of objective measures of complexity to explain performance in DDM scenarios. The combination of the nonlinear and non-parametric models with an analysis of group mean and median performances exhibited an even higher tendency to overfitting, with extremely low scores of the standard error of the regression. While there were sensible differences between the $RMSE$ indicators relative to the nonlinear and non-parametric methods and the use of either group mean or median performances, those differences are relatively insubstantial compared to their collective differences with the original goodness of fit indicators for the candidate MLR models. An analysis of estimator bias in nonlinear and non-parametric models may not be as straightforward as it is in traditional linear regression analysis, but the reduction in model error attributable to differences in variance has produced some interesting insights as to which statistical regression models for the complexity of DDM problems stand as better candidates to explain the human performance data.

3.4.3 Limitations and way ahead

Based on prior research results (Pronovost et al, 2014, 2015, as well as the previous chapter), there were expectations that the functional form of the regression analysis could possibly follow a rapidly declining curve, whereby an increase in the parameters of complexity capturing the intrinsic features of dynamic decision-making would lead to a sharp decline in human performance. Yet instead of obtaining a nonlinear threshold function similar to the Weber–Fechner law (Weber, 1851, Fechner, 1860) or Stevens' power law (1957) in psychophysics, we obtained partial linear, quasi-linear, and robust linear functions bridging the predictors of candidate models for the complexity of DDM with the response variable. This might be seen as less controversial after all, in light of the high intra- and inter-individual differences inherent to the macrocognitive phenomena involved in the systems thinking and complex problem-solving applied to DDM. The universal, microcognitive processes involved in attention, memory, and perception are pitched at a level of analysis closer to the border between biological and cognitive 'bands' of human information processing, while the higher cognitive functions are situated along the rational band of an integrated cognitive system consisting of a human agent interacting with a task environment (Marr, 1982, Newell, 1990, Anderson, 2002). Measures of accuracy and latency for human performance in DDM tasks are expected to exhibit considerably more variance than similar measures for tasks involving more primitive cognitive processes (Schoelles, Neth, Myers, & Gray, 2006, Myers, Gluck, Gunzelmann, & Kruskmark, 2010, Gray, 2012).

We strongly believe that the question of model accuracy to account for human performance in DDM scenarios should be based on the comparison of multiple modeling approaches and the use of various algorithms, insofar as the high variance in human performance can cause problems for traditional modeling methods such as multiple linear regression using the ordinary least squares computation to produce a linear trend in the multivariate data. Playing the bias-variance trade-off game, we have decided to attempt to reduce the error source imputable to the model's variance, by limiting the impact of multivariate outliers on the regression models, and the robust regression method indeed produced higher efficiency across all candidate models for our objective measures of complexity. Even better models of multiple linear regression were produced by using our parameters of complexity as predictors of group median performances, and extremely low model residuals were produced by combining the group median performances data with nonlinear and non-parametric methods of regression. What about our models' bias? A model's bias represents how close the functional form of the model can get to the true relationship between the predictors and the outcome. We did not tackle this issue, as we believe that we have insufficient data points to explore the entire feature space of the objective measures of complexity for DDM performance. It was in fact a deliberate choice to include nonlinear and non-parametric methods with the aim of increasing the bias of our candidate models, as they tend to reduce model variance, particularly in the presence of high multicollinearity between the regression predictors (Kuhn & Johnson, 2013).

Reducing the standard error of the regression by whichever means (penalized least squares in linear regression, nonlinear and non-parametric methods, robust and resistant regression, etc.) might help reduce a model's variance while increasing the model's bias, but we were not particularly interested in any of the candidate regression models' coefficients to begin with, besides the positive or negative relationship of individual predictors with the response variable. We were primarily interested in the possibility of producing models of complexity for DDM which could explain a good proportion of the variance in human performance, and to be able to make predictions based on such models. Future work might be interested in minimizing a particular model's bias through the use of techniques such as partial least squares, or regularization methods such as penalized least squares (ridge regression, least absolute shrinkage and selection operator, elastic nets, etc.), but this is a concern of secondary interest relative to the minimization of error due to the high variance in the human performance data.

The question remains as to which model, among the set of candidate models tested through the many types of regression analyses, is preferable? The encouraging results from the robust regression analysis and the MLR analysis using group median performances suggest that candidate models featuring C_{DDM} metrics with the information I_{DDM} and the difficulty D_{DDM} parameters are more appropriate to explain and predict human performance in DDM scenarios. We prefer the *system dynamics complexity* metric to use as the structural complexity measure. SDC_{DDM} is inspired by models from algorithmic computational complexity and structural systemic complexity (Kinsner, 2010), but also represent the particular features of DDM variables and relationships such as a differentiation criterion between endogenous and exogenous variables, as a complex system's many variables may have more than one role within a DDM problem.

On the other hand, the nonlinear and non-parametric methods, by themselves or combined with the group mean and median response variable analysis, still suggest that most if not all of the predictors are relevant in explaining and predicting human performance for DDM problems. Based on the exhaustive analyses of the previous chapter, it is expected that a research program could answer such a question by using an array of DDM scenarios where combinations of subsets of the parameter values would produce a minimum set of DDM problems sufficient to weight the impact of such predictors on human performance. The following chapter discusses our research objectives and our experimental findings, with a particular emphasis on the necessity to establish a research program implementing various configurations of the feature space of the objective parameters of complexity. This final section lays down the necessary and sufficient criteria to explore the impact of individual parameters of complexity on DDM performance, and ultimately to select the combination of parameters which would constitute the most accurate measure of complexity.

Conclusion

This chapter presents a synthesis of the findings concerning our objective measures of complexity for human performance in dynamic decision-making scenarios. We will first discuss whether the results have met the objectives and answered our hypotheses, and acknowledge the limitations of our research project. Another section presents the theoretical, methodological, and practical contributions of the concepts, models, and results we have documented in this dissertation, with an emphasis on ideas for future research and development projects. A final section presents a research program designed to explore the parameter space of objective measures of complexity for DDM scenarios in more breadth and depth, so as to weight which combinations of the characteristics of complex decision problems are determinant for the human comprehension and control of complex and dynamical systems.

Summary of the objectives

The main objective of this research has been to establish a relationship between the objective characteristics of complex decision-making and the ability of human decision makers to understand and influence the dynamical systems underlying such decision tasks. In order to achieve objective measures of complexity, we first had to identify the type of cognitive processes and functions involved in complex decision-making, and then assess among the available methods in experimental psychology which of the options were best suited to elicit human performance data in chapter 1. We scoured the literature concerning the complexity of various phenomena in order to settle for a number of concepts, models, and calculations of the complexity of a task, and then built models suitable for the tasks in our experimental design. In chapter 2, we used our objective measures of complexity to evaluate their explanatory and predictive capability, by first selecting a number of subsets of the parameters through a three-pronged approach of methodological, statistical, and machine learning criteria. Cross-validation and out-of-sample validation methods were also utilized in order to verify the generalizability of the models to another DDM scenario. Chapter 3 endeavored to assess the accuracy and overall goodness of fit of the candidate regression models, through a comparison with other regression modeling approaches operating on different assumptions than the ordinary least squares method common in experimental psychology. This third chapter also constitutes an assessment of the hypothesis of a different functional form (i.e., a nonlinear relationship) between the parameters of complexity and human performance in dynamic decision-making problems.

Summary of the results

We have found a number of characteristics of complexity for complex decision-making problems which offer reasonable grounds to explain the ability of decision makers to understand and influence complex and dynamic decision tasks. In chapter 1 we first identified the macrocognitive nature of the higher-level cognitive functions involved in complex decision-making, such as the metacognitive aspects of systems thinking. We then selected an experimental approach best suited for the macrocognitive functions of interest, namely the use of simulation-based experimentation through microworlds. Another achievement is the selection and validation of models and metrics of complexity from the literature about complex systems theory, computational complexity theory, and cognitive systems engineering, applicable to our dynamic decision-making problems.

The results concerning the modeling, analysis, and validation of objective measures of complexity relative to the empirical results were presented in chapter 2. A first issue encountered in the modeling phase was the high collinearity exhibited between the parameters of complexity describing the DDM scenarios, where some of the parameters were either intrinsically collinear as they attempted to explain redundant features of the DDM problems, while some other parameters were accidentally collinear, in that they exhibited proportionality in the increase or decrease over certain parameters merely by the design of the scenarios. It was found that a small number of 3-parameter multiple linear regression models could explain approximately 46% of the variance in the human performance data for the five DDM scenarios in the CODEM simulation environment. When the Ecopolicy test scenario was included for the purposes of cross-validation, the candidate models could explain between 37% and 40%. The analyses from chapter 2 produced MLR models with a lot of violations of the assumptions for traditional linear regression modeling using the ordinary least squares computation for the residuals, on the one hand, and those candidate models for the objective measures of complexity exhibited high scores in residuals because of the variance in the results, on the other hand.

Chapter 3 proposed alternative ways to model the relationship between the parameters of complexity and the performance scores, by means of nonlinear and non-parametric models of regression which do not have to fulfill the standard assumptions of classical linear regression. Those nonlinear and non-parametric approaches do not make assumptions about the shape of variable distributions, they are more flexible in the search for a functional form in the regression analysis, and they employ alternative algorithms to reduce the error function in computing a model's residuals. It was found that those models, while mostly successful in reducing the absolute measure of the standard error of the regression, did not so in a substantial fashion over the MLR models using the OLS error calculation. Another approach to regression modeling was used, the robust and resistant regression method using MM-estimation, was found to be substantially more effective in reducing the $RMSE$ indicator for the same candidate models of complexity measures, nearly halving the root-mean-square error relative to the original MLR models. The MM-estimation algorithm produced robust coefficients of determination ranging between 70% and 72% for the 3-parameter models (and between 65% and 79% for the same models using the Ecopolicy performance data directly in the analysis). A third method employed a restricted subset of data points (one per DDM scenario) as a response variable, using group mean and median performances, which yielded considerably higher goodness of fit indicators, but where only the models featuring the parameters of structural complexity, information complexity, and difficulty could produce both explanatory and predictive goodness of fit indicators of sufficient accuracy, and strictly as predictors of group median performance. Finally, a combination of both the nonlinear/non-parametric models with the analysis of the group mean/median performances produced the lowest standard error of the regression scores, suggesting that the nonlinear modeling of the relationships between the subset of OMC models of interest with human performance was a more accurate methodology than the OLS approach of traditional linear regression modeling and analysis.

Those results suggest that the enterprise of investigating objective measures of complexity to explain the variance in human performance in dynamic decision-making scenarios is not only hopeful from a qualitative stance, they also constitute a first attempt at quantifying the relationship between the complexity of dynamical systems and the ability of human decision makers to understand and influence such systems.

Research hypotheses and answers

Have the results been conductive of support for our hypotheses concerning the relationship between objective measures of complexity for dynamic decision problems and human performance? Yes and no. The primary objective of this research project was to develop objective measures of complexity based on characteristics of dynamic decision-making problems which can explain the variance in human performance scores, and there were two hypotheses guiding our objective.

A first hypothesis was that the objective measures of complexity would function as parameters in a multiple linear regression model explaining the variance in the human performance data, whereby higher net complexity for a DDM scenario over a given subset of complexity metrics would yield lower performance scores, while lower complexity ratings will yield higher performances. Insofar as the empirical results suggest that higher values on multiple parameters of complexity produce lower performance scores, this appears to be a reasonable assumption. However, there is no definitive result supporting this hypothesis for one parameter of complexity over another. Looking at the COIN 1 and COIN 2 scenarios, for example, the performance scores are considerably different, yet the two DDM scenarios bear exactly the same high values for structural and information complexity. They in fact differ only on measures of difficulty, action complexity, uncertainty, and system instability. So the answer is yes, the overall complexity across a number of parameters fo DDM scenarios tend to translate in a decrease in human performance in DDM scenarios, but the differences in values across the different parameters of complexity may loom larger in their effect on performance, which is an research question for which we do not possess sufficient data to provide an answer.

A second hypothesis concerned the functional form of the relationship between the objective parameters of complexity and the performance scores in DDM scenarios. Based on earlier experimental results (Pronovost et al, 2014) and on assumptions about the nature of human cognitive discrimination capabilities, such as the differential thresholds, or limen, of perception in psychophysics (Weber–Fechner law, Weber, 1851, Fechner, 1860, and Stevens' power law, 1957), it was expected that the relationship between the parameters of complexity and the performance in DDM scenarios might not be linear in form. Yet the candidate models of multiple linear regression using performance group medians exhibited very high goodness of fit indicators, on the one hand, and their combination with nonlinear and non-parametric regression models only improved the goodness of fit in a marginal fashion. Moreover, the robust regression analysis in chapter 3 for the same candidate models produced lower residuals and higher coefficients of determination than in the original MLR analysis in chapter 2, suggesting that the decreased sensitivity to influential observations enhanced the likelihood of the functional form being a linear relationship. So the answer to that second hypothesis is negative, insofar as we were not able to produce a better fit representing something similar to a rapidly decreasing threshold function, given the structure of the observations.

Tackling dynamic decision-making with simulation-based experimentation

Has the conduct of experimental research using microworlds facilitated our objectives and produced useful results? The study of the relationship between the objective measures of complexity for dynamic decision-making problems and human performance could not have been conducted in any other way. Complex decision-making manifests itself in the problems and everyday challenges faced by novices and experts alike, in the complex sociotechnical systems of the modern day workplace. This type of phenomenon is best described at the level of macrocognition, that is, spanning the levels of analysis of low to high level cognitive processes and functions, at the frontier of the 'cognitive band' of elementary tasks occurring over seconds, to the 'rational band' of complex tasks, strategies, and procedures occurring over minutes, hours, and even days. The traditional experimental approach in cognitive psychology would circumscribe the study of complex decision-making to simplified, controlled, and often unrealistic settings.

For example, earlier versions of serious games aiming to educate students about policymaking, such as Ecopolicy, were first based on card games and boardgames (Vester, 1983). It would not be surprising thus, given the nature of the experimental approach, if the observations and analyses at the time were focused on the results of complex decision-making, rather than on the entire process. Moving from microcognition to a macrocognitive framework, simulation-based experimentation introduces many desirable properties of naturalistic settings where important features of complex decision-making occur, such as metacognition and systems thinking. Dynamic decision-making is precisely interested in complex decision-making from the point of view of a processed-based experimental approach, focused on information-rich, complex sequences of interactive behavior between a decision maker and a challenging, ever-changing, and sometimes opaque task environment. Only synthetic environments such as microworlds can afford this level of fidelity for a realistic, immersive, and interactive task environment suitable for research.

One particular aspect of the use of microworlds for research on DDM which has not yet been emphasized in the research project is the emotional, motivational, and physiological features of DDM as they relate to cognition and behavior. DDM scenarios are used to collect data through a simulation environment about a decision maker's information-seeking behaviors, such as the types of information sought and accompanying latencies, the choices of interventions, as well as the type of, and time spent on feedback from decisions, etc. For all of the benefits of the simulation environment capabilities, such as data collection, scenario editing and parameter tuning, as well as affording information-rich representations, DDM problems are complex beasts. They may involve long experiments, they may induce boredom and complacency, and they may even be so challenging that they become frustrating for the decision maker.

Those undesirable emotional, motivational, and physiological states (from the point of view of both the researcher and the participants!) can be mitigated by the gamification of the DDM scenarios, which is why those experiments are framed as serious games, whereby a ludic component is added to the framing of the DDM problem. Whether the microworld is used for research or for pedagogical purposes, another benefit of simulation-based experimentation is that it can not only be tailored to the requirements and level of fidelity proper to a given task of interest, but it can also facilitate the engagement and involvement of participants in a way that distinguishes it from research experiments in traditional laboratory settings. DDM scenarios are presented as turn-based strategy games, often in an experimental or pedagogical design of incremental challenge and complexity, soliciting from the decision-maker the resourcefulness, the motivation, and the attention that match the kind of engagement that real-world complex problem-solving demands from us.

Simulation-based experimentation on dynamic decision-making through the use of gamified scenarios in microworld environments allows the experimenter to collect meaningful data about the evolution of the DDM problems over time, including the system's states, as well as the information-seeking behaviors of participants, their decisions, and overall performance. It takes advantage of the same benefits as the classical laboratory-based experimentation, such as a tight control over the experimental conditions, leading to the reproducibility of the experimental design and the possibility of corroborating results, as well as the tuning and refinement of hypotheses in order to further the research. On the other hand, it affords a realism and richness nearly matching the intricacies and immersion of real-world problems, supporting greater ecological and external validity without sacrificing the prospect of internal validity.

Objective measures of complexity for dynamic decision-making problems

What are the objective measures of complexity for dynamic decision-making problems, which ones were kept in the modeling and analysis endeavors, which ones were dropped, and why? The OMC metrics were derived from a literature review on three primary sources of information about complex and dynamic systems: namely, from (i) branches of algorithmic (computational) and structural (systems) complexity theory (Kinsner, 2010), from (ii) recent studies in cognitive informatics interested in the impact of complex informational structures on the human perception of complexity (Wang, Kinsner, Anderson, et al, 2009, Kinsner, 2010, De Silva, Kodagoda, & Perera, 2012, De Silva & Kodagoda, 2013, De Silva, Weerawarna, Kuruppu, et al, 2013), and from (iii) a review of the literature on task complexity from the point of view of experimental cognitive psychology and cognitive ergonomics (Liu & Li, 2011, 2012, 2014, Bedny, Karwowski, & Bedny, 2012, Stouten & Größler, 2017).

The objective measures of structural complexity (cyclomatic complexity, coefficient of network complexity, and system dynamics complexity) are the most straightforward and intuitive metrics, as the graph-based representation of the relationships between variables is both ubiquitous and evocative of the various degrees of structure in a stock and flow model (Kinsner, 2010). We have emphasized the importance of structural complexity C_{DDM} throughout this document, and it turns out that all three variants are good predictors of performance in DDM scenarios. However, since the system dynamics complexity metric SDC_{DDM} was tailored specifically with dynamic decision-making tasks in mind (as opposed to the algorithmic complexity metrics), it is semantically more favorable to our designs. Moreover, the MLR models using SDC_{DDM} were more appropriate in capturing the differences between DDM scenarios based on the number of endogenous and exogenous variables, such as in the case of the particular discrepancies between the CC_{DDM}, CNC_{DDM}, and SDC_{DDM} measures for the *Cybernetia* scenario in Ecopolicy.

The information complexity metrics I_{DDM} and A_{DDM} are also inspired and motivated by the literature on algorithmic and sytems complexity measures, but with a particular focus on informational processes pertaining to human activity and organization (Henry & Kafura, 1981, Cardoso et al, 2006, Liu & Li, 2012). Having a metric quantifying the number of inputs, interventions, and outputs for a stock and flow model, a business process model, or a task analysis model is critical to predicting human comprehension and influence over complex and dynamic systems. The information flow metric I_{DDM} was a successful predictor of human performance when combined with a structural complexity C_{DDM} parameter and the very different notion of difficulty, or D_{DDM}. The objective measure of difficulty is a completely different category of intrinsic complexity in achieving a DDM task, insofar as it is more a measure of distance to goal achievement (and to the opposite end state of failure). A complex system, whether minimally or highly complex in structure or information flow, may lie in a state of proximity to an ideal state leading to goal completion, or much closer to critical states leading to failure (Funke, 1988, Gonzalez et al, 2005, Bedny et al, 2012).

The same notion of distance to desirable and undesirable system states can be used to describe a system's propensity to inertia, or inversely, to catastrophic changes, as captured by the stability metric S_{DDM}, inspired by Osman and Speekenbrink (2011). The stability of a DDM scenario is determined by its evolution, or lack thereof, in variable states in absence of an agent's interventions. The intransparency, fuzziness, and presence of singular occurrences in a complex and dynamic system were summarized as a measure of uncertainty U_{DDM}, a rather popular feature of experimental research on DDM (Osman et al, 2015), and all quantifiable sources of uncertainty were included in the present research project (uncertainty pertaining to variable states, in relationships and consequences from interventions, as well as surprise events affecting the DDM scenarios). A few measures were dropped along the analyses and the modeling endeavor, including a measure of nonlinearity L_{DDM} for the relationships among, and interventions on, a system's variables (inspired by Sterman, 1989, Soyer & Hogarth, 2015, and Özgün & Barlas, 2015), which was meant to capture the proportion of relative departure from linearity for those relationships. Unfortunately, L_{DDM} was dropped for quantitative and statistical purposes, as the DDM scenarios did not adequately differentiate among values for this parameter, and the correlation with performance scores did not warrant further investigation.

The OMC parameters were based on metrics involving compounds of features, originally put together in research on cognitive informatics, such as the Cognitive Functional Size (Shao & Wang, 2003), the Cognitive Information Complexity Measure (Kushwaha & Misra, 2006), and the Cognitive Systems Complexity (Wang, 2007, 2009). Those models of complexity were based on structural and informational characteristics in the domain of engineering of algorithms, as well as the associated cognitive load in recognizing them. The concatenative nature of those measures is reflected in the OMC parameters throughout this document, even though the OMC models reflect an effort at breaking down the above-mentioned compound measures of algorithmic complexity. Our parameters of complexity nevertheless constitute a degree of concatenation among the many elements featured in dynamic decision-making problems, as a complete breakdown of those features would amount to an explosion of parameters for the multiple linear regression analyses. Indeed, the isolation of each and every quantifiable characteristic and property of a DDM scenario would not make sense to analyze on their own. We thus attempted to maintain a certain degree of semantic consistency among our objective measures of complexity, so as to group the isolated features into cogent, meaningful units of organization. That said, such parameters could certainly be arranged differently, which is reflected in the way the three variants of the structural complexity metrics, the two information complexity measures, and the cognitive weight complexity model feature redundant characteristics of DDM problems. For example, what we portray here as structural complexity and informational complexity metrics mirror the models of operational complexity and architectural complexity in Wang's Cognitive Systems Complexity (2009), in a imperfectly overlapping fashion. That is, the two metrics we use have a "many-to-many" mapping of characteristics with Wang's two sub-metrics composing his overall model of cognitive complexity for information systems. What goes into a given metric depends not only on aiming towards meaningful units of organization, it also requires a bit of wrangling in order to accommodate a domain of knowledge or application. Concerns about construct, content, and criterion validity are therefore still an open question.

Did the modeling of dynamic decision-making scenarios according to a set of objective parameters of complexity support our goal of relating DDM problem complexity to the observed variance in human performance? The answer is definitely positive, although with multiple caveats. The quantification effort over ten parameters of complexity led to multiple analyses, from a correlational analysis to a principal component analysis in order to first observe the redundancies between the objective parameters. A preliminary analysis of the relationship between the parameters of complexity and the performance data revealed that while most parameters could predict the order of group mean and median performances by DDM scenario (based on the results from Spearman rank correlations, where negative correlations between monotone decreases in performance for increases in the parameters of complexity were expected), some parameters did not warrant their utilization in regression analyses, such as the aforementioned situation concerning the nonlinearity L_{DDM} metric. The difficulty D_{DDM} parameter was also problematic for the prediction of the scenario performance rank order, but was kept for further analyses based on its semantic relevance in prior research (Pronovost et al, 2014) and the intuition that the difficulty ratings for our DDM scenarios interacted with the structural complexity measures in affecting the performances.

Since the quantification of the objective features of complexity was inexorably bound to the particular design of the few DDM scenarios available for the experimentation, some redundancies between the parameter sets were found to be highly collinear. A first phenomenon explaining the presence of this collinearity is intrinsic, or proper to the characterization of the measures of complexity themselves, as those measures were established in a way to reflect earlier categorizations and taxonomies of the features of complexity for anything from the structure of graphs to the information flow in an algorithm. The main reason as to why the cognitive weight complexity metric W_{DDM} was ultimately dropped from the analysis and modeling endeavor was precisely because of this intrinsic collinearity with parts of the three measures of structural complexity, as well as parts of the two information complexity measures. The second type of collinearity was the accidental collinearity attributable to the design of the DDM scenarios. Scenarios which exhibited high measures of structural and information complexity tended to feature high scores on the other types of complexity measures. Such collinearity could be much lower in future DDM scenario designs. Additional scenarios could feature different magnitudes and ratios over select parameters of complexity in order to compare the scenarios among themselves.

To mitigate the problem of redundancy-induced collinearity, an exhaustive search of the parameter values for the DDM scenarios was operated over six subsets each containing five parameters of complexity in a linear regression analysis, sampling the parameter space of complex decision problems as it relates to performance. Six MLR models were retained for their relative index of goodness of fit in explaining the variance in human performance. Those subsets of 3-parameter candidate models of regression were able to explain nearly half of the variance in the performance variable, while the inclusion of another scenario in a stratified n-repeated k-fold cross-validation method predicted nearly 40% of the variance in performance when an arbitrary scenario was held out as predicted data from the training set of the other five DDM scenarios. Since the parameter set for each DDM scenario is a set of numbers held constant, even if it was sampled from a hypothetically indefinite pool of random effect factors, the task of finding a multiple linear regression model for the performance scores amounts to drawing a line across a 2-axis Cartesian chart where the domain (the x axis) features the scenarios ordered in a certain way to increase the 'overall' complexity (the multiple regressors) and where the y axis features the means for the performances in the five or six scenarios. We therefore had every reason to hope for simpler regression models featuring less parameters than groups of data points. The MLR models representing the objective measures of complexity were thus chosen as the best possible increase of fitness relative to their complexity, retaining a minimum of predictors where possible.

Judging from the results of the regression analysis relative to the distribution of performance scores for the DDM scenarios, it appears that differentiating between (i) similar performance scores for scenarios of considerable differences in most parameters of complexity on the one hand, and (ii) different performance scores for scenarios which are nevertheless 'closer' in the parameter space of complexity on the other hand, rests upon a complex interaction between the parameters of complexity for a DDM task. That is, the performance outcome which can be anticipated for a DDM problem as a function of the relative proximity in values for one or more given parameter(s) does not warrant knowledge about the outcome in a similar DDM problem. Extremely small differences in initial conditions for the COIN 1 and COIN 2 scenarios over four parameters of complexity produced wildly different performance distributions. The human performance in DDM scenarios exhibits a high sensitivity to small nuances in parameters of complexity that could appear quite secondary to the intuitively important notions of structural and information complexity. The distance between the initial values of a dynamical system and the values representing both success and failure in a DDM problem can facilitate, or completely undermine, a decision maker's likelihood of achieving his or her goals. Similarly, given two DDM problems of equal structure, information flow, and difficulty, toying around with the relative nonlinearity of the relationships, with the amount of opacity, fuzziness, and uncertainty in the available information, or with the likelihood of a system to barely change under inertial conditions can be expected to lead to vastly different central tendencies and degrees of dispersion in the human performance scores. For all of the aforementioned reasons, the entire set of objective measures of complexity should be considered in future research investigating the relationship between the complexity of DDM tasks and human performance.

Working around the assumptions of linear regression analysis while dealing with the bias-variance trade-off

Were the attempts to circumvent the violations of assumptions for the linear regression analysis and to reduce the standard error of the estimate successful? In order to be reliable, the multiple linear regression models must observe a certain number of prior assumptions about the distributions and relationships between variables. Our linear regression analyses exhibited non-normality in our candidate models' residuals, non-constant error variance, or heteroscedasticity, and a number of multivariate outliers, or influential observations. Additionally, the high levels of heteroscedasticity, collinearity among independent variables, and overall variance in the human performance data induced high values for the standard error of the estimates, thereby making the accuracy of the models questionable. Compared to run-of-the-mill studies of experimental psychology applied to cognition, where low-level, microcognitive processes are studied at the boundary between the 'biological band' and the 'cognitive band' and produce low variance in intra- and inter-individual studies with regards to accuracy and latency, the macrocognitive processes and functions involved in complex decision-making exhibit high degrees of intra- and inter-individual differences.

A first attempt at circumventing the assumptions of the MLR models for the objective measures of complexity was to depart from the error function reduction method of ordinary least squares in favor of a robust and resistant regression analysis, which combines the statistical efficiency of the linear regression models with a high breakdown point in the face of influential outliers (Yohai, 1987). The statistical mean has a very poor breakdown point of 0%, i.e., a single, highly influential outlier can significantly affect the regression model's slope, whereas a robust estimator uses a MM-estimator, a maximum-likelihood estimation with a breakdown point similar to that of a statistical median, namely 50%. Distant outliers can yield unduly large residuals since they are squared in the OLS computation, as the linear regression analysis seeks to account for every data point without any consideration for their relevance or influence on the overall model. The optimization algorithm for the reduction of the error function in the robust regression analysis uses an iteratively reweighted least squares computation for the influential observations, minimizing their impact on the regression line. The robust regression analysis produced a positive account of the relationship between the objective measures of complexity and the human performance data, insofar as it circumvented the assumptions of the linear regression analysis and reduced the standard error of the regression models considerably. Robust coefficients of determination for regression suggested that the 3-parameter candidate models from the MLR analysis greatly improved model fit with the performance. This was achieved at the price of a 'selective insensitivity' to approximately 10% of the data points in our observations, which were weighted down by the IRLS optimization algorithm to values which are rejected in fitting the regression line to the observations.

Another way of circumventing the assumptions of the MLR models was to use group mean and median performances in order to literally eliminate all sources of variances stemming from differences within-condition, or intra-group differences. The inflated scores for the goodness of fit indicators reflected positively on the analysis, particularly with respect to explanations and predictions of the variance in performances using the high breakdown estimator that is the group median. Those group mean and median performances were also used in combination with nonlinear and non-parametric methods for regression analysis, inspired by the literature of computational learning theory. The main highlight of this approach is the possibility of avoiding the violations of assumptions in the traditional linear regression analysis by relaxing the constraints on the way the machine learning algorithms model the relationships among variables, and the way they compute the statistical error reduction function which produces the regression model residuals. The mathematical exploits of nonlinear and non-parametric models stem from their capability to approximate quasi-linear functional forms, transform the irregularities of regression models into linearizable feature spaces through kernel functions, or classify multivariate data based on similarity functions operating in higher dimensional space. It was found that most nonlinear and non-parametric methods of regression used in combination with the group mean and median performances produced lower residuals across all models. The results from all three alternative regression modeling endeavors strongly support the idea that using the objective parameters of complexity to explain and predict human performance in complex decision-making can produce efficient statistical estimators.

Contributions

There are a few literature reviews on the specific topic of dynamic decision-making (Frensch & Funke, 1995, Hsiao & Richardson, 1999, Qudrat-Ullah, Spector, & Davidsen, 2008) documenting the wealth of research on the relationships between complex and dynamical systems, simulation-based experimentation, and dynamic decision-making performance, yet for all that research, the topic remains seriously underdetermined according to Karakul and Qudrat-Ullah (2008). In the introduction to Qudrat-Ullah, Spector, and Davidsen's *Complex Decision Making: Theory and Practice* (2008), Bar-Yam mentions that

> The field of complex systems provides a number of sophisticated tools, some of them conceptual helping us think about these systems, some of them analytical for studying these systems in greater depth, and some of them computer based for describing, modeling or simulating them.

It is hoped that our modest characterization of the relationship between the objective parameters of complexity for dynamic decision-making and human performance will generalize not only to other DDM research projects, but may be used for decision complexity and task complexity in diverse areas, such as in evaluating the complexity of business process models, the design of algorithms, and even in the evaluation of the complexity of human-technology interaction in human-systems integration projects, in human factors and cognitive ergonomics evaluations, and in the overall practice of cognitive systems engineering. The following sections highlight some ideas for the potential contributions of the objective measures of complexity in dynamic decision-making research from the point of view of the theory, the methodology, and practical applications.

Theoretical contribution

Complexity theory has as many models as it has applications, which complicated our literature review in the matter (Coveney & Highfield, 1995, Heylighen, Bollen, & Riegler, 199, Northrop, 2010). We have borrowed the concepts of algorithmic computational complexity (from algorithmic information theory and computational complexity theory) and of structural systemic complexity (from fields such as graph theory, programming complexity, and network complexity), in order to parameterize the structural and informational characteristics of dynamic decision-making problems, based on Kinsner's taxonomy (2010). Since our interest was primarily on the features of the DDM task proper which would bring about differential results in human performance scores, we followed Karakul and Qudrat-Ullah's (2008) taxonomy of *decision task characteristics* and *decision-making environment characteristics*. We have therefore included structural measures of complexity, metrics for the information flow and the cognitive weight, as well as additional parameters to account for difficulty, nonlinearity, uncertainty, and instability.

Yet using the measures of complexity from information theory, the theory of computation, and graph theory for DDM problems would not have been possible without the prior attempts to adapt complexity metrics and models to business processes (Cardoso et al,2006), the evaluation of algorithms from a cognitive point of view (Shao & Wang, 2003, Wang, 2009, De Silva et al, 2012), and earlier attempts to frame the concept of the complexity of a task (Liu & Li, 2011). From a review of complexity models based on the viewpoints of structure, resource requirement, and interaction, Liu and Li extracted six components of a task model (goals, inputs, process, outputs, modeled system structure, time) and ten dimensions for a model of the complexity of a task (size, variety, ambiguity, relationship, variability, unreliability, novelty, incongruity, action complexity, and temporal demand). Stouten and Größler (2017) mapped those criteria and characteristics to stock and flow problems as they appear in dynamic decision-making tasks, which further inspired and encouraged our ongoing investigation into the objective features of complexity for DDM scenarios as they relate to human performance.

Taking stock of Karakul and Qudrat-Ullah's, Liu and Li's, and Stouten and Größler's prior characterization of the objective features of complexity which could be leveraged to measure the impact of dynamic decision-making tasks on human performance, we have isolated ten parameters of complexity which represent objective characteristics of our DDM scenarios. We have validated the idea of using different combinations of some of those parameters to explain and measure the impact of decision complexity on performance through our findings that even with a few DDM problems and a small set of observations, multiple linear regression models could explain a significant proportion of the variance in performance. Our objective measures of complexity are thus not only corroborating Karakul and Qudrat-Ullah's, Liu and Li's, and Stouten and Größler's ideas concerning the characterization of the complexity of decision tasks from a qualitative point of view, they also constitute a first answer to Stouten and Größler's (2017) open challenge for laboratory experiments that aim at assessing the understanding of, and influence over dynamical systems from a detailed, fully parameterized quantitative approach. Taken together then, those ten measures of complexity constitute both adaptations of existing metrics (structural and information complexity, cognitive weight complexity, the measure of uncertainty) and novel ways (system dynamics complexity, as well as the measures of difficulty, nonlinearity, and stability in their current form) of assessing the complexity of dynamic decision-making scenarios through simulation-based experimentation.

Methodological contribution

Modeling the objective characteristics of complexity for dynamic decision-making problems could be greatly beneficial to research psychologists who are targeting a precise feature of complex decision-making and want to control for the other aspects of complexity as well as for potentially confounding factors in their experimental design. A matrix of objective parameters of complexity, and knowledge about their relative contribution to human performance, could therefore help in designing DDM experiments tailored to the particular needs of researchers, in order to support the accuracy of research questions and the resolution of the hypothesis and variables of interest involved in the DDM experimentation. A particular example of interest to the author is the possibility of mapping the features of dynamic decision problems which beget a breakdown in human performance, to the minimal complex systems of Funke (2014, see also the Figure 1 in the introduction that illustrates the MCS approach). Funke's method of assessment of the impact of DDM complexity on human performance is incremental over the number of variables and relationships, and complementing his method with additional parameters representing objective features of complexity could potentially yield interesting results, if only to control for features such as the nonlinearity of the relationships between variables. Progressively introducing those features of complexity in the MCS scenarios could potentially reveal a certain frontier, or threshold, at which the "minimal complex systems" actually become "complex problem solving" due to changes in one or more parameter(s).

Another methodological aim which could benefit research on DDM is through the support of the pedagogy concerning systems thinking and metacognition in general (Gonzalez & Dutt, 2011, Sterman, 2014, Fischer, Greiff, Wüstenberg, et al, 2015, Qudrat-Ullah, 2015). Training on dynamic decision-making could be leveraged by incorporating the concepts of the objective measures of complexity in an intelligent tutoring system featured in interactive learning environments. Scenarios of incremental overall complexity could be fine tuned by targeting particular features of DDM problems which are to be the focus of a given lesson, such as the introduction of feedback loops from one scenario version to another, then conditional relationships, delayed relationships, etc. Previous research (Pronovost et al, 2015, Lafond et al, 2016) has demonstrated that the performances of decision makers in three incrementally more complex versions of the same DDM scenario (the Arctic 1, 2, and 3 scenarios of chapter 2) with an ITS were comparable (for the final, more challenging version of the scenario) with performances in an experimental condition of 'implicit learning", where the decision makers were only told to play with that same scenario (Arctic 3) for three hours. In the ITS condition, the decision makers were only exposed twice to each of the three versions of the Arctic scenario, and their final performance in Arctic 3 was not significantly different from performances of decision makers in their final trials for the implicit learning condition. This suggests that there can be a bootstrapping effect on DDM performance if decision makers are trained on systems thinking, using incrementally more complex DDM scenarios. The objective measures of complexity could enhance the tuning of DDM scenarios in support of teaching and training for systems thinking precisely because they not only break down the intrinsic characteristics of the decision task, they also do so in a quantitative approach, beyond a qualitative assessment.

The nonlinear and non-parametric regression analyses from chapter 3 also promoted the idea of applying machine learning methods to computational statistics. The field of computational learning theory has given statisticians the opportunity of using inductive methods for the estimation of relationships among variables through supervised learning algorithms. The results in chapter 3 support the idea of exploring the goodness of fit of regression functions approximated through sophisticated pattern recognition methods. Since the machine learning algorithms exhibit different properties in the way they model variable relationships and statistical error, those nonlinear and non-parametric models can capture variable relationships in a way that is not always amenable to traditional linear regression using the ordinary least squares computation of the discrepancy between observations and predictions. We presented methods using piecewise regression modeling with splines, interpolation and weighting through smoothing kernel functions, mappings of multivariate data combinations into linearizable feature spaces, intermediate transition functions to map nonlinear inputs to linearized output functions, and the classification of multivariate data based on similarity functions such as the Euclidean distance. There are plenty more options available for statistical modeling and analysis, such as decision tree learning, reinforcement learning algorithms, genetic algorithms, Bayesian networks, and a great number of variations of the artificial neural network approach, such as the "deep learning" algorithms featuring multiple, hierarchically structured hidden layers. The traditional tools of mathematical psychology, favored in psychophysics, behaviorism, and psychophysiology could therefore be complemented by the modern tools and models of computational statistics (Gentle, 2002), statistical learning theory and machine learning (Vapnik, 1998, Friedman, Hastie, & Tibshirani, 2001, Kuhn & Johnson, 2013), and computational psychology (Boden & Mellor, 1984, Sun, 2008, Busemeyer, Wang, & Townsend, 2015).

Practical contribution

Stouten and Größler (2017) suggest that research about the complexity of dynamic decision-making tasks could lead to improvements in task design, or in the development of more appropriate strategies in the context of such tasks. This could be extrapolated to better interaction design, the improvement of standard operating procedures for critical infrastructures, as well as research and development for the design and evaluation of more ergonomic technologies in complex sociotechnical systems, such as dynamic information visualization solutions for complex problems. Assessing software complexity was the purpose of many of the complexity metrics reported in the introduction, such as the cyclomatic number, the coefficient of network complexity, Halstead's software metrics, and the four variants of cognitive weight complexity, while the interface complexity measure was a suggestion by Cardoso et al (2006) to assess business process complexity. By that logic, any task design, such as the more formal definition of interaction design in the applied sciences of industrial design, human-technology interaction, and human factors and ergonomics, could leverage knowledge about the features of complexity that characterize their domain of application. Abstract or concrete, the products of research and design such as a business procedure, a set of tasks to be performed to achieve a result, the multiple ways of interacting with a particular piece of technology, the architecture of a software application, or even a strategic plan for policymaking or wargaming, etc, could capitalize on qualitative and quantitative assessments of the objective measures of complexity proper to the challenges they raise for a decision-maker. Beyond research and development, even the tools used in the analysis of human-technology interaction could incorporate metrics for decision complexity, such as hierarchical task analysis, cognitive work analysis, business process modeling, etc. There is even an interest in modeling human behavior in accordance with complex and dynamic decision-making models for the agent-based modeling (ABM) and simulation of virtual agents in synthetic environments (Pronovost & West, 2007, 2008, Pronovost, 2009, 2010, 2012, West & Pronovost, 2009, Zobarich, Miller, Kramer, Pronovost, & Kelsey, 2011, Zobarich, Pronovost, Torunski, & Unrau, 2011ab, Handel, 2016).

Another practical application of the objective measures of complexity for dynamic decision-making is the development of competency evaluations programs. Using interactive learning environments to facilitate knowledge acquisition and general performance in DDM scenarios can be repurposed to the evaluation of systems thinking as a required competency for certain work positions (Spector, 2000). Serious games are being utilized for professional training, skill-based education, and competency evaluation throughout the world, as featured in showcases and events such as the Serious Games Showcase & Challenge (SGS&C) at the Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), an annual event organized by the US National Training and Simulation Association. The use of modeling and simulation for training has been around for decades, particularly in the defense and aerospace domains, and span simulation-based training environments ranging across all of Gray's (2002) taxonomy, from high-fidelity simulations of complex systems to scaled worlds and microworlds. Ecopolicy is a particular example of a serious game developed to train people on systems thinking as part of a UNESCO program. Originally attached to a study named *Urban systems in crisis - Understanding and Planning Human Living Spaces by the Biocybernetic Approach*, Ecopolicy has been used in certain parts of the world (notably in Europe) to train students and policymakers about the complex relationships between socioeconomic issues and the wellfare of the ecological environment (Harteveld & Drachen, 2015). Karakul and Qudrat-Ullah (2008) suggest that while knowledge about human-technology interaction can help design better interactive learning environments suitable for dynamic decision-making, more research on the nature of DDM itself is required to support the evaluation of competencies. The objective measures of complexity could be leveraged in the evaluation of what the authors distinguished as the three targets of DDM research, namely task performance, task knowledge acquisition, and transfer learning. A psychometric approach to DDM which incorporates notions of performance, knowledge acquisition, and transfer learning relative to measures of complexity, difficulty, uncertainty, and other parameters could be tailored to training needs analyses (Moore & Dutton, 1978), for personnel selection, based on the knowledge (technical and non-technical), skills, and aptitudes taxonomy of researchers interested in organizational and team training (Salas, Cannon-Bowers, & Johnston, 1997, Salas, Burke, & Cannon-Bowers, 2002, Cunningham, 2008).

A third area of practical application for the objective measures of complexity is in the design and implementation of decision support systems for problems exhibiting features of DDM in complex sociotechnical systems (Gonzalez, 2005). Complex decision-making processes supported by computational decision models featuring information on feedback processes, nonlinear relationships, and delayed effects could potentially improve a decision-maker's performance, and Qudrat-Ullah, Davidsen, and Spector (2008) suggest that:

> One of the most important sources of such information is the outcome of both the model building process and the application of the model of the complex system. Modeling supports decision-making by providing specific "what-ifs" scenario analysis opportunity to the decision makers in a "non-threatening" manner.

Decision support systems of various degrees of information and decision automation and transparency (Parasumaran, Sheridan, & Wickens, 2000, Rovira, McGarry, & Parasuraman, 2007, Hancock, Jagacinski, Parasuraman, Wickens, Wilson, & Kaber, 2013) exist in diverse areas such as emergency response management and contingency planning (Greenley, Pronovost, Race, Kelleher, Graham, & Chawla, 2009), clinical support in medical diagnosis (Wright & Sittig, 2008), policymaking (Buurman & Babovic, 2016), etc. Perceived task complexity has already been accounted for as a determinant of the efficiency of information systems (Marshall & Byrd, 1998, DeRosa, Grisogono, Ryan, & Norman, 2008, Aboutaleb & Monsuez, 2015, Salado & Nilchiani, 2015, Wu, Fookes, Pitchforth, et al, 2015), and task complexity is a primary topic of interest for decision support designers from a cognitive systems engineering perspective (Hollnagel & Woods, 2005, Lintern, 2005, Militello, Dominguez, Lintern, & Klein, 2010). Dynamic decision-making is a growing topic of interest in the development of decision support technologies to enhance the analytical reasoning skills of decision makers involved in complex sociotechnical systems (Tremblay, Gagnon, Lafond, et al, 2017). The objective measures of complexity could be integrated as parameters in complex decision models to evaluate options for a decision maker, or in the design of support technologies which takes those parameters as constraints on human-systems integration efficiency (Parasuraman, Sheridan, & Wickens, 2008, Righi & Saurin, 2015, Schöttl & Lindemann, 2015).

Limitations and way ahead

A number of constraints and shortcomings were encountered during the investigation of objective measures of complexity for dynamic decision-making performance. We present a summary of those drawbacks relative to their appearance in this document. Concerning methodological issues, the experimental design and the nature of the tasks prohibited the use of a counterbalanced order for the DDM scenarios. The length of scenario experimentation is a prohibitive factor to data collection, and the order of execution of DDM scenarios must inexorably involve a path from simpler to more complex scenarios in order for participants to learn the basic concepts of systems thinking, as well as to familiarize themselves with the interface of a microworld. Running participants through the more complex scenarios first, such as COIN, is useless for the purposes of experimentation, as the performances are assured to be abysmal.

Learning effects are unavoidable, yet for highly complex DDM scenarios, this is not a real problem, as multiple experimental conditions using facilitation (through the support of intelligent tutoring systems, pre- and post-experimental briefings, pedagogical material, training and rehearsal, etc.) have not been found to significantly increase the average performances for scenarios like COIN (Sterman, 1989, 1994, Paich & Sterman, 1993, Pronovost et al, 2015, Lafond et al, 2016). We have nevertheless taken the additional precaution of using the performance data for participants with the least prior exposure to DDM scenarios, given the different experimental conditions to which the participants were assigned.

In the DDM scenarios used for the analyses in chapters 2 and 3, some of the features of complexity turn out to be semantically related, and had to be segregated into subsets of models in the regression analysis to be compared between themselves. Additionally, some parameters are also accidentally highly collinear due to the design of DDM experiments, such as in the case of Arctic 3 and COIN 2, where structural complexity, information complexity, and measures of difficulty were highly correlated between themselves. The parameters for which the collinearity is accidental were preserved for the entire analyses, while the semantically related parameters were isolated from one another as best as was possible.

Another issue is the violations of the assumptions for classical linear regression because of the non-normality of the model residuals, the non-constant error variance, and the presence of influential observations. On the one hand, we have found no particular reason to eliminate the observations which were causing the heteroscedasticity, nor the multivariate outliers. The high degree of inter- and intra-individual variability in the human performance data for dynamic decision-making naturally exhibits such undesirable phenomena in small samples, so there was no cogent methodological motivation to eliminate the outliers. Issues were raised in the literature on statistical analysis concerning the relevance of using coefficient of determination for ordinary least-squares regression (or R^2_{OLS}) in the context of comparing models of MLR between themselves and/or with other types of regression analyses, such as our nonlinear and non-parametric models (Huber, 1964, Kvålseth, 1985, Kuhn & Johnson, 2013).

A statistical method of robust regression was therefore employed to mitigate those shortcomings, favoring the more resistant properties of sample medians to outliers with their high breakdown point, and the computation of a coefficient of determination using weighted least-squares, or R^2_{WLS} (Kvålseth, 1985, Willett & Singer, 1988, Renaud & Victoria-Feser, 2010). Yet the measures produced through the various calculations of coefficients of determination are merely measures of correlation, and not measures of accuracy (Kuhn & Johnson, 2013). Those measures of squared multiple correlation are relative metrics to compare similar models (relative to the number of predictors, and sample size), so an absolute measure of model residuals such as the standard error of the regression ($RMSE$) is more favorable to compare different regression models (Tabachnick & Fidell, 2012, Kuhn & Johnson, 2013).

Complications arose in the regression model validation phase. The dataset prevented any attempt at hold-out sampling, bootstrap, or leave-one-out ($LOOCV$) validation, because of the rank-deficient fit of the variance-covariance matrix for the independent variables. That is, the parameters of complexity for each and all scenarios are a small set of parameters which are held at constant values, relative to the variance in the dependent variable. Moreover, the highly collinear nature of the parameters further exacerbates the rank deficiency phenomenon. Adding another scenario with a few constant parameter values did not facilitate the validation procedure to fit a MLR model to the data, as there was insufficient information contained in the data to estimate such models. A stratified, n-repeated k-fold cross-validation method was employed instead. The use of group mean and median performances for the MLR and nonlinear/non-parametric analyses of chapter 3 eliminated the issue of rank deficiency, as the multicollinear nature of the regression parameters no longer underdetermined the possibility of conducting a computation for the maximum likelihood estimation of obtaining the observations, given the set of parameters.

Another hypothetical limitation concerning the interpretation of the results is that the nonlinear and non-parametric regression methods are sometimes overly complex for the task at hand (the support vector regression algorithm and the artificial neural network approach tend to complexify analyses for which a simpler linear regression analysis could potentially yield reasonably accurate models), sometimes too opaque (the black box models of artificial neural networks are hard to scrutinize in order to find a cogent functional form for the relationship between variables), or perhaps even too liberal in some cases (less constraints of the function approximation and a multiplicity of tuning parameters could potentially overfit the relationships in the data, such as in the LOESS algorithm). A larger issue is that there is no way to know so, short of a complete literature review on the benefits of machine learning techniques relative to classical linear regression, and experimentation using multiple layered comparisons of each machine learning method with an array of tuning parameters to test on larger datasets.

In the present research project, some of the machine learning algorithms could not even be used with the suggested default tuning parameter values (e.g., the LOESS method) because of non-invertible matrices in computing maximum likelihood-type estimators for the reduction of the error function. Similar to the rank-deficient fit issue above, the high collinearity combined with fixed (constant) parameter values across DDM scenarios (even if they are, in principle, random effect factors) causes the variance-covariance matrices to be singular (i.e., non-invertible). As most of the nonlinear and non-parametric models managed only modest improvements relative to the original MLR models (outside of their pairing with the analyses using group mean and median performances), it was deemed that they were approximating the relationship between the parameters of complexity and the human performance data in a reasonable fashion.

With regards to the high accuracy of the results for the robust and resistant regression models, were able to achieve such low ratings for the absolute measure of the standard error of the regression ($RMSE$) and high scores for the relative measure of the robust adjusted R^2 by weighting down some observations all the way to a weight value of nearly 0, effectively eliminating those influential observations altogether. The proportion of unused observations in determining the linear regression was around approximately 10%. This was nevertheless appraised as a reasonable compromise given the high variance-induced residuals in the dataset. Without any methodological motive to eliminate the some of the influential observations, we relied on the iteratively reweighted least squares method in order to find an estimator which maximizes the likelihood of making the observations given the parameters. The MM-estimation method of computation in robust regression appears to be a viable compromise between undue reduction of the model variance and the possibility of overfitting.

The largest issue is ultimately that the parameter space for the objective measures of complexity in DDM scenarios is too limited. The parameter values implemented in the five DDM scenarios of the present research project are assumed to be random samples in a larger pool of unrestricted domains, i.e., they represent random effects for the purposes of statistical inference tests such as analyses of variance and measures of association such as general linear models. All of the analyses performed through the exploration of the parameter space of the objective measures of complexity use variance components models, that is, they are random effects models. A larger set of DDM scenarios would help find the functional form of the relationship of the regression model using the objective measures of complexity as parameters to explain the variance in human performance, in a way to capture whether it is indeed a linear relationship or a nonlinear one. We anticipated some sort of a threshold value, representing a "wall of complexity", that should theoretically map the functional form between small increments in complexity for DDM and a rapidly declining performance, in the likeness of an exponential decay function. The next final section lays down the necessary means to explore the impact of the parameters of complexity on DDM performance in a more comprehensive way, in order to select the combination of parameters which would constitute the most accurate measure of complexity for dynamic decision-making scenarios.

Additional parameters of complexity and response variables

Although the objective measures of complexity for dynamic decision-making problems presented herein are seen as critical characteristics of human performance in the comprehension and control of complex systems, they are not an exhaustive list of DDM task features nor of the task environment factors which may affect performance. One particular aspect of complex decision-making which is known to affect performance is temporal pressure (Kerstholt, 1994, 1996, Gonzalez, 2004, Karakul & Qudrat-Ullah, 2008, Stouten & Größler, 2017). While the research project presented in this document did have an overall time limit for the completion of DDM scenarios, it was not particularly prohibitive nor impactful on the participants' performance. The effects of time pressure as an objective measure of complexity on human performance could be set up experimentally through the use of varying durations for information acquisition and/or decision evaluation and feedback phases, with a working hypothesis that shorter latencies would negatively affect not only performance but also stress and motivation.

Response variables could be nuanced too, as the focus on human performance in the research presented herein is but one facet of the impact of complex decision-making on the everyday lives of humans in sociotechnical systems. As mentioned in the introduction, Qudrat-Ullah (Qudrat-Ullah, Spector, & Davidsen, 2008, Qudrat-Ullah, 2014, 2015) suggests on the one hand that the concept of performance is a theoretical construct which can be measured in various ways, such as the minimization or maximization of certain quantities associated with variables, predictions or achievements of different system states, and the control of a complex system in accordance with a set of goals. The author therefore distinguished between complementary research targets for DDM, besides task performance (an operationalized concept aiming to explain how decision makers achieve a certain degree of success in controlling a dynamical system, understood as a DDM problem), such as task knowledge (the extent to which a decision maker exhibits an understanding of the dynamical system, i.e., the accuracy of his or her mental model of the DDM problem), and transfer learning (the notion of whether learning about, and performing in, DDM problems, dynamical systems, and systems thinking, actually generalizes this knowledge, competence, and performance to other DDM problems).

Simulation-based experimentation in interactive learning environments allow the measurement of multidimensional criteria relative to a decision-maker's involvement in a complex decision task, such as distinctions between structural knowledge and heuristics knowledge (Qudrat-Ullah, 2015). A poor understanding of task structures leads to a poor task performance. The structural knowledge, which may be elicited from participants through questionnaires, concerns the facts and rules underlying a complex system's model. The heuristics knowledge concerns the strategies and systems thinking skills of a participant, with regards to his or her comprehension of the causal relationships among the complex system's variables. The consistency of strategies employed by participants (through the measurement of fluctuations in their decision patterns, for example, Qudrat-Ullah, 2015) are hints about their metacognitive skills, along with other behavioral markers such as the time and frequency spent on information acquisition related to situation assessment and feedback on decision outcomes (Kleinmuntz, 1985, Qudrat-Ullah, 2015, Pronovost et al, 2015, Lafond et al, 2012, 2016). Qudrat-Ullah (2015) also suggests that time spent on information and decision-making is a measure of cognitive effort, which could be used as a response variable on its own to compare between DDM scenarios of various degrees of complexity. The use of transfer learning has, for its part, been investigated for its capacity for near and far transfer across similar tasks over variable periods of time as well as over similar tasks across novel and unfamiliar task environments or interfaces (Hayes & Broadbent, 1988, Bakken, Gould & Kim, 1994, Huber, 1995, Barnett & Ceci, 2002, Mayer, Dale, Fraccastoro, et al, 2011, Gegenfurtner, Veermans, & Vauras, 2013, Pronovost et al, 2015). While learning effects can be used to predict success in similar DDM tasks over periods of time, far transfer to unfamiliar tasks or increasingly complex scenarios has not been observed (Pronovost et al, 2015).

The literature on the relationships between objective measures of task performance and perceived, or subjective, measures of complexity does not suggest favorable applications for the latter in research on dynamic decision-making. Feldman (2000, 2003a, 2003b) observed the existence of a strong relationship between the subjective complexity of logical concepts and the complexity of Boolean (logical) operations, which could be summarized as a simple, universal law akin to findings in psychophysics, whereby the subjective difficulty of a concept is directly proportional to its Boolean complexity (Shepard, Hovland, & Jenkins, 1961, Neisser & Weene, 1962, Goodwin & Johnson-Laird, 2013). In Feldman's view, a participant's ability to learn a Boolean concept depends on its intrinsic complexity. The difficulty of learning a concept is associated with the number of trials required to correctly categorize a Boolean concept of arbitrary size (i.e., its Boolean complexity, given the most compact representation of a concept in the Boolean algebra, or propositional logic). This type of structural and information complexity measure is directly related to the minimum description length (MDL, similar to Kolmogorov complexity, 1965) found in algorithmic information theory. Feldman initially (2000) thought that this Boolean algebra using a dichotomic classification of binary representations could be extended to algebraic complexity as a subjective measure of difficulty (where the author clearly does not make a distinction between complexity and difficulty as used throughout this document) for non-binary (continuous) representations.

Yet Lafond, Lacouture, and Mineau (2007) observed a discrepancy between the logical complexity of Boolean structures and operators on the one hand, and the psychological complexity, i.e., the complexity involved in learning and understanding that logical complexity, on the other hand. Whereas Feldman claimed to be able to predict the subjective measure of difficulty through the proportion of correct recall based on the minimally complex expression of a Boolean concept, Lafond, Lacouture, and Mineau found that participants tend to use non-minimal rules of Boolean concept classification. The authors suggested that the perceived complexity of concept acquisition did not necessarily follow the learning model accounted for by Feldman's complexity minimization principle. Other studies on the relationship between subjective measures of complexity and the structural and informational complexity of algorithms suggest that there are strong relationships with the perceived complexity, where this perceived complexity monotonically increases as a funtion of increments in various models and measures of objective complexity (De Silva & Kodagoda, 2013, De Silva, Kodagoda, & Perera, 2012, De Silva, Weerawarna, Kuruppu, Ellepola, et al, 2013). We have observed that subjective measures of complexity, using self-reported ratings on a Likert scale after the first turn of play and after the end of the scenario, are indeed positively correlated with objective parameters of complexity for highly complex DDM scenarios in earlier studies (Pronovost et al, 2014, see also Özgün & Barlas, 2015).

What could be, then, a more adequate measure of human comprehension and control over dynamic decision-making problems? Lafond, Lacouture, and Mineau (2007) ask for a "psychologically relevant description" of complexity for human learning and categorization, but is there a coherent and principled, best practice to the evaluation of the impact of decision complexity on human cognition? The jury is still out even on the very definition of complex problem solving (CPS), as an ongoing argument between some authors polarizes two positions between those who would see CPS as a latent, high-level ability, or a metacognitive construct (Greiff & Martin, 2014) while others envision it as a multifaceted cognitive process (Funke, 2014). Schoppek and Fischer (2015) argue that those two positions rest on a having two very different methodological angles on CPS, namely a psychometric research interest vs. a process-oriented approach. Greiff, Stadler, Sonnleitner, et al (2015) prefer a fine-tuned approach to CPS using minimal/multiple complex systems (MCS), a competency which is seen as an overall latent predictor of scores in reasoning tests and school grades, while Funke, Fischer, and Holt (2017) regard the MCS approach as simple and unrealistic examples of CPS ability, which undermines its external and construct validity as a means to assess the rich, multifaceted metacognitive functions involved in dynamic decision-making. While distinctions and nuances are essential to this research paradigm with regards to conjectures about multiple competencies and cognitive processes on the one hand, and to the use of multiple response variables in experimental designs on the other hand, there is no doubt that objective measures of both the characteristics of a complex system and of task performance in DDM scenarios will remain a critical feature of experimentation in future research about complex decision-making.

A research program for objective measures of complexity

This final section details a research program for the objective measures of complexity of dynamic decision-making problems aiming to explore the feature space for parameters to be included in regression models of human performance. Considerations about the objective parameters to be preserved are discussed, and a detailed breakdown of the feature space for a full-fledged model of DDM complexity is presented.

Candidate parameters of complexity and models of human performance in dynamic decision-making

In chapters 2 and 3, we adopted a three-pronged selection method for the OMC metrics as regression parameters: a first, semantic (or methodological) selection based on a qualitative and quantitative assessment of the commonalities of the complexity measures; a second, statistical method of parameter selection was used to observe their individual relationship with the response variable of performance; and finally, a machine learning approach to parameter selection was used to narrow down which combinations of the parameters of complexity were best suited as multiple linear regression models. Early on, we were faced with the intrinsic and accidental collinearity in the parameter space representing the DDM scenarios used for the empirical investigation. Given the high intrinsic and accidental collinearity values for many of the parameters involved in our exploratory analysis, combined with the low number of data points in our five DDM scenarios, our experimental design could not cover an in-depth exploration of the feature space to model the impact of complexity on DDM performance. Recall from chapter 2 that the DDM scenarios exhibited some clusters across the parameter space, as seen in Figure 60 showing the principal component analysis for the independent variables.

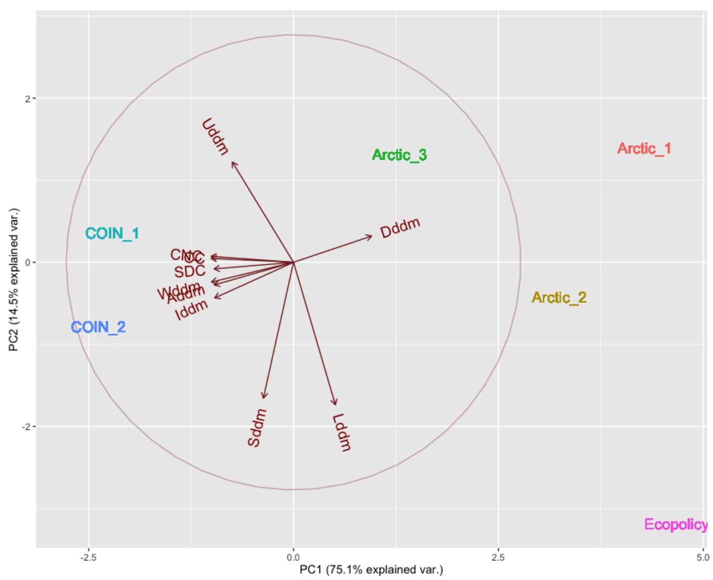

Figure 60 – A bivariate plot representing the principal component analysis loadings of the ten parameters of complexity for the DDM scenarios used in chapters 2 and 3.

A unified model for decision complexity could in principle be composed of all the parameters exhibiting less redundant characteristics of DDM complexity (accidental collinearity) and comparisons could be made between those models across different parameter values. This would require a multiple-year research design involving high numbers of DDM scenarios and participants, since the parameter values for DDM scenarios are random effect factors drawn out of an indefinite number for the scalars (the magnitudes expressing structural, information, and cognitive complexity), and from 0 to 1 for the measures expressing relationships (the ratio values of the parameters of difficulty, nonlinearity, uncertainty, and instability). A more suitable approach would be a stratified, non-probabilistic sampling method (or "purposive" sampling) to parameter values selection, whereby the parameter values would be sampled in high contrast zones, such as in a "low vs. high" design.

Feature space and dynamic decision-making models

A regression model of explanatory variables using our objective measures of complexity for human performance in dynamic decision-making problems can be understood as a vector in a feature space. Each feature (an objective parameter of complexity) can be thought of as a dimension of a dynamic decision-making problem represented in an Euclidean space. The idea of a feature space is a concept often used in the machine learning literature. A common challenge of machine learning is feature extraction and/or selection (Bishop, 2006), hence we view all the independent variables of our objective measures of complexity as features in a problem space. The present research project can thus be seen as an endeavor of *feature engineering*, or applied machine learning, whereby explanatory variables are individual measurable properties (features) of a problem (the feature space), and the multiple linear regression models are feature vectors combined with weights (predictor coefficients) aiming to predict a pattern or an outcome variable. An instance of this multidimensional feature space can then be considered as a "point" in an n-dimensional feature space, where this point corresponds to a particular dynamic decision-making scenario.

Chapter 2 presented six subsets of the objective parameters of complexity which sampled unique metrics for the structural complexity C_{DDM} (between the three variants of CC_{DDM}, CNC_{DDM}, and SDC_{DDM}) and the information complexity I_{DDM} (among two versions, I_{DDM} and A_{DDM}). Those subsets of parameters were used to conduct an exhaustive search of the parameter space using machine learning techniques for stepwise multiple linear regression modeling. The selection of the best candidate parameters was not motivated strictly by the statistical accuracy of the estimators in MLR models, it was also driven by semantic considerations. That is why the SDC_{DDM} metric is preferred to the other two variants of structural complexity, as the models involving all three variants actually produced comparable results. With the remaining five objective parameters of complexity that produced the best candidate models of regression analysis in chapters 2 and 3, namely SDC_{DDM} for structure, I_{DDM} for information, and the measures of difficulty, uncertainty, and instability, we can suggest a research program fulfilling the proposal in the previous section for a stratified, non-probabilistic sampling of the parameters in a "low vs. high" design. A matrix of the possible DDM scenarios in order to test the 3-valued parameter range over the five features would produce a set of $2^5 = 32$ possible combinations in the overall multidimensional feature space of the objective measures of complexity.

The suggested values are based on the results in the present research project, and are meant to exacerbate any potential difference in the impact of individual parameters on human performance in dynamic decision-making scenarios. Figures 82 to 86 in appendix B represent those parameter values for each objective measure of complexity, where the two values (low and high complexity ratings) for each of the parameters would be represented in the 32 prospective DDM scenarios. Table 28 below presents those ratings in a more concise format for each

OMC parameter, while a bivariate plot in Figure 61 suggests that the parameter values across those features of complexity are indeed remote enough from one another to avoid accidental collinearity, in the eventuality of each parameter combination being represented by a unique dynamic decision-making scenario. It is expected that a research program involving those scenario combinations would yield valuable insights on the relationship between the features of complexity of dynamic decision problems and the decline of human performance facing such complexity. A longer term research effort could potentially reveal the conjectured nonlinear nature of this relationship, where a threshold, or "wall of complexity" is found to limit human comprehension and influence on complex and dynamic systems.

Table 28 – Values for the objective measures of complexity in DDM scenarios for a future experimental design.

parameter	**low parameter value**	**high parameter value**
SDC_{DDM}	500	2000
I_{DDM}	50	350
D_{DDM}	.10	.50
U_{DDM}	.10	.30
S_{DDM}	.10	.50

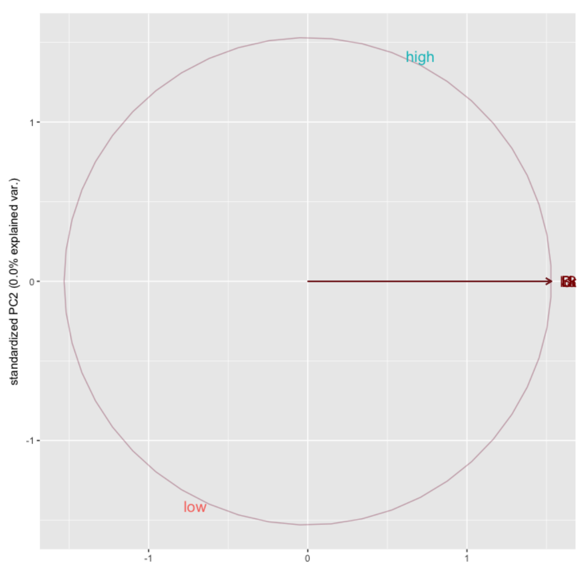

Figure 61 – A bivariate plot representing the principal component analysis loadings of the candidate parameters of complexity for the prospective DDM scenarios.

Epilogue

Complexity is a considerable frontier to human cognition. Complex problem solving, dynamic decision-making, and systems thinking are all fundamentally limited by, and informed through, the intricacies of information, structure, and difficulty in coping with a complex and dynamic system. Our understanding and influence of complex decision problems cannot be helped by heuristics commonly used in tackling complicated and laborious information processing tasks, such as through the use of means-ends analysis and hierarchical task decomposition (Newell, Shaw, & Simon, 1958, Fikes & Nilsson, 1971, Tate, 1976), heuristic search (Pólya, 1945, Newell & Simon, 1976, Pearl, 1984), and dynamic optimization (Bellman, 1957, Dijkstra, 1959, Howard, 1960). We contrast this type of *complicated* tasks with the *complex* problems defined in this dissertation, namely tasks involving the nonlinear evolution of a system's states over time, changing on its own due to intrinsic dynamics, and featuring uncertainty in states, relationships, as well as in feedback over decision outcomes.

There are many ways to define and characterize complexity (Kinsner, 2010). We have focused on structural and information complexity metrics from the onset of this research project, as such models as ubiquitous in the natural sciences, in engineering specifications, and in the more recent fields of the information sciences. Models and measures of complexity for dynamic decision-making problems are still in their infancy, as previous studies focused on qualitative assessments of the characteristics affecting human comprehension and control of stock and flow models, with some studies in cognitive psychology tweaking a few parameters in experimental designs using analyses of variance to compare the effects of different features of complexity.

This research project combined fundamental, experimental, and analytical aspects of an overall scientific inquiry into the identification and analysis of the features of what makes complex decision problems so hard to cope for human decision makers. Although there was no direct practical considerations in the pursuit of this knowledge, the research was nevertheless conducted with practical contributions in mind, such as the development of concepts, strategies, and tools to support human cognition in the face of the throes of complexity. Complex problem solving and decision-making should be supported by pedagogical material and means, such as notions of systems thinking and feedback from intelligent tutoring systems. They should also be facilitated by decision support systems, such as ergonomic knowledge management tools and interactive information visualizations. We say that they *should* support human cognition because we don't yet know for sure that they *can* do so in their current forms, although there are researchers who are optimistic in that endeavor (Qudrat-Ullah, 2015). The frontier of complexity has to be challenged by all means possible, as the information age has produced an unprecedented combinatorial explosion of information intractable for the human mind. Human cognition should not be thought of as a limitation to a sociotechnical system's proper function. With the help of properly engineered technological support, it can achieve incredibly complex tasks. The recent escalation in computing performance and the development of artificial intelligence tailored to support and enhance human cognition for everyday tasks in expert domains (Kelly, 2015) will likely push the boundaries of complexity a little further.

References

Aboutaleb, H., & Monsuez, B. (2015). Measuring the complexity of a higraph-based system model: formalism and metrics. *Procedia Computer Science, 44*, 11-20.

Ackermann, E. K. (1993). *Tools for constructive learning: Rethinking Interactivity.* Epistemology and Learning Group, MIT Media Laboratory.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics, 16*(1), 125-127.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician, 46* (3): 175–185.

Anderson, J. R. (2010). *Cognitive Psychology and Its Implications.* New York, NY: Worth Publishers.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?.* Oxford: Oxford University Press.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science, 26*(1), 85-112.

Anderson, J. R. (1990). *The adaptive character of thought.* Psychology Press.

Baddeley, A., & Bernses, O.A. (1989). *Cognitive Psychology: Research Directions In Cognitive Science: European Perspectives, vol 1.* East Sussex, UK: Lawrence Erlbaum Associates Ltd.

Bakken, B. E. (2008). On improving dynamic decision-making: implications from multiple-process cognitive theory. *Systems Research and Behavioral Science, 25*(4), 493-501.

Bakken, B. E. (1993). Learning and transfer of understanding in dynamic decision environments. *Unpublished doctoral dissertation.* MIT, Boston.

Bakken, B. E. (1989). Learning in dynamic simulation games; using performance as a measure. In Milling, P. E., & Zahn, E. O. K. (eds.), *Computer-Based Management of Complex Systems* (pp. 309-316). Springer Berlin Heidelberg.

Bakken, B., Gould, J. & Kim, D. (1994). Experimentation in learning organizations: A management flight simulator approach. In J. Morecroft & J. Sterman (Eds.) *Modeling for learning organizations* (pp. 243–266). Portland, Or.: Productivity Press.

Banks, J., & Carson, J. S. (1984). *Discrete-Event System Simulation.* Prentice Hall, Englewood Cliff, NJ.

Bar-Yam, Y. (2003). *Dynamics of complex systems.* Westview Press.

Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with computers, 23*(1), 4-17.

Bechtel, W. (1994). Levels of description and explanation in cognitive science. *Minds and Machines, 4* (1), 1-25.

Bedny, G. Z., Karwowski, W., & Bedny, I. S. (2012). Complexity evaluation of computer-based tasks. *International Journal of Human-Computer Interaction, 28*(4), 236-257.

Bellman, R. E. (1957). *Dynamic Programming.* Princeton University Press, Princeton, NJ.

Bernoulli, D. (Originally published in 1738, translated by L. Sommer, 1954). Exposition of a New Theory on the Measurement of Risk (Latin: Specimen theoriae novae de mensura sortis). *Econometrica, 22* (1): 22–36.

Bertsekas, D.P. (1976). *Dynamic Programming and Stochastic Control.* Academic Press, N.Y.

Bisantz, A. M., Kirlik, A., Gay, P., Phipps, D. A., Walker, N., & Fisk, A.D. (2000). Modeling and analysis of a dynamic judgment task using a Lens model approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 6*, 605–616.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford university press.

Boden, M. A., & Mellor, D. H. (1984). What is computational psychology? *Proceedings of the Aristotelian Society, Supplementary Volumes, 58*, 17-53.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).

Brehmer, B. (1992). Dynamic decision-making: Human control of complex systems. *Acta Psychologica, 81*(3), 211–241.

Brehmer, B. (1990). Strategies in real-time dynamic decision-making. In R. M. Hogarth (Ed.), *Insights in decision-making* (pp. 262–279). Chicago: University of Chicago Press.

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance 11*: 1–27.

Brehmer, B., & Allard, R. (1991). Dynamic decision-making: The effects of task complexity and feedback delay. In Rasmussen, J., Brehmer, B., & Leplat, J. (Eds.), *Distributed decision-making: Cognitive models of cooperative work*. Chichester: Wiley.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9*, 171–184.

Breiman, L., & Spector, P. (1992). Submodel selection and evaluation in regression. The X-random case. *International statistical review*, 291-319.

Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica. 47* (5): 1287–1294.

Brewer, M. B. (2000). Research design and issues of validity. In Reis, H. T., & Judd, C. M. (eds.), *Handbook of research methods in social and personality psychology.* Cambridge University Press.

Briand, L. C., Freimut, B., & Vollei, F. (2004). Using multiple adaptive regression splines to support decision-making in code inspections. *Journal of Systems and Software, 73*(2), 205-217.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence, 47* (1), 139-159.

Busemeyer, J. R., Wang, Z., Townsend, J. T., & Eidels, A. (2015). *The oxford handbook of computational and mathematical psychology.* Oxford University Press.

Bußwolder, P. (2015). The Effect of a Structured Method on Mental Model Accuracy and Performance in a Complex Task. *Systems, 3*(4), 264-286.

Buurman, J., & Babovic, V. (2016). Adaptation Pathways and Real Options Analysis: An approach to deep uncertainty in climate change adaptation policies. *Policy and Society, 35* (2), 137-150.

Cacciabue, P. C., & Hollnagel, E. (1995). Simulation of cognition: Applications. In Cacciabue, P. C., & Hollnagel, E. (eds.), *Expertise and technology: Cognition and human-computer cooperation.* Lawrence Erlbaum Associates.

Card, S. K., Moran, T. P. & Newell, A. (1983). *The psychology of human-computer interaction.*

Lawrence Erlbaum Associates, Hillsdale, N.J.

Cardoso, J., Mendling, J., Neumann, G., & Reijers, H. A. (2006). A discourse on complexity of process models. In *Business process management workshops, LNCS vol. 4103* (pp. 117-128). Springer Berlin Heidelberg.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press, New York, NY, USA.

Chevan, A., & Sutherland, M. (1991) Hierarchical Partitioning. *The American Statistician 45*, 90–96.

Church, A. (1936). A Note on the Entscheidungsproblem. *Journal of Symbolic Logic (1)*: 40–41.

Clements, D. (1989). *Computers in elementary mathematics education.* Englewood Cliffs, NJ: Prentice Hall.

Cleveland, W. S. (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician, 35* (1): 54.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association, 74* (368), pp. 829–836.

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83* (403), 596-610.

Cook, S. (1971). The complexity of theorem proving procedures. *Proceedings of the Third Annual ACM Symposium on Theory of Computing.* pp. 151–158.

Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1990). *Introduction to Algorithms (1st ed.).* MIT Press and McGraw-Hill.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Couture, M. (2007). *Complexity and chaos: State-of-the-art formulations and measures of complexity.* Defence R&D Canada-Valcartier, ON: Technical Note TN 2006-451.

Coveney, P., & Highfield, R. (1995). *Frontiers of Complexity the Search for Order in a Chaotic World.* Ballantine Books.

Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13* (1): 21–27.

Craik, K. J. W. (1943). *The Nature of Explanation.* Cambridge: Cambridge University Press.

Cronin, M., Gonzalez, C., & Sterman, J.D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes 101*, 116–130.

Croux, C., Dhaene, G., & Hoorelbeke, D. (2003). Robust standard errors for robust estimators. *Discussions Paper Series 03.16. K.U. Leuven, Center for Economic Studies.*

Cunningham, I. (2008). Are "skills" all there is to learning in organizations? The case for a broader framework. *Development and Learning in Organizations, 22*(3), 5-8.

Dawson, M. (1998). *Understanding Cognitive Science.* Blackwell Publishing.

DeRosa, J. K., Grisogono, A. M., Ryan, A. J., & Norman, D. O. (2008, April). A research agenda for the engineering of complex systems. In *Systems Conference, 2008 2nd Annual IEEE* (pp. 1-8). IEEE.

De Silva, D. I., & Kodagoda, N. (2013). Applicability of Weyuker's Properties Using Three Complexity Metrics. In *proceedings of the 8th International Conference on Computer Science & Education.*

De Silva, D. I., Kodagoda, N., & Perera, H. (2012). Applicability of three complexity metrics. In *Advances in ICT for Emerging Regions (ICTer)*, (pp. 82-88).

De Silva, D. I., Weerawarna, N., Kuruppu, K., Ellepola, N., & Kodagoda, N. (2013). Applicability of Three Cognitive Complexity Metrics. In *proceedings of the 8th International Conference on Computer Science & Education* (ICCSE 2013).

Diamantidis, N. A., Karlis, D., & Giakoumakis, E. A. (2000). Unsupervised stratification of cross-validation for accuracy estimation. *Artificial Intelligence, 116* (1), 1-16.

Diehl, E., & Sterman, J. D. (1993). Effects of feedback complexity on dynamic decision-making. *Organizational Behavior and Human Decision Processes, 62*(2), 198-215.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik, 23* (3): 269–271.

Dörner, D. (1989). *Die Logik des Mißlingens.* Reinbek bei Hamburg: Rowohlt.

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulations and Games, 11*, 8–106.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis.* Wiley & Sons.

Duch, W., Setiono, R., & Zurada, J. M. (2004). Computational intelligence methods for rule-based data understanding. *Proceedings of the IEEE, 92*(5), 771-805.

Edmonds, B. (1999). What is complexity? The philosophy of complexity *per se* with applications to some examples in evolution. In: Heylighen, F., Bollen, J., Riegler, A. (eds.) *The Evolution of Complexity*, p. 296. Springer, New York.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. *Human Factors 4*, 59–73.

Edwards, W. (1961). Behavioral decision theory. *Annual review of psychology, 12*(1), 473-498.

Edwards, W. (1954). The theory of decision-making. *Psychological bulletin, 51*(4), 380.

Ellsberg, D. (1961). "Risk, Ambiguity, and the Savage Axioms". *Quarterly Journal of Economics 75* (4): 643–669.

Evans, J. S. B., Newstead, S. E., & Byrne, R. M. (1993). *Human reasoning: The psychology of deduction*. Psychology Press.

Eysenck, M.W. (1990). *Cognitive Psychology: An International Review*. West Sussex, England: John Wiley & Sons, Ltd.

Fechner G.T. (1860). *Elemente der Psychophysik*. Leipzig, Breitskopf and Härtel.

Feldman, J. (2003a). A catalog of Boolean concepts. *Journal of Mathematical Psychology, 47(1)*, 75-89.

Feldman, J. (2003b). The simplicity principle in human concept learning. *Current Directions in Psychological Science, 12*(6), 227-232.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*(6804), 630-633.

Fikes, R. E., & Nilsson, N. J. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial intelligence, 2*(3-4), 189-208.

Fischer, H., & Gonzalez, C. (2015). Making Sense of Dynamic Systems: How Our Understanding of Stocks and Flows Depends on a Global Perspective. *Cognitive Science, 40*(2), 496-512.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving 4*, 19–42.

Fischer, A., Greiff, S., Wüstenberg, S., Fleischer, J., Buchwald, F., & Funke, J. (2015). Assessing analytic and interactive aspects of problem solving competency. *Learning and Individual Differences, 39*, 172-179.

Fischhoff, B. (2002). heuristics and biases in application. In Gilovich, T., Griffin, D., & Kahneman, D. (eds.), *Heuristics and biases*. New York, NY: Cambridge University Press.

Fitch, F. (1963). A Logical Analysis of Some Value Concepts. *Journal of Symbolic Logic, (28)* 2, pp. 135–142.

Fix, E., & Hodges, J. L. (1951) Discriminatory analysis, nonparametric discrimination: Consistency properties. *Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.*

Flavell, J.H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-development inquiry. *American Psychologist, 34* (10): 906–911.

Ford, D. N., & Sterman, J. D. (1998). Expert knowledge elicitation to improve formal and mental models. *System Dynamics Review, 14*(4), 309-340.

Forrester, J. W. (1989). The beginning of system dynamics. *International meeting of the System Dynamics Society*, Stuttgart, Germany.

Forrester, J. W. (1985). The" model versus a modeling "process. *System Dynamics Review, 1*(1), 133-134.

Forrester, J. W. (1968). *Principles of Systems.* Cambridge: The MIT Press.

Forrester, J. W. (1961). *Industrial dynamics.* Productivity Press: Cambridge, MA.

Frensch, P. A., & Funke, J. (1995). *Complex problem solving: The European perspective.* Psychology Press.

Friedman, J. (1991). Multivariate Adaptive Regression Splines. *The Annals of Statistics, 19*(1), 1–141.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning (vol. 1).* Springer, Berlin: Springer series in statistics.

Funke, J. (2014) Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Front. Psychol. 5*:739.

Funke, J. (2013). Problem solving: What are the important questions? *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 493-498). Austin, TX.

Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing, 11*(2), 133-142.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & reasoning, 7*(1), 69-89.

Funke, J. (1995). Experimental Research on Complex Problem Solving. In P. Frensch & J. Funke (Eds.) *Complex problem solving: The European perspective* (pp. 3–25). NJ: Lawrence Erlbaum Associates Publishers.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R.J. Sternberg & P.A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Funke, J., Fischer, A., & Holt, D. V. (2017). When Less Is Less: Solving Multiple Simple Problems Is Not Complex Problem Solving — A comment on Greiff et al. (2015). *Journal of Intelligence, 5*(1), 5.

Gegenfurtner, A., Veermans, K., & Vauras, M. (2013). Effects of computer support, collaboration, and time lag on performance self-efficacy and transfer of training: a longitudinal meta-analysis. *Educational Research and Reviews, 8*, 75–89.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association, 70*(350), 320-328.

Gentle, J. E. (2002). *Elements of Computational Statistics.* Springer.

Gentner, D., & Stevens, A. L. (1983). *Mental models.* Psychology Press.

Gibson, J.J. (1977). The Theory of Affordances. In R. Shaw & J. Bransford (eds.). *Perceiving, Acting, and Knowing: Toward an Ecological Psychology.* Hillsdale, NJ: Lawrence Erlbaum. pp. 67–82.

Gigerenzer, G., & Selten, R. (2002). *Bounded Rationality: The Adaptive Toolbox.* MIT Press.

Gilovich, T., Griffin, D., & Kahneman, D. (eds.). (2002). *Heuristics and biases.* New York, NY: Cambridge University Press.

Myers, C., Gluck, K., Gunzelmann, G., & Krusmark, M. (2010). Validating Computational Cognitive Process Models across Multiple Timescales. *Journal of Artificial General Intelligence, 2* (2) 108-127.

Goldfeld, Stephen M., & Quandt, R. E. (1965). Some Tests for Homoscedasticity. *Journal of the American Statistical Association. 60* (310): 539–547.

Gonzalez, C. (2012). Training decisions from experience with decision-making games. *Adaptive technologies for training and education*, 167-178.

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes, 96*, 142–154.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors, 46*(3), 449-460.

Gonzalez, C., & Dutt, V. (2011). A generic dynamic control task for behavioral research and education. *Computers in Human Behavior, 27*, 1904–1914.

Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in dynamic decision-making. *Cognitive Science, 27*, 591–635.

Gonzalez, C., & Quesada, J. (2003). Learning in dynamic decision-making: the recognition process. *Computational and Mathematical Organization Theory 9*, 287–304.

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision-making. *Intelligence, 33*(2), 169-186.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision-making. *Computers in Human Behavior, 21*(2), 273–286.

Goodwin, G. P., & Johnson-Laird, P. N. (2013). The acquisition of Boolean concepts. *Trends in cognitive sciences, 17*(3), 128-133.

Gray, W. D. (2012). *The Cognitive Science of Natural Interaction or Why Suboptimal Performance Results when Human Systems Violate the Design Assumptions of Nature*. HFES Webinar. Retrieved from http://homepages.rpi.edu/~grayw/pubs/papers/2012/120924-HFES-webinar.pdf.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied. *Cognitive Science Quarterly, 2*, 205–227.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 449–510.

Greenley, A., Pronovost, S., Race, P., Kelleher, D., Graham, J., & Chawla, B. (2009). *Identification of Decision Support Concepts for the Planning of Air Force Immediate Contingencies Operations*. Research report conducted with CAE Professional Services for DRDC Valcartier.

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence, 41*(5), 579-596.

Greiff, S., & Martin, R. (2014). What you see is what you (don't) get: a comment on Funke's (2014) opinion paper. *Frontiers in psychology, 5*, 1120.

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: an analysis of behavioral data from computer-generated log files. *Computers in Human Behavior, 61*, 36-46.

Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C., & Martin, R. (2015). Sometimes less is more: Comparing the validity of complex problem solving measures. *Intelligence, 50*, 100-113.

Größler, A., Rouwette, E., & Vennix, J. (2016). Non-Conscious vs. Deliberate Dynamic Decision-Making — A Pilot Experiment. *Systems, 4*(1), 13.

Gruszka, A., Matthews, G., & Szymura, B. (2010). *Handbook of Individual Differences in Cognition*. New York: Springer.

Hackman, J. R. (1969). Toward understanding the role of tasks in behavioral research. *Acta Psychologica, 31*(0): 97-128.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., & Lagnado, D. (2010). Spontaneous causal learning while controlling a dynamic system. *The Open Psychology Journal, 3*, 145-162.

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis Seventh Edition*. Prentice Hall.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in cognitive sciences, 11*(6), 236-242.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in cognitive sciences, 14*(11), 497-505.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*(06), 803-831.

Halstead, M. H. (1977). *Elements of Software Science (Operating and programming systems series)*. Elsevier Science Inc.

Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F., & Kaber, D. B. (2013). Human-automation interaction research past, present, and future. *Ergonomics in Design, 21* (2), 9-14.

Handel, O. (2016). Modeling Dynamic Decision-Making of Virtual Humans. *Systems, 4*(1), 4.

Harary, F. (1969). *Graph Theory*. Addison–Wesley, Reading, MA.

Harteveld, C., & Drachen, A. (2015). Gaming on Environmental Issues. In Ruth, M. (ed.): *Handbook of Research Methods and Applications in Environmental Studies*. Edward Elgar Publishers.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world*. Thousand Oaks, CA: Sage Publications, Inc.

Hastie, T., Tibshirani, R., & Friedman, J.H. (2009). *The Elements of Statistical Learning,* *2nd edition.* Springer.

Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition, 28*, 249–276.

Henrion, M., Breese, J. S., & Horvitz, E. J. (1991). Decision analysis and expert systems. *Artificial Intelligence Magazine, 12*(4): 64–91.

Henry, S., & Kafura, D. (1981). Software structure metrics based on information flow. *Software Engineering, IEEE Transactions on, (5)*, 510-518.

Heylighen, F., Bollen, J., & Riegler, A. (1999). *The evolution of complexity.* Springer.

Hilbert, D., & Ackermann, W. (1928). *Grundzüge der theoretischen Logik (Principles of Mathematical Logic).* Springer-Verlag.

Hinton, G., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In Rumelhart, D. E., & McClelland, J. L. (eds.) *Parallel Distributed Representations: Explorations in the Microstructure of Cognition, Vol. 1.* Cambridge, MA, MIT Press, pp. 77–109.

Holland, J. (1994). *Adaptation in Natural and Artificial Systems.* MIT Press, Cambridge.

Hollnagel, E. (2001). Extended cognition and the future of ergonomics. *Theoretical Issues in Ergonomics Science, 2*, 309-315.

Hollnagel, E. & Woods, D. D. (2005). *Joint cognitive systems : Foundations of cognitive systems engineering.* Boca Raton, FL: CRC Press.

Hollnagel, E. & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18*, 583-600.

Hotaling, J., Fakhari, P., & Busemeyer, J. R. (2015). Dynamic Decision Making. In Wright, J. D., (ed.), *International Encyclopedia of the Social & Behavioral Sciences, 2nd Edition.* Elsevier. pp.709-714.

Howard, A. H. (1960). *Dynamic Programming and Markov Processes.* The M.I.T. Press.

Howard, R. A. (1988). Uncertainty about probability: A decision analysis perspective. *Risk Analysis 8*(1): 91–98.

Hsiao, N., & Richardson, G. P. (1999). In Search of Theories of Dynamic Decision Making: A Literature Review. In *Systems Thinking for the Next Millennium: Proceedings of the 17th International Conference of the System Dynamics Society.*

Huber, O. (1995). Complex problem solving as multistage decision-making. In P. Frensch & J. Funke (Eds.) *Complex problem solving: The European perspective* (pp. 151–173). NJ:

Lawrence Erlbaum Associates Publishers.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics, 1* (5), 799-821.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics, 35*(1), 73-101.

Hutton, R. J., & Klein, G. (1999). Expert decision-making. *Systems Engineering, 2*(1), 32-45.

Issacs, W. & Senge, P. (1994). Overcoming limits to learning in computer-based learning environments. In J. Morecroft and J. Sterman (Eds.) *Modeling for learning organizations* (pp. 267–287). Portland, Or.: Productivity Press.

Jagacinski, R. J., & Hah, S. (1988). Progression-regression effects in tracking repeated patterns. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 77–88.

Jagacinski, R. J., & Miller, R. A. (1978). Describing the human operator's internal model of a dynamic system. *Human Factors, 20*, 425–433.

James, W. (1893). *The principles of psychology.* New York: H. Holt and Company.

Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press.

Kahneman, D., Slovic, P. & Tversky, A. (eds.) (1982). *Judgment under Uncertainty: Heuristics and Biases.* Cambridge, UK: Cambridge University Press.

Kaimann, R. A. (1974). Coefficient of network complexity. *Management Science, 21*(2), 172-177.

Karakul, M., & Qudrat-Ullah, H. (2008). How to improve dynamic decision-making? Practice and promise. In Qudrat-Ullah, H., Spector, J. M., & Davidsen, P. I. (eds.) *Complex Decision Making: Theory and Practice.* Springer Berlin Heidelberg. pp. 3-24.

Karatzoglou, A., Smola, A., & Hornik, K. (2016). '`kernlab`': *Kernel-Based Machine Learning Lab.* `R` package. Version 0.9-25. https://cran.r-project.org/web/packages/kernlab/index.html.

Keil, F. C., & Wilson, R. A. (2000). *Explanation and cognition.* MIT press.

Kelly III, J. E. (2015). *Computing, cognition and the future of knowing.* IBM Research: Cognitive Computing. IBM Corporation.

Kerstholt, J. H. (1996). The effects of information costs on strategy selection in dynamic tasks. *Acta Psychologica 94*, 273–290.

Kerstholt, J. H. (1994). The effect of time pressure on decision-making behavior in a dynamic task environment. *Acta Psychologica, 86*, 89–104.

Khamis, H., & Kepler, M. (2010). Sample size in multiple regression: 20 + 5k. *Journal of Applied Statistical Science, 17*, 505–517

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction, 4*(12), 391-438.

Kieras, D., & Polson, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies, 22*(4): 365-394.

Kinsner, W. (2010). System complexity and its measures: How complex is complex. In *Advances in Cognitive Informatics and Cognitive Computing*(pp. 265-295). Springer Berlin Heidelberg.

Kirlik, A., Miller, R. A., & Jagacinski, R.J. (1993). Supervisory control in a dynamic and uncertain environment: process model of skilled human-environment interaction. *IEEE Transactions on Systems, Man, and Cybernetics 23*, 929–952.

Klein, G. (2008). Naturalistic decision-making. *Human Factors, 50* (3), 456-460.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *Intelligent Systems, IEEE, 18* (3), 81-85.

Klein, D. E., Klein, H. A., & Klein, G. (2000). Macrocognition: Linking Cognitive Psychology and Cognitive Ergonomics. *Proceedings of the 5th International Conference on Human Interactions with Complex Systems*, Univ. of Illinois at Urbana-Champaign.

Kleinmuntz, D. (1993). Information processing and misperceptions of the implications of feedback in dynamic decision-making. *System Dynamics Review 9*, 223–237.

Kleinmuntz, D. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science 31*, 680–702.

Kleinmuntz, D., & Thomas, J. (1987). The value of action and inference in dynamic decision-making. *Organizational Behavior and Human Decision Processes 39*, 341–364.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th international joint conference on Artificial intelligence vol. 2* (pp. 1137-1143).

Koller, M., & Stahel, W.A. (2011). Sharpening Wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis, 55* (8), 2504–2515.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission, 1* (1), 1–7.

König, D. (1936). *Theorie der endlichen und unendlichen Graphen.* Leipzig: Akademische Verlagsgesellschaft.

Kopainsky, B., & Alessi, S. M. (2015). Effects of structural transparency in system dynamics simulators on performance and understanding. *Systems, 3*(4), 152-176.

Kopainsky, B., & Sawicka, A. (2011). Simulator-supported descriptions of complex dynamic problems: experimental results on task performance and system understanding. *System Dynamics Review 27*, 142–172.

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of cheminformatics, 6* (1), 1.

Kuhn, M. (2016). '`caret`': *Classification and Regression Training.* `R` package. Version 6.0-72. https://cran.r-project.org/web/packages/caret/index.html.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling.* New York: Springer.

Kushwaha, D.S., and Misra, A.K. (2006). A moditied cognitive information complexity measure of software. *ACM SIGSOFT Software Engineering Notes, 31* (1), 1-4.

Kvålseth, T. O. (1985). Cautionary note about R 2. *The American Statistician, 39*(4), 279-285.

Lafond, D., DuCharme, M. B., Gagnon, J. F., & Tremblay, S. (2012). Support requirements for cognitive readiness in complex operations. *Journal of Cognitive Engineering and Decision Making, 6*(4), 393-426.

Lafond, D., Gagnon, J.-F., Pronovost, S., DuCharme, M. B., & Tremblay, S. (2016). Behavioral Test for Prediction of Individual Differences in Dynamic Decision Making Ability. *Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics.*

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology, 51*(2), 57-74.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial intelligence, 33*(1), 1-64.

Latva-Koivisto, A. M. (2001). *Finding a complexity measure for business process models.* Helsinki University of Technology, Systems Analysis Laboratory.

Lebiere, C., & Best, B. J. (2009). From Microcognition to Macrocognition: Architectural Support for Adversarial Behavior. *Journal of Cognitive Engineering and Decision Making, 3*, 176-193.

Lindeman, R.H., Merenda, P.F., & Gold, R.Z. (1980) *Introduction to Bivariate and Multivariate Analysis.* Glenview IL: Scott, Foresman.

Lintern, G. (2005). Integration of cognitive requirements into system design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49* (3), pp. 239-243).

Liu P, & Li, Z. (2014). Comparison of task complexity measures for emergency operating procedures: convergent validity and predictive validity. *Reliability Engineering & System Safety 121*: 289–293.

Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics, 42*(6): 553-568.

Liu P, & Li, Z. (2011). Toward understanding the relationship between task complexity and task performance. In Rau, P. L. P. (ed.), *Internationalization, Design and Global Development.* Springer: Berlin Heidelberg; 192–200.

Lloyd, S. (2008). *Measures of complexity: A non-exhaustive list.* http://web.mit.edu/esd.83/www/notebook/Complexity.pdf

Mackinnon, A. J., & Wearing, A. J. (1985). Systems analysis and dynamic decision-making. *Acta Psychologica 58*, 159–172.

Mackinnon, A. J., & Wearing, A. J. (1980). Complexity and decision-making. *Behavioral Science, 25*(4), 285–292.

Mandler, G. (2002). Origins of the cognitive (r)evolution. *Journal of the History of the Behavioral Sciences, 38*, 339–353.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods. Wiley Series in Probability and Statistics.* Wiley & Sons.

Maronna, R. A., & Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference, 89*, 197–214.

Marr, D. (1982). *Vision.* Freeman: New York.

Marshall, T. E., & Byrd, T. A. (1998). Perceived task complexity as a criterion for information support. *Information & management, 34*(5), 251-263.

Mayer, W., Dale, K., Fraccastoro, K., & Moss, G. (2011). Improving transfer of learning: relationship to methods of using business simulation. *Simulation & Gaming, 42*(1), 64–84.

McCabe, T. J. (1976). A complexity measure. *Software Engineering, IEEE Transactions on, (4)*, 308-320.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: knowing about knowing.* Cambridge, MA: MIT Press.

Milborrow, S. (2016). '`earth`': *Multivariate Adaptive Regression Splines.* `R` package. Version 4.4.7. https://cran.r-project.org/web/packages/earth/index.html.

Militello, L. G., Dominguez, C. O., Lintern, G., & Klein, G. (2010). The role of cognitive systems engineering in the systems engineering design process. *Systems Engineering, 13*(3), 261-273.

Misra, S. (2006). Complexity measure based on cognitive weights. *International Journal of Theoretical and Applied Computer Sciences, 1* (1), 1-10.

Moore, M. L., & Dutton, P. (1978). Training needs analysis: Review and critique. *Academy of Management Review, 3*(3), 532-545.

Myers, C., Gluck, K., Gunzelmann, G., & Krusmark, M. (2010). Validating computational cognitive process models across multiple timescales. *Journal of Artificial General Intelligence, 2*(2), 108-127.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* New York: Freeman.

Neisser, U. (1967). *Cognitive Psychology.* Englewood Cliffs, NJ: Prentice Hall.

Neisser, U. & Weene, P. (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology, 64*, 640-645.

Newell, A. (1990). *Unified Theories of Cognition.* Harvard University Press. Cambridge, Massachusetts.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review, 65*(3), 151.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM, 19*(3), 113-126.

Nicholson, B., & O'Hare, D. (2014). The effects of individual differences, prior experience and cognitive load on the transfer of dynamic decision-making performance. *Ergonomics, 57*(9), 1353-1365.

Nordlie, E., & Plesser, H. E. (2010). Visualizing neuronal network connectivity with connectivity pattern tables. *Frontiers in neuroinformatics, 3*, 39.

Norman, D. A. (1988). *The psychology of everyday things.* Basic books.

Northrop, R. B. (2010). *Introduction to complexity and complex systems.* CRC press.

Omodei, M. M., & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision- making behavior. *Behavior Research methods, Instruments, & Computers, 27*, 303–316.

Osman, M., Glass, B. D., & Hola, Z. (2015). Approaches to learning to control dynamic uncertainty. *Systems, 3*(4), 211-236.

Osman, M., & Speekenbrink, M. (2011). Cue utilization and strategy application in stable and unstable dynamic environments. *Cognitive Systems Research, 12*(3), 355-364.

Özgün, O., & Barlas, Y. (2015). Effects of systemic complexity factors on task difficulty in a stock management game. *System Dynamics Review, 31*(3), 115-146.

Paich, M., & Sterman, J. D. (1993). Boom, bust, and failures to learn in experimental markets. *Management Science, 39* (12), 1439–1458.

Papert, S. (1980). Computer-based microworlds as incubators for powerful ideas. In Taylor, R. (ed.), *The computer in the school: Tutor, tool, tutee* (pp. 203–210). New York: Teacher's College Press.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation : Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A : Systems and Humans, IEEE Transactions on, 30* (3), 286-297.

Park, J. (2014). Investigating the TACOM measure as a general tool for quantifying the complexity of procedure guided tasks. *Reliability Engineering & System Safety, 129*, 66-75.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker.* Cambridge University Press, NY.

Pearl, J. (1984). *Heuristics: intelligent search strategies for computer problem solving.* Addison-Wesley Pub. Co., Inc. Reading, MA.

Piaget, J. (1957). *Construction of reality in the child.* London: Routledge & Kegan Paul.

Pólya, G. (1945). *How to Solve It.* Princeton University Press. Princeton, New Jersey.

Pronovost, S. (2012). Using Micro-Architectures of Cognition to Model Macro-Cognitive Systems: A Warfighting Video Game Example. *Proceedings of the 21st Behavior Representation in Modeling & Simulation (BRIMS) Conference*.

Pronovost, S. (2010). Improving Usability and Integration of Human Behavior Representation Engineering across Cognitive Modeling, Human Factors, and Modeling and Simulation Best Practices. *Proceedings of the 19th Behavior Representation in Modeling & Simulation (BRIMS) Conference*.

Pronovost, S. (2009). Bridging Cognitive Modeling and Human Behavior Representation Engineering. *Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*.

Pronovost. S, Gagnon, J.-F., Lafond, D., & Tremblay, S. (2014). Objective measures of complexity for dynamic decision-making in an interactive learning environment. *Proceedings of the 1st Asia-Pacific System Dynamics Conference of the System Dynamics Society*, Senshu University, Tokyo.

Pronovost, S., St-Louis, M.-E., Lafond, D., Gagnon, J.-F., DuCharme, M. B., & Tremblay, S. (2015). Implicit learning in dynamic decision-making: A glass-box approach. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Pasadena, California.

Pronovost, S., & West, R. L. (2008). Bridging cognitive modeling and model-based evaluation: extending GOMS to model virtual sociotechnical systems and strategic activities. *Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society*.

Pronovost, S., & West, R. L. (2007). A GOMS Model of Virtual Sociotechnical Systems: Using Video Games to Build Cognitive Models. *Proceedings of the European Conference on Cognitive Ergonomics*.

Qudrat-Ullah, H. (2015). *Better Decision Making in Complex, Dynamic Tasks: Training with Human-Facilitated Interactive Learning Environments*. Springer International Publishing Switzerland.

Qudrat-Ullah, H. (2014). Yes we can: improving performance in dynamic tasks. *Decision Support Systems, 61*, 23-33.

Qudrat-Ullah, H., Spector, J. M., & Davidsen, P. I. (2008). *Complex Decision Making: Theory and Practice*. Springer Berlin Heidelberg.

`R` Core Team and contributors worldwide. (2016). '`stats`': *The R Stats Package*. `R` package. Version 3.4.0. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html.

Rao, C. R. (1965). *Linear statistical inference and its applications.* Wiley & Sons.

Rapoport, A. (1975). Research paradigms for studying dynamic decision behavior. In Wendt, D., & Vlek, C. (eds.), *Utility, Probability, and Human Decision Making.* Reidel, Dordrecht-Holland, pp. 347–369.

Rapoport, A. (1967). Dynamic programming models for multistage decision-making. *Journal of Mathematical Psychology 4*, 48–71.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive Systems Engineering.* John Wiley & Sons.

Renaud, O., & Victoria-Feser, M.-P. (2010). A robust coefficient of determination for regression, *Journal of Statistical Planning and Inference, 140*, 1852-1862.

Rieber, L. P. (1996) Microworlds, in Jonassen, D. H. (ed.) *Handbook of research on educational communications and technology. Handbook of Research for Educational Communications and Technology, 2nd edition.* Taylor & Francis.

Righi, A. W., & Saurin, T. A. (2015). Complex socio-technical systems: Characterization and management guidelines. *Applied ergonomics, 50*, 19-30.

Ripley, B. D. (2016). '`nnet`': *Feed-Forward Neural Networks and Multinomial Log-Linear Models.* `R` package. Version 7.3-12. https://cran.r-project.org/web/packages/nnet/index.html.

Ripley, B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press, New York, NY, USA.

Rouse, W.B. (1980). *Systems Engineering Models of Human–Machine Interaction.* North-Holland, NY.

Rouse, W. B., & Rouse, S. H. (1979). Measures of complexity of fault diagnosis tasks. *IEEE Transactions on Systems, Man, and Cybernetics 9*(11): 720–727.

Rousseeuw, P. J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association, 79*, pp. 871–880.

Rousseeuw, P. J., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., & Maechler , M. (2016). '`robustbase`': *Basic Robust Statistics.* `R` package. Version 0.92-6. https://cran.r-project.org/web/packages/robustbase/index.html.

Rousseeuw, P. J.. & Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley.

Rousseeuw, P., & Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis, vol. 26 of the series Lecture Notes in Statistics*, pp 256-272. Springer, New York.

Rouwette, E. A., Größler, A., & Vennix, J. A. (2004). Exploring influencing factors on rationality: a literature review of dynamic decision-making studies in system dynamics. *Systems Research and Behavioral Science, 21*(4), 351-370.

Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision-making in a simulated command and control task. *Human Factors, 49*, 76–87.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323* (6088): 533–536.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge: MIT Press.

Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach, 3rd Edition.* Pearson.

Salado, A., & Nilchiani, R. (2015). A Research on Measuring and Reducing Problem Complexity to Increase System Affordability: From Theory to Practice. *Procedia Computer Science, 44*, 21-30.

Salas, E., Burke, C. S., & Cannon-Bowers, J. A. (2002). What we know about designing and delivering team training: tips and guidelines. In Kraiger, K., *Creating, Implementing, and Managing Effective Training and Development: State-of-the-Art Lessons for Practice.*

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team? Emerging training strategies. In Zsambok, C. E., & Klein, G. (eds.), *Naturalistic Decision Making.* Lawrence Erlbaum Associates, Mahwah, NJ, pp 359-370.

Salibian,, M. & Yohai, V.J. (2006). A Fast Algorithm for S-Regression Estimates, *Journal of Computational and Graphical Statistics, 15* (2), 414-427.

Savage, L. J. (1954). *The foundations of statistics.* Courier Corporation.

Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine Learning, 13* (1), 135-143.

Schaffernicht, M., & Groesser, S. N. (2009). What's in a mental model of a dynamic system? Conceptual structure and model comparison. In *Actas del 27th International Conference of the System Dynamics Society*, Albuquerque.

Schoelles, M. J., Neth, H., Myers, C. W., & Gray, W. D. (2006). Steps toward integrated models of cognitive systems: A level-of-analysis approach to comparing human performance to model predictions in a complex task environment. *Paper presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, BC, Canada.*

Schoppek, W., & Fischer, A. (2015). Complex problem solving - single ability or complex phenomenon?. *Frontiers in psychology, 6*: 1669-1669.

Schöttl, F., & Lindemann, U. (2015). Quantifying the complexity of socio-technical systems – a generic, interdisciplinary approach. *Procedia Computer Science, 44*, 1-10.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science, 26*: 113–125.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association, 63* (324), pp. 1379–1389.

Senge, Peter (1990). *The Fifth Discpline.* Doubleday.

Shafir, E. (1999). Decision-making. In Wilson, R. A., & Keil, F. C. (Eds.), *The MIT Encyclopedia of the Cognitive Sciences.* Cambridge, MA: MIT Press.

Shafir, E., & Tversky, A. (1995). *Decision-making.* The MIT Press.

Shao, J., & Wang, Y. (2003). A new measure of software complexity based on cognitive weights. *Electrical and Computer Engineering, Canadian Journal of, 28*(2), 69-74.

Shepard, R., Hovland, C. L. & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl. 75*, 1-42.

Simon, H. A. (1956). "Rational Choice and the Structure of the Environment". *Psychological Review 63* (2): 129–138.

Simon, H. A. (1957). "A Behavioral Model of Rational Choice", in (eds.) *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting.* Wiley.

Sipser, M. (1996), *Introduction to the Theory of Computation.* PWS Publishing Co.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Peformance 6*, 649–744.

Soyer, E., & Hogarth, R. M. (2015). Learning from experience in nonlinear environments: Evidence from a competition scenario. *Cognitive psychology, 81*, 48-73.

Spector, J. M. (2000). System dynamics and interactive learning environments: Lessons learned and implications for the future. *Simulation & Gaming, 31*(4), 528–535.

Sterman, J. (2014). Interactive web-based simulations for strategy and sustainability: The MIT Sloan LearningEdge management flight simulators, Part I. *System Dynamics Review, 30*(1-2), 89-121.

Sterman, J.D. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review 26*, 316–334.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world.* McGraw-Hill.

Sterman, J. D. (1994). Learning in and around complex systems. *System Dynamics Review, 10* (2–3), 291–323.

Sterman, J. D. (1989). Misperceptions of feedback in dynamic decision-making. *Organizational Behavior and Human Decision Processes, 43* (3), 301–335.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64* (3): 153–181.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 111-147.

Stouten, H., & Größler, A. (2017). Task Complexity in Individual Stock Control Tasks for Laboratory Experiments on Human Understanding of Dynamic Systems. *Systems Research and Behavioral Science, 34* (1), 62-77.

Sun, R. (2008). *The Cambridge handbook of computational psychology.* Cambridge University Press.

Susanti, Y., & Pratiwi, H. (2014). M Estimation, S Estimation, And MM Estimation In Robust Regression. *International Journal of Pure and Applied Mathematics, 91*(3), pp. 349-360.

Sylvester, J. J. (1878). Chemistry and Algebra. *Nature 17*: 284.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics, 6th edition.* Pearson.

Tate, A. (1976). *Project Planning Using a Hierarchic Non-linear Planner.* D.A.I. Research Report No. 25, Department of Artificial Intelligence, University of Edinburgh.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. I, II, III. *Nederl. Akad. Wetensch. Proc. Ser. 53*, pp. 386-392 (Part I), 521-525 (Part II), 1397-1412 (Part III).

Tremblay, S., Gagnon, J. F., Lafond, D., Hodgetts, H. M., Doiron, M., & Jeuniaux, P. P. (2017). A cognitive prosthesis for complex decision-making. *Applied Ergonomics, 58*, 349-360.

Trist, E.L., & K.W. Bamforth (1951). Some Social and Psychological Consequences of the Long-Wall Method of Coal-Getting. *Human Relations, 4*, 3-38.

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 2* 42: 230–65.

Tversky, A., & Kahneman, D. (1974), "Judgments Under Uncertainty: Heuristics and Biases", *Science 185* (4157): 1124–1131.

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological science, 3*(6), 358-361.

Tyler, L. E. (1947). *The psychology of human differences.* New York: Appleton-Century-Crofts/Prentice-Hall.

Vapnik, V. N. (1998). *Statistical learning theory.* New York: Wiley.

Vespignani, A. (2012). Modelling dynamical processes in complex socio-technical systems. *Nature Physics, 8*(1), 32-39.

Vester, F. (1999). *The Art of interconnected thinking: Tools and concepts for a new approach to tackling complexity.* Munich, Germany: MCB Verlag.

Vester, F. (1985). *Unsere Welt, ein vernetztes System.* Deutscher Taschenbuch Verlag.

Vester, F. (1983). *Urban Systems in Crisis.* Munich, Germany: MCB Verlag.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* London: Lawrence Erlbaum Associates.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton university press.

Wang, Y. (2009). On the cognitive complexity of software and its quantification and formal measurement. *International Journal of Software Science and Computational Intelligence (IJSSCI), 1*(2), 31-53.

Wang, Y. (2007). The theoretical framework of cognitive informatics. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI), 1*(1), 1-27.

Wang, Y., Kinsner, W., Anderson, J. A., Zhang, D., Yao, Y., Sheu, P., ... & Chan, C. (2009). A doctrine of Cognitive Informatics (CI). *Fundamenta Informaticae, 90(3)*, 203-228.

Weber, E. H. (1851). *Die lehre vom tastsinne und gemeingefühle auf versuche gegründet.* Friedrich Vieweg und Sohn.

Weinberger, K. Q., & Saul, L. K. (2009). Distance Metric Learning for Large Margin Classification. *Journal of Machine Learning Research, 10*: 207–244.

West, R. L., & Pronovost, S. (2009). Modeling SGOMS in ACT-R: Linking Macro- and Microcognition. *Journal of Cognitive Engineering and Decision Making, 3* (2): 194-207.

White, B. Y. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction, 10*(1), 1–100.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). Engineering psychology and human performance. Psychology Press.

Willett, J. B., & Singer, J. D. (1988). Another cautionary note about R 2: Its use in weighted least-squares regression analysis. *The American Statistician, 42*(3), 236-238.

Williamson, T. (2000). *Knowledge and Its Limits.* Oxford: Oxford University Press.

Wilson, B. (1995). Metaphors for instruction: Why we talk about learning environments. *Educational Technology, 35*(5), 25-30.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37*(1): 60-82.

Wright, A., & Sittig, D. (2008). A framework and model for evaluating clinical decision support architectures. *Journal of Biomedical Informatics, 41*: 982–990.

Wu, P. P. Y., Fookes, C., Pitchforth, J., & Mengersen, K. (2015). A framework for model integration and holistic modelling of socio-technical systems. *Decision Support Systems, 71*, 14-27.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics, 15*, 642-656.

Zhang, H., & Singer, B. H. (2010). *Recursive Partitioning and Applications, 2nd edition.* Springer.

Zionts, S., & Wallenius, J. (1976). An Interactive Programming Method for Solving the Multiple Criteria Problem. *Management Science, 22* (6): 652–663.

Zobarich, R., Miller, J., Kramer, C., Pronovost, S., & Kelsey, S. (2011). *Dynamic Synthetic Environment & Human Behavior Representation Analysis (Milestone 1).* Research report conducted with CAE Professional Services and Presagis for CAE Inc.

Zobarich, R., Pronovost, S., Torunski, E., & Unrau, D. (2011a). *Dynamic Synthetic Environment & Human Behavior Representation Analysis (Milestone 2). COMF (Command and Organizational Modeling Framework) and HBR Integration.* Research report conducted with CAE Professional Services, Presagis, and CAE Australia for CAE Canada Inc.

Zobarich, R., Pronovost, S., Torunski, E., & Unrau, D. (2011b). *Dynamic Synthetic Environment & Human Behavior Representation Analysis (Milestone 3). Technology Investigation of Modeling Tools for Enhanced AI and Behavior Modeling Capabilities.* Research report conducted with CAE Professional Services, Presagis, and CAE Australia for CAE Inc.

Appendix A

Dynamic Decision-Making Scenarios

A.1 Arctic 1

A.1.1 Variables

Table 29 – Variables in Arctic 1.

variable	details	initial value	winning value	failure value
sustainability objective	no color (mediating)	0		
crime rate	uncertainty $\pm$ 2	7	<= 9	
environment		12	>= 14	
Arctic infrastructure		9	>= 10	
resource development		4	>= 13	
socio-economic welfare		6	>= 15	0

A.1.2 Relations between variables

Table 30 – Relations between variables in Arctic 1.

source	values	target	values	details
crime rate	0 to 20	socio-economic welfare	0 to -3	
		resources development	0 to -3	
environment	0 to 20	socio-economic welfare	-3 to 1	
Arctic infrastructure	0 to 20	socio-economic welfare	-1 to 2	
		resources development	0 to 2	
resource development	0 to 20	socio-economic welfare	0 to 2	
		environment	0 to -4	conditional env. policies <= 6
		environment	0 to -2	conditional env. policies > 6 and <= 13
		environment	0 to 3	conditional env. policies > 13
socio-economic welfare	0 to 20	crime rate	3 to 2	

A.1.3 Interventions and influences from interventions

Table 31 – Interventions and influences from interventions in Arctic 1.

source	values	target	values	details
law enforcement	0 to 10	resources development	-1 to 2	
		environment	0 to -2	
		crime rate	1 to -4	
infrastructure development	0 to 10	environment	0 to -3	
		socio-economic welfare	0 to 4	
		Arctic infrastructure	0 to 3	
environmental protection (request support)	0 to 10	environment	0 to 3	
economic stimulus package	0 to 10	resources development	-1 to 3	
		socio-economic welfare	0 to 4	

A.1.4 Miscellaneous

action points begin with 16
carryover half of the unused actions points to the next turn

total number of turns maximum of 12 turns

turns required to win minimum of 4 turns

scenario events 2 events

A.2 Arctic 2

A.2.1 Variables

Table 32 – Variables in Arctic 2.

variable	details	initial value	winning value	failure value
sustainability objective	no color (mediating)	0		
crime rate	uncertainty $\pm$ 2	7	<= 9	
environment		12	>= 14	
environmental policies	no color (mediating)	5		
Arctic infrastructure		9	>= 10	
resource development		4	>= 13	
socio-economic welfare		6	>= 15	0

A.2.2 Relations between variables

Table 33 – Relations between variables in Arctic 2.

source	values	target	values	details
crime rate	0 to 20	crime rate	0 to -2	
		socio-economic welfare	0 to -3	
		resources development	0 to -3	
environment	0 to 20	environment	-3 to 2	
		socio-economic welfare	-3 to 1	
		environmental policies	2 to -2	hidden
environmental policies	0 to 20	environmental policies	-1 to -1	conditional env. policies $>$ 10
		environmental policies	0 to 0	conditional env. policies $=$ 10
		environmental policies	1 to 1	conditional env. policies $<$ 10
Arctic infrastructure	0 to 20	socio-economic welfare	-1 to 2	
		resources development	0 to 2	
		Arctic infrastructure	0 to -3	
resource development	0 to 20	socio-economic welfare	0 to 2	
		environment	0 to -4	conditional env. policies $<=$ 6
		environment	0 to -2	conditional env. policies $>$ 6 and $<=$ 13
		environment	0 to 3	conditional env. policies $>$ 13
		resources development	0 to 3	conditional env. policies $<$ 11
		resources development	0 to 2	conditional env. policies $>$ 10
socio-economic welfare	0 to 20	socio-economic welfare	-1 to 1	
		crime rate	3 to 2	

A.2.3 Interventions and influences from interventions

Table 34 – Interventions and influences from interventions in Arctic 2.

source	values	target	values	details
law enforcement	0 to 10	resources development	-1 to 2	
		environment	0 to -2	
		crime rate	1 to -4	
infrastructure development	0 to 10	environment	0 to -3	
		socio-economic welfare	0 to 4	
		Arctic infrastructure	0 to 3	
environmental protection (request support)	0 to 10	environmental policies	0 to 3	
		environment	0 to 3	
economic stimulus package	0 to 10	resources development	-1 to 3	
		socio-economic welfare	0 to 4	

A.2.4 Miscellaneous

action points begin with 16
carryover half of the unused actions points to the next turn

total number of turns maximum of 12 turns

turns required to win minimum of 8 turns

scenario events 2 events

A.3 Arctic 3

A.3.1 Variables

Table 35 – Variables in Arctic 3.

variable	details	initial value	winning value	failure value
incident management	no color (mediating)	4		
sustainability objective	no color (mediating)	0		
unresolved incidents		5		
crime rate	uncertainty $\pm$ 2	7	$<=$ 9	
environment		12	$>=$ 14	
environmental policies	no color (mediating)	5		
Arctic infrastructure		9	$>=$ 10	
resource development		4	$>=$ 13	
socio-economic welfare		6	$>=$ 15	0

A.3.2 Relations between variables

Table 36 – Relations between variables in Arctic 3.

source	values	target	values	details
incident management	0 to 20	incident management	0 to -20	
		unresolved incidents	0 to -20	
unresolved incidents	0 to 20	resources development	1 to -6	delay 1
		environment	0 to -4	
		socio-economic welfare	0 to -6	
crime rate	0 to 20	crime rate	0 to -2	delay 1
		socio-economic welfare	0 to -3	delay 1
		resources development	0 to -3	
environment	0 to 20	environment	-3 to 2	delay 1
		socio-economic welfare	-3 to 1	
		environmental policies	2 to -2	hidden
environmental policies	0 to 20	environmental policies	-1 to -1	conditional env. policies > 10, delay 1
		environmental policies	0 to 0	conditional env. policies = 10, delay 1
		environmental policies	1 to 1	conditional env. policies < 10, delay 1
Arctic infrastructure	0 to 20	socio-economic welfare	-1 to 2	
		resources development	0 to 2	
		Arctic infrastructure	0 to -3	
resource development	0 to 20	socio-economic welfare	0 to 2	
		environment	0 to -4	conditional env. policies <= 6
		environment	0 to -2	conditional env. policies > 6 and <= 13
		environment	0 to 3	conditional env. policies > 13
		resources development	0 to 3	conditional env. policies < 11, delay 1
		resources development	0 to 2	conditional env. policies > 10, delay 1
socio-economic welfare	0 to 20	socio-economic welfare	-1 to 1	delay 1
		crime rate	3 to 2	

A.3.3 Interventions and influences from interventions

Table 37 – Interventions and influences from interventions in Arctic 3.

source	values	target	values	details
law enforcement	0 to 10	resources development	-1 to 2	
		environment	0 to -2	
		crime rate	1 to -4	
incident management budget (resupply)	0 to 10	incident management	0 to 10	
infrastructure development	0 to 10	environment	0 to -3	
		socio-economic welfare	0 to 4	
		Arctic infrastructure	0 to 3	
environmental protection (request support)	0 to 10	environmental policies	0 to 3	
		environment	0 to 3	delay 1
economic stimulus package	0 to 10	resources development	-1 to 3	delay 1
		socio-economic welfare	0 to 4	delay 1

A.3.4 Miscellaneous

action points begin with 20
carryover half of the unused actions points to the next turn

total number of turns maximum of 12 turns

turns required to win minimum of 6 turns

scenario events 13 events

A.4 COIN 1

A.4.1 Variables

Table 38 – Variables in COIN 1.

variable	details	initial value	winning value	failure value
criminality		13	>= 9	
host nation governance		6	>= 8	
infrastructures		7	>= 6	
insurgent forces	uncertainty ± 1	12	>= 15	
local forces		4	>= 6	
local media		11	>= 6	
population allegiance		15	>= 7	0
socio-economic welfare		5	>= 7	
cultural understanding	no color (mediating)	3	>= 15	

A.4.2 Relations between variables

Table 39 – Relations between variables in COIN 1.

source	values	target	values	details
criminality	0 to 20	insurgent forces	-2 to 0	
		local media	-5 to 5	delay 1
		population allegiance	-3 to 1	
		socio-economic welfare	-5 to 0	
host nation governance	0 to 20	host nation governance	-1	
		infrastructures	0 to 3	delay 1
		local forces	0 to 3	delay 1
		local media	-3 to 2	
		population allegiance	-5 to 5	
		socio-economic welfare	-2 to 4	delay 1
infrastructures	0 to 20	infrastructures	-1	delay 1
		population allegiance	-2 to 2	
		socio-economic welfare	-3 to 3	
insurgent forces	0 to 20	criminality	-3 to 0	hidden
		infrastructures	-3 to 0	hidden
		local forces	-4 to 0	hidden
		local media	-3 to 0	hidden, delay 1
		population allegiance	-3 to 0	hidden
		host nation governance	-3 to 0	hidden, conditional pop. allegiance < 10
		host nation governance	-3 to 0	hidden, conditional pop. allegiance >= 10
		insurgent forces	-4 to 0	hidden, conditional pop. allegiance < 10
		insurgent forces	-2 to 0	hidden, conditional pop. allegiance >= 10
local forces	0 to 20	criminality	0 to 3	
		local forces	-1	delay 1
		population allegiance	-5 to 2	
		insurgent forces	0 to 2	conditional insurgent forces > 12
		insurgent forces	0 to 3	conditional insurgent forces < 13
local media	0 to 20	population allegiance	-4 to 4	
population allegiance	0 to 20	criminality	-2 to 1	delay 1
		insurgent forces	-5 to 2	
		local forces	-1 to 1	
		local media	-2 to 2	
		population allegiance	-1 to 1	
socio-economic welfare	-5 to 0	criminality	1 to -6	
		local media	-3 to 3	
		population allegiance	-4 to 2	
		insurgent forces	0 to -2	conditional pop. allegiance < 10
		insurgent forces	0	conditional pop. allegiance >= 10

A.4.3 Interventions and influences from interventions

Table 40 – Interventions and influences from interventions in COIN 1.

source	values	target	values	details
cultural training	0 to 10	cultural understanding	0 to 4	
gov. capacity building	0 to 10	host nation governance	0 to 5	cond. cultural understanding < 7
		host nation governance	0 to 6	cond. cultural understanding > 13
		host nation governance	0 to 5	cond. cultural understanding >= 7 & <= 13
		population allegiance	0 to 2	
humanitarian aid	0 to 10	insurgent forces	0 to -2	
		local media	-5 to 2	cond. socio-economic welfare <= 10
		local media	0	cond. socio-economic welfare > 10
		population allegiance	-2 to 3	cond. socio-economic welfare <= 10
		population allegiance	0 to 2	cond. socio-economic welfare > 10
		socio-economic welfare	0 to 5	
influence operations	0 to 10	insurgent forces	0 to 1	delay 1, cond. cultural understanding < 7
		insurgent forces	0 to 3	delay 1, cond. cultural understanding > 13
		insurgent forces	0 to 2	delay 1, cond. cultural understanding >= 7 & <= 13
		local media	0 to 2	delay 1, cond. cultural understanding > 10
		local media	0 to 1	delay 1, cond. cultural understanding <= 10
		population allegiance	0 to 2	delay 1, cond. cultural understanding < 7
		population allegiance	0 to 3	delay 1, cond. cultural understanding > 12
		population allegiance	0 to 2	delay 1, cond. cultural understanding >= 7 & <= 12
infrastructure dev.	0 to 10	infrastructures	0 to 5	delay 1
		population allegiance	0 to 2	delay 1, cond. cultural understanding < 7
		population allegiance	0 to 3	delay 1, cond. cultural understanding >= 7 & <= 11
		population allegiance	0 to 5	delay 1, cond. cultural understanding > 11
security operations	0 to 10	criminality	0 to 1	cond. cultural understanding < 8
		criminality	0 to 3	cond. cultural understanding >= 8
		insurgent forces	0 to 5	cond. cultural understanding <= 12
		insurgent forces	0 to 2	cond. cultural understanding > 12
		local media	0 to -4	
		population allegiance	-4 to -10	cond. cultural understanding < 7
		population allegiance	-4 to -6	cond. cultural understanding > 13
		population allegiance	-4 to -10	cond. cultural understanding >= 7 & <= 13
train/supply forces	0 to 10	local forces	0 to 6	delay 1, cond. cultural understanding > 9
		local forces	0 to 5	delay 1, cond. cultural understanding <= 9

A.4.4 Contribution of interventions to action points

Table 41 – Contribution of interventions to action points in COIN 1.

source	values	target	values
host nation governance	0 to 20	action points	7 to 0
local media	0 to 20	action points	0 to 5
population allegiance	0 to 20	action points	0 to 6
socioeconomic welfare	0 to 20	action points	9 to 5

A.4.5 Miscellaneous

action points begin with 22
carryover half of the unused actions points to the next turn

total number of turns maximum of 7 turns

turns required to win minimum of 4 turns

scenario events 6 (counting 1 counter-action by insurgents after each turn)

A.5 COIN 2

A.5.1 Variables

Table 42 – Variables in COIN 2.

variable	details	initial value	winning value	failure value
criminality		13	>= 9	
host nation governance		6	>= 8	
infrastructures		7	>= 6	
insurgent forces	uncertainty $\pm$ 1	12	>= 15	
local forces		4	>= 6	
local media		11	>= 6	
population allegiance		14	>= 7	0
socio-economic welfare		5	>= 7	
cultural understanding	no color (mediating)	3	>= 15	

A.5.2 Relations between variables

Table 43 – Relations between variables in COIN 2.

source	values	target	values	details
criminality	0 to 20	insurgent forces	-2 to 0	
		local media	-5 to 5	delay 1
		population allegiance	-3 to 1	
		socio-economic welfare	-5 to 0	
host nation governance	0 to 20	host nation governance	-1	
		infrastructures	0 to 3	delay 1
		local forces	0 to 3	delay 1
		local media	-3 to 2	
		population allegiance	-5 to 5	
		socio-economic welfare	-2 to 4	delay 1
infrastructures	0 to 20	infrastructures	-1	delay 1
		population allegiance	-2 to 2	
		socio-economic welfare	-3 to 3	
insurgent forces	0 to 20	criminality	-3 to 0	hidden
		infrastructures	-3 to 0	hidden
		local forces	-4 to 0	hidden
		local media	-3 to 0	hidden, delay 1
		population allegiance	-3 to 0	hidden
		host nation governance	-3 to 0	hidden, conditional pop. allegiance < 10
		host nation governance	-3 to 0	hidden, conditional pop. allegiance >= 10
		insurgent forces	-4 to 0	hidden, conditional pop. allegiance < 10
		insurgent forces	-2 to 0	hidden, conditional pop. allegiance >= 10
local forces	0 to 20	criminality	0 to 3	
		local forces	-1	delay 1
		population allegiance	-5 to 2	
		insurgent forces	0 to 2	conditional insurgent forces > 12
		insurgent forces	0 to 3	conditional insurgent forces < 13
local media	0 to 20	population allegiance	-4 to 4	
population allegiance	0 to 20	criminality	-2 to 1	delay 1
		insurgent forces	-5 to 2	
		local forces	-1 to 1	
		local media	-2 to 2	
		population allegiance	-1 to 1	
socio-economic welfare	-5 to 0	criminality	1 to -6	
		local media	-3 to 3	
		population allegiance	-4 to 2	
		insurgent forces	0 to -2	conditional pop. allegiance < 10
		insurgent forces	0	conditional pop. allegiance >= 10

A.5.3 Interventions and influences from interventions

Table 44 – Interventions and influences from interventions in COIN 2.

source	values	target	values	details
cultural training	0 to 10	cultural understanding	0 to 4	
gov. capacity building	0 to 10	host nation governance	0 to 5	cond. cultural understanding < 7
		host nation governance	0 to 6	cond. cultural understanding > 13
		host nation governance	0 to 5	cond. cultural understanding >= 7 & <= 13
		population allegiance	0 to 2	
humanitarian aid	0 to 10	insurgent forces	0 to -2	
		local media	-5 to 2	cond. socio-economic welfare <= 10
		local media	0	cond. socio-economic welfare > 10
		population allegiance	-2 to 3	cond. socio-economic welfare <= 10
		population allegiance	0 to 2	cond. socio-economic welfare > 10
		socio-economic welfare	0 to 5	
influence operations	0 to 10	insurgent forces	0 to 1	delay 1, cond. cultural understanding < 7
		insurgent forces	0 to 3	delay 1, cond. cultural understanding > 13
		insurgent forces	0 to 2	delay 1, cond. cultural understanding >= 7 & <= 13
		local media	0 to 2	delay 1, cond. cultural understanding > 10
		local media	0 to 1	delay 1, cond. cultural understanding <= 10
		population allegiance	0 to 2	delay 1, cond. cultural understanding < 7
		population allegiance	0 to 3	delay 1, cond. cultural understanding > 12
		population allegiance	0 to 2	delay 1, cond. cultural understanding >= 7 & <= 12
infrastructure dev.	0 to 10	infrastructures	0 to 5	delay 1
		population allegiance	0 to 2	delay 1, cond. cultural understanding < 7
		population allegiance	0 to 3	delay 1, cond. cultural understanding >= 7 & <= 11
		population allegiance	0 to 5	delay 1, cond. cultural understanding > 11
security operations	0 to 10	criminality	0 to 1	cond. cultural understanding < 8
		criminality	0 to 3	cond. cultural understanding >= 8
		insurgent forces	0 to 5	cond. cultural understanding <= 12
		insurgent forces	0 to 2	cond. cultural understanding > 12
		local media	0 to -4	
		population allegiance	-4 to -10	cond. cultural understanding < 7
		population allegiance	-4 to -6	cond. cultural understanding > 13
		population allegiance	-4 to -10	cond. cultural understanding >= 7 & <= 13
train/supply forces	0 to 10	local forces	0 to 6	delay 1, cond. cultural understanding > 9
		local forces	0 to 5	delay 1, cond. cultural understanding <= 9

A.5.4 Contribution of interventions to action points

Table 45 – Contribution of interventions to action points in COIN 2.

source	values	target	values
host nation governance	0 to 20	action points	7 to 0
local media	0 to 20	action points	0 to 5
population allegiance	0 to 20	action points	0 to 6
socioeconomic welfare	0 to 20	action points	9 to 5

A.5.5 Miscellaneous

action points	begin with 20 carryover half of the unused actions points to the next turn
total number of turns	maximum of 10 turns
turns required to win	minimum of 5 turns
scenario events	9 (counting 1 counter-action by insurgents after each turn)

A.6 *Cybernetia* in Ecopolicy

A.6.1 Variables

Table 46 – Variables in Ecopolicy.

variable	initial value	winning value	failure value
sanitation	1	17	0
production	9	11 to 16 (quadratic)	0
environmental stress	13	10	0
education	4	21	0
quality of life	9	21	0
growth rate	20	13 to 19 (quadratic)	0
population	23	7 to 32 (quadratic)	0
policy	0	16	-11

A.6.2 Relations between variables

Table 47 – Relations between variables in Ecopolicy.

source	values	target	values	details
environmental stress	0 to 29	environmental stress	-3 to 0	
		quality of life	-25 to 2	
population	0 to 48	quality of life	-10 to 0	
growth rate	0 to 29	population	-4 to 3	delay 1

A.6.3 Interventions and influences from interventions

Table 48 – Interventions and influences from interventions in Ecopolicy.

source	values	target	values	details
sanitation	0 to 29	sanitation	-3 to 0	
		environmental stress	-9 to 2	
production	0 to 29	production	0 to 2	delay 1
		environmental stress	0 to 22	
quality of life	0 to 29	quality of life	-2 to 2	
		growth rate	-15 to 2	
		policy	-10 to 5	
education	0 to 29	education	-1 to 2	
		growth rate	0 to 4	
		quality of life	-2 to 6	

A.6.4 Contribution of interventions to action points

Table 49 – Contribution of interventions to action points in Ecopolicy.

source	values	target	values
policy	-11 to 37	action points	-6 to 3
population	0 to 48	action points	0 to 9
production	0 to 29	action points	-4 to 10
quality of life	0 to 29	action points	-6 to 5

A.6.5 Miscellaneous

action points	begin with 8
total number of turns	maximum of 12 turns
turns required to win	minimum of 9 turns
scenario events	0 events **(NOTE: random events turned off)**

Appendix B

Objective Measures of Complexity Calculations

Table 50 – Values of the parameters of complexity OMC_{DDM} for the DDM scenarios.

parameter	Arctic 1	Arctic 2	Arctic 3	COIN 1	COIN 2	Ecopolicy
CC_{DDM}	8	20	32	50	50	9
CNC_{DDM}	24.20	88.17	171.13	277.78	277.78	28.13
SDC_{DDM}	200	756	1600	2128	2128	784
I_{DDM}	36	40	70	378	378	60
A_{DDM}	576	640	1400	8316	7560	480
W_{DDM}	57	88	142	245	245	90
D_{DDM}	.5016	.5016	.5016	.1834	.2072	.5272
L_{DDM}	.0795	.1076	.0913	.0882	.0882	.2007
U_{DDM}	.0944	.0944	.2306	.1957	.2028	.0000
S_{DDM}	.5016	.5443	.5228	.5217	.5657	.5476

B.1 Structural complexity C_{DDM}

B.1.1 Cyclomatic complexity CC$_{DDM}$

$$CC_{DDM} = v(G) = e - n + 2p$$

Table 51 – Cyclomatic complexity CC$_{DDM}$ for the DDM scenarios.

scenario	variables	relationships	connected components	CC$_{DDM}$
Arctic 1	5	8 + 3 conditional	1	8
Arctic 2	6	15 + 8 conditional	1	20
Arctic 3	8	20 + 9 delay + 8 conditional	1	32
COIN 1	9	34 + 8 delay + 8 conditional	1	50
COIN 2	9	34 + 8 delay + 8 conditional	1	50
Ecopolicy	8	14 + 1 delay	1	9

B.1.2 Coefficient of network complexity CNC$_{DDM}$

$$CNC_{DDM} = \frac{e^2}{n}$$

Table 52 – Coefficient of network complexity CNC$_{DDM}$ for the DDM scenarios.

scenario	variables	relationships	CNC$_{DDM}$
Arctic 1	5	8 + 3 conditional	24.20
Arctic 2	6	15 + 8 conditional	88.17
Arctic 3	8	20 + 9 delay + 8 conditional	171.13
COIN 1	9	34 + 8 delay + 8 conditional	277.78
COIN 2	9	34 + 8 delay + 8 conditional	277.78
Ecopolicy	8	14 + 1 delay	28.13

B.1.3 System dynamics complexity SDC$_{DDM}$

$$SDC_{DDM} = endogenous\ variables \text{ x } exogenous\ variables \text{ x } relations$$

Table 53 – System dynamics complexity SDC$_{DDM}$ for the DDM scenarios.

scenario	endogenous vars	exogenous vars	relations	SDC$_{DDM}$
Arctic 1	4	5	10	200
Arctic 2	6	6	21	756
Arctic 3	8	8	25	1600
COIN 1	8	7	38	2128
COIN 2	8	7	38	2128
Ecopolicy	8	7	14	784

B.2 Information complexity I_{DDM}

B.2.1 Information complexity variant I$_{DDM}$

$$I_{DDM} = interventions \text{ x } (influences\ from\ interventions + contributions\ to\ action\ points)$$

Table 54 – Information complexity I$_{DDM}$ for the DDM scenarios.

scenario	interventions	influences from interventions	contrib. to action points	I$_{DDM}$
Arctic 1	4	9	0	36
Arctic 2	4	10	0	40
Arctic 3	5	11 + 3 delay	0	70
COIN 1	7	17 + 6 delay + 27 cond.	4	378
COIN 2	7	17 + 6 delay + 27 cond.	4	378
Ecopolicy	4	10 + 1 delay	4	60

B.2.2 Action complexity variant A_{DDM}

A_{DDM}= *action points* x *interventions* x *(influences from interventions* + *contributions to action points)*

Table 55 – Action complexity A_{DDM} for the DDM scenarios.

scenario	action points	interventions	influences from interventions	contrib. to action points	I_{DDM}
Arctic 1	16	4	9	0	576
Arctic 2	16	4	10	0	640
Arctic 3	20	5	11 + 3 delay	0	1400
COIN 1	22	7	17 + 6 delay + 27 cond.	4	8316
COIN 2	20	7	17 + 6 delay + 27 cond.	4	7560
Ecopolicy	8	4	10 + 1 delay	4	480

B.3 Cognitive weight W_{DDM}

W_{DDM} = *(relations* x W_i*)* + *(influences from interventions* x W_i*)* + *(contributions to action points* x W_i*)*

Table 56 – Cognitive weights of basic control structures (BCS), from Shao and Wang (2003).

BCS	notation	calibrated cognitive weight w_i
sequence	$\rightarrow$	1
branch	$\mid$	2
case	$\mid \ldots \mid \ldots$	3
for-loop	R^i	3
repeat-loop	R^*	3
while-loop	R^*	3
function call	$\rightarrowtail$	2
recursion	$\circlearrowleft$	3
parallel	$\parallel$ or $\Box$	4
interrupt	$\lightning$	4

Table 57 – Cognitive weight W_{DDM} for the DDM scenarios (columns feature the total cognitive weight for each characteristic) *.

scenario	**W$_i$ relations**	**W$_i$ infl. from interventions**	**W$_i$ contrib. to action points**	**W$_{DDM}$**
Arctic 1	3	17	0	57
Arctic 2	8	24	0	88
Arctic 3	8	30	12	142
COIN 1	31	48	13	245
COIN 2	31	48	13	245
Ecopolicy	28	(included in relations)	2	90

* details of influences on variables:

— conditional relationships are considered as the category '*branch*' (such as a *if-then-else* function), with a BCS weight of $w_i = 2$

— all variable value intervals (e.g., [0,10]) are considered as the category '*case*' (evaluates each input value), with a BCS weight of $w_i = 3$

— delayed relationships are considered as the category '*loop*' (a type of iterative process), with a BCS weight of $w_i = 3$

B.4 Difficulty D_{DDM}

$$D_{DDM} = \textit{(distance to win + (1 - distance to fail)) / 2}$$

where

$$\textit{distance to win} = \left\{ \sqrt{1/n \times \sum_{i=v_1}^{v_n} \left(\frac{G_{success_i} - G_{initial_i}}{G_{optimal_i}} \right)^2} \right\}$$

and

$$\textit{distance to fail} = \left\{ \sqrt{1/n \times \sum_{i=v_1}^{v_n} \left(\frac{G_{initial_i} - G_{failure_i}}{G_{range_i}} \right)^2} \right\}$$

Table 58 – Difficulty D_{DDM} for the DDM scenarios.

scenario	RMS to win	RMS to fail	D_{DDM}
Arctic 1	.288963666	.285714286	.50162469
Arctic 2	.288963666	.285714286	.50162469
Arctic 3	.288963666	.285714286	.50162469
COIN 1	.081009259	.714285714	.183361772
COIN 2	.081009259	.666666667	.207171296
Ecopolicy	.34463722	.290322581	.527157319

B.5 Distance to linearity L_{DDM}

L_{DDM} = (distance to linearity for relations + distance to linearity for influences from interventions) / 2

where

$$distance\ to\ linearity\ for\ relations\ = \left\{ 1/n * \sum_{i=rel_1}^{rel_n} (1 - |L_{a_i}|) \right\}$$

and

$$distance\ to\ linearity\ for\ influences\ from\ interventions\ = \left\{ 1/n * \sum_{i=int_1}^{int_n} (1 - |L_{a_i}|) \right\}$$

Table 59 – Distance to linearity L_{DDM} for the DDM scenarios (excluding relationships and interventions where y is constant).

scenario	mean of regression slopes	L_{DDM}
Arctic 1	.920473684	.0795
Arctic 2	.892407407	.1076
Arctic 3	.908666667	.0913
COIN 1	.911814286	.0882
COIN 2	.911814286	.0882
Ecopolicy	.799333333	.2007

B.6 Uncertainty U_{DDM}

$$U_{DDM} =$$

$$\left\{ \frac{\frac{fuzzy_{var}+hidden_{var}}{variables} + \frac{fuzzy_{\ rel}+hidden_{rel}}{relations} + \frac{fuzzy_{int}+hidden_{int}}{infl.\ on\ variables} + \frac{fuzzy_{ap}+hidden_{ap}}{infl.\ on\ action\ points} + \frac{\#\ events}{\#\ turns}}{\#\ of\ categories\ above\ with\ non-zero\ values} \right\}$$

Table 60 – Uncertainty U_{DDM} for the DDM scenarios.

scenario	hidden/uncertain vars	for influences	for action points	events	U_{DDM}
Arctic 1	2	0	0	2	.0944
Arctic 2	2	1	0	2	.0944
Arctic 3	2	1	0	13	.2306
COIN 1	1	7	0	6	.1957
COIN 2	1	7	0	9	.2028
Ecopolicy	0	0	0	0	0

B.7 Distance to stability S_{DDM}

S_{DDM} = *(distance to end game inertial values + turn handicap to inertial failure) / 2*

where

$$distance\ to\ end\ game\ inertial\ values\ = \left\{ \sqrt{1/n \times \sum_{i=v_1}^{v_n} \left(\frac{S_{inertial_i} - S_{initial_i}}{S_{range_i}} \right)^2} \right\}$$

and

turn handicap to inertial failure = (maximum # turns - # turns to fail from inertia) / maximum # turns

Table 61 – Distance to stability S_{DDM} for the DDM scenarios (the turns handicap to inertial failure is determined by the ratio of the maximum number of turns minus the number of turns to fail in an idle state, to the maximum number of turns).

scenario	RMS (inertial - initial values)	turns handicap to inertial failure	S_{DDM}
Arctic 1	.336615427	$(12-4)/12$	.5016
Arctic 2	.25521115	$(12-2)/12$	.5443
Arctic 3	.29559567	$(12-3)/12$	.5228
COIN 1	.329149863	$(7-2)/7$	0.5217
COIN 2	.331438302	$(10-2)/10$	0.5657
Ecopolicy	.261797943	$(12-2)/12$	0.5476

B.8 Objective parameters of complexity for the DDM scenarios in CODEM

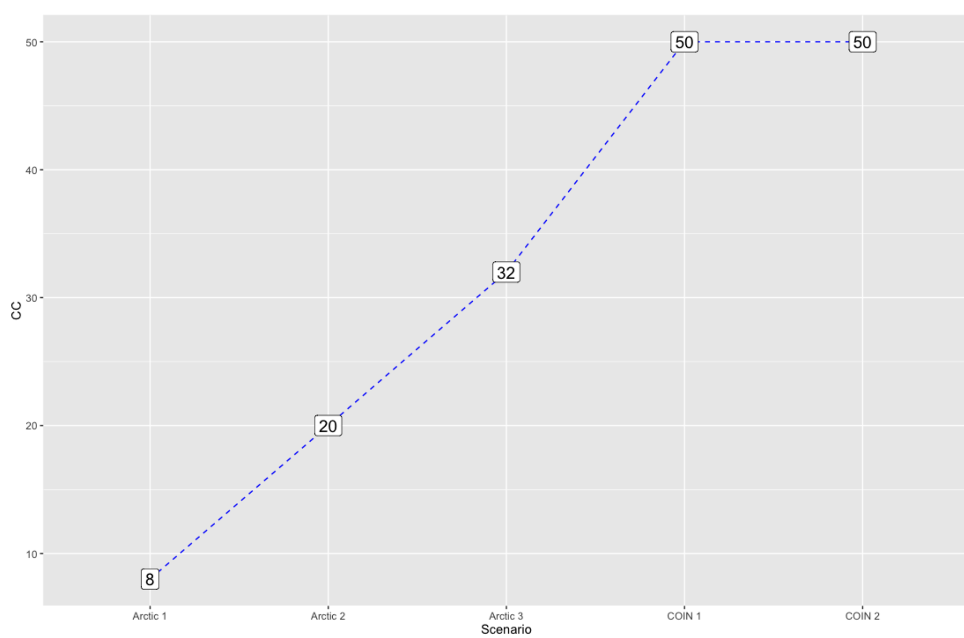

Figure 62 – Measures of structural complexity based on the CC calculation for the five DDM scenarios.

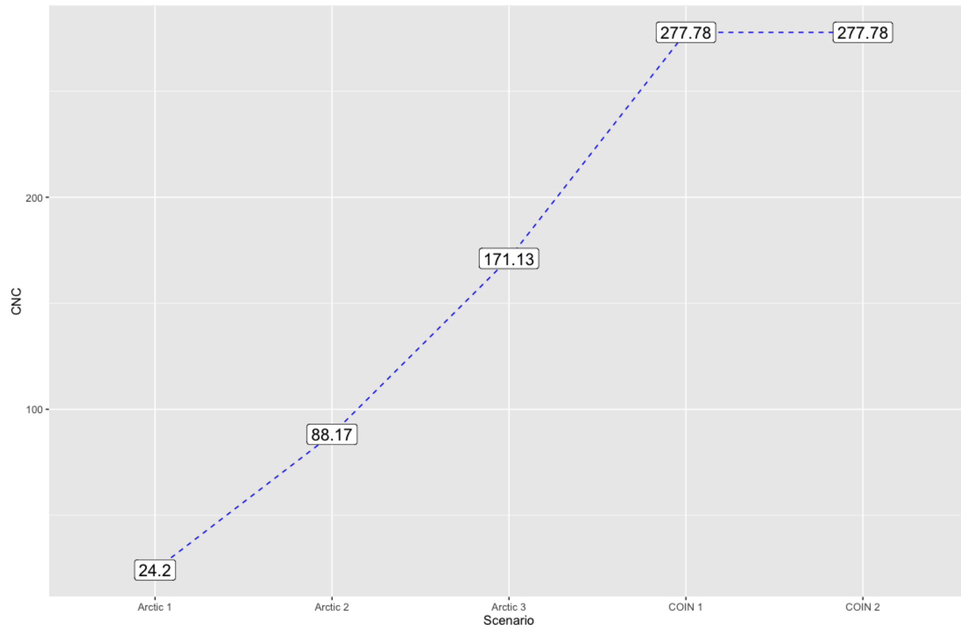

Figure 63 – Measures of structural complexity based on the CNC calculation for the five DDM scenarios.

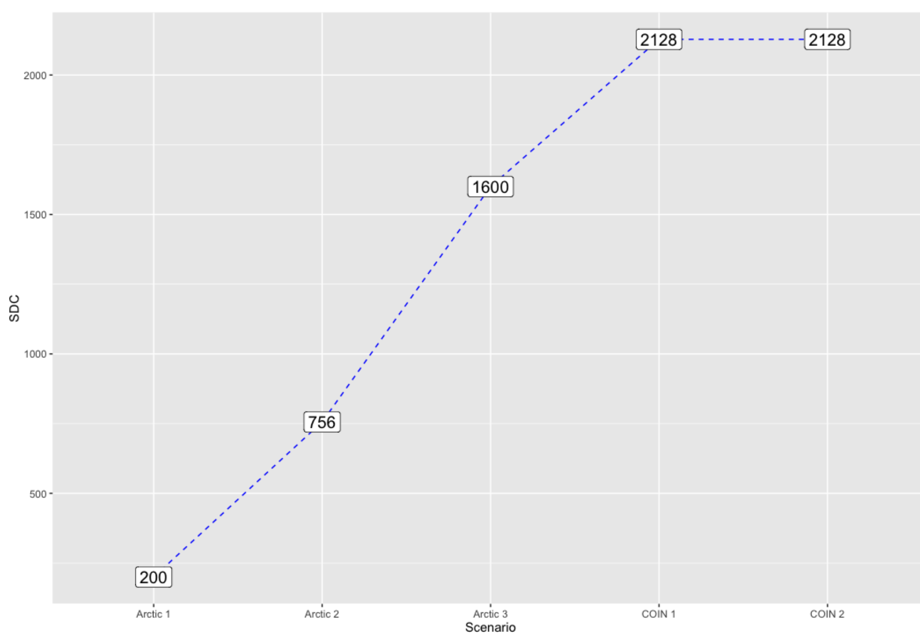

Figure 64 – Measures of structural complexity based on the SDC calculation for the five DDM scenarios.

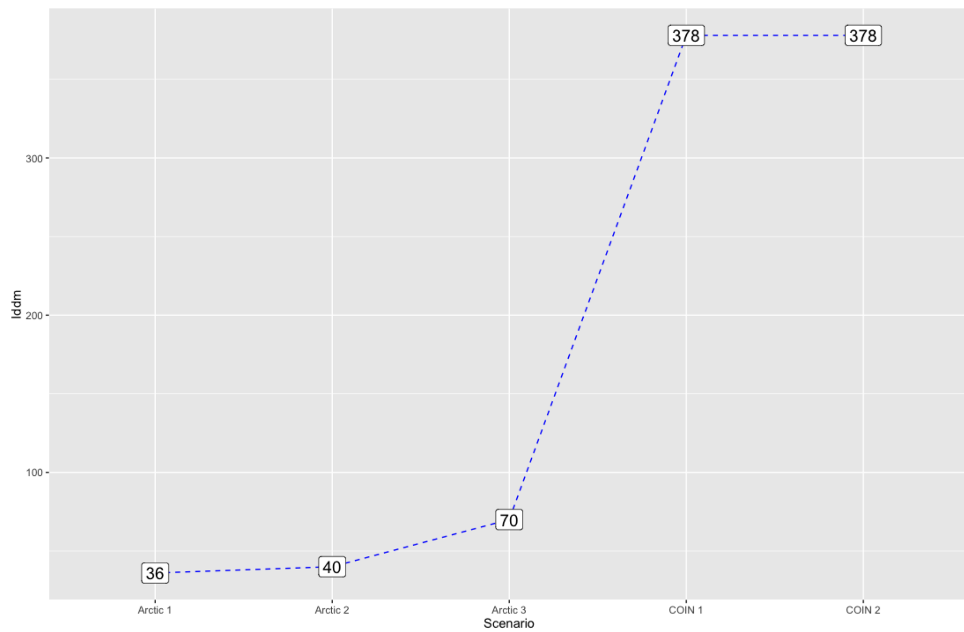

Figure 65 – Measures of information complexity (I_{DDM}) for the five DDM scenarios.

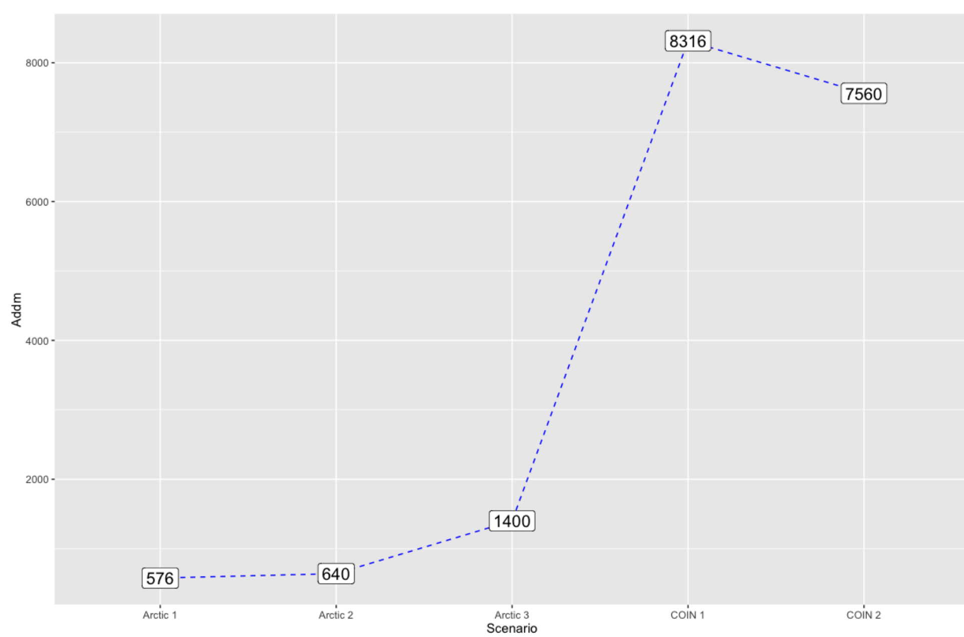

Figure 66 – Measures of action complexity (A_{DDM}) for the five DDM scenarios.

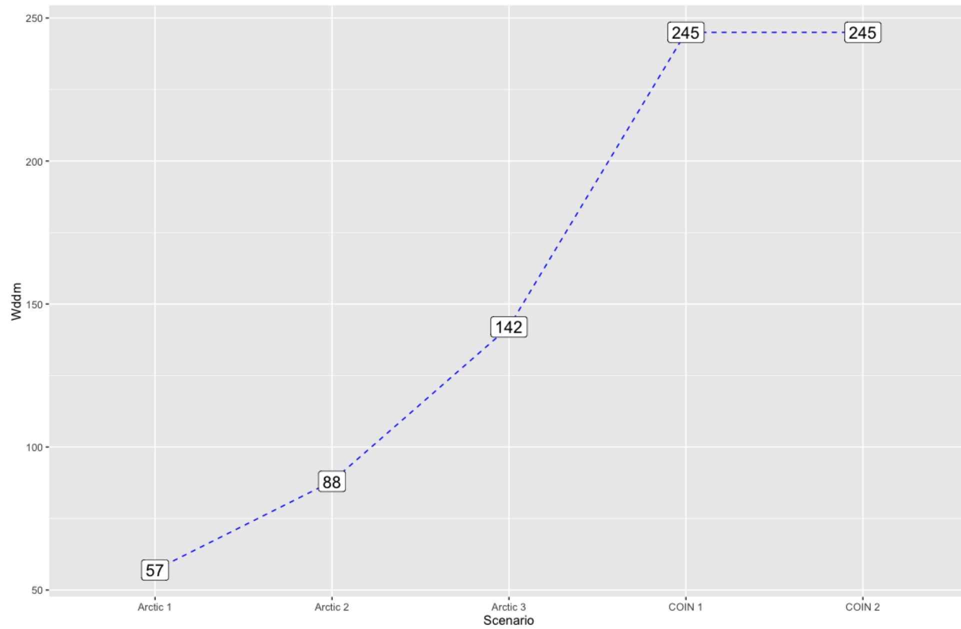

Figure 67 – Measures of cognitive weight (W_{DDM}) for the five DDM scenarios.

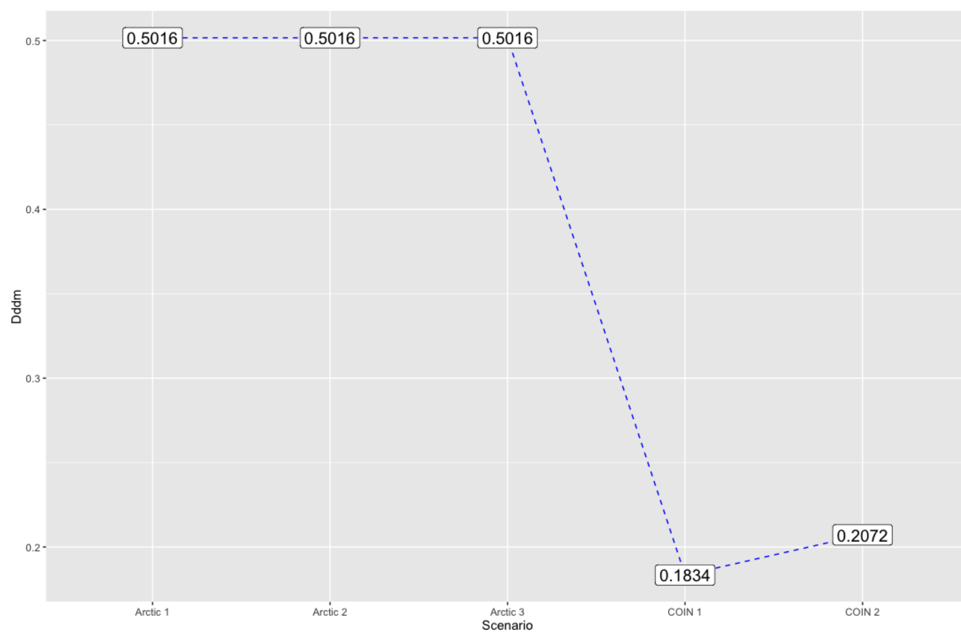

Figure 68 – Measures of difficulty (D_{DDM}) for the five DDM scenarios.

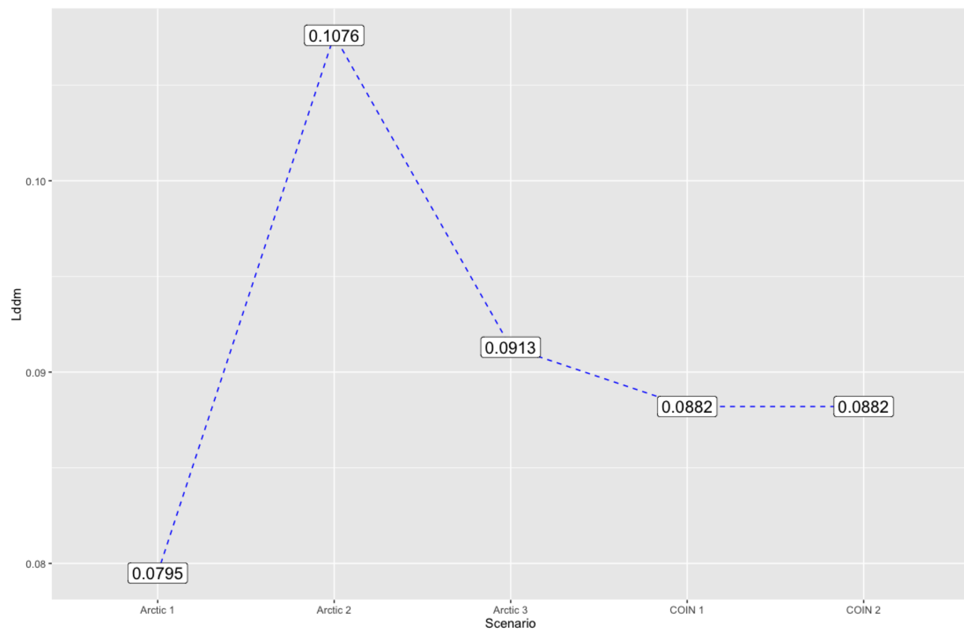

Figure 69 – Measures of nonlinearity (L_{DDM}) for the five DDM scenarios.

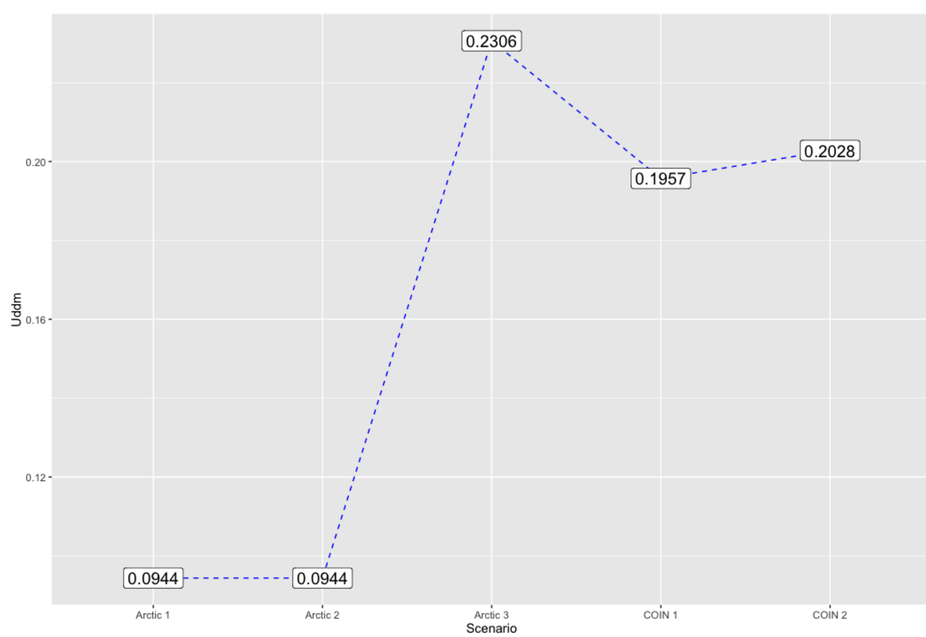

Figure 70 – Measures of uncertainty (U_{DDM}) for the five DDM scenarios.

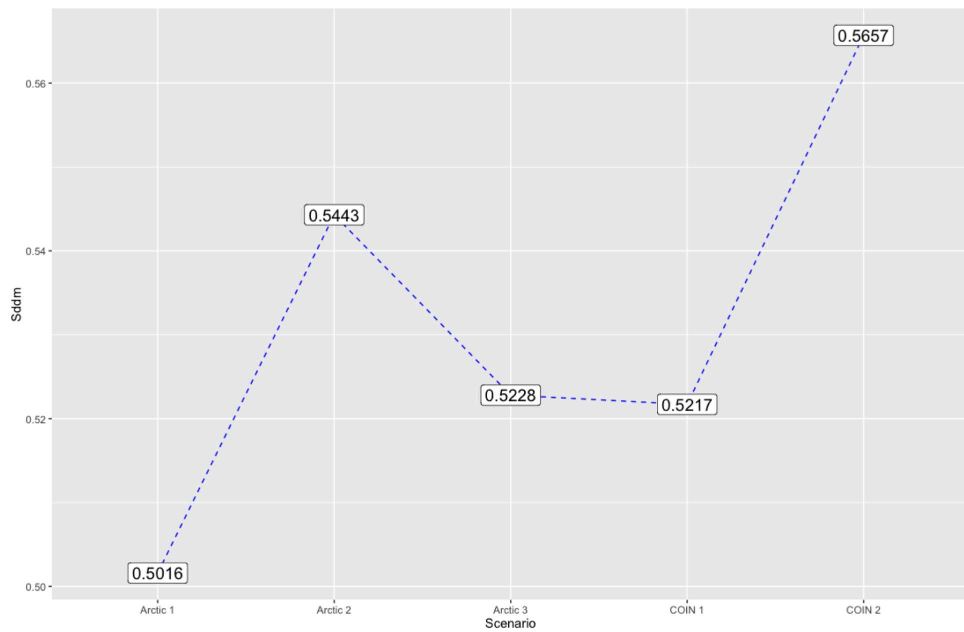

Figure 71 – Measures of system stability (S_{DDM}) for the five DDM scenarios.

B.9 Objective parameters of complexity for the DDM scenarios including Ecopolicy

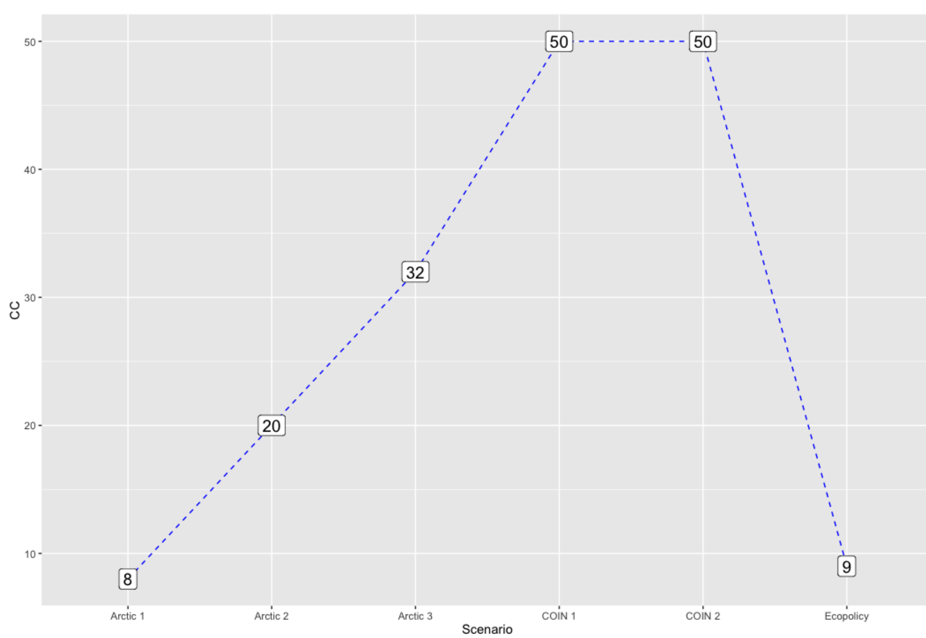

Figure 72 – Measures of structural complexity based on the CC calculation for the DDM scenarios including Ecopolicy.

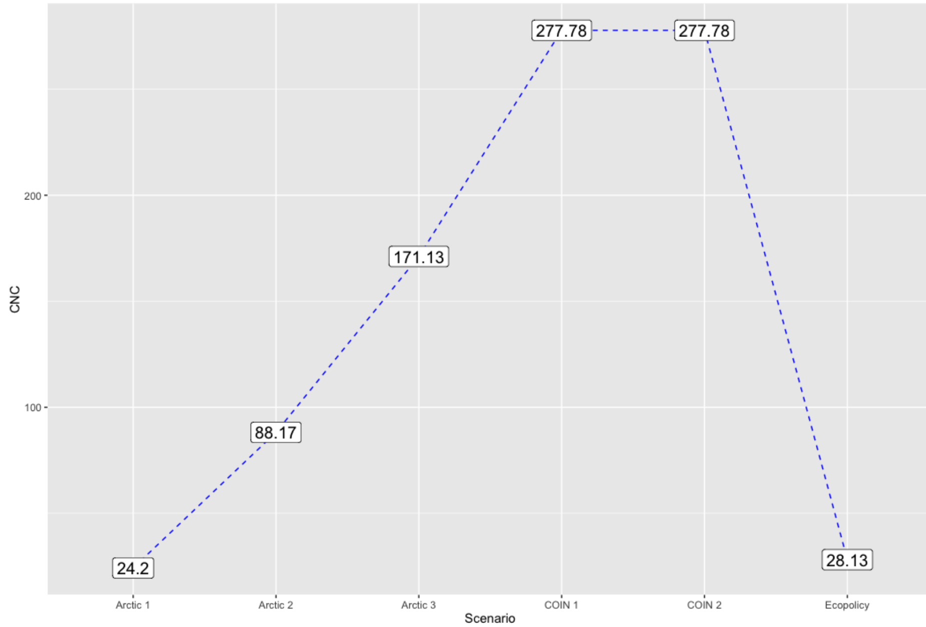

Figure 73 – Measures of structural complexity based on the CNC calculation for the DDM scenarios including Ecopolicy.

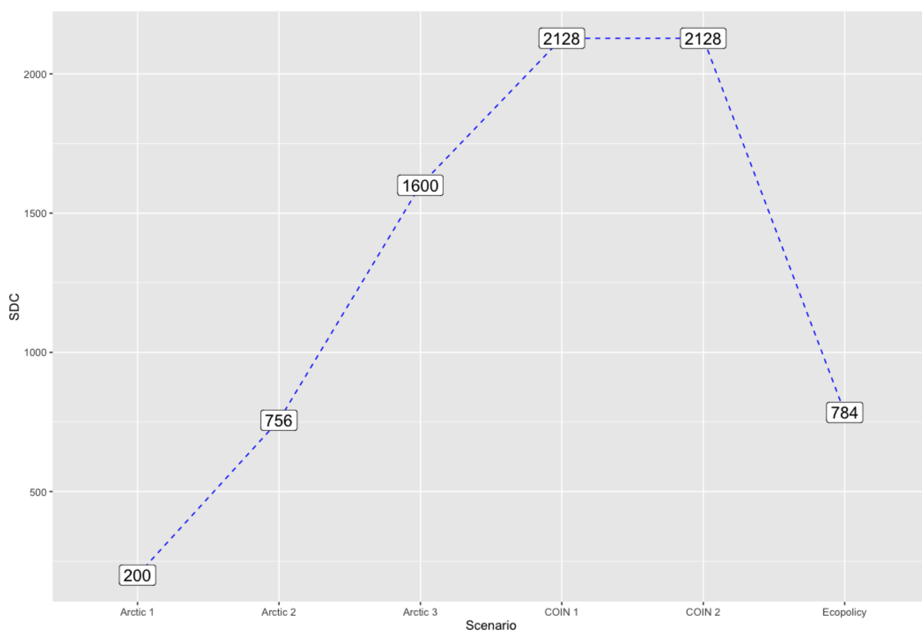

Figure 74 – Measures of structural complexity based on the SDC calculation for the DDM scenarios including Ecopolicy.

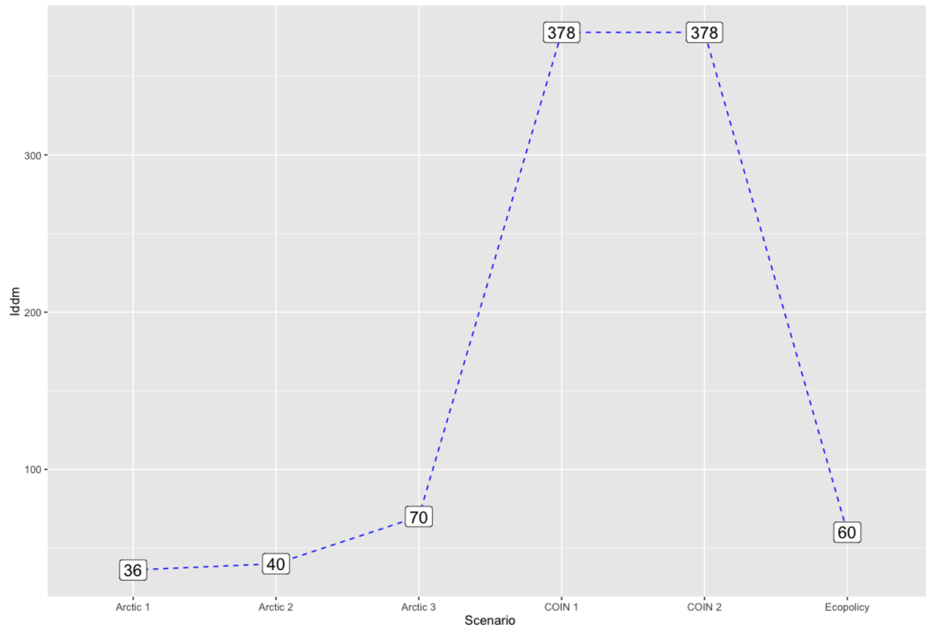

Figure 75 – Measures of information complexity (I_{DDM}) for the DDM scenarios including Ecopolicy.

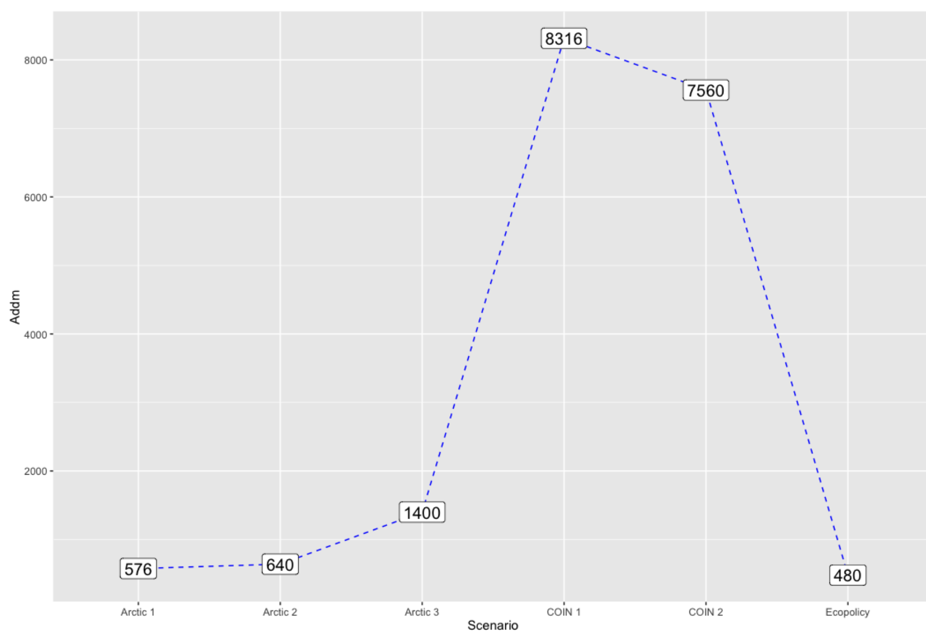

Figure 76 – Measures of action complexity (A_{DDM}) for the DDM scenarios including Ecopolicy.

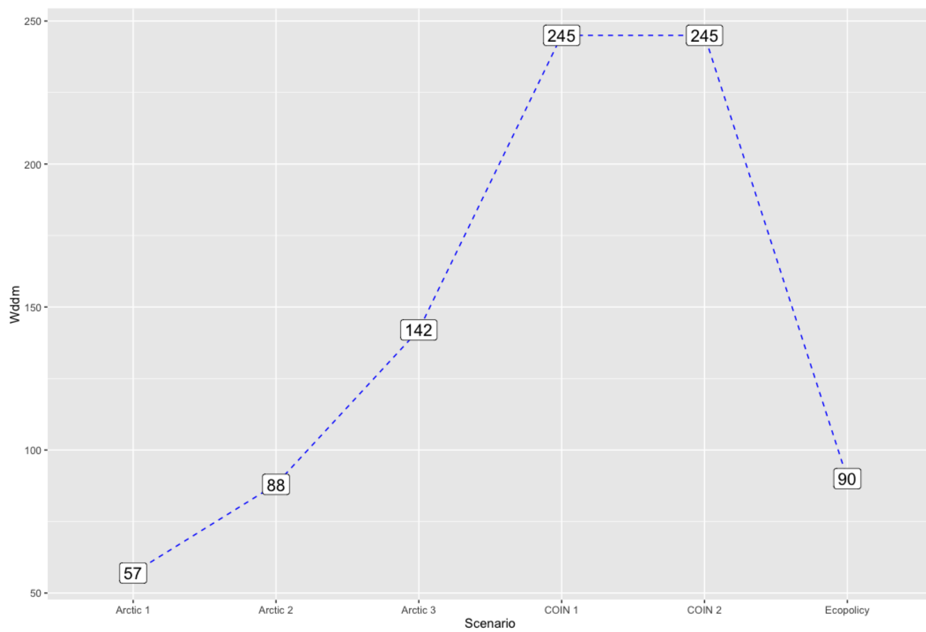

Figure 77 – Measures of cognitive weight (W_{DDM}) for the DDM scenarios including Ecopolicy.

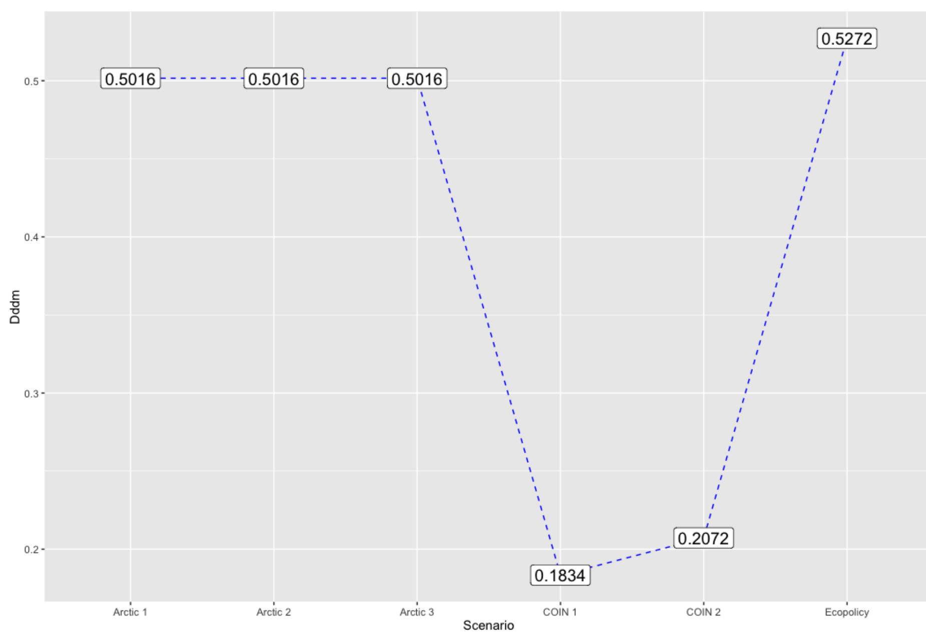

Figure 78 – Measures of difficulty (D_{DDM}) for the DDM scenarios including Ecopolicy.

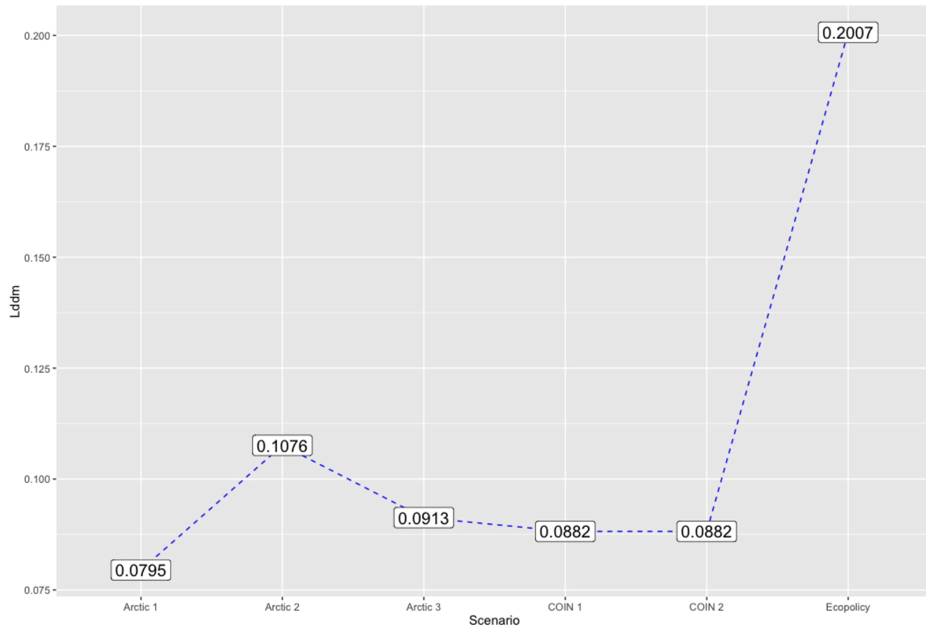

Figure 79 – Measures of nonlinearity (L_{DDM}) for the DDM scenarios including Ecopolicy.

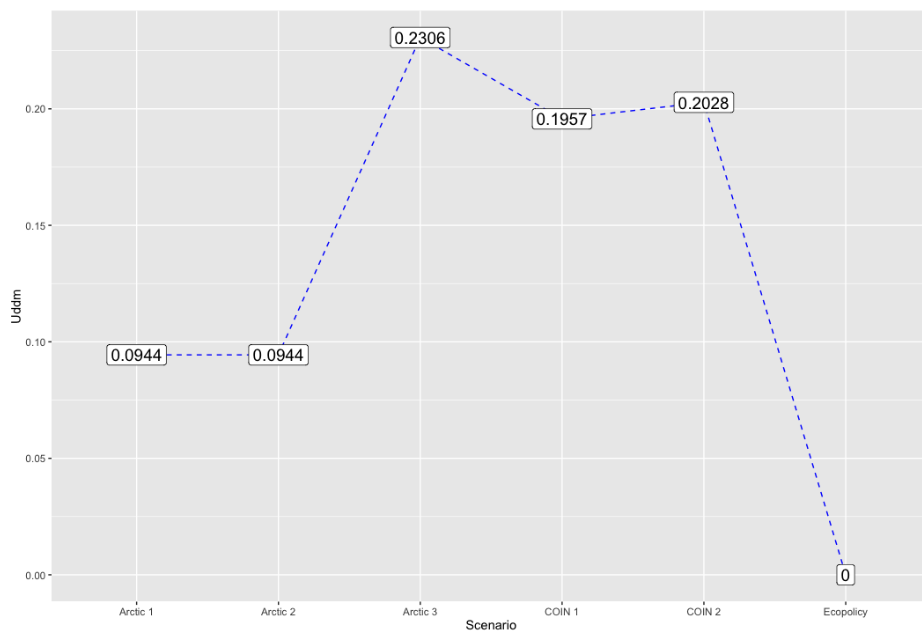

Figure 80 – Measures of uncertainty (U_{DDM}) for the DDM scenarios including Ecopolicy.

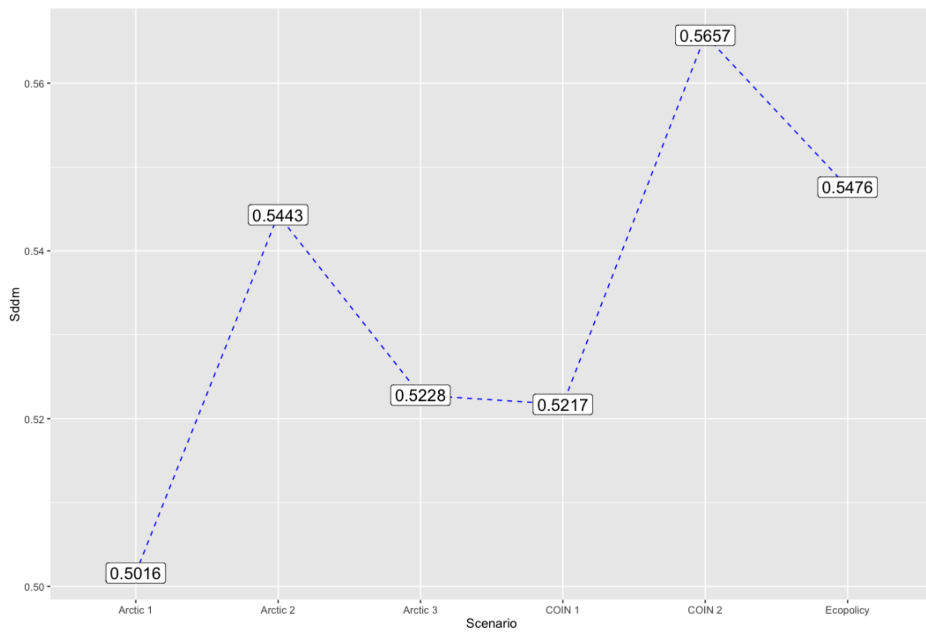

Figure 81 – Measures of system stability (S_{DDM}) for the DDM scenarios including Ecopolicy.

B.10 Objective parameters of complexity for DDM scenarios suitable for a future research program

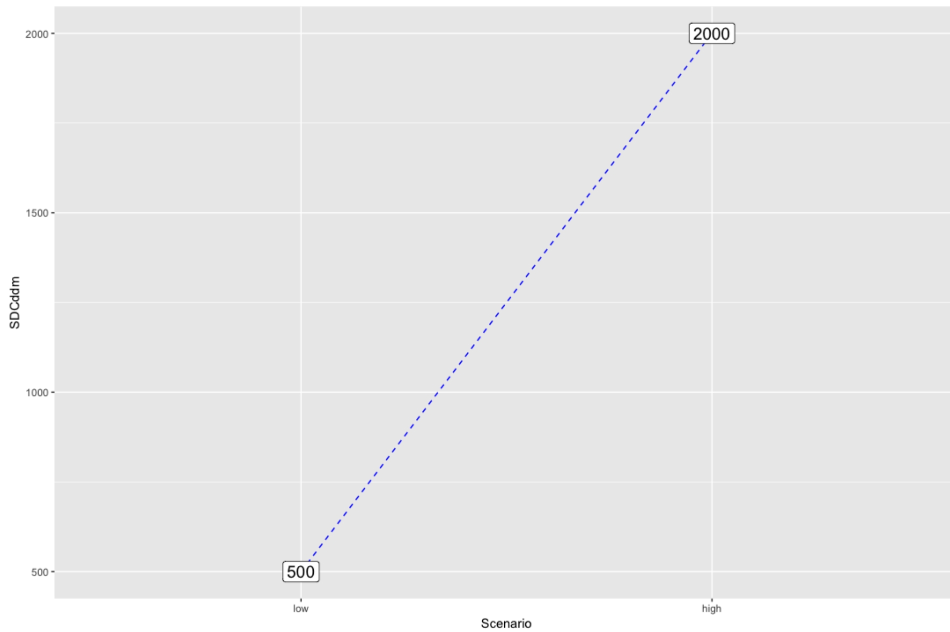

Figure 82 – Structural complexity SDC_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance.

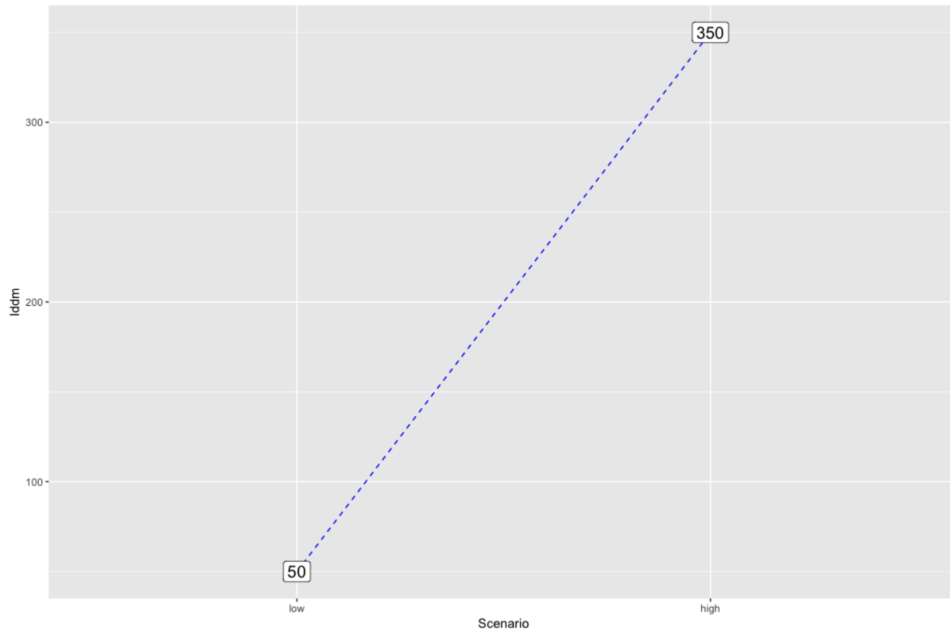

Figure 83 – Information complexity I_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance.

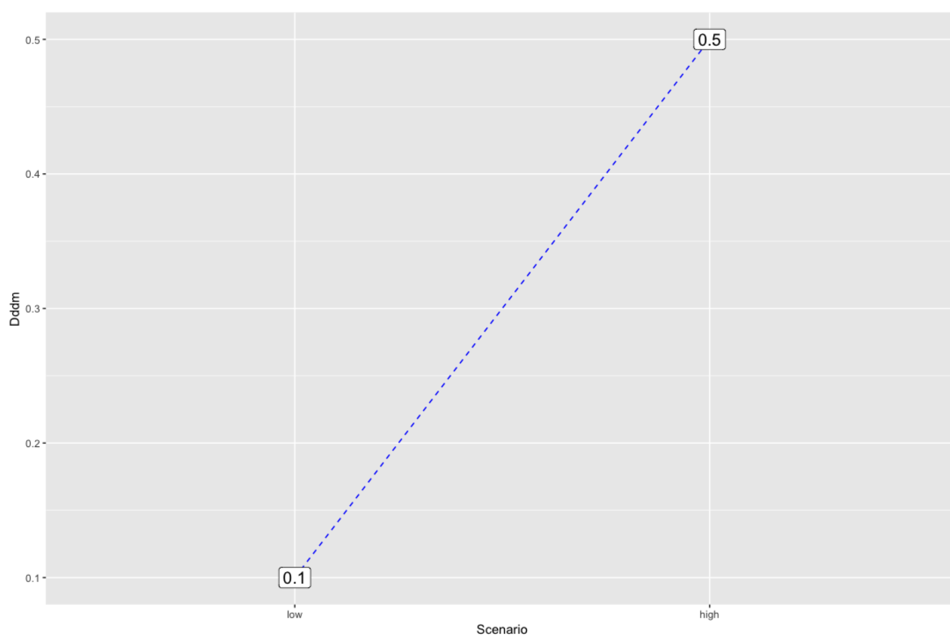

Figure 84 – Difficulty D_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance.

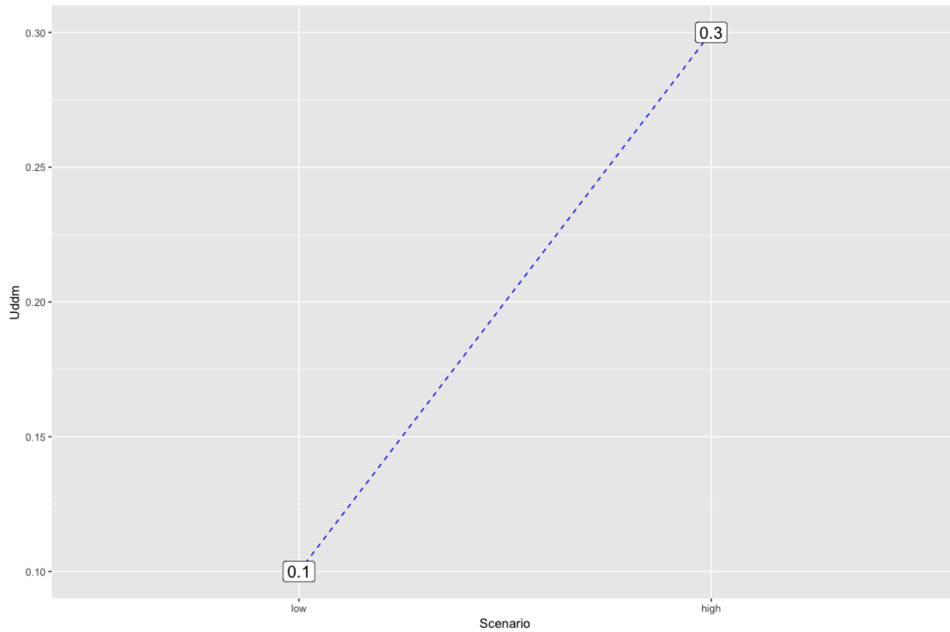

Figure 85 – Uncertainty U_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance.

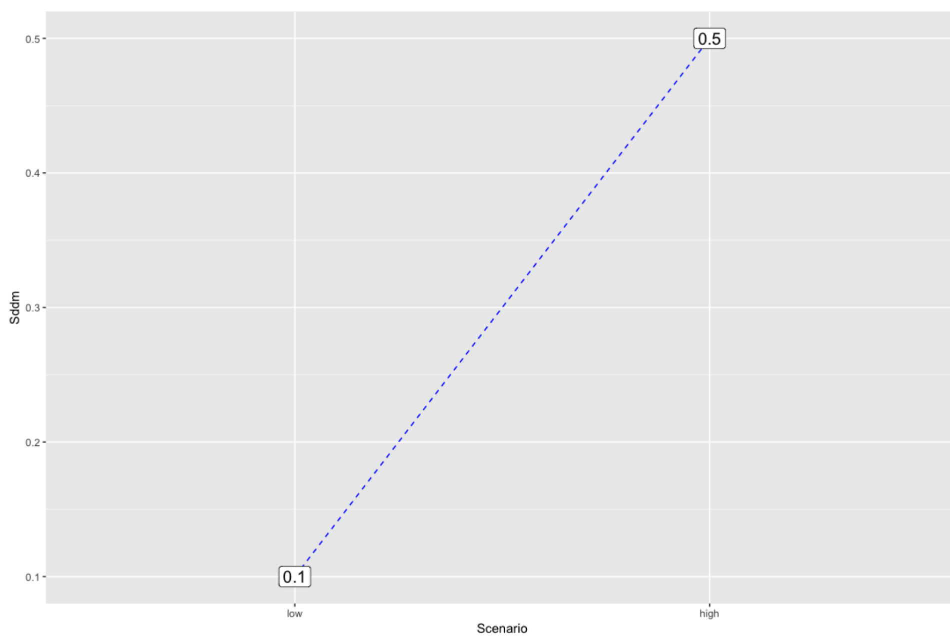

Figure 86 – Instability S_{DDM} ratings for future modeling and simulation efforts to relate the objective parameters of complexity of DDM problems to human performance.

Appendix C

Details for the Regression Models created through Machine Learning

This appendix presents the exhaustive search for the best fit multiple linear regression models based on the six subsets of parameters of complexity, conducted with the help of the `leaps` and `glmulti` packages in `R`. An exploration of the model-averaged variable importance for the parameters of complexity is presented for the exhaustive search phase, as well as a bootstrap validation [1] of the relative importance of the regression model variables for the selected candidate models.

1. Based on 1000 samples, using the *LMG* method, where the $R2$ contribution is averaged over orderings among regressors.

C.1 Exhaustive search with `leaps` using the adjusted R^2 selection criterion

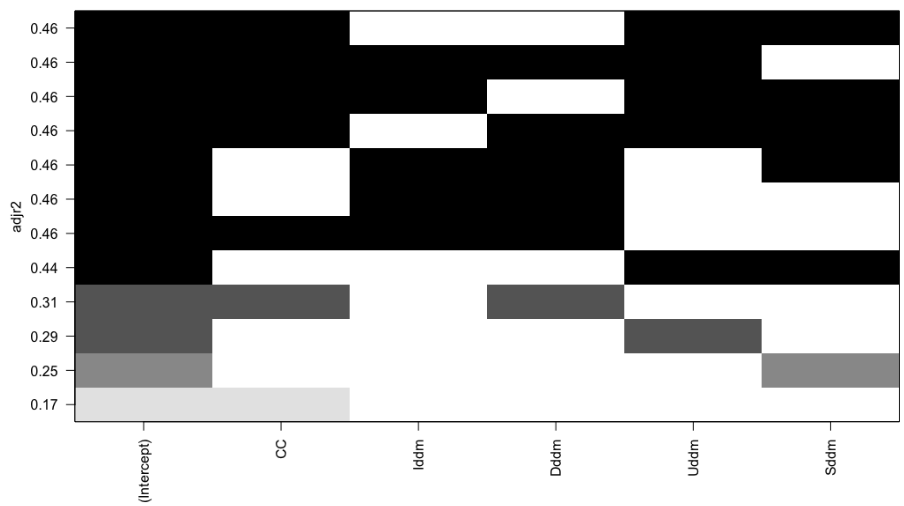

Figure 87 – Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

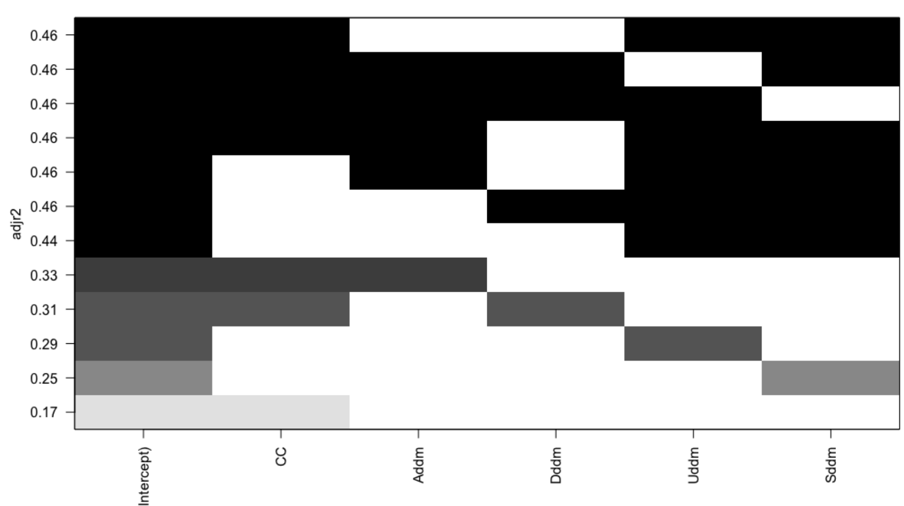

Figure 88 – Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

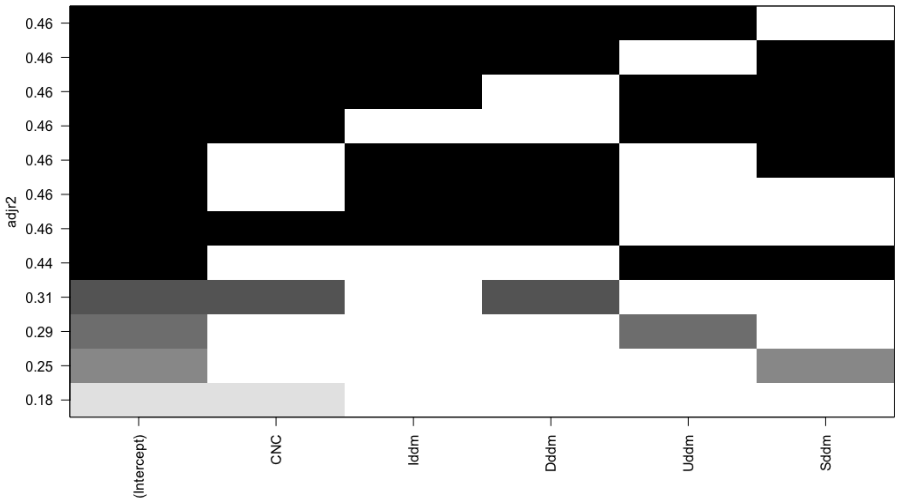

Figure 89 – Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

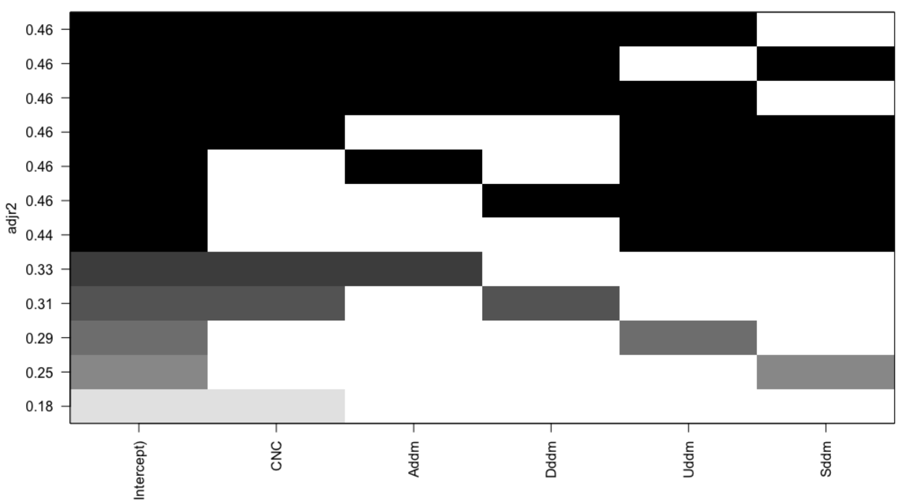

Figure 90 – Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

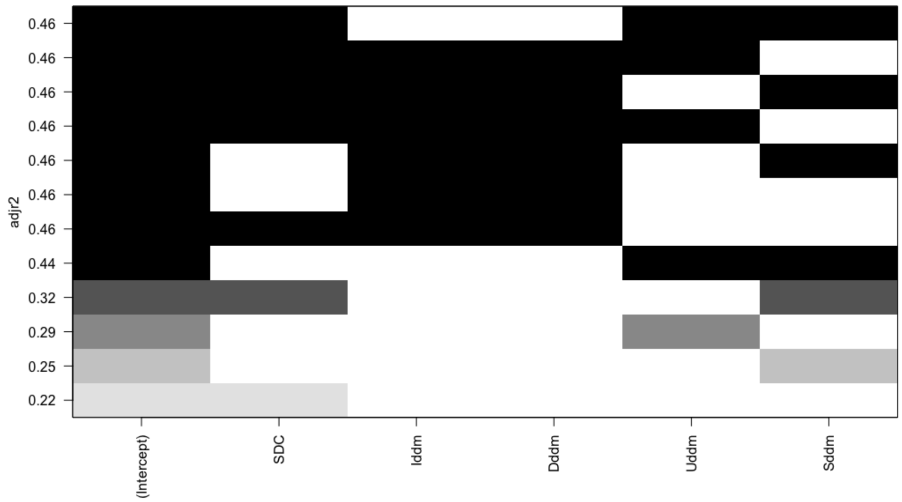

Figure 91 – Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

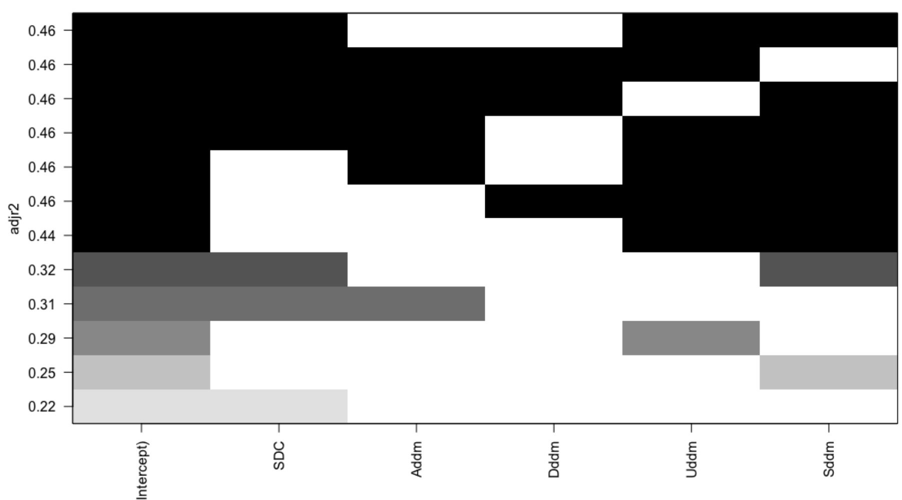

Figure 92 – Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

C.2 Exhaustive search with `glmulti` using the $AICc$ selection criterion

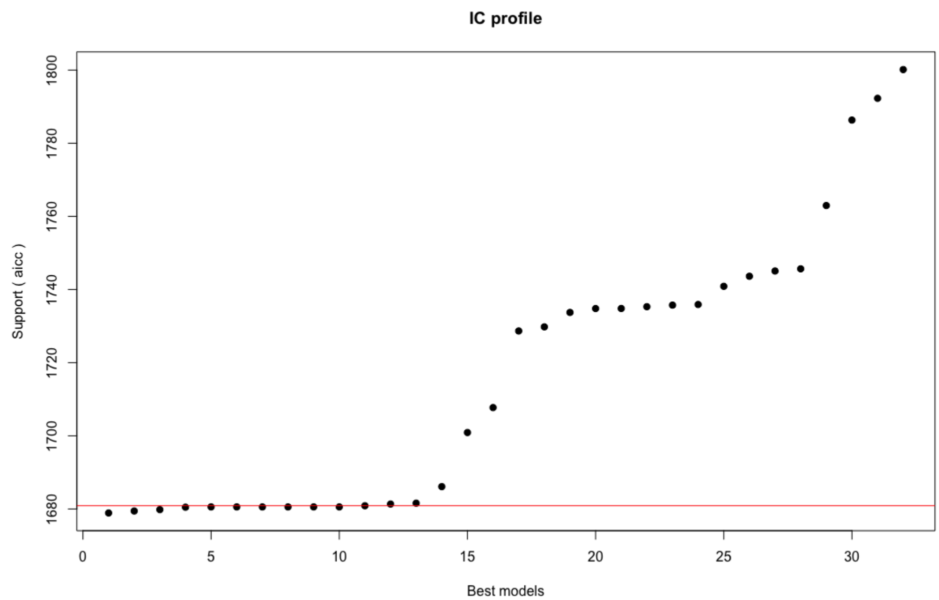

Figure 93 – Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

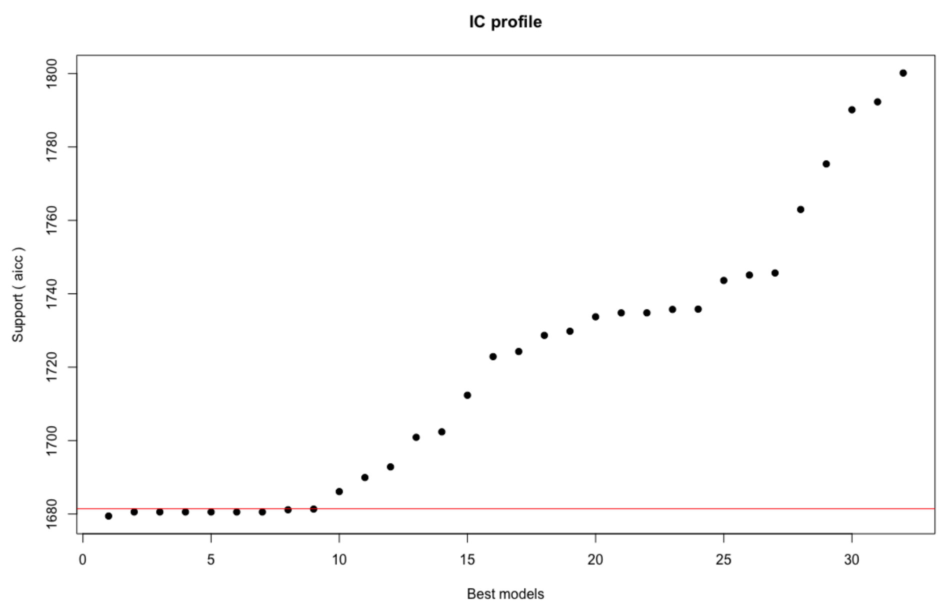

Figure 94 – Exhaustive search for a multiple linear regression model using CC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

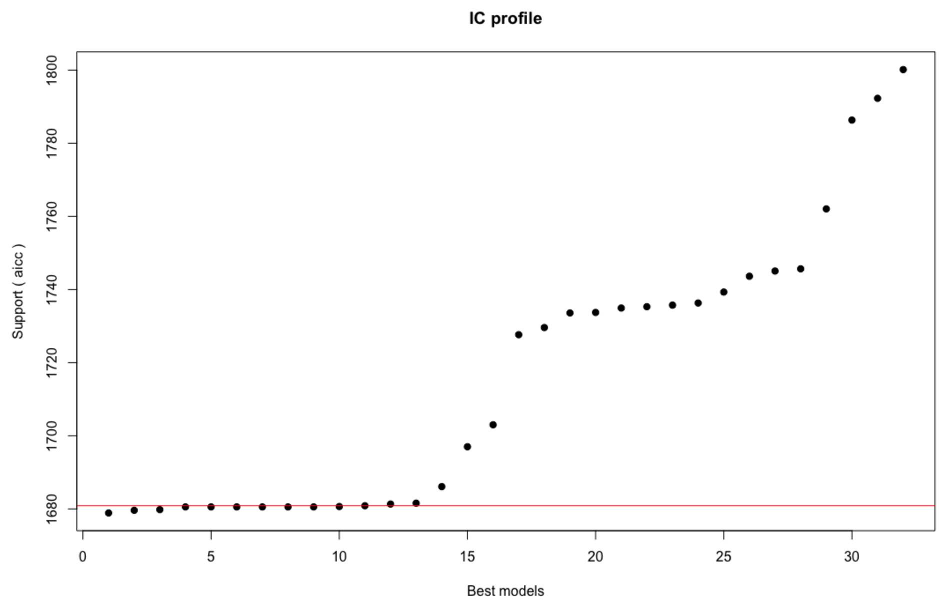

Figure 95 – Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

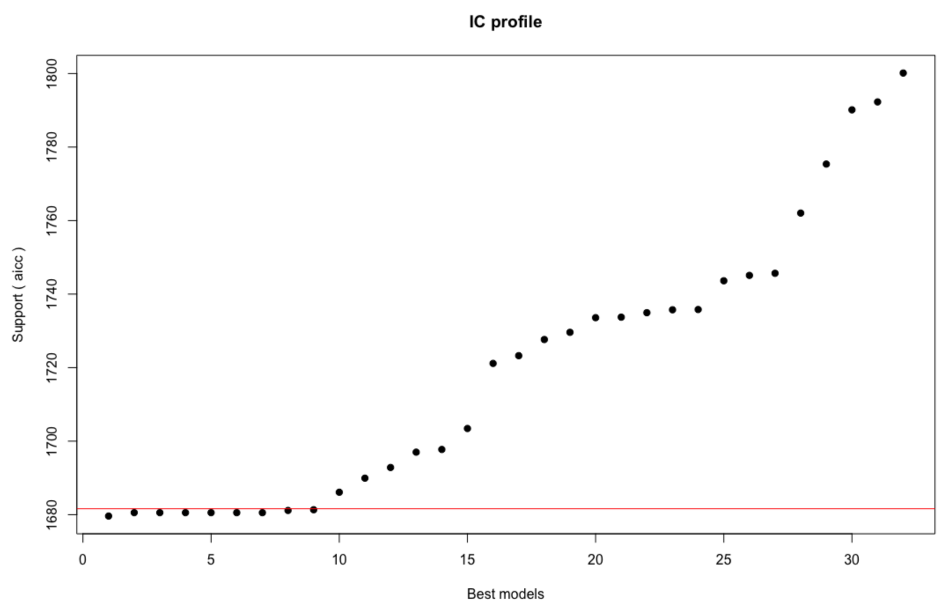

Figure 96 – Exhaustive search for a multiple linear regression model using CNC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

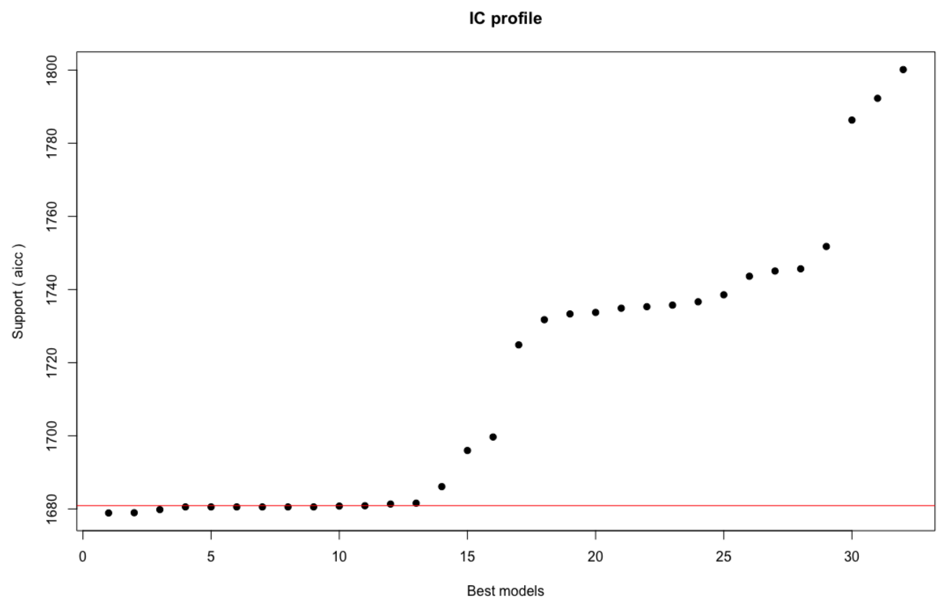

Figure 97 – Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

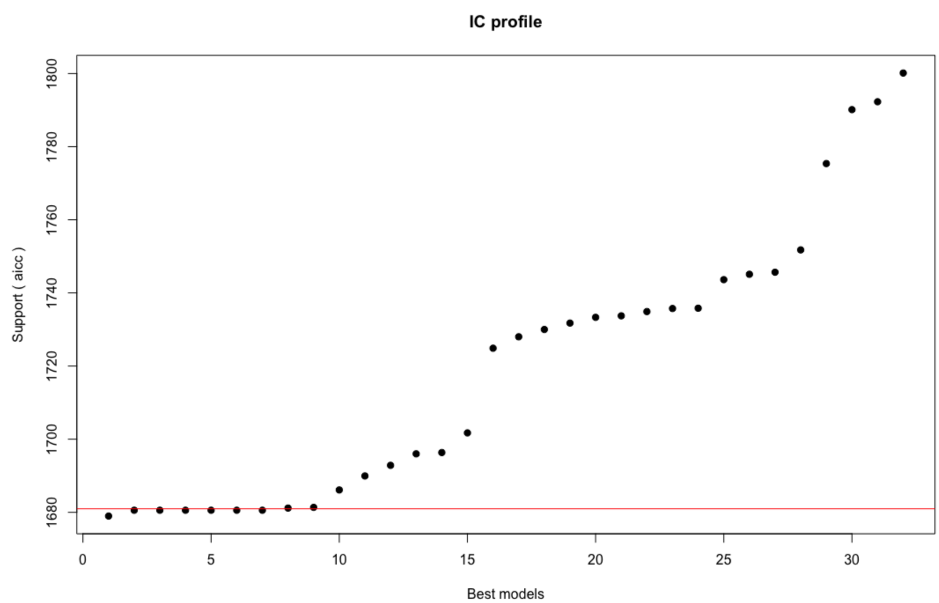

Figure 98 – Exhaustive search for a multiple linear regression model using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

C.3 Model-averaged importance of the parameters of complexity selected through the exhaustive search

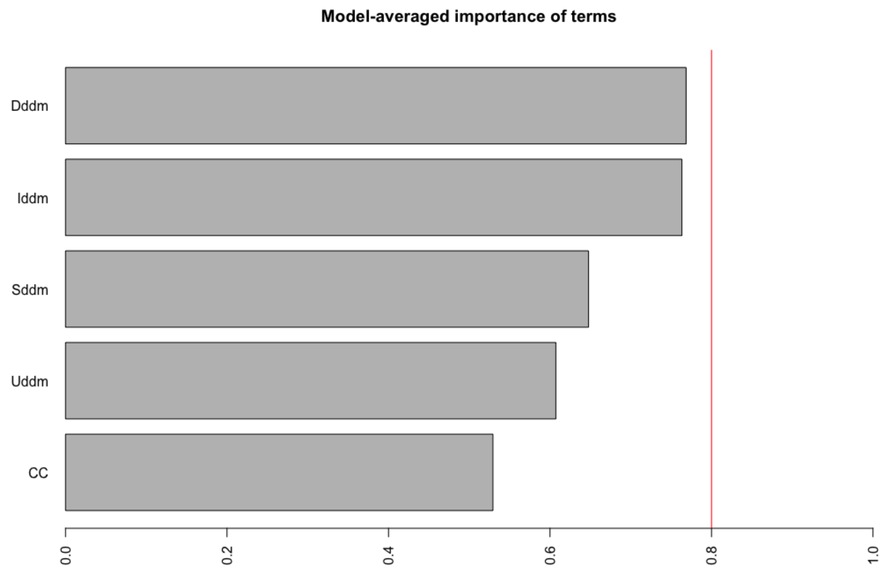

Figure 99 – Model-averaged variable importance using CC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

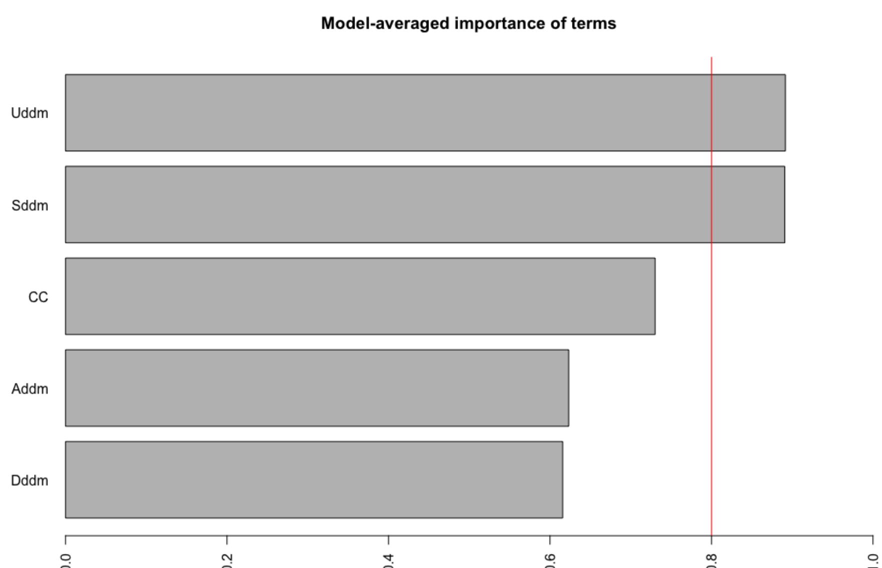

Figure 100 – Model-averaged variable importance using CC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

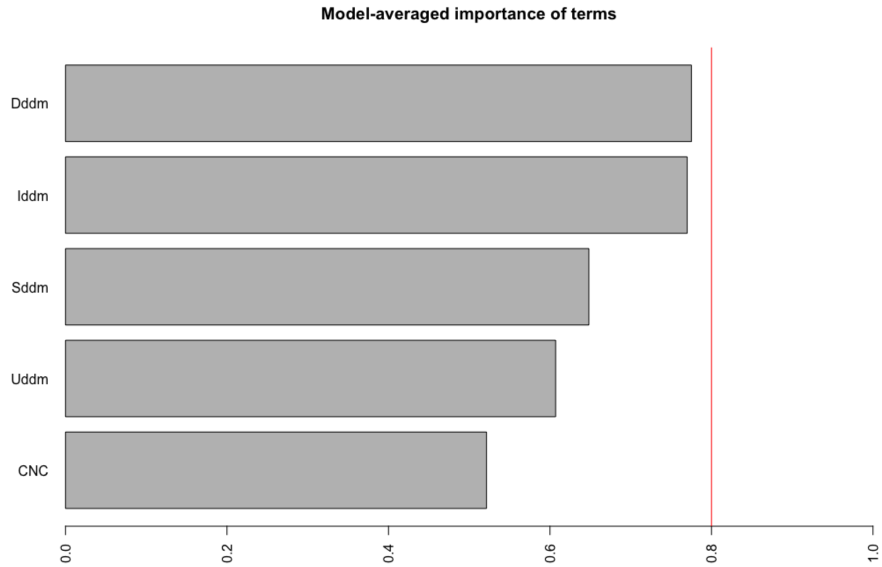

Figure 101 – Model-averaged variable importance using CNC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

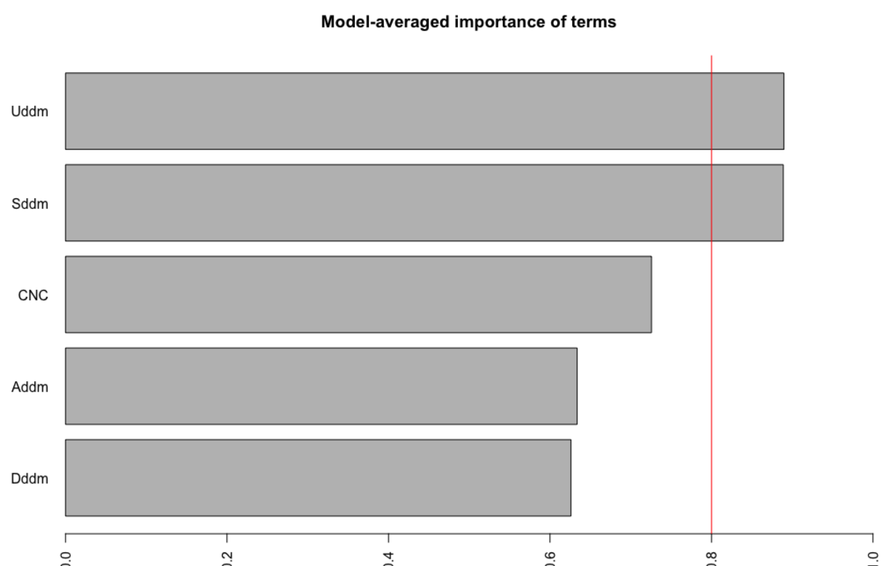

Figure 102 – Model-averaged variable importance using CNC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

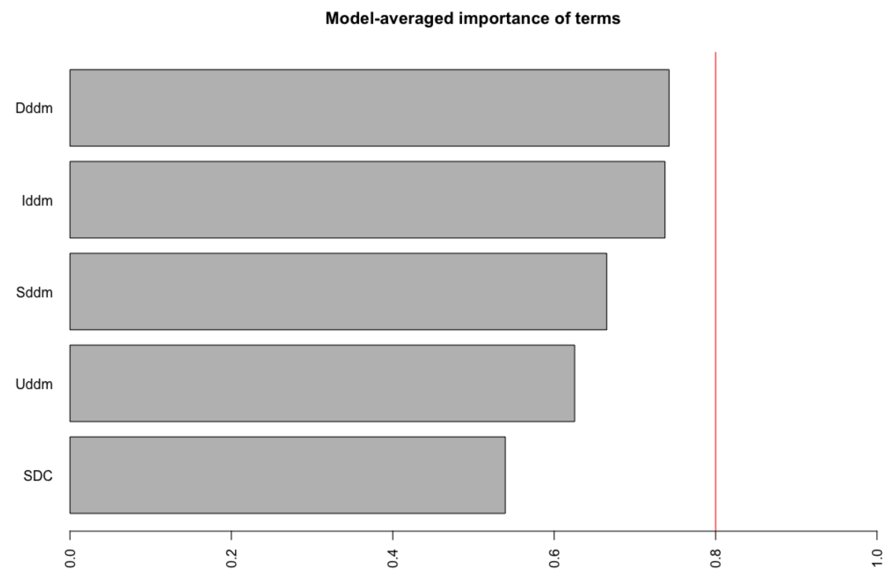

Figure 103 – Model-averaged variable importance using SDC_{DDM} as the structural complexity parameter and the information complexity I_{DDM}.

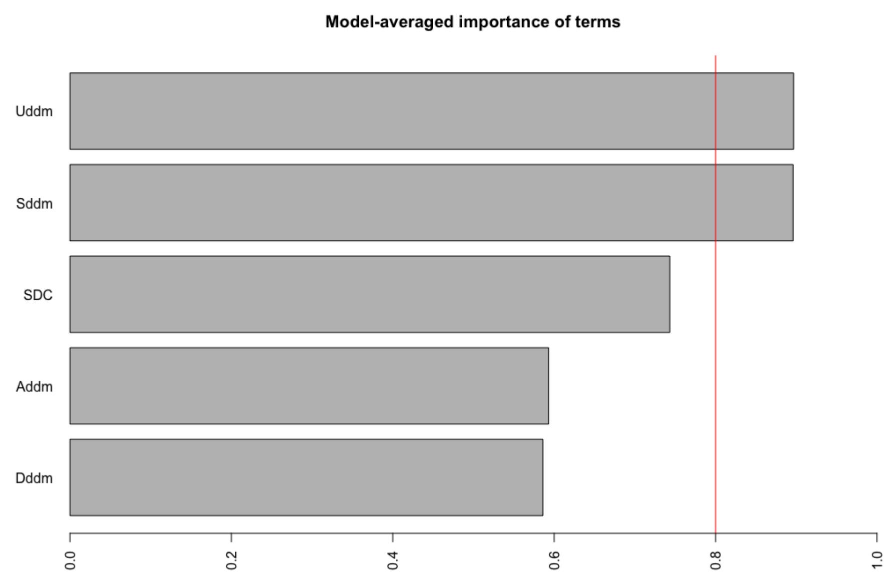

Figure 104 – Model-averaged variable importance using SDC_{DDM} as the structural complexity parameter and the action complexity A_{DDM}.

C.4 QQ plots of the residuals against the fitted values for the candidate MLR models used for regression diagnostics

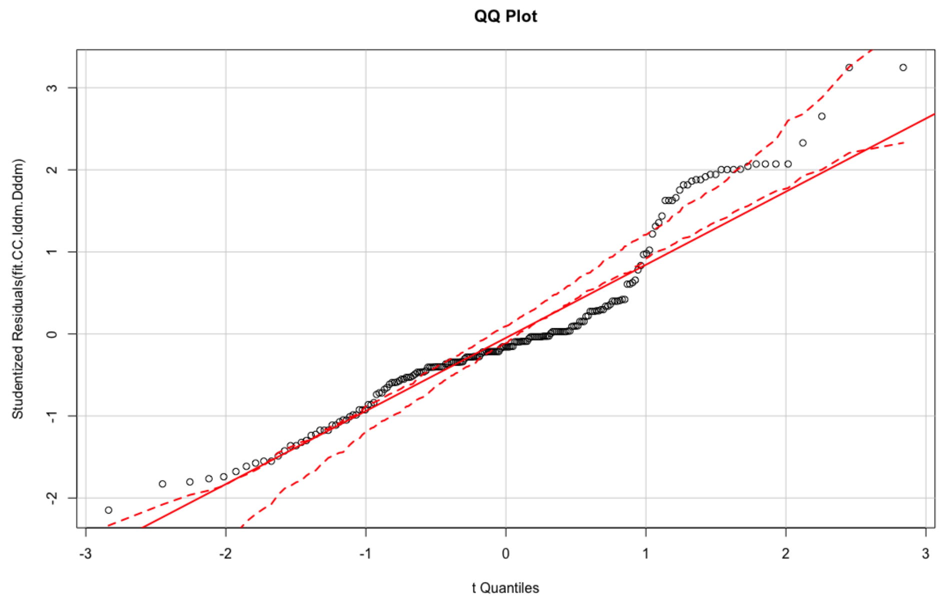

Figure 105 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CC_{DDM}, I_{DDM}, D_{DDM} parameters.

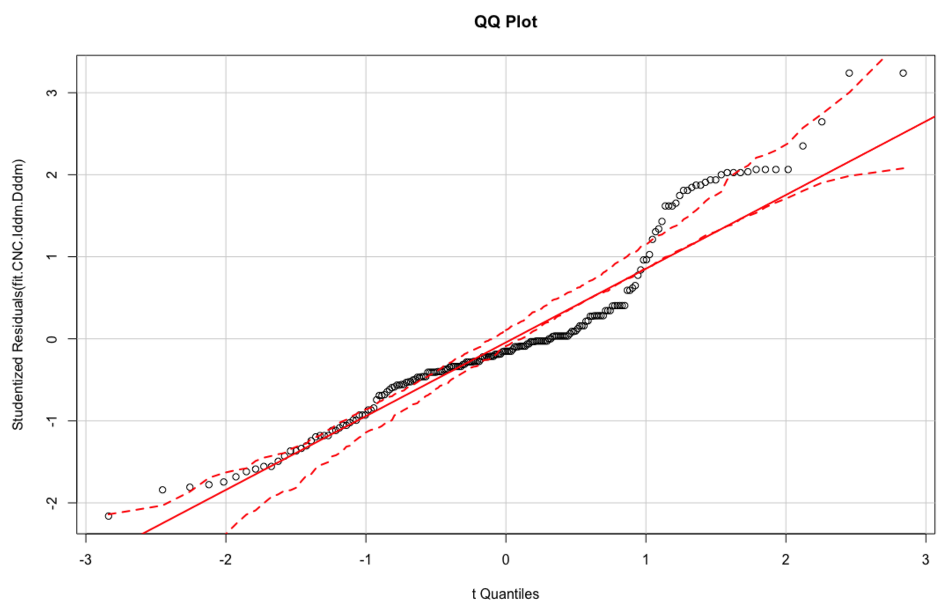

Figure 106 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CNC_{DDM}, I_{DDM}, D_{DDM} parameters.

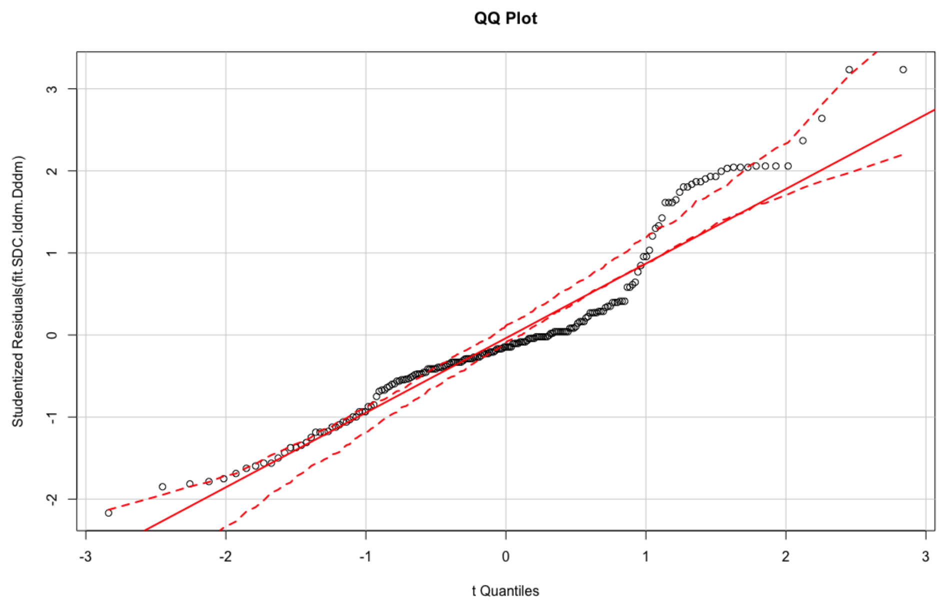

Figure 107 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters.

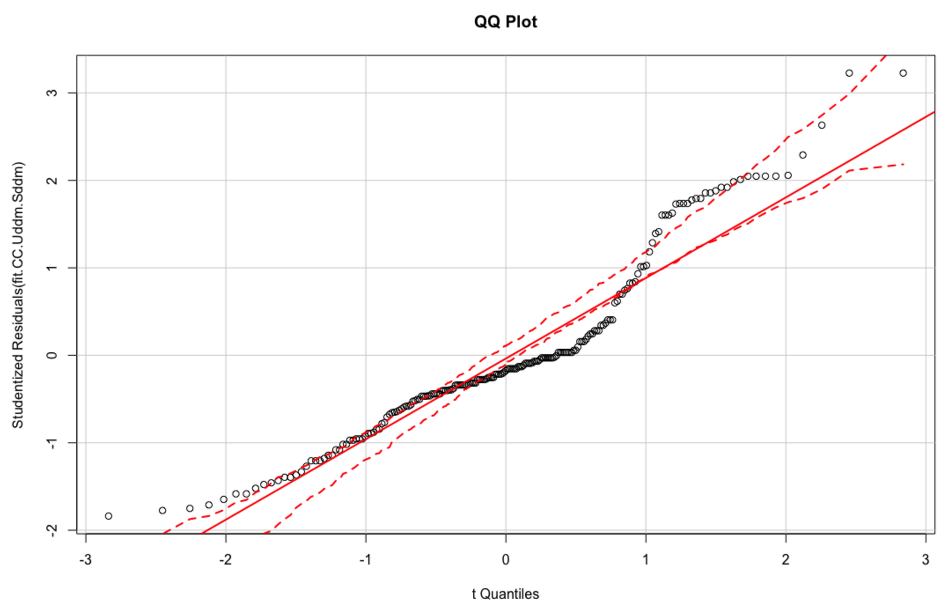

Figure 108 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CC_{DDM}, U_{DDM}, S_{DDM} parameters.

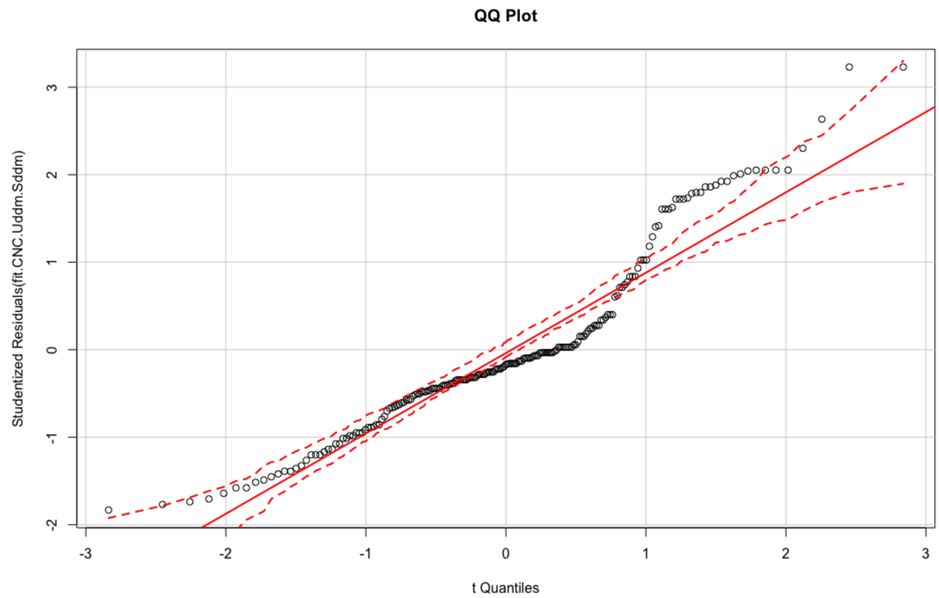

Figure 109 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the CNC_{DDM}, I_{DDM}, S_{DDM} parameters.

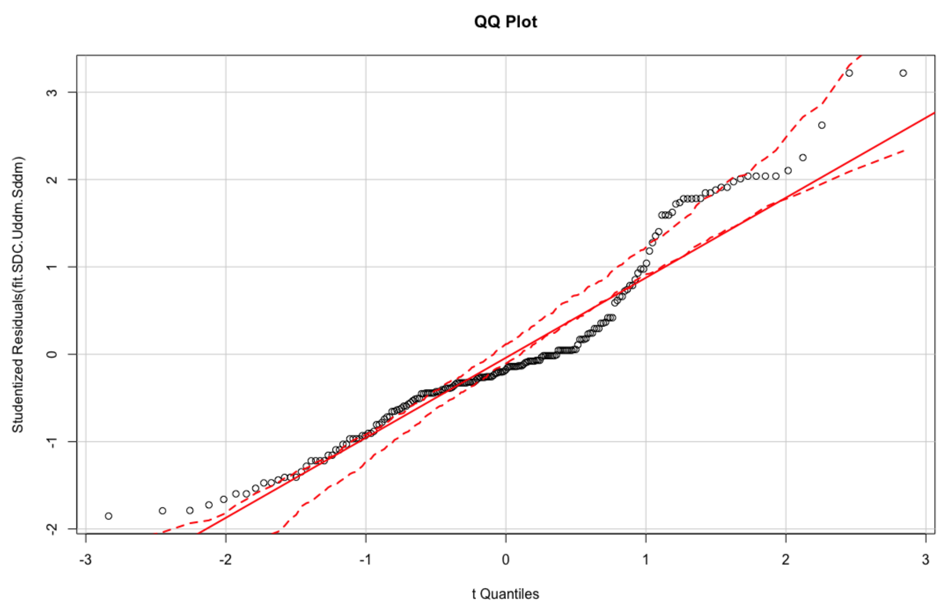

Figure 110 – A quantile-quantile (q-q) plot of the residuals against the fitted values for the MLR model using the SDC_{DDM}, U_{DDM}, S_{DDM} parameters.

C.5 Graphs of Cook's distance for the candidate MLR models used for regression diagnostics

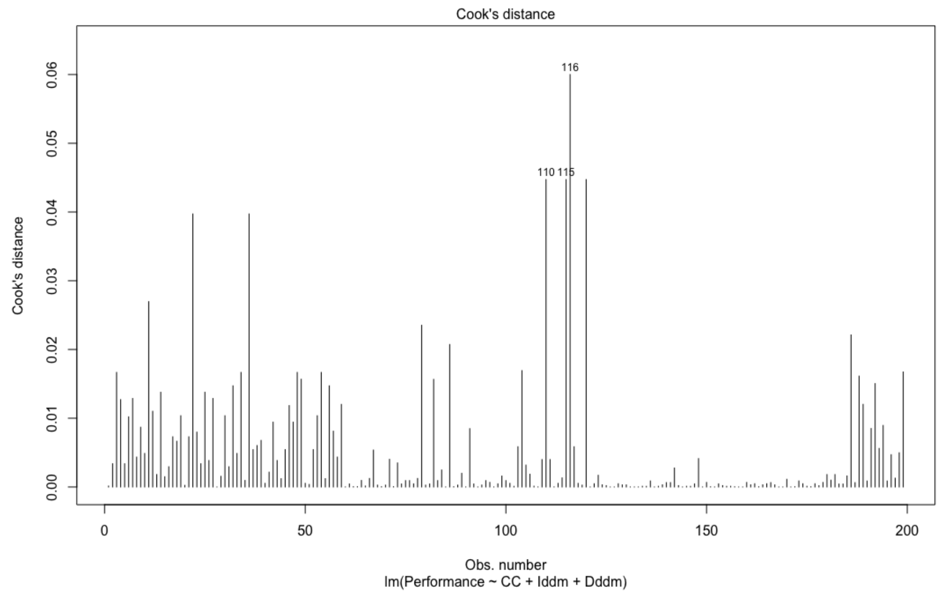

Figure 111 – Cook's distance D plot for the MLR model using the CC_{DDM}, I_{DDM}, D_{DDM} parameters.

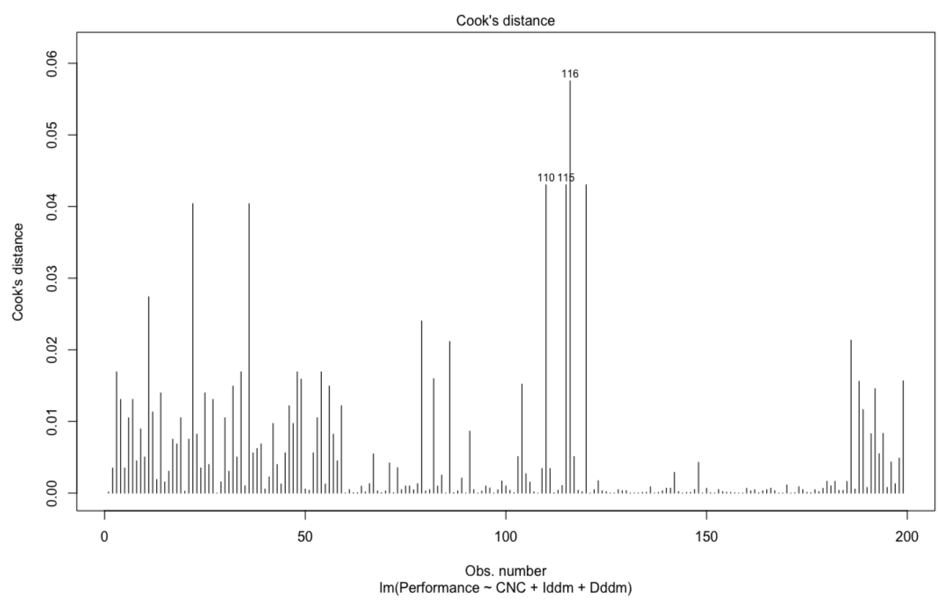

Figure 112 – Cook's distance D plot for the MLR model using the CNC_{DDM}, I_{DDM}, D_{DDM} parameters.

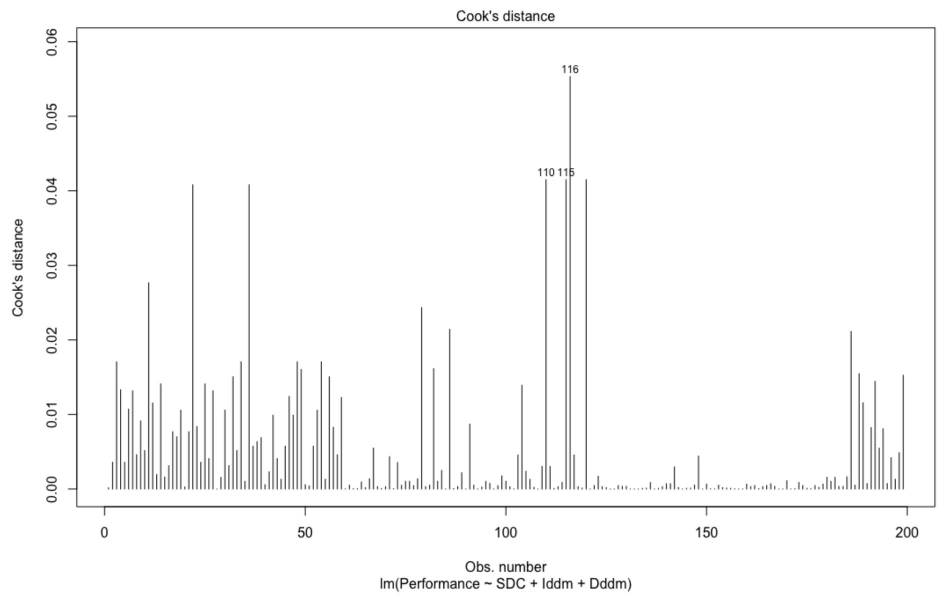

Figure 113 – Cook's distance D plot for the MLR model using the SDC_{DDM}, I_{DDM}, D_{DDM} parameters.

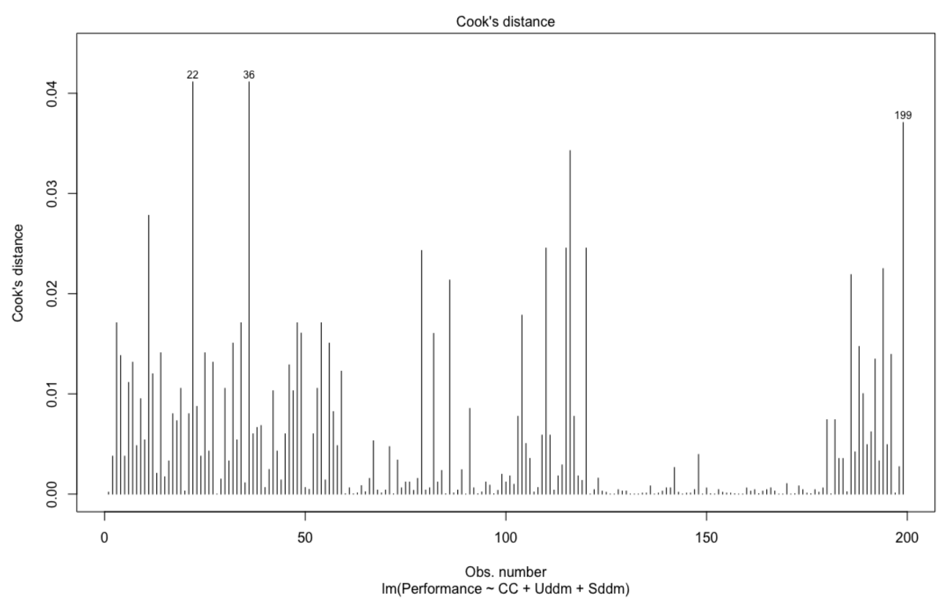

Figure 114 – Cook's distance D plot for the MLR model using the CC_{DDM}, U_{DDM}, S_{DDM} parameters.

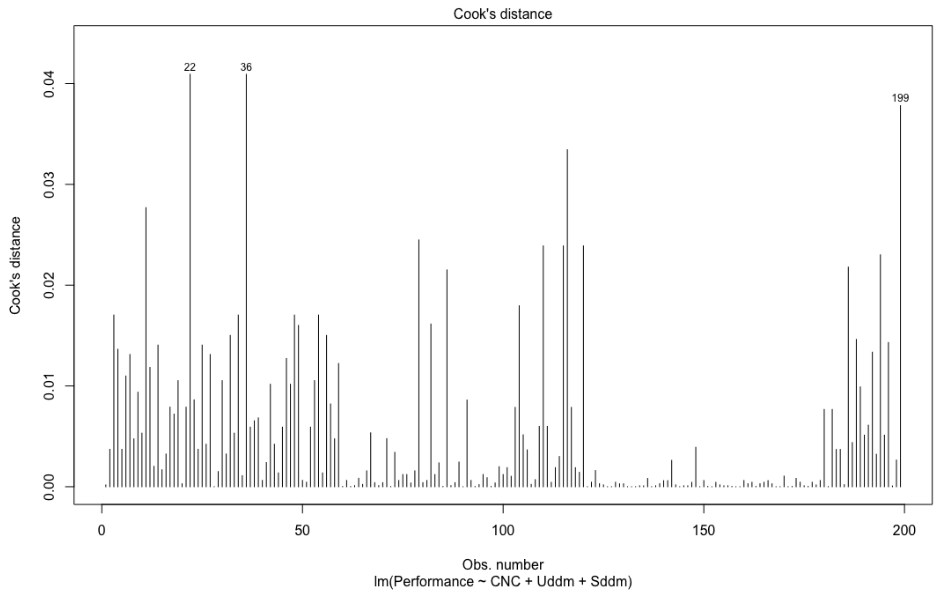

Figure 115 – Cook's distance D plot for the MLR model using the CNC_{DDM}, U_{DDM}, S_{DDM} parameters.

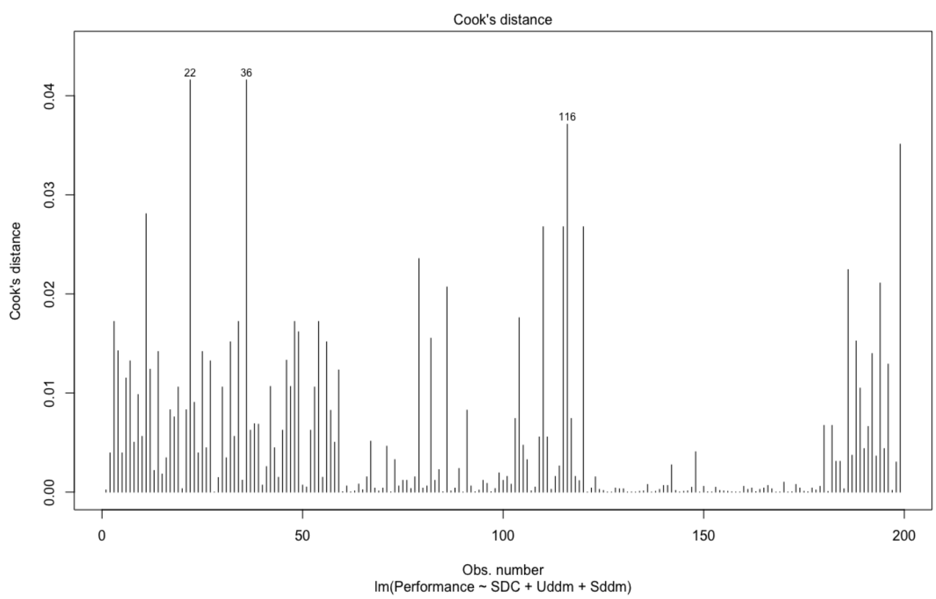

Figure 116 – Cook's distance D plot for the MLR model using the SDC_{DDM}, U_{DDM}, S_{DDM} parameters.

C.6 Relative importance of the regression model variables for the selected candidate models using 1000-sample bootstrap validation

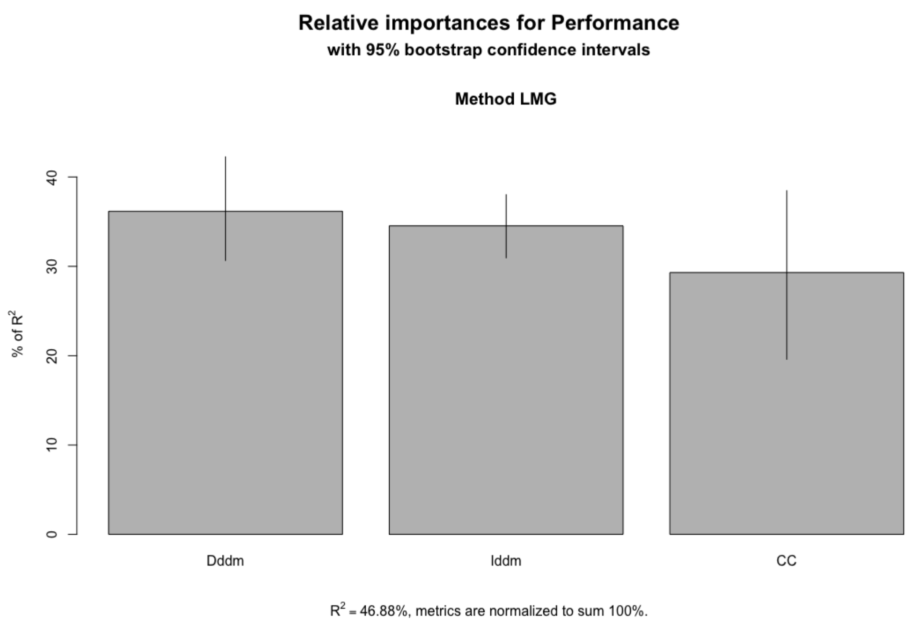

Figure 117 – Relative importance of the model variables for CC_{DDM}, I_{DDM}, D_{DDM}.

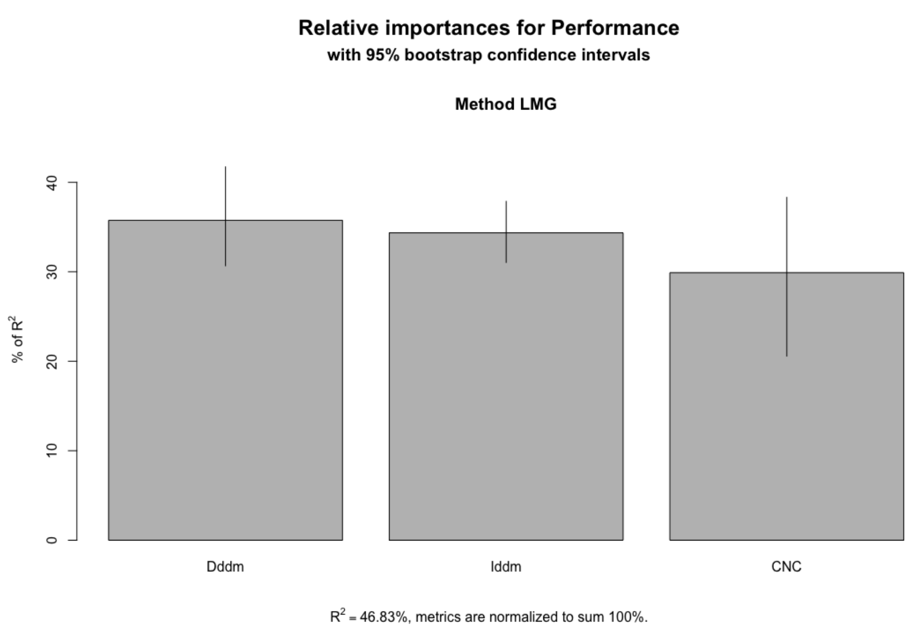

Figure 118 – Relative importance of the model variables for CNC_{DDM}, I_{DDM}, D_{DDM}.

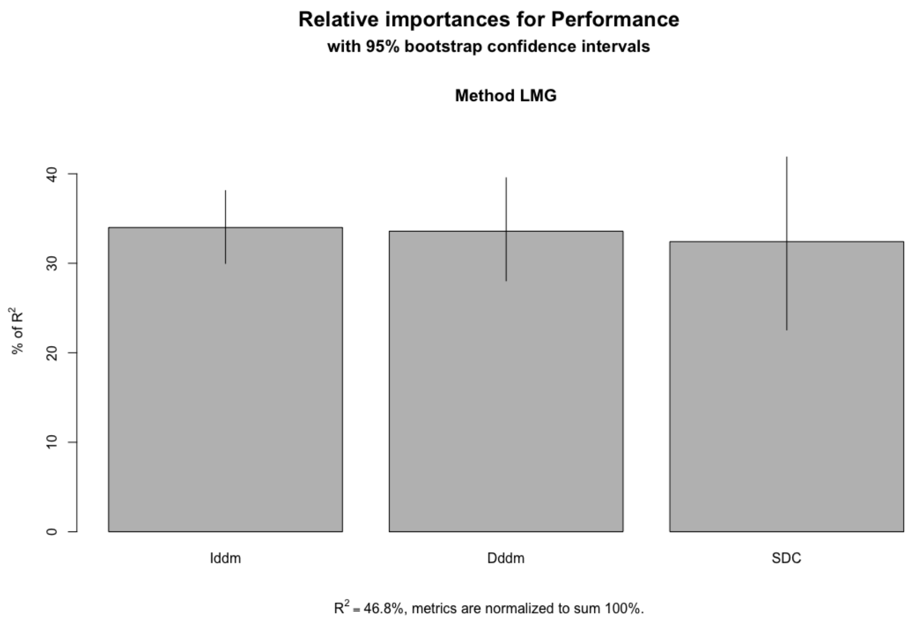

Figure 119 – Relative importance of the model variables for SDC_{DDM}, I_{DDM}, D_{DDM}.

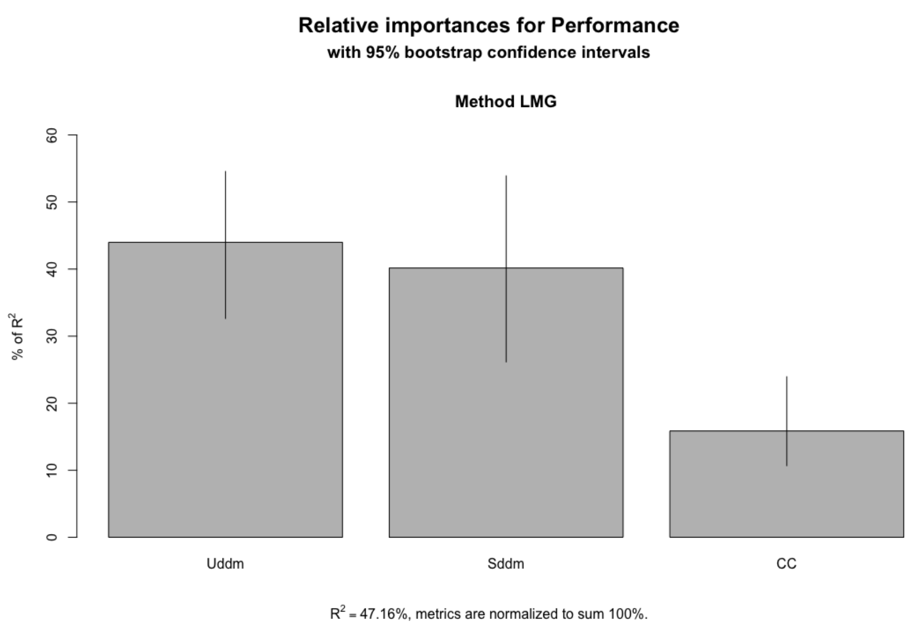

Figure 120 – Relative importance of the model variables for CC_{DDM}, U_{DDM}, S_{DDM}.

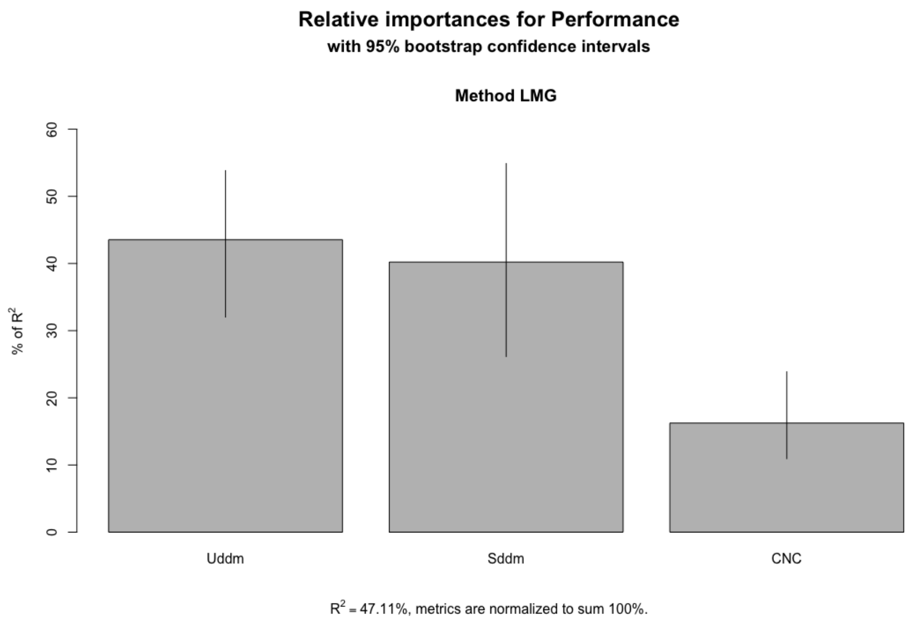

Figure 121 – Relative importance of the model variables for CNC_{DDM}, U_{DDM}, S_{DDM}.

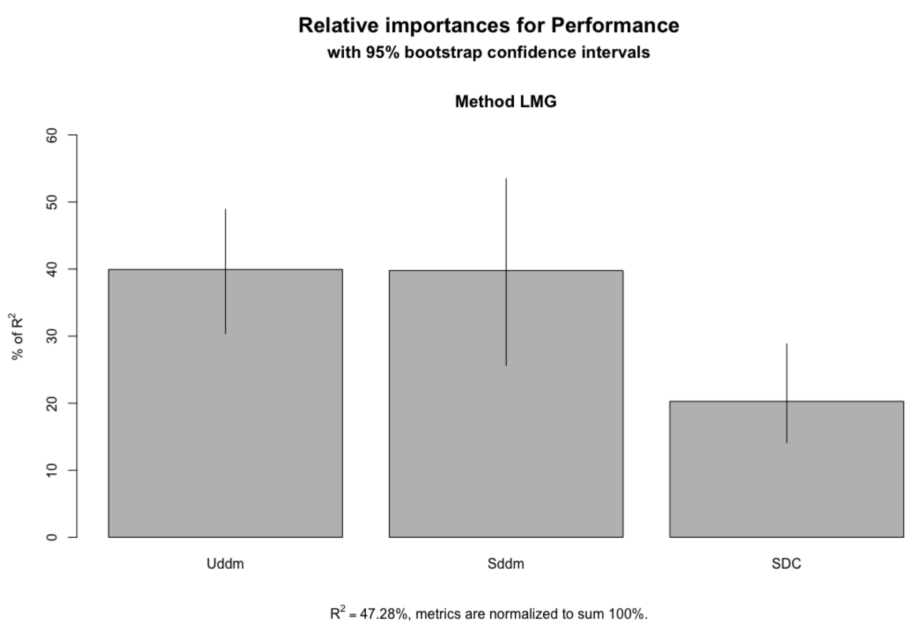

Figure 122 – Relative importance of the model variables for SDC_{DDM}, U_{DDM}, S_{DDM}.