Spring 2015

Toward a Theory of Multi-Method Modeling and Simulation Approach

Mariusz A. Balaban
Old Dominion University

TOWARD A THEORY OF MULTI-METHOD MODELING AND SIMULATION APPROACH

by

Mariusz A. Balaban
B.S. June 2001, Lublin Technical University
M.E. June 2001, Lublin Technical University
M.E. June 2003, Lublin Technical University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MODELING AND SIMULATION

OLD DOMINION UNIVERSITY
May 2015

Approved by:

Patrick Hester (Director)

Catherine M. Banks (Member)

er)

Saikou Diallo (Member)

Saikou Diallo (Member)

ABSTRACT

TOWARD A THEORY OF MULTI-METHOD MODELING AND SIMULATION APPROACH

Mariusz A. Balaban
Old Dominion University, 2015
Director: Dr. Patrick Hester

The representation via simulation models can easily lead to simulation models too simple for their intended purpose, or with too much detail, making them hard to understand. This problem is related to limitations of the modeling and simulation methods. A multi-method Modeling and Simulation (M&S) approach has the potential for improved representation by taking advantage of methods' strengths and mitigating their weaknesses. Despite a high appeal for using multiple M&S methods, several related problems should be addressed first. The current level of theoretical, methodological, and pragmatic knowledge related to a multi-method M&S approach is limited. It is problematic that there is no clearly identified purpose and definition of the multi-method M&S approach. Theoretical and methodological advances are vital to enhancing the application of a multi-method M&S approach to address a broader range of scientific inquiries, improve quality of research, and enable finding common ground between scientific domains. This dissertation explored theoretical principles and research guidelines of a multi-method M&S approach.

The analyzed literature offered perspectives related to the purpose, terms, and research guidelines of a multi-method M&S approach. A pragmatic philosophical stance was used to provide the basis for the choice of terms and definitions relevant to a multi-method M&S approach were proposed. The degrees of falsifiability are adapted to the

M&S domain, which allowed for developing complementarity principles as the theoretical basis of a multi-method M&S approach. Next, a blueprint of a multi-method M&S approach called method formats was derived, because transitions toward formats must seek justifications in order to increase research objectivity and transparency.

A sample set of methods was explored in the context of a proposed sample set of criteria. None of the methods were evaluated with the maximum score for every criterion, which implied that if all those characteristics were required within a research context, then, none of the methods could provide the highest possible score without combining methods. Finally, a case study that included a multi-method simulation model was developed, providing a data layer for evaluation of complementarity principles. The case study contributed to the credibility of complementarity principles as a reason to use a multi-method M&S approach and value of pseudo-triangulation as a mean of verification of a selected approach.

With gratefulness, for my beloved wife

ACKNOWLEDGMENTS

There are many whose assistance and encouragement contributed greatly to the completion of this work. First and foremost, I offer my sincerest gratitude to my supervisor Dr. Patrick Hester, and committee members Dr. Saikou Diallo, Dr. Catherine Banks, and Dr. John Sokolowski for their guidance throughout the development of my thesis. The valuable suggestions offered by Dr. Jose Padilla and Dr. Andreas Tolk are deeply appreciated. Many thanks to Dr. Rick McKenzie, Dr. Yuzhong Shen and all MSVE department faculty members for help and inspiration. Moreover, special thanks to Mr. Trey Mayo for all the guidance, insight, and impeccable academic advisory.

I would like to thank my friends and colleagues, including my fellow students in the MSVE department for sharing the journey during good and better moments. Special thanks to Dr. Krzysztof Rechowicz for all his assistance, help and encouragement. I would like to extend my gratitude to my employers and co-workers from MYMIC for a friendly work-environment and support, and for giving me flexibility in work schedule.

I also need to mention someone who will never read this, but whose joyful face, loving eyes, and happily wiggling tale was there, always by my side, under the desk, keeping up with my demanding schedule and (not always) patiently waiting for some cheering puppy-time and refreshing walks. Thanks to my parents-in-law for bright times of laughter that kept my spirit up. Finally, the biggest acclamation goes to my wife, Dorota, for her patience, especially during listening to my countless and endless monologues. The love and care, support and help, inspiration and encouragement I experienced from her was the key to the completion of this dissertation.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Macal and North [1] referred to the use of the computer simulation as a third way of doing science in addition to deductive and inductive reasoning. Constructive simulations, as a new way of conducting science, could be characterized as inductive and deductive at different stages of a study. The creation of a constructive, virtual world with often deductive rules follows inductive analysis of output data or patterns, which in turn might lead to insight into consequences of assumptions of studied phenomenon [2], leading to the question of whether the deductively created virtual world is an adequate representation to produce valid information for further inductive analysis of phenomena. Unfortunately, there is no easy answer to this question at this point[1].

The need to use combined discrete event and continuous simulation was introduced by Fahrland [3] more than four decades ago. He suggested application of both discrete and continuous methods to model different parts of systems. For instance, in batch-processing chemical plants, discrete process could help investigate policies that pertain to scheduling, inventory and resource use, while continuous view of chemical reactions could describe mass balance. In automobiles, traffic queuing and driver decisions would be discrete while vehicle dynamics remain continuous. In neuro-muscular systems, task sequencing and impulses would be discrete while muscle mechanics and biochemical reactions remain continuous. The approach with multiple methods has gained momentum, already representing various phenomena in manufacturing [4]; healthcare [5, 6]; and supply chain systems [7]. Discrete Event

[1] IEEE Transactions and Journals style is used in this dissertation for formatting figures, tables, and references.

Simulation (DES) and System Dynamics (SD) methods often complement each other. For instance, DES offers a better representation of detail complexity, and SD allows for easier representation of dynamic "feedback" effects [8, 9].

The working definition of a multi-method M&S approach is offered based on Balaban and Hester [10] as a combination of at least two M&S methods that combined allow for a unique system or phenomena representation and execution. Mingers points at two main reasons for using a multi-method approach: "It is both desirable and feasible to combine together different research methods to gain richer and more reliable research results." [11] He refers to the principle of complementarity in which "no one paradigm is superior, but that their individual rationalities should be respected within the discipline as a whole."[11]

During the development of a simulation model, a modeler mostly operates on abstraction and refinement processes, which can lead to a model that lacks the required fidelity by building it too simple for the purpose. The opposite could also be true when the representation carries too much detail, making it hard to understand. Sylvan and Voss discussed the relationship between the quality of a problem representation and the quality of a solution that was summarized as follows: "...more specific representations led to more specific solutions. Indeed, in general, the quality of the solution was a function of the quality of representation." [12] This finding contributes to the discussion on how much detail is enough to understand phenomena studied. The context given by Sylvan and Voss could be interpreted contrary to what M&S practitioners often claim as a general modeling rule: 'KISS (Keep it Simple, Stupid)'. Despite many advantages of simulation, the scientific community faces problems of phenomena representation where the 'KISS' approach often does not work. Schwandt [13] conveyed the problem of representation with the declaration that no interpretative account could ever properly, directly or completely, capture lived experience. As a realization of the need for a more descriptive approach, Edmonds and Moss [14] proposed a new approach under the saying "Keep it Descriptive, Stupid (KIDS)". This helped to limit oversimplification to simulation practices overall.

The presumed or trivial representation of a phenomenon that does not cover important aspects of the underlying phenomena can lead to a solution, insight, or decisions that are inaccurate and miss important alternatives. One possible cause of oversimplifications is related to limitations of used modeling and simulation methods. More flexibility and creativity to represent various phenomena using an M&S approach seems desirable. On the other hand, human capacities to comprehend and computer power limitations can inhibit the usefulness of complex simulation models [15], and this is very much applicable to complex multi-method simulation models.

1.1 PROBLEM

Presently, the practice of combining methods has become more popular as more simulation tools offer capabilities beyond the original idea of combining continuous and discrete parts. Different M&S methods can contribute their advantages, forms of expressiveness, and different perspectives on capturing complexity of phenomena. For instance, SD seems more suitable for capturing dynamic complexity [16], Agent Based Modeling (ABM) seems more suitable for representing complexity arising from individual behavior and interactions [1], and DES can well capture "black box" process complexity [17]. Bayesian Networks (BN) offer a unique probabilistic view, where posterior probabilities can measure the degree of belief based on evidence and can be used to represent e.g. beliefs of agents [18, 19]. Fuzzy Method (FM) allow for capturing vagueness of phenomena systematically [20] and can be useful in social simulations [21]. Triangulation or use of multiple methods within a single or multiple simulation models could be valuable. The outcome of a study based on a simulation model can indicate the value of an approach chosen, for instance, the level of gained understanding for investigated phenomena. Tashakkori and Teddlie argue, "...there is an iterative process between considering the research purpose and the research question. Out of this iterative process come decisions about methods. We make the case that when the purpose is complex (as it often is), it is necessary to have multiple questions, and this frequently necessitates the use of mixed methods." [22] If possible, projection of this argument into the multi-method simulation based research could empower and encourage the use of this approach. Swinerd and McNaught acknowledged that it may be challenging to employ a single method approach to represent complex, modern systems, and that the use of multiple methods "could provide a simpler, more natural or more efficient solution." [23] They have emphasized coupling between different scales of a system and representing cross-scale dynamics as a potential value added, but noticed a need for more research.

Despite a high appeal to using multiple M&S methods to represent various phenomena, it is problematic that possible reasons and justifications have not been thoroughly explored to provide a solid theoretical base. Because the use of multiple methods seems intuitively more difficult, the trade-offs would be systematically deliberated. The concept of complementarity of methods originated from the complementarity theory postulated by Bohr [24]. In an M&S context, it is often given in the context of justification for the use of more than a single method. This is observed in M&S [9, 25] and close to M&S for instance information systems (IS) and management sciences [11, 26], but also in more distant empirical social sciences [27]. The idea of the complementarity of methods pertains to taking advantage of methods strengths and mitigation of their weaknesses. The question arises as to if and how complementarity could provide a general overarching reasoning for the use of more than a single method.

Different terms, definitions, and knowledge exist within branches of the multidisciplinary M&S field, which may be due to a variety of M&S methods more or less applicable within different domains [28]. M&S is a fast-growing discipline, and it may take time to clarify, refine and categorize all terms. Different terms are used e.g. method, paradigm, technique, formalism or methodology to describe the DES, SD, ABM and other M&S approaches [8, 10, 29, 30]. Similar problems exist when terms are used to describe approach with more than a single method, e.g. multi-method, multi-methodology, multi-paradigm, hybrid, mixed-method, multi-model and multi-formalism [23, 31-33]. Sometimes a single term is used, sometimes multiple terms are used within a single piece of work as synonyms solely for readability purposes, and still other times, different meanings of those terms are intended to convey. In many cases, the purpose of using multiple terms is difficult to determine, which can lead to confusion and should be further explored and corrected. The lack of agreed-upon terms that may or may not mean the same thing can cause consistency problems and should be clearly understood to provide a useful, clear, and holistic terminology accepted by M&S community.

Chahal [8] proposed the reasoning for the use of hybrid SD/DES models in the context of healthcare setting. Unfortunately, this approach has limitations related to the scope of methods considered because only two methods were used. For this reason, the use of this framework to other methods, or for more than two methods is problematic. Chahal [8] disintegrated objectives in order to determine if both DES and SD are needed for a representation. If criteria for different objectives aligned with different methods, then multiple methods were used. Unfortunately, the ability to assign clear qualitative boundaries for criteria of a given objective may not be always possible for subjective phenomena. The criteria would be unable to expose methods' uniqueness in a particular context due to their limited precision.

Currently, methodological guidelines for multi-method M&S approach focus on methods considered, study problems, and system at consideration [8, 34], but are often method or domain (or both) specific [8, 23, 31, 35]. When a modeling framework prescribes a set of methods, it can lack flexibility and constrain conceptual modeling. Moreover, the problem should not be adjusted to the known methods, but handled by the most appropriate one [8]. On the other hand, in a realistic situation a modeler may not be even aware of, or familiar enough with the appropriate method(s). In this case, guidelines could only direct to the method(s) from a set of methods available within the software used and known, or those that could be learned within time frame available. Depending on the circumstances, a modeler could learn new methods, but must know which one should be used, which leads us back to the original point. Unless an updated knowledge base of all known M&S methods existed, there is always a possibility of choosing not appropriate method(s). To the best knowledge, no such a repository of knowledge currently exists. Even when assuming that all knowledge that pertains to methods was accumulated in the repository, should this enable full objectivity during the selection process? A general set of criteria may not provide sufficient threshold to decide which method is better in a given case. For instance, Glazner [35] used three methods, SD, DES, and ABM to represent different parts of the system. Glazner noticed that two out of three subsystems could be modeled using either of three methods. The decision, which method to use in each case, was a combination of the modeler preference and expected modeling effort. The only part that was directly leaning toward the use of ABM was "organizational unit", characterized by individual behavior, which could not be sufficiently represented using either SD or DES. This example indicates that in some cases, there is a gray area for choosing a method, but in other situations, there is a clear choice due to the requirements of the modeling effort.

A better understanding of subjectivity that influences method(s) choice is desirable. Multiple aspects, for instance, limited knowledge about methods, systems and phenomena, and lack of guidelines are probable factors that all tie to human subjectivity. The ultimate goal to eliminate subjectivity may not be achievable, but ability to limit and to communicate it using more holistic, transparent yet systemic, research guidelines would be beneficial. The general guidelines for a multi-method M&S approach should not prescribe methods within its guideline core. However, they should provide a balanced systemic process to determine satisfactory method(s) based on multiple elements e.g. research questions, merits of methods, modeler's knowledge of methods, and availability of software.

In summary, the current theoretical basis, and guidelines to conduct a multi-method study are limited. The lack of a theoretical basis to a multi-method M&S approach relates to a taxonomy, purpose and affects methodological guidelines mentioned above. It is expected, that exploration of complementarity of methods can contribute to a more sound theory of multi-method M&S approach and methodological guidelines. This is vital to enhance application of M&S to a broader range of scientific inquiries, improve quality of research, and enable finding common ground between scientific domains. In this dissertation, development of theoretical basis leading to methodological guidelines for the use of multiple M&S methods is pursued.

1.2 RESEARCH QUESTION

The research explores the theoretical basis and research guidelines for a multi-method M&S approach. The research question is: *What is the theoretical basis for a multi-method M&S approach*? The proposed answer presented in this work consists of:

- a set of relevant definitions,
- principles guiding multi-method M&S approach,
- general method formats, and
- multi-method M&S research guidelines.

The research method and approach for each of these elements are discussed in the following sections.

1.3 RESEARCH OBJECTIVES

This section identifies main research objectives.

Objective 1: Explore Purpose, Terms, and Methodological Aspects of Multi-Method M&S Approach

A literature review was conducted in order to examine components of a theoretical basis of a multi-method M&S approach. The scope consists of the purpose(s) for the use of multiple methods, relevant terminology, and methodological guidelines.

Objective 2: Propose Definitions for Multi-Method M&S Approach

This objective is about clarifying important terms. First, a pragmatic philosophical view will be used to provide a basis for the definition of a multi-method M&S approach and its derivative terms. The proposed definitions will supply an ontological base for theory of the multi-method M&S approach.

Objective 3: Propose Theoretical Principles Guiding the Use of Multiple Methods

Complementarity, Falsifiability, Commensurability, and Triangulation will be used to propose theoretical principles of multi-method M&S approach. Next, these principles will be utilized to develop building blocks called method formats, which provide an abstracted view of methods and their relationships.

Objective 4: Evaluate Multi-Method M&S Approach

In order to explore and assess the plausibility of proposed theoretical developments, a sample set of methods in the context of criteria for method selection will be analyzed. Next, a case study research format will be applied. Within this case study, research guidelines based on a theoretical principle for a multi-method M&S approach will be proposed. The developments undertaken based on a real-world problem by using these research guidelines will serve as data for the evaluation layer.

1.4 RESEARCH METHOD

Induction and deduction are often considered the most popular scientific research approaches. Induction directs research from "specifics to general" relying on observation and then inferring, which could lead to generalization. Deduction directs research from "general to specific" and often relies on rigid assumptions and testing their consequences. If an area of research was not adequately covered in the related literature, the inductive approach is usually a better choice [36]. Because the topic of the multi-method M&S approach was not broadly debated in relevant literature, it lends itself to an inductive approach. Moreover, inductive research is often associated with qualitative data i.e. non-numerical data. The analysis of qualitative data can lead to a theory, often seen as an outcome of research [37].

Adams and Buetow said: "It is tempting to assume that when good method and good processes are adequately assembled, good theory will follow. This act of faith fails to recognize the constitutive and multilayered contribution of theory." [38] It may be helpful to use a background theory (a starting point for further enquiry) sufficient to provide a basis in the context of a research thesis. It is also desirable to reach beyond background theories toward a grand theory. As pointed by Adams and Buetow: "While not every enquiry is compelled to explore its grand theory roots, a major enquiry, such as a PhD thesis, is vastly enriched when it tracks back to these origins."[38] Padilla et al. said that "M&S is the study of conceptualizations, their theory, analysis, design, efficiency, implementation, validity and verification, and application." [39] Theory building process can be based on M&S process [40], hence in order to investigate theoretical principles of multi-method M&S approach a higher order of analysis is needed.

This research could be characterized by both inductive literature analysis, complemented by learning by doing approach through exploration and development, preferred by the author and advocated by one of the most influential social scientists Herbert A. Simon (1916-2001) [41]. The learning by doing approach empowers the inductive approach by generating necessary observations for the evaluation. The theoretical principles created within this work will be reexamined using a case study [42]. The case study will be used to develop research guidelines from theoretical principles and apply these guidelines to a sample real-world problem for evaluation determinations.

The identified research gap summarized the lack of theoretical principles behind the multi-method M&S approach. Although multiple methods can be used for theory development [43], this dissertation research aims to close this gap with the inductive based research that includes exploration via literature review, application of relevant theoretical concepts within multi-method M&S approach context, and evaluation that involves M&S-based case study.

1.5 RESEARCH APPROACH

Figure 1 illustrates the research approach undertaken for this dissertation that leads toward the development of a theoretical basis of multi-method M&S approach. The research consists of three main sections and starts within the top large section. The literature review and analysis explores M&S relevant literature and synthesizes results into coherent perspectives. Four questions were explored:

1. What is the purpose of multi-method M&S approach?
2. What does exist within multi-method M&S approach?
3. How does one employ the multi-method M&S approach?
4. How does one evaluate the multi-method M&S approach?

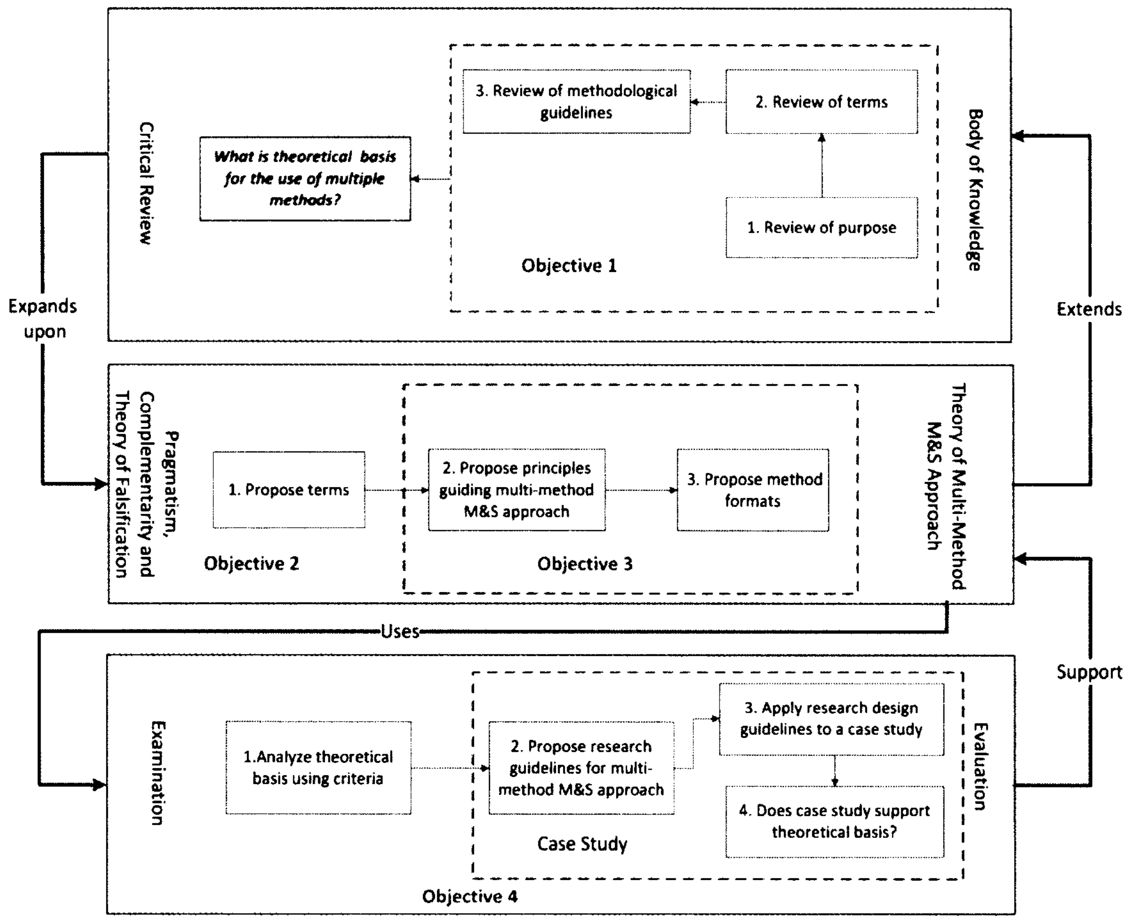

Figure 1. Research approach.

The first question explores the purpose of multi-method M&S approach. Exploring the purpose of using multiple M&S methods can contribute to a better understanding of its theoretical basis, exposing dimensions and criteria for deciding whether the use of multiple methods is the right choice in comparison to a single method approach. Two perspectives were examined. The first perspective consists of the review and analysis of relevant M&S literature, to summarize current reasoning behind the use of multiple methods. The second perspective uses theoretical principles related to the purposes of the mixed-method approach according to Greene [27], and projects them onto the M&S domain. Greene et al. [44] developed a mixed-method conceptual framework from the theoretical literature and refined it based on analysis of 57 empirical mixed-method evaluations. A key question to support this perspective is whether a more established mixed method could offer its knowledge, experiences, and principles to guide the research of the multi-method M&S approach. The second query reviews and discusses relevant terms related to the use of more than a single method e.g. what approaches that use multiple methods were called and why. The aim is to explore and determine if ontological ambiguity is present within the M&S field in the context of using more than a single method, which is necessary in order to analyze and develop more consistent ontological basis for multi-method M&S approach. The third query will review important and relevant research guidelines. This will include a review of method formats as a structural guiding aid in multi-method research, problem of method selection often related to criteria, and general research dimensions relevant to multi-method M&S. Finally, the last query focuses on research evaluation guidelines with a special focus on aspects applicable to multiple methods. A background for assessing quality and validity of a study that employs multiple methods based on social science perspective was discussed and different M&S perspectives on Verification and Validation (V&V) are briefly introduced.

This research expands upon critical review to propose the theoretical basis of a multi-method M&S approach, which can be considered a major step toward answering a research question of this dissertation. First, the analysis of the most important work terms is conducted. The author takes a pragmatic philosophical stance to provide basis for the choice of terms and proposes a definition of the multi-method M&S approach and its derivatives.

The second small block in the middle section develops principles guiding a multi-method M&S approach. Although theory of falsification developed by Popper [45] provides a very strict and anti-induction perspective, in the author's view it is a suitable starting point because it conveys an idea of a falsifiable statement, which reflects the idea of testability. The idea of a falsifiability of methods and commensurability of methods are defined and used in this dissertation to analyze complementarity of methods in the context of the purpose of multi-method M&S approach. This, in turn, will be helpful during theorizing about multi-method M&S approach in the context of its dimensions e.g. origination, methods, systems and/or phenomena at consideration, and human dimension. It is emphasized that the author of this dissertation is a proponent of a pragmatic philosophical stance, which does not constrain views about methods and theories that including both inductive methods and theory of falsification. Both perspectives, although quite far in their canons, are useful and play an important role in this research. However, the principles of pragmatism for using multiple methods in the context of M&S field may need more guidelines related to structured and well-defined purpose, especially when looking at reasoning to use different constructive methods. For that reason, degrees of falsifiability seem to be a good choice for exploring the theoretical principles of multi-method M&S approach.

Based on developed principles, the final block of the middle section develops method formats (MFs). In short, MFs pertain to a generalized view that consists of methods and system and/or phenomena. A set of transitions of model component(s) toward MF(s) can be used to design a multi-method M&S approach including a simulation model structure that can involve multiple modeling methods. Multiple sources for this derivation were used.

- Balaban and Hester [10] proposed an initial concept of MFs derived from empirical mixed method approach based on Greene [27]
- Review of M&S literature and the use of UML relations led to the specification of three general relations for a multi-method simulation model
- Proposed theoretical principles from the previous section

The bottom section focuses on an evaluation of theoretical principles. First, a sample set of methods and criteria for method selection will be used as a data layer during analysis of the theoretical basis. The goal is to gain insight into the relationship between commensurability and complementarity related to the purpose of multi-method M&S approach and problem of method(s) selection. Next, a case study is developed to look into a practical application of theoretical basis using falsifiers instead of criteria. Research guidelines are proposed based on previously developed theoretical basis, and are embedded into a case study format [42]. The conducted case study provides a mechanism to evaluate plausibility of the theoretical basis and their implications, examining whether theoretical basis will have the potential to improve decisions for choosing methods. This case study employees "*learning by doing*" approach, which seems suitable for the practical investigation of theoretical principles of multi-method M&S approach [41]. The case study could also serve as a model of how one can conduct multi-method M&S study. The case study will have three hierarchal dimensions:

- Dimension describing multi-method M&S approach research guidelines
- Dimension driven by the purpose of a real-world problem studied using proposed multi-method M&S research guidelines
- Overarching evaluation dimension, which will serve as a platform for assessment of the two other dimensions

The case study dimension will examine the theoretical basis to generate insight into plausibility of theoretical developments and will provide a valuable lesson to refine multi-method M&S research guidelines itself. A detail view of the decision to select method(s) will indicate areas prone to subjectivity. The case study will include implementation of a multi-method simulation model, which will allow for additional stimuli for the evaluation. The simulation model will be used for experimentation to explore a real-world problem, and additionally to evaluate the purpose of multi-method simulation model by assessing the insight generated. For instance, it will be examined if the use of multiple methods can be justified by examining if similar insight could potentially be generated without using a multi-method simulation model. This could show a case demonstrating complementarity principle indicating benefit of multi-method M&S over a single method in answering a research question. A detailed description of the case study dimensions are provided in the introductory section of Chapter 5.

1.6 SUMMARY AND OUTLINE OF DISSERTATION

This chapter introduced the main problem being addressed in this dissertation as a lack of theoretical basis of multi-method M&S approach, which led to determining a research question, which was followed by research method, objectives, and outlined approach. Finally, limitations of this work were briefly discussed. This dissertation has five chapters. Chapter 2 provides literature review and analysis related to Objective one. Chapter 3 focuses on theoretical basis of multi-method M&S approach, which aligns with Objectives two and three. Chapter 4 uses criteria for analysis of the proposed theoretical basis. Chapter 5 consists of a case study, which proposes multi-method research guiltiness and subsequently develops a multi-method simulation model. Both Chapter 4 and 5 contribute to Objective four of this dissertation. Finally, Chapter 6 provides a review of how the research question was answered with the stated objectives, and how this research contributed to the body of knowledge. Moreover, possible directions for future work are identified.

CHAPTER 2

BACKGROUND OF THE STUDY

More than four decades ago, Fahrland [3] introduced the notion of combined discrete event and continuous simulation. Presently, the practice of combining methods has matured and more simulation platforms offer capability beyond the original idea of combining two main modeling methods. Mingers [11] points at two main reasons for using a multi-method approach: "It is both desirable and feasible to combine different research methods to gain richer and more reliable research results." [11] He refers to the principle of complementarity in which "no one paradigm is superior, but, that their individual rationalities should be respected within the discipline as a whole."[11] Detailed definitions provided in the next chapter are guided by this chapter, but for clarity's sake, in this chapter "mixed method" refers to social science approaches and "multi-method" refers to M&S approaches that use more than a single method.

Tolk [46] pointed to ontology, epistemology, and teleology as enablers of a holistic view of M&S as a discipline. This view motivates development of the basis for a multi-method M&S approach in the context of teleological, ontological, epistemological, and axiological beliefs as shown in Figure 2. Tolk et al. [47] emphasized simulation philosophy as a key to the determination of whether or not current philosophy of science is sufficient, or a new pragmatic philosophy of simulation is needed. Moreover, Tolk et al. [47] pointed at the need "...to develop methodologies and standards for the use of simulation in scientific research." [47] Figure 2 is used as a guideline in this chapter, which consists of four main sections.

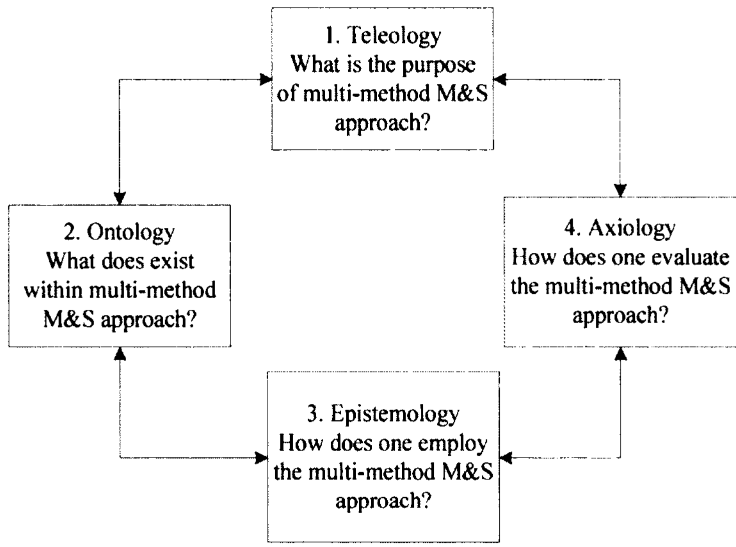

Figure 2. Basis for the M&S methodological developments.

In the first section, relevant literature is reviewed to examine the reasoning behind the use of multiple methods. The first part explores M&S relevant literature, while the second part uses theoretical principles related to the purposes of the mixed-method approach according to Greene [27], and projects them onto the M&S domain. The second section discusses terms related to the use of more than a single modeling and simulation (M&S) method, which aims to explore ontological ambiguity present within the M&S field in the context of using more than a single method. The third section investigates available research guidelines involving multiple methods including method selection and integration. Moreover, it explores objectivity, quality, legitimacy, and validity in context of evaluation of multi-method M&S, including approaches to, and evaluation of research. The last section ends with a summary of the findings.

2.1 PURPOSE FOR USING MULTIPLE METHODS

An initial review that could justify future work on the theory of multi-method clearly depends on the support of reasoning why should one consider using approaches that consist of multiple methods. The first part of this section explores purpose of the use of multiple methods based on M&S relevant literature. The review process is directed at finding different views, perspectives or reasoning for the use of simulation models that employ more than a single method. The second part of this section analyzes the purposes for mixing methods according to Greene [27], which are projected onto the M&S domain.

2.1.1 Purpose of the Use of Multiple Methods in M&S Field

The following are the main purposes for the use of multiple methods in M&S field found in M&S relevant literature.

The *complementarity* of methods presumably mitigates assumptions prescribed within methods that allow for shaping research approaches that are more flexible. Eldabi et al. have gathered information on the direction of M&S domain in the healthcare context in the form of synthesis of the trends identified by experts in the field. The reasons for combinations of methods and the need of hybrid methodologies given by respondents referred to "move[ing] away from perception that one method fits all" [48], a need for a holistic view of the complex interconnected systems, and a need to include human elements. Brailsford et al. [25] have demonstrated *complementarity* of SD and DES in inclusion of different system factors in relation to perception of components being inside or outside of the system. Similarly, Morecroft and Robinson [9] noticed the *complementarity* of SD and DES. They observed that DES effectively captures detail complexity by tracking and analyzing of individual entities, but does not handle dynamic complexity easily because implementation of feedback loops is less intuitive and more difficult to build. The opposite is also true for SD. Zulkepli et al. [49] reflected that combined Operations Research (OR) and M&S techniques might reduce the limitations and increase capabilities of the individual methods e.g. passive individuality requirements for DES, and feedback elements of SD. Kott and Corpac [50] noticed that no single modeling method is truly relevant to the entire Diplomatic, Information, Military and Economic (DIME) and Political, Military, Economic, Social, Infrastructure, and Information (PMESII) dimensions. This indicates the *complementarity* reason with the emphasis on the system context as the main reason directing toward the multi-method M&S approach. The use of multiple complimentary methods may carry additional abduction risks. Abduction can be considered a third way of research, besides induction and deduction, and it pertains to finding causes for a certain effect by assuming that a specific resulting regularity are adequate (which is uncertain) [34]. Lorenz and Jost [34] described this risk to be more dangerous if the implicitly accepted combination of assumptions carried by different methods are not well understood, or cannot be stated explicitly, leading to higher uncertainty of the observed regularities. Level of coupling between complementary methods depends on the level of required interaction between methods. Fahrland [3] has considered use of a multi-method methods within a single simulation model in cases where representation of system elements not only required different methods, but additionally a strong interaction between these methods. Similarly, Helal [51] has considered application of multiple methods dependent on the presence of strong coupling between methods. Subsequently, Chahal [8] developed a framework in which the need for multi-method simulation model is reliant on strong dependency between methods. This reasoning indicates the possibility of different levels of coupling between methods e.g. methods that do not interact, or methods that interact during a simulation run.

Multilateral problems. Djanatliev et al. [52] decided to employ multi-method SD and ABM to cover both a globally aggregated level and more detailed workflows. They believe that a combination of methods could profit in assembling complex, large-scale simulation architectures, and that taking advantage of different modeling methods could help them in answering multiple questions about economic prognoses and impacts of different factors on patient's health. Currently, multi-method simulations are employed more often because more complex problems are being targeted [23].

Modeler preference and skills. It is clear that modeler preference plays a role in the use of a multi-method approach. Viana et al. [53] do not elaborate much on why they decided to use multi-method approach, explaining that each subsystem was implemented using the best method, with the "best" meaning the method that most closely aligned with the mental models of designers. Glazner [35] used SD, DES, and ABM to represent views of different parts of the system but noticed that two out of three views could be modeled using either of three methods. The decision on which method to use in each case was a combination of the modeler preference and expected modeling effort most likely related to proficiency in using a modeling method. Only one view could be clearly determined for the use of ABM because the individual behavior of the organizational unit could not be sufficiently represented using either SD or DES methods. This example indicates that in some cases, there is a gray area for choosing a method, but in other situations, there is a clear choice due to capabilities needed. A modeler needs to make a decision about which method, or combination of methods, is the best or satisfactory choice for a given purpose. On one hand, a modeler's expertise is often the determining factor for a method choice [54]. However, if a modeler is unfamiliar with some crucial method, there is a risk of using a suboptimal method by adjusting problems to methods with which the modeler is more acquainted. According to either Chahal [8] and Lorenz and Jost [34], the opposite, choosing method to fit the problem, is the right approach.

Stakeholder acceptability. Viana et al. point out that by using different methods suited better for different tasks, "the stakeholders have gained greater buy-in and understanding, where the stakeholders included both the problem owners (health care and social care professionals) and those members of the project team who are unfamiliar with the techniques."[53] Similar reasons, oriented toward acceptability of simulation models by stakeholders were given by Sachdeva et al. [55]. The results from their study indicated that a mixture of hard and soft OR methods allowed for better understanding, acceptance, and willingness to implement results by stakeholders.

Data availability. Lättilä et al. [30] suggest that data availability could also be a factor for choosing multi-method approaches. Because data availability often depends on phenomena studied, and because different data could align better with different methods, a multi-method approach could allow alignment with available data from different parts of the system.

Validity. Could advantages of a multi-method simulation model be based on validation merits? Parunak et al. [56] pointed out that validation at multiple levels of analysis might be more difficult, but could deliver a more accurate model. Following this idea further, if a multi-method approach can facilitate adequate modeling at multiple levels of analysis, it is possible that this leads to models that are more accurate as well. Crespo and Ruiz [32] have combined DES and ABM with a goal of obtaining estimation that is more accurate and a more realistic model of the CMMI process. The innovative part in this model included the use of ABM to represent the project coding process, including the project team behavior from the participants' perspective. Similarly, Siebers claimed that a combination of DES with ABS had a positive impact on the model accuracy and allowed for "proactive behavior in service system models." [18]

Unique representation. Lättilä et al. [30] determined problematic situations where a combination of ABM and SD are needed in order to create models that are more realistic: 1) different actors, e.g., in SD, actors are homogeneous, in ABM, they are heterogeneous; 2) data availability; 3) system structure, e.g., in SD is fixed, in ABM it can change; 4) complexity of events; and 5) policy representation. The need for more sophisticated modeling approaches to represent proactive behavior was the reason for extending the Commander's Model Integration and Simulation Toolkit (CMIST) [57]. CMIST is a multi-method modeling environment integrating so far three modeling methodologies: SD, ABM, and derivatives of Bayesian approaches, namely Dynamic Bayesian Network (DBN), and Bayesian Knowledge Bases (BKB). The addition of BKB had the intention to support advanced intent modeling for inference of goals and beliefs of an agent. This extension allows for the representation of more proactive agents. The agents were capable of simulating the simplified model of the already simulated world, projecting the future state of the simulated world, including for instance adversary behavior. Lieberman [19] also used DBN as a method for representation of an agent's internal Beliefs, Values, and Interests (BVI), which is an interesting direction to enhance representativeness of an agent by capturing a change of perspectives, values of prior probabilities, and likelihood function probabilities to accommodate for new information. Kott and Corpac [50] presented Conflict Modeling, Planning and Outcomes Experimentation (COMPOEX) as an integrated set of decision aids to assist leaders in planning and executing campaigns. The COMPOEX described a set of interacting models, developed with appropriate paradigms, required to represent the environment defined by all of the DIME and the PMESII dimensions. The COMPOEX engages many different modeling methods, e.g. concept maps, social networks, influence diagrams, differential equations, causal models, BN, Petri-Nets (PN), SD, DES, and ABM to facilitate unique representation of individual dimensions leading to a better representativeness of large, complex systems.

Emergent phenomena. Kott and Corpac remark on multi-method M&S, "A family of interacting models have the potential to produce surprisingly unanticipated results due to effects of cascading." [50] A cascade reaction is a result of interactions between models that can produce an emerging situation that a single model by itself could not. This reaction indicates the purpose of surprising discovery, but the important question to answer is determining whether or not this reasoning was conceptualized at the origination of the model's concept or if it was realized because such an interesting effect was observed and then considered desirable to facilitate understanding and stakeholders' discussion. Please refer to [58, 59] for a discussion and classification of emergence types.

Dimensions and criteria. Different criteria and dimensions provide more systemic view of purpose and were applied to justify the usage of multi-method M&S. For instance, Brailsford and Hilton [54] focused on technical differences, whereas Lane [60] focused on conceptual differences. Sweetser [61] used a structure, mental model, system orientation, role of simulation, and validity as criteria to differentiate between SD and DES methods. Axelrod [62] provided criteria for choosing modeling methods in relation to a modeler: construction time and effort needed by modeler to build a useful model, and flexibility and ease to modify it; a user: user prerequisites, time to learn, transparency to discover bias; and a method itself: mathematical rigor, predictive value, and heuristic value. Behdani [63] characterized SD, DES, and ABM methods in accordance with their ability to represent complexity at micro and macro levels. These two levels are further divided into criteria, which provide guidance for selection of one or more methods. Lorenz and Jost [34] proposed three dimensions that should be aligned in order to choose the suitable modeling approach: purpose, object, and methodology. Chahal [8] took this idea further. He used three different perspectives to describe and differentiate between SD and DES methods: the methodology perspective that covered criteria based on assumptions, capabilities, and unique aspects of methods; the system perspective that was concerned with the real system under investigation; and the problem perspective that focused on why a method might be useful for studying a problem. Each of these perspectives carried a set of criteria, which revealed possible reasons for choosing between SD and DES. The choice based on the criteria could also reveal that single method models were not the appropriate choice.

Missing consideration of "the why" question. Waltz [64] provides a discussion that categorizes PMESII elements with the detailed model components, along with the methods used for each component, and its modeled function. It seems that each category of PMESII has a dominant method, e.g. Political – ABM, Social – BN, Economic – SD, Infrastructure – SD, Information – DES, and Military – mainly SD with some BN use [64]. The strengths of the four major categories of modeling approaches (ABM, SD, BN, and DES) used in COMPOEX are briefly presented in Tables 2-4 [64], but the lack of discussion about reasoning and the justification for combining these methods should be mentioned. It is a problematic situation to provide the "what", but ignore the "why" questions in methodological reasoning about a multi-method approaches. Glazner [35] noticed that decision regarding which method to use was subjective in two out of three cases, indicating that there is a gray area for choosing a method. On the other hand, there was a clear choice favoring use of ABM due to capabilities offered by this method.

The presented perspectives on reasoning behind the use of multi-methods can provide a starting point that can shape the direction of this research. It is noticeable that relatively young multi-method M&S field has limited scientific literature. The reasoning for the use of multi-method simulation models that were found in the literature relate to the complementary nature of methods with the additional need for methods coupling, data availability and usability, skills and preference of a modeler, stakeholder acceptability, emergent phenomena, enhanced with the very diverse needs related to understanding, credibility, validity, and complexity of models. Dimensions and criteria provided by Lorenz and Jost [34], Chahal [8], and Behdani [63] are a good starting point to provide more systemic perspective, but generalizability of a single set of criteria should be further reconsidered.

A limited use of a multi-method approach at the end of the twentieth century could be attributed to narrow the focus of educational institutions, lack of textbooks, and lack of appropriate tools [65, 66]. This situation has improved with more software and educational resources available. Additionally, multi-method simulations are employed more often because more complex nature of problems are being targeted [23]. Lorenz and Jost [34] stated that modelers could overlook modeling methods when deciding which one(s) suit the purpose because they are not very familiar with them or have biased preferences, which can lead to an inability to compare alternative approaches and to choose methods based on insufficient judgment. It is possible that some scientists are not acquainted with more than one simulation method, and they might not be able to explore the potential for more flexibility and creativity by integration of multiple simulation methods.

Viana et al. have pointed at more difficulties and challenges

> "…in designing sub-components and their interactions so that they represent the real-world complexity without overwhelming the model with impenetrable detail. Moreover, this process is both enriched and made more challenging by the combination of disciplines involved. The work required a marriage of an OR stakeholder-driven approach, with the 'empirical eye' of social statisticians and the micro-level theories of complexity science. Social statistics helped make informed decisions on where mechanisms could be abstracted from relationships in empirical data, rather than having the causal mechanisms modeled explicitly.

> However, the latter is a strong current in complexity science, and promises to help in better modeling individuals' adaptation to changing social and technological contexts, which the scenarios explored here represent." [53]

This citation indicates that with more complexity involved in the project came more work, more people with different backgrounds involved, and most likely a need for better methods and tools. When looking at organizations that started an application of multi-method frameworks these are usually big sponsoring organizations involved in larger projects [19, 50, 57, 67]. On the other hand, cheaper multi-method simulation tools and better research guidance should change this situation. Unfortunately, there are not many modeling platforms allowing for easy use of multi-method M&S, and appropriate tools like AnyLogic® are rather expensive. Currently, the lack of more explicit reasoning displaying advantages and purposefulness of multi-method M&S can add up the need to overcome the difficulties related to tool availability. Addressing the purposefulness of multi-method M&S requires a tangible reasoning why multi-method approach is needed to support the decision to use a multi-method simulation based on some merits. Additionally, multi-method M&S could be described as theoretical guidelines within a set of general formats.

The literature reviewed so far showed existing reasoning for the use of the multi-method M&S approach based on M&S literature within socio-technical context. This reasoning can provide a starting point for exploration of the usefulness and purpose of applying a multi-method approach. The studies that use a multi-method approach as a research method often consider both social and technical phenomena (see Appendix A for definitions), but subjectivity of social phenomena generate difficulty to more objectively analyze merits of this approach. One can observe growing trend of trying to incorporate social phenomena into more descriptive simulations that including combining them with technical phenomena [18, 19, 49, 52, 53, 68, 69]. Because of a scarcity of implemented and analyzed in detail multi-method M&S studies that consists of social phenomena the reasoning for the use of multiple methods to represent social phenomena may be more challenging. The idea of using multiple different methods has is also present in empirical mixed methods, which is a well-established field with a dedicated journal, *Journal of Mixed Methods Research*. Because there is no well-established equivalent research within M&S field, the view of purposes related to social phenomena research will be analyzed based on mixing methods from empirical social science domain.

2.1.2 Purpose for Mixing Methods

Starting in 1970s, research paths of using mixed methods in social science began to emerge, and "started to blossom at the turn of the century." [27] Mixing methods in social inquiry could be described as invitation of different mental models into the same inquiry space with plurality of philosophical paradigms, theoretical assumptions, methodological approaches, formal techniques, and with inclusion of subjectivity reflecting the human perceptions.

A key question to support this work's research approach is whether a more established mixed method could offer its knowledge, experiences, and principles to guide the research of the usefulness of the multi-method M&S approach. This section uses the purposes for mixing methods in empirical social science and explores their analogies within the M&S domain. This is facilitated by exploration and translation of mixed methods' perspectives covered by Greene [27] into the area of multi-method M&S.

An initial comparison of both the mixed method approach and the multi-method M&S approach should begin by discussing the context of their emergence, and similarities and differences. Both multi-method and mixed method views emerged as an alternative to the single method approach. There is a claim for more creativity in mental processes using mixed method approach because of the abilities to connect many conceptual dimensions through multiplicity of methods used [27], which also seems plausible for the use of multi-method M&S, but this notion is not supported by scholarly literature. It seems that both approaches can suffer from more difficulty in the design, development, and analysis. Practical aspects of mixed methods are more difficult than theoretical ones [27]. This statement may be not so obvious for multi-method simulation models. The development of a multi-method simulation model can often be considered difficult, but theoretical and axiomatic aspects are also problematic. In mixed method social study, a "wider toolbox" increases flexibility and chances of a broader view of phenomena. Similarly, a researcher engaged in multiple dimensions of building, testing, analyzing of a multi-method simulation model could draw mental models represented differently with each method. The availability of the "right" simulation method and skills required for multi-method M&S could facilitate broader modeling perspectives on a system. This may increase chances of building a model that is adequate for its purpose. Propelling modelers' generative abilities may be the most important advantage of the multi-method M&S approach. On the other hand, this fact could be very difficult to prove. It's emphasized that the generative mental state of a modeler during modeling is considered here, which is not to be confused with the "generative growth" approach that considers generative aspects of a simulation model [70]. On the other hand, if a multi-method approach allowed increased creativity in modeler's mental states, this ultimately could also yield more generative simulations. A mixed method approach has been established and growing fast and many research guidelines were proposed [27, 71-74]. Recently, the multi-method M&S approach has become more popular, but the lack of a more general and systematic approach in the form of research guidelines or a framework is problematic.

Greene et al. [44] developed conceptual framework aimed at mixed method approach. It is based on theoretical principles from the literature with addition of the analysis of 57 empirical mixed-method evaluations. They identified five purposes for engaging in mixed-method approach. The exploration of these purposes could provide an important direction for the evaluation of the usefulness of the multi-method M&S approach, especially in the context of representing social phenomena. The following is the summary of these purposes and their projections reflecting the M&S multi-method context.

Figure 3 illustrates the ideas covered during the discussion.

1) *Triangulation* uses different measures for the purpose of investigation of the same phenomenon with offsetting biases of different methods, with the ability to identify irrelevant sources of variation, observing consistency based on comparison of results from different methods. It captures a phenomenon through different lenses but with the same conceptualization. This has a goal of increased validity and credibility. In the M&S field, this may be conceptualized as building two or more models using

different methods, maybe by different parties, to increase the validity of results or to represent phenomenon through different lenses of abstraction (e.g. specific or general).

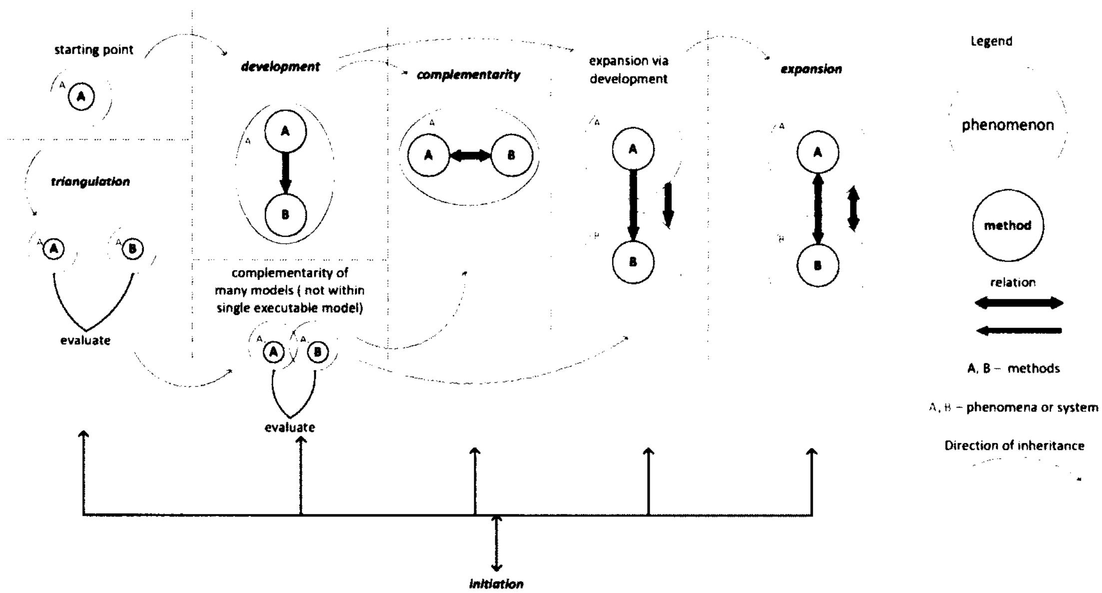

Figure 3. Graphical representation of mixed method projected onto multi-method M&S based on [44].

2) Triangulation could also be considered in the context of different models built with the same method. The main idea behind *triangulation* refers to the possibility of the comparison of two or more models, for our consideration (focus) developed with different methods. The models are not designed in order to interact together during the models execution.
3) *Complementarity* focuses on broader, deeper, and more comprehensive facets through additional development, initiation, and expansion of the same complex

phenomenon. Different methods are employed because they complement each other. This approach projected onto the M&S field might be translated as the addition of elements or views realized at a different or the same level of analysis by using different methods needed for better representation of a phenomenon for a given purpose. A somewhat similar idea in the M&S community can be called a pluralistic perspective and was advocated by Helbing, who wrote that this approach "should lead to a better quantitative fit or prediction than most (or even each) model in separation, despite the likely inconsistency among the models." [75] Helbing considered usefulness of different models to represent different aspects or parts of the system (which may overlap) by creation of the analytical structure made of different models that increases validity of the insight. Unfortunately, he did not consider merging models into a single executable model. What follows, an opportunity for increased usefulness of combined methods should be considered as a driver for the use of integrated models. Axelrod recognized the scale of difficulty in the process of developing a combined M&S model: "The most ambitious method is to develop a single unified outline of a comprehensive model… This method of linking two or more models is substantially more ambitious than merely requiring that one model's output to serve as another model's input, since the comprehensive method requires that the parts work together in many different ways." [62] In the M&S field the *complementarity*-based simulation model should be implemented in the form of views that can be integrated, allowing for more holistic view of the system or phenomenon. Because the focus of this work is the multi-method M&S approach, *complementarity* refers to methods, not models. Two forms can also be distinguished that are important to consider in the M&S field. The first form should consider execution of complementarily viewed parts with different methods within a single model. The second approach focuses on the use of complementary models with separate methods that are not executed together and used, e.g., via analytical evaluation that provides a more holistic view. Hence, the major difference between two complementarity views in M&S lies at the level of binding: executable as a single model or not. A tight analytical structure for evaluation of complementary models as proposed by Helbing [75] is closer to triangulated and complementary models that are built with analytical binding, because Helbing did not consider a single executable model built with different methods. Obviously, there can be many models of phenomenon built with the same method, which relates to a broader human perspectives' on complementarity and triangulation, providing different viewpoints based on each modeler's views and views of many modelers as well. The combination of model, human, and method dimensions creates possible combinations of how one can understand complementarity. Because this work focuses on purposes of the multi-method approaches, a methods' complementarity is given the most consideration at this moment.

4) *Development's* main idea lies in the sequential alignment of different methods with their inherent strengths, where one method is used to inform and help in the development of the follow up work that employs another method. In M&S, this could mean that an output from the first model represented with one method is used as an input to the second model using a different method. The frequency of updating between methods defines time complexity of this unilateral binding. Other options

explaining projection of *development* into the M&S field is the purpose of the systematic increase of the phenomenon understanding, facilitated by using different methods at different stages of modeling and validation of a conceptual model with an intermediate method [76]. This option would not require methods to be integrated, but be only related by a sequential function in the simulation-based research process. In order to distinguish this purpose from the complementarity purpose, it is specified that interaction flow (conceptual or numerical) is unidirectional (no feedback).

5) *Initiation* induces paradox, contradiction, divergence, dissonance, and disagreement in order to create different perspectives and important insights, and allows for discovering the need for further analysis. It is similar to complementarity but with the concept of looking at a broader scope of disagreement and divergence. In multi-method M&S, initiation may be realized when applied additional different method is leading to contradiction, surprising results, or unexpected insight in comparison to the single method original model. Even if this seems more an effect than a purpose, use of, for instance, ABM, in social science is especially focused on *initiation*. Unfortunately, social scientists, in large measure, are not concerned with the possibility of multi-methods M&S as the additional driver of this effect.

6) *Expansion* calls for the use of different methods to capture different phenomena, which extends scope, breadth, and range of a study. It focuses on the use of the most appropriate method for different constructs. In multi-method M&S, this may be represented as the combination of different modeling methods to capture different phenomena.

All of the presented purposes for engaging in mixed method approach have feasible explanations or projections to simulation-based studies. The focus of this dissertation is a multi-method M&S approach, including simulation models where methods exchange or does not exchange data during their execution. The purpose of *development* is limited to the sequential character and could be derived in the M&S context from the purpose of *complementarity* or *expansion*, which makes it a subset of them. The focus of *triangulation* in the context of methods is their separate use for validation purposes via comparison.

The purpose of *initiation* seems applicable to all of the other purposes as the desirable feature, but it is a very abstract concept that exists at human dimension and therefore it is more difficult to represent graphically. The exploration of emergent phenomena can often be surprising, and social scientists are engaged with simulation techniques to get that "wow" moment that could be described by the *initiation* purpose. Most likely origination of the study directs the use of multi-method M&S approach by purposes of complementarity, development, or expansion that could lead to the *initiation* effect. Furthermore, it would be problematic to assume that the multi-method approach would bring constructive disagreement from the beginning of the model design. The purpose of *initiation* needs further research in M&S science, especially because it can be considered a higher-level purpose for explaining social phenomena. The above discussion about the purposes of multi-method M&S based on purposes for mixing methods provided by Greene [27] will be narrowed temporarily to *complementarity* and *expansion*. The purpose of *initiation* is an abstract concept that can exist within any other purposes considered here.

The *complementarity* and *expansion* elements as purposes for using mixing method approach in social science are relevant when projected onto the reasoning for the use of multi-method M&S. However, these are high-level purposes that need to be interpreted through more detail M&S dimensions and criteria. In order to justify the choice of using a multi-method M&S approach in a given study context, this choice should show its superiority over a single method model by providing supporting dimensions and criteria. It should be shown that a single method model could not provide the same results or insight as model obtained based on the *complementarity* or *expansion* purposes. For instance, the need of *expansion* of a model to embed additional phenomena can lead to requirements identifying multi-method M&S as the preferred approach, thereby prohibiting the choice of expanded model using a single method as sufficient to capture multiple phenomena. Similarly, additional insight into a phenomenon through refinement or generalization should be shown impossible with the single method approach. Obviously, these cases should not be considered as the general rule, but as prove of concept showing the need for of a multi-method approach in some cases. Hence, *expansion* or *complementarity* could take the multi-method route, but depending on some additional dimensions or criteria that would have regarded the single approach as inferior. The purposes of *expansion* and *complementarity* can sometimes become vague depending on a subjective definition of phenomena. When analyzing Greene's definition, the *expansion* could not be conceptualized as *complementarity* purpose because it is directed toward additional phenomena. On the other hand, when considering concept of M&S methods' complementarity only, this difference could be omitted because a phenomenon does not have to be considered as a unit of analysis to distinguish between the *complementarity* and *expansion* reasons. From the M&S perspective, it is possible that different methods complement each other in order to expand the simulation model inward or outward through refinement and generalization. In this context, complementarity is required to expand a view on a phenomenon or extend a model with a new phenomenon. It does not seem sufficient to say that different methods are always required, but they may be required to complement each other. With this in mind, it is possible to combine social science purposes of *complementarity* and *expansion* perspectives and M&S's method complementarity perspective to describe *complementarity of methods.*

Definition 1

Complementarity of methods *is a purpose for using different methods within mental, analytical or simulation space to enhance the expansion of studied phenomena or systems inward (generalization or refinement), or enhance the expansion outward to combine different phenomena or systems (scope). Multiple inward and outward expansions are possible.* This definition is refined based on work of Balaban and Hester [10]. The complementarity of methods can also be internally driven by a set of practical reasons, e.g. required computational efficiency, data availability, skills and preferences of modeler, and origination of research related to and managerial and organizational circumstances, e.g. preferences of stakeholders [10].

Because another echelon of reasoning for the use of multiple methods is related to triangulation, the following discussion explores triangulation in the context of M&S study, especially looking into the context of methods.

Triangulation is a strategy for increasing the validity of evaluation and research findings [77]. Triangulation can be used in a context of a purpose or a study type to investigate the same phenomenon through replication of study results using the same or different methods [27]. Denzin [78] specifies four types of triangulation: data triangulation, investigator triangulation, theory triangulation and methodological triangulation. These types can be generalized in the context of M&S by introducing a concept of a level of triangulation. Moreover, triangulation conducted by separate modelers should be distinguished from pseudo-triangulation conducted by the same individual modeler.

Definition 2

Pseudo-triangulation *is triangulation that is conducted by the same individual who conducts the original research.*

Triangulation level describes a phase of a replication study based on how it is conducted, which influences level of variability allowed in the triangulation. Within an M&S based study, stakeholders and modelers would have at least two study decision points affecting triangulation: method(s) and a starting point of triangulation. These decisions would affect comparability of results, closeness of results of compared studies, in turn reflecting on credibility of triangulation. The aspects of triangulation level are introduced next.

Within M&S, levels of triangulation could be established based on a generic M&S research process, e.g., purpose, research question(s), concept of phenomena and system(s), simulation model, experimentation, and analysis. This way, when specifying a level of triangulation, it is assumed that the previous levels are asserted, and relevant knowledge base is available. Each of the proposed levels could serve as a starting point of a triangulation study. The steps of a study would align with the order of decreasing permitted level of variability of the triangulation. The purpose level permits the highest level of variability and in principle should generate the most credible results, with these results decreasing when starting at lower levels.

During triangulation at the study purpose level, modelers share only the purpose and work in total isolation without sharing information about all the following phases, i.e., development of research question(s), concept, and analysis. Triangulation at the research question(s) level would assume the same research questions and purpose, while triangulation at the concept level would assume the same purpose, research question, and concept as a starting point. Following the same logic, one arrives finally at the analysis level, where triangulation would utilize the same design of experiments, and modelers would analyze output data, describe insights, and recommended decisions. Moreover, the proposed levels can be refined into smaller levels based on the desired insight to be gained from the modeling effort. For instance, the concept level could be informally separated into high and low levels. High-level could provide an overview of phenomena and system and depending on the model's purpose, it could provide some dependencies between them, e.g., a causal loop diagram, whereas a low-level would operate on constructs of constructive simulation methods like DES, ABM, SD or implemented Statechart (SC). Moreover, these levels serve as a general overview only, and study-specific triangulation levels can be derived.

Another option that can influence research design is preselecting methods. This means that the methods may be artificially imposed, which can influence the rest of the process. Method(s) could be preselected in order to lower variability of solution by considering the same method(s) at the purpose, research question(s), or concept levels. It is important to note that triangulation at the concept level (low-level) can be considered as a threshold point at which methods must be selected in order to implement a simulation model.

Pseudo-triangulation is more problematic. At the purpose level, additional research questions would mainly expand the research scope conducted by an individual within established purpose. At the research question(s) level, multiple concepts of the same phenomenon and the resulting development paths could be considered by an individual based on the same or different method(s) for established research question(s). Because the concepts are created by the same individual, they cannot be derived independently and the objectivity of resulting triangulated views can be affected. Some expansion of the original concept is very likely depending on methods used and pseudo-triangulated views can refine the representation of system or phenomena. For instance, possible alternative simulation models can be considered by an individual based on implementation options related to method(s) for an established conceptual model. At the simulation model level, experiment level, and analysis level, pseudo-triangulations are even more questionable, because alternative designs of experiments or additional analysis would expand the research effort, rather than serve as confirmation of the results as is the case when conducted by separate modelers.

As discussed above, triangulation at different levels could provide benefits to compare different research paths. The question is also if preselecting method(s) at early stages of study, e.g. at purpose or research question levels, is a justifiable practice when considering how this can limit possible variability.

2.2 AMBIGUITY OF TERMS USED

In order to develop an ontological basis for a multi-method M&S approach, this section discusses terms related to the use of more than a single modeling and simulation (M&S) method. The aim is to explore the ontological ambiguity currently present within the M&S field in the context of using more than a single method.

Hofmann [79] distinguishes two classes of ontologies in modeling and simulation: methodological, which defines methods, and referential, which focuses on representing real-world systems. Partridge et al. [80] discussed briefly historical background and different aspects of the use of the word ontology. For instance, they referred to Honderich [81] who described derivative use of ontology to describe things that exist within a theory. This top-level meta-methodological context is adopted, providing context for the word ontology in this work, and a base for the clarification of terms relevant to a multi-method theory.

As with many fast-growing application fields, it takes time to clarify and categorize terms, definitions, and knowledge of new branches of the multidisciplinary M&S field. This is also due to a variety of applicable M&S methods in different domains [28]. DES, SD, ABM and other approaches are called methods, paradigms, techniques, formalisms and methodologies. The literature consists of different terms describing concepts related to the situation where more than a single method is used, e.g. multi-method, multi-methodology, multi-paradigm, hybrid, mixed-method, multi-model and multi-formalism. Most often, several of these terms are used as synonyms solely for readability purposes, while sometimes only single term is used, and still other times, different meanings of those terms are intended. In many cases, the purpose of using multiple terms is difficult to determine. Below are a few examples presented to show the need for more consistency in using different terms that may or may not mean the same thing in the M&S field. The following review is only a sample of the vast extent of relevant literature. It is hoped that this short review illustrates the scope of this problem. It is stressed that the purpose here is not to criticize, but to present the current situation, discuss it, and, later on, propose a more unified taxonomy.

Balaban and Hester [10] use the terms method and paradigm without discussing possible differences between them. Chahal [8] refers to hybrid simulations and models as integrated DES and SD and described hybrid simulation as a form of mixed methods. He also uses the term multi-method in sentence "Through an extensive review of existing literature in hybrid simulation, the thesis has also contributed to knowledge in multi-method approaches." [8] This may indicate a parent-child relation between multi-method approaches to a hybrid. Finally, Chahal referred also to SD and DES as paradigms, e.g., "...deployment of SD and DES in an integrated way, where both paradigms symbiotically enhance..." [8] Rabelo et al. [82] and Rabelo et al. [4] call SD and DES methods, but also a methodology and integrated SD and DES a hybrid or a methodology. Glazner refers to DES, SD, and ABM as simulation methodologies, but also as paradigms: "In other cases, this paradigm might not make sense..."[35] It is difficult to determine if he equates the words "hybrid" and "multi-methodology" by saying: "others have gone on to argue that a portfolio of stand-alone simulation models does not accurately convey the system's dynamics, and that a hybrid, multi-methodology approach to simulation should be used."[35]

Helal [51] refers to hybrid as a more than a single form of abstraction used to represent e.g. cars, robots, cell phones, digital watches, medical devices microwaves, washing machines because they fall under a hybrid systems umbrella. He defines a hybrid simulation as "combined discrete-continuous simulations, which gives modelers the ability to reach better fidelity and fit the characteristics of all sections of the system being modeled." [51] Moreover, he refers to SD and DES as a methodology or a method, but the word "method" is also used to connote numerical methods, HLA calling methods, and synchronization methods in distributed simulations. Martin and Raffo [83] described a hybrid as a combined continuous and discrete models and two main modeling paradigms, allowing to examine phenomena that are not reproducible in either continuous or discrete models alone. Choi et al. [84] describe the combination of SD and DES paradigms as a hybrid, whereas the word method was used in reference to numerical integration. Levin and Levin [85] use a word paradigm to refer to continuous differential equation and discrete finite state machine (FSM) parts. They use the word hybrid based on "...hybrid system theory [that] connects two models of change, one described by continuous differential equations and the other by discrete logical transitions." [85] Osgood [86] uses the word hybrid to mean combined discrete and continuous rules and hybrid automata from analog-digital control theory and refers to SD and ABM as paradigms. Henzinger [87] defined a hybrid system a dynamical system with both discrete and continuous components and developed a formal model of a mixed discrete-continuous system called hybrid automaton. Rossiter and Bell [88] call workflow hybrid an integrated multi-model, multi-paradigm simulation framework, and call SD, DES paradigms. Setamanit et al. [89] call hybrid a combined SD and DES. Swinerd and McNaught [23] call SD and ABM as both paradigms and methodologies, while combined SD and ABM hybrid or multi-methodology. They define "hybrid approaches [as those] which combine at least two of the three methodologies discussed [SD, ABM, DES]."[23] Venkateswaran and Son refer to hybrid simulation as "the work carried out in using together discrete and continuous aspects for analyzing a system."[90] Wakeland et al. [91] call hybrid combined SD and DES, and machine learning approaches are called methods. Heath et al. [29] refer to SD, DES, and ABM as paradigms, and examine "cross-paradigm" modeling. In the same paper, the word method is used for DES and a naïve Euler, Runge-Kutta algorithms. Hassan et al. [21] use the word "paradigm" in the context of the individual social agent, while the ABM is seen as a tool that executes several individual agents. Peña-Mora et al. [92] refer as hybrid to combined SD and DES. Rabelo et al. [93] describe initially a hybrid approach as a combination of SD and DES, while analytic hierarchy approach (AHP) is listed as separate item, not as a part of hybrid. In the conclusion of their paper, the authors change this structure: "This paper presents a preliminary analysis of the potentials of integrating the group analytic hierarchy process (AHP) technique, system dynamics (SD) and discrete-event simulation (DES) in a comprehensive hybrid approach." [93] They refer to AHP as a methodology but also as a method and technique. This example may indicate evolution of the use of the word hybrid beyond continuous and discrete methods. In this context, the term hybrid can be synonymous to the term multi-method since it has evolved from its original meaning as the combination of two discrete and continues views into more general meaning. Zulkepli et al. also expand meaning of original world hybrid to include OR/ simulation methods "such as Optimization, Markov Chains, Linear Programming, DES, SD, Forecasting, Just-In-

Time, Decision Trees and Soft Systems Analysis, to facilitate better and more informed decision making."[49]

Lee et al. [94] call integrated SD and DES simply combined SD and DES. They also used combination of different words like hybrid, method, paradigm, and technique for writing convenience: "This hybrid algorithm is developed to combine the nested partitions methods with the paradigm of an efficient ranking and selection technique."[94] This shows how puzzling the writing about application of multiple methods can become. Hester and Tolk [28] discussed M&S methods in the context of their use for systems engineering (SE), providing an overview of M&S methods. The two sentences "...(M&S) methods in support of complex systems engineering has become integral part of the "toolbox" used today by engineers."[28] and "...the different M&S methods used to improve systems engineering efforts are often perceived to be based on fundamentally different paradigms" [28] indicate that paradigm can be seen as a more established method, but both terms are used later in the paper often as synonyms.

Zeigler et al. refer to use of different methods (formalisms) as multi-formalism: "...they require a combined discrete/continuous modeling and simulation methodology that supports a multi-formalism modeling approach..."[95] Moreover, "...a model that subsumes several different models is termed multi-model. The DEV&DESS formalism is an appropriate means to implement multi-models." [95] Fishwick refers to a multi-model as "...a collection of individual models, each characterizing an abstraction level-connected together in a seamless fashion to promote level traversal" [96], and mixes the words model and method: "It is better to choose a variety of well-utilized and proven modeling methods and then search for ways to glue them together to yield a multi-model rather than always to view the world to be modeled through a single-model colored lens perspective." [96] The use of term multi-model clearly indicates model as its level of analysis, which does not convey the idea of using multiple methods within research or simulation model. This means that multi-model is not necessarily a multi-method approach. Holm et al. define multi-methodology as "...the combination of methodologies, often from different paradigms." [97] They discussed combination of hard positivistic method e.g. DES with interpretivistic soft method e.g. Soft Systems Methodology (SSM). provides a summary of the review. Different terms used to convey meaning that pertains to the use of more than a single method can create ambiguity. The presented literature demonstrated the need for more consistency in using different terms, because they may or may not mean the same thing. The problematic situation of the lack of agreed upon terms displayed above is analyzed in Chapter 3, also proposing definition of the multi-method M&S approach and its derivative terms.

Table 1. Summary of reviewed terms.

Author(s)	Terms used										
	Multi-methodology	Multi-method	Multi-paradigm	Multi-model[ing]	Hybrid	Mixed-method	Cross-paradigm	Multi-formalism	Method	Paradigm	Methodology
Balaban and Hester [10]		✓				✓			✓	✓	✓
Baoding and Yian-Kui [98]					✓						
Baskent and Keles [99]					✓				✓	✓	
Behdani [63]									✓	✓	

Table 1 (cont.)

Author(s)	Terms used										
	Multi-methodology	Multi-method	Multi-paradigm	Multi-model[ing]	Hybrid	Mixed-method	Cross-paradigm	Multi-formalism	Method	Paradigm	Methodology
Birle et al. [100]					✓				✓		
Borshchev [31]		✓	✓						✓	✓	
Chahal [8]		✓			✓	✓			✓	✓	✓
Chahal et al. [101]		✓									
Choi et al. [84]					✓				✓	✓	
Crespo and Ruiz [32]			✓		✓					✓	
D'Ambrosio [102]					✓				✓		
Donzelli and Iazeolla [103]					✓				✓		
Fishwick [96]				✓					✓		
Glazner [35]	✓				✓					✓	✓
Hassan et al. [21]										✓	
Heath et al. [29]					✓		✓		✓	✓	✓
Helal [51]					✓				✓		✓
Henzinger [87]					✓						
Hester and Tolk [28]									✓	✓	✓
Holm et al. [97]	✓										✓
Kotiadis and Mingers [104]	✓		✓						✓	✓	✓
Lättilä et al. [30]			✓						✓	✓	✓

Table 1 (cont.)

Author(s)	Terms used										
	Multi-methodology	Multi-method	Multi-paradigm	Multi-model[ing]	Hybrid	Mixed-method	Cross-paradigm	Multi-formalism	Method	Paradigm	Methodology
Lee et al. [94]					✓				✓	✓	✓
Levin and Levin [85]					✓					✓	
Lorenz and Jost [105]			✓						✓	✓	✓
Mao [106]		✓							✓		✓
Martin and Raffo [83]					✓					✓	
Mingers [26]	✓	✓							✓	✓	✓
Mustafee and Bischoff [107]									✓		
Osgood [86]					✓					✓	
Peña-Mora et al. [92]					✓						
Rabelo et al. [4], [82]					✓				✓		✓
Rabelo et al. [93]					✓				✓		✓
Rossiter and Bell [88]			✓	✓	✓					✓	
Setamanit et al. [89]					✓						
Sokolowski and Banks [41]						✓			✓	✓	✓
Swinerd and McNaught [23]	✓				✓						✓
Tolk [108]									✓		✓
Tolk et al. [33]			✓							✓	
Vangheluwe et al. [109]			✓								

Table 1 (cont.)

Author(s)	Terms used										
	Multi-methodology	Multi-method	Multi-paradigm	Multi-model[ing]	Hybrid	Mixed-method	Cross-paradigm	Multi-formalism	Method	Paradigm	Methodology
Venkateswaran and Son [90]					✓						✓
Wakeland et al. [91]					✓				✓	✓	
Zeigler et al. [95]				✓				✓			
Zulkepli et al. [49]					✓				✓	✓	✓

2.3 REVIEW OF RESEARCH GUIDELINES

Current research guidelines for multi-method approach are often method or domain (or both) specific [8, 23, 31, 35]. Because this view can constrain method selection and conceptualization flexibility more general and flexible guidelines for multi-method conceptualizations seems desirable. The choice of dimensions and criteria is important for deciding if the multi-method M&S approach should be used in a study, but deciding which criteria to choose can be problematic. This sections reviews research guidance in the area related to M&S multi-method approaches. It displays current approaches related to method formats, criteria, and dimensions. These elements provide some basic insights into how to choose method(s), which also influence choice between single and multi-method approaches. Furthermore, this also motivates discussion related to question if single and unique set of criteria and dimensions could be defined and used within multi-method M&S approach research guidelines.

2.3.1 Multi-method Simulation Model Structure

This subsection reviews method formats found in literature and generalizes them based on three UML concepts. Transitions of model component(s) toward atomic MFs could specify structural research characteristics that involve multiple modeling methods. In order to advance discussion related to multi-method research guidelines definition of method format is proposed first.

Definition 3

A method format (MF) is defined as a basic arrangement of method(s) and their relations overlaid with systems (or their components) and/or phenomena.

Chahal [8] propose three formats for SD and DES. Three formats pertain to combined SD+DES when methods required interacting. The hierarchical format could be used for analysis of the vertical interactions between different levels for "Setting strategic targets and evaluating their feasibility," "Simultaneous generation of strategic plan and operational schedules," and "Evaluation of resource allocation policies from operational perspective" [8]. The process performance–environment format could be used for "re-engineering of process or operations department [and] long-term consequences of interventions" [8]. Finally, the process–environment format was conceptualized with the purpose of "evaluating the interactions between environmental context and process activities; for example evaluating the impact of qualitative factors such as experience, motivation, schedule pressure etc. on process performance" [8]. DES was considered useful for capturing operational and processing view of systems and SD in representing either a strategic level or environmental factors.

Swinerd and McNaught [23] propose three common formats for SD and ABM (called hybrid design classes). An agent with rich internal structure is a format where SD method is used within an agent. In a stocked agents format, SD method is used to bound aggregate measure of agents. Finally, in the parameter with emergent behavior format, the aggregate measure of agents is used to influence a parameter within SD method.

Borshchev [31] discussed six common formats for combined SD, ABM, and DES methods (called architectures): 1) agents interacting with SD method, 2) SD method inside agents, 3) agents interacting with DES method, 4) DES method inside agents, 5) DES method interacting SD method, and 6) agents persistence thought their DES presence.

Each of the presented approaches provides insights, but they do not offer a general view for MFs. Please refer to Figure 4 during the following discussion about generalized relations A, B, and C. All formats for combining methods proposed by Chahal [8] can be generalized as a format in which methods are associated to exchange data within their interaction points during simulation. Formats proposed by Swinerd and McNaught [23] add a special case of association where embedding of one method into another takes place. Subsequently, Borshchev [31] specified beyond those two formats, adding the concept of dual existence of an actor within different methods.

It is not difficult to map those formats to UML notations [110]. In UML terms, relation A is the most general association where data exchange takes place. Relation B is a more specific association where aggregation describes how parts relate to the whole, components have their own identity, may be owned by more than a single aggregate, and their ownership can change over time. Relation C is a more specific form of relation B and restricts identity of components to the composite, so components must be referenced and owned by a unique composite. All three formats will be used as MFs in Section 3.3

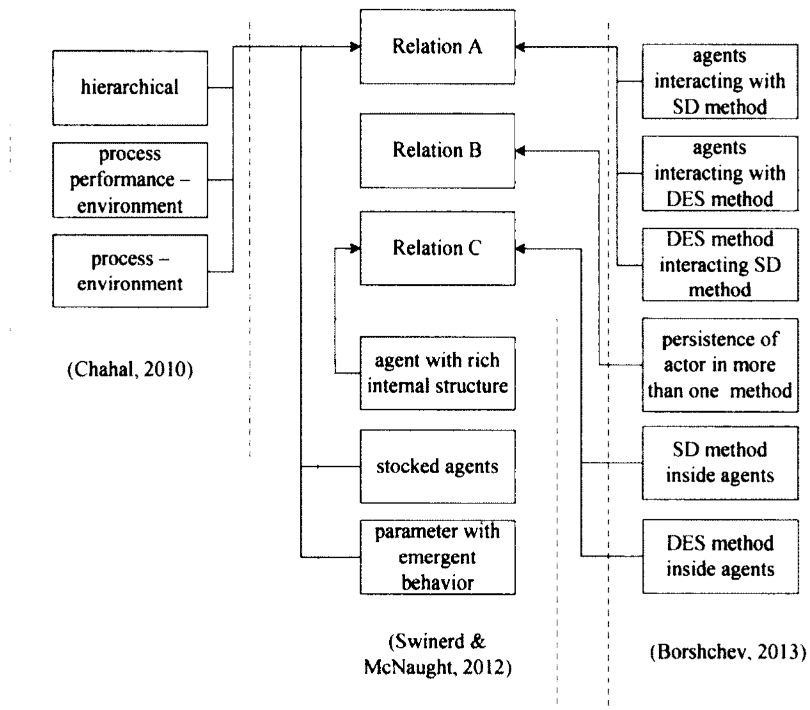

Figure 4. Generalized formats.

2.3.2 Research Design

2.3.2.1 Interaction between Methods

Chahal [8] defined three types of relationships between interaction points of methods. In direct replacement, the equivalent variables exist within both methods; values of one variable are replaced by equivalent variable defined using a different method. In aggregation/disaggregation type, the same conceptual elements are present in both methods, but do not exist at the same level. Finally, according to Chahal [8] in a causal type interaction points influence each other. The last concept seems more problematic as being described. If both values generated by each method affect each other this is a feedback between methods that consists of at least two one-directional interaction points. From the perspective of merit to using different methods, interaction point of simple variables will be one-directional unless the merits of using the values generated by two methods is not constant but changes during simulation run and is controlled by additional logic. If, on the other hand, causal type interaction pertains to data transformation, this is a one-directional data exchange where the meaning of data being passed between two methods has different dimensions e.g. space, time, probability. For a clarification, an interaction point where exchanged data have different dimensions within interacting methods will be called transformation. Another interaction type that should be added to this list is triggering and listening to conditions (including messages). These interactions generate asynchronous discrete events, which, in turn, can cause state transition or data exchange. If autonomous atomic structures such as ABM agent are implemented as discrete event versus discrete time (clock ticks), they are often associated with internal asynchronous events. The last important consideration to developing multi-method SD/DES given by Chahal [8] is mode of interaction: cyclic or parallel. In cyclic mode, models developed with different methods do not interact during run but the information has to be transported manually, while in parallel mode they interact during run time automatically. Level of coupling between methods depends on level of required interaction between methods.

2.3.2.2 Criteria for Method Choice

Selection of appropriate methods is one of the hardest problems in the M&S field; "…the hardest general problem in simulation is determining the exact method that one should use to create a model." [96] Practitioners need criteria that provide orientation for when to apply which method or methods [34]. The criteria for method choice should be able to expose methods' uniqueness in a particular context. This can help to select method(s) and a need for the use of multiple methods. The criteria for method choice often considered features of method, system, complexity, modeler, and a user in different contexts [34], [62], [8], [60], [111], [51], [112], and [63]. The development of criteria itself is a subjective endeavor, but can enrich research and justify context for methods chosen in the study, providing higher-level reasoning. For instance, validity limitations could be traced back to certain criteria not considered, avoiding pointless validation effort of implications arising from assumptions that cannot support representation of a given phenomenon.

Chahal [8] advised to disintegrate objectives allowing a modeler to determine if a multi-method approach is needed. If criteria for different objectives aligned with different methods, then simulation model with multiple methods should be employed. Chahal [8] extracted a set of the criteria based on relevant literature that could be used to choose between SD and DES using problem and system perspectives. These criteria are further extended with other methods (ABM, BN, FM, and SC) and are shown in Table 2. The *problem perspective* identifies purpose, importance of randomness, importance of interaction between individual entities, and required level of detail. The *system's perspective* distinguishes system view, complexity of importance, evolution over time, and control parameter. Through these criteria Chahal [8] assigned methods to disaggregated objectives. Both SD and DES were selected if they were interacting to achieve separate objectives. Table 3 presents criteria for selection between of SD, ABM, and DES methods provided by Behdani [63] and extended for additional methods like BN, FM, and SC.

Table 2. Criteria for selection between SD and DES proposed by Chahal [8] extended for ABM, BN, FM, and SC.

Method	DES	SD	ABM	BN	FM	SC
Criteria	Problem perspective					
Purpose	Decision optimization, prediction and comparison	Policy making, overall understanding	Decision optimization, prediction and comparison, Policy making, overall understanding	Policy making, overall understanding, inference	Policy making, overall understanding, description	Logic description
Importance of randomness	High	Low	High	High	High	High
Importance of interaction between individual entities	High	Low	High	Low	Low	Low
Required level of resolution	Detailed microscopic view	Aggregate, high level	Detailed microscopic view and aggregate, high level	Aggregate, high level	Aggregate, high level	Aggregate, high level
	System' perspective					
System view	Detailed microscopic view	Holistic telescopic view	Detailed microscopic view and holistic telescopic view	Holistic telescopic view	Holistic telescopic view	Holistic telescopic view
Complexity of importance	Detail complexity	Dynamic complexity	Detail complexity and dynamic complexity	Conditional complexity	Fuzzy complexity	Logic complexity
Evolution over time	Discontinuous event based	Continuous	Discontinuous event based and continuous	Discontinuous event based	Discontinuous event based	Discontinuous event based
Control parameter	Holding (queues)	Rates (flows)	Population, agents	Node, states, connections	Membership function	States, transitions

Behdani [63] characterized SD, DES, and ABM methods in accordance to their ability to represent complexity at micro and macro levels. These two levels are further divided into criteria. Because criteria proposed by Behdani [63] are concerned with levels of complexity, they may provide additional value for determination of the need for multi method M&S to represent social phenomena beyond criteria selected by Chahal [8].

Robinson [113] provided guidelines for a conceptual modeling and proposed a set of factors to assess model meets requirements. These factors can be divided into four groups as shown in Table 4. The groups were adapted to describe method context as follows. Conceptual validity (CV) is defined as perception, on behave of modeler(s) that the developed with proposed method(s) component can be at sufficient accuracy for the purpose.

Table 3. Criteria for selection between DES, SD, and ABM proposed by Behdani [63] extended for BN, FM, and SC.

Level	Method	DES	SD	ABM	BN	FM	SC
Micro level complexity	Numerousness and heterogeneity	Distinctive and heterogeneous entities in the technical level	No distinctive entities, working with average system observables (homogenous entities)	Distinctive and heterogeneous entities in both technical and social level	No distinctive entities, working with probabilistic system observables (homogenous entities)	No distinctive entities; working with fuzzy system observables (homogenous entities)	No distinctive entities, working with state and transition view of system observables (homogenous entities)
	Local Interactions	Interactions in technical Level	Average value for interactions	Interactions in both social and technical level	Probabilistic value for interactions	Fuzzy value for interactions	Interaction between statcharts
	Nestedness	Not usually presented	Hard to present	Straightforward to present	Hard to present	Hard to present	Straightforward to present
	Adaptiveness	No Adaptiveness at individual level	No Adaptiveness at individual Level	Adaptiveness as agent property	No Adaptiveness at individual level	No Adaptiveness at individual Level	No Adaptiveness at individual level
Macro level complexity	Emergence	Debatable because of pre-designed system properties	Debatable because of lack of modeling more than one system level	Capable to capture because of modeling system in two distinctive levels	Not present	Not present	Not present
	Self-organization	Hard to capture due to lack of modeling the individual decision making	Hard to capture due to lack of modeling the individual decision making	Capable to capture because of modeling autonomous agents	Hard to capture due to lack of modeling the individual decision making	Hard to capture due to lack of modeling the individual decision making	Hard to capture due to lack of modeling the individual decision making
	Co-evolution	Hard to capture because processes are fixed	Hard to capture because system structure is fixed	Capable to capture because network structure is modified by agents interactions	Hard to capture because networks are fixed	Hard to capture because membership functions are fixed	Hard to capture because statecharts are fixed
	Path dependency	Debatable because of no explicit consideration of history to determine future state	Debatable because of no explicit consideration of history to determine future state	Capable to capture because current and future state can be explicitly defined based on system history	Debatable because of no explicit consideration of history to determine future state of network	Debatable because of no explicit consideration of history to determine future membership function	Capable to capture because current and future state can be explicitly defined based on system history

Conceptual credibility (CC) is defined as perception, on behave of client(s) that the developed with proposed method(s) component can be at sufficient accuracy for the purpose. Conceptual feasibility (CF) is defined as perception, on behave of modeler(s) and client(s) that component developed with proposed method(s) can be useful during experimentation phase and for a possible reuse.

Table 4. Factor for model requirements assessment based on Robinson [113] adapted to methods.

Groups	Description	Factors
Conceptual validity	Perception, on behalf of a modeler that the developed with proposed method(s) component can be at sufficient accuracy for stimulating for the purpose	Accuracy
Conceptual credibility	Perception, on behalf of the clients that the developed with proposed method(s) component can be at sufficient accuracy for the purpose	Accuracy
Conceptual utility	Perception, on behalf of modeler and the clients that component developed with proposed method(s) can be useful during experimentation phase and for a possible reuse	Ease to use and flexibility
		Run-speed
		Visual display
		Reuse
Conceptual feasibility	Perception, on behalf of modeler and the clients that component can be developed with proposed method(s) into a simulation model with resource including skills, data, and time available	Resources and skills
		Data
		Time

Finally, conceptual utility (CU) is defined as perception, on behave of modeler(s) and the client(s) that component can be developed with proposed method(s) into computer model with resource, data, and time available.

Table 5 presents a set of proposed criteria by the author in relation to following methods: DES, SD, ABM, BN, FM, and SC. This set was assembled mostly based on analysis of criteria provided by [34], [62], [8], [60], [111], [51], [112], [63], and the author's M&S practical experience. Criteria for methods' choice should display unique characteristics of considered methods to distinguish their merits. The work to assemble the above criteria was motivated by the initial belief that a unique set of criteria could be created. For instance, Table 6 provides a second version of criteria that adds Petri Nets (PN) method.

Table 5. Proposed criteria for method choice considering DES, SD, ABM, BN, FM, and SC.

Method / Criteria	DES	SD	ABM	BN	FM	SC
Representation of individual behavior as part of a larger system	Correlations of passive entities create view of system	Limited due to structural constraints	Can focus both on internal and external behavior for passive, reactive and proactive agents, their internal behavior, and correlations and interactions with other agents	Limited due to structural constraints	Limited due to structural constraints	Limited due to structural constraints
Ability to operate on aggregates	Possible, but often limited	Holistic view through causality and feedback	Desirable for multilevel models , and used during experimentation	Network nodes as aggregates	Fuzzy view or perspective about aggregated system	States as aggregates
Ability to handle uncertainty	Within constant structure, usually to represent time dimension or routing options	Not as its core but possible within predefined structure	At structural and behavioral level	Bayesian perspective within predefined structure	Fuzzy set concept	Transitions between states as probabilities
Interaction	Limited to correlations of reactive entities	Based on causality, limited by predefined structure	Interactions between agents, environment, and between elements	Probabilistic value for interactions	Fuzzy assessment of interaction level	Interaction between statecharts
Unique features	Effective process description, visual animation	Ease to construct feedback loops	Agents types can be designed to different levels of specification	Unique approach to inference	Fuzzy set perspective	Combines different triggering options condition, rate, timeout, message
Form of descriptive usage	Predictive analysis based on empirical input from the system	Calibration, stylized facts	Calibration, stylized facts, predictive analysis based on empirical input from the system	Utilization of empirical data to generate prior knowledge CPTs and inferences	Utilization of empirical data to describe fuzziness	Logic of mechanisms or systems
Form of theoretical usage	Process concept testing	Causal structure evaluation	Individuality and interaction based	Conceptual inference	Understanding if phenomenon have fuzzy properties	Logic of phenomena
Relevant to represent complexity type	Limited to structure and input	Time complexity, but limited to structure and input	Both structural and behavioral	Limited to structure and input propagation	Limited to structure and input	Logical complexity limited to structure and types of transitions

Table 6. Proposed criteria for method choice considering DES, SD, ABM, BN, FM, and SC.

Criteria/Method	**DES**	**SD**	**ABM**	**BN**	**FM**	**SC**	**PN**
Representation of individual behavior as part of a larger system	significant	minimal	essential	none	none	none	moderate
Ability to operate on aggregates	none	essential	significant	essential	moderate	moderate	moderate
Ability to handle uncertainty	essential	minimal	significant	significant	significant	significant	significant
Interaction	significant	none	essential	none	none	moderate	significant
Descriptive usage	essential	minimal	moderate	moderate	moderate	significant	significant
Theoretical usage	minimal	essential	essential	significant	significant	moderate	moderate
Emergence	minimal	minimal	essential	none	none	none	moderate
Ability to represent active behavior	none	none	essential	none	none	significant	significant

The criteria were estimated using scale ranging from essential, through significant, moderate, minimal, to none. The perspective about finding a unique set of criteria has changed during this research. Unfortunately, the ability to find a unique set of criteria for all methods may not be always possible. The criteria may not be applicable in the context of methods examined nor cover sufficiently the considerations deliberated as vital to the modeling effort.

Moreover, if social phenomena are present, their subjective character can complicate matching criteria to objectives. The first question is, if considering division of objectives can really lead to sufficiently granular options directing the choice of using multiple methods. The framework proposed by Chahal [8] specifies that multi-method is chosen when different methods address different objectives, and there is a strong relationship between parts represented with different methods. It is problematic to use objectives or questions as a unit of analysis during selection of methods. For instance, what if a question or an objective cannot or should not be divided, but still requires a multi-method approach? The approach proposed by Chahal [8] can be useful, but may not work in every case. The set of criteria assembled above as well as any other set presented or cited before can also be useful in analyzing methods' choices, but may not necessarily include all considered phenomena and system's contexts.

Because the criteria for method choice originate from human deliberation and change with scientific advancements, a human interpretation about method choice seems to be the ultimate stage. A set of unique criteria for method selection could limit methodological ambiguity, but there can be no unique perspective on research that covers all methods considered, problems, phenomena, and systems. The aspect of a set of methods considered is also important and problematic. Because perception on usefulness and applicability of simulation methods to different kind of problems evolves as new practices and functionalities are established and implemented into software criteria should naturally adapt to reflect this progress. On the other hand, work toward a unique set of criteria for method choice should not be discouraged by their current and future limitations, but propelled by that fact, allowing for subsequent improvements of multi-method scientific practice.

2.3.2.3 Proposed Research Dimensions

Balaban and Hester [10] have extended work of Lorenz and Jost [34] beyond object, method, and purpose as three main dimensions. A graphical representation shown in Figure 5 identifies high-level dimensions discussed during this literature review. The following extensions are considered:

- Human dimension pertains to subjectivity of method choice
- Origination of study is related to managerial and organizational circumstances affecting human dimension. This can include relation with a sponsor and what project constraints we have e.g. software, methods

A decision to employ multi-method should involve projecting the simulation study steps on the reasons derived during analysis of M&S literature conducted in Section 2.2.1.

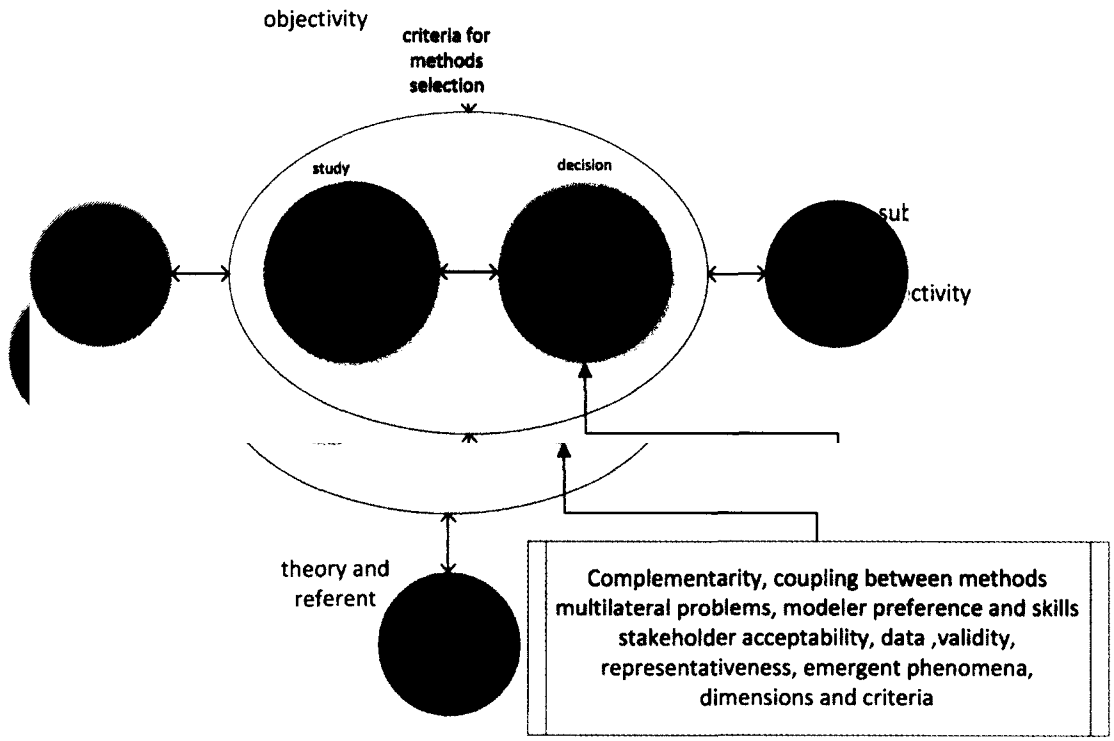

Figure 5. General research dimensions relevant to multi-method M&S [10].

Figure 6 highlights specific important links between dimensions. It is desirable that criteria for method selection expose methods' uniqueness in the context of all other dimensions specified. It is pointed out that considered criteria originate from human deliberation so they are also subjective. On the other hand, they could permit better understanding of subjectivity by disclosing deliberation given during research design, leading to more objectivity.

The improved reasoning for employing the multi-method M&S approach should include process for method(s) selection that identifies the human subjectivity and discloses it. This can also improve research credibility because of the ability to evaluate layers of considerations given to the dimensions.

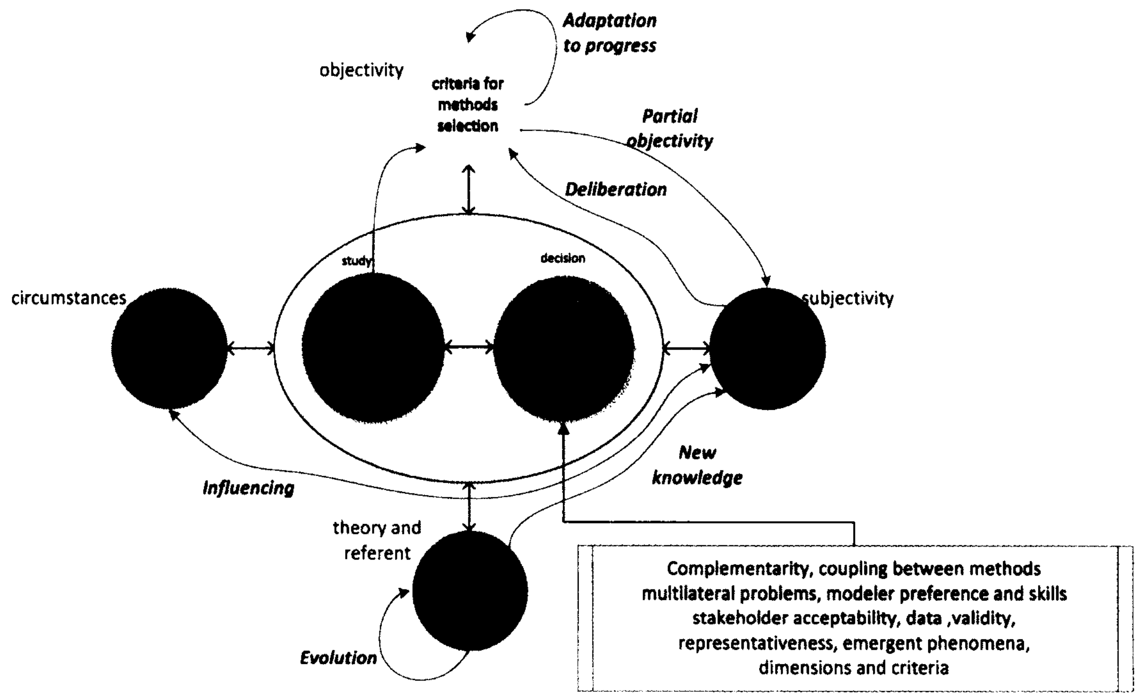

Figure 6. Important relationship between dimensions.

Unfortunately, large scopes within each of the dimension and multiple dependencies between the dimensions limits feasibility to consider all their knowledge bases in relation to the objectivity of research. First, how should be criteria developed? It is evident throughout the M&S domain that many scholars attempted to devise a method-specific single set of criteria [8, 34, 51, 60, 62, 63, 111, 112]. On the other hand, it is problematic to assume that the same set of criteria is applicable the same way for every scientific community, every study, every set of methods considered, and every modeler. Criteria developed once in the past may not be adequate because of evolution of the dimensions. A classical approach to criteria can have limitations. Considered systems and/or phenomena, origination concerns, and human contexts evolve, and with them our limited, yet increasing, knowledge how to conduct research. Careless adoption of criteria from the past research is not objective unless those criteria are universal and true in every context, which as discussed above, would not be the case unless the world had stopped and all possible useful knowledge was formal.

A large and growing number of M&S methods impose additional implications. Perception about the usefulness and applicability of simulation methods to different purposes evolves as new practices and functionalities are established and implemented into software. Criteria should naturally adapt to take advantage of this progress.

It may be especially problematic to devise criteria related to the human dimension because of its subjectivity. Preferences of modelers to use particular methods often relate to their knowledge, modeling skills and various constraints, which, in the end, influence time to build a simulation model. Pragmatic considerations that often arise within origination dimension may often be useful. For instance, general factors like flexibility, run-speed, visual display, and reuse can impose requirements helpful to plan modeling tasks and discover feasibility constraints [113]. Moreover, the origination of a study can shape the character of a study toward expansion, comparison, or both. Tight dependencies between researcher and stakeholders are possible. Both stakeholders and researchers often follow rules, policies, and organizational objectives, which can affect the research process. For instance, a researcher may have to follow rules related to method(s) that should be considered, or must follow directives about the level of exploration within a study. Often, the decision about a research approach is made by both stakeholders and researchers, but the level of independence of the researcher can influence research objectivity. For instance, the scope of a study can be affected by preselecting method(s) and assertions about knowledge bases. If a researcher is independent, the research process is more internally controlled and affected by knowledge available, the initial researcher's knowledge, which also consists of knowledge of method(s), researcher effort, and resources.

The guidance for designing a study without limiting it to a particular methods and a set of criteria could enhance multi-method M&S approach. Because there are already many M&S methods, and the list will most likely grow, it seems challenging to develop one specific set of agreed upon criteria addressing all methods, formats, phenomena, and system contexts. A more general process for the development of multi-method research seems to be a valuable approach. For instance, a project/study would start with considerations given to its origination, propelling the design of a simulation study, including both reuse and/or development of criteria for selection of method(s) based on research question(s). A conceptual model would emerge along the process with the considerations given to different methods, and according to criteria chosen and developed. This would allow for a balanced systemic approach, yet sufficient modeling freedom during conceptualization that do not constrain views up front, and gradually considers the use of different methods to describe system and phenomena. The proposed dimensions influence reasoning about the use of a multi-method approach and provide a theoretical path for the current and future research related to this topic. The scientific goal of achieving objectivity by supporting a human dimension with a set of criteria that satisfy the study aims is an idealized case scenario worthwhile of pursuing.

2.3.3 Research Evaluation: Objectivity, Quality, Legitimacy, and Validity

A general view on objectivity introduces the topic, and it is followed by a quality and legitimacy of a study with multiple methods employed. Validity concepts in relation to M&S filed follow next. This section ends with a discussion about evaluation of multi-method M&S approach.

2.3.3.1 Research Quality and Legitimacy

Post-positivists' view on objectivity accepts that theories, values and knowledge of the scientist could influence what is perceived, but with minimization of an inquirer and methodological bias in the quest for truth [114]. The feminist tradition of objectivity emphasizes challenging of prevailing but false assumptions [115]. Democratic objectivity evaluation criteria reflect ideals of fairness and equity, advancing the well-being of the most underserved and giving voice to all legitimate perspectives and interests rather than privileged ones [116]. Philosophical views such as interpretivism and constructivism present objectivity as unattainable and negotiated through dialog, hence, subjective. Defensible knowledge could be attained by closeness, engagement, and sufficient time to understand different perspectives.

Judging the quality of a model from a social science perspective could be conducted with many criteria, e.g., data representativeness, generality of inferences, richness of samples, contextual meaningfulness of inquiry, actionability of an inquiry and knowledge generated [27]. Tetlock pointed out that: "political psychology poses greater-then-usual scientific challenges that require us to model the mindsets not just of research participants but of the researchers themselves." [117] Likewise, during the assessment of the quality of a simulation model, especially representation of social phenomena, we should be concerned with not only the simulation model but also the mindset of the scientists who build and interpret the model. It seems intuitive that the quality and validity of a study depends on a person conducting the research, and there could be an immense difference in the results even if multiple scientists use the same paradigms and methodology. Similarly, judgment of both quality and validity depends on personal views and knowledge of evaluators. Qualitative assessment of the merits of a simulation model capturing social phenomena poses a challenge because it can be very subjective. Greene [27] indicated that with only one paradigm there is one set of criteria for warranting the use of the method and the study's outcome, which makes this process simpler comparing to a mixed method approach.

The study can often be judged quantitatively based on accuracy, reliability, and precision of results but qualitative measures are also important. Tashakkori and Teddlie [74] provide the following criteria for inference of mixed methods' quality:

- Conceptual consistence is a degree of agreement between inferences and between knowledge and theory that pertains to the inferences.
- Interpretative agreement is a degree of consistency of interpretations between, e.g., scholars and Subject Matter Experts (SME).
- Interpretive distinctiveness is a degree of difference between inference and alternative possible interpretations; meaning rival explanations are ruled out.

Greene [27] provided four elements to consider for the warranting quality of inferences of a mixed method approach:

- Data choice for inferences should be assessed based on how different paradigms handle data, because different methods facilitate use of different data. This can allow for minimizing prejudice and bias, and maximizing data merit.

- Criteria of methodological assessment should be utilized in an integrative, coordinated and synthetized way, as integrated judgments based on inquiry findings from multiple paradigms views and perspectives. Conflicts, contrasts, and tensions between findings from different methods are welcomed aspect, which can provide for an additional insight.
- Persuasive power of deliberation, emphasizing coherence, expansiveness, interpretive insight, relevance, rhetorical force, appeal, and texture of argument can be used as a measure of quality. Even if the different views could be considered adversary, they could engage possible dissonance in judgments yielded by multiple criteria leading to a dialog, and in fact contribute to understanding.
- Determination of additional insight and understanding that is reached with mixed method design that is not attained within a single method.

Different types of legitimacy that can provide insight into mixed method validation were proposed by Onwuegbuzie and Johnson [118]:

- Sample integration should yield quality meta-inferences and valid statistical generalizations.
- Different perspectives should be utilized, e.g., insider's and observer's views.
- Weakness mitigation could be accomplished when the weakness from one approach is compensated by the advantages from the other approach.
- Potential problems based on sequence and structure of methods utilization should be considered and minimized.

- Mapping of data, such as quantitizing or qualitizing, should yield quality meta-inferences.
- The researcher uses mixed methods based on his/her beliefs that methods are combined and blended with sufficient epistemological, ontological, axiological, methodological, and rhetorical justification.
- Commensurability should be reflected by utilization of mixed world-views based on the cognitive process of switching between methods and integration of scientist's perspectives.
- Multiple views of validity should be incorporated based on different validation approaches to different methods.
- Validity can be partially inferred based on how much the consumers of mixed methods research value the research results.

2.3.3.2 Validity in M&S

The Department of Defense defines validation as the process of determining the degree to which a model or simulation and their associated data are an accurate representation of the real world from the perspective of the intended uses of the model. Validation should answer the following questions:

- Did we build the right thing?
- Does our simulation do what it is supposed to do?

The validation of a simulation model is described in Law's textbook as the process of determining if a simulation model is an accurate representation of the system based on requirements specified for a particular problem [17]. The validation process divides the error of a simulation study into the validation error in translating system to a model and the output analysis error. If the first term can be called accuracy, the second align better with a definition of precision provided in [119]. Closely related to validation term is fidelity, which encompasses other more specific terms like accuracy, precision and resolution [119].

Accuracy is defined by Joint Committee for Guides in Metrology (JCGM) as a degree of closeness of the measurements of a quantity to that quantity's actual value [120]. Accuracy is defined in [119] as the degree to which a parameter or variable or set of parameters or variables within a model or simulation conform exactly to reality or to some chosen standard or referent This can be interpreted in a simulation validation study as a measure of closeness between a system and its simulation model output. However, the lack of accuracy between the system and its model can be subjective and does not imply directly its lack of validity, because this also depends on the context of the simulation use. Evaluation based on accuracy in the M&S validation study can also be subjective because there are neither a perfect nor a one hundred percent accurate representation of a system. Modeling is a process of abstracting elements of reality based on the purpose of a simulation model.

Harmon and Youngblood [121] have defined validation as the process of generating information in the quest for truth. They have discussed the risk of the validation process as dependent on the quality of information, which is based on truthfulness as an essential measure of validation of information. Moreover, they used objectivity, repeatability, timeliness, completeness, and accuracy as attributes of information quality. Grüne-Yanoff [122] argues that the full explanation cannot be fully supported because of the validation issues. For instance, social phenomena may not be directly observable, and are usually characterized by subjective empirical data and lack of full causal understanding. If simulation regenerates limited quality empirical data and lacks solid, theoretical grounds, then the simulation model cannot be considered a true explanation of social phenomena, but it may contribute to increased understanding. Balci [123] offers 15 Verification, Validation and Testing (VV&T) principles. First and foremost, it is crucial that VV&T must be conducted throughout all phases of a simulation study. He presented also and discussed main groups of VV&T techniques: informal, static, dynamic and formal, with multiple possible techniques within each group. Sterman [16] provides extensive discussion and guidance on conducting VV&T in his book about SD, which complements work of Balci [123] by providing insight on validation of theoretical models.

Rossiter et al. [124] extended the model-centered approach of McKelvey [125] to application in simulation studies. This perspective on scientific exploration via simulation models can be used as a high-level validation model (shown in Figure 7). Rossiter et al. [124] defined different types of adequacy testing that can be used in qualitative and quantitative assessment of simulation models that consists of both descriptive and theoretical components. Use of different methods in a simulation model may often be spurred by the need for modeling both more theoretical and more descriptive components.

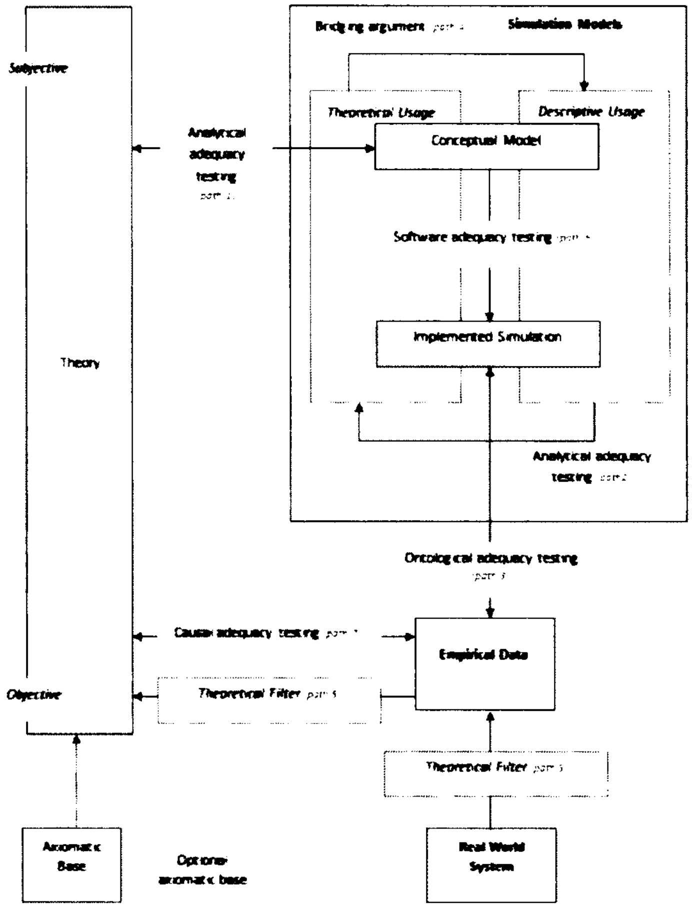

Figure 7. Validation of simulation models based on Rossiter et al. [124].

Assuming this, the approach provided by Rossiter et al. [124] can be applied as a useful validation construct. Following is a brief introduction and discussion about this validation model. *Analytical adequacy testing* (AAT) can be conducted as theory-versus-conceptual-model validation (path 1, causality) or as the validation of a theory with a descriptive simulation model (path 2, quantitative). Theoretical models are often built with a focus on exploration of theory and its consequences. Analytical adequacy relates to the model as a representation of theory. *Ontological adequacy testing* (OAT) is conducted by comparison of empirical data versus simulation model output (path 3). It focuses on determination of how closely observed phenomena is reproduced by a simulation model based on quantitative basis, but it does not imply validity on its own.

A descriptive model provides the researcher with information that a given model setting allows fitting empirical data with a certain degree, but does not necessarily imply its validity. Both descriptive and theoretical usages of a model can be useful for different purposes and at different stages of research, and both usages can require different methods. Exploration of various phenomena; technical, social, or mixed may require both theoretical and descriptive perspectives. A transition from theoretical to descriptive usage called *bridging argument* (path 4) can be described as formation of a hypothesis on how the real world works based on some theory. The opposite, conversion through path 2, indicates an approach to validation of the existing theoretical context via a descriptive model (real world setting). The combination of these two approaches can create a loop allowing for iterative refinement and generalization of a phenomenon representation, which in turn can lead to a better understanding, theoretical contributions, and decisions. *Theoretical filter* (path 5) indicates subjectivity in choice of empirical data. *Software adequacy testing* is the verification process of a model translation from a conceptual to the computational form; here the methods used and their interaction are an important consideration. *Causal adequacy testing* (CAT) provides additional support to AAT by analyzing and comparison of theoretical aspects of a model with the real world.

2.3.3.3 Discussion

Many of the items in the lists provided in Section 2.4.3.1 have their analogy in the M&S domain, indicating ties between these fields. A higher legitimacy [118] based on combined insider and observer's views can be facilitated within a single simulation model by combining different methods e.g. SD as an observer view, and ABM as insider view. Complementarity of methods can also be mapped to improved legitimacy because weaknesses mitigation from one approach can be compensated by the advantages from the other approaches [8, 118]. Improved degree of 'interpretative agreement' with the application of multi-method suggests that the use of multiple methods has potential to improve quality of simulation model [53, 55, 74]. Flexibility of data mapping can be considered as a factor for minimizing bias and maximizing merit, hence, improvement of quality and legitimacy [118]. Similar reasoning is also given as a purpose for the use of multi-method by a member of simulation community [27, 30, 118]. Value added related to 'initiation' [27] is closely related to 'emergent phenomena' [10]. Both can be generalized as inclusion of different methods into mental process during research design, communication, modeling and experimentation. This aspect is difficult to measure, hence the research investigating influence of 'thinking in multi-method way' on cognitive abilities translated into the quality of research or validity of simulation model would be desirable. If multiple methods allowed for a better access to different levels of analysis, translated into better mapping of the researched phenomena [56], this could have positive effects on 'conceptual consistence' and 'interpretative distinctiveness', factors of research quality described by Tashakkori and Teddlie [74]. On the other hand, improvements to 'conceptual consistence' and 'interpretative distinctiveness' have limits related to human capabilities. Complex models can extend structure, behavior, and experimental framework beyond human comprehensibility, hence "depth" (conceptual consistence) and "breadth" (interpretative distinctiveness) of a simulation model should be balanced.

Moreover, concerns related to methods interaction, structure, criteria for validity of all methods included in a simulation model should be considered. Finally, a measure of usefulness related to insight and understanding that is reached with a simulation study that employs multiple methods may not be always be a good indication of quality or validity of a simulation model, but can be valuable regardless of that, similarly as in mixed method study [27].

The challenge of evaluation of combined elements represented by different methods within one model in large relates to the difficulty with synthesizing validation requirements for different modeling methods. For instance, common use of DES and SD methods may lead to different perspectives on what a valid model should look like when the methods are combined. Most likely combination of standards should be used, although some may be contradictory. For instance, assumption of input being independent, and identically distributed (IID) may be violated if SD method controls creation of entities while the feedback loop of SD depends on DES process itself. On the other hand, if the error between referent of a real or imaginary system and simulation output had decreased this would have increased credibility of simulation model at the cost of violation of statistical assumptions. When adding DES to SD this would have most likely extended validation requirements set by a common SD approach [16], because results could have varied across replications. This may indicate that the level of validity might depend on the standards used, which in turn should define validation limitations and the proper context of simulation model usage. The key aspect to consider when specifying validation requirements for a multi-method simulation model is the fulfillment of the purpose with a sound and holistic perspective on a study. Verification and Validation (V&V) processes expose scientists' skills about a subject studied and M&S skills for representation of researched system of phenomena.

2.4 SUMMARY

After a brief introduction, Section 2.2.1 provided a review of relevant M&S literature and analysis of the rationale for the use of multi-method. The justifications for using multi-method M&S that were found were characterized as complementarity of methods, multilateral problems, modeler preference and skills, stakeholder acceptability, data availability, validity, unique representation, emergent phenomena, and more generally, dimensions and criteria. Ideally, the purposefulness of multi-method M&S should be based on more solid theoretical base propelled by guidelines that support the decision to use a multi-method M&S approach. In Section 2.2.2, justifications for mixing method in empirical social science field according to Greene [27] were projected onto the M&S domain and analyzed, allowing for deeper understanding of complementarity through human, model and method lenses, and leading to a proposed definition of complementarity of methods. Finally, levels of triangulation were analyzed in the context of M&S study, including multi-method cases.

Presented in Section 2.3, the literature demonstrated the need for more consistency in using different terms related to approaches that use more than a single method.

Section 2.4 generalized multi-method simulation model structure based on reviewed MFs. Section 2.4.2.1 discussed and extended interaction points between methods. Discussion in Section 2.4.2.2 led to conclusions that the ability to find clear boundaries of criteria for method selection may not always be possible making translation of phenomena and system, with consideration to methods' complementary set of assumptions, into required simulation model difficult and possibly subjective. Section 2.4.2.3 identified high-level dimensions for multi-method M&S approach. Section 2.4.3 provided a theoretical background and insight into evaluation of quality and validity of the of multi-method M&S research. The reviewed and analyzed literature provides foundations for the developments undertaken in Chapter 3. The next chapter will develop and propose a theoretical basis of multi-method M&S approach to fulfil the research gap.

CHAPTER 3

THEORETICAL BASIS OF MULTI-METHOD M&S APPROACH

This section is divided into three main parts. The first discusses and analyzes the most important to this work's terms in order to provide basis for proposed definition of multi-method M&S approach and its derivative terms. The second uses concepts of falsifiability [45], commensurability, complementarity and triangulation of methods to search for principles of multi-method M&S approach. The last part proposes method formats.

3.1 PROPOSED DEFINITION OF TERMS

Different terms, definitions, and knowledge exist within branches of the multidisciplinary M&S field. This may be due to a variety of applicable M&S methods in different domains [28]. Sometimes a single term is used, sometime multiple terms are used with a single piece of work as synonyms solely for readability purposes, and still other times, different meanings of those terms are intended. In many cases, the purpose of using multiple terms is difficult to determine. Presented in Section 2.3, the literature demonstrated the need for more consistency in using different terms. This section analyzes the most important to this work terms. The first part takes pragmatic philosophical view to provide basis for proposed in part two definition of multi-method M&S approach and its derivative terms.

3.1.1 Pragmatic Stance on Terms

Three ways of looking at the term *methodology* are presented by Mingers [11]. The first refers to methodology as a study of methods [126, 127]. The second meaning is the most specific and pertains to a particular research study (see [73]), while the third one is a generalization of the second. Using the word "methodology" as a combination of methods or techniques is more general and less prescriptive but "it can be difficult to precisely delineate the boundaries between method and methodology." [11] He also states that the use of the terms methodology and multi-methodology in the United Kingdom are synonymous with method and multi-method, respectively.

Mingers defines the term *paradigm* as "a construct that specifies a general set of philosophical assumptions covering, for example, ontology (what is assumed to exist), epistemology (the nature of valid knowledge), ethics or axiology (what is valued or considered right), and methodology." [11] For example, research paradigms in social science are positivism, post positivism, interpretivism, and pragmatism. These were characterized through the dimensions of fundamental beliefs that affect ways to conduct research, i.e., ontology, epistemology, axiology, and methodology [126]. Moreover, Mingers argues that "the paradigm concept is useful as a shorthand for a particular constellation of assumptions, theories, and methods, but it is purely a heuristic device." [11] This means that we can "detach research methods (and perhaps even methodologies) from a paradigm and use them, *critically and knowledgeably,* within a context that makes different assumptions." [11] This concept is examined by Lorenz and Jost [34], who analyzed assumptions of DES, SD and, ABM and differences between them. The authors leave the modeler with two options: first, to use methods within a single established methodology, or second, to combine methods within methodologies of different paradigms. This can be pictured as a possibility to combine methods between different paradigms. Lorenz and Jost add that a paradigm "...is characterized by the fact that it is to a large extent not questioned within its scientific community." [34] This means that the assumption of whether a method becomes a paradigm can be questioned by an individual's personal set of beliefs, but what really matters is that the supporting community agrees upon terms and definitions and shares fundamental beliefs that affect ways of conducting research.

M&S theory and practice echelons need to provide more guidelines on what should be considered a paradigm and why and whether this term is even correct to convey what is meant. Considering ABM as a paradigm can be problematic because it has not reached the point of sufficient agreement about its epistemological and axiological bases as compared to SD and DES. On the other hand, it would be easier to assume SD and DES as paradigms because these methods have a long tradition and dedicated development communities, e.g., System Dynamics Society and SIGSIM PADS (recently extended to other areas), respectively. When looking more formally at methods Zeigler et al. [95] distinguish three main formalisms: discrete event system specification (DEVS), discrete time system specification (DTSS), and differential equation system specification (DESS). They are used to provide general dynamic system formalism. Moreover, the authors give examples of SD and Bond Graph methods as sub-formalisms of DESS, and Petri Nets and Statecharts as sub-formalisms of DEVS. Combination of different methods is called multi-formalism modeling. Within this theoretical, formal view, DEVS, DESS, and DTSS could be considered M&S paradigms, while SD, Bond

Graph, Statecharts, and Petri Nets would be sub-paradigms or methods. On the other hand, Fishwick [96] discounts continuous and discrete time simulations as main categories and focuses on distinctions that pertain to modeling, i.e. conceptual, declarative, functional, constraint, and spatial. The groups provide different ways to categorize simulation methods as compared to Zeigler et al. [95]. Fishwick [96] indicates that formal specification can be very useful to convey M&S bases and it is mathematically pleasing, but the use of formalisms by scientists and modelers is less intuitive and can be even deceiving. The inclusive character of the word "method" versus the philosophical-assumption-constrained "paradigm" can be beneficial in this context. Additionally, the use of the word "multi-formalism" or "multi-specification," grounded in predicate logic or a mathematical theory, can be less intuitive to modelers and scientists.

Many methods, e.g., Bayesian Networks, Neural Networks, and Fuzzy Methods, can be complementary within simulation-based methods, and should not be excluded during conceptualization. It is important to point at the inclusive character of the word method as a unit of consideration in description of a multi-method M&S approach. For instance, because methods evolve, the word multi-method seems more inclusive and specific over multi-paradigm because the considered method may not be established in the M&S field as a paradigm, yet it can contribute its desirable unique characteristic. Besides, the unique paradigmatic perspectives are not always desirable, but only some methods within an M&S paradigm are complementary and may not change the perspective of the original complemented part. In this case, we can draw a relation that a paradigm is or has one or more methods, while a method is not necessarily a paradigm.

Sokolowski and Banks [41] refer to the combination quantitative and qualitative data gathering as mixed-method research pointing at M&S for the quantitative part. When considering M&S as a multi-disciplinary field built from different domains, pragmatism seems the most appropriate paradigm to follow because it integrates quantitative methods (simulation model) and often-qualitative methods (conceptual model). Expansion of simulation research to other domains of science, e.g., social sciences, can be a little confusing if methods are called paradigms, because the word paradigm was used there at a different, higher-level. For instance, if M&S is a part of mixed method research that exists within a pragmatic paradigm, naming SD and DES paradigms within the same piece of work can be confusing. Clearly, some sort of structure to terminology is needed to avoid using the same terms at different levels.

Mingers [26] uses the term multi-method in reference to a general plurality of methods and techniques, both qualitative and quantitative, and within a real-world intervention. He pointed at many logical possibilities about whether methods come from different paradigms, are combined within the same intervention, and if methods may be combined. This work adopts the position on paradigms proposed by Mingers [11], which allows us to remove constraints related to paradigms at the level of methods, while assuming a pragmatic paradigm within the whole M&S domain. This directs the focus on M&S methods, whether taken from an established M&S method, often called a paradigm, or not. Obviously, commensurability of methods is not assumed in all cases, because not all methods can be used together. This also depends on method computability and the study context itself. Reducing level of analysis from a multi-paradigm to multi-method M&S approach allows it to be more flexible, specific and inclusive.

3.1.2 Definitions

The following set of definitions provides a starting point for multi-method M&S approach theory development [128]. Figure 8 illustrates dependencies between these terms.

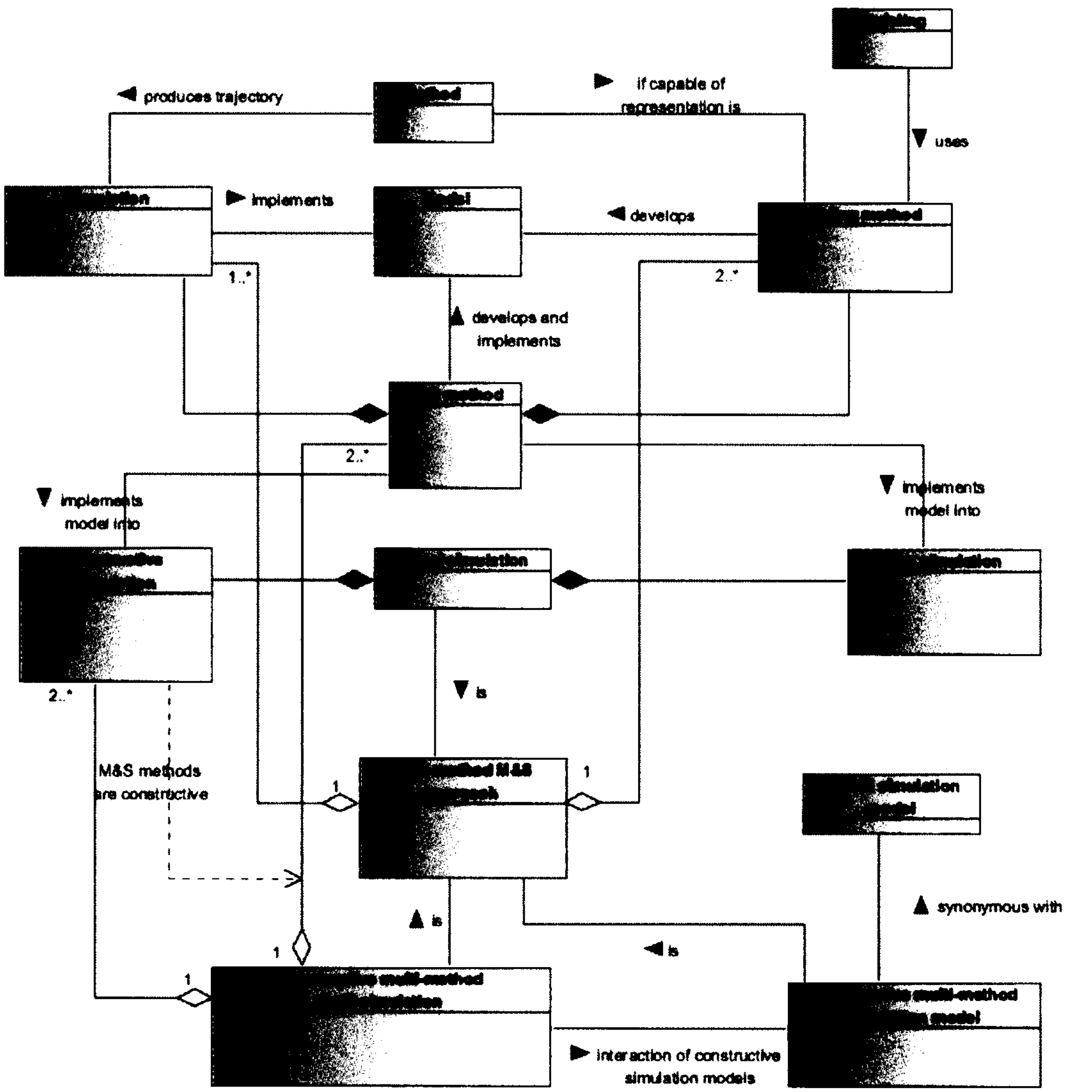

Figure 8. Dependencies between definitions and relevant terms [128].

Definition 4

*A **methodology** is the ideological and theoretical foundation of a method* [126].

Implications:

Methodology as a model to conduct a research within the context of a particular paradigm is closer to research practice than philosophical concepts found in paradigms. It can properly refer to the theoretical analysis of the methods appropriate to a field of study or to the body of methods and principles particular to a branch of knowledge. It does not set out to provide solutions but offers a theoretical underpinning for understanding which method, or which set of methods, can be useful to a specific case.

Definition 5

*A **method** is a systematic procedure, technique, or mode of inquiry employed by or proper to a particular discipline* [129].

Implications:

This is a broad and general definition providing a starting point for the discussion related to the use of multiple methods. It is pointed out that methods can be more or less specific. For instance, they can be characterized by a systematic way of instruction or representation.

Definition 6

*A **model** is a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process* [130].

Examples:

An example of a physical model is a plastic car. A mathematical model can consist of mathematical symbols and their relationships, e.g., as in mathematical equations. A logical model can consist of a set of interdependent, logical statements.

Definition 7

Modeling *is a mental process that, combined with a modeling method, is used to develop a model.*

Implications:

It is noted that modeling as a mental process carries a notion of a method itself since it can be described as a process. For instance, Hester and Tolk [28] defined modeling as the process of abstracting, theorizing, and capturing the resulting concepts and relations in a conceptual model. Modeling pertains to abstraction where systems or phenomena are mental projections made by a modeler related to the highest level of human consciousness [131]. The projections become models by using a modeling method.

Definition 8

A ***modeling method*** *is a method capable of representation.*

Implications:

There are different modeling methods depending on a type of model being developed and its purpose. For instance developing a scaled physical model could consist of a process of building and assembly of components. Building a 3D constructive model would include development of a shape and its protrusions using 3D modeling software. On the other hand, development of a logical model of system behavior could consist of the representation of its important factors and their dependencies. Each of these examples would need a method for building a model. Conceptual method is a type of modeling method, and it is an import one in the M&S discipline. Different levels of specification are offered by different conceptual methods. For instance, a causal loop diagram is a higher-level conceptual method in comparison to an activity diagram or stock and flow diagram. Analytical methods are another relevant category of modeling methods. Analytical methods are modeling methods producing a closed form models (without simulation). Hester and Tolk [28] considered analytical models as the foundation for simulation models usually limited to relatively simple models used for deterministic analysis and static scenarios.

Definition 9

*A **simulation** is a method for implementing a model over time [132].*

Implications:

Three main simulation types are considered within the M&S community. **Live simulation** involves real people operating real systems [133]. For instance, live simulation consists of phases and events of an exercise. Its purpose is training within an environment closely resembling reality. C**onstructive simulation** involves simulated people operating simulated systems [132] that is Turing-computable [134]. For a definition of a model as a computable function and its implications see Weisel et al. [135] and Weisel et al. [136]. **Virtual simulation** involves real people operating simulated systems [132]. It is a combination of live and constructive simulation. It has both constructive simulation and human operator that connects both live and constructive environments. A flight simulator is an example of virtual simulation.

Definition 10

*A **M&S method** is a method that consist of both a modeling method and simulation.*

Implications:

The boundary between modeling used to produce a model and a simulation is problematic in the pragmatic context when trying to describe common methods like DES, SD, ABM, and PN. It is impractical or even inappropriate to refer to them as modeling or simulation only, when they are used for both parts. Although modeling and simulation are considered as separate parts, there is a need for a term that acknowledges methods that are commonly used for both modeling and simulation. This work proposed M&S method to combine elements of both modeling and simulation under a single term. When discussing certain specifics or formal descriptions related to either modeling or simulation, one can always use terms modeling, modeling method or simulation instead of the aggregated M&S method term.

Definition 11

*A **multi-method M&S approach** consists of at least two modeling methods, where at least one of them is an M&S method.*

Implications:

Combined modeling methods should allow for a unique system or phenomena evaluation, representation or insight. Epistemologically, a multi-method M&S embraces complementarity of methods and triangulation as its research justification. At a more abstract mental dimension, the multi-method M&S approach could be perceived as a way of diverse representation through different mental models that direct to use of different

M&S methods. Combined methods are chosen from a set of a total of n methods that is greater or equal to the number of methods known and used during conceptualization.

Definition 12

*A **constructive multi-method multi-simulation** consists of at least two constructive simulation models represented by different methods or different combination of methods, which do not interact.*

Example:

For instance, triangulation using two methods, e.g., ABM and SD of the same phenomenon, or a set of two simulation models developed using, e.g., DES and SD that separately represent a phenomenon are examples of constructive multi-method multi-simulations. No exchgange of data between both M&S methods exists during their separate runs.

Definition 13

*A **constructive multi-method simulation** is a special case of constructive multi-method M&S approach in which methods interact during the computer simulation run controlled by a simulation engine.*

Example:

An example of constructive multi-method simulatin is combined DES with SD, where both M&S methods exchange data during a simulation run controlled by a simulation engine.

A multi-method M&S approach is focused on the M&S field and requires at least one M&S method, but it is philosophically synonymous with a pragmatism-based mixed method approach [22, 27, 44, 73, 74, 137]. Moreover, combination of a conceptual method, e.g. qualitative analysis, and a simulation model [41] can be considered a mixed method approach as defined by Greene [27]. One can see, in this context, the major difference between mixed method and multi-method M&S approaches. A mixed method is a multi-method M&S approach if, among all methods mixed, at least a single simulation method is used. Additionally, the definition of a multi-method M&S approach specifies an important aspect that can distinguish the terms multi, mixed, or hybrid. Terms mixed or hybrid capture a study characteristic where methods are combined, while multi indicates multiplicity of methods considered, but not necessarily determines their status. The definition of multi-method M&S approach combines both aspects: multiplicity of methods considered, and, actually, mixed methods as its subset. For instance, if M&S methods considered and methods used are equal, then a multi-method M&S approach is also mixed or hybrid. Moreover, a single method simulation model is a special case of a multi-method simulation model. A single method simulation model can also be part of a multi-method M&S approach depending on conceptual method used.

This section has explored the problem of ontological ambiguity for the use of the term "multi-method M&S approach." Current reasoning or often lack of it, demonstrated in Section 2.3, displayed perspectives on different terms used to convey meaning that pertains to the use of more than a single method. This section discussed philosophical stance adopted about chosen terminology providing basis for defining multi-method M&S approach and relevant terms. The provided above definitions and work related to purpose of multi-method M&S approach included in Section 2.2 direct a path for research related to the epistemological and axiological aspects of the multi-method M&S approach. The complementarity of methods will be explored next to shape theory of multi-method M&S approach as a base to develop method formats, which in turn could provide a bridge between theoretical and applied parts of this work.

3.2 THEORETICAL BASIS

The M&S literature reviewed in Section 2.2.1 indicated various purposes for the use of a multi-method M&S approach. Most of the identified purposes such as multilateral problems, unique representation, data availability, validity, and emergent phenomena relate to the complementarity of methods as an overarching purpose. This section uses concepts of falsifiability, commensurability, complementarity and triangulation of methods to develop the principles of a multi-method M&S approach.

Definition 14

A falsifier is a basic statement that can be falsified (evaluated false).

According to Popper [45], a statement is falsifiable if it is possible to conceive an observation or an argument, which proves the statement (falsifier) in question to be false. Because scientific theories are formed from falsifiers, they must be accepted or rejected by scientists. A theory can be falsifiable to various degrees depending on chosen falsifiers. It must be at least theoretically possible to question falsifiers so that they can come into conflict with observation. The aspect of observation during M&S method choice is problematic in this context because the purpose of knowing which method or combination of methods to use in addressing a research question would require testing all possible configuration if an empirical approach, as conveyed by Popper [45], was assumed within the methodological context. The meta-analysis of modeling methods is clearly necessary in the context of method selection for a multi-method M&S approach whether considered as scientific, philosophical, or somewhere in between, e.g., as proposed by Mingers [11] by removing constraints related to established "paradigms" by separation of research methods from paradigms. This may require a higher-level analysis as compared to a level at which theories are described, for instance those theories that could be developed using M&S methods. The concept of commensurability of methods will be introduced and discussed later in this chapter to provide insight into a possible resolution to this concern. Popper [45] admitted that the one method of rational discussion is "that of stating one's problem clearly and of examining its various proposed solutions critically" [45]. The analysis of M&S methods should follow this advice to avoid naïve falsification in relation to method selection, e.g., by examining multiple falsifiers in the study context.

Even though a pragmatic view on multi-method is necessary, achievement of higher objectivity through a better understanding of subjective dimensions with a set of transparent falsifiers is considered paramount. The concept of falsifiability will be adapted to a multi-method M&S approach. The original context of falsifiability pertains to empirical content; hence, some adaptation to meta-analysis of simulation-based concerns is required. The concept of falsifiability of method is introduced next. Falsifiability of method is divided into internal and external falsifiability.

Internal method falsifiability is conceptualized as a characteristic of a method that describes whether a method can facilitate achievement of research objectives as seen by a modeler. For instance, if a method could not represent a phenomenon or a system with a required fidelity it would not yield a sufficiently valid simulation model. This, in turn, would disallow to answer research question(s) based on conducted experiments. Such a situation could be translated as an insufficient falsifiability of method expressed in Popper's terms as both inability of a method to represent system or phenomena at desirable level of universality (scope), and its insufficient precision (accuracy, resolution, and precision). Popper's universality in a simulation study is adapted to a scope in M&S. Similarly, precision is adapted by multiple factors such as accuracy, precision, and resolution. Precision in this work pertains to units of simulation trajectory (most often time) and when considering stochastic simulation an analysis of stochastic output [17] e.g. measured by variance [119]. Resolution is the degree of detail used to represent aspects of the real world or a specified standard or referent by a model or simulation [119].

External method falsifiability as seen by the scientific community or stakeholders relates to credibility of the study in the context of deliberation about quality of study in the context of a method or methods employed, and considerations about a method or methods that could have been used instead. The external falsifiability is more subjective. The often-qualitative external falsifiability requires confirmation from scientific communities. The multidisciplinary character of the M&S field makes this requirement more problematic because currently there are no agreed upon mechanisms for communicating subjectivity that could satisfy different scientific communities.

ABM may be more falsifiable than DES if used, for instance, to capture complex phenomena beyond DES passive entity capabilities (see Appendix A for definition of passive entity). Less falsifiable would mean more predictable and less variable description of phenomenon, but less probable as a sufficient outcome of phenomenon representation if a higher degree of falsifiability was desirable as expressed by higher scope, accuracy, precision, and resolution. If specific method(s) choices are inferior based on required level of falsifiability, the ability to choose adequately would make a research design more objective.

Based on concept of falsifiability of methods, factors derived from the literature in Section 2.2.1, and factors proposed by Robinson [113], a set of falsification criteria is proposed next. These are scope, accuracy, resolution, precision, data, run speed, visual display, reuse, and time to build a model. They can provide insight into the considered in Section 2.4.2 high-level dimensions i.e. on origination of study, methods considered, system and phenomena, and human dimension in relation to a conducted study. This work will especially focus on the evaluation context during method selection. Given a set of considered falsification criteria does not mean that all of them are applicable to all model components and study perspectives with the same magnitude. Falsification criteria are next divided into two groups based on strength of their relation with the concept of falsification to form a theoretical and hierarchical view that provides some initial guidelines on which of these are considered more important and why.

The first group of falsifiers consists of scope, accuracy, resolution, precision, and data availability. All these factors with the exception of data availability aimed at Popper's universality and precision, and they are directly associated with falsifiability. Because a lack of the proper data can inhibit calibration and validation of a simulation model this factor is included in the first group. Data availability should be considered especially in cases where descriptive model is important, and where different methods can be less prohibitive to generate an insight with limited amounts of data. Secondary, more pragmatic factors include simulation run speed, visual display, reuse, and time to build, which both can less directly affect testability using a simulation model. For instance, a visual display of entities and various visual analytics can be very helpful in observing patterns of behavior, while run speed can constrain an experimental framework due to computational complexity. The reuse of a previous model or building a model with a consideration for future reuse and time to build can influence research design decisions within origination of study dimension. These falsification criteria would pertain to the more holistic view of a research project, which indirectly influences testability. For instance, if model of a phenomenon was developed using one method, it is quite natural to consider its reuse for a similar project even if this necessitates combination of different methods. Although it is critical to confirm its usefulness within the research context, reuse could provide a solid starting point. The addition of a secondary set of falsifiers is justified based on adopted pragmatic philosophical stance as discussed in Section 3.2.1. Moreover, both the knowledge of methods and resources are important considerations. Resources and skills are not included as a falsification criterion because they shape the project's scope. Because of that, relevant information about available software and methods should be disclosed within the origination of study discussion. The modeler and stakeholders must choose a level of falsifiability that gives the most opportunity for success with the given study constraints. The path to knowledge generation may need a less falsifiable model at first, in order to realize how more falsifiability can benefit in gaining more understanding later.

According to Glazner [35], the decision to choose among methods may have some grey areas where no single method have unquestionable advantage. The ability to explore this situation necessitates the flexibility of representing study purpose using specific to a research question statements called falsifiers. Popper [45] discusses types of statements and their relation to falsifiability and verifiability (see the compiled view attached in Table 7). A specific or singular statement refers only to a finite class of specific elements within a finite individual spatio-temporal region so, according to Popper [45], they are not falsifiable. Moreover, universal statements should refer to any place and time hence they are falsifiable, and for the same reason they are not verifiable. On the other hand, strictly existential statements cannot be falsified, but can be verified. This is based on Popper's [45] logic that no singular statement can contradict the existential statement because they are limited to space and time.

Table 7. Types of statements according to Popper [45].

Type of statement	Example	Falsifiable	Verifiable
Numerically universal statement	Of all human beings now living on the earth it is true that their height never exceeds 8 feet	No (within space and time region)	Yes
Strictly or purely universal statement	All ravens are black	Yes (any place and time)	No
Strictly or purely existential statements	There are black ravens or there exists at least one black raven	No (no empirical/ metaphysical)	Yes
Negations of strictly existential statements	There is no perpetual motion machine	Yes (any place and time)	Yes

"We cannot search the whole world in order to establish that something does not exist, has never existed, and will never exist."[45] It is noted that the negation of a purely universal statement is always equivalent to a strictly existential statement and vice versa. In development of falsifiers used during method selection, the strict view of falsifiable statement is influential (objectivity), but may be prohibitive. Since only negations of strictly existential statements are both falsifiable and verifiable, the goal is to develop method falsifiers in the sentence format, *"there is no method that represents...,"* which is both verifiable and falsifiable. If one finds a method that falsifies the critical falsifiers, this method has desirable characteristics. By forming sentences as negations of strictly existential statements, we direct the focus to the required characteristics of the tested methods and nothing more beyond these boundaries.

Potential falsifiers used for exploration of a purpose of a multi-method M&S approach should be examined in the context of enhancing falsification criteria. In order to aid during the exploration of purposefulness of multi-method M&S approach, including evaluation of method(s) selection process it is preferable that these falsifiers can be used to eliminate methods, but if there is no clear distinction how they can also be used to rank methods by modelers and/or stakeholders within pragmatic stance advocated in this work. Popper [45] provided an example of deducibility relations between following four statements:

A. All orbits of heavenly bodies are circles.
B. All orbits of planets are circles.
C. All orbits of heavenly bodies are ellipses.
D. All orbits of planets are ellipses.

Statement A has the highest degree of universality and precision, and all other statements follow from it. Similarly, falsifiers generated based on study requirements, and enhanced within the context of falsification criteria, could provide a base for exploration of potential of simulation model falsifiability with a given method or a set of methods. Falsifiers could have advantage over criteria proposed in literature (e.g. refer to criteria in Section 2.4.2). They can be derived in context of appropriate universality (scope) and precision in the context of system and/or phenomena as seen through the lenses of a modeler. Summarizing, falsifiers will be considered as a means to capture desirable degree of falsifiability.

Definition 15

The degree of falsifiability is defined by the universality and precision of falsifier(s).

Definition 16

The gain of falsifiability (GOF) is the difference between higher and lower degree of falsifiability.

Kuhn stated "...that men who hold incommensurable viewpoints be thought of as members of different language communities and that their communication problems be analyzed as problems of translation." [138] The measure of commensurability is in large part still a philosophical concept that is difficult to assess or even describe, but it can offer an additional interesting perspective on multi-method M&S approach, hence an attempt to define it for the purpose of this work.

One can compare things or phenomena to search for similarity, differences, and a mix of both. The value of similarity and difference often depend on the context. If something is similar in a given context, it is often not different and vice versa, although crisp boundaries are not always easily distinguishable, and this situation is called fuzzy. Commensurability reflects ability to compare at language level. The context of comparison can be the language itself, which could provide value if more precisely stated in relation to the purpose of comparison. For instance, if comparing languages pertains to the purpose of comparing theories (models) arising from the language then commensurability can be analyzed in the context of closeness between theories in relation to the language that was used to describe them. If comparing languages pertains to the purpose of expanding theory that have better potential for closeness of theory (model) to system or phenomena, then commensurability can be better analyzed in the context of language uniqueness.

The first purpose aligns with triangulation, while the second with complementarity. The purpose of complementarity of methods is used for expansion, while pseudo-triangulation can be seen as unattainable, and often leads to expansion. If triangulation produces perfectly the same results, one can say that confirmation produced view of phenomena is more credible. In the situation when different methods continuously produce the same or very similar results based on the same situation, it may be claimed, to a degree, that the measures arising from different methods are suitable to triangulate given situation [139]. This way one could approach confirmation of correctness of triangulation of a given method/measure. On the other hand, if a triangulation study produces some differences, the expanded view based on differences in results necessitates further exploration. Because the differences in methods could cause different results, the comparison of methods would be a part of explaining the differences in produced theories.

The uniqueness of methods dominates the region of commensurability that is characterized by expansion, while closeness between theories dominates region of commensurability that is characterized by triangulation. The ability to point to methods uniqueness and theory closeness is a convention for differentiation between meaning of commensurability in relation to the context of its purpose i.e. the ability to compare at the language level. Finding uniqueness in the context of lack of similarity can be misleading and vice versa. A lack of similarity does not guarantee uniqueness, and a lack of uniqueness does not guarantee similarity. The distance between these extreme poles is what makes the gray area so large.

The difference between commensurability of models (e.g. a theory) and commensurability of methods will be explained first. As a convention, these terms have opposing meanings because of their different purposes. Commensurability of models pertains to commonality of language that permits or does not permit comparison of models (theories). Kuhn described incommensurability using the phrase 'no common language':

"...theories are incommensurable is then the claim that there is no language, neutral or otherwise, into which both theories, conceived as sets of sentences, can be translated without residue or loss." [140]

Because a method is a form of a language [33], the phrase 'no common language' can be stated as 'no common method'. The commensurability as defined originally by Kuhn [138] means that different methods can produce sentences that are incommensurable because of a translation problem (leading to misinterpretation). From this perspective, methods that are more similar could produce sentences, in relation to a theory, that are more similar and incommensurability of two models representing the same theory should be less probable given that these theories are meant to be the same. A notion of commensurability of methods is proposed at one level higher over the commensurability of a model. If the previous logic is applied, one can say in the context of commensurability of methods 'no common language about/of method'. If one considers choosing method(s) from a set of methods, determination of their commensurability could pertain to their characteristics and ability to find a common language that consists of sentences that would allow finding their required unique characteristics. If the goal of comparison of methods is to choose a method or a set of methods, falsifiers that compare methods characteristics should focus on their uniqueness in the context of study purpose. From this point of view, if methods have unique characteristics they would be more comparable hence more commensurable (proposed as convention). The difference between commensurability of models and commensurability of methods relates to the purpose of comparison versus purpose of expansion. Using this perspective, when methods possess their necessary unique characteristics, these will be considered as methods that are more commensurable, but may not necessarily imply less commensurable models of the same theory. When methods are more alike for a given purpose, it implies a better chance for commensurable models of the same theory, but does not focus on unique characteristics of methods. This leads to a definition of commensurability of methods and models.

Definition 17

Commensurability of methods and models are characteristics that determine the existence of falsifiers, allowing for either complementarity of methods, triangulation, or both.

The relationship between GOF for methods M1 and M2 and commensurability of models is proposed in Figure 9. The shape of the graph is assumed for illustration to display a decrease of GOF along the commensurability of models axis. The challenge is to find falsifiers that make less commensurable models more commensurable, which would enable advancements of theories. The relationship between GOF, commensurability of methods, and complementarity is proposed in Figure 10. The shape of the graph is assumed for illustration, and it displays increase of GOF along the commensurability of methods axis. The problem surfaces with the practical aspects of measuring gain of falsifiability, and commensurability of models and methods, which may be subjective because they depend on developed falsifiers and their evaluation.

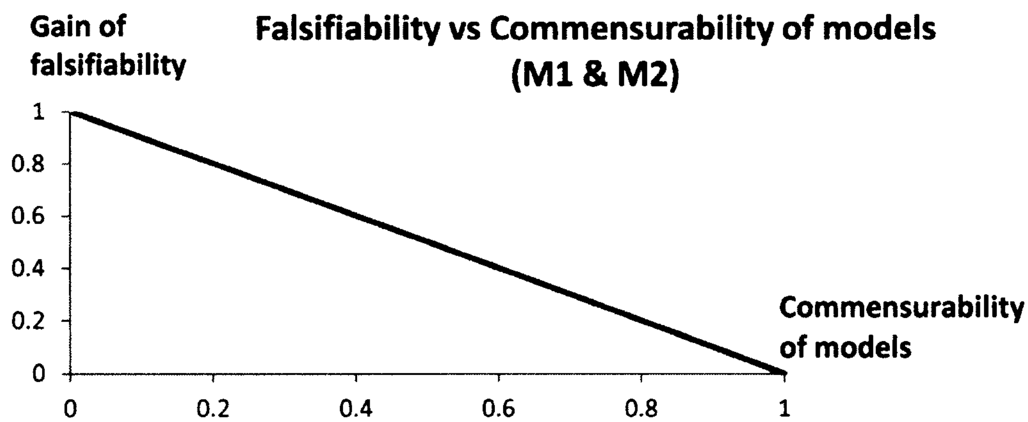

Figure 9. Conceptualization of relationship between falsifiability and commensurability of models for two methods M1 and M2.

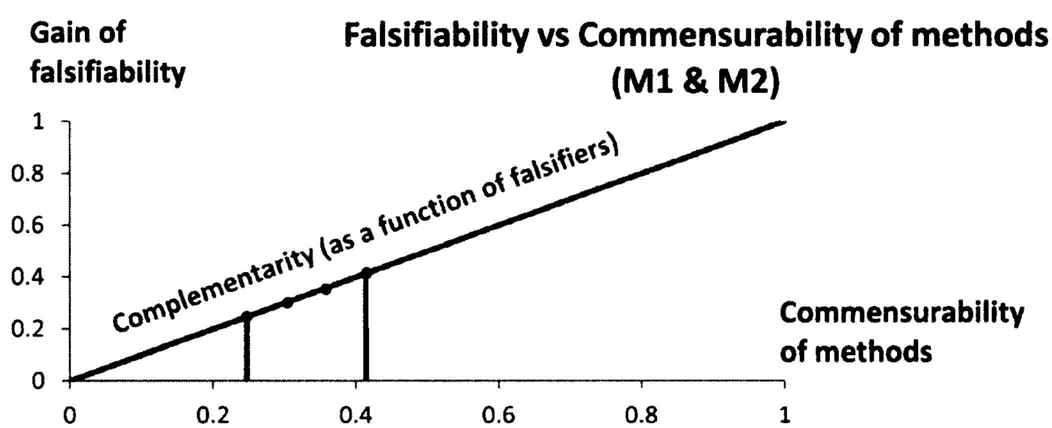

Figure 10. Conceptualization of relationship between falsifiability and commensurability of models for two methods M1 and M2.

The question is how to determine the level of commensurability of methods and how it can be used. This work next proposes how to identify and estimate commensurability of methods. The approach is based on ability of methods to have unique characteristics like assumptions, unique measures, and unique language in relation to purpose of representation manifested e.g. into model components. Having common criteria/falsifiers that can be used to compare methods is a requirement of commensurability of methods. The degree of difference between alignments to criterion for compared methods will be used to assess commensurability of methods.

The measure of commensurability of methods in relation to criteria/falsifiers is described by Equations 1 and 2. Equation 1 describes commensurability C of methods A and B for a given criterion or falsifier i. Equation 2 calculates the average over n number of given criteria.

$$C_i = |C_{iA} - C_{iB}| \tag{1}$$

$$C_n = \frac{\sum_{i=1}^{n} C_i}{n} \tag{2}$$

Commensurability is calculated as an absolute difference between criterion/falsifier scores of two methods. The larger difference between the methods' characteristics means more distinguishable, hence commensurable, methods. Similarly, if methods are similar for a given criterion they are more difficult to distinguish, hence lower commensurability of methods. Equation 2 is used to calculate average commensurability of methods for a given set of criteria.

A purpose of complementarity of methods will be used to explain the purpose of multi-method M&S approach. The concepts of falsifiability (testability) [45] and commensurability of methods will be used to explore complementarity of methods.

The concept of sub-falsifiability is defined in the context of the study boundaries to choose methods that can facilitate a desirable level of falsifiability based on a study's purpose.

Definition 18

The sub-falsifiability score is a partial degree of falsifiability, evaluated in relation to characteristic(s) defined by a falsifier(s) reflecting desirable degree of falsifiability.

Definition 19

The complementarity of methods score (CoMS) is a gain of the sub-falsifiability score calculated as a difference in sub-falsifiability scores between better adequate and less adequate methods for a given falsifier or a set of falsifiers.

Complementarity Principle 1

If a higher degree of sub-falsifiability is desirable, and if, for considered falsifiers, multiple methods used together facilitate CoMS above zero, a multi-method M&S approach is justifiable.

Complementarity Principle 2

If a higher degree of sub-falsifiability is desirable, an approach with higher CoMS for considered falsifiers is more justifiable.

Complementarity Principle 3

If, for considered falsifiers, CoMS equals zero, a multi-method M&S approach is not justifiable except for a pseudo-triangulation. This would mean that, if, for a given falsifier, sub-falsifiability of each method used in total separation is the same, then both methods have equivalent characteristics for a given purpose. Pseudo-triangulation between views created with method(s) at the same level of falsifiability for a given purpose may be conducted in cases where methods are the same and are adequate for the purpose. In these cases, methods or a combination of methods used separately should be able to realize the same concepts and possibly subsequent results for comparison. In this case, a single modeler could to some degree benefit from pseudo-triangulation, but engaging different modelers would facilitate more objective triangulation.

Complementarity Principle 4

If, for considered falsifiers, neither of the classes of falsifiers of considered methods could include the other(s) as a partial subclass, the methods have non-comparable falsifiers thus complementarity and pseudo-triangulation are impossible. In this case, methods cannot be used for comparison or expansion because they do not have relevant mental, numerical, or language domains of consideration.

It would be not be appropriate to use a single falsifier; hence, a different set of considered falsifiers can yield different CoMS results, even for the same set of methods considered. This requires some elaboration. The devised falsifiers can influence research objectivity and communicate its subjectivity. It is prohibitive to use a set of methods based on a single falsifier (naïve falsifiability). A set of methods may be used both for complementarity and pseudo-triangulation reasons given different falsifier are considered.

In reality, the decision to choose methods during research design may be more blurred due to limited knowledge about systems and phenomena, and some approximation made by researcher/modeler will be required. For instance, if CoMS is insignificant for a given falsifier or a set of falsifiers, expansion may be less valuable given the effort to build a more complex multi-method simulation model. In this case, the effort would end up closer to pseudo-triangulation with some possible expansion. If methods have desirable unique characteristics, using multiple methods has the potential to boost CoMS and in turn may improve the developed theory.

3.3 PROPOSED METHOD FORMATS

Balaban and Hester [10] proposed an initial concept of MFs derived from empirical mixed method approach based on Greene [27]. In Section 2.4.1, a definition of MFs have been proposed and three general relations were derived from M&S relevant literature with support of UML relations. In short, MFs pertain to generalized view that consists of methods and system and/or phenomena. A set of MFs and transitions of model component(s) toward atomic MF(s) can be used to design a simulation based research that can involve multiple modeling methods. A theoretical principles proposed in previous section will aid in the process. Please refer to Figure 11 during the following discussion. Each MF and its transition(s) are described next.

3.3.1 Special Case Transitions: Single Method

Transition 1 toward MF I conveys the idea that in order to triangulate a view of phenomenon and/or system A, while using the same method(s) two modelers M1 and M2 are needed. This is the most proper triangulation because the same method(s) are used.

This transition pertains to the case where CoMS and commensurability of methods is zero in a given context because both methods are the same.

Transition 2 toward MF II conveys the concept of refinement or extension of system and/or phenomenon with an addition of a new component developed with the same method. This is the case where multi-method approach is not needed because a single method is at sufficient degree of sub-falsifiability to expand the system or phenomenon to fulfill the research purpose. An additional method would not have facilitated a CoMS above zero.

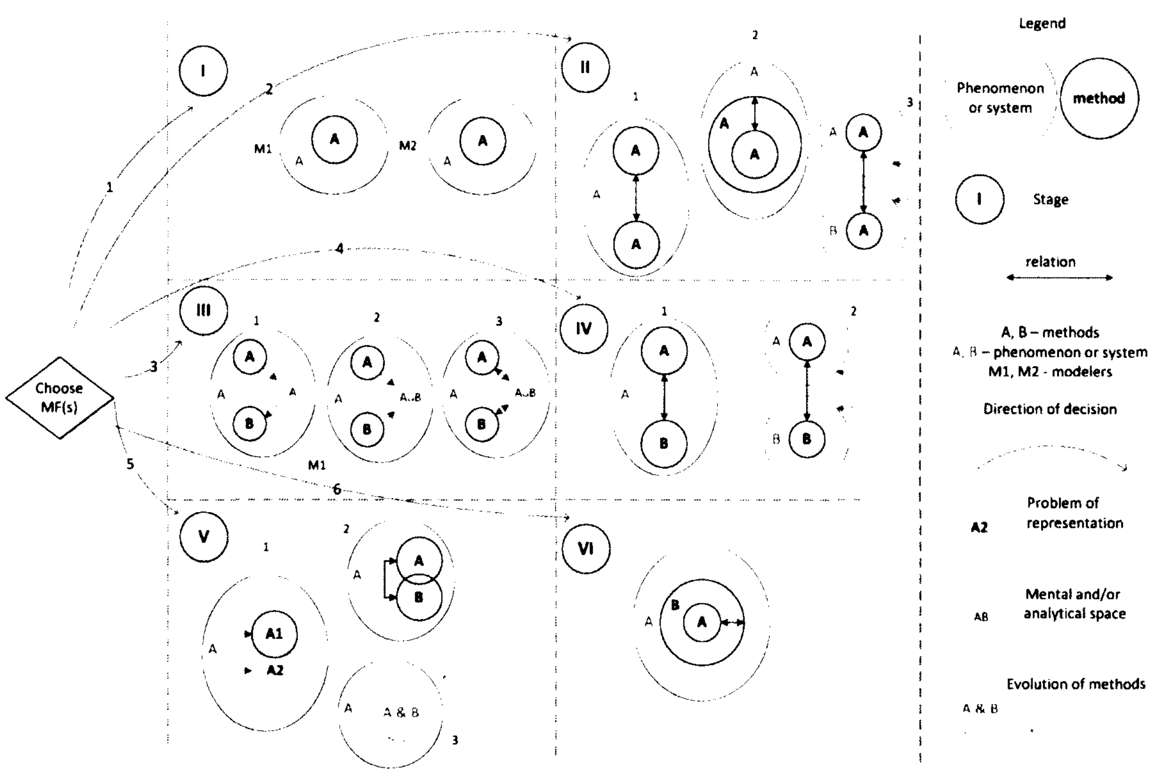

Figure 11. MFs and transitions based on complementarity of methods.

3.3.2 Multi-Method M&S Approach

Transition 3 toward MF III displays three multi-method M&S situations. At the beginning, it is worth to emphasize that all three versions are not multi-method simulation models. The arrows that depict relations point toward mental (learning) and/or analytical spaces; hence, methods are not bounded by a simulation engine. In the first situation supported by Complementarity Principles 1 and 2, only one-directional, sequential data and/or insights flow takes place. This means that the first method can be used e.g. to conceptualize and generate parameter values, while the second method can be used to expand concept upon its accuracy, resolution, precision, computational efficiency, or to balance them. This MF should be characterized but CoMS above zero for a given purpose.

The second situation within MFIII shows pseudo-triangulation of phenomena or system through the lenses of different methods to identify irrelevant sources of variation and observe consistency of two models. A one-directional flow of data and/or insights aims at comparison of results, hence sufficient similarity of methods and CoMS that equals zero of combined methods is desirable (Complementarity Principle 3).

For instance, within ABM, an agent's states can be mapped properly to stocks of SD method, or if within ABM, an agent's states can be mapped properly to the process view of DES blocks for a given phenomenon as demonstrated by Borshchev and Filippov [141]. In practice, expansion can also take place because of methods' differences, which can mean two things: 1) modeling error leading to unnecessarily inflated purpose, or 2) discovery of desirable expansion unforeseen by the original purpose.

Balaban et al. [76] used MFIII.2 first by employing first a Bayesian Network (BN) model as a way to increase conceptual validity of a causal diagram, by using a point estimation results of BN. In the following step model was expanded by building a more accurate simulation trajectory using SD as MFIII.1. Both phases can be seen in Figure 12.

In the third case of MFIII, data and insights can be exchanged between complementary methods in two directions. This could allow for expansion of partial to different methods phenomena and systems based on gain of falsifiability or allow for a pseudo-triangulation of results of sufficiently similar in the context of comparison methods. This situation is a combination of earlier presented situations one and two of MFIII, which most likely would occur at different stages of a study.

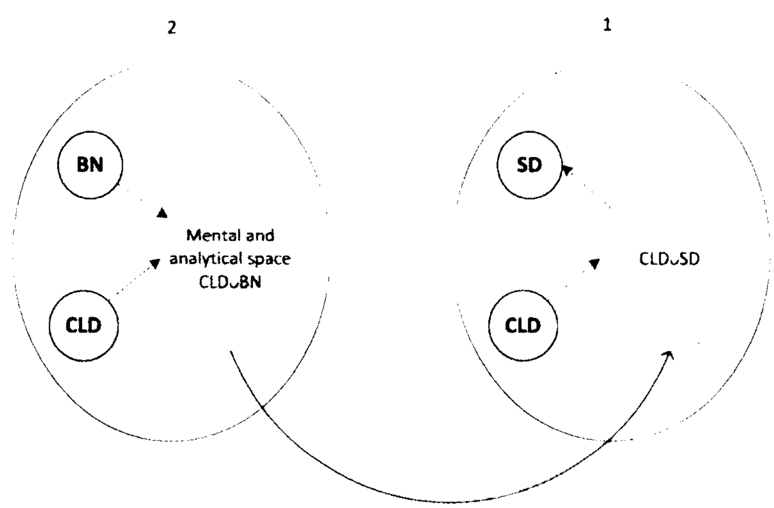

Figure 12. MFs showing two phases: conceptual validation and expansion.

For instance, Calanni Fraccone et al. [142] separately used two methods ABM and Stochastic Petri Nets (SPN) in a two-step methodology. The first phase focused on exploration of hazardous scenarios with ABM method, while the second phase used SPN to quantify the risks of identified hazards. This approach falls into transition 3 toward MF III. ABM was used to model a portion of the National Airspace System (NAS), e.g., aircraft trajectories, actions of pilots and ATCs in order to explore hazard scenarios, and extract traffic parameters and conditions. The insight and outputs from ABM were then used in SPN model, which allowed for higher level of abstraction of the environment, while preserving crucial aspects of system and human errors necessary to capture hazardous scenarios. The efficiency of SPN allowed for faster exploration of sensitivity of various parameters:

> "…running 10,000 Monte Carlo simulations for 10,000 s of operation for each set of parameters, which takes about 15 s of total simulation using a MacAir laptop; this can be contrasted with about 5 hrs. of agent-based simulation required to produce the same results." [142]

Moreover, ABM was also used for pseudo-triangulation (called validation) with a more abstract SPN model. This methodological approach showed the value of complementarity of methods where SPN method was able to preserve accuracy of ABM while significantly decreasing experimentation time. The authors used both transitions MFIII.1 and then MFIII.2, which could also be displayed as MFIII.3. Both ways to represent this multi-method study is shown in Figure 13. This example shows case where the authors used the same two methods for both expansion inward (generalization) and for comparison.

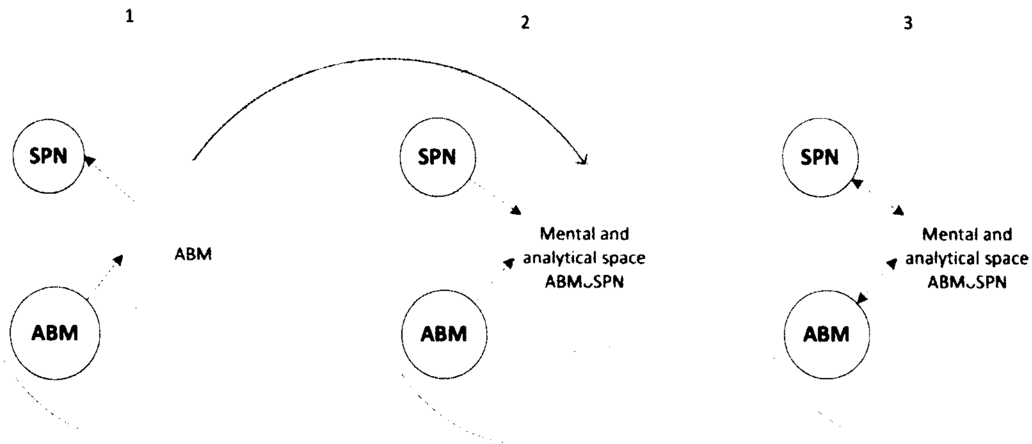

Figure 13. MFs on left shows two phases: Expansion and comparison, while on the right MFIII.3 is shown as an aggregated equivalent view.

The main point to understand within MFIII situations is that the first situation lends itself toward a CoMS above zero, while the second makes more sense for triangulation. In the third case, methods are used for both expansion and pseudo-triangulation at different stages of study.

3.3.3 Transitions Specific to Multi-Method Simulation Models

MFs IV, V and VI come to existence by the addition of a second constructive method, which creates multi-method simulation model that can support sequential (not shown) or bidirectional interactions between methods controlled by a simulation engine. The transition toward these MFs should be supported by CoMS above zero.

Transition 4 toward MF IV is realized by a combination of constructive methods that exchange or manipulate data, trigger events, or allow for transitions between forms of representation. This MF is the most general association (relation A from Figure 4), which also explains its large scope of possible interactions between methods and possible subjectivity to decide which method(s) to use. This MF has two different cases related to flexibility of phenomena conceptualization. Case 1 is described by a single phenomenon and two methods. Case 2 defines a second phenomenon, as an expansion outward [10] (scope). This is a pragmatic, but subjective, matter related to how social phenomena are defined; hence, it must assume interpretivism as a philosophical stance for support, which is allowed from within a pragmatic stance as long as it aligns with the purpose of a study. A level of justification for using a second method as captured by CoMS in both cases depends on falsifiers used. Because of its general characteristic and subjective phenomenon or phenomena structure this format can reflect the problems during method choice as pointed by Glazner [35], and requires special attention.

Transition 5 toward MF V represents "fuzzy complementarity" (case 2). Desirable higher degree of sub-falsifiability would be infeasible by adding component A2 with the same method (see Figure 11 case 1 of MF V) because of CoMS would equal zero (Complementarity Principle 3). Case 2 with the overlapping methods A and B creates some unique and separable complementary representation, which can produce cascading effects and allows for dualism of conceptually atomic unit. Both methods add their unique behavior under combined element. This MF is equivalent to relation B in Figure 4 (Section 2.4.1).

In a simulation model, separable views can influence each other via two directional interaction points, which requires both having complex structures and often behaviors (not simple variables). For instance, a service system can be conceptualized with the dual view of customer as an entity and an agent. DES offers easy and more efficient view of service process, while ABM allows for representation of active or proactive behaviors (see Appendix A for definitions) e.g. balking from queue at any time based on internal event generation while being in the DES process. DES alone would allow only for decision at exact points (gates) of the process.

Transition 5 is added to discuss important aspects of multi-method practice: methods evolution, which was already briefly indicated in Section 2.4.2.3. The level of integration and evolution of methods can take different routes. If each method can be considered a system, than an evolutionary character of multi-method M&S can be viewed as concept similar to evolution of SoS [143]. For instance, a simulation vendor AnyLogic® offers "Road library" as integrated DES with functions of motion optimized toward mimicking highway traffic, street traffic, and parking lots. This allows more efficient physical and queuing modeling of systems with vehicles, roads, and lanes but hides to the user some access to original DES or solution functions used for car motion separately. Lost independence of two methods is partially compensated by restricted compatibility with DES library. Similar route took other simulation software vendor Emulate3D by offering physics based DES. The integration process is often realized by simulation software vendors because of their competition, driven by pull from customers and simulation community. One of the challenges M&S field faces is to find mechanisms that allow for easy model development as in fully integrated methods and flexibility to use methods separately when needed, e.g. allowing for easy aggregation and disaggregation of libraries at different level of abstraction. It is also added that from a computational stand point of view, both Case 2 and Case 3 configurations in MF V could generate the same results, but efficiency, model creation process, and reuse at different stages of integration could differ significantly (more efficient, easier to develop, and less flexible for more integrated version).

Transition 6 toward MF VI, where one method is enclosed within other method, can be helpful for creating different multi-method simulation architectures based on embedding methods into one another. This is relation C in Figure 4. Currently, this concept is mainly used to embed different methods (SD, BN, and DES) into an agent (object) within ABM. From the software engineering point of view, it is possible to implement embedding of different methods into each other, but why would one need to do it is less obvious. For instance, ABM with its characteristics benefits from embedding, while one at this point simply does not know if e.g. BN or SD could benefit from it. Moreover, interaction of methods within this format is optional and it can be one or two directional (Figure 11 shows two directional).

In aggregation/disaggregation, interaction points of simple variables will be always one-directional. Only in the cases where complex structures interact, they can be devised in the way that their internal views can be codependent bi-directionally.

3.3.4 Demonstration: MFs of Multi-Method Simulation Model

The purpose of this demonstration is to examine the ability to map a multi-method simulation model using MF II, IV, V and VI (methods interact during a simulation run). The example is based on a multi-method simulation model developed by Balaban and Mastaglio [144]. The study identified potential cases that call for the application of simulation-based decision support system in the context of short sea shipping at both strategic and operational level. A simulation model discussed employed multiple M&S

methods i.e. DES, SD, ABM, BN, SC, and Road Library (RL). The components represented by using different methods will be retrofitted using proposed in previous sections MFs. A brief introduction of system will be followed by an overview of main components from which the MFs will be drawn.

Graphical depiction of the discussed next Roll-On-Roll-Off (RoRo) system is shown in Figure 14.

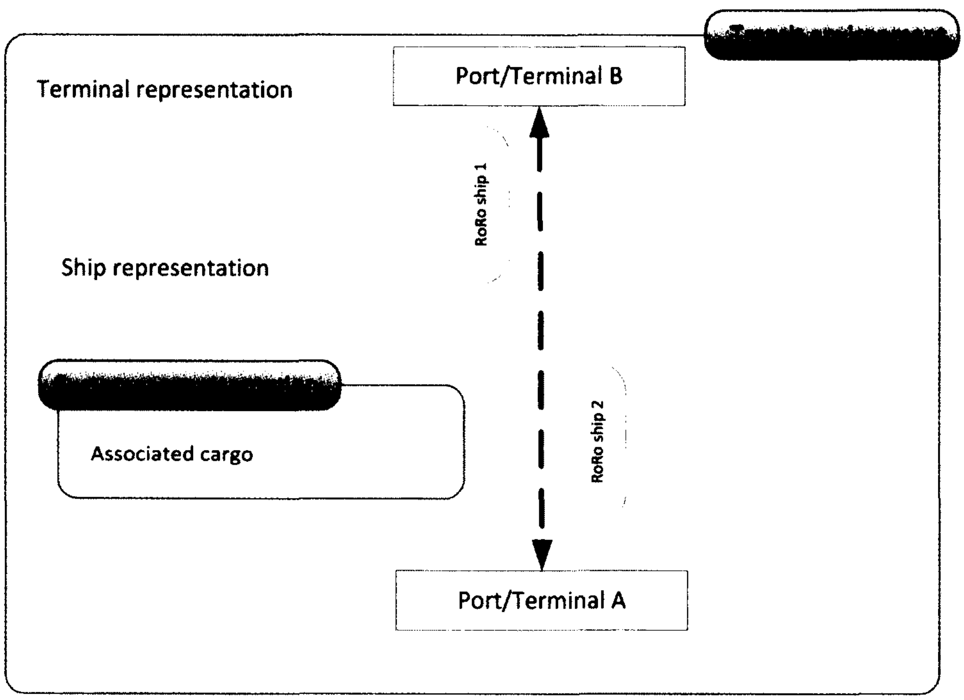

Figure 14. View of RoRo system concept.

Two ships operate between Port A and Port B. These ships are under schedule constraints correlated with speed of vessels required to support the schedule. The cost of ship operation is largely dependent on fuel economy that is dependent mostly on ferry speeds during transit. The transit reliability depends on port, ship, transit conditions including weather, and human elements. The unforeseen variability of ship transit could be potentially compensated on the terminal side of the RoRo system by higher capacity for cargo and more time efficient loading and unloading processes allowing for punctual arrival at destination. Moreover, time flexibility during transit should allow lower fuel consumption and cost. Passengers are aware of the schedule and are prompted to arrive within certain period before the ferry departure. Passenger must decide at what time to arrive before departure, and this decision can affect congestion during processing at terminal. Customers arrive depending on terminal demand as a population of potential customers. Seasonal and weekly fluctuation of demand for various cargo types affects arrival rate. Arriving at a port, customers are processed to an access area, where they wait for permission to enter a ship. A ship has back ramp used for loading and unloading vehicles. After loading, a ship departs e.g. Port A. Transit conditions can generate speed fluctuations. A loaded ship arrives at the terminal of Port B, moors, its ramps are deployed, and terminal cargo operations begin. Alternative sequences of cargo loading and unloading can be tested. When all cargo is unloaded and loaded, the ship prepares to depart by closing ramps and cruises back toward Port A. The cycle repeats based on schedule of daily departures. The customer is modeled throughout round-trip, and it is permitted to decide on mode of transport during each phase. The hierarchical structure of model is shown in Figure 15.

The transit environment (TE) modeled as ABM includes two other main components: ports and fleet, and is a placeholder for weather and map components. The map consists of an accurate scaled route representation as a transit path. The simplified map displays a spatial view of ports and moving ships. A weather component represented as SC generates weather oscillations to represent ships' transit conditions that influence their speeds.

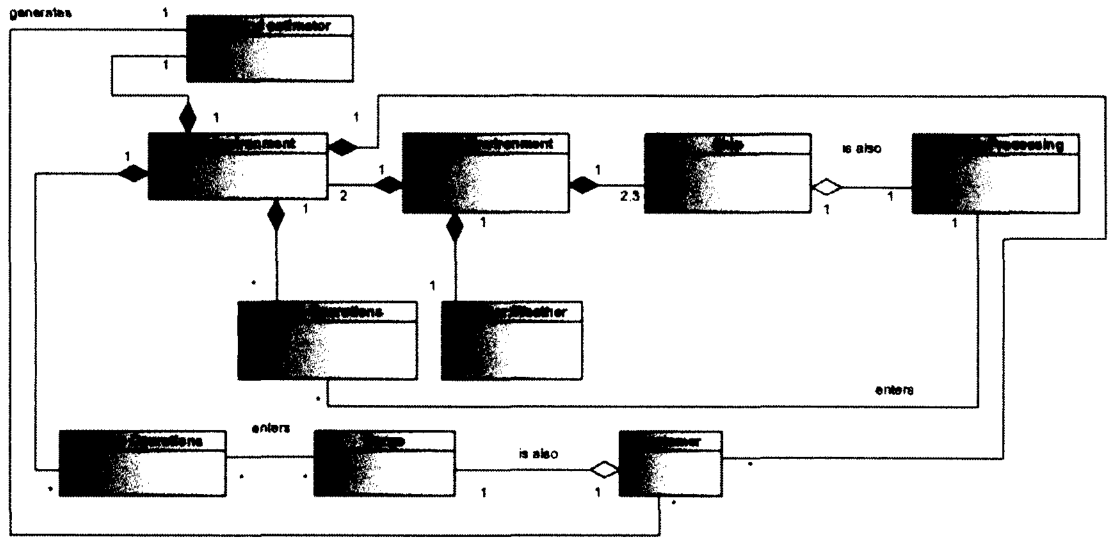

Figure 15. A diagram of main components and subcomponents.

The port environment (PE) is modeled as ABM, and it is a subcomponent of TE. It consist the following components: layout of terminal, ship operations (DES), and loading and unloading processes, which have conceptually very similar functionality hence are considered together as cargo operations. Additionally, a demand estimator component developed using SD generates terminal demand as a population of potential customers. Seasonal and weekly fluctuation of demand for various cargo types affects arrival rate. PE creates and maintains customer representation component and its transitions between components throughout its persistence.

Cargo operations are modeled using RL and are subcomponents of PE, and are crucial activities of the system. They represent cargo movements within a terminal's layout as loading and unloading processes. The differences in loading and unloading that range only in minutes can translate into significant ship's cost reduction. For this reason, these processes require high fidelity mapping of real system to generate data reflecting accurate time based measures related to reliability of RoRo terminal operations. Accurate examination and detection of minute-range time differences of different layouts and operational rules for access areas requires high accuracy representation of moving cargo with scaled physical dimensions. It is possible to adjust logic of cargo flow to test different options by controlling individual lanes.

Ship operations are modeled using DES and take place within the PE as its subcomponent. They include ship time-based processes such as mooring, ramp operations, departure, and coordination with cargo operations.

Customer is modeled as ABM and it is a subcomponent of PE, and its main components: *Cargo, Behavior and Satisfaction Construct* are shown in Figure 16.

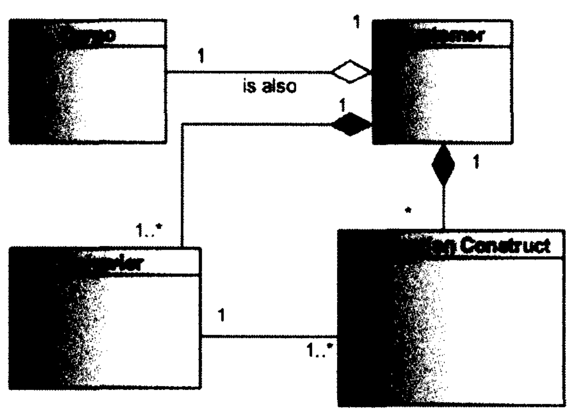

Figure 16. A diagram of main subcomponents of Customer component.

The *Customer* component can be of type private or commercial. A customer is also a *Cargo* (modeled as RL), which can have also two types: commercial truck or private car. The customer representation persists throughout two-way travel, which is captured within *Behavior* component modeled as SC method. The *Behavior* component is also necessary to represent potential shift of its original travelling mode plans. If the customer decided to choose RoRo, it is scheduled to arrive at terminal. Once customer arrives at terminal, it is processed as *Cargo* according to terminal's rules depending on type of cargo associated with it (e.g. car or truck). The customer follows its scheduled time to come back, and chooses the mode of transport once again. It must coordinate its activities with terminals and ships. The customer experience translates into satisfaction within *Satisfaction Construct* modeled as BN, which could also be used to represent intention for choosing between modes of transportation.

RoRo ships are modeled using ABM and are subcomponents of TE. They need the following components: *ship's logic* (modeled as SC), ship's *operational status* (modeled as SC), *movement* and *fuel-consumption* (modeled as SD) sub models as shown in Figure 17. The ship represents cargo transporting between ports. It is susceptible to weather conditions during transit, and can break. Additionally, ship is involved in *ship operations* (DES) like mooring, ramp and departure processes. The ship's goal is to follow the schedule while minimizing fuel consumption. The ship represents patterns of transit speeds to calculate transit time and cost of fuel. It must coordinate activities with *PE*, *cargo*, and *customers*.

In order to draw MFs of this model, first all components with embedding relation were connected, followed by components with duality relation, and finally with exchange relation. Then all methods and interaction between methods were acknowledged.

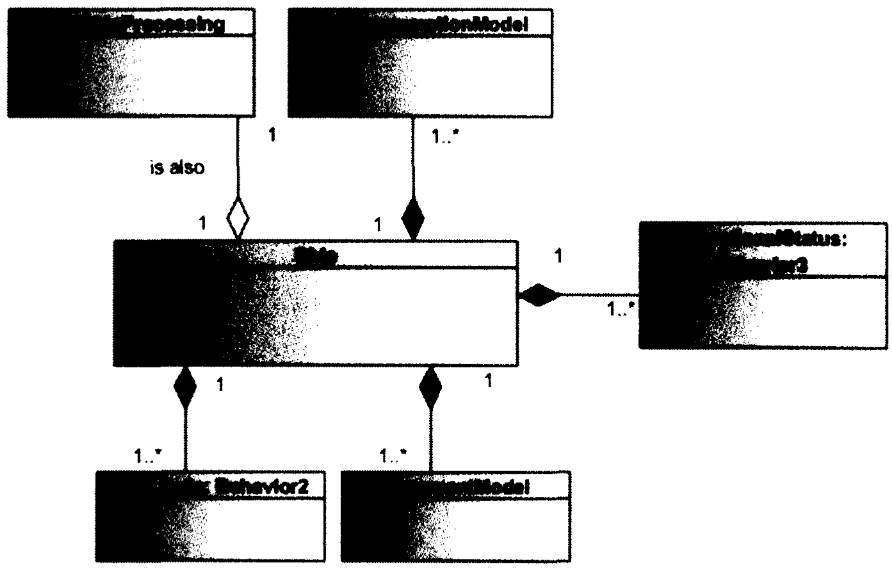

Figure 17. A diagram of main subcomponents of the Customer component.

Developed by Balaban and Mastaglio [144] a multi-method simulation model is characterized by the MFs shown in Figure 18. ABM view enhanced with SC and BN provided more flexibility in representing individual customers and their behavior, whereas processing view of RL simplified representation of high fidelity cargo flow through terminals. DES was used to capture ship's operations within a port, yet ABM was used to capture the ship's transit between ports because this allowed the representation of internal fuel consumption and movement dynamics using SD. SD was also used to estimate high-level demand. The combination of modeling methods with their unique characteristics facilitated the representation of both aggregated and individual levels. The funnel-like estimated demand of customers permitted to include both effects of large scope of considered population of customers and higher fidelity of the operational view.

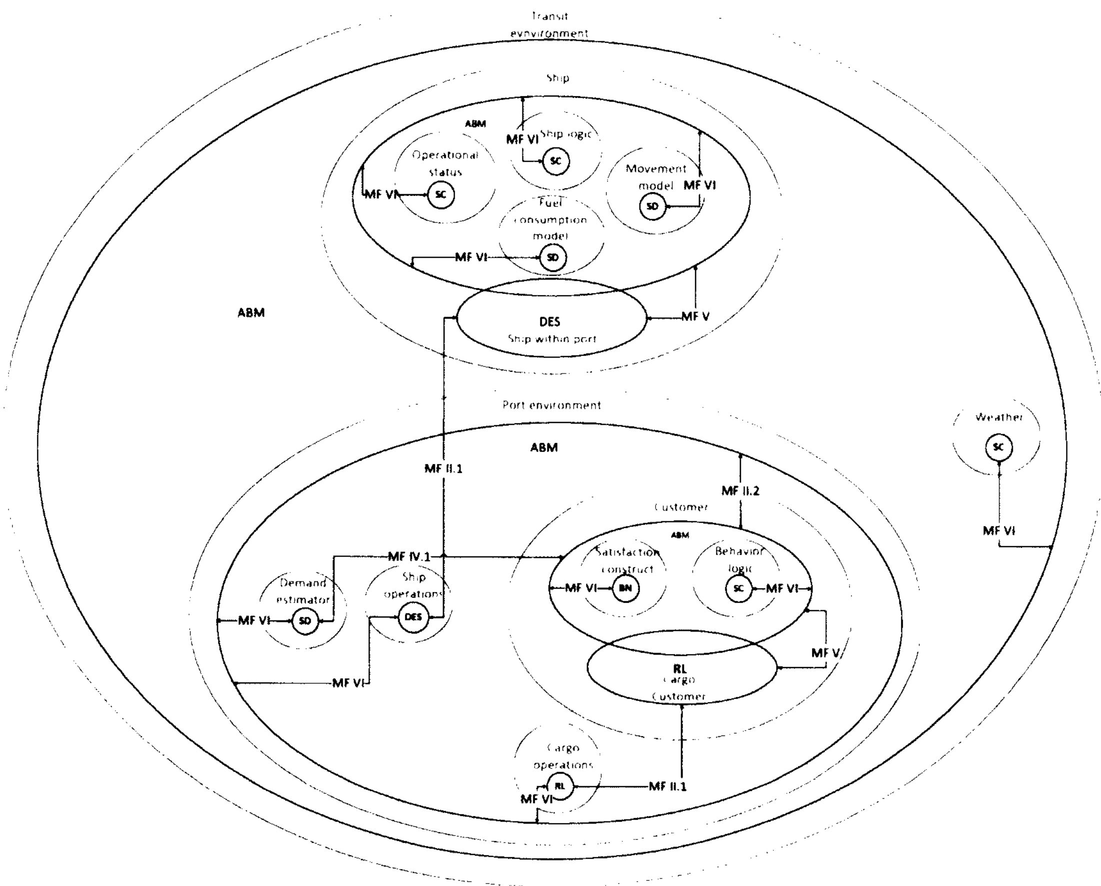

Figure 18. RoRo multi-method simulation model displayed as MFs.

3.3.5 Summary of Method Formats

Each proposed MF could be distinguished from the others based on its unique characteristics as summarized in Table 8. Moreover, the ability to map MFs to both multi-method M&S approach in which methods do not interact and interact were demonstrated using examples.

Table 8. Summary information for model formats (MF).

Model Format	Summary Characteristics
I	Depicts the purpose of triangulation using the same method or a set of the same methods, while more than a single modeler is available.
II	Depicts the expansion of system and/or phenomena representation by adding a component/subcomponent using the same method as it suffices for the purpose.
III	Depicts the expansion (MFIII.1), pseudo-triangulation (MFIII.1), or pseudo-triangulation with expansion (MFIII.3) of system and/or phenomena representation using a multi-method M&S approach realized by different methods not joined by a simulation engine.
IV	Depicts the expansion of a system and/or phenomena representation using a multi-method simulation model realized as direct replacement, aggregation/disaggregation, trigger events, or transformation.
V	Depicts the expansion of a system and/or phenomena representation using a multi-method simulation model, with at least single duality realized using different methods.
VI	Depicts the expansion of a system and/or phenomena representation using a multi-method simulation model, where one method is embedded within other method.

3.4 SUMMARY

The first section of this chapter explored the problem of ontological ambiguity for the use of the term "multi-method M&S approach". The philosophical discussion clarified stance adopted about chosen terminology and provided basis for defining multi-method M&S approach and relevant terms.

In the second section, a search of the theoretical basis was conducted to move toward answering the research question. The complementarity principles were derived based on theory of falsification as a mechanism for reasoning about method choice that can facilitate desirable level of sub-falsifiability in relation to a study purpose. In this context, CoMS was proposed as a measure used to justify the use of multiple methods to enhance representation. Moreover, the exploration of commensurability provided an additional dimension of the analysis of complementarity.

In the last section, MFs were derived as a blueprint of multi-method M&S approach. MFs III, IV, V, and VI fall under a multi-method M&S approach. The principles of complementarity direct to appropriate MFs. MF I and MF II provide an alternative to a multi-method path. MF I depicts the purpose of triangulation with the same method or a set of the same methods. MF III is realized by different methods not joined by a simulation engine. MFs II, IV, V, and VI can be used to create a larger structure of a multi-method simulation model. This means that MFs IV, V, and VI can be used multiple times by different components of a multi-method simulation model, and can be combined with MF II. On the other hand, MF II on its own is a single method simulation model. In order to increase research objectivity and transparency transitions toward formats must seek justification as directed by complementarity principles. For instance, the criteria for method(s) selection or falsifiers could be used to evaluate methods. The evaluation using falsifiers is expected to provide a way to select a viable configuration in the study context. Falsifiers could highlight unique aspects of methods, explaining specific merits of multi-method M&S approach and possible configurations. In the next chapter, the criteria will be used as a proxy for falsifiers to explore the relationship between commensurability and CoMS, which is related to the purpose of multi-method M&S approach and problem of method(s) selection.

CHAPTER 4

ANALYSIS OF THEORETICAL BASIS USING CRITERIA

The goal of this chapter is to gain insight into commensurability of methods and CoMS and their relationship. This will help to assess plausibility of the theoretical basis proposed in Section 3.2 related to the purpose of multi-method M&S approach and problem of method(s) selection. A sample set of criteria for method selection proposed in Section 2.4.2.2 will be used as a data layer during the analysis. The criteria will be used as a proxy for falsifiers.

4.1 ANALYSIS

The proposed criteria in Section 2.4.2.2 aligned with seven methods were estimated using scale ranging from none through minimal, moderate, significant, to unique, as shown in Table 9.

Table 9. Ordinal scale between criteria and a set of seven sample methods.

Criteria/Method	DES	SD	ABM	BN	FM	SC	PN
Representation of individual behavior as part of a larger system	significant	minimal	unique	none	none	none	moderate
Ability to operate on aggregates	none	unique	significant	unique	moderate	moderate	moderate
Ability to handle uncertainty	unique	minimal	significant	significant	significant	significant	significant
Interaction	significant	none	unique	none	none	moderate	significant
descriptive usage	unique	minimal	moderate	moderate	moderate	significant	significant
theoretical usage	minimal	unique	unique	significant	significant	moderate	moderate
emergence	minimal	minimal	unique	none	none	none	moderate
Ability to represent active behavior	none	none	unique	none	none	significant	significant

Next, each item was quantified using Table 10. Table 11 shows the realization of Equation 1 and the quantified scale from Table 10.

Table 10. Quantified ordinal scale.

unique	1
significant	0.75
moderate	0.5
minimal	0.25
none	0

Table 11. Possible values of commensurability of methods.

scales	unique	significant	moderate	minimal	none
unique	0	0.25	0.5	0.75	1
significant	0.25	0	0.25	0.5	0.75
moderate	0.5	0.25	0	0.25	0.5
minimal	0.75	0.5	0.25	0	0.25
none	1	0.75	0.5	0.25	0

Using Equations 1 and 2, commensurability of methods for each criterion and each pair of methods were calculated. For instance, Table 12 shows commensurability of methods scores for DES with all other methods considered, while Appendix B provides scores for the rest of the methods.

It was proposed that commensurability of methods is based on methods' uniqueness in relation to units of common language. If the difference in methods' alignments with a criterion is high, they are highly commensurable, which is viewed in Table 12 as a higher score. For instance, if one of assessed methods within a pair being compared does not align with a criterion at all, while the other method is evaluated as unique, the methods are highly commensurable. They are comparable based on their differences for that characteristic and it is easy to determine which method to use if one method does not align with a criterion at all and the other has a unique characteristic. For instance, commensurability between DES and SD methods for *ability to operate on aggregates* is one.

Table 12. Commensurability of DES with six other methods for given criteria.

Criteria/Method	DES/SD	DES/ABM	DES/BN	DES/FM	DES/SC	DES/PN
Representation of individual behavior as part of a larger system	0.5	0.25	0.75	0.75	0.75	0.25
Ability to operate on aggregates	1	0.75	1	0.5	0.5	0.5
Ability to handle uncertainty	0.75	0.25	0.25	0.25	0.25	0.25
Interaction	0.75	0.25	0.75	0.75	0.25	0
Descriptive usage	0.75	0.5	0.5	0.5	0.25	0.25
Theoretical usage	0.75	0.75	0.5	0.5	0.25	0.25
Emergence	0	0.75	0.25	0.25	0.25	0.25
Ability to represent active behavior	0	1	0	0	0.75	0.75
Sum	4.5	4.5	4	3.5	3.25	2.5
Average	**0.56**	**0.56**	**0.5**	**0.44**	**0.41**	**0.31**

It would be a clear choice between the two if one needed to use aggregated values as a single criterion. If one of methods did not align with a criterion at all, it would not be applicable to concepts within the context of the criterion, which would have permitted its elimination from consideration as a viable option clarifying situation in the context of method selection e.g. for a given component.

The criterion *ability to represent active behavior* is precise for methods like DES and ABM, making them highly commensurable. Passive entities in DES clearly do not allow for *active behavior*, while agents in ABM clearly do.

In a situation when neither of the two methods aligns with a criterion or they are assessed at the same level, their commensurability of methods is zero. On the other hand, the different cases where commensurability of methods is evaluated to zero should be noticed. In the first case, both scores are evaluated at zero because methods do not align with criterion. They are not appropriate for pseudo-triangulation in the context of that criterion. In the second case, especially where both methods scored high on criterion (e.g. 0.75 or 1) they seem more appropriate for pseudo-triangulation. The quantitative to qualitative interpretation scale of commensurability of methods is proposed in Table 13.

Table 13. Scale for commensurability of methods in relation to criterion/falsifier considered.

Level of methods commensurability	Estimate	Description
Not decidable	0	Methods are incommensurable for a given criterion/falsifier
Minimally decidable	0.25	Methods are minimally commensurable for a given criterion/falsifier
Moderate decidable	0.5	Methods are moderately commensurable for a given criterion/falsifier
Significantly decidable	0.75	Methods are significantly commensurable for a given criterion/falsifier
Fully decidable	1	Methods are completely commensurable for a given criterion/falsifier

Table 14 shows aggregated scores for commensurability of methods between all methods considered. The relatively lower overall scores for commensurability of methods between PN and SC, ABM, and DES reflect similarity of some of their characteristics. Based on this analysis, PN has potential for pseudo-triangulation where scores are 0 or 0.25 in Table 15, and additionally the PN scores in Table 17 are 0.75 or 1 (*descriptive usage* for PN and DES, and SC, and *interaction* for PN and ABM and DES).

Table 14. Aggregated scores for commensurability of methods for seven methods considered.

Methods	DES	SD	ABM	BN	FM	SC	PN
DES		0.5625	0.5625	0.5	0.4375	0.40625	0.3125
SD	0.5625		0.5625	0.1875	0.25	0.46875	0.5
ABM	0.5625	0.5625		0.5625	0.5625	0.46875	0.3125
BN	0.5	0.1875	0.5625		0.0625	0.28125	0.4375
FM	0.4375	0.25	0.5625	0.0625		0.21875	0.375
SC	0.40625	0.46875	0.46875	0.28125	0.21875		0.15625
PN	0.3125	0.5	0.3125	0.4375	0.375	0.15625	

Table 15. Commensurability of methods between PN and other methods.

Criteria/Method	PN/SC	PN/FM	PN/BN	PN/ABM	PN/SD	PN/DES
Representation of individual behavior as part of a larger system	0.5	0.5	0.5	0.5	0.25	0.25
Ability to operate on aggregates	0	0	0.5	0.25	0.5	0.5
Ability to handle uncertainty	0	0	0	0	0.5	0.25
Interaction	0.25	0.75	0.75	0.25	0.75	0
Descriptive usage	0	0.25	0.25	0.25	0.5	0.25
Theoretical usage	0	0.25	0.25	0.5	0.5	0.25
Emergence	0.5	0.5	0.5	0.5	0.25	0.25
Ability to represent active behavior	0	0.75	0.75	0.25	0.75	0.75
Sum	1.25	3	3.5	2.5	4	2.5
Average	**0.16**	**0.37**	**0.44**	**0.31**	**0.5**	**0.31**

Sub-falsifiability scores for all methods and CoMS in relation to maximum option are displayed in Table 16.

Table 16. Criteria scores for all methods.

Criteria/Method	DES	SD	ABM	BN	FM	SC	PN	Max	Highest scored method(s)
Representation of individual behavior as part of a larger system	0.75	0.25	1	0	0	0	0.5	1	ABM
Ability to operate on aggregates	0	1	0.75	1	0.5	0.5	0.5	1	SD/ABM
Ability to handle uncertainty	1	0.25	0.75	0.75	0.75	0.75	0.75	1	DES
Interaction	0.75	0	1	0	0	0.5	0.75	1	ABM
Descriptive usage	1	0.25	0.5	0.5	0.5	0.75	0.75	1	DES
Theoretical usage	0.25	1	1	0.75	0.75	0.5	0.5	1	SD/ABM
Emergence	0.25	0.25	1	0	0	0	0.5	1	ABM
Ability to represent active behavior	0	0	1	0	0	0.75	0.75	1	ABM
Sum	4	3	7	3	2.5	3.75	5	8	na
Average	0.50	0.38	0.88	0.38	0.31	0.47	0.62	1.00	na
CoMS in relation to max option	**0.50**	**0.62**	**0.12**	**0.62**	**0.69**	**0.53**	**0.38**	**0.00**	**na**

One should notice that none of the methods was evaluated with the highest score for every criterion. This implies that if all presented characteristics were required within a research context none of the methods could have provided the highest possible score without combining them. It is visible that different methods could complement each other to enhance overall approach, which is in accordance with complementarity principles. One should also notice that some of the methods would not be selected based on the highest score even once e.g. FM, SC, PN. This can be viewed as both limitations of these methods in comparison to the "winning" method and limitations of criteria to display their unique characteristics.

In Chapter 5, a case study will explore the use of falsifiers instead of criteria. If falsifiers could be developed more precisely in the context of the research requirements, they could mitigate these limitations.

A scale for CoMS is shown in Table 17. It has a purpose to give a qualitative degree of justification to different configurations with multiple methods. If CoMS is estimated as none, there is no justification to utilize multiple methods based on gain of sub-falsifiability. On the other extreme, if CoMS is evaluated to critical it means that original method(s) was/were insufficient for the falsifier or falsifiers considered.

Table 17. Scale for CoMS.

Degree of justification	CoMS Value	Description
None	$CoMS = 0$	There is no gain of sub-falsifiability when combined methods are used
Minimal	$0 < CoMS \leq 0.25$	A gain of sub-falsifiability is minimal when combined methods are used
Moderate	$0.25 < CoMS \leq 0.5$	A gain of sub-falsifiability is moderate when combined methods are used
Significant	$0.5 < CoMS \leq 0.75$	A gain of sub-falsifiability is significant when combined methods are used
Critical	$0.75 < CoMS \leq 1$	A gain of sub-falsifiability is critical when combined methods are used

The added method has then a critical effect to enhance the approach toward a desirable level of falsifiability (sub-falsifiability). In the following discussion, CoMS is estimated for a given set of methods based on a sample set of criteria to illustrate the idea using

CoMS as a degree of justification for the use multiple methods. This idea will be extended during a case study in the next chapter where falsifiers will be used instead of criteria.

CoMS as a gain of sub-falsifiability can be calculated at individual criterion/falsifier level or at a set of criteria/falsifiers level. At the individual level, CoMS is calculated as a difference between better adequate and less adequate method. For instance, for a given criterion/falsifier a score for the first method is 0.5, while the score for the second method is 0.75. CoMS would be 0.25 if the second method was used instead of the first one, but zero other way around. When comparing configurations based on cumulative scores across multiple criteria/falsifiers, CoMS can be calculated for a considered relation e.g. between methods or sets of methods, or between methods and the highest achievable sub-falsifiability for a set of considered falsifiers. CoMS could be displayed using a perspective of adding each of methods to another or as a combined view. Columns 2 and 3 in Table 18 display scores for DES and SD methods for each criterion respectively. Columns 4 and 5 display CoMS seen as adding SD to DES, and DES to SD respectively. Column 6 displays mutual CoMS, while Columns 7 and 8 display the highest possible score for DES/SD combination and methods with the highest score for each criterion respectively. It is noticed that none of the methods could support representation of active behavior. Total sub-falsifiability of combined DES/SD is 0.72. When SD and DES were added CoMS was 0.22, and 0.34 respectively. Mutual CoMS is 0.56 meaning that the gain of sub-falsifiability is evaluated as significant for the considered criteria (based on Table 17). It is noted that mutual CoMS is calculated with the same equation as commensurability of methods, which means that commensurability of methods is proportional to mutual CoMS. Figure 19 shows graphs of mutual CoMS for DES in relation with all other methods assembled based on Table 12.

Table 18. Different scoring views for choice between DES and SD.

Criteria/Method	DES	SD	SD added CoMS	DES added CoMS	Mutual CoMS	Final scores for criteria	Choice
Representation of individual behavior as part of a larger system	0.75	0.25	0	0.5	0.5	0.75	DES
Ability to operate on aggregates	0	1	1	0	1	1	SD
Ability to handle uncertainty	1	0.25	0	0.75	0.75	1	DES
Interaction	0.75	0	0	0.75	0.75	0.75	DES
Descriptive usage	1	0.25	0	0.75	0.75	1	DES
Theoretical usage	0.25	1	0.75	0	0.75	1	SD
Emergence	0.25	0.25	0	0	0	0.25	DES/SD
Ability to represent active behavior	0	0	0	0	0	0	none
Sum	4	3	1.75	2.75	4.5	5.75	
Average	0.5	0.38	0.22	0.34	0.56	0.72	

Each radar graph has eight vertices. The top vertex represents the first criterion from Table 12, and subsequent criteria are assigned clockwise. Each vertex has mutual CoMS and a line is drawn between vertices creating an area. This graph provides a visual representation of complementarity of DES with other methods, which can be informally perceived as the size of the area. Figure 20 combines the graphs from Figure 19 into a single radar graph. The outer vertices indicate the highest complementarity between DES and other method(s) for a given criterion.

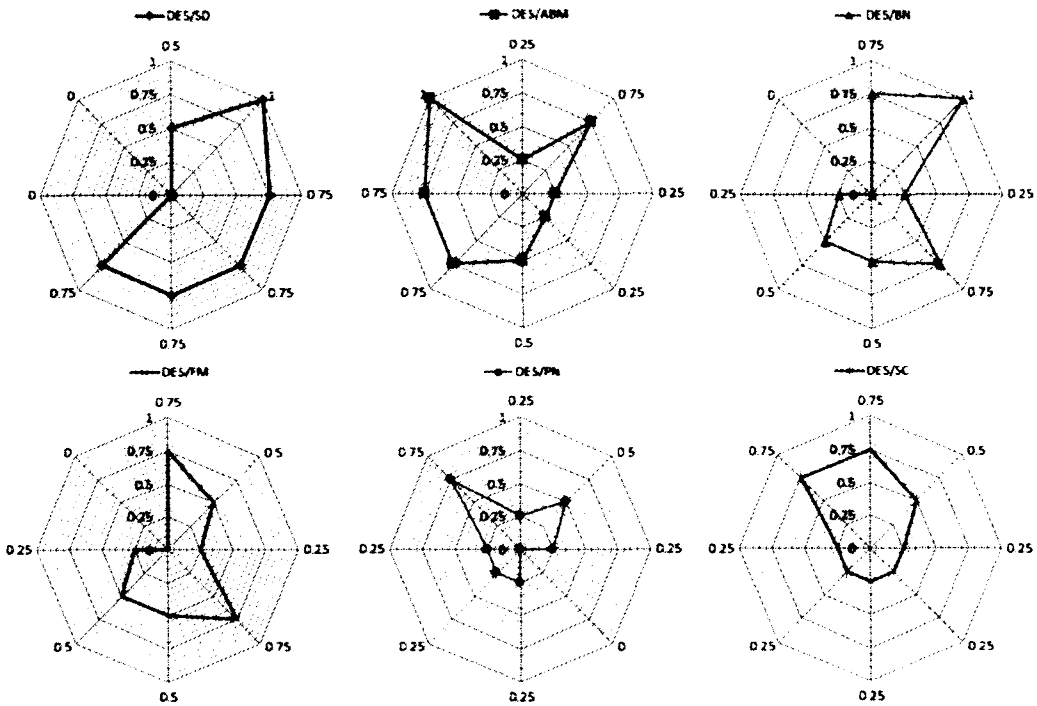

Figure 19. Mutual CoMS for DES in relation with other methods.

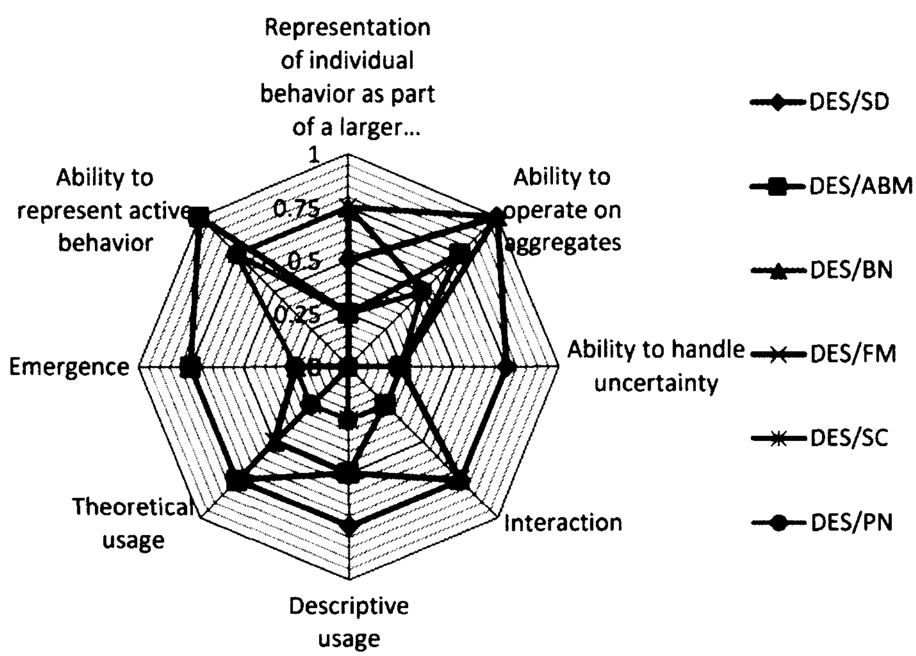

Figure 20. Mutual CoMS for DES in relation with other methods (combined).

The requirements could be translated into falsifiers and used to select method(s). CoMS can be useful to display numerical difference between configurations with different methods and to evaluate a single versus multi-method options. The problem surfaces with the practical aspects of measuring sub-falsifiability and commensurability of methods, which may be subjective because they depend on subjectivity of human during development and evaluation of falsifiers. Ranking of methods against falsifiers could lead to a better insight about which method or a set of methods is more appropriate. Complementarity of methods can be seen as a fuzzy purpose for using multiple methods within gain of sub-falsifiability and commensurability of methods boundaries, where fuzziness is related to subjectivity of knowledge about systems and phenomena at consideration, which in turn translates into research question, M&S requirements, and conceptualization.

According to Popper [45], the smaller the range (less statements permitted by a theory), the better the theory. In the context of multi-method M&S theory, this pertains to the quality of developed falsifiers to facilitate unambiguous choice of method(s). In other words, the less methods are permitted (but at least one) to a developed falsifier used for method(s) selection, the less ambiguous choice of method(s), which should translate into a better theory or an answer to a research question. This can also be used to observe evolution of M&S methods.

This also brings the idea of satisfactory level of sub-falsifiability as seen by a modeler, which is related to limitations of methods. The known fact in M&S field is that there exist no perfectly valid models, yet models can be sufficiently valid for a given purpose. In order to define sub-falsifiability one must develop a set of falsifiers, and may enhance them in the context of a set of falsification criteria. The falsifiers in the study context are falsifiable statements that describe the requirements and permit or do not permit for selection of method(s) adequate in the context of study purpose. Moreover, criteria referenced and provided in Section 2.4.2.2 can be helpful as a general view of unique characteristics of methods.

4.2 SUMMARY

A sample set of methods and criteria were analyzed in the context of commensurability of methods and CoMS. The analysis revealed that commensurability of methods is proportional to mutual CoMS. None of the methods was evaluated with a maximum score for every criterion, which contributed to credibility of complementarity principles. If all considered characteristics were required within a research context, none of the methods could have provided the highest possible sub-falsifiability score without combining methods and the resulting CoMS would have been above zero. Section 2.4.2.3 discussed the evolutionary character of methods, systems, phenomena, and unique study contexts as seen through a human dimension. Because the possibility of devising a "one size fits all" criteria that would fit different studies capturing specific requirements was ruled out, the next chapter will look into practical application of the theoretical basis using falsifiers instead of criteria.

CHAPTER 5

CASE STUDY

This chapter builds on previously developed theoretical basis of multi-method M&S approach to propose and examine multi-method M&S approach research guidelines. Concurrently developed simulation model related to a real-world problem will serve as an evaluation case. This chapter uses a case study format, which could also serve as a model of how to explore multi-method M&S approach in the future. The case study will have three dimensions:

1. A dimension driven by a multi-method M&S approach research guidelines
2. A dimension driven by the purpose of a real-world problem explored by using proposed multi-method M&S approach guidelines
3. An overarching evaluation dimension, which will assess the two other dimensions in the context of developed complementarity principles

As discussed in Section 2.4.2.3, the evolutionary character of methods, systems, phenomena, and unique study contexts as seen through a human dimension constraints the possibility of devising "one size fits all" criteria that would fit different studies, and capture specific requirements. The emphasis on objectivity and better understanding of subjectivity on one hand, and evolutionary character of dimensions on the other, lead to contradicting options to assume a constant or study-dependent set of criteria, respectively. This can lead to an inevitable conundrum, and a pragmatic view about criteria use was deliberated and adopted. The idea is that criteria can be divided into separate parts i.e. structural and behavior falsifiers tailored to each study needs and enhanced using falsification criteria from Section 3.3. The use of these types serves here as a core of the multi-method conceptualization. Moreover, complementarity principles are operationalized within multi-method M&S approach research guidelines.

The remainder of this chapter provides the following sections. In the first section, research guidelines are discussed. The next section discusses the choice of a problem that will be addressed within M&S development. In the third section, an evaluation plan is proposed and discussed. The fourth section explores a real-world problem using multi-method way of conceptualization. Finally, the next section evaluates M&S development in the context of proposed guidelines, and the final section ends this chapter with a summary.

5.1 RESEARCH GUIDELINES

Current research guidelines for multi-method approach are often method or domain (or both) specific [8, 23, 31, 35]. This view can cause constrained, domain based conceptualizations, and specific to a given set of methods assumptions. The guidelines aim at devising a process that facilitates enhanced conceptualization by providing an option for employing multi-method M&S approach, and consequently arrival at more desirable falsifiability level (sub-falsifiability). The development is supported with seemingly opposing goals: devising a robust, systemic approach, and better flexibility and creativity of modeling process. Both opposing aspects can be important within multi-method study at different stages, facilitating better chances of insight into research questions and solution(s) to problem(s).

Section 2.2 of this dissertation analyzed rationale for the use of multi-method M&S approach given by M&S community. This led to the conclusion that purposefulness of a multi-method M&S should be based on a tangible reasoning propelled by guidelines that support the decision to choose multi-method simulation over a single method based on some merits related to unique characteristic of combined methods. The overarching ideas that support use of multiple methods was characterized as complementarity of methods and it was further explored in Section 3.3 using concepts of falsifiability and commensurability. These analyses provide insight helpful during development of both systemic and general guidelines related to questions why and how to use multi-method M&S approach.

The following multi-method M&S guidelines aim to direct toward specification of MFs developed in Section 3.4 to describe multi-method M&S approach. The general steps for the guidelines are proposed as shown in Figure 21. The process can often be iterative, but for clarity, the phases are presented in a linear manner.

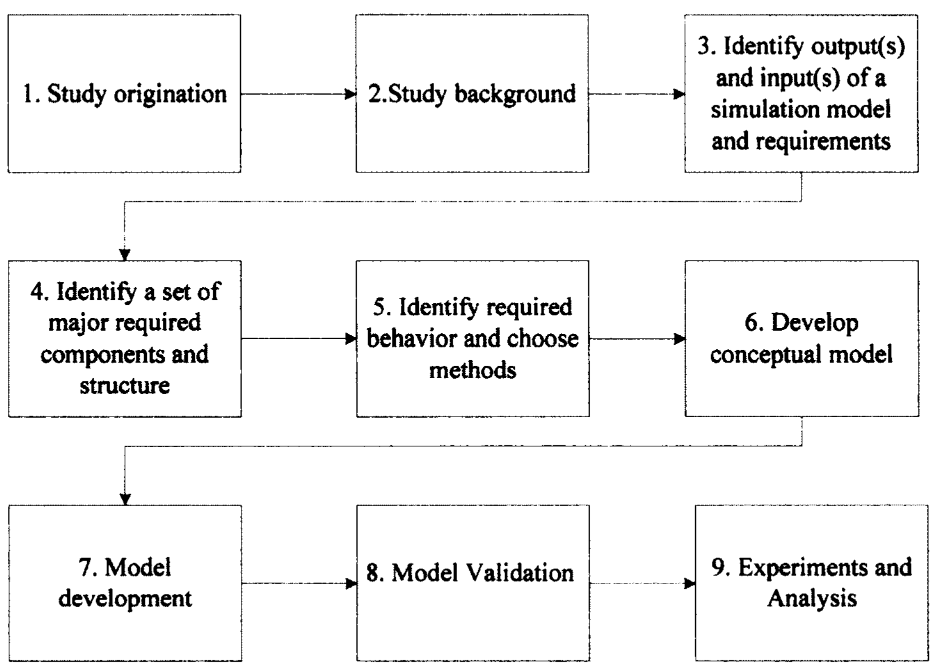

Figure 21. Proposed steps of guidelines.

The theoretical basis of multi-method M&S refers to the complementarity principles proposed in Section 3.3. The analysis of the complementarity of methods will be pursued in this case study because it aligns with confirmation of proposed theoretical basis. Because the choice of multiple methods can have confirmatory effect in relation to the purpose of complementarity as defined by Complementarity Principle 1, the reasoning for method choice is the main consideration within our guidelines. Although all steps of the process shown in Figure 21 are important, because the choice of methods from a set of considered methods occurs before low-level conceptual model specification it narrows the main focus of the considered guidelines to the first six steps. More specifically, steps one to three should provide a high-level conceptual view, steps four and five guide toward low-level conceptual model, and step six develops low-level conceptual model. The proposed guidelines provide discussion about the elements to consider and the directions that are more specific depending on the step. This discussion leads to the following main requirement of the research guidelines:

- Guide the development of reasoning to support, or not support, the use of multi-method M&S approach including development of falsifiers for method(s) selection, subsequently leading to a developed method format(s)

A general guidance considered could be used by an M&S practitioner during multi-method conceptualization, but could also be helpful to stakeholders and scientific communities by providing mechanisms for evaluation of multi-method M&S based research. For instance, disclosing that methods were selected mainly based on skills and preferences of modelers can affect perceived quality of research.

Figure 22 displays use cases of multi-method M&S approach guidelines. A conceptualization with multiple methods should be at the core of these guidelines, because all other use cases depend on it. A modeler is displayed as a part of scientific community. All, the modeler, the scientific community, and a stakeholder should be able to use these guidelines as an aid in reasoning about methods. A modeler focuses on developing new research, or retrofitting and possibly extending existing studies. Both cases should disclose information allowing for external method falsifiability, which in turn depends on multi-method way of conceptualization. This should permit for a better understanding of subjectivity of the emerging multi-method M&S approach.

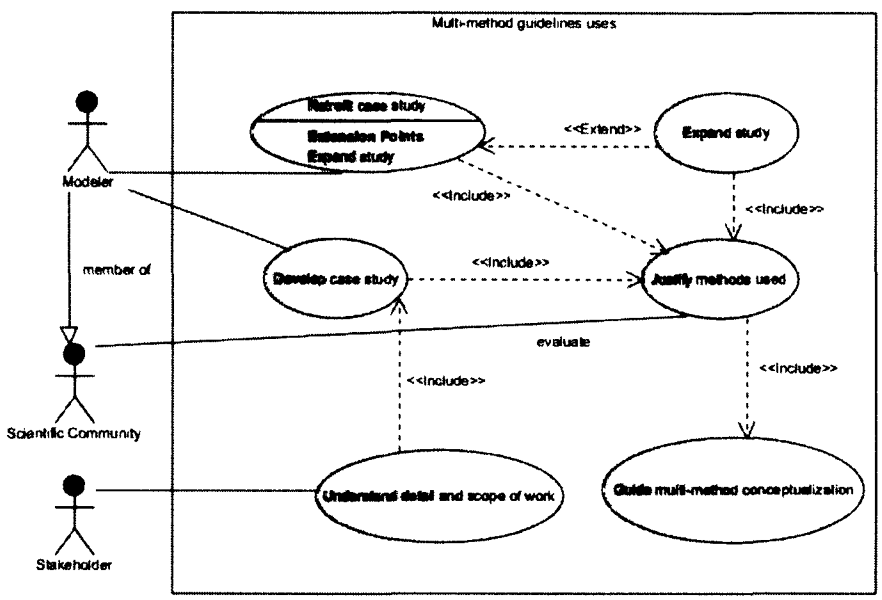

Figure 22. Use cases for multi-method M&S research guidelines.

The guidelines are developed with the focus on answering dissertation's research question and are considered a work in progress. Nonetheless, it is hoped that they poses practical value for the M&S community. Following next, steps one to six are described, while steps seven through nine are briefly introduced and the references are provided.

5.1.1 Study Origination

Aspects related to a stakeholder may or may not need to be considered, which depends on situation. Assuming that there exists at least one stakeholder other that modeler, the first step of the research guideline process should aim to understand stakeholders' expectations. Initial consultation and problem definition is rarely stated clearly. Because of various possible levels of expectations and generalization of problems Robinson [113] advised to clearly identify the purpose. Depending on the type of a problem, this step may require iterative approach in form of meetings with project manager, analyst, and subject matter experts (SMEs). It is appropriate to decide if the simulation is the right choice. The dialog between all parties often involving going over help lists that consist of important for modeling and simulation (M&S) practitioner questions should result in definition of a problem, which is understandable to all. That leads to definition of overall objectives of the project, like what type of questions should the project answer, for instance:

How can system performance be improved?

How can future design problems be avoided?

How can true system requirements be predicted?

How can system behavior be understood?

This step should also display how human dimension influences the study design process. It is necessary to define stakeholder requirements and determine necessary resources. Stakeholders may be interested in various types of the study to conduct 1) an atomic study that explores a system or a phenomenon, 2) an atomic study that attempts to confirm other past research related to a system or phenomenon via triangulation, or 3) a concurrent triangulated study (multi-case) 4) multi-phase study that includes both exploration and triangulation. Options one and two can be considered as simplified cases of option 4. The multitude of ways that the study can be designed based on MFs provides general guidelines for multi-method research.

Interaction with a stakeholder to a various degree spans throughout all the phases of study. Stakeholder can also influence research methods, which can constrain or change the research. Finally, the desirable or required time to conduct the study is an important factor, and it is often provided or known based on the type of activity.

The availability of software can significantly influence available methods, hence possible options should be considered as early as possible. Simulation software often allows using a single, two, and rarely multiple M&S methods. Usually, software with more open IDE architecture permits to extend software capabilities. Considered software can also characterize some specific capabilities of methods, for instance their visual display capabilities.

Both theoretical (e.g. mathematical, logical, and formalism) and practical (programming, scripting, visual modeling) knowledge about methods are needed to be able to successfully utilize M&S methods. It is noted that if a very large number of methods is considered this could have also negative consequences related to complexity of possible options. For instance, if only five main components and five methods are considered, and assuming that each component can be modeled only using a single method, initial number of combinations is 3,125 (5^5). If the number of components is increased to six this would have generated 15,625 combinations (5^6).

5.1.2 Study Background: Phenomena, System, and Research Question(s)

This step usually requires conducting qualitative analysis to define, redefine, clarify and analyze phenomena and system(s) involved in order to understand study purpose, problem(s), and to develop RQ(s). For instance, an M&S practitioner tasked with modeling of system that involves social phenomena conducts qualitative analysis, which can involve "information gathering by direct observation, analysis of documents and sources, and interviews" in order to gain familiarity about phenomena [41]. If previous theoretical work and implementation of simulation model is found in literature, it is examined for reuse within the study. Moreover, depending on type of research this step can also involve creation of high-level conceptual models showing necessary dependencies. This approach allows for flexibility during initial conceptualization within study background. It would not be appropriate to constrain conceptual derivation. Clearly, different high-level conceptual methods can be helpful and used to aid in this process. If primary data collection is possible, modeler can be more creative when developing RQs.

5.1.3 Identifying Outputs and Inputs of a Simulation Model and Requirements

Ören [145] offers a systemic view on using simulation to finding values of two out of three types: output, input, or state variables given that two out of three types are known. Three types of objectives are identified and relate to an *analysis* problem, i.e.

generating model behavior (output), a *design* problem (states satisfying input/output pair), or a *control* problem (searching for a necessary system input control). What can be defined as input, output, or states depends on purpose, not a simulation model. In a real-world (especially when involved with social phenomena), two out of three elements are often not given and solving a problem may involve iterative filling gaps both at conceptual and simulation experiment levels. Robinson [146] defined inputs as model elements that are manipulated during simulation run to obtain desired effects determined by study objectives, while outputs are the results from simulation run to see if objectives are achieved and if not, why. This step aims at identifying outputs from the simulation model that are required to explore and/or answer RQs, and identifying input(s) to the simulation model, which should be used to manipulate model's conditions in order to explore and/or answer RQs. Next, a set of requirements should align study background with input(s) and output(s).

5.1.4 Identification of Major Required Components and Structure

The multi-method M&S approach requires looking beyond the concept, while still developing it to generate insight into which methods to use. The main differences between conceptual modeling for a single method approach and multi-method M&S approach are reflected in the necessity of a multi-method way of conceptualization. Steps four and five are proposed to guide toward a low-level conceptual model and should end up with a specification of MFs. This part of multi-method conceptualization investigates scope and structural dependences of components of a model and employs constraints related to structural characteristics, at the same time guiding toward appropriate methods.

This step should identify major simulation model components and their subcomponents that can realize output, input, and phenomena. The steps shown in

Figure 23 are iterative, and only displayed as linear for clarity. The process starts with identifying major simulation model components based on analysis of RQ(s), output(s), input(s) and requirements. The most important components should allow manipulations related to input requirements to produce simulation output that have potential to answer RQ(s). If they do not, this indicates a need to go back to the study background section to explore further phenomena.

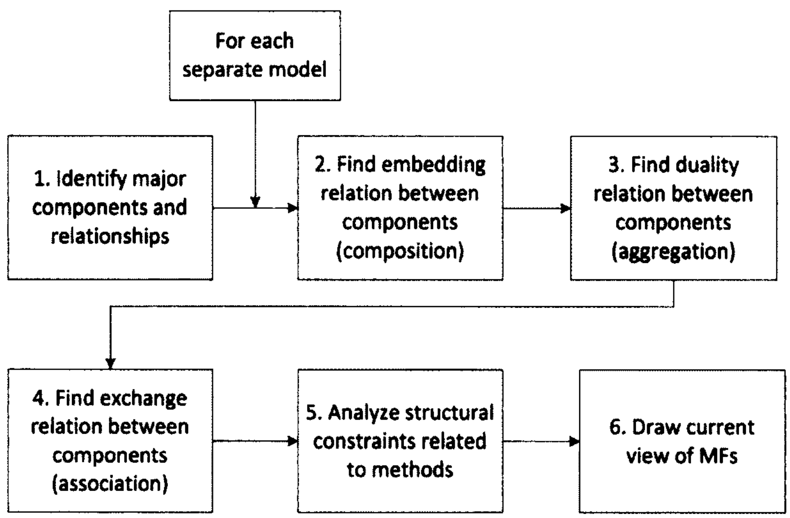

Figure 23. Steps to defining model structure.

Only compositions and aggregations create constraints related to structural falsifiers. Association is the most general relation and does not have any constraints related to a structure, but provides information about required connectivity between components. It is noted that if some components are not connected, they become separate models. Similarly, there may be a situation where two or more separate assemblies that each consists of multiple components are created, which also indicates number of separate models needed. All compositions, aggregations, and associations should be conceptually acknowledged providing high-level structure that can represent output, input, and phenomena. Next, structural characteristics are turned into structural falsifiers and methods are evaluated against them. A falsifier has to be evaluated as false in order to accept considered method as an option for a component. Finally, structural falsifiers related to method characteristics must be developed and analyzed and current structural view using MFs must be drawn.

5.1.5 Required Behavior and Choice of Methods

This part focuses on conceptualization of behavioral requirements of components, including interaction points between components. The aim is to identify critical characteristics of methods needed in order to capture behavior of components and their interaction that allows answering RQ, which at the same time can facilitate insight into method choices and development of MFs. The main interaction types were recognized in Section 2.4.2.1. These were focused on data exchange by replacement, aggregation/ disaggregation, transformation, and triggering and listening to conditions. The main steps for this phase are shown in Figure 24.

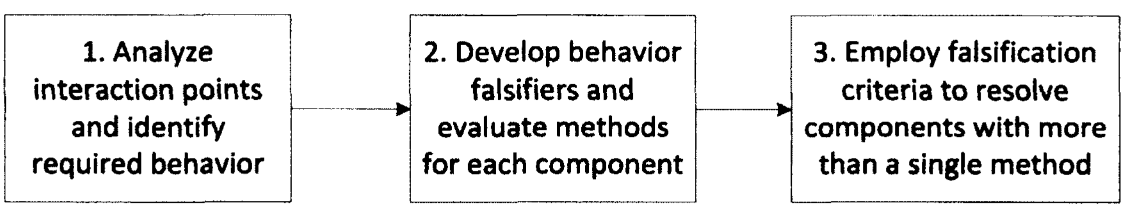

Figure 24. Analysis of behavior.

The first step combines conceptualization of behavior and interaction points between components. This should lead to critical characteristics of each component. Next, these characteristics are turned into behavior falsifiers and methods are evaluated against them. Falsifier has to be evaluated false in order to accept considered method as an option for a component. In the last step, ties must be analyzed if more than a single method evaluates falsifier false. It is proposed to combine falsifiers with falsification criteria discussed in Section 3.3 and evaluate scores for each method. the scope was already covered in the section related to the structure of a model. Moreover, because methods can be associated with externally developed graphical representation, the evaluation of visual display criterion can be constrained to the prebuilt features of modeling elements associated with considered methods. Ease of developing a simulation model affects time to develop; hence, only time to develop will be evaluated as a falsification criterion.

The scoring (level of disagreement) of falsifiers in the context of falsification criteria is proposed in Table 19, while importance of falsification criterion is scored using Table 20. The overall sub-falsifiability score is a multiplication of both scores.

Table 19. Scale for scoring of methods in a context of falsification criteria and falsifiers.

Level of disagreement	Score
Very strongly disagree	1
Strongly disagree	0.75
Disagree	0.5
Somewhat disagree	0.25
Do not disagree	0

Table 20. Level of importance of falsification criterion on component.

Level of importance	Weight	Description
Insignificant	0	The falsification criterion bears insignificantly on the component
Minimal	0.25	The falsification criterion bears minimally on the component
Moderate	0.5	The falsification criterion bears moderately on the component
Significant	0.75	The falsification criterion bears significantly on the component
Essential	1	The falsification criterion is essential for the component

The outline for calculation of CoMS is proposed as follows. For each falsifier f, falsification criterion j, and each method i calculate sub-falsifiability score

$$\boldsymbol{F_{fji} = I_{fji} * (F_e)_{fij}}, \tag{3}$$

An importance of falsification criterion on component is represented as I_{fji}, and $(F_e)_{fij}$ stands for evaluation score (level of disagreement). Second, for each falsifier f and each method i calculate cumulative sub-falsifiability score

$$\boldsymbol{(F_c)_{fi} = \sum_{j=0}^{j=m} F_{fji}}, \tag{4}$$

A size of falsification criteria set is represented by a letter m. Next, calculate cumulative sub-falsifiability score for each method i

$$\boldsymbol{(F_c)_i = \sum_{f=0}^{f=p} (F_c)_{fi}} \tag{5}$$

The number of falsifiers is represented by a letter p. Then, using a cumulative sub-falsifiability score for each falsifier f, assess the maximum sub-falsifiability score.

$$\boldsymbol{(F_{Max})_f = Max((F_c)_i)} \tag{6}$$

Finally, assess the maximum cumulative sub-falsifiability score.

$$F_{Max} = \sum_{f=0}^{f=p} (F_{Max})_f \tag{7}$$

Similarly, calculate the overall maximum achievable sub-falsifiability score, which is a maximum score that methods could have been evaluated, including importance of falsification criterion.

$$Max(F_c)_{fi} = \sum_{j=0}^{j=m} Max(F_{fji}) \tag{8}$$

$$Max(F_c)_i = \sum_{f=0}^{f=p} Max(F_c)_{fi} \tag{9}$$

$$Max(F) = \sum_{i=0}^{i=n} Max(F_c)_i \tag{10}$$

Given all the scores calculated above, calculate the CoMS as a difference between the actual maximum cumulative sub-falsifiability score and the cumulative sub-falsifiability score for each method i

$$CoMS_i = F_{Max} - (F_c)_i \tag{11}$$

This score can be used to get an insight into advantage of using particular options. This is also operationalized complementarity principle, which can be used e.g. to show advantage of multi-method M&S approach over a single method. Also, calculate the CoMS as a difference between maximum achievable sub-falsifiability score and actual maximum cumulative sub-falsifiability score to see how closely the methods fulfilled the requirements.

$$CoMS_T = Max(F) - F_{Max} \tag{12}$$

It is also advised to normalize results using $Max(F)$ as maximum value for clearer comparison.

The problem of subjectivity during multi-method conceptualization relates to the fact that simulation models with different configurations of methods do not exist, and cannot be fully observed unless implemented. The proposed approach that employed falsifiers and falsification criteria is only approximations based on reasoning on a future simulation model as discussed by Robinson [113]. As mentioned earlier, because low-level conceptual models contain strong implications about M&S methods, the choice of methods from the set of considered methods should occur in steps four and five, which means that at this point method formats should be determined.

5.1.6 Low-level Conceptual Model(s) Development

Once the methods are assigned and MFs developed, the low-level conceptual model can also be developed. During the method choice process, highlighted conceptual features should be expanded upon in this step. Upon reaching this point of the process, the structure and behavior should be known enough to develop a low-level conceptual model using modeling elements typical for selected methods.

Different methods have different modeling blocks and/or characteristics. For instance, Robinson [113] specified four types of components within DES as entities, activities, queues, and resources. SC would consist of states (simple, composite, final, and history), transitions, initial points, and branch objects. SD often uses causal loop diagrams during conceptualization, while the main implementation building blocks of SD are stock, flow, general variable, and SD often supports lookup tables. There are a few possible ways of looking at developing conceptual model within ABM [1, 147], but common component types of ABM is a population of agents and agent specification itself. BN has typically nodes (chance, decision, and value) and arcs that connect them. Moreover, types of components have dependencies, which must be obeyed.

5.1.7 Steps Seven to Nine

The development, design of experiment, and analysis of multi-method simulation model depend on selected methods and software used. If a conceptual model was developed using M&S software, this provides a smooth transition into specifying all the necessary code that connects components structure and behavior into an executable simulation model reflecting desired inputs, states transitions, and outputs. A detailed guideline for model development, design of experiment, and analysis is outside of the scope of this work. Please refer to Section 2.4.2.1 and M&S textbooks and publications for more information, for instance see [16, 17, 31, 96, 148, 149].

Similarly, it is difficult to prescribe detail guidelines on validation that can be applied to multi-method M&S approach. Please refer to Section 2.4.3 for discussion about simulation model validation. As pointed there, the key aspect to consider when conducting validation of a multi-method simulation model is the fulfillment of the purpose with a sound and holistic perspective on a study. This may indicate that the level of validity might depend on the standards used, which should define their limitations and the proper context of their usage. Verification and Validation (V&V) processes expose scientists' skills about a subject studied and M&S skills for representation of the researched phenomena.

5.2 SELECTION OF A PROBLEM FOR M&S DEVELOPMENT

An undertaken real-world problem through proposed research guidelines serves as a data layer for evaluation of multi-method M&S approach. The choice of the real-world problem selection is discussed first.

The simulation model should be able to explore non-trivial situation, where more than a single method seems appealing but also problematic. The multi-method M&S approach has already been used to represent technical phenomena, for instance: in manufacturing [4]; healthcare [5, 6]; and supply chain systems [7]. In hybrid Discrete Event Simulation/System Dynamics (DES/SD) models, methods often complement each other. For instance, DES offers better representation of detail complexity, and SD allows for easier representation of dynamic "feedback" effects [8, 9]. Technical phenomena that are considered in these studies appear within well-bounded levels of analysis. A situation is more problematic if social and technical phenomena are mixed. Social phenomena can be very difficult to understand. A social system is often characterized by high complexity that arises from more than one level of analysis and fuzzy boundaries [150]. A representation of social phenomena with a combination of different methods seems intuitively appealing especially in the situations where questions pertain to descriptive as well as theoretical aspects of a social phenomenon. It is likely that some M&S methods would serve better in addressing theoretical, and some answering descriptive questions. There is also a possibility that single research questions pertain to both theoretical and descriptive characteristics. Recently, communities that were usually focused on representation of technical phenomena also started exploration of social concepts [18, 63, 151]. Unfortunately, the usefulness of a multi-method M&S approach to represent social phenomena is not well understood and demonstration of its advantages is not an easy endeavor. For all the above reasons, the case study will focus on a real-world problem that consists of both technical and social phenomena.

The real-world problem will pertain to a return to work phenomenon of Social Security Disability Insurance (SSDI) enrollees in the United States. A multi-method simulation model may or may not be needed to capture system conditions and enrollee behavior.

A general purpose of this work exists in a higher logic of evaluation of research guidelines and theoretical basis, while at the lower level of this scenario it pertains to the use of a simulation model to investigate aspects of the return to work phenomenon. The problem with representation of enrollee's decision to work is not trivial. In order to get insight into the "why" enrollees consider to work, a common approach involves qualitative analysis of issues related to disability to provide further hypotheses for quantitative analysis [152]. A different research path to the above approach is proposed in this case study. A simulation-based study is considered, in which generated pseudo qualitative data will be processed at a higher level, providing a more holistic view of the system. This approach could potentially complement traditional data collection and analysis. With the current state-of-the-art, simulation at multiple levels of analysis can combine aspects of both qualitative and quantitative empirical worldviews. The developed simulation model will be used to answer a sample research query developed in Section 4.4.2 that examines how the attitude of an enrollee toward work incentives related to health improvements, money, and vocational assistance can affect the return to work phenomenon for 18 to 39 year old SSDI enrollees (at enrollment). The choice of the population type is related to availability of relevant research literature and data. Moreover, young population of enrollees has the most potential to return to work and value for the stakeholders, because if they do not return to work they will use SSDI funds for a very long time. The research question will be measured as the total percentage of population with benefits terminated for work.

5.3 EVALUATION PLAN

The evaluation will have two major sections. The first section uses evaluation questions and statements, while the second one attempts to pseudo-triangulate the developed simulation model using a single method that was not selected.

5.3.1 Questions and Statements

The response to questions and statements will look directly at how well the requirements of research guidelines are supported by the developed process. The proposed steps within the research guidelines are based on common M&S steps, but additional multi-method features during conceptualization make them unique. These additional features are the main subject of the analysis, but evaluation of the scenario is also included. The following questions and evaluation statements will be explored.

1. Did the research guidelines facilitate generation of information for external method falsifiability?

This question examines if the proposed steps (especially structural and behavior conceptualization that involves multiple methods) can provide a good base for a critique of the selected methods. This is important to evaluate and improve a particular study and multi-method M&S research practice in general.

2. Within this case study, did MFs provide high-level description of multi-method M&S approach?

This question reviews whether MFs supported these research guidelines. No other guidelines were found that would support this generalized way of representing multi-method M&S research approach including a structure of a multi-method simulation.

3. Was the use of multiple methods justified based on operationalized theoretical developments from Section 3 related to complementarity principles?

This question explores developments from Section 3. The choice of methods from the set of considered methods occurs somewhere between the research background step and before the conceptual model is fully developed. The theoretical developments in Section 3.3 led to proposed approach that estimates complementarity of methods using CoMS, which will be examined within this question and evaluated in the context of complementarity principles. This question examines also how selective were developed falsifier during choosing methods and falsification criteria.

4. Based on the case study, evaluate the sentences that apply to the multi-method approach:

    a. The use of multiple methods was justified by examining if a similar insight could not be generated without using a multi-method simulation model (adapted from Greene [27]).

b. The use of multi-method M&S approach facilitated the use of different perspectives e.g., insider and observer's views (adapted from Onwuegbuzie and Johnson [118]).

c. The weakness from one approach were compensated by the advantages from the other approach (adapted from Onwuegbuzie and Johnson [118]).

Questions 1 through 3 and Statements 4a, 4b, and 4c investigate aspects of complementarity of methods in different contexts. Moreover, the described real-world problem used in this case study was summarized as a paper and submitted to a Winter Simulation Conference 2014 to obtain some independent external evaluation source [153]. Three reviewers evaluated the paper by answering questions 5, and 6. The evaluation questions 5 and 6 can provide also insight into value of multi-method M&S approach as seen by M&S experts.

5. What is the novelty presented in the paper?
6. How do you evaluate the potential impact on the application field?

5.3.2 Demonstration

A structural and behavioral analysis should lead to selection of a method or a set of methods for conceptualized components that will be used to develop a simulation model. The question arises if the selection process provided a sufficient justification. The purpose of this demonstration is to pseudo-triangulate the developed simulation model using a method not previously selected to explore possible biases during the selection process. This effort can demonstrate whether or not the selected solution can capture inputs, outputs, and possibly the results of the developed simulation model. It examines feasibility of representation in spite of demonstrated weaknesses of a not selected method in the research context. Pseudo-triangulation of all possible configurations that were not selected is outside of the scope of this work (this would require 624 configurations). A single simulation model that includes a not selected method will be developed and evaluated.

5.4 M&S DEVELOPMENT

This section employs proposed in Section 5.1 research guidelines to explore return to work phenomenon of SSDI enrollees in the U.S.

5.4.1 Study Origination

As discussed in Section 5.1.1, the study origination will consider aspects related to stakeholders expectations, choice and availability of modeling software, and considered methods.

- Stakeholder and/or researcher expectations

This is a very important step in real M&S practice. In this case, the researcher expectation is to use a real-world problem to provide a data layer for evaluation driven by the purpose of dissertation requirements. Dissertation committee members could be considered as stakeholders in this case because they evaluate this work's quality. Because the atomic study that explores a phenomenon is considered, triangulation that involves additional modeler(s) can be eliminated. This means MFI is not applicable. The total time to conduct the study is assumed below two months.

- Choice and availability of modeling software

The comparison of different modeling platforms and differences in implementation of methods within these environments could produce a separate dissertation, and it was considered outside of the scope of this work. AnyLogic® simulation software (version 6.9) is used for this case study for two main reasons. The first reason is related to its multi-method capabilities. This IDE based software includes ABM, DES, SD, SC, among other domain specific libraries, and permits to include additional methods because it is open to Java based code libraries. The second reason pertains to accessibility to a "student version" of this software. It is pointed out that used software provides graphical layer associated with methods (modeling blocks). For instance, SD and SC have full spectrum of graphical blocks available e.g. stocks, flows, dynamic variables, and states, transitions, branching block respectively. ABM provides a basic structure (shell) that holds agents, where agents can be associated with a picture that may be located in chosen graphical space, but it does not have to have graphical representation. Multiple options of using ABM make the visual representation of internal behavior and external view of agent customizable to the particular purpose, which often requires more coding than other methods. BN is not included in AnyLogic®. It can be implemented in Java code within AnyLogic® (which can be time consuming) or integrated with implemented Java based software or library. The second approach will be considered (using Genie ® tool) during method selection. The Genie® tool is free and provides an easy-to-use graphical interface to develop networks. It can be easily imported to AnyLogic® as a jar library.

- Considered methods

The number of methods considered is limited to five to avoid too many possible combinations. ABM, DES, SD, SC, and BN are considered because of a good theoretical and practical knowledge about them. The brief description of methods considered is provided in Appendix C (mainly as implemented in AnyLogic®). The summarized view using criteria developed in Section 2.4.2.2 provides additional insight. It is noted that some methods offer low-level conceptual view (DES, SC, BN) while others guide toward low-level from high-level conceptual view e.g. causal loop diagram toward SD.

5.4.2 Study Background: Phenomena, System, and Research Question(s)

Social Security Disability Insurance (SSDI) is a benefit available in the U.S. to people with disabilities. It can often be combined with Supplemental Security Income (SSI), Medicaid and Medicare [152]. Statistical data from 2010 indicate 64.9% of SSDI recipients aged 21 – 44 years, 50% of SSDI recipients, aged 45 – 54 years, and 31.4% aged 55 – 64 years, were also Medicaid/Medicare beneficiaries, translating into significant spending [152]. In 2008, estimated federal expenses on workers with disabilities were approximate at $357 billion while state spending estimated $71 billion (90% of which on Medicaid) [154]. The cost of SSDI benefits for workers with disabilities and their dependents was $127.9 billion [154]. Difficult economic situations can increase the rate of application for disability benefits. By the end of 2001, 5.3 million disability benefits were provided by Social Security Administration (SSA) with an average of 57,600 new recipients per month [155]. This number increased to 7.1 million by the end of 2007 with an average of 68,900 new recipients per month, and to 8.8 million in mid-2012 with an average of 82,400 new recipients each month [156]. A typical SSDI enrollee stays in the program for many years. Three major paths to exiting the program are as follows: 1) death; 2) reaching full retirement age; and 3) no longer meeting medical disability standards. Data show that in 2004, 12% of beneficiaries left the program for the above reasons [157]. The 10-year follow-up study of SSDI enrollees provides information that benefits were terminated for 3.7% of recipients after they found work [158]. Moreover, the data indicate that a majority of SSDI enrollees who found work while using work incentives do so in the first five years from being awarded [158].

Upon award of SSDI benefits, a disabled person becomes eligible for federal and state programs that include vocational rehabilitation and employment assistance. There are four major work incentive programs: 1) Work Incentive Planning and Assistance Program (WIPA); 2) Protection and Advocacy for Beneficiaries of Social Security Program (PABSS); 3) Ticket to Work Program; and 4) Social Security / Vocational Rehabilitation Program. Kregel [159] provided an overview of outcomes from current research related to these programs. Livermore et al. [160] found a consistent and significant relationship between the receipt of WIPA services and an increased likelihood that a beneficiary will be employed and experience a reduction in benefits in the future. Once SSDI is awarded, there is a 24-month waiting period for Medicare entitlement.

Development of a coordinated and comprehensive system of incentives is challenging because it must be tailored to many different groups with specific needs and characteristics, but may bring profits for recipients and providers, assuring that the money is spent wisely, bringing savings for the budget and at the same time giving the best possible care and options for the disabled population. As pointed out by Kennedy et al. [152] evidence from outside of the U.S. suggests that an introduction of vocational rehabilitation and return-to-work goals at the beginning of an SSDI determination process can encourage successful workforce reintegration. Because the amount of outpatient services used for Medicare enrollees is negatively associated with employment, one should expect savings of government money [161]. Moreover, findings support efforts to encourage work because of associations between employment with better health, healthy behaviors, and lower costs [162].

Liu and Stapleton [158] discussed the problem of short term evaluations based on “cross-sectional” statistics, and highlighted the need for a more detailed view of beneficiaries leaving SSDI for a work through longitudinal studies. They discussed the complexity of capturing dynamic changes in possible multiple transitions from significant gainful activity (SGA), non- SGA, and an unemployed status. These transitions depend on the amount of enrollee earnings each month. Kennedy et al. [152] point out the need to measure enrollee employment in terms of earnings, which would require merging (e.g. unemployment and Medicare) data.

In their analysis, Liu and Stapleton [158] identified five stages: SSDI awarded, first time employed, trial work period (TWP) completed, benefits suspended after finding work, and benefits terminated after finding work. Percentages of SSDI awardees for each stage from 1996 to 2006 were traced, giving also a cumulative percent at the 10-year mark. Liu and Stapleton [158] focused mostly on reporting what has happened using data analysis, which provided initial clues about how the system behaves. Ben-Shalom and Mamun [163] focused on four milestones: service enrollment, start of TWP, TWP completion month, and the first suspension month. They used only “complete” cohort data for 60 months after the first SSDI award, taking the research one-step further and providing estimated probabilities of service enrollment at considered stages, as a function of age and type of disability. However, this more informative approach still lacks the answer to the question "why" enrollees would consider returning to work. The reviewed literature provides insight into systemic rules, but it does not offer a theoretical model that can be reused in simulation of return to work phenomenon. No implementation of simulation model was found that considers this topic altogether, hence the reuse or extension of previous M&S work is not applicable in this case.

In order to get insight into the "why" question, a common approach involves qualitative analysis of issues related to disability to provide further hypotheses for quantitative analysis [152]. A different approach is proposed here i.e. a simulation based study is employed to generate data that can be processed providing a more holistic view.

Evidence indicates that younger beneficiaries who have received benefits for a shorter period are more likely to become employed [159]. Work incentive programs focus on different aspects (e.g. health improvement, money incentives, and vocational assistance), which can contribute to a return-to-work. The research question examines *how much attitude toward incentives related to health improvement, money, and vocational assistance affect return to work phenomenon.* This will be measured for the younger population (18 to 39 year old) of SSDI enrollees' as the difference in percentage that remained "on the rolls".

5.4.3 Identifying Outputs and Inputs of a Simulation Model and Requirements

The following RQ was identified in the previous step: *how much attitude toward incentives related to health improvement, money, and vocational assistance affect return to work phenomenon?* This will be measured for the younger population (18 to 39 year old) of SSDI enrollees' as the percentage that remained "on the rolls".

Using this information the output value is specified as a *percentage of population (of interest) with benefits terminated for work after given period*, which then can be used to compare differences between different cases. Based on the study background section, a few requirements that pertain to output can be identified. The output should arise from simulated (18 to 39 year old) population or a sample population of SSDI enrollees. The time span will be from 1996 to 2006 based on available data from relevant research, e.g. [164]. The output should be calculated based on stochastic character of individual SSDI process related to systemic phases and transitions as described by Liu and Stapleton [158] and Ben-Shalom and Mamun [163], and corresponding decision points of individual active enrollees to look for a job, and enrollee work status (e.g. SGA, non-SGA). Descriptive accuracy of the output is important but not at cost of detail. For instance, the highly accurate predictive algorithm with a low resolution is not desirable at this point of analysis, but rather a model in which we can identify factors and processes relevant to the theoretical view. The simulation should explore stochastic characteristics of output hence output analysis should determine 95% confidence of the output at the end time mark (1996-2006) based on 30 runs or more, and determine variability of the process via 2D histogram. Moreover, it is desirable that simulation model provided visual display of SSDI process and work status for validation purposes.

In order to get insight into RQ, *attitude toward systemic incentives related to vocational assistance, money, and health improvement programs* should be represented as parameters within some identified adequate scale. For instance, inputs of enrollee attitude toward different incentives could be scaled on ordinal scale from sufficient to insufficient. Psychological level should be considered e.g. beliefs, attitudes, and intentions. Inputs will reside in the enrollee's modeled attitude and relate to incentives in relation to the consideration of work. Low-level input is important to investigate hidden aspects related to transitions controlled by human behavior.

5.4.4 Identification of Major Required Components and Structure

The following four main components were identified based on requirements: population of individual enrollees, SSDI process, enrollee work status, enrollee attitudes. Because the population of enrollees consists of individual enrollees as subcomponents this is considered as an additional relationship.

- Component 1 describes population or a sample population of SSDI enrollees 18 to 39 year old as an individual active decision makers
- Component 2 describes SSDI process of phases and transitions describing return-to-work phenomenon
- Component 3 describes work status of enrollee e.g. working / not working state
- Component 4 describes enrollee's attitudes toward incentives on the scale from sufficient to insufficient

The relationships between components are identified as shown in Table 21 and graphically in Figure 25 as compositions or associations using UML class diagram notation.

Table 21. Relationships between main components.

Components/ subcomponents	Population of enrollees	Enrollee	SSDI process	Enrollee work status	Enrollee attitudes
Population of enrollees		has			
Enrollee	is within		follows	has	has
SSDI process		is for		influences	is influenced
Enrollee work status		is within	influences		
Enrollee attitudes		are within	influences		

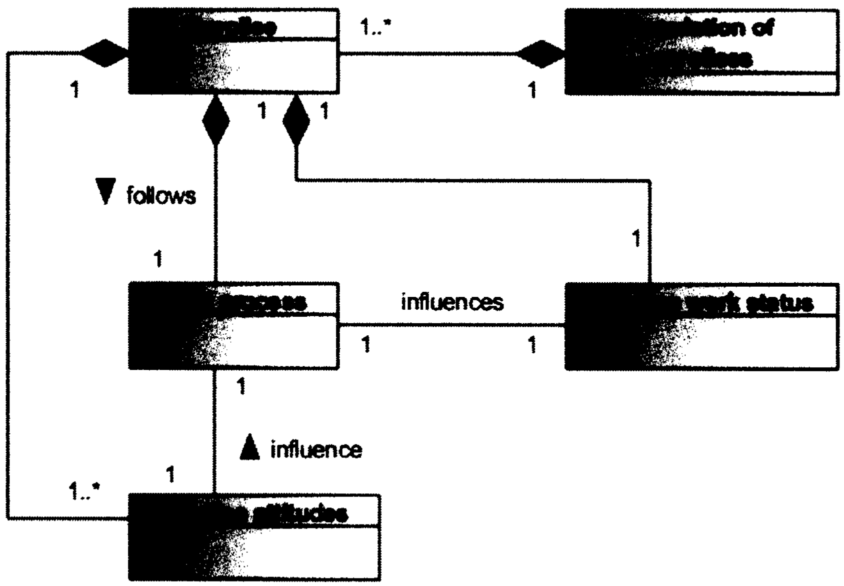

Figure 25. Relationships between main components.

This structural view provides also a set of requirements based on which one can develop structural falsifiers in relation to methods considered for each relevant component. First, statements describing structural requirements are presented.

1. Population of enrollees is a composition of enrollees
2. Individual enrollee is a composition of SSDI process, enrollee work status, and enrollee attitudes
3. SSDI and enrollee work status influence each other (association)

4. Enrollee attitude influences SSDI process (association)

Next, considering the above statements as a base, structural falsifiers in relation to methods considered for each relevant component will be developed. Component 1 uses composition relation (MFVI) which can constrain the possible solution, whereas other components use association relation (MFIV) which is the most general and do not constrain the structure. Because of this, only structural requirements one and two are useful to develop structural falsifiers in the context of Component 1.

Structural falsifiers:

- No method represents a population of individual enrollees as a composition of individual enrollees
- No method represents individual entities as a composition of SSDI process, enrollee work status, and enrollee attitudes

The following are the results of analysis in relation to the falsifiers:

Component 1:

DES

- DES represents a population of individual enrollees as a composition of individual enrollees (entities)
- DES does not represent individual entities as a composition of SSDI process, enrollee work status, and enrollee attitudes

SD

- SD represents a population of individual enrollees as a composition of individual enrollees (an array of stock and flow models)

- SD represents individual entities as a composition of SSDI process, enrollee work status, and enrollee attitudes. Individual stock and flow models are used as a composition of SSDI process, enrollee work status, and enrollee attitudes

ABM

- ABM represents SSDI population of individual enrollees as a composition of enrollees (agents)
- ABM represents individual entities as a composition of SSDI process, enrollee work status, and enrollee attitudes. Individual agents are used as a composition of SSDI process, enrollee work status, and enrollee attitudes.

BN

- BN does not represent a population of individual enrollees as a composition of individual enrollees
- BN does not represent individual enrollees as a composition of SSDI process, enrollee work status, and enrollee attitudes

SC

- SC does not represent a population of individual enrollees as a composition of individual enrollees
- SC does not represent individual enrollees as a composition of SSDI process, enrollee work status, and enrollee attitudes

Table 22 displays results of analysis of structural requirements. Number one means that the answer to a falsifier is false (negation of sentence that is set as negation), zero means true.

Table 22. Results of analysis of structural requirements on methods.

Component/Method	DES	SD	ABM	BN	SC
Population of enrollees	0	1	1	0	0
SSDI process	1	1	1	1	1
Enrollee work status	1	1	1	1	1
Enrollee attitudes	1	1	1	1	1

Analysis of structural falsifiers allowed eliminating DES, BN and SC as possible options for Component 1. Figure 26 displays partial MFs. After structural analysis, methods are not fully determined for all components indicated as Xi.

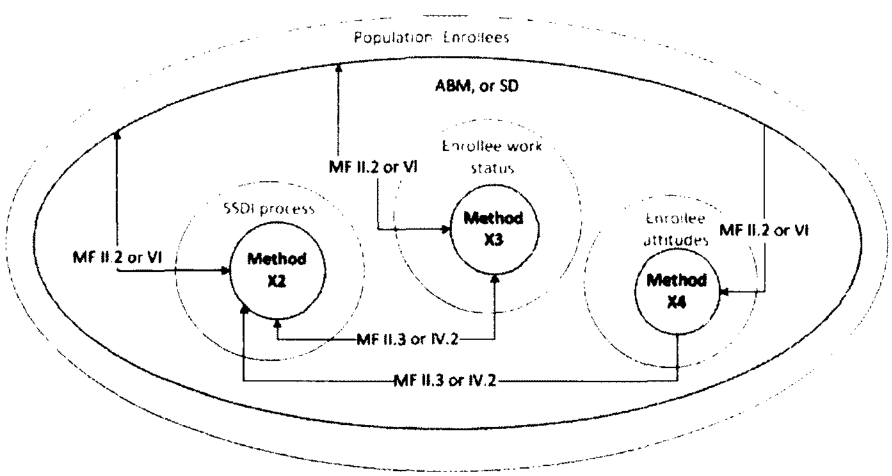

Figure 26. MFs with placeholders for methods.

5.4.5 Required Behavior and Choice of Methods

1. Analyze interaction points and identify required behavior

An SSDI awarded enrollee enters SSDI system, where they may consider the "working path". Their attitude to work will affect the decision to work. This interaction was specified in previous section as Component 4 influencing Component 2. This interaction point is critical because it could facilitate insight into investigation of hidden aspects related to transitions controlled by human behavior. This sets the requirement on Component 1, which means that enrollees must be able to make decisions by internally generated decision events.

Once enrollees decide to work, they will look for work, which will trigger their work status (Component 3). This interaction was specified in previous section at high level as Component 2 (SSDI process) influencing Component 3 (work status). This was described in Section 2.4.2.1 as triggering and listening to interaction type. Enrollees move to TWP stage if they find a job and make more than TWP income limit. This interaction was specified in previous section at a high level as Component 3 (work status) influencing Component 2 (SSDI process). Additionally, in order to represent this behavior the work status component must represent working and not working states, and SSDI component must represent transition that depends on state of Component 3 and the TWP income limit. This is again triggering and listening to interaction type. The enrollee stays in TWP stage of SSDI process until a total of 9 months has accumulated. This leads to the requirement that the Component 2 must represent a condition that counts the TWP months during which the enrollee made more than the TWP income limit. The amount of monthly income and its consistency determines length of stay in TWP.

After enrollees accumulate 9 months above the threshold, they enter the Extended Period of Eligibility (EPE) stage. During this time if the enrollee makes over $980 per month, which is considered SGA, the financial benefits are withheld for this month. This requires representation of condition within SSDI process that checks work status (SGA and NSGA) to determine if this month's money is withdrawn, and adds time spent within EPE stage. If 36-months have passed and the enrollee's work status is SGA, the financial benefits will be terminated for work. This, again, requires representation of the condition within SSDI process that checks status of work status (SGA and NSGA). There is also a need for a behavior that can represent amount of earnings, which may be represented using probability distribution function (PDF).

In the Section 5.4.3, requirements related to input were specified for Component 4. The incentives to work at psychological level scaled from sufficient to insufficient were of special interest.

Based on described requirements behavioral characteristics of individual components and their relations are identified as follows.

- Component 1: Enrollee is an individual, active decision maker.
- Component 2: SSDI process are to be represented as stochastic, capable to trigger and capable to be triggered. For instance, with transitions between SSDI stages are based on time stochastic functions, be responsive to external triggers, and are able to trigger work status (Component 3).
- Component 3: Work status is represented as able to trigger and response to an external change. For instance, it is able to be activated by other components. It represents enrollee working and not working states, and states related to working state i.e. SGA and NSGA.
- Component 4: It represents enrollee attitudes at psychological level to uncover hidden aspects related to decision to work.

2. Develop behavior falsifiers and evaluate methods for each component

Based on the above requirements, behavior falsifiers are developed. Although more than a single falsifier per component is permitted, each component has in this case only a single falsifier.

Component 1:

No method represents enrollee as an individual active decision maker.

Component 2:

No method represents SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered (listen to other conditions).

Component 3:

No method represents work status component with behavior that can trigger and is capable to be triggered (listen to other conditions).

Component 4:

No method represents enrollee attitudes at psychological level.

Table 23 shows the results of applying behavioral falsifiers to methods.

Table 23. Analysis of behavioral falsifiers.

Behavior falsifier/Method	DES	SD	ABM	BN	SC
No method represents enrollee as an individual active decision maker	x	0	1	x	x
No method represents SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered	1	0	1	0	1
No method represents work status component with behavior that can trigger and is capable to be triggered	1	0	1	0	1
No method represents enrollee attitudes at psychological level	1	1	1	1	1

The number one means that the answer to a falsifier is false, zero means true, and x means that this option was eliminated in the previous structural analysis. Component 1 should be developed using ABM because SD cannot represent individual active decision-making behavior logic. Although DES was already eliminated, this falsifier would have been also eliminated here because DES can create only individual passive entities (see Appendix A for definition). DES, ABM and SC can describe required by Component 2's behavior i.e. stochastic transitions related to systemic phases and transitions, and behavior that can trigger and is capable to be triggered. SD cannot describe SSDI stochastic process related to systemic phases and transitions because it is based on differential equations, which are inherently deterministic and would not be able to convey concepts of transition between phases but rather flows. Similarly, BN cannot describe SSDI stochastic process because it does not convey concepts of transition between phases but probabilistic relations between nodes. Component 3 has very similar requirements to Component 2 (except stochastic behavior requirement), and the same methods can be applied to Component 3. The behavior falsifier for Component 4 was not able to eliminate any of the methods because it is not precise enough to clearly associate characteristics of methods with behavior.

3. Employ falsification criteria context within falsifiers to resolve components with more than a single method

The criterion of scope was previously examined as the structural falsifiers in Section 4.5. Because no relevant simulation model was found in the literature (Section 4.3), and because reuse is not a concern of this implementation the criterion of reuse is not applicable and it is skipped. Similarly, the run speed mainly depends on sample size of population of enrollees and is not of concern for behavior falsifiers. The sample size will be adjusted to fulfill run speed tolerance. The falsification criteria evaluated will consist of resolution, accuracy, precision, available data, visual display, and development time.

Table 24 provides evaluation scores of falsifiers in the context of required resolution for considered methods. Because both falsifier one and two are very similar in this context they are discussed together. All methods are close contestants for falsifiers one and two. SC has characteristics that strongly falsify this sentence.

Table 24. Analysis of methods using falsifiers in the context of resolution.

Falsifier	No method represents the SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered with required resolution				No method represents the work status component with behavior that can trigger and is capable to be triggered with required resolution				No method represents the enrollee attitudes at psychological level with required resolution			
Method	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score
DES	1	1	1	1	1	1	1	1	1	1	0.25	0.25
SD	x	x	x	x	x	x	x	x	1	1	0.75	0.75
ABM	1	1	1	1	1	1	1	1	1	1	0.5	0.5
BN	x	x	x	x	x	x	x	x	1	1	1	1
SC	1	1	1	1	1	1	1	1	1	1	0.5	0.5

The ability to have a composite structure especially adds possibilities to represent levels of detail related to triggering behaviors that can produces higher resolution. ABM itself would require a lot of coding to represent required resolution, and which would likely end up as implemented in code state machine. DES would require workarounds to translate process view to represent considered behavior at sufficient resolution. On the other hand, all three methods could produce the required resolution. Although SC may be easier and faster to use, this is not a concern of this falsification criterion.

For the third falsifier, all methods seem possible but some seem less adequate for the purpose. For instance, it is difficult to conceptualize attitudes as discrete entities and events associated with DES processes. This is possible option but at this stage of research the resolution is too high to build a useful model. SC seems a little more applicable than DES since its holistic view may offer better match to desired resolution, but it is difficult to conceptualize the representation of multiple attitudes as states affected by multiple factors at the same time (network view). ABM could represent enrollee attitudes as individual competing agents. This is an interesting but very challenging approach, and it would produce too much detail for the purpose at this stage of the research. SD should be considered, but similarly to ABM it would require assumption of some parameters related to dynamic relations, which at this point would be too difficult to implement numerically. BN denies the third falsifier, because it is very intuitive and descriptive in the context of detail qualitative attitudes, which can be represented as a network of prior and conditional probabilities. At the current state of knowledge of return to work phenomena, this method has potential to produces desirable resolution and gain insight into the phenomenon.

Table 25 provides evaluation scores of falsifiers in the context of the required accuracy for considered methods. It seems that the accuracy context, the same as resolution, directs slightly more toward SC then ABM or DES, but this cannot be clearly determined. Although SC has the ability to describe triggering behavior as its key element, allowing for asynchronous transitions between states to produce high accuracy, and although ABM has to be programed and DES needs workarounds it could not be determined that these methods would make difference in the context of accuracy.

Table 25. Analysis of methods using falsifiers in the context of accuracy.

Falsifier	No method represents the SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered with required accuracy				No method represents the work status component with behavior that can trigger and is capable to be triggered with required accuracy				No method represents the enrollee attitudes at psychological level with required accuracy			
Method	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score
DES	1	1	1	1	1	1	1	1	1	0.5	0.25	0.125
SD	x	x	x	x	x	x	x	x	1	0.5	0.75	0.375
ABM	1	1	1	1	1	1	1	1	1	0.5	0.5	0.25
BN	x	x	x	x	x	x	x	x	1	0.5	1	0.5
SC	1	1	1	1	1	1	1	1	1	0.5	0.5	0.25

The lower score of importance for the third falsifier reflects the direction given during the analysis of requirements in Section 4.4.2 that accuracy is less important than resolution. Enrollee attitudes should be first represented well as a theoretical concept before engaging with a more detail view, and DES is not a good candidate for theorizing the multi-factor concept with its "black box" viewpoint. Accuracy of DES would be more questionable at this point of research because of lack of top-level concept of phenomenal and related data. A similar problem may pertain to ABM, although this approach can be used to theorize at the micro and macro levels. This could be enabled in the future to provide a detail specification of entities and decision processes, but not at this stage of knowledge. SC may have more potential to represent more accurately attitudes at aggregated level, but its discrete event character requires specifying a detailed view of conditions leading to transitions, which is problematic without sufficient knowledge. SD and BN have the most potential for "accurate" results; hence they better falsify the falsifier. This can be viewed as alignment with their close to theoretical use characteristics, which also matches current state of knowledge about return to work phenomenon. Because enrollee attitudes are not explored, the view that is less granular can also be more accurate in this case. Because SD requires more data to implement its dynamic view, this can also influence accuracy, whereas BN intuitively can capture concept without need to specify detailed dynamic relationships.

Table 26 provides evaluation scores of falsifiers in the context of required precision for considered methods. Because all three methods are considered for the same simulation software, they are equivalent when considering precision for falsifier one and two in relation to simulation time unit.

Table 26. Analysis of methods using falsifiers in the context of precision.

Falsifier	No method represents the SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered with required precision				No method represents the work status component with behavior that can trigger and is capable to be triggered with required precision				No method represents the enrollee attitudes at psychological level with required precision			
Method	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score
DES	1	1	1	1	1	1	1	1	1	0.25	0.25	0.063
SD	x	x	x	x	x	x	x	x	1	0.25	0.75	0.188
ABM	1	1	1	1	1	1	1	1	1	0.25	0.5	0.125
BN	x	x	x	x	x	x	x	x	1	0.25	1	0.25
SC	1	1	1	1	1	1	1	1	1	0.25	0.5	0.125

Similarly, all three methods can have the same precision in relation to estimation of stochastic output. The precision of simulation time unit can be important for statistical accumulators, but not to internal events, which will be asynchronous for all three methods when developed as internal method within an agent. The easier modeling of asynchronous events using SC could have indirect impact on precision but this cannot be proved.

The third falsifier has less emphasis on precision, because the knowledge about the underlying phenomenon is not sufficient to demand higher precision. High precision of enrollee attitudes representation is not realistic at this stage of research. Because of the long simulation period, obtaining insight into possible enrollee attitudes in relation to return to work phenomenon does not focus on high precision. For this reason, a method that allows representation of attitudes that would not require precise implementation of dynamics seems better adequated. BN allows representation of the attitudes along the probability scale, which can simplify initial conceptualization of qualitative factors. SD could provide future enhancements of BN, but at this point of research, it requires too detailed specification in relation to time, while not even a structural view exists.

Table 27 provides evaluation scores of falsifiers in the context of available data for considered methods.

Table 27. Analysis of methods using falsifiers in the context of available data.

Falsifier	No method represents the SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered with available data				No method represents the work status component with behavior that can trigger and is capable to be triggered with available data				No method represents the enrollee attitudes at psychological level with available data			
Method	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score
DES	1	0.75	1	0.75	1	0.75	1	0.75	1	1	0	0
SD	x	x	x	x	x	x	x	x	1	1	0.5	0.5
ABM	1	0.75	1	0.75	1	0.75	1	0.75	1	1	0.25	0.25
BN	x	x	x	x	x	x	x	x	1	1	0.75	0.75
SC	1	0.75	1	0.75	1	0.75	1	0.75	1	1	0	0

Because the systemic view that pertains to the SSDI system is clearly defined and available, this, although important, does not affect greatly methods for the first falsifier as all three could produce systemic SSDI view. The second falsifier pertains to enrollee work status, which although can be clearly defined is driven by enrollee earnings data, which are thus far not found. Because of lack of enrollee earnings data, they must become a calibration input factor. This provides insight into model development, but does not change evaluation of falsifier two in relation to three methods considered. The third falsifier is the most difficult to estimate, because no data at this level were found and there is not clear structure of model at this time. Because BN facilitates qualitative view this enables more flexibility to represent qualitative phenomenon with limited amount of data, and aid to approximate data as a proof of concept.

Table 28 provides evaluation scores of falsifiers in the context of required visual display for considered methods.

Table 28. Analysis of methods using falsifiers in the context of visual display.

Falsifier	No method represents the SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered with required visual display				No method represents the work status component with behavior that can trigger and is capable to be triggered with required visual display				No method represents the enrollee attitudes at psychological level with required visual display			
Method	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score
DES	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	0.75	0.375
SD	x	x	x	x	x	x	x	x	1	0.5	0.5	0.25
ABM	1	0.5	0.5	0.25	1	0.5	0.5	0.25	1	0.5	0.25	0.125
BN	x	x	x	x	x	x	x	x	1	0.5	0.25	0.125
SC	1	0.5	1	0.5	1	0.5	1	0.5	1	0.5	0.75	0.375

In AnyLogic®, methods could be associated with additionally developed graphical representation. Because of this, the evaluation is focused on the prebuilt features of modeling elements associated with considered methods. DES and SC deny falsifiers one and two because their visual representation of building blocks provides easy to follow interactive (DES) or animated (SC) graphical view into their dynamic transitions. Because of its characteristic, ABM would require coding to achieve the same visual effects, hence its lower score. Attitudes as the BN model can be quickly and intuitively developed using selected Genie ® tool and imported to AnyLogic®, but this approach does not have dynamic visual representation prebuilt within AnyLogic® and requires its implementation. All other methods except ABM have visual display blocks that would allow monitoring dynamic changes related to the components representing attitudes.

Table 29 provides evaluation scores of falsifiers in the context of required development time for considered methods.

Table 29. Analysis of methods using falsifiers in the context of development time.

Falsifier	No method represents the SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered with required development time				No method represents the work status component with behavior that can trigger and is capable to be triggered with required development time				No method represents the enrollee attitudes at psychological level with required development time			
Method	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score	Incl.	Impo.	Eval.	Score
DES	1	1	0.25	0.5	1	1	0.5	0.5	1	1	0.25	0.25
SD	x	x	x	x	x	x	x	x	1	1	0.25	0.25
ABM	1	1	0	0	1	1	0	0	1	1	0	0
BN	x	x	x	x	x	x	x	x	1	1	1	1
SC	1	1	1	1	1	1	1	1	1	1	0.25	0.25

This falsification criterion has significant importance because of time constraints related to this study. For falsifiers one and two, SC is the most time efficient method in the context of implementation, DES requires some workarounds, while ABM would require significant time to code required behaviors. Using BN for the development of attitudes seems the most time effective, because the development of network is the most time consuming yet very quick and intuitive using Genie ®. On the other hand, detailed dynamic views offered by all other methods in context of representing enrollee attitudes seem to be more time consuming, and are very problematic at this point, which negatively affect development time.

Table 30 provides methods chosen and CoMSs for considered methods derived using falsifiers in the context of falsification criteria using proposed in Section 5.1.5 equations.

Table 30. Choice of methods and analysis of CoMS for considered methods and falsifiers in the context of falsification criteria.

Falsifier	Falsifier for SSDI process	Falsifier for work status	Falsifier for enrollee attitude	Falsifiability score	Relative to max	CoMS
DES	4.5	4.5	1.06	10.06	0.68	**0.28**
SD	x	x	2.31	2.31	0.16	**0.80**
ABM	4	4	1.25	9.25	0.63	**0.33**
BN	x	x	3.63	3.63	0.25	**0.71**
SC	5.25	5.25	1.50	12.00	0.81	**0.14**
Highest score	**5.25**	**5.25**	**3.63**	**14.13**	**0.96**	**0.04**
Method choice	**SC**	**SC**	**BN**	**SC and BN**		

Based on cumulative score from all falsifiers, SC is selected for SSDI and work status components, while BN is selected for representation of attitudes of enrollees. When both SC and BN methods are selected instead of only SC or only BN, SC are complemented by BN with the CoMS of 0.14, while BN is complemented by SC with the score 0.71. CoMS for combined SC and BN versus DES or ABM is 0.28 and 0.33 respectively. Table 31 provides CoMSs for all considered methods and both structural and behavior falsifiers.

Table 31. Choice of methods and analysis of CoMSs for all considered methods and all falsifiers (both structural and behavioral).

All Falsifiers/Method	DES	SD	ABM	BN	SC	Max	Highest scored method(s)
No method represents a population of individual enrollees as a composition of individual enrollees	x	1.00	1.00	x	x	1.00	**SD/ABM**
No method represents individual entities as a composition of SSDI process, enrollee work status, and enrollee attitudes	x	1.00	1.00	x	x	1.00	**SD/ABM**
No method represents enrollee as an individual active decision maker	x	x	1.00	x	x	1.00	**ABM**
No method represents SSDI process component with stochastic transitions and behavior that can trigger and is capable to be triggered	0.86	x	0.76	x	1.00	1.00	**SC**
No method represents work status component with behavior that can trigger and is capable to be triggered	0.86	x	0.76	x	1.00	1.00	**SC**
No method represents enrollee attitudes at psychological level	0.25	0.54	0.29	0.85	0.35	0.85	**BN**
Sub-falsifiability score	1.96	2.54	4.82	0.85	2.35	5.85	na
Score relative to max score	0.34	0.43	0.82	0.15	0.40	0.98	na
CoMS	0.64	0.55	0.16	0.83	0.58	na	na

The order of reasoning is kept in agreement with the developments i.e. first structural falsifiers, then behavioral falsifiers, and finally scores for behavioral falsifiers in the context of falsification criteria. The symbol x indicates that the method was already eliminated. ABM has the overall highest sub-falsifiability score, and its CoMS when SC and BN are added is 0.16. From the perspective of SC the CoMS is 0.58, and finally from the perspective of BN CoMS is 0.83. Figure 27 displays complete MFs. The analysis of falsifiers led to selecting methods for all four components. Component 1 will be developed using ABM, Components 2 and 3 using SC and Component 4 using BN.

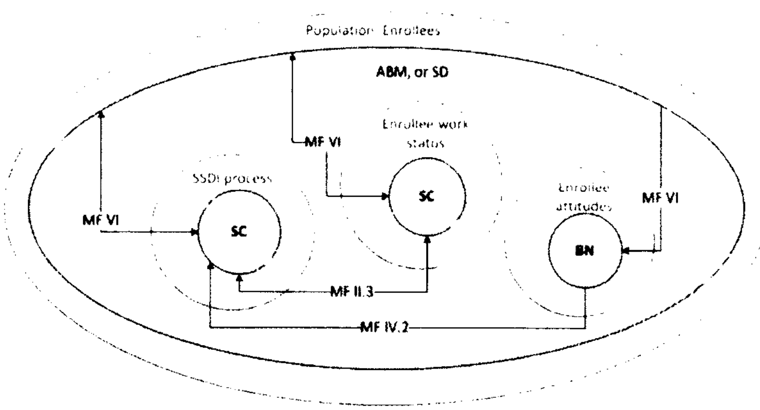

Figure 27. Completed MFs.

5.4.6 Low-level Conceptual Model(s) Development

ABM is used to develop Component 1, which provides an overarching structure for the agent's behavior (Components 2, 3, and 4). Components 2 and 3 are developed first using two inter-reliant SCs. Finally, Component 4 will be developed using BN.

Figure 28 presents an SSDI enrollee (agent) passing through different stages related to the return-to-work phenomenon. *SSDI behavior* and *Work Status* are two main parts (state charts) visible within the enclosed blue dashed lines. The goal is to represent

SSDI systemic conditions related to the return-to-work phenomenon and corresponding behavior of enrollees.

An SSDI awarded enrollee enters the initial composite state: *SSDI awardee,* where he or she may consider working, represented as *Transition 1* from *Awarded* state to *Decided To Work* state. This transition triggers *Condition 1* for moving the enrollee from *awarded* state to *Look For Job* state, both located within composite state *Not Working.*

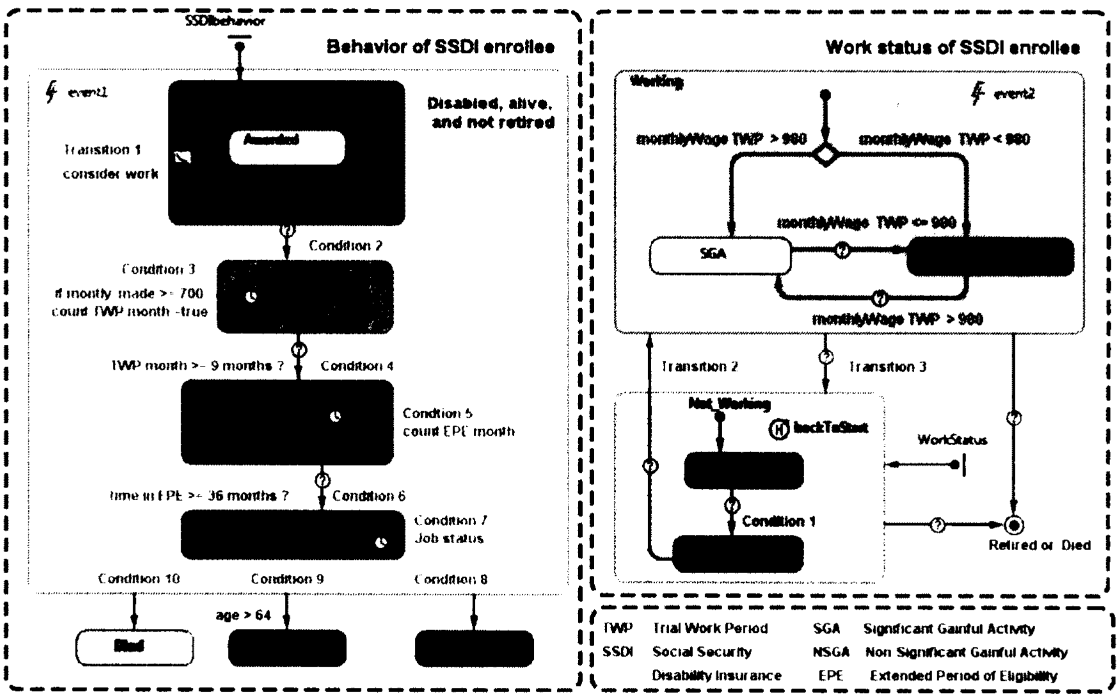

Figure 28. SSDI process and work status.

If the enrollee finds a job (*Transition 2),* he or she enters the composite *Working* state significant gainful activity (*SGA*) or not significant gainful activity (*NSGA*) at the same time triggering *Condition 2.* If the enrollee makes more than TWP income limit, he or she moves to the *TWP Start* state. This counts as the first month of TWP [163]. During this time, the enrollee can make as much as he or she wants-without financial reductions of

SSDI payments [155]. Internal *Condition 3* counts the TWP months in which the enrollee made more than the TWP income limit. The enrollee stays in *TWP Start* state until a total of 9 months has accumulated. The amount of monthly income and its consistency determines length of stay in *TWP Start* state. Upon accumulating 9 months (*Condition 4),* the enrollee enters *Extended Period of Eligibility* (EPE) state, and the internal *Condition 5* starts adding time. During this time if the enrollee makes more than $980 per month, which is considered SGA for non-blinded enrollee [163], the financial benefits are withheld for this month. This is shown as transitions between *SGA* and *NSGA* states within *Working* state based on monthly income of enrollee. After the 36-month mark is reached, the enrollee's financial benefits will be terminated for work after the first *SGA* month (*Condition 6* - completion of EPE and *SGA* state). Otherwise, the enrollee stays in EPE state indefinitely. If the enrollee is terminated for work, *Condition 7* checks the enrollee's job status and if other than *SGA* (e.g. *NSGA*, *Look for job*), the benefits will be reinstated. In addition, the enrollee enters *Medical Reason* state if SSDI is terminated for medical reasons (*Condition 8)*, enters *Retired* state after becoming 64 years old (*Condition 9*), or enters *Died* state (*Condition 10*) when deceased. The C*onditions* 1 to 10 represent the system process. Hall et al. [162] reported that enrollees are being discouraged from working by medical professionals and federal disability policies. Transitions 1 to 3 define enrollee behavior and job related factors. Prediction of the system behavior and subsequent experimentation of alternative solutions (interventions and/or programs) could only be accomplished based on the gained understanding of transitions as prerequisite.

Both medical condition and internal attitude about the system are determinants for a decision to work. In the system, the monthly income of enrollee determines SGA or NSGA state, which in turn determines transitions to TWP, EPE, and termination of benefits because of work. Understanding the relationship between enrollees' attitude about level of income that is sufficient to encourage working behavior, and minimizing adoption of the patient role can provide insight into a possible design of return-to-work programs. The systemic conditions presented using SCs above provide a high-level view that needs to be expanded to uncover hidden aspects related to transitions controlled by human behavior.

Please refer to Figure 29 during the discussion that explains the concept of the Component 4.

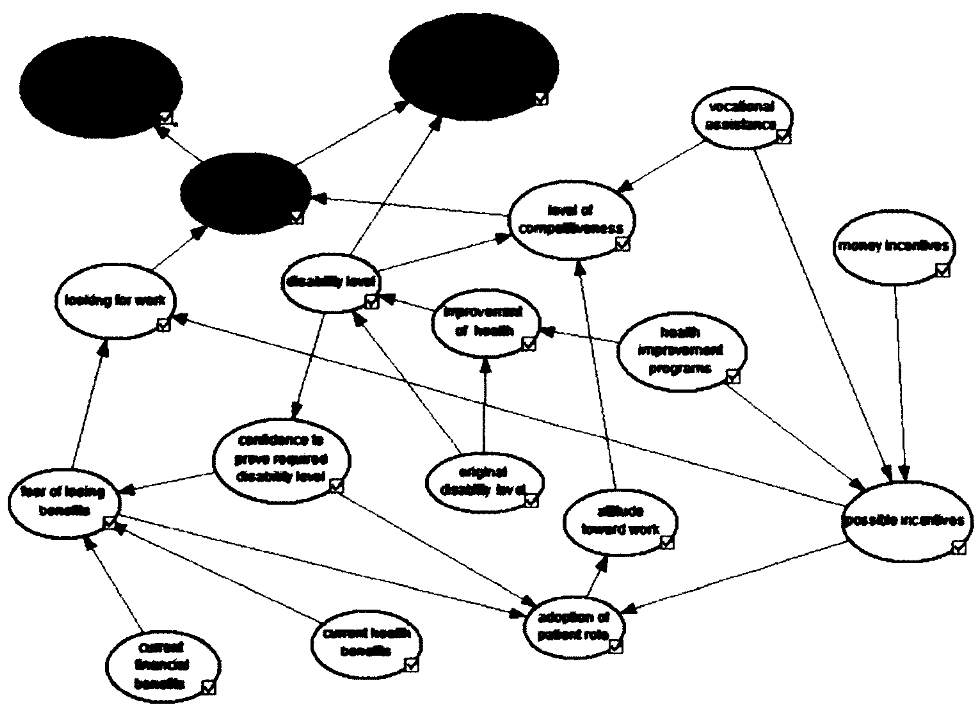

Figure 29. Factors affecting enrollee-working behavior.

Genie ® software was used to develop a conceptual model as a BN. The factors relations are derived based on work of Hall et al. [162], Kennedy et al. [152] and Thomas and Ellis [161]. This BN model aims at capturing financial and health factors on the probability of an enrollee to consider work, and subsequently be able to work. *Possible incentives* can encourage working behavior, but *Fear of losing current benefits* create a detrimental effect to *looking for work* factor, especially because of a high effort to obtain benefits in the first place [152]. The scope and availability of *current health benefits*, significance of *current financial benefits*, and *confidence to prove disability* (which at least in principle, are proportional to actual *disability level*) influence *fear of losing benefits*. This fear may be offset by providing *possible incentive,* for instance *vocational assistance* (such as vocational rehabilitation, personal assistance, and adaptive technologies and transportation), *money incentives,* and *health improvement programs* at a sufficient level as seen by the enrollee (sufficient, insufficient, or a degree sufficiency).

Adoption of patient role can change an enrollee's *attitude toward work* and influences *level of competitiveness* [161]. *Level of competitiveness* is also influenced by *vocational assistance* and *disability level*, while *disability level* may be improved by *health improvement programs.* Both, the *level of competitiveness* and *looking for work* factors influence *probability of enrollee to work*, which in turn influences medical and financial independence of enrollee. C*urrent financial benefits* and *current health benefits* can work as a mental inhibitor to work, while incentives can offset this attitude.

Assuming that, for some enrollees, their health condition permits work, the question is how to establish the system of benefits and incentives to prevent *adoption of* *patient role*, which can prohibit enrollees to better their lives through work and subsequent financial and medical independence. Improvement of their health through *health improvement programs* can improve *level of competitiveness*, at the same time decreasing *confidence to prove disability,* hence increasing *fear of losing benefits.* This can be especially true when *current health benefits* are already provided (2 year waiting period).

The selected methods resulted in development of a conceptual model. As seen by the modeler, a combination of ABM, SC, and BN provided a satisfactory choice to describe both systemic transitions of SSDI process and theoretical relations related to individual enrollee's decision to work.

5.4.7 Model Development

A multi-method simulation model that comprises ABM, SC, and BN methods was developed using AnyLogic® modeling software. It implements details of system phases and enrollee behavioral factors developed during conceptual modeling as an agent behavior. The upper-level, which consists of Component 1, is developed as an ABM. It is used mainly as a shell for other components to collect statistics about internal states of agents providing information relevant to the RQ e.g. numerical counters and graphs depicting cumulative view of SSDI process for the sample population of enrollees. A custom distribution of the population ages 18 to 39 was created based on the Annual Statistical Report by the Social Security Disability Insurance Program [155] for the 1996 population (beginning of enrolment). State charts developed during conceptual modeling were used as a blueprint for a sample population of SSDI awardees. Software adequacy testing (see Section 2.4.3.2) was conducted to verify model translation from a low-level conceptual to the computational form; here the methods used and their interaction points were important considerations. The model was calibrated using historical data of return to work phases for the population ages 18 to 39 [164] (see Figure 30).

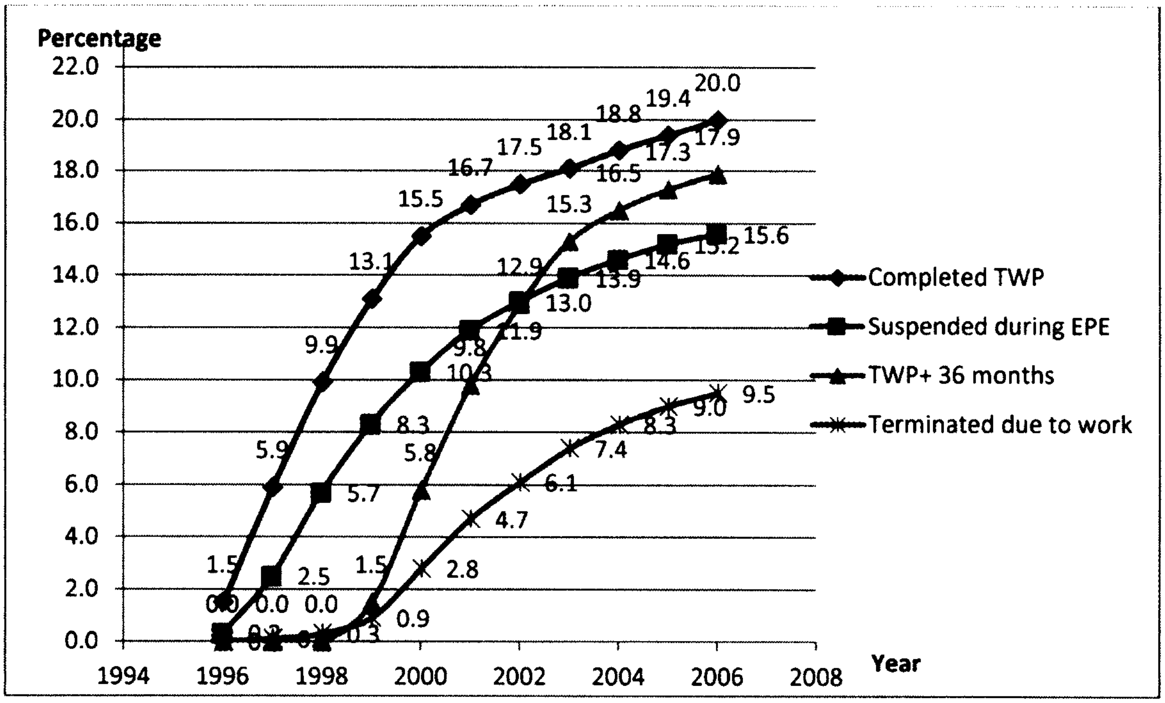

Figure 30. Cumulative longitudinal work incentive statistics for 1996 awardees age group 19-39, 1996–2006.

The values of conditional probability tables (CPT) were estimated and are available in Appendix D. Empirical derivations of CPTs based on qualitative interview data are desirable, but this was not possible for this study. Useful guidelines and examples for using interviews to build CPT values can be found in literature [148, 165, 166].

In the first phase, *Transition 1* was represented as a rate. The rate is determined by a scaled *looking for work* factor of BN. The time-series probability tables of *vocational assistance*, *money incentives,* and *health improvement programs* factors within BN are zero in the base case scenario and will serve as an input variable during experimentation. Six CPT values (other six are equal to 1- probability of the first one) of *looking for work* factor representing a person decision to work are derived through a calibration experiment that minimized error between enrolees that entered *TWP Start* state and historical data for the completed TWP phase [164]. The meaning of the calibration curves shown in the bottom-right part of Figure 31 is as follows: "standard" is the historical data, "current" is the last run, and "best feasible" means the curve with the overall lowest error. The top left of Figure 31 displays an extract of CPT for the *looking for work* factor, while symbols HH, HM, HL… are its parameters used during calibration. For instance, HH is its conditional probability value that person is seeking job given that both the *fear of losing benefits* and the *possible incentives* are high (see CPT table). Support arrays (1) store values of input and output variables for each agent's BN, which connect (2) to the experimentation framework through a set of parameters (3 and 4) as depicted in Figure 31. *Transitions 2* and *3,* and internal transitions between *SGA* and *NSGA* states within *Working* state depends on the amount of money made by an enrollee. The money made varies with the amount, and with the frequency of changes represented as a dynamic event setting different probability distribution functions (PDF) for each enrollee's phase. The monthly amounts of money made are represented using Beta PDFs, which were also derived through calibrating the experiment against historical data for percentage of enrollees that completed EPE phase [164]. Different PDFs were tested (e.g. uniform, triangular, truncated normal, exponential, and beta). None of them was an exact fit, but the closest matches were obtained using Beta PDFs (see Figure 32). Two Beta

PDFs were used within the EPE phase. The triggering point to switch between first and second Beta PDFs in this phase was developed in an effort to represent approaching the end of EPE phase and the possibility of losing benefits. Further experimentations were conducted to alter Beta PDF and to create custom distributions to minimize calibration error.

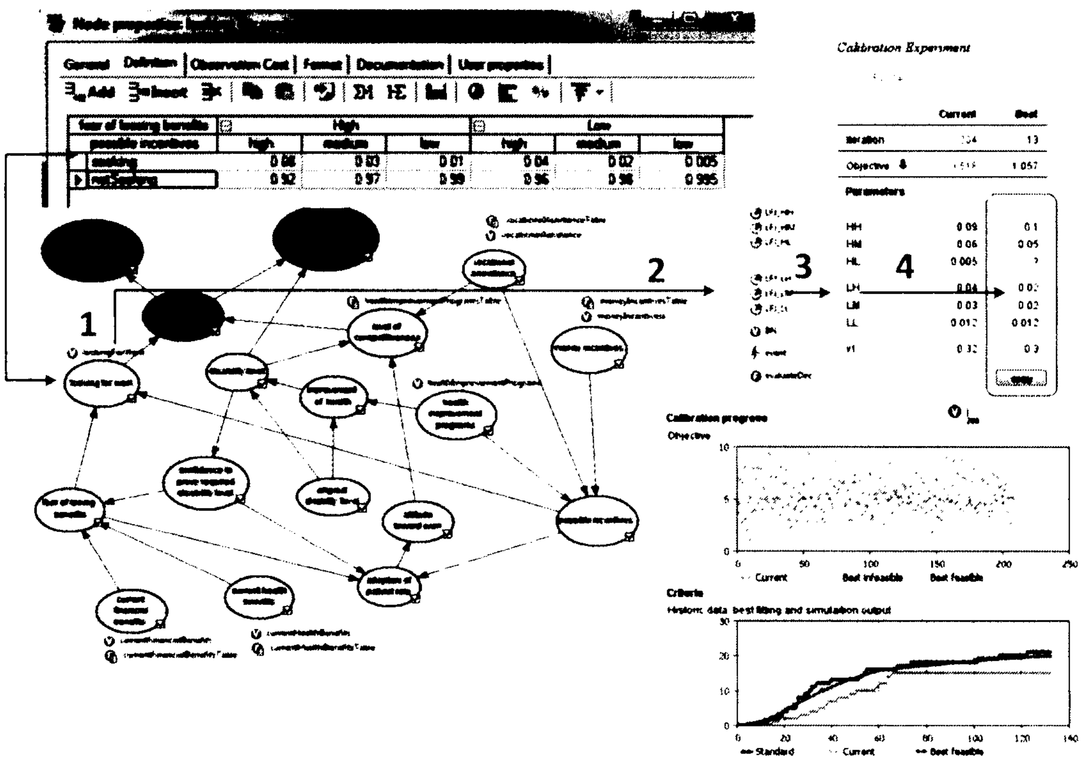

Figure 31. Calibration of CPT *of looking for work* factor.

Only one of the custom PDFs (for *Decide To Work* and *TWP Start* states) was finally used because it had a smaller error as compared to Beta PDF. Ideally, all transitions should be calibrated at once, but this was not possible only because seven parameters maximum can be optimized using the educational version of AnyLogic®. The professional or "university researcher" versions of AnyLogic® does not have these limitations.

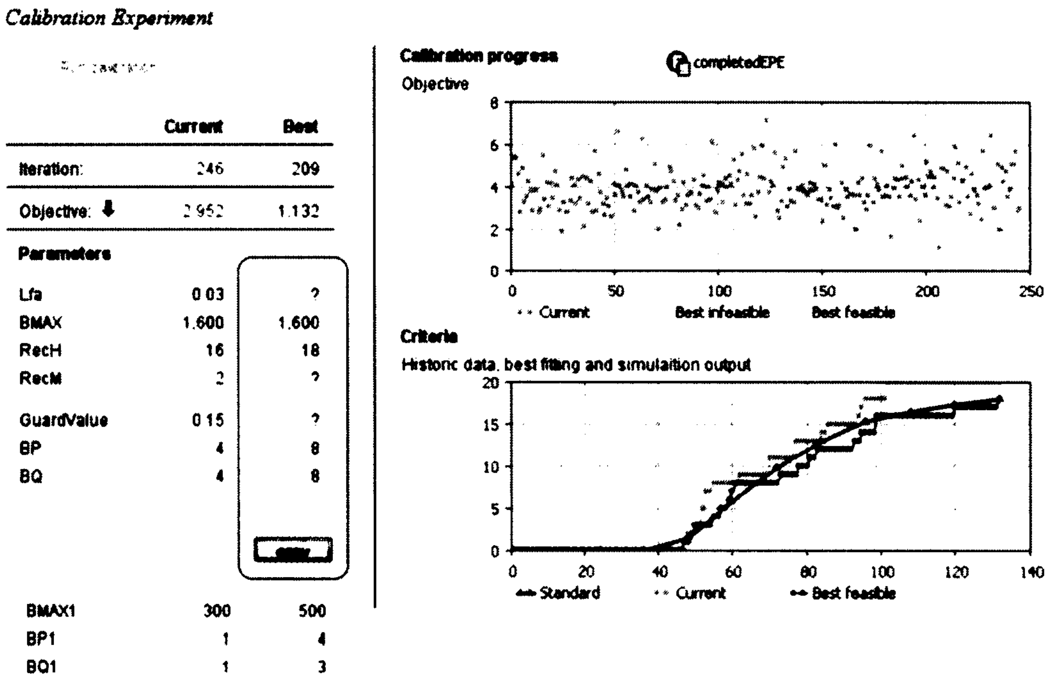

Figure 32. Calibration of Beta PDF for transition to EPE phase.

5.4.8 Model Validation

A validation process aims at the determination of feasibility of the developed model to conduct necessary experimentations to answer the sample research question. It was a challenge to capture phenomenon related to termination of benefits using aggregated PDFs representing money made by enrollees during period of *Extended Period of Eligibility* due to high variability (see

Figure 33). This is most likely because of enrollees' awareness of possible imminent termination of benefits (highly variable human behavior). A 2D histogram was built based on 200 runs. It displays variability levels for a percentage of benefits terminated. It ranges from 0 to 132 on the x axis and 0 to 25 on the y axis, with 132 and 125 intervals, respectively. According to the graph, rarely, percentage of benefits terminated could reach 20 percent at the high end and 3 at the low end (less than 10).

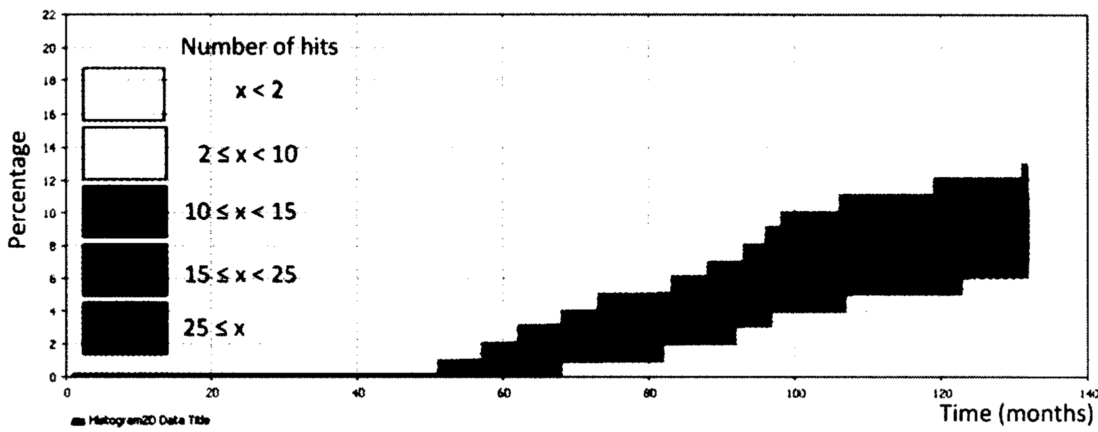

Figure 33. 2D histogram showing variability levels for percentage of benefits terminated based on 200 runs.

More likely outcomes (darker color) can range between 6 and 12 percent. The end of EPE phase, similarly to initial decision to start TWP, should be represented in the future in more detail, for instance similarly as *Transition 1* using a BN. The simulation model output was compared to the real historical data [164]. Figure 34 shows percentages for four phases of a sample population of 3000 enrollees within the return-to-work process, generated by the simulation model (purple), in comparison to the historical data (green). The visual inspection indicates correct trend lines of the model. Additional calibrations and refinement related to the EPE phase could improve this model further. A 200-simulation run experiment with 100 enrollees per each run was used for validation. Ontological adequacy testing (OAT) is conducted by comparison of empirical data versus simulation model output (see Section 2.4.3.2). A sample mean of benefits terminated was used for statistical validation of output using the test statistic value $z = \frac{\bar{x} - \mu_0}{\delta/\sqrt{n}}$, where $\bar{x} = 9.36$ is a sample mean of percent of benefits terminated; $\mu_0 = 9.5$ is the historical value of benefits terminated; $\delta = 3.1$ is the sample standard deviation, and $n = 200$ is the sample size.

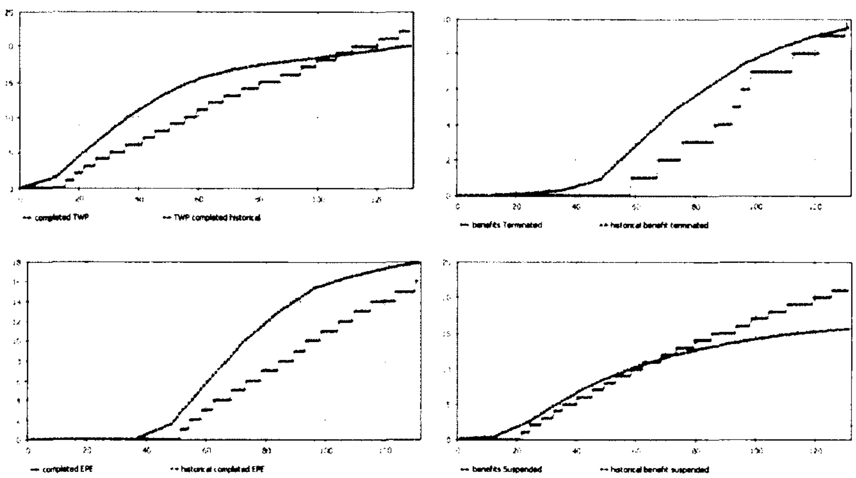

Figure 34. The output from the simulation vs the historical data for return to work phases.

The test conducted is based on a two-tailed $z_{0.025} = 1.96$ at a significance level of 0.05. The resulting $z = 0.62 \leq 1.96$ so the model cannot be proven to produce results different from statistical and historic data. Finally, it is pointed out that except for *looking for work* factor values of the rest of CPT were not fully calibrated nor derived based on interviews or surveys, hence results should be considered as a proof of concept and not real values.

Analytical adequacy testing (AAT) (see Section 2.4.3.2) was used to assess theory-versus-conceptual-model validation and validation of theory with a descriptive simulation model. The BN and SC mix allowed to map theoretical conceptual model (based on relevant literature) related to enrollee's attitudes toward return to work at desirable resolution. BN probabilistic value was transformed into time event generated by a time based (rate) transition within SC. Analytical adequacy could be significantly improved if CPTs were derived based on qualitative data gathering e.g. interviews to better grasp variability related to indigenous factors. This problem can also be viewed as theoretical filter (path 5 on Figure 7), which indicates subjectivity related to lack of empirical data of enrollee attitudes. Moreover, transformation interaction point should have been more thoroughly investigated because changes to CPTs affect transformation pattern. Because no empirical CPT values were collected here, calibration was conducted based on a single CPTs set. Validity within a range of the transformations of the interaction point should be established if real qualitative data were available by using sensitivity analysis of CPT values versus accuracy of generated output from SC transition. This would be very importation to increase analytical adequacy as a representation of return to work theory. The model could also be extended in relation to money made by enrollees. Probability to find a job and money made by enrollee are influenced by a specific job market. The relevant factors and appropriate data could enhance current model and have positive impact on its validity.

5.4.9 Experiments and Analysis

A developed simulation model will be used to conduct experiments to create insight into the research question. In order to answer the research question, enrollee attitude about three types of incentives: *vocational assistance, money incentives,* and *health improvement programs* was varied. The base case was compared to four scenarios, in which the effects of enrollees' increased attitude toward incentives was assessed as the difference in the total percentage that remained "on the rolls" at the end of year 2006. The first three scenarios used a single incentive, while the last one used all three incentives combined. Prescribed yearly "levels of incentives" are represented as time series prior probabilities of sufficient incentives as seen by enrollee (see Figure 35) and are set with the same values for all incentives to enable their comparison. There is a single output measure captured: percentage of population with benefits terminated for work.

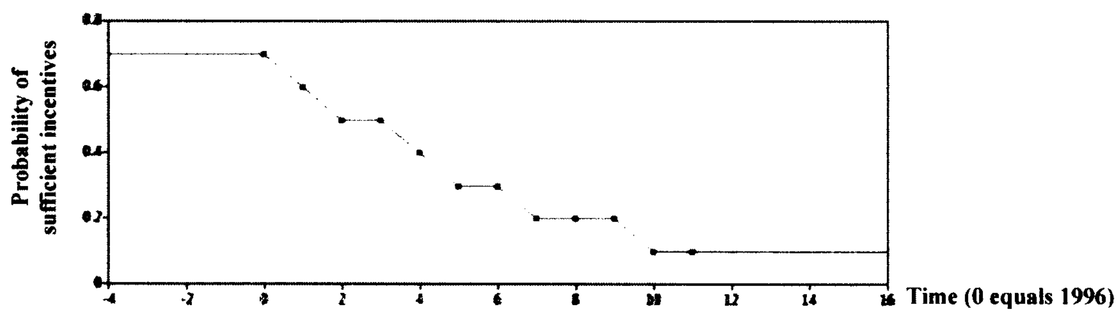

Figure 35. A sample input probabilities for incentives used for analysis.

Figure 36 shows the results the sample scenario. All incentives were statistically significant at 95% as compared to the base simulation. The difference with the base simulation ranges for *vocational assistance* between 6.80 and 8.28, for *money incentives* between 6.52 and 7.86, and for *health improvement* between 7.32 and 8.74. All three incentives resulted in similar values with only *money incentives* and *health improvement* statistically different (1.60; 0.08), but as mentioned in the validation section, these results should be considered with caution. When all incentives were combined, this resulted in a difference between 16.90 and 18.44 as compared to the base scenario.

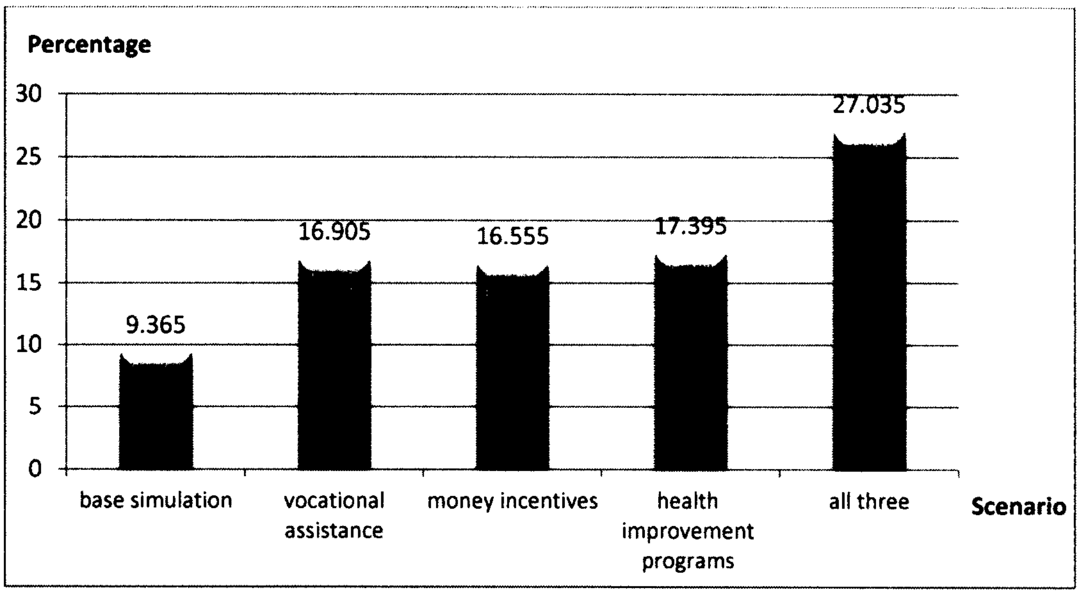

Figure 36. Percentage of population with benefits terminated based on incentives used.

According to the simulation output, prescribed levels of incentives significantly increased percentage of population with benefits terminated for work as compared to the base case with no incentives. The relationship with the cost for this effect was not considered and should be included in the future research. According to Kregel [159], annual savings from the WIPA program accounted for about 20 percent of the program cost itself, which although it seems modest, can accrue over time. Better understanding of costs related to incentives could provide improved view on financial tradeoffs for decisions related to which programs and incentives are implemented.

5.5 EVALUATION

This section employs the evaluation plan proposed in Section 5.3.

5.5.1 Questions and Statements

The evaluation is conducted using questions and evaluation statements proposed in Section 5.3.

Question 1: Did the research guidelines facilitate generation of information for external method falsifiability?

The information provided during origination of study step can be helpful in clarifying subjective human dimension necessary to better understanding possible problems related to scope of methods considered in subsequent developments. For instance, the first step disclosed subjective aspects of methods considered related to software used and methods used. The implementation of methods within software can have differences, which will affect perception about what the method is and what it can do. A choice of AnyLogic® and Genie® as considered modeling environments without consideration of other software like Arena®, Simio, Repast, and NetLogo, Extend®, Netica™ , BayesiaLab or even general programming platforms like Visual Studio® or Eclipse® is in large a subjective aspect of the process.

The first step also disclosed all preselected methods. Clearly, other modeling methods such as PN and FM could have been considered but were not, which is subjective. Some software offers low-level conceptual methods that can be implemented into computable form, while other software may only offer conceptual view e.g. software that can support UML modeling like MS Visio, Virtual Paradigm etc. High-level conceptual methods, e.g., causal loop diagram, or use case diagram must be extended toward low-level concepts forms during steps 4 and 5 of the guidelines. The guidelines do not constrain or specify a set of high-level conceptual methods as a prerequisite to low-level method selection. This allows for freedom of conceptualization at a high-level. For instance, Figure 25 depicts a high-level method capturing structural dependencies.

The information about methods and software disclosed the boundary of this particular multi-method M&S approach, which could be used for external critique by M&S community. This has the potential to improve multi-method M&S practice. The set of multiple methods considered for multi-method conceptualization is indicated up front, which provides a new perspective on multiplicity of methods considered for conceptualization versus the actual set of methods used to develop a simulation model.

The developed RQ provides information that can be helpful in understanding subjectivity related to the author's perception about what is important as a research topic. This directly affects methods that may have better chances to be selected. Many other RQs related to this system and phenomena e.g. *how enrollee characteristics such as age, health, and profession influence return to work*, or *which strategies have more promising outlook on return to work programs* can be raised. Clearly, the choice of RQ can affect methods used.

The identified inputs and outputs needed to get insight into the RQ define subjective requirements as seen by a researcher and provide insight helpful in clarifying the subjective human dimension related to methods chosen in the subsequent steps. The disclosed information related to specified requirements can be used as a base for a critique. Specified requirements can influence method choice. Although some of them are chosen based on the literature, others reflected the desire of the author to direct the research deeper into a psychological level of SSDI enrollees. For instance, the focus on understanding of internal drivers of enrollee at psychological level within return to work phenomenon can be argued as infeasible, yet it carries weight toward methods that can facilitate particular modeling view.

During the structural analysis step, the insight is related to subjectivity of specifying structure of a simulation model. The methods chosen are based on structural falsifiers, hence objectivity of each statement can be traced to principles of falsification, which discloses degree of falsifiability (universality) as information that can facilitate insight for external critique. On the other hand, the structural relations (see Table 21) can be subjective if dependencies between social phenomena are considered at the structural level in more detail (because they are unknown, or at best, fuzzy). Structural falsifiers could also differ depending on the software used. For instance, if simulation software adheres to a non-object oriented modeling concepts structural falsifiers may not be visible and even relevant. Moreover, as a particular critique it could be argued that implementation of SD as an array of stock and flow diagrams would depend on software used or even a version of the same software (e.g. commercial versus free). The use of an array view in SD has become more popular, but may or may not be considered as a part of the SD specification. The array view extension was taken into consideration in this case, but could have been skipped with the given above argumentation.

The analysis of behavior provides a final step during which methods were selected. At the same time, this produced insight into subjectivity during specifying behavior of a simulation model. The methods chosen were based on behavioral falsifiers, hence objectivity of each statement can be traced to principles of falsification (precision), which discloses degree of falsifiability as an information that can facilitate insight for external method critique. On the other hand, derivation of behavioral falsifiers is subjective to a modeler's view about what it takes to build a model that has potential to generate an insight into the RQ.

The evaluation scores of falsifiers in the context of falsification criteria, although supported by reasoning, consist of a subjective view of a modeler on desirable characteristics of a simulation model. The CoMS calculated based on falsifiers in the context of falsification criteria facilitated the ability to select methods, but the objectivity of the reasoning can be questioned. Moreover, finding clear boundaries for characteristics of methods can be difficult. For instance, it could be argued that SD can represent active events using dynamic variables i.e. external logic could be attached to dynamic variable via functions, which could facilitate active behavior. This characteristic is not how SD is specified but it is related to SD implementation within AnyLogic®. Moreover, inability to differentiate score for ABM, DES, and SC methods for Components two and three in the context of accuracy, precision and resolution should be pointed out. The clearest differentiator was a development time as falsification criterion, which is related to the necessary coding in ABM and workarounds in DES. It is pointed that development time is a pragmatic falsification criterion, and it was not identified in Section 3.2 as a core criterion. Within ABM, one can implement a state machine pattern. In DES, workarounds can make process behave like a state machine as well. Since those developments or changes ultimately lead to mimicking SC characteristics the question one should ask is whether they are still considered original methods (ABM and DES), or a SC version built or transformed with other methods.

The scoring process specified that each component could be developed with a single method, which may not always be the case. For instance, with more than a single falsifier describing the behavior of a component, falsifiers can point toward the use of more methods, each for a separate falsifier that scored higher. Logically in such a case, component could be disaggregated into separate components with their unique characteristics.

The main critique of a developed simulation model will most likely pertain to estimated CPTs. On the other hand, the ability to describe dependencies of attitudes related to return to work has value on its own. For instance, the simulation model could serve as a blueprint for a design of interventions and evaluation projects.

Question 2: Within this case study, did MFs provide high-level description of multi-method M&S approach?

During structural analysis phase, a general view of MFs with placeholders for methods offered an overview of multi-method forming study. MFs offered a view of proposed multi-method M&S research design, which in this case was equivalent to a structure of multi-method simulation model. In particular, after structural analysis MFs have shown the structure of a simulation model and both placeholders and possible choices of methods based on research scope and structural requirements. During the analysis of behavior, MFs were filled with methods and relations between components. This offered a high-level view on multi-method simulation model and, in this case, also a study. This view can be naturally extended providing an easy overview of the multi-method M&S approach with consecutive phases of research. For instance, within this case study pseudo-triangulation studies could be considered for different considered methods for different components. MFs would offer an overview of this effort, and the possible subsequent expansions of a simulation model.

Question 3: Was the use of multiple methods justified based on operationalized theoretical developments from Section 3 related to complementarity principles?

The structural falsifiers allowed eliminating three methods from Component 1. This was helpful during further analysis because 625 original possible arrangements (five methods and four components equals 5^4 and assuming that each component can only be developed with a single method) were lowered to 250 arrangements ($2*5^3$). Nonetheless, the structural falsifiers had limited power, which reflects similarity of methods in the context of the desirable structure. For instance, SD seems less appropriate to use than ABM, but it can be used as stated by structural falsifiers. At this point, the use of multiple methods was not justifiable, because there are two methods that could be used as a single method model. On the other hand, at least in principle it is possible that more uniquely structured model would have benefitted more from structural falsifiers in the context of method selection. For instance, if two different methods were evaluated exclusively as false based on structural falsifiers for two different components and all other methods evaluated to true, structural falsifiers would have been able to justify the use of multiple methods. Structural falsifiers can also be useful to demonstrate situations where no method will evaluate a falsifier to false, which means that the methods considered cannot represent a required structural configuration.

First step of behavior analysis investigated known facts of system and phenomena from the perspectives of critical behavioral elements required in a simulation model. This allowed for identification of behavioral characteristics of individual components. In the second step, behavioral falsifiers were derived. They allowed eliminating five options (one from Component one, and two from Components two and three each). This was helpful during further analysis because 250 combinations were lowered to 45 possible arrangements. Nonetheless, the behavioral falsifiers as true and false statements had limited power, which reflected similarity of methods in the context of the desirable behavior. The difficulty in deciding which method to use when more than one was still possible was explored in the third step. The reasoning for choosing methods based on falsifiers in the context of falsification criteria provided additional context to enhance conceptualization and scoring. This facilitated the ability to reason in cases where a true or false approach based on structural and behavioral falsifiers yielded ties. Scoring different methods via extended falsifiers required critical thinking about merits of each method in the context of both behavior characteristics and falsification criteria. This also improved the understanding of the purpose by showing desirable, not idealized, view of a simulation model through the falsification criteria (especially for Component four).

The CoMSs that included all falsifiers provided a numerical value as an argument that multiple methods are used in this situation to complement each other (see Table 31). For instance, falsifiers for Components one, two, and three were zero for BN. On the contrary, falsifier for Component four had the highest score for BN. This leads to a conclusion that BN needed to be complemented with another method(s) that have a non-zero score for falsifiers one, two, and three.

Operationalized theoretical developments i.e. evaluated structural and behavioral falsifiers, and behavioral falsifiers in the context of falsification criteria have helped to build confidence in the solution that comprises multiple methods. The overall score of combined ABM, SC, and BN was estimated superior to any other individual method or combination of methods. The scores indicated that structural characteristics of ABM are complemented by behavioral advantages of SC and BN, and vice versa. ABM is the only method that could be used for all falsifiers, but with lower scores for three out of six falsifiers. This confirms finding from Chapter 3, where the exploration of general criteria led to the same general conclusions about complementarity of methods as a principle that could guide multi-method M&S approach.

Statement 4a : The use of multiple methods was justified by examining if a similar insight could not be generated without using a multi-method simulation model (adapted from Greene [27]).

It is difficult and problematic to prove this statement. First, it is difficult to prove that any insight could not have been generated in spite of doing any research at all. In order to be able to discuss this question, more pragmatic philosophical view must be assumed. One pragmatic way would require employing a pseudo-triangulation (see Section 2.2.2 and Section 3.3.1). MFIII.2 would be required, because different methods (in this case a single method simulation model) would have to be developed, experimented with, and compared to the insights generated with the multi-method simulation model. This could establish limitations of particular solutions, which could be partially tied to generated insights. If the used method cannot generate inputs or outputs at the desired scope, resolution, accuracy, or precision the statement becomes much easier to evaluate. The insights that arise from the conceptualization process itself are difficult to pseudo-triangulate and would require more than a single researcher.

According to the simulation output, prescribed levels of incentives significantly increased percentage of population with benefits terminated for work as compared to the base case with no incentives. Although results are based on some estimated by the author CPT values of BN, similar insight related to ability to inspect different type incentives like vocational assistance incentives, money incentives, and health improvement incentives would not have been possible within time afforded if e.g. DES, ABM, or SC were employed. More specifically, *look for work* transition in SC was represented by BN to disaggregate the meaning of this transition into a theoretical model with multiple input variables.

The multiple methods enabled the ability to combine systemic and theoretical levels of return to work phenomenon. The generated insight pertains also to plausibility of using M&S approach to enhance decision-making. The results showed promising use of this approach, which would not have been possible when a single method was used. The only method that had a potential to realize the simulation model as conceptualized was ABM. Because ABM would have required coding all the parts, and the given time constraints as two months, it would be highly unlikely to represent SSDI process, work status of enrollees and their attitudes with the same resolution, accuracy and precision to generate comparable insight. If considering any other method other than ABM, a single method solution appears limited. The use of BN could provide insight into type of primary data to collect during interventions and evaluations. When considering amount of time spent on modeling, it is unlikely to achieve this level of conceptual understanding of return to work phenomenon without using BN combined with other methods. The only comparable to BN method would be SD, because of its capabilities to represent theoretical concepts. Moreover, the multi-method simulation model propelled ideas to use M&S as a part of bigger research efforts. M&S activity could be in the future a part of a larger multistage, both mixed-method and multi-method M&S research projects as shown conceptually in Figure 37. For instance, when more data becomes available at different stages of research new relationships can be investigated or validated, allowing shifting inputs and outputs within the model.

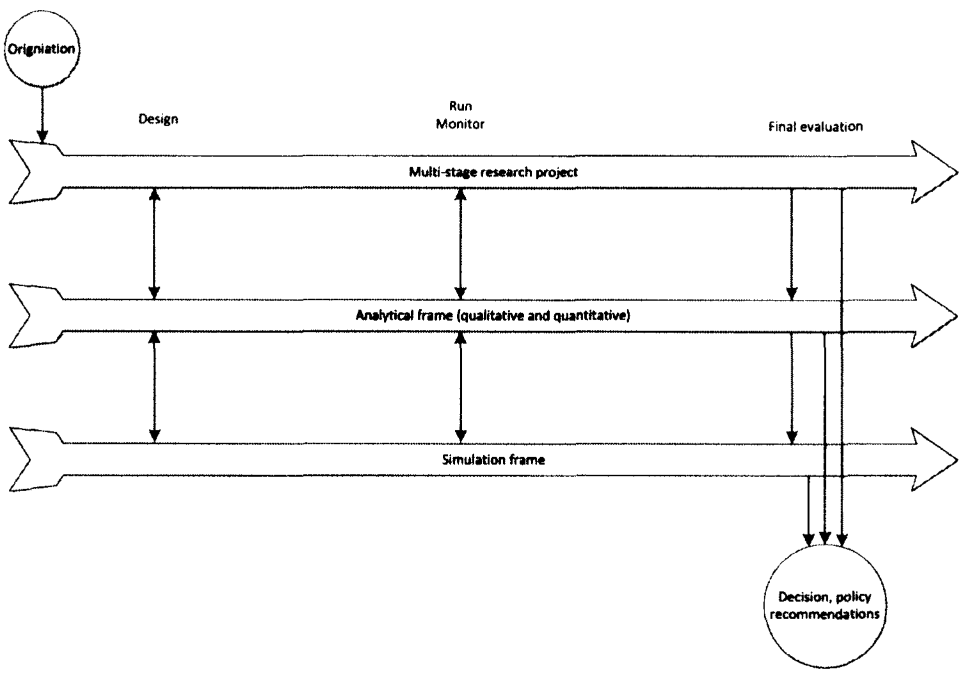

Figure 37. M&S activity within a larger multistage, both mixed-method and multi-method M&S research project.

A multi-method M&S approach seems especially appealing in this context, allowing for more flexible model expansion and analysis when additional data become available. This could enhance both descriptive view and allow for additional testing of theoretical assumptions, having positive impact on model, experimentation, and ultimately improving research objectivity and decision-making.

Statement 4b: The use of multi-method M&S approach facilitated the use of different perspectives e.g., insider and observer's views (adapted from Onwuegbuzie and Johnson [118]).

With the current state-of-the-art simulation at multiple levels of analysis could combine aspects of both qualitative and quantitative empirical worldviews. Based on the case study, the use of multiple modeling and simulation (M&S) methods within a single simulation model can be helpful to gain insight for the system at multiple levels of abstraction. For instance, the developed within this case study simulation model permitted for both insider's and observer's views. ABM provided a macro-level shell of enrollees, which facilitated observer's view in terms of aggregated characteristics of enrollees as statistical counters and graphs (see Figure 38). SC and BN enhanced the insider's view (see Figure 28 and Figure 29). SC facilitated internally viewed SSDI process and behavior of individual enrollee, while BN enhanced representation of factors related to attitudes and intentions. It is noted that ABM on its own could also facilitate both the observer's and the insider's view, but SC and BN complemented ABM to facilitate enhanced insiders' views.

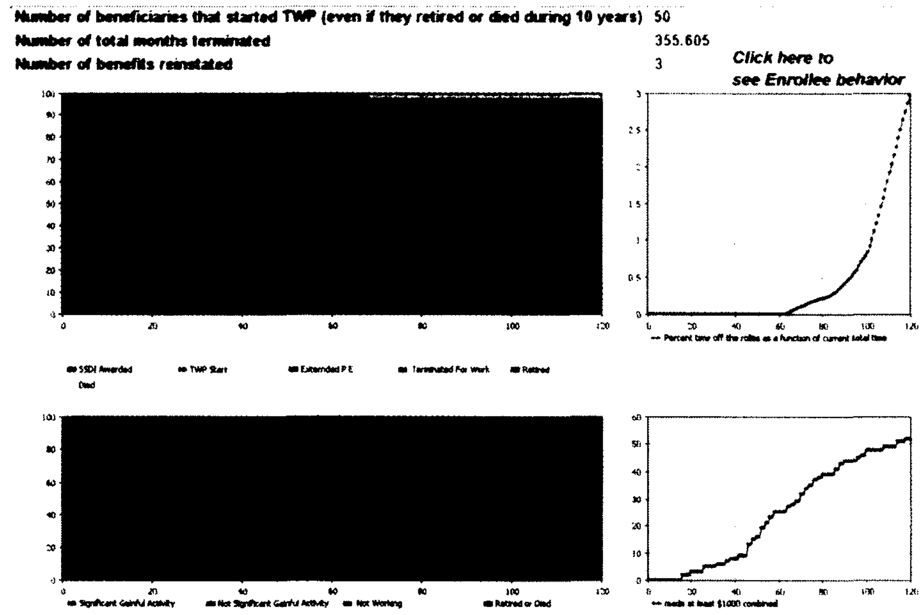

Figure 38. Aggregated view of enrollees' characteristics as statistical counters and graphs.

Statement 4c: The weakness from one approach were compensated by the advantages from the other approach (adapted from Onwuegbuzie and Johnson [118]).

The SC and BN had structural weakness related to representation of population of enrollees and were eliminated as possible solutions during selection process. On the other hand, they outscored other methods based on their advantages related to effective modeling of internal behavior of enrollees.

Introduction to the external evaluation (Questions 5 and 6).

The described in Section 5.4 case study was summarized in a paper format and submitted to Winter Simulation Conference 2014. For this reason, the external evaluators refer to this piece of work as a paper and not as a dissertation. It is pointed out that the paper reviewed by M&S experts did not contain all information included in this work due to space limitations. All reviewers recognized novelty and value of modeling using multi-method M&S approach. Reviewers 1 and 2 noticed also a possibility of application of this multi-method approach to other domains. Reviewer 2 mentioned important aspects related to possible collaboration and extending M&S to different domains of science (proposed to engage with Statisticians). The positive impact was evaluated mainly within application of this approach domain, because the analysis domain is limited due to mentioned earlier lack of empirical data for CPTs. The main critique came from reviewer 3 pointing at CPTs problem, and extensive calibration of the simulation model. The following is the exact evaluation text provided by M&S experts.

Question 5: What is the novelty presented in the paper?

Reviewer 1:

This is a nice application of the ABM methods together with a Bayesian-inspired model for driving agent activity. This is useful in terms of potential policy application, though the paper does not focus on that dimension as much as on the modeling process.

Reviewer 2:

[The novel]… Is the development of an ABM with the aim of understanding the return to work decision of those in receipt of SSI in the USA. The representation of individual agent level behavior through the use of embedded BNs within each agent is also novel. The conceptual model of how and why the ABM was combined with individual BNs can be applied in other domains.

Reviewer 3:

The combination of a Bayesian Network and Agent-based simulation. The application of the combined method in the context of social security insurance scheme.

Question 6: How do you evaluate the potential impact on the application field?

Reviewer 1:

It seems to be a competently done analysis of mixing the agent based modeling and the Bayesian belief network ideas for agent state in an interesting application. There are probably a number of applications where this approach might also be useful. The impact is more likely to be in the application domain more so than the analysis domain.

Reviewer 2:

The combination of BN and ABM is a novel idea and the paper suggests that they have been successfully integrated. Typically, the agent behavior in ABMs is captured using state charts informed by behavioral rules, which can range from very simple to very complex. The BN approach proposed in this paper adds to the existing methods of how to represent behavior in ABMs. It is also an opportunity to engage with Statisticians to discuss simulation modeling approaches such as ABMs as the introduction of BNs into the ABM is a technique that many statisticians will be familiar with.

Reviewer 3:

The potential impact is probably minimal given that this appears to be a poorly described proof-of-principle study with the value of many of the input parameters determined by calibration. In addition, no conditional probabilities that are vital for the Bayesian Network are reported in the paper, other than a statement that sample values were used. Whether an empirical derivation of these probabilities is feasible or not is not discussed. Finally, the paper is difficult to follow as there is no clear structure. There is no section with modeling requirements or assumptions and as such, it is difficult to evaluate the quality of the behavior modeling as this is presented by the two Statecharts.

5.5.2 Demonstration

The structural and behavioral analysis conducted in Sections 5.4.4 and 5.4.5 did not lead to selection of DES for any of the components. In turn, ABM, SC and BN were chosen and used to develop a multi-method simulation model. The question arises if the decision to discard DES as an overarching method was sufficiently justified. The purpose of this demonstration is to attempt a pseudo-triangulation of the developed multi-method simulation model using DES. This effort could demonstrate whether DES-based solution can capture inputs, outputs and possibly the results of the multi-method simulation model. It examines the feasibility of representation and accuracy using DES in comparison to selected multi-method configuration in spite of pointed out in Sections 5.4.4 and 5.4.5 weaknesses of DES. The MFs for this pseudo-triangulation are shown in Figure 39.

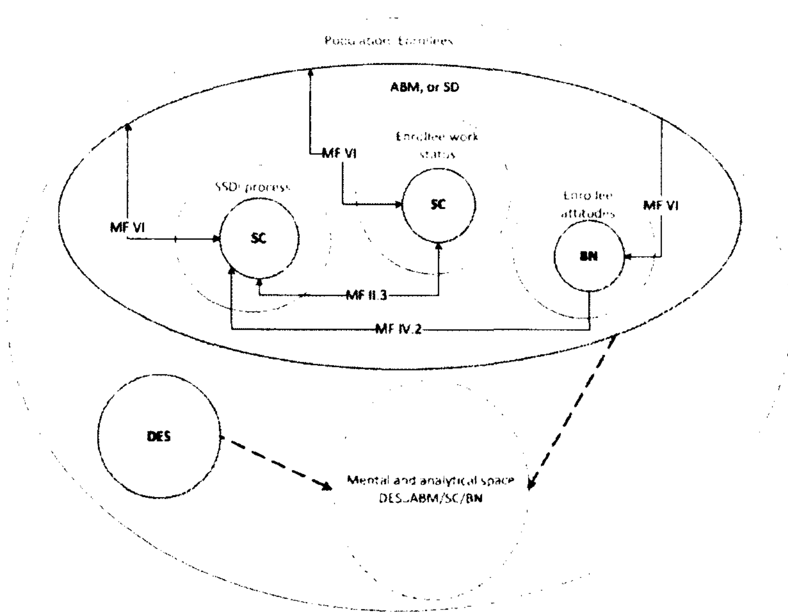

Figure 39. MFs of demonstration.

During the analysis of structural falsifiers, DES was eliminated as a candidate for Component one because it was determined that each enrollee as DES entity cannot facilitate composition of SSDI process, enrollee work status, and enrollee attitudes.

Based on the definition of passive entity (see Appendix A), a DES enrollee cannot create an active behavior related to, e.g., triggering characteristics. This means that an entity can only use its attributes to store information, which can be reevaluated at the processing blocks. Because of these and possibly other constraints, some workarounds will likely be necessary to represent all components using DES. This situation is problematic because pseudo-triangulation at the low-level conceptual method level is not directly possible. This requires reformatting the low-level conceptual model using DES characteristics. Two options were considered to represent transitions between phases in DES as shown in Figure 40.

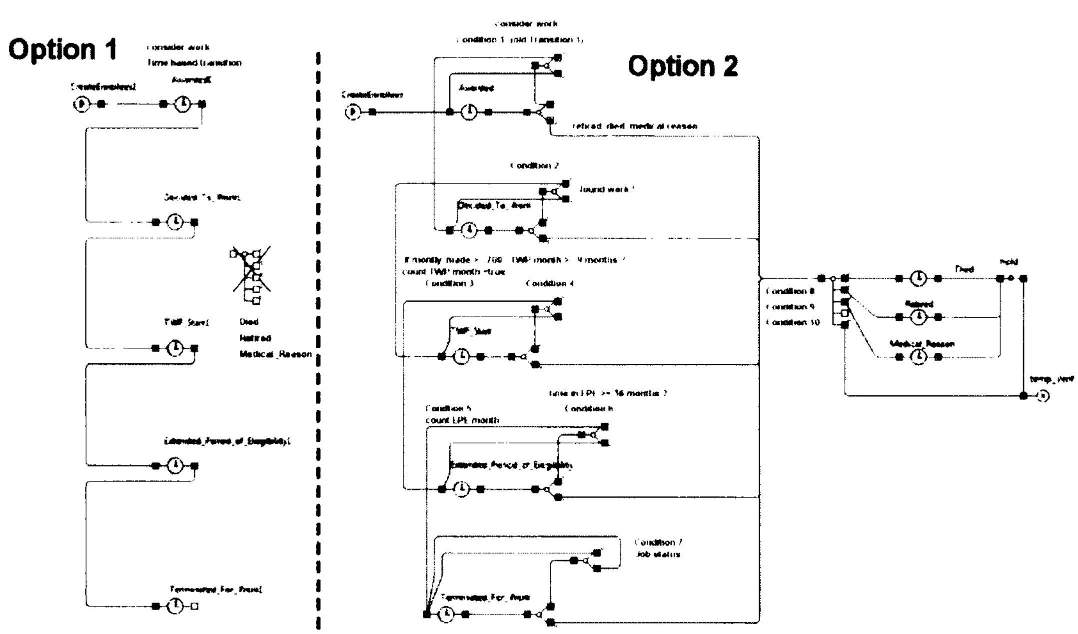

Figure 40. Modeling options considered using DES.

The first option uses time-based delay blocks, e.g., using a PDF as a single-tick transition between phases. For instance, a PDF would represent a transition when enrollee decides to work from *Awarded1* to *Decided_To_Work1* phases. The problem with this representation is that, while the entity is within the delay block, it does not evaluate other possible transitions, e.g., in order to go the *Died*, *Retired*, or *Medical Reason* stages when appropriate. Moreover, when looking at TWP and EPW phases, this approach is not appropriate because entity cannot evaluate its status in the context of systemic rules, e.g., work status, including monthly salary, or time spent in the phase. Option 2 is based on an artificial loop represented using delay block set to a time interval and two subsequent select path objects. This looping of entities allows for reevaluation of various conditions at gates. For instance, the condition would test every month, or other desired time interval, to see if enrollee decided to work. This also allows evaluating if the enrollee belongs in the *Died*, *Retired*, or *Medical Reason* stages. These loops aim to mimic active behavior of agents represented in the multi-method version using SC. Option 1 permits the lumping of the transition as a single time-based event, but it does not permit representation critical to model the systemic rules that evolve with time, hence a second option is required. It is noted that in DES there is no possibility to take the best out of the two options. The solution where work status is represented as a function influencing attributes of enrollee at gates lacks flexibility due to constrained to gates process. This would be even more problematic if multiple different time-interval decisions are considered by the entity. In the multi-method view with SCs inside of an agent, one can use both options because of an unconstrained ability to represent behavior at various desirable frequencies as internally generated events at multiple hierarchical levels of SC (behavior resolution).

Whereas representation of output variable is straightforward in DES model, the input variables are problematic. The theoretical model represented using BN offers an insight into hidden aspects related to transitions within SSDI process controlled by human behavior. DES does not provide components that could facilitate easy representation of an enrollee's attitude toward work. This means that the enrollee's decision to work would have to be generated based on some developed function that would include necessary inputs (*vocational assistance, money incentives,* and *health improvement programs*), and would be evaluated at gates. This algorithm should also include all internal variables and their relations as used in BN. Creating an ad hoc method using Java to capture these dependencies would require developing some method that could capture inputs, outputs and dependencies as conceptualized, which will not be necessary if DES were sufficient.

Precision of DES representation related to the considered evaluation period of systemic SSDI process rules is the same as in the multi-method simulation model because monthly evaluation serves as a period in both cases. On the other hand, there is a difference in representation of work status, where DES has limitations related to precision. In the original simulation model, salaries were modeled based on dynamic events internal to each enrollee (agent), where a reevaluation period was part of the calibration. This straightforward implementation is more difficult in the DES version, especially if less than monthly intervals would be needed. In the DES solution, all behaviors have to be represented in the context of the lowest considered evaluation period, so any asynchronous behavior with precision less than considered evaluation period would have lower precision. On the other hand, one can lower evaluation period to the lowest desirable interval, but subsequent adjustments and changes in DES solution must be coordinated to adhere to the new period. This limitation may be less visible through numerical accuracy because of theoretical focus of the representation, and asynchronous dynamic events with precision based on monthly intervals.

Calibration against historical data across different phases was conducted for DES simulation model in a similar fashion as in the first multi-method simulation model, but using specific parameters. These parameters were used within functions controlling transitions between phases and within probability distribution functions (Beta was used as in the first model) to generate monthly earnings of enrollees. Calibrations of DES and multi-method simulation models are shown in Figure 41. Calibration errors for multi-method simulation model were a little lower than DES, i.e., TWP complete phase was 3.111 for DES (see Figure 41, segment A) versus 1.057 for the original model; EPE completed phase error was 1.834 (see Figure 41, segment B) versus 1.132; and benefits terminated phase was 1.005 (see Figure 41, segment C) versus 0.0772. The run speed of DES was faster, although overall amount of time spent on calibration of both models was not measured. Because model logic in DES was largely different from the multi-method simulation model, most of the parameters are not comparable. Moreover, without the theoretical model considered, only partial analysis of DES simulation results was possible. A point estimator of percent-terminated enrollees because of work was compared with a historical data point.

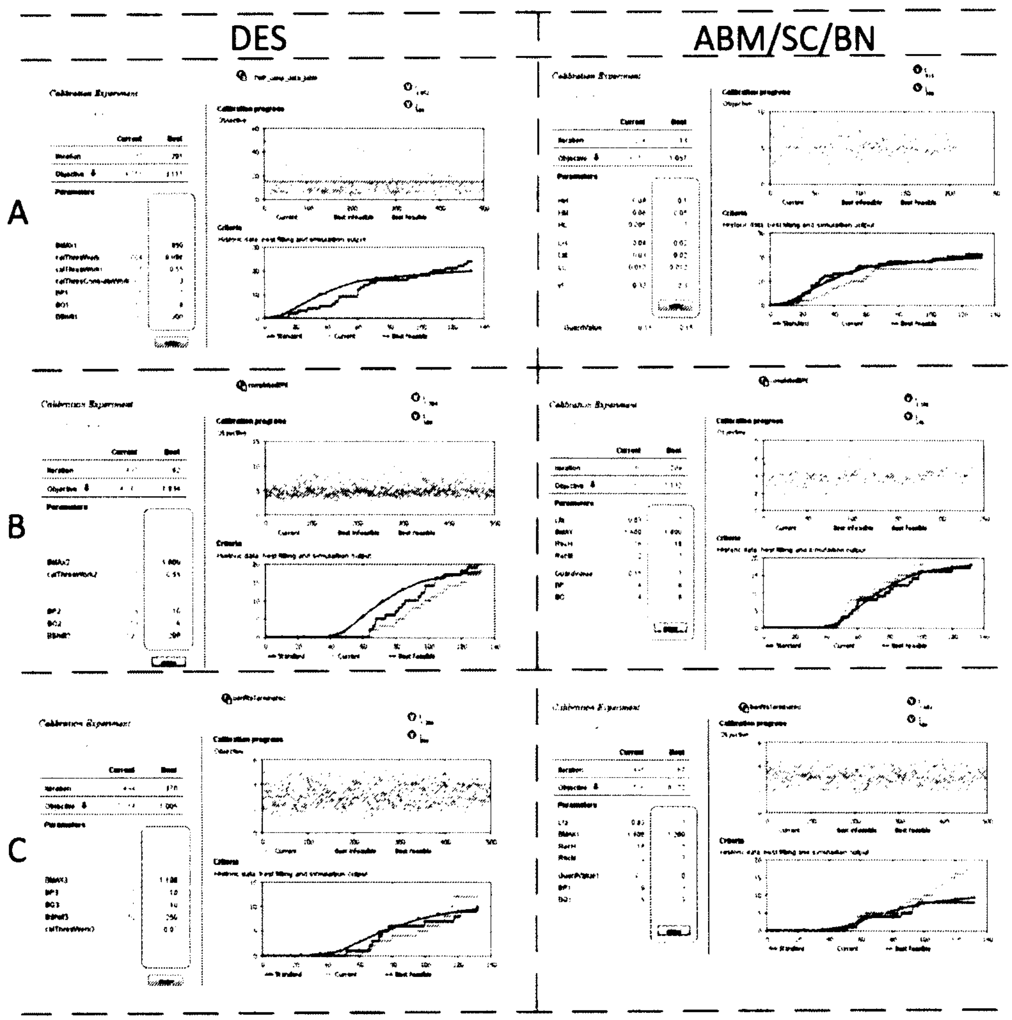

Figure 41. Calibrations of DES and multi-method simulation models.

Initial analysis based on calibrated DES model showed significant difference, hence additional calibration (see Figure 42) was conducted to verify a possibility to produce a statistically valid point estimator. A sample mean of benefits terminated was used for statistical validation of output from DES simulation model using the test statistic value $z = \frac{\bar{x} - \mu_0}{\delta/\sqrt{n}}$, where $\bar{x} = 9.357$ is a sample mean of percent of benefits terminated; $\mu_0 = 9.5$ is the historical value of benefits terminated; $\delta = 1.16$ is the sample standard deviation, and $n = 200$ is the sample size. The conducted test is based on a two-tailed

$z_{0.025} = 1.96$ at a significance level of 0.05. The resulting $z = 1.74 \leq 1.96$, so the model cannot be proven to produce results different from statistical historical data.

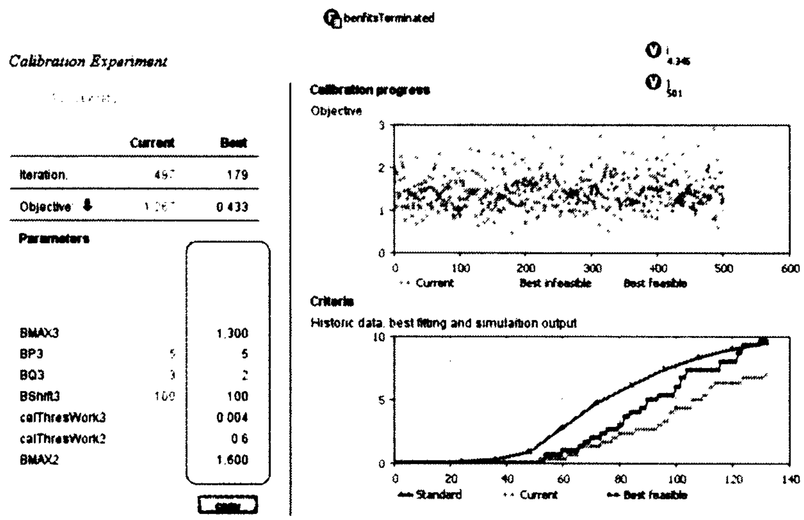

Figure 42. Additional calibration.

Comparison between two simulation models also did not show a statistical difference. A paired-t approach was used with $\bar{Z}(n) \mp t_{n-1,1-\alpha/2}\sqrt{\frac{\sum_{j=1}^{n}[Z_j-\bar{Z}(n)]^2}{n(n-1)}}$ at level 0.05 for n = 200, t = 1.96. Because the resulting interval (-0.45, 0.46) contained zero, the difference between benefits terminated for both simulation models cannot be shown as significant. The practical significance of this result is low since the additional calibration of DES model toward a point estimator was necessary. A conducted validation against the point estimator did not provides sufficient credibility. On the other hand, trajectory data provides more information about accuracy of simulation models, facilitating higher credibility. Figure 43 presents four phases of the original multi-method simulation model, while Figure 44 presents DES version. Each version is based on a population of 1000 enrollees. Mean Absolute Error (MAE) between simulation trajectory and historical data was calculated for both simulations and t-paired approach for n = 132 was used to compare them (see Table 32).

The trajectories resulting from DES simulation model appeared less fitted and linear when compared to the multi-method simulation model. This could be traced back to the assumptions and parameters of each model and possibly mitigated by spending more time on development. MAE for ABM/SC/BN is lower than for DES.

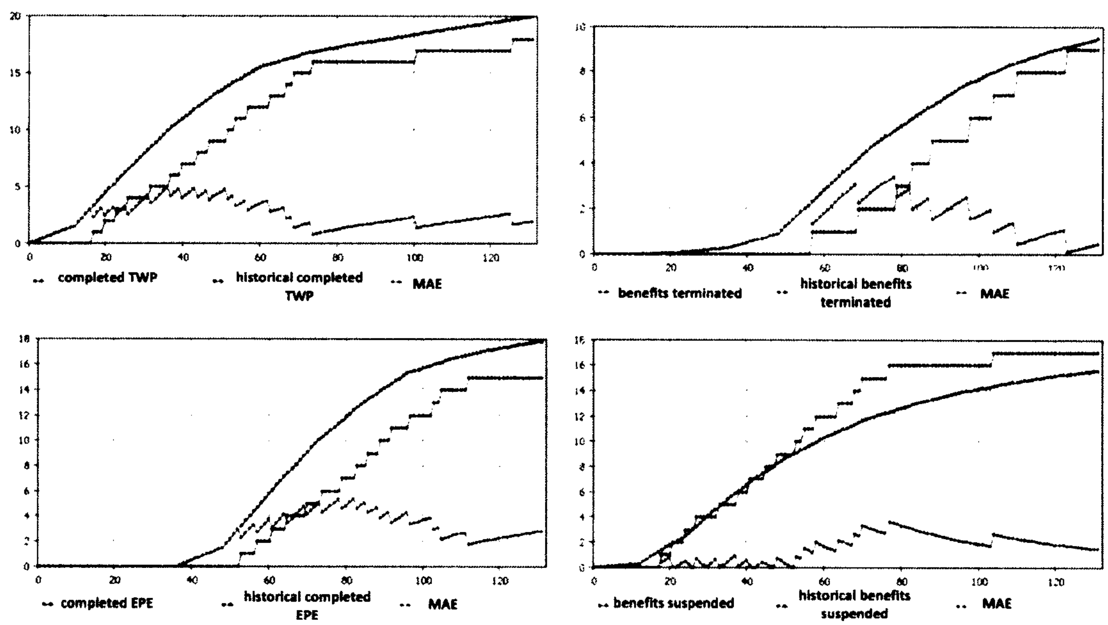

Figure 43. MAE for the multi-method simulation model.

The difference is rather small although statistically significant for all phases. Based on visual inspection of both graphs, DES appears more precise when compared to ABM/SC/BN. This is the effect of monthly reevaluation, and it is only apparently more precise due to functions reevaluated each month by DES.

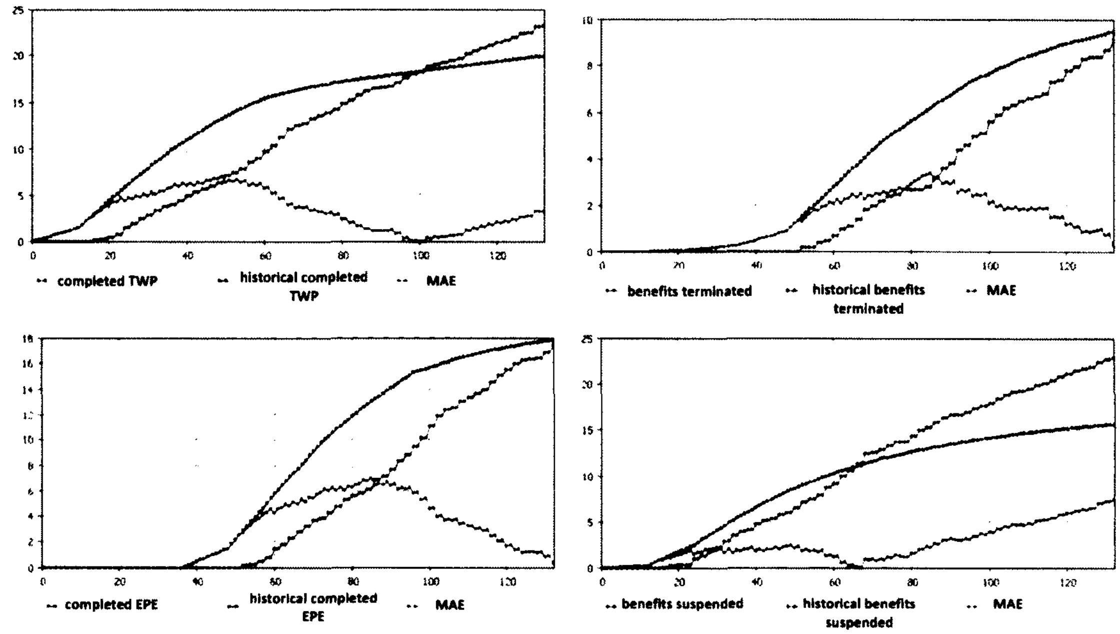

Figure 44. MAE for the DES simulation model.

Table 32. Analysis of trajectories for both ABM/SC/BN and DES simulation models.

Phase	MAE for ABM/SC/BN	MAE for DES	Interval difference at 0.05 level
TWP complete	2.44	3.18	(0.51, 0.97)
EPE complete	2.25	2.77	(0.35, 0.69)
Benefits suspended	1.47	2.58	(0.77, 1.44)
Benefits terminated	1.14	1.39	(0.17, 0.33)

The DES simulation model was faster. A single run simulation with 100 entities took on average 0.3 seconds, while the multi-method simulation model with 100 agents took about 4.0 seconds. It was more difficult to track codependency of SSDI process and could be easily tracked by visual inspection of SC, which helped during verification of the simulation model.

As noted in Section 5.5.1, objectivity of each structural falsifier can be traced to degree of falsifiability as information that can facilitate insight for an external critique. The original decision to discard DES during structural analysis because it could not facilitate composition relation is problematic. The DES entity, as implemented in AnyLogic®, could facilitate composition of objects, e.g., a BN method object, which was found during development of this demonstration. The use of the embedded objects within entity is permitted, but limited to types of object and their use at gates. Active objects cannot be used from within entity reflecting its passive character. The structural falsifier two from Section 5.4.4 should have specifically stated the composition as it pertained to the active objects. Moreover, DES implementation demonstrated that desirable degree of falsifiability in the context of active behavior was slightly inflated, which is visible when comparing DES with the multi-method simulation model. For instance, by using recurring function calls at the gates, DES mitigated the active behavior requirements in this particular case. On the other hand, it is better to set the bar too high rather than too low. Moreover, the processing view is less practical and intuitive for the research that focuses on individual behavior. Adding further internal logic of an enrollee could make the use of DES even more difficult. DES results were a little less accurate against historical data when compared to the multi-method simulation model. On the other hand, the comparison of results is irrelevant if one of the simulation models could not represent desirable unique characteristics. A single method DES cannot directly mimic desirable characteristics captured by all selected methods, especially the theoretical construct captured by BN. Although the DES model closely replicated the historical output data for a base case scenario, its inflexibility to represent the underlying theoretical construct does not make it very useful in the context of the research question. On the other hand, if the DES representation was enhanced using BN, and mindful of the DES limitations, DES instead of ABM would have seemed a more reasonable option.

The original model was developed with considerations related to desirable degree of falsifiability as defined by falsifiers, which if underestimated can lead to a workable solution, but a more constraining view, and conceptualization options that can compromise fidelity of the representation. Moreover, slightly overestimated degree of falsifiability may not necessarily have a negative impact as shown in this case (except for simulation run speed). The DES model had to be developed differently to compensate for characteristics of ABM/SC/BN, which required some workarounds. Although, at the current stage of research, it provided a similar numerical accuracy, going forward with the DES model would be challenging, e.g., adding new considerations and codependences related to the limited passive entities. It is noted that the conducted pseudo-triangulation was not possible at the low-level conceptualization using DES, but was based on the already developed high-level concept translated into DES simulation model. DES showed potential for a high-level concept representation using its constructs in spite of its falsifier structural weakness. This demonstrated the value of pseudo-triangulation as a mean of verification of selected approach. It would also be interesting to conduct triangulation at the level beyond high-level concept using independent modelers and explore differences of the conceptualization processes and outcomes using DES only and multi-method view (ABM/SC/BN).

5.6 SUMMARY

This case study developed a multi-method simulation model related to a return to work phenomenon. The M&S process was used as a data layer for evaluation. The complementarity principles were applied during choosing method process using developed structural and behavioral falsifiers, and later using behavioral falsifiers estimated in the context of falsification criteria. The evaluation indicated a value of using principles of complementarity to aid during selection of methods. The structural and behavioral falsifiers were able to eliminate methods, but fell short to specify a final view of MFs because ties between methods for a given falsifier were still present. Subsequent scoring of behavioral falsifiers in the context of falsification criteria allowed to select methods, and development of unique MFs for multi-method simulation model, but the objectivity of the reasoning could be argued.

Sub-falsifiability scores and CoMS calculated based on falsifiers in the context of falsification criteria facilitated ability to reason in the cases where true, false approach based on structural and behavioral falsifiers yielded ties. The scoring process required critical thinking about merits of each method in the context of both behavior characteristics and falsification criteria. CoMS for combined ABM, SC, and BN methods were superior to any other individual method or combination of methods, which helped to build confidence before developing such a solution. The scoring method facilitated ability to select methods, but the reasoning can be questioned, which is also valuable in the context of external method falsifiability. The sub-falsifiability scores and CoMS indicated that structural characteristics of ABM are complemented by behavioral advantages of SC and BN, and vice versa. ABM was the only method that could be used for all falsifiers, but with lower scores for three out of six falsifiers. This demonstrates value of developments from Chapter 3 building credibility in complementarity principles as a basis for the use of a multi-method M&S approach. The proposed guidelines have potential to improve multi-method conceptualization and decision to choose appropriate methods. They improved level of confidence to support the use of multi-method M&S approach in this case. Other important element of practice, which exists within multi-method, pertains to the disclosed information about methods and software. This provides the boundary of a particular multi-method M&S approach, which is important to understand better methodological scope, and allows for an external critique by M&S community.

The application of multi-method simulation model provided an insight about phenomenon explored, which can be partially attributed to methods' unique characteristics. The developed simulation model was used to mimic a return-to-work phenomenon. A multi-method simulation model that consisted of ABM, SC, and BN was used in an attempt to capture system conditions and enrollee behavior. The RQ led to a simulation model that connected attitude of enrollee toward work incentives and percentage of benefits terminated. The simulation model was validated and experimentation led to conclusion that prescribed levels of incentives significantly increased percentage of population with benefits terminated for work as compared to the base case with no incentives. To improve understating of enrollee behavior, it would be desirable to employ in the future qualitative data collection, and use them within CPT of BN to provide results that are more valid and credible. The growth of the use of a multi-method M&S approach still trails empirical mixed methods in healthcare, but both methodological views are built on similar pragmatic philosophical beliefs, and a combination of both will be the natural next step in the evolution of scientific endeavors.

Although ABM could have been considered for all four components, it would not have allowed for developing the same simulation model with the time allotted because it would require coding most of the characteristics, which is much more time-consuming than reusing modeling blocks existing within the selected methods. The case study demonstrated pseudo-triangulation of the developed multi-method simulation model using DES in spite of its acknowledged weaknesses. DES resulted with trajectories a little less accurate against historical data as compared to a multi-method simulation model. Moreover, DES could not represent desirable unique characteristics, especially theoretical construct captured by BN. Creating an ad hoc method using Java to capture inputs, outputs and dependencies as conceptualized by BN will not be necessary, if the DES was sufficient. The structural falsifier that led to the elimination of DES for Component 1 was imprecise, and its desirable degree of falsifiability in the context of active behavior was somewhat inflated. Summarizing, the demonstration showed that the DES approach would not have produced a comparable simulation model given the same circumstances, although if enhanced with BN it could be considered.

The evaluated research guidelines provide also a high-level insight, which could propel and advance the discussion about relations between objectivity and subjectivity and broadening philosophical views about M&S field as a part of scientific community.

CHAPTER 6

CONCLUSIONS

The concluding chapter of this dissertation provides a review of how the research question was answered with stated objectives, and how this research contributed to the body of knowledge. Moreover, limitations of this research and possible directions for future work are identified.

6.1 CONTRIBUTIONS

The research has explored query about theoretical principles of multi-method M&S approach. The literature review added to the body of knowledge by showing the existing reasoning for the use of the multi-method M&S approach based on M&S literature within a socio-technical context. The found reasoning related to the complementary nature of methods with the additional need for methods coupling, data availability and usability, skills and preference of a modeler, stakeholder acceptability, expectation of unique insight, enhanced with the very diverse needs related to understanding, credibility, validity, and complexity of models. Moreover, projection of purposes for mixing methods from empirical social science contributed by offering perspectives describing two main purposes of multi-method M&S approach i.e. complementarity and pseudo-triangulation. Another contribution to the body of knowledge relates to the demonstrated need for more consistency in using different terms and general guidelines of how to conduct multi-method M&S studies. These contributions were related to finding theoretical basis on multi-method M&S approach, needed as a base for answering research question. The discussed literature review and analysis fulfilled objective one of this dissertation.

A pragmatic philosophical stance provided foundation for the choice of terms and definitions relevant to a multi-method M&S approach. Proposed definitions clarified what multi-method M&S approach is by providing its major derivative definitions and supporting terms. For instance, complementarity of methods was identified as a main reason behind the use of multi-method M&S approach. The analysis and proposed definitions for multi-method M&S approach fulfilled objective two of this dissertation.

Another contribution in the context of the research question pertains to the complementarity principles. They were derived based on theory of falsification as a mechanism for reasoning about method choice that can facilitate desirable level of falsifiability in relation to a study purpose. In this context, the sub-falsifiability score and CoMS were derived as measures supporting complementarity principles. Moreover, the exploration of commensurability provided an additional dimension of analysis of complementarity. MFs contributed to the body of knowledge by providing a way of creating a general blueprint of multi-method M&S approach. Transitions toward formats must seek justifications to increase research objectivity and transparency. The MFs offer overview of the research, and can be used to describe phases of research and a structure of methods within a simulation model(s). The presented theoretical developments fulfilled objective three of this dissertation.

A final contribution to the body of knowledge includes evaluation of the theoretical principles proposed in this dissertation. First, a sample set of methods and criteria were analyzed in the context of commensurability of methods and CoMS. The analysis revealed that commensurability of methods is proportional to mutual CoMS. Moreover, if the considered characteristics were required within a research context, none of the methods could have provided the highest possible sub-falsifiability score without combining methods and the resulting CoMS would have been above zero. Next, a case study offered insights in relation to theoretical basis. A multi-method simulation model that pertained to a return to work phenomenon was developed and used as a data layer for evaluation of complementarity principles. The case study demonstrated the plausibility of complementarity principles as a way to reason about the use of a multi-method M&S approach. The proposed research guidelines offered a scoring approach that involved structural and behavioral falsifiers, in addition to using behavior falsifiers in the context of falsification criteria. Moreover, the application of a multi-method simulation model provided a novel way to explore return to work phenomenon for the disabled population in the U.S. The demonstration attempted to pseudo-triangulate the developed multi-method simulation model using DES in spite of its acknowledged weaknesses. It showed that a DES approach would not have produced a comparable simulation model given the same circumstances, although if enhanced with BN it could be considered. This demonstration uncovered a limitation of devised structural falsifier two. DES was eliminated for Component 1 due to both an imprecise falsifier and a somewhat inflated desirable degree of falsifiability. The analysis demonstrated the value of pseudo-triangulation as a means of verification of a selected approach. The developed case study fulfilled objective four of this dissertation.

6.2 LIMITATIONS AND DELIMITATIONS

The research covered theoretical and methodological challenges and contributed to theoretical basis of multi-method M&S approach.

One of the limitations of this research is related to the evolutionary character of methods as a trend of merging more established methods to create new versions. This can confuse practitioners and researchers. Similarly, a multi-method M&S approach is still evolving and crude, time-consuming to develop, can be problematic, and needa further exploration, which can be visible throughout this dissertation. Although this research has attempted to discover all the different MFs, the ever-evolving aspect of MFs is beyond the control of the author. Moreover, the problem of subjectivity during multi-method conceptualization within research guidelines relates to the fact that simulation models with different configurations of methods do not exist, and cannot be fully observed unless implemented. The goal of the proposed approach is to mitigate this requirement, but it allows only for approximations based on reasoning about future simulation model as discussed by Robinson [113]. On the other hand, if all possible configurations were to be implemented to confirm the reasoning this would have defeated the purpose of having method selection guidelines for studies with multiple methods in the first place.

The major delimitation pertains to the use of a single case study and proposed guidelines. The guidelines were developed with the focus on answering the dissertation's research question and served only as a proof of concept in relation to the theoretical basis. At their core, the guidelines were used as a means of conceptualization for multiple methods, with the reasoning for choice of methods leading to MFs. They therefore do not provide a definite solution that is proven to work in all cases neither to provide a solution for all the methodological problems related to multi-method M&S approach. The method for calculating CoMS was used to operationalize complementarity principles within the narrow scope related to this case study. Moreover, not all MFs were utilized for the case study because they were not selected. Many combinations and hierarches of MFs can be assembled, but examination of all possible cases is outside the scope of this research.

A subjective human conceptualization at various stages was discussed in Section 4.5, and it is added to a list of delimitations. Many factors that affect human subjectivity were exposed and discussed, for instance, methods considered, software considered, type of system and phenomena studied all affected developed falsifiers. For instance, the finite number of methods (SD, ABM, DES, BN, and SC) considered within the case study, limits the evaluation scope in the relation to theoretical principles. On the other hand, disclosed information showed the need and value of additional transparency in multi-method studies e.g., pertaining to the analysis of falsifiers.

Although the underlying motivation for multi-method M&S is based on the principle of complementarity of methods, its overall theoretical basis is concerned with both complementarity and triangulation (including pseudo-triangulation). The final delimitation pertains to the limited exploration of the triangulation echelon within the case study. The next section highlights the need to explore this echelon in the future.

6.3 FUTURE WORK

The explorations of methods in Section 3.3 and a case study analyzed complementarity of methods and its main principles but did not exhaustively examine triangulation. The pseudo-triangulation using different methods may result in expanding the representation of system or phenomena. It would be interesting to experiment more with pseudo-triangulation and examine if methods expand system or phenomena representation at different levels of triangulation (see Section 2.2.2). This could yield better insight into relation between complementarity and triangulation in M&S.

The case study showed that specific method(s) choices were inferior based on the required level of falsifiability. Subsequently, adequate choices were made. The disclosed analysis of the research design options adds objectivity to the research. Another step in the research could focus on CoMS for a set of methods in relation to level of falsifiability of different research questions considered. This could permit the extension of the scope of this research to investigate if CoMS in relation to methods facilitate generation of models (theories) at higher level of falsifiability viewed as higher universality and precision of research questions (see Section 3.3) that could be tackled.

An exploration, review, and categorization of M&S methods in the context of their unique characteristics could provide a knowledge database to a practitioners and researchers. Related research could involve finding differences in implementation of the same methods in different commercial modeling software (or event their subsequent versions) as a way to understand evolution of M&S methods.

The commensurability of methods and models deserves more investigation. For instance, in depth exploration of commensurability of methods in relation to the interaction points seems interesting. The reconceptualization of commensurability in this context seems to have practical value. For instance, the analysis of transformation interaction point could be especially valuable in this context because it could provide a way to categorize method pairs in the context of their possible use.

As the case study showed subjectivity of method choice related to mental conceptualization phase, the understanding of multi-method way of conceptualization could be undertaken by employing in the future brain exploration devices e.g. functional magnetic resonance imaging (FMRI), electroencephalography (EEG), near-infrared spectroscopy (NIS) or hybrid brain-computer interfaces to study brain activities during conceptualization processes. For instance, understanding subjectivity patterns and the main places in the brain where subjective thoughts occurs could lead to better understanding why they happen in relation to methods used, which could make an important contribution to M&S practice (objectivity).

REFERENCES

[1] C. M. Macal and M. J. North, "Tutorial on agent-based modeling and simulation," presented at the Winter Simulation Conference, 2005.

[2] R. Axelrod, "Simulation in the social sciences," in *Handbook of research on nature inspired computing for economy and management*, J.-P. Rennard, Ed., ed Hershey, PA: Idea Group Inc., 2007, pp. 90-100.

[3] D. A. Fahrland, "Combined discrete event continuous systems simulation," *SIMULATION*, vol. 14, pp. 61-72, 1970.

[4] L. Rabelo, M. Helal, A. Jones, J. Min, Y. J. Son, and A. Deshmukh, "A hybrid approach to manufacturing enterprise simulation," presented at the Winter Simulation Conference, 2003.

[5] S. Brailsford, S. M. Desai, and J. Viana, "Towards the holy grail: Combining system dynamics and discrete-event simulation in healthcare," presented at the Winter Simulation Conference, 2010.

[6] K. Chahal and T. Eldabi, "Applicability of hybrid simulation to different modes of governance in UK healthcare," presented at the Winter Simulation Conference, 2008.

[7] Y. H. Lee, M. K. Cho, S. J. Kim, and Y. B. Kim, "Supply chain simulation with discrete–continuous combined modeling," *Computers & Industrial Engineering*, vol. 43, pp. 375-392, 2002.

[8] K. Chahal, ""A generic framework for hybrid simulation in healthcare"," Ph.D. thesis, Department of Computing and Mathematics, Brunel University, West London, 2010.

[9] J. Morecroft and S. Robinson, "Explaining puzzling dynamics: comparing the use of system dynamics and discrete-event simulation," in *23rd International Conference of the System Dynamics Society*, 2005, pp. 17-21.

[10] M. Balaban and P. Hester, "Exploration of purpose for multi-method simulation in the context of social phenomena representation," presented at the Winter Simulation Conference, 2013.

[11] J. Mingers, "Combining IS research methods: towards a pluralist methodology," *Information Systems Research,* vol. 12, p. 240, 2001.

[12] D. A. Sylvan and J. F. Voss, *Problem representation in foreign policy decision-making*: Cambridge University Press, 1998.

[13] T. A. Schwandt, *Dictionary of qualitative inquiry.* Thousand Oaks, CA: Sage Publications, Inc., 2001.

[14] B. Edmonds and S. Moss, "From KISS to KIDS–an 'anti-simplistic'modelling approach," in *Multi-Agent and Multi-Agent-Based Simulation.* vol. 3415, P. Davidsson, B. Logan, and K. Takadama, Eds., ed Berlin Heidelberg: Springer 2005, pp. 130-144.

[15] R. J. Brooks and A. M. Tobias, "Choosing the best model: Level of detail, complexity, and model performance," *Mathematical and computer modelling,* vol. 24, pp. 1-14, 1996.

[16] J. Sterman, *Business dynamics.* U.S.: McGraw-Hill, 2000.

[17] A. M. Law, *Simulation modeling and analysis*, 4th ed.: New York: McGraw-Hill, 2007.

[18] K. Hoad and C. Watts, "Are we there yet? Simulation modellers on what needs to be done to involve agent-based simulation in practical decision making," *Journal of Simulation,* vol. 6, pp. 67-70, 2012.

[19] S. Lieberman, "Extensible software for whole of society modeling: Framework and preliminary results," *Simulation,* vol. 88, pp. 557-564, 2012.

[20] J. Verkuilen and M. Smithson, *Fuzzy set theory: Applications in the social sciences* vol. 147: Sage Publications, Inc., 2006.

[21] S. Hassan, L. Garmendia, and J. Pavón, "Agent-based social modeling and simulation with fuzzy sets," *Innovations in Hybrid Intelligent Systems,* vol. 44, pp. 40-47, 2007.

[22] A. Tashakkori and C. Teddlie, *Handbook of mixed methods in social & behavioral research.* Thousand Oaks: Sage Publications, Inc., 2002.

[23] C. Swinerd and K. R. McNaught, "Design classes for hybrid simulations involving agent-based and system dynamics models," *Simulation Modelling Practice and Theory,* vol. 25, pp. 118-133, 2012.

[24] N. Bohr, *The quantum postulate and the recent development of atomic theory*: Printed in Great Britain by R. & R. Clarke, Limited, 1928.

[25] S. Brailsford, L. Churilov, and S. K. Liew, "Treating ailing emergency departments with simulation: An integrated perspective," in *Health Sciences Simulation*. vol. 25, J. Anderson and E. Katz, Eds., ed San Diego, USA: Society for Modeling and Computer Simulation, 2003.

[26] J. Mingers, "Variety is the spice of life: combining soft and hard OR/MS methods," *International Transactions in Operational Research,* vol. 7, pp. 673-691, 2000.

[27] J. C. Greene, *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass, 2007.

[28] P. T. Hester and A. Tolk, "Applying methods of the M&S spectrum for complex systems engineering," presented at the Proceedings of the 2010 Spring Simulation Multiconference, San Diego, CA, 2010.

[29] S. K. Heath, S. C. Brailsford, A. Buss, and C. M. Macal, "Cross-Paradigm Simulation Modeling: Challenges And Successes," presented at the Winter Simulation Conference, 2011.

[30] L. Lättilä, P. Hilletofth, and B. Lin, "Hybrid simulation models – When, Why, How?," *Expert Systems with Applications,* vol. 37, pp. 7969-7975, 2010.

[31] A. Borshchev, *The Big Book of Simulation Modeling*: AnyLogic North America, 2013.

[32] D. Crespo and M. Ruiz, "Decision making support in cmmi process areas using multiparadigm simulation modeling," presented at the Winter Simulation Conference, 2012.

[33] A. Tolk, S. Y. Diallo, J. Padilla, and H. Herencia-Zapana, "Reference modelling in support of M&S—foundations and applications," *Journal of Simulation,* vol. 7, pp. 69-82, 2013.

[34] T. Lorenz and A. Jost, "Towards an orientation framework in multi-paradigm modeling," presented at the Proceedings of the 2006 International Conference of the System Dynamics Society, Nijmegen, The Netherlands, 2006.

[35] C. Glazner, "Understanding Enterprise Behavior Using Hybrid Simulation of Enterprise Architecture," ed. Engineering Systems Division, Massachusetts Institute of Technology, Cambridge, MA, 2009.

[36] P. Lewis, M. N. Saunders, and A. Thornhill, *Research methods for business students*. Rotolito Lombardia, Italy: Pearson, 2009.

[37] M. N. Saunders, M. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students, 5/e*: Pearson Education India, 2011.

[38] P. J. Adams and S. Buetow, "The place of theory in assembling the central argument for a thesis or dissertation," *Theory & Psychology,* vol. 24, pp. 93-110, 2014.

[39] J. J. Padilla, S. Y. Diallo, and A. Tolk. (2011) Do we need M&S science? *SCS M&S Magazine*.

[40] E. R. Smith and F. R. Conrey, "Agent-Based Modeling: a new approach for theory building in social psychology," *Personality and Social Psychology Review,* vol. 11, pp. 87-104, February 1, 2007 2007.

[41] J. A. Sokolowski and C. M. Banks, *Modeling and simulation for analyzing global events*. Hoboken, NJ: Wiley, 2009.

[42] R. K. Yin, *Case study research: Design and methods* vol. 5: SAGE Publications, Inc., 2008.

[43] R. J. Torraco, "Research methods for theory building in applied disciplines: A comparative analysis," *Advances in Developing Human Resources,* vol. 4, pp. 355-376, 2002.

[44] J. C. Greene, V. J. Caracelli, and W. F. Graham, "Toward a conceptual framework for mixed-method evaluation designs," *Educational Evaluation and Policy Analysis,* vol. 11, pp. 255-274, 1989.

[45] K. Popper, *The logic of scientific discovery*: Routledge, 2002.

[46] A. Tolk, "Truth, trust, and turing – Implications for modeling and simulation," in *Ontology, Epistemology, and Teleology for Modeling and Simulation*. vol. 44, A. Tolk, Ed., ed: Springer Berlin Heidelberg, 2013, pp. 1-26.

[47] A. Tolk, B. Heath, M. Ihrig, J. Padilla, E. Page, E. Suarez, *et al.*, "Epistemology of modeling and simulation," presented at the Winter Simulation Conference, 2013.

[48] T. Eldabi, R. J. Paul, and T. Young, "Simulation modelling in healthcare: reviewing legacies and investigating futures," *Journal of the Operational Research Society,* vol. 58, pp. 262-270, 2006.

[49] J. Zulkepli, T. Eldabi, and N. Mustafee, "Hybrid simulation for modeling large systems: an example of integrated care model," presented at the Winter Simulation Conference, 2012.

[50] A. Kott and P. S. Corpac, "COMPOEX technology to assist leaders in planning and executing campaigns in complex operational environments," in *12th International Command and Control Research and Technology Symposium*, Fairfax, VA, 2007.

[51] M. Helal, "A hybrid system dynamics-discrete event simulation approach to simulating the manufacturing enterprise," Master's thesis, Department of Industrial Engineering and Management Systems, University of Central Florida, Orlando, Florida, 2008.

[52] A. Djanatliev, R. German, P. Kolominsky-Rabas, and B. M. Hofmann, "Hybrid simulation with loosely coupled system dynamics and agent-based models for prospective health technology assessments," presented at the Winter Simulation Conference, 2012.

[53] J. Viana, S. Rossiter, A. A. Channon, S. Brailsford, and A. Lotery, "A multi-paradigm, whole system view of health and social care for age-related macular degeneration," presented at the Winter Simulation Conference, 2012.

[54] S. Brailsford and N. Hilton, "A comparison of discrete event simulation and system dynamics for modelling health care systems," in *Planning for the Future: Health Service Quality and Emergency Accessibility*, J. Riley, Ed., ed Glasgow Caledonian University, 2001.

[55] R. Sachdeva, T. Williams, and J. Quigley, "Mixing methodologies to enhance the implementation of healthcare operational research," *Journal of the Operational Research Society,* vol. 58, pp. 159-167, 2006.

[56] H. V. D. Parunak, R. Savit, and R. Riolo, "Agent-based modeling vs. equation-based modeling: A case study and users' guide," in *Multi-Agent Systems and Agent-Based Simulation*, 1998, pp. 277-283.

[57] N. J. Pioch, J. Melhuish, A. Seidel, E. Santos, D. Li, and M. Gorniak, "Adversarial intent modeling using embedded simulation and temporal Bayesian knowledge bases," in *Modeling and Simulation for Military Operations IV*, 2009, pp. 73480E-1.

[58] J. Duinhoven - Van Steenbergen, "Emergence in Multi-Agent Systems," 2012.

[59] F. Jochen. (2005). *Types and Forms of Emergence*. Available: http://arxiv.org/abs/nlin/0506028

[60] D. C. Lane, "You just don't understand me: Modes of failure and success in the discourse between system dynamics and discrete event simulation," ed: London School of Economics, 2000.

[61] A. Sweetser, "A comparison of system dynamics (SD) and discrete event simulation (DES)," in *17th International Conference of the System Dynamics Society*, 1999, pp. 20-23.

[62] R. Axelrod, *Comparing modeling methodologies*: University of Michigan, Gerald R. Ford School of Public Policy, 2004.

[63] B. Behdani, "Evaluation of paradigms for modeling supply chains as complex sociotechnical systems," presented at the Winter Simulation Conference 2012.

[64] E. Waltz, "Situation analysis and collaborative planning for complex operations," in *13th ICCRTS: C2 for Complex Endeavors*, 2008.

[65] D. C. Lane, "With a little help from our friends: how system dynamics and soft OR can learn from each other," *System Dynamics Review,* vol. 10, pp. 101-134, 1994.

[66] R. G. Sargent, "A historical view of hybrid simulation/analytic models," presented at the Winter Simulation Conference, 1994.

[67] D. Helbing, "Introduction: The FuturICT knowledge accelerator towards a more resilient and sustainable future," *European Physical Journal Special Topics,* vol. 214, pp. 5-9, 2012.

[68] R. Bhavnani and H. J. Choi, "Modeling civil violence in Afghanistan: Ethnic geography, control, and collaboration," *Complexity,* vol. 17, pp. 42-51, 2012.

[69] B. S. S. Onggo, "Simulation modeling in the social care sector: A literature review," presented at the Winter Simulation Conference, 2012.

[70] J. M. Epstein, *Generative social science: Studies in agent-based computational modeling*: Princeton University Press, 2007.

[71] J. W. Creswell, A. C. Klassen, V. L. Plano Clark, and K. Smith Clegg, "Best practices for mixed methods research in the health sciences," 2011.

[72] M. D. Fetters, L. A. Curry, and J. W. Creswell, "Achieving Integration in Mixed Methods Designs—Principles and Practices," *Health Services Research,* vol. 48, pp. 2134-2156, 2013.

[73] A. Tashakkori and C. Teddlie, *Mixed methodology: Combining qualitative and quantitative approaches* vol. 46. Thousand Oaks: Sage Publications, Inc., 1998.

[74] A. Tashakkori and C. Teddlie, "Quality of inferences in mixed methods research: Calling for an integrative framework," in *Advances in mixed methods research*, M. M. Bergman, Ed., ed Thousand Oaks: Sage Publications Inc, 2008, pp. 101-119.

[75] D. Helbing, "Pluralistic modeling of complex systems," *Science and Culture,* vol. 76, July 18, 2010 2010.

[76] M. Balaban, C. M. Banks, and J. A. Sokolowski, "Vietnam 1969-1973: Engaging M&S to characterize cause and effect patterns of US withdrawal," *SpringSim 12 : Emerging Applications of M&S in Industry and Academia (EAIA'12) Symposium,* 2012.

[77] S. Mathison, "Why triangulate?," *Educational researcher,* vol. 17, pp. 13-17, 1988.

[78] N. K. Denzin, "The research act in sociology: A theoretical introduction to sociological methods," 1970.

[79] M. Hofmann, "Ontologies in modeling and simulation: An epistemological perspective," in *Ontology, Epistemology, and Teleology for Modeling and Simulation*, A. Tolk, Ed., ed: Springer, 2013, pp. 59-87.

[80] C. Partridge, A. Mitchell, and S. de Cesare, "Guidelines for developing ontological architectures in modelling and simulation," in *Ontology, Epistemology, and Teleology for Modeling and Simulation*, A. Tolk, Ed., ed: Springer, 2013, pp. 27-57.

[81] T. Honderich, *The Oxford companion to philosophy*. Oxford, New York: Oxford University Press, 1995.

[82] L. Rabelo, M. Helal, A. Jones, and H. S. Min, "Enterprise simulation: A hybrid system approach," *International Journal of Computer Integrated Manufacturing,* vol. 18, pp. 498-508, 2005.

[83] R. H. Martin and D. Raffo, "A model of the software development process using both continuous and discrete models," *Software Process: Improvement & Practice,* vol. 5, pp. 147-157, 2000.

[84] K. Choi, D. Bae, and T. Kim, "An approach to a hybrid software process simulation using the DEVS formalism," *Software Process: Improvement & Practice,* vol. 11, pp. 373-383, 2006.

[85] T. Levin and I. Levin, "Integrating hybrid modelling with system dynamics," presented at the Proceedings of the 21st System Dynamics Society International Conference, 2003.

[86] N. Osgood, "Using traditional and agent based toolsets for system dynamics: present tradeoffs and future evolution," *Proceedings of the 25th International Conference of the System Dynamics Society,* July 29 - August 2 2007.

[87] T. A. Henzinger, "The theory of hybrid automata," *Verification of Digital and Hybrid Systems,* vol. 170, pp. 265-292, 2000.

[88] S. Rossiter and K. R. W. Bell, "A workflow hybrid as a multi-model, multi-paradigm simulation framework," presented at the ESM 2010: The 2010 European Simulation and Modelling Conference, 2010.

[89] S. Setamanit, W. W. Wakeland, and D. Raffo, "Using simulation to evaluate global software development task allocation strategies," *Software Process: Improvement & Practice,* vol. 12, pp. 491-503, 2007.

[90] J. Venkateswaran and Y. J. Son, "Hybrid system dynamic—discrete event simulation-based architecture for hierarchical production planning," *International Journal of Production Research,* vol. 43, pp. 4397-4429, 2005.

[91] W. W. Wakeland, R. H. Martin, and D. Raffo, "Using design of experiments, sensitivity analysis, and hybrid simulation to evaluate changes to a software development process: a case study," *Software Process: Improvement & Practice,* vol. 9, pp. 107-119, 2004.

[92] F. Peña-Mora, S. Han, S. Lee, and M. Park, "Strategic-operational construction management: Hybrid system dynamics and discrete event approach," *Journal of Construction Engineering & Management,* vol. 134, pp. 701-710, 2008.

[93] L. Rabelo, H. Eskandari, T. Shaalan, and M. Helal, "Value chain analysis using hybrid simulation and AHP," *International Journal of Production Economics,* vol. 105, pp. 536-547, 2007.

[94] Y. Lee, M. Cho, and Y. Kim, "A Discrete-Continuous Combined Modeling Approach for Supply Chain Simulation," *SIMULATION,* vol. 78, pp. 321-329, 2002.

[95] B. P. Zeigler, H. Praehofer, and T. G. Kim, *Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems*. San Diego, California: Academic Press, 2000.

[96] P. A. Fishwick, *Simulation model design and execution: building digital worlds*. Upper Saddle River, New Jersey: Prentice Hall, 1995.

[97] L. B. Holm, F. A. Dahl, and M. Barra, "Towards a multimethodology in health care–synergies between Soft Systems Methodology and Discrete Event Simulation," *Health Systems,* vol. 2, pp. 11-23, 2012.

[98] L. Baoding and L. Yian-Kui, "Expected value of fuzzy variable and fuzzy expected value models," *IEEE Transactions on Fuzzy Systems,* vol. 10, pp. 445-450, 2002.

[99] E. Z. Baskent and S. Keles, "Spatial forest planning: A review," *Ecological Modelling,* vol. 188, pp. 145-173, 2005.

[100] S. Birle, M. A. Hussein, and T. Becker, "Fuzzy logic control and soft sensing applications in food and beverage processes," *Food Control,* vol. 29, pp. 254-269, 2013.

[101] K. Chahal, H. Daly, T. Eldabi, and R. J. Paul, "Developments in modeling organizational issues in healthcare: Multi method modeling," in *Handbook of Research on Advances in Health Informatics and Electronic Healthcare Applications: Global Adoption and Impact of Information Communication Technologies*, K. Khoumbati, A. Srivastava, Y. K. Dwivedi, and B. Lal, Eds., ed Hershey, New York: IGI Global, 2009, pp. 190-204.

[102] B. D'Ambrosio, "A hybrid approach to reasoning under uncertainty," *International Journal of Approximate Reasoning,* vol. 2, pp. 29-45, 1988.

[103] P. Donzelli and G. Iazeolla, "A hybrid software process simulation model," *Software Process: Improvement & Practice,* vol. 6, pp. 97-109, 2001.

[104] K. Kotiadis and J. Mingers, "Combining PSMs with Hard OR Methods: The Philosophical and Practical Challenges," *The Journal of the Operational Research Society,* vol. 57, pp. 856-867, 2006.

[105] T. Lorenz and A. Jost, "Towards an orientation framework in multi-paradigm modeling," in *Conference Proceedings of the System Dynamics Society*, Nijmegen, The Netherlands, 2006, pp. 23-27.

[106] Y. Mao, "Use multiple modeling approaches to study sustained online communities " Master of Science, Department of Computer Science, University of Saskatchewan, Saskatoon, Canada, 2008.

[107] N. Mustafee and E. E. Bischoff, "A multi-methodology agent-based approach for container loading," presented at the Winter Simulation Conference, 2011.

[108] A. Tolk, "M&S body of knowledge: Progress report and look ahead," *SCS M&S Magazine,* vol. 1, 2010.

[109] H. Vangheluwe, J. De Lara, and P. J. Mosterman, "An introduction to multi-paradigm modelling and simulation," presented at the Proceedings of the 2002 AI, Simulation and Planning, Lisbon, 2002.

[110] H. E. Eriksson, M. Penker, B. Lyons, and D. Fado, *UML 2 Toolkit* vol. 26: John Wiley & Sons, 2003.

[111] N. Schieritz and P. M. Milling, "Modeling the forest or modeling the trees," in *21st International Conference of the System Dynamics Society*, 2003, pp. 20-24.

[112] J. Finnigan, "The science of complex systems," *Australian science,* vol. 26, pp. 32-34, 2005.

[113] S. Robinson, "Conceptual modelling for simulation Part II: A framework for conceptual modelling," *Journal of the Operational Research Society,* vol. 59, pp. 291-304, 2007.

[114] C. Robson, *Real world research: A resource for social scientists and practitioner-researchers* vol. 2. Oxford: Blackwell 2002.

[115] S. Harding, "Rethinking standpoint epistemology: What is "strong objectivity"," in *Feminist epistemologies*, L. Alcoff and E. Potter, Eds., ed New York and London: Routledge 1993, pp. 49-82.

[116] K. R. Howe and E. House, *Values in evaluation and social research.* Thousand Oaks, CA: Sage Publications, Inc., 1999.

[117] P. E. Tetlock, "Psychology and politics: The challenges of integrating levels of analysis in social science," in *Social Psychology: Handbook of basic principles*, A. W. Kruglanski and E. T. Higgins, Eds., 2nd ed New York, NY: Guilford Press, 2007, pp. 888-912.

[118] A. J. Onwuegbuzie and R. B. Johnson, "The validity issue in mixed research," *Research in the Schools,* vol. 13, pp. 48-63, 2006.

[119] D. C. Gross, "Report from the fidelity implementation study group," in *Fall Simulation Interoperability Workshop Papers*, 1999.

[120] JCGM, "OIML 2007 International vocabulary of metrology—basic and general concepts and associated terms (VIM)." in *Geneva ISO/IEC Guide 99: 2007, and JCGM 200: 2008*, ed, 2008.

[121] S. Harmon and S. M. Youngblood, "A proposed model for simulation validation process maturity," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology,* vol. 2, pp. 179-190, 2005.

[122] T. Grüne-Yanoff, "The explanatory potential of artificial societies," *Synthese,* vol. 169, pp. 539-555, 2009.

[123] O. Balci, "Verification, validation, and testing," in *Handbook of Simulation.* vol. 10, J. Banks, Ed., ed, 1998, pp. 335-393.

[124] S. Rossiter, J. Noble, and K. R. W. Bell, "Social Simulations: Improving Interdisciplinary Understanding of Scientific Positioning and Validity," *Journal of Artificial Societies and Social Simulation,* vol. 13, 2010.

[125] B. McKelvey, "Model-centered organization science epistemology," in *Companion to organizations*, J. A. C. Baum, Ed., ed Thousand Oaks, CA: Sage, 2002, pp. 752-780.

[126] D. Wahyuni, "The research design maze: Understanding paradigms, cases, methods and methodologies," *Journal of Applied Management Accounting Research,* vol. 10, pp. 69-80, 2012.

[127] P. Checkland, *Systems thinking, systems practice*. New York: John Wiley & Sons, 1981.

[128] M. Balaban, P. Hester, and S. Y. Diallo, "Towards a theory of multi-method M&S approach: Part I," presented at the Winter Simulation Conference, Savannah, GA, 2014.

[129] Merriam-Webster.com. (2014, 24 Nov. 2014). *"Method"*. Available: http://www.merriam-webster.com/dictionary/method

[130] DoD, "5000.61," in *M&S VV&A*, ed, 1996.

[131] M. Kaku, *The future of the mind: The scientific quest to understand, enhance, and empower the mind*: Random House LLC, 2014.

[132] DoD, "M&S Human Capital Strategy," 2007.

[133] DoD, "5000.59, M&S Management," ed: January, 1998.

[134] A. Turing, "Intelligent machinery (1948)," in *The Essential Turning: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life: Plus The Secrets of Enigma*, B. J. Copeland, Ed., ed New York: Oxford University Press, 2004, pp. 395-432.

[135] E. W. Weisel, M. D. Petty, and R. R. Mielke, "Validity of models and classes of models in semantic composability," presented at the Proceedings of the Fall 2003 Simulation Interoperability Workshop, 2003.

[136] E. W. Weisel, M. D. Petty, and R. R. Mielke, "A comparison of DEVS and semantic composability theory," presented at the Proceedings of the Spring 2005 Simulation Interoperability Workshop, San Diego CA, 2005.

[137] M. D. Fetters, L. A. Curry, and J. W. Creswell, "Achieving Integration in Mixed Methods Designs-Principles and Practices," *Health Services Research,* vol. 48, pp. 2134-2156, 2013.

[138] T. S. Kuhn, *The structure of scientific revolutions*: University of Chicago press, 2012.

[139] O. Ghrayeb, P. Damodaran, and P. Vohra, "Art of triangulation: An effective assessment validation strategy," *Global Journal of Engineering Education,* vol. 13, pp. 96-101, 2011.

[140] T. S. Kuhn, "Commensurability, Comparability, Communicability," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association,* vol. 1982, pp. 669-688, 1982.

[141] A. Borshchev and A. Filippov, "From system dynamics and discrete event to practical agent based modeling: Reasons, techniques, tools," in *Proceedings of the 22nd International Conference of the System Dynamics Society*, 2004.

[142] G. Calanni Fraccone, R. Valenzuela-Vega, S. Siddique, V. Volovoi, and A. Kirlik, "Nested Modeling of Hazards in the National Airspace System," *Journal of Aircraft,* vol. 50, pp. 370-377, 2013.

[143] A. P. Sage and C. D. Cuppan, "On the systems engineering and management of systems of systems and federations of systems," *Information, Knowledge, Systems Management,* vol. 2, pp. 325-345, 2001.

[144] M. Balaban and T. W. Mastaglio, "RoPax/RoRo: Exploring the use of simulation as decision support system," in *Oceans 13*, San Diego, CA, 2013.

[145] T. I. Ören, "Modeling and simulation: A comprehensive and integrative view," *Agent-directed simulation and systems engineering,* pp. 3-36, 2009.

[146] S. Robinson, "Conceptual modelling for simulation Part I: Definition and requirements," *Journal of the Operational Research Society,* vol. 59, pp. 278-290, 2007.

[147] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, pp. 7280-7287, 2002.

[148] J. Cain, *Planning improvements in natural resources management* vol. 124, 2001.

[149] R. P. Kelton, R. P. Sadowski, and D. T. Sturrock, *Simulation with Arena*: New York: McGraw Hill, 2007.

[150] B. Y. Márquez, M. Castañon-Puga, J. R. Castro, and E. Suarez, "Methodology for the modeling of complex social system using neuro-fuzzy and distributed agencies," *Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Software Engineering (JSSE),* 2011.

[151] S. Brailsford, E. Silverman, S. Rossiter, J. Bijak, R. Shaw, J. Viana, *et al.*, "Complex systems modeling for supply and demand in health and social care," presented at the Winter Simulation Conference, 2011.

[152] J. Kennedy, G. Gimm, and E. Blodgett, "Return to work: A critical aspect of care coordination for younger dual eligibles," *Disability and Health Journal,* vol. 6, pp. 95-99, 4// 2013.

[153] M. Balaban, "Return to work behavior of people with disabilities: A multi-method approach," presented at the Winter Simulation Conference, Savannah, GA, 2014.

[154] G. Livermore, D. C. Stapleton, and M. O'Toole, "Health care costs are a key driver of growth in federal and state assistance to working-age people with disabilities," *Health Affairs,* vol. 30, pp. 1664-1672, 2011.

[155] SSA, "Annual statistical report on the Social Security Disability Insurance Program," SSA, Baltimore, MDJuly 2012 2011.

[156] J. Sherk. (2013, 3/23/2014). *Not looking for work: Why labor force participation has fallen during the recession.* Available: http://www.heritage.org/research/reports/2013/09/not-looking-for-work-why-labor-force-participation-has-fallen-during-the-recession

[157] M. Duggan, "The growth in the social security disability rolls: A fiscal crisis unfolding," National Bureau of Economic Research, Cambridge, MA Working Paper 12436, 2006.

[158] S. Liu and D. Stapleton, "How many SSDI beneficiaries leave the rolls for work? More than you might think," *Benefits,* vol. 60, p. 10.3, 2010.

[159] J. Kregel, "Work incentives planning and assistance program: Current program results document the program's ability to improve employment outcomes, reduce dependence on benefits, and generate cost savings for SSA," *Journal of Vocational Rehabilitation,* vol. 36, pp. 3-12, 2012.

[160] G. Livermore, S. Prenovitz, and J. Schimmel, "Employment-related outcomes of a recent cohort of work incentives planning and assistance (WIPA) program enrollees," *Mathematica Policy Research. Washington D.C.,* 2011.

[161] K. C. Thomas and A. R. Ellis, "Patterns of healthcare use and employment among people with disabilities," *Disability and Health Journal,* vol. 6, pp. 133-140, 4// 2013.

[162] J. P. Hall, N. K. Kurth, and S. L. Hunt, "Employment as a health determinant for working-age, dually-eligible people with disabilities," *Disability and Health Journal,* vol. 6, pp. 100-106, 4// 2013.

[163] Y. Ben-Shalom and A. Mamun, "Return-to-work outcomes among Social Security Disability Insurance Program beneficiaries," Princeton, New JerseyJune, 21, 2013 2013.

[164] D. C. Stapleton, S. Liu, D. Phelps, and S. Prenovitz, "Longitudinal statistics on work activity and use of employment supports for new Social Security Disability Insurance beneficiaries," *Social Security Bulletin,* vol. 71, 2011.

[165] S. Nadkarni and P. P. Shenoy, "A causal mapping approach to constructing Bayesian networks," *Decision Support Systems,* vol. 38, pp. 259-281, 2004.

[166] Z. Sun and D. Müller, "A framework for modeling payments for ecosystem services with agent-based models, Bayesian belief networks and opinion dynamics models," *Environmental Modelling & Software,* 2012.

[167] P. Kasaie, D. Dowdy, and D. Kelton, "An agent-based simulation of a tuberculosis epidemic: Understanding the timing of transmission," presented at the Winter Simulation Conference, Washinton D.C., 2013.

[168] P. Patlolla, V. Gunupudi, A. R. Mikler, and R. T. Jacob, "Agent-based simulation tools in computational epidemiology," in *Innovative Internet Community Systems,* T. Böhme, V. M. Larios Rosillo, H. Unger, and H. Unger, Eds., ed Berlin Heidelberg: Springer, 2006, pp. 212-223.

[169] A. Djanatliev and R. German, "Prospective healthcare decision-making by combined system dynamics, discrete-event and agent-based simulation," presented at the Winter Simulation Conference, 2013.

[170] M. Bayes and M. Price, "An essay towards solving a problem in the doctrine of chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S," *Philosophical Transactions (1683-1775),* vol. 53, pp. 370-418, 1763.

[171] S. M. Berry, B. P. Carlin, J. J. Lee, and P. Muller, *Bayesian adaptive methods for clinical trials*. Boca Raton, FL: CRC Press, 2010.

[172] D. J. Spiegelhalter, K. R. Abrams, and J. P. Myles, *Bayesian approaches to clinical trials and health-care evaluation* vol. 13. West Sussex, England: Wiley, 2004.

[173] M. Fishbein and I. Ajzen, *Belief, attitude, intention and behavior: an introduction to theory and research.* Boston, MA: Addison-Wesley Pub, 1975.

[174] J. M. Garcia, *Theory and practical exercises of system dynamics*: MIT Sloan School of Management, 2006.

APPENDICES

APPENDIX A

SECONDARY DEFINITIONS

A passive entity is an entity that flows through predefined system structure, without individuality that would allow them acting according to internally generated behavior.

Technical behavior is an observation engineered by humans or animals that abstracts away active or proactive behavior.

Active behavior is a change-oriented and self-initiated behavior in situations that involves acting rather than just reacting to it.

Proactive behavior is an anticipatory, change-oriented, and self-initiated behavior in situations that involves acting in advance of a future situation, rather than just reacting to it.

A social event phenomenon is an observation produced because of active or proactive behavior(s) of entity as an individual, in groups, or as a group.

APPENDIX B

EXERCISE TABLES FOR COMMENSURABILITY OF METHODS

Table 1. Commensurability of SD with six other methods for given criteria.

Criteria/Method	SD/DES	SD/ABM	SD/BN	SD/FM	SD/SC	SD/PN
Representation of individual behavior as part of a larger system	0.5	0.75	0.25	0.25	0.25	0.25
Ability to operate on aggregates	1	0.25	0	0.5	0.5	0.5
Ability to handle uncertainty	0.75	0.5	0.5	0.5	0.5	0.5
Interaction	0.75	1	0	0	0.5	0.75
Descriptive usage	0.75	0.25	0.25	0.25	0.5	0.5
Theoretical usage	0.75	0	0.25	0.25	0.5	0.5
Emergence	0	0.75	0.25	0.25	0.25	0.25
Ability to represent active behavior	0	1	0	0	0.75	0.75
Sum	4.5	4.5	1.5	2	3.75	4
Average	**0.56**	**0.56**	**0.19**	**0.25**	**0.47**	**0.5**

Table 2. Commensurability of ABM with six other methods for given criteria.

Criteria/Method	ABM/SD	ABM/DES	ABM/BN	ABM/FN	ABM/SC	ABM/PN
Representation of individual behavior as part of a larger system	0.75	0.25	1	1	1	0.5
Ability to operate on aggregates	0.25	0.75	0.25	0.25	0.25	0.25
Ability to handle uncertainty	0.5	0.25	0	0	0	0
Interaction	1	0.25	1	1	0.5	0.25
Descriptive usage	0.25	0.5	0	0	0.25	0.25
Theoretical usage	0	0.75	0.25	0.25	0.5	0.5
Emergence	0.75	0.75	1	1	1	0.5
Ability to represent active behavior	1	1	1	1	0.25	0.25
Sum	4.5	4.5	4.5	4.5	3.75	2.5
Average	**0.56**	**0.56**	**0.56**	**0.56**	**0.47**	**0.31**

Table 3. Commensurability of BN with six other methods for given criteria.

Criteria/Method	BN/ABM	BN/SD	BN/DES	BN/FM	BN/SC	BN/PN
Representation of individual behavior as part of a larger system	1	0.25	0.75	0	0	0.5
Ability to operate on aggregates	0.25	0	1	0.5	0.5	0.5
Ability to handle uncertainty	0	0.5	0.25	0	0	0
Interaction	1	0	0.75	0	0.5	0.75
Descriptive usage	0	0.25	0.5	0	0.25	0.25
theoretical usage	0.25	0.25	0.5	0	0.25	0.25
Emergence	1	0.25	0.25	0	0	0.5
Ability to represent active behavior	1	0	0	0	0.75	0.75
Sum	4.5	1.5	4	0.5	2.25	3.5
Average	**0.56**	**0.19**	**0.50**	**0.06**	**0.28**	**0.4375**

Table 4. Commensurability of FM with six other methods for given criteria.

Criteria/Method	FM/BN	FM/ABM	FM/SD	FM/DES	FM/SC	FM/PN
Representation of individual behavior as part of a larger system	0	1	0.25	0.75	0	0.5
Ability to operate on aggregates	0.5	0.25	0.5	0.5	0	0
Ability to handle uncertainty	0	0	0.5	0.25	0	0
Interaction	0	1	0	0.75	0.5	0.75
Descriptive usage	0	0	0.25	0.5	0.25	0.25
Theoretical usage	0	0.25	0.25	0.5	0.25	0.25
Emergence	0	1	0.25	0.25	0	0.5
Ability to represent active behavior	0	1	0	0	0.75	0.75
Sum	0.5	4.5	2	3.5	1.75	3
Average	**0.06**	**0.56**	**0.25**	**0.44**	**0.22**	**0.375**

Table 5. Commensurability of SC with six other methods for given criteria.

Criteria/Method	SC/FM	SC/BN	SC/ABM	SC/SD	SC/DES	SC/PN
Representation of individual behavior as part of a larger system	0	0	1	0.25	0.75	0.5
Ability to operate on aggregates	0	0.5	0.25	0.5	0.5	0
Ability to handle uncertainty	0	0	0	0.5	0.25	0
Interaction	0.5	0.5	0.5	0.5	0.25	0.25
Descriptive usage	0.25	0.25	0.25	0.5	0.25	0
Theoretical usage	0.25	0.25	0.5	0.5	0.25	0
Emergence	0	0	1	0.25	0.25	0.5
Ability to represent active behavior	0.75	0.75	0.25	0.75	0.75	0
Sum	1.75	2.25	3.75	3.75	3.25	1.25
Average	**0.22**	**0.28**	**0.47**	**0.47**	**0.41**	**0.16**

Table 6. Commensurability of PN with six other methods for given criteria.

Criteria/Method	PN/SC	PN/FM	PN/BN	PN/ABM	PN/SD	PN/DES
Representation of individual behavior as part of a larger system	0.5	0.5	0.5	0.5	0.25	0.25
Ability to operate on aggregates	0	0	0.5	0.25	0.5	0.5
Ability to handle uncertainty	0	0	0	0	0.5	0.25
Interaction	0.25	0.75	0.75	0.25	0.75	0
Descriptive usage	0	0.25	0.25	0.25	0.5	0.25
Theoretical usage	0	0.25	0.25	0.5	0.5	0.25
Emergence	0.5	0.5	0.5	0.5	0.25	0.25
Ability to represent active behavior	0	0.75	0.75	0.25	0.75	0.75
Sum	1.25	3	3.5	2.5	4	2.5
Average	**0.16**	**0.38**	**0.44**	**0.31**	**0.50**	**0.31**

APPENDIX C

METHODS USED IN CASE STUDY

ABM

Agent Based Modeling (ABM) is a bottom-up method that aims at capturing interactions by using computer created entities called "agents". These individual agents are assigned attributes, states, rules of behaviors, and often interactions. ABM is more suitable in representing complexity arising from individual behavior and interactions [1]. ABM can facilitate insider views into an agent, as well as observer based-views on overall group emerging characteristics. Model development can consists of developing agent's attributes, behavioral rules, memory specification, decision-making capability, adjustment behavior properties, supporting data, relationships in form of methods between agents and environment.

ABM implemented in AnyLogic® is implemented as array list or linked hash set of agents. The macro view of this implementation is discrete. Agent can have both discrete and continuous internal representation, which is determined by methods used (see MFVI in Figure 11). In AnyLogic® ABM can be easily enhanced with other methods as its internal behavior e.g. SC, SD, or DES but should be considered as a separate method since they do not need ABM to produce simulation models outside scope of ABM. Moreover, ABM on its own can produce simulation models by implementing agent behavior rules using Java code.

ABM in healthcare is often used to simulate epidemiological phenomena [167, 168], but also can be used to more operational setting, for instance, to test medical innovations or interventions [169].

BN

A Bayesian method is based on the principles of Bayes' Theorem by Thomas Bayes (1702-1761), where a probability is represented as the likelihood that a statement is true, given the prior information [170]. The Bayesian method is extensively used in the healthcare setting for design and inference of clinical trials [171], and healthcare evaluations [172]. TheBayesian method can be useful in drawing inferences with a quantified degree of confidence, based on some prior known evidence. This method has evolved significantly e.g. into the Bayesian Network (BN) method, providing easy-to-use tools with graphical interfaces allowing to quickly develop BNs which are widely utilized within many scientific, business and government communities. BN can be used in a simulation model to capture behavior intentions at a psychological level; for instance, based on a theory of reasoned action [173] as proposed by Balaban and Mastaglio [144]. BN can be used as a modeling method (see Definition 8) or as an M&S method (see Definition 10), which would be at that point considered a Dynamic Bayesian Network.

SD

The System Dynamics method consists of a feedback loops in form of differential equations that provide for building relations between variables. It is useful for studying complex nonlinear systems, especially finding cause and effect relationships. Models built with this methodology can help framing issues and problems, revealing dynamics related to change imposed on the system. Models are typically used to show trends of relationships and not always precisely computing specific values. This method can be an invaluable tool in assessing big picture of a problem, testing alternative policies and strategies at the governance or enterprise level [174]. SD model development usually involves creating a causal diagram. Causal diagram represents elements of system and relations between them in form of links that end with arrows indicating what influences what. Positive or negative relationship are marked as "+" or "-" on the line. A plus symbol means that a second variable follows the direction of change of the first variable, while the negative means that a direction is opposite to the direction of change. At least two links are needed to form a feedback loop, which can be positive or negative as well. Feedback loops define type of feedback behaviors, positive loops (plus sign or R for reinforcement) propels the change and negative feedback (minus sign or B for balancing) is the cause for stability in the system. During model development, a flow diagram is created out of a causal diagram. From the practical perspective on how to build a model, one should know about so-called stocks or levels, flows, auxiliary variables (dynamic variables), and constants as the main blocks providing for metaphors. Levels are used to model accumulation and depend on function of flows. Auxiliaries and constants are parameters that serve as additional information needed to specify flows, which in turn affect levels. Creating model structure is done by connecting blocks and arrows to form desirable relationships.

DES

In a discrete event simulation model, changes can occur only at separate points in time. These changes are called events, and everything in the model is related to them in one way or another. DES model consists of entities flowing through designed by modeler process. Events are stored in a calendar, which contain information that allows model to be executed in accordance to its logic. The central idea of DES is that variables of the model will not change between successive events. In addition, important to understand element of DES are queues. They are just like lines in the store, and can be define depending on the system's real queue, which may have limitations for number of elements that can fit into it, and has different rules reflecting priority of leaving it by stored elements e.g. FIFO means first in first out and LIFO meaning last in first out. Other essential components of DES are resources, which may be personnel, equipment etc. They are used by entities while going through process. DES is often used to capture stochastic behavior of the system but can model deterministic events as well. DES is usually capturing anything that can be described in a processing way, and often is used at operational and detailed level.

SC

SC is an implemented version of UML-based state chart diagram within AnyLogic®. SC consists of states (simple, composite, final, and history), transitions, initial point, and branch objects. Transitions are triggered by defined conditions, messages, timeouts, or rates. Important feature of SC is ability to represent hierarchy of states e.g. where a composite state consists of one or more states. SC represents discreet events, and its implementation within AnyLogic® provides ability to define deterministic and probability based time transitions between states that if used within ABM allow representing stochastic behavior similarly to DES. Please see Borshchev [31] for a in-depth introduction to SC.

APPENDIX D

CONDITIONAL PROBABILITY TABLES

Possible incentives

health improvement programs	sufficient				insufficient			
money incentives	sufficinet		insufficient		sufficinet		insufficient	
vocational assistance	sufficient	insufficient	sufficient	insufficient	sufficient	insufficient	sufficient	insufficient
high	0.7	0.4	0.4	0.2	0.4	0.2	0.2	0
medium	0.25	0.4	0.4	0.4	0.4	0.4	0.4	0.05
low	0.05	0.2	0.2	0.4	0.2	0.4	0.4	0.95

Adoption of patient role

possible incentives	high				medium				low			
confidence to prove required disability level	confident		notConfident		confident		notConfident		confident		notConfident	
fear of loosing benefits	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
high	0.6	0.5	0.5	0.4	0.8	0.7	0.65	0.6	0.95	0.9	0.8	0.5
low	0.4	0.5	0.5	0.6	0.2	0.3	0.35	0.4	0.05	0.1	0.2	0.5

Attitude toward work

adoption of pa...	high	low
enthusiastic	0.3	0.7
pesimistic	0.7	0.3

Level of competitiveness

disability level	high				medium				low			
perception ab...	enthusiastic		pesimistic		enthusiastic		pesimistic		enthusiastic		pesimistic	
vocational assi...	sufficient	insufficient	sufficient	insufficient	sufficient	insufficient	sufficient	insufficient	sufficient	insufficient	sufficient	insufficient
high	0.4	0.3	0.1	0.01	0.45	0.4	0.2	0.05	0.6	0.4	0.25	0.1
medium	0.3	0.4	0.3	0.3	0.35	0.3	0.4	0.35	0.3	0.45	0.45	0.4
low	0.3	0.3	0.6	0.69	0.2	0.3	0.4	0.6	0.1	0.15	0.3	0.5

Improvement of health

health improve...	sufficient			insufficient		
original disabilit...	High	Medium	Low	High	Medium	Low
high	0.2	0.3	0.4	0.01	0.1	0.15
medium	0.3	0.4	0.5	0.1	0.15	0.2
low	0.5	0.3	0.1	0.89	0.75	0.65

Disability level

original disability level	High			Medium			Low		
improvement of health	high	medium	low	high	medium	low	high	medium	low
high	0.3	0.5	0.8	0	0	0	0	0	0
medium	0.6	0.3	0.15	0.5	0.65	0.8	0	0	0
low	0.1	0.2	0.05	0.5	0.35	0.2	1	1	1

Confidence to prove required disability level

disability level	high	medium	low
confident	0.85	0.6	0.2
notConfident	0.15	0.4	0.8

Fear of losing benefits

confidence to prove required disability level	confident				notConfident			
current financial benefits	available		notAvailable		available		notAvailable	
current health benefits	available	notAvailable	available	notAvailable	available	notAvailable	available	notAvailable
High	0.6	0.5	0.55	0.3	0.99	0.8	0.8	0.3
Low	0.4	0.5	0.45	0.7	0.01	0.2	0.2	0.7

VITA

Mariusz Balaban received his two Masters in Mechanical Engineering from Lublin Technical University, Poland. He is an M&S Scientist/Developer at MYMIC. His current work focus is M&S area including R&D, training and maintenance of simulation models and software. Before joining MYMIC Team, he spent three years conducting research, writing papers, and developing simulation models at ODU in the area of M&S, particularly related to multi-method M&S approach, innovation, R&D governance, global events modeling, social systems, transportation and logistics, healthcare, system of systems, and enterprise decision support systems. Previously he was an active member of US Navy. Beside his M&S based research interests, Mr. Balaban has strong awareness and interest of new IT technologies. Mr. Balaban is a member of SCS, ACM SIG, MTS, and GK International. He is an author and coauthor of multiple papers in the area of M&S.