Journal of the Association for Information Systems (2019) **20**(6)**,** 735-786
doi: 10.17705/1jais.00550

RESEARCH PAPER

ISSN 1536-9323

Information Systems as Representations: A Review of the Theory and Evidence

Jan Recker[1], Marta Indulska[2], Peter F. Green[3], Andrew Burton-Jones[4], Ron Weber[5]

[1]University of Cologne, Germany / Queensland University of Technology, Australia, jan.recker@wiso.uni-koeln.de
[2]University of Queensland, Australia, m.indulska@business.uq.edu.au
[3]Queensland University of Technology, Australia, p.green@qut.edu.au
[4]University of Queensland, Australia, abj@business.uq.edu.au
[5]Monash University, Australia / University of Queensland, Australia, ron.weber@monash.edu.au

Abstract

Representation theory proposes that the basic purpose of an information system (IS) is to faithfully represent certain real-world phenomena, allowing users to reason about these phenomena more cost-effectively than if they were observed directly. Over the past three decades, the theory has underpinned much research on conceptual modeling in IS analysis and design and, increasingly, research on other IS phenomena such as data quality, system alignment, IS security, and system use. The original theory has also inspired further development of its core premises and advances in methodological guidelines to improve its use and evaluation. Nonetheless, the theory has attracted repeated criticisms regarding its validity, relevance, usefulness, and robustness. Given the burgeoning literature on the theory over time, both positive and negative, the time is ripe for a narrative, developmental review. We review representation theory, examine how it has been used, and critically evaluate its contributions and limitations. Based on our findings, we articulate a set of recommendations for improving its application, development, testing, and evaluation.

Keywords: Representation Theory, Representation Model, State-Tracking Model, Good-Decomposition Model, Information Systems Analysis and Design, Conceptual Modeling, Grammar Expressiveness, Ontology, Literature Review.

Roger Chiang was the accepting senior editor. This research article was submitted on May 22, 2017, and went through two revisions.

1 Introduction

One of the few theories consistently named as native to the information systems (IS) field is *representation theory* (RT) (Gregor, 2006; Straub, 2012). RT posits that the essential purpose of an IS is to provide a faithful representation of some focal real-world phenomena, thereby assisting its users to track states and state changes (events) in the phenomena it represents (e.g., Wand & Weber, 1988, 1990b, 1990a, 1993, 1995; Weber, 1987). In this way, an IS obviates the need for its users to follow the often more costly alternative of directly tracking the focal real-world phenomena themselves (Weber, 2003).

We have had an enduring engagement with RT. Initially, we focused on using it to account for IS analysis and design phenomena. Subsequently, we noted and sometimes participated in RT's deployment across increasingly diverse, sometimes unanticipated areas—for instance, IS use, enterprise systems fit, IS security, and human resource management (Burton-Jones & Grange, 2013; Sia & Soh, 2007; Strohmeier &

Röhrs, 2017; Strong & Volkoff, 2010; Thomas & Dhillon, 2012). We also witnessed and sometimes took part in discourse about criticisms of RT (e.g., Allen & March, 2012; Lyytinen, 2006; Shanks & Weber, 2012; Wyssusek, 2006).

Some time ago, we concluded that research on RT was reaching an impasse. We saw a growing tension between its longstanding and more recent varied use within the IS field and several escalating, sometimes vehement, disputes about its value. For some protagonists in the debates, our view was their positions had become entrenched. We decided, therefore, to review the research done on RT in the hope we could find ways to help resolve the impasse and inform further debates about RT. Overall, we wanted to be in a better position to advise colleagues and students about whether they should engage with RT and, if so, how they should do so. We also wanted to know whether we should persist with our own work on RT.

When we began our review work in 2013, we were aware of only one other review of RT that had been done—namely, Saghafi and Wand's (2014) meta-analysis of the results of 12 laboratory experiments that had been conducted to test some of RT's predictions about the ability of users to understand different types of conceptual models. This review informed but did not cover the broader objectives we had contemplated for our review of RT—for instance, to determine its scope and application, to identify the outcomes it had produced, to examine the research approaches used to test it, and to evaluate the likely merits of continuing with it.

In light of our broad objectives, we decided to embark upon a *narrative, developmental* review of research on RT (Templier & Paré, 2015, pp. 118-119). This type of review summarizes published research on a topic and seeks to provide new directions for the research. As our review work unfolded, however, we realized we first needed to answer a specific question—namely, has research on RT shown it to be a success or a failure? If RT had been debunked, continuing with a narrative, developmental review seemed pointless. We concluded, therefore, that we first needed to undertake another type of review—namely, a *focused, thematic, cumulative* review (Templier & Paré, 2015, p. 120). Such a review draws overall conclusions about a *specific* research question (Arksey & O'Malley, 2005)—in our case, whether research on RT has succeeded or failed. Thus, we pursued this objective first.

Upon completing the focused, thematic, cumulative review, we concluded that a clear verdict on the success of RT could not yet be reached (Burton-Jones, Recker, Indulska, Green, & Weber, 2017). We found the research done to date shows signs of success as well as failure, but both types of research remain too limited to reach clear-cut conclusions. Moreover, neither work on pursuit of success nor work on pursuit of failure has proceeded systematically enough. Consequently, the results obtained are often piecemeal, disjointed, and conflicting.

Given the inconclusive outcome of our focused, thematic, cumulative review, we felt the question of how RT research should continue had become even more important. We were also mindful of the concentrated scope of our first review—from an initial pool of several hundred papers[1] that engaged with RT in some way, we focused on 69 papers only. Moreover, during the course of writing the review, research on RT had continued. Many new papers have been published since our initial review work in 2013.

Accordingly, we decided to continue work with our original focus on a narrative, developmental review to complement and extend the findings in Saghafi and Wand (2014) and Burton-Jones et al. (2017). Using several criteria employed to distinguish between various types of reviews, (Paré, Trudel, Jaana, & Kitsiou, 2015; Rowe, 2014; Templier & Paré, 2015; Webster & Watson, 2002), Table 1 shows how our current review differs from those in Saghafi and Wand (2014) and Burton-Jones et al. (2017). Specifically, we expanded the scope of our work to appraise *all* literature on RT (from its inception until 2016) and not just (1) selected experimental studies on RT (from 2001 through 2012) (Saghafi & Wand, 2014), or (2) papers that engaged only with RT's success or failure (from its inception until 2013) (Burton-Jones et al., 2017). Also, we changed how we reviewed the literature from a concept-centric approach (Webster & Watson, 2002) to a paper-centric approach (e.g., Vessey, Ramesh, & Glass, 2002).

---

[1] We use the term "paper" generically to cover journal publications, book chapters, conference papers, monographs, etc.

Table 1. Position of this Paper in Relation to Other Reviews of RT

Criteria / **Paper**	**Saghafi & Wand (2014)**	**Burton-Jones et al. (2017)**	**This Review**
Type and Aim of Review	*Aggregative:* synthesize available experimental evidence through meta-analysis	*Cumulative:* draw specific conclusions to support further theory development	*Narrative:* summarize previously published research to identify substantial future research opportunities and inquiries
Scope	*Experimental* papers about user understanding of conceptual models (n = 12)	*Selected* papers focused on the success and failure of RT (n=69)	*All* papers (n = 365)
Timeframe	2001-2012	From inception until 2013	From inception until 2016
Approach	*Statistical, quantitative* review of reported experimental results	*Concept-centric, theoretical* review of success and failure	*Paper-centric, summative* classification of papers to show their development, by application domains, methods, evidence, and critiques

For the narrative, developmental review that is the focus of this paper, we chose five goals:

- examine the scope of RT as manifested in those areas where it has been applied;
- determine the quantity and quality of the empirical evidence in support of RT;
- identify the extent of theoretical and methodological advances made to RT;
- understand and classify the critiques made of RT;
- pinpoint opportunities for further research on RT.

Combined with our focused, thematic, cumulative review (Burton-Jones et al., 2017), we believe our narrative, developmental review provides a foundation for improved programs of research on RT—hopefully, programs that will lead to clearer, more compelling, and more innovative outcomes and to a conclusive verdict about RT's merits and deficiencies.

Our paper proceeds as follows. First, we provide a brief summary of RT. Second, we describe and explain how we identified papers that reference RT, the ways we coded them, and the basis we used to choose those papers included in our review. Third, we examine RT from the perspectives of areas in which it has been applied, theoretical and methodological advances that have been made, critiques of its theory and methods, and opportunities for further research. Finally, we summarize our findings, discuss the limitations of our review, and present some brief conclusions.

2 Brief Summary of RT

The primary motivation for the initial work on RT was to derive a theory to account for the fundamental nature of high-quality IS (e.g., Wand & Weber, 1988, 1990b, 1990a, 1990c, 1995; Weber, 1987). RT has evolved since it was first proposed in the late 1980s and early 1990s. Thus, the theory's components are scattered across multiple publications (Wand & Weber, 1988, 1990b, 1995; Weber, 1987, 1997). As a result, different publications usually provide only a partial rather than a complete view of RT.

In Burton-Jones et al. (2017), we present our understanding of RT as it currently stands, reflect briefly on its origins and development, and examine some refinements that have occurred to its core concepts and premises. We also provide a description of RT in Appendix A. Note that we do not provide dense citations in Appendix A to show the historical evolution of RT. Rather, we seek to lay a foundation to better understand subsequent sections of our paper where we reach conclusions about RT based on our literature review. Thus, we cite only a few key papers in our explanation of RT.

As we explain in Burton-Jones et al. (2017, pp. 1309-1310), RT's primary focus is the *deep structure* of an IS—those characteristics of an IS that manifest the meaning (as perceived by stakeholders) of the real-world phenomena it is intended to represent (Wand & Weber, 1995, pp. 205-207). RT's primary concern is *the extent to which the deep structure of an IS provides and remains a faithful representation of the meaning of the focal real-world phenomena* (Wand & Weber, 1995, p. 207). In seeking to understand how the deep structure of an IS might provide and remain a faithful representation, Wand and Weber developed three models.

1. *Representation model* (RM): Seeks to account for the ability of IS grammars to generate scripts that are faithful representations of the focal real-world phenomena.

2. *State-tracking model* (STM): Imposes a set of conditions on scripts that have been used to enact (make operational) an IS if the IS is to remain a faithful representation of the focal real-world phenomena as the things in the real world undergo change.
3. *Good-decomposition model* (GDM): Imposes a set of conditions on the scripts used to develop and implement an IS if the scripts are to communicate more meaning about the focal real-world phenomena (and, by extension, facilitate stakeholders assessing the faithfulness of the representation).

Note that the RM, STM, and GDM each focus on somewhat different but related aspects of an IS's deep structure. Together, however, they are intended to provide both a theory for explanation and prediction (Gregor, 2006, p. 628) and a theory for design and action (Gregor, 2006, pp. 628-629). They provide an *explanation* of why certain properties of an IS script either enhance or detract from its deep-structure representational fidelity. Based on these properties, they provide *predictions* about whether an IS actualized via the script will be deemed useful. The properties also provide guidance for *design and action*. They indicate how IS scripts must be prepared to preserve deep-structure representational fidelity, thereby enhancing the usefulness of the IS they actualize.

3 Literature Identification and Coding

Our review of the literature that references the RM, STM, and GDM drew on several established approaches (Paré et al., 2015; Rowe, 2014; Vessey et al., 2002; Webster & Watson, 2002). Based on these approaches, we decided to proceed in three steps: (1) identify relevant literature via citation analysis, (2) develop a coding scheme to categorize the literature, and (3) analyze the literature within each category (Vessey et al., 2002).

In the subsections below, we describe how we carried out each step. Our goal was to achieve "systematicity" in relation to our review—in other words, "reproducibility through documenting the search process and potentially indicat[ing] comprehensiveness" (Rowe, 2014, p. 246).

3.1 Literature Identification

To identify relevant literature, we first debated which publications about the RM, STM, and GDM were seminal until we reached consensus. The agreed-upon set comprised three journal papers: Wand and Weber (1990b, 1993, 1995). We then used Harzing's (2010) Publish or Perish tool to retrieve the lists of citations to these three papers.

By July 2013, we had obtained 1,022 records of papers citing the three papers we deemed seminal. Because we were interested in whether works using the RM, STM, and GDM have been useful to other researchers, we removed all records of citing papers that were not themselves cited (i.e., records of papers that prima facie had not made any impact). This action resulted in a list of 770 records, which still included duplicate records for those papers citing more than one of the three seminal works. Removal of duplicates (53 records), as well as records of papers not written in English (42 records), resulted in a list of 675 records.

As a next step, we considered how to account for the impact of papers published recently, versus those published some time ago. By consensus, we decided papers published over five years ago (i.e., before 2009) required a minimum of 10 citations (at least two citations per year) to be deemed impactful. In contrast, other than the original requirement of at least one citation, we did not place a minimum on citations for papers published during or after 2009 (because these papers have not had the same exposure as the pre-2009 papers).

Through analysis of citation numbers of the citing papers, we eliminated 202 records of papers published prior to 2009 (because they had less than 10 citations). This process resulted in 473 records. To ensure that this elimination round was not excessive, we reviewed the 202 records based on our familiarity with the research and the researchers. We felt some papers with a citation count of less than 10 still might have a long-run impact. Through a voting process involving all authors in which a paper required at least one vote to be retained, 29 of the 202 deleted records were returned to the list. This step resulted in 502 records of potentially relevant papers.

The papers corresponding to the 502 records were then downloaded from various publication databases and the World Wide Web. In some cases, we requested copies of papers directly from their authors (for difficult-to-find papers). Ultimately, we were unable to obtain copies of six papers (these papers were not available online, and the authors did not respond to email requests for copies of their papers). We included the remaining 496 papers in our analysis.

We processed all papers for optical character recognition (OCR) to enable full-text search capabilities. These capabilities are important because they allowed us to reduce the number of papers requiring full analysis by identifying those that referenced the seminal works for purposes not relevant to our study (e.g., to support definitions of "ontology" but not to use the three models substantively, such as employing them analytically or testing them empirically).

The search term "Weber" was identified as the best term to identify the number of relevant references in each paper, as well as the context of related discussion.

Using this term with a full-text search, we identified 162 papers that cited the seminal works for purposes unrelated to our review. We then analyzed these papers to confirm their lack of relevance. The final set comprised 334 papers. Each was stored and linked to its EndNote record.

At this stage, our analysis of the 334 papers proceeded in the direction of a focused, thematic, cumulative review (Templier & Paré, 2015, p. 120). The aim was to address the question of whether research had shown RT to be a success or failure. This work resulted in a separate paper (Burton-Jones et al., 2017). As our work on this paper neared completion in January 2016, we returned to our original review purposes. We then embarked on a second phase of our study to identify more recent papers that cited the three seminal publications. Following our earlier processes, we identified 124 new papers published since the first phase of our study. Four could not be obtained. Also, recall that in the first phase of our study we excluded papers that had been published in the last five years but had fewer than 10 citations by 2013. Thirty-two of these papers now had enough citations to meet our inclusion criteria. Thus, we had 152 additional papers to examine. We again searched each paper using the term "Weber". Three of us also independently read each paper to determine whether its use of RT was substantive or cursory. We identified 59 new papers that required coding; 28 were removed subsequently because they were deemed irrelevant for the purposes of our review (see Subsection 3.3 below).

In the second phase of our study, we also noted that 36 papers published in 2009 or 2010 and initially included in our first-phase analysis had now been available for over five years but did not have 10 citations. Nonetheless, because we had considered them in our earlier analysis, we decided not to exclude them. Therefore, our analysis is based on 365 papers.

3.2 Coding Scheme

To understand the full extent of the applications, tests, advances, and critiques of the RM, STM, and GDM, we developed a coding scheme. We used dimensions typically employed in literature reviews, such as research approach (Vessey et al., 2002), research method (Chen & Hirschheim, 2004), research topic (Galliers & Whitley, 2007), application domain and focal element (Wand & Weber, 2002), concept or premise investigated (Recker, 2011), and quality of evidence (GRADE Working Group, 2004). We felt a broad classification scheme would help us better understand the work conducted on RT and enable us to further review particular categories of literature (e.g., application domains or empirical studies).

Our coding scheme started with a basic set of codes related to the purposes of our paper. It then evolved over three rounds of pilot tests. During each pilot test, four of us coded a subset of randomly selected papers. We then reviewed our coding, addressed inconsistencies, and reflected on whether our coding scheme was sufficient to meet the goals of our study. This process highlighted the importance of having code definitions and examples of a code's use to ensure all coders had a consistent understanding of the codes. It also resulted in removal, addition, and modification of several codes. After the third pilot test, we were satisfied our coding scheme allowed us to address our review goals.

To ensure we adequately covered the scope of RT's use, we developed codes to capture whether a study applied the RM, STM, or GDM and its focus. We also included a code to capture how a study used RT—whether it formed the conceptual foundation, whether it was tested, whether it was extended, whether it was the focus of a critique or used in other ways, and which of its premises were examined. We included another code for methodological advances to identify research that had enhanced approaches to applying RT.

For conceptual-modeling studies, we used additional codes to capture whether the study focused on data modeling, object-oriented modeling, or process modeling, and whether the unit of analysis was a method, grammar, or script (or something else) (Wand & Weber, 2002). We also included a "conceptual-modeling approach" code, which allowed us to record the conceptual-modeling grammar studied.

To ensure we had sufficient data to determine the quantity and quality of empirical evidence, we included codes to capture the research method used in a study (e.g., conceptual analysis, survey, experiment, field study, case study, interviews, design science), details of any empirical work (e.g., number of student and/or practitioner participants), and the outcomes of the study. Finally, a code that focused on application areas allowed us to determine the main research domains in which RT had been applied (as well as outlier applications).

The final coding scheme (Table B1) was then used to analyze/code literature identified in both phases of the study. We conducted first-phase coding in 2013 and second-phase coding in late 2016 and early 2017. We had a different set of coders in the two phases (see below).

3.3 Literature Coding

In the first phase of our literature analysis, a postdoctoral researcher familiar with the RM, STM, GDM, and much of the literature (he had completed his PhD using RT) coded all 334 papers during 2013 and 2014. Before he commenced coding, we explained the coding scheme to him during several meetings. We then conducted two pilot tests to evaluate the quality of his coding. The pilot tests used the sets of papers we had employed earlier to undertake the initial pilot tests of the coding scheme. After each pilot test, we compared our coding with the postdoctoral researcher's coding. Where inconsistencies arose, we clarified the coding scheme with him until we were confident he understood it thoroughly.

The coding process took over four months to complete. It produced a spreadsheet of 334 coded papers based on a full-text reading. It also resulted in the identification of some papers that the coder indicated were not relevant to our study (e.g., citing the seminal papers but using the RM simply as an example of an ontology). One of the authors subsequently reviewed these papers to determine whether they were indeed irrelevant for our purposes. As a result, 22 papers were removed, which resulted in a final set of 312 papers.

To ascertain the reliability of the coding, we employed a second coder to code a random sample of 52 papers (just over 20%). This coder was also a postdoctoral researcher familiar with the three models and much of the literature on RT. We followed the same preparation process that we used with the first coder.

Finally, we compared the two sets of coding and calculated Cohen's (1960) Kappa as a measure of intercoder reliability. Kappa understates agreement when a specific coding category is more prevalent than others—a problem that occurred in two of the 17 coding categories (where Kappa values were < 0.05). In these cases, the intercoder agreements were > 92%, which indicates high agreement between the coders. When these two categories were excluded, the average Kappa was 0.69, indicating adequate levels of agreement.[2]

In the second phase of our study, the postdoctoral researcher was no longer available; therefore, one of us read and coded all 59 papers. The papers and their assigned codes were then divided equally among the remaining authors to read and verify the initial coding. During this process, six papers were identified as not relevant to our study (on the same grounds as in the first phase), resulting in a total of 53 new papers. Where coding discrepancies arose (19 out of 53 papers—35%), the two codes and the paper were given to a third author, who then made a judgment in consultation with a fourth author about the most appropriate code. In most cases, the discrepancies were minor omissions (e.g., missing classification of a design science paper).[3] Combining the two phases resulted in a full coding of 365 papers (312 from the first phase and 53 from the second phase).

4 Applications of RT

In our first analysis, we examined the domains where RT had been applied. We wished to determine the scope of RT as evidenced by its use. We classified papers into application domains by examining their full text. Our coding scheme in Table B1 shows that not all papers relevant to our study can be considered as applications of the three models. For example, theory critiques and methodological advances are relevant, but they might not be applications of the RM, STM, or GDM. For purposes of identifying application domains, such papers were excluded. Accordingly, we removed 158 papers from our overall set, which left 207 papers to be classified into application domains.

To identify application domains, we used two iterative steps. Because we are familiar with much of the RT literature, we knew its main application domain was conceptual modeling, especially data and process modeling. Hence, our initial analysis of application domains used only three codes—namely, data modeling applications, process modeling applications, and "other" applications. The results obtained revealed several themes in the "other" category. Accordingly, we refined the coding scheme to include two additional application domains—namely, "object-oriented modeling" and "ontology".

Using the updated codes, we then conducted a second round of coding. Because most papers we reviewed had a singular application domain focus, we decided to assign all papers to one application domain code only. We also concluded that little information about applications domains would be lost by following the simpler coding approach of using one domain only. Where a paper covered more than one domain (e.g., object-oriented modeling and process modeling), the main application domain was determined by reading the paper and making a judgment about its main focus. In this round of coding, each paper classified as "other" was analyzed based on its stated goals to determine its application domain.

Table 2 shows our overall results. Most (66%) of the 207 papers used the RM in one of four domains—namely, data modeling, object-oriented (OO) modeling, process modeling, and ontology (Table 2). Other domains in which the RM has been applied include IS security (e.g., Thomas & Dhillon, 2012), business-IT alignment (e.g., Singh & Woo, 2009), and data quality (e.g., Ram & Liu, 2007). The STM and GDM have received less attention than the RM. Across different application domains, we found only 12 instances of their use (Table 2). Three papers we examined (Thomas & Dhillon, 2012; Tollington & Spinelli, 2012; Wand & Weber,

---

[2] No specific pattern of disagreement among the coders was evident.

[3] We did not calculate *kappa* because only verification of one author's coding occurred (not independent coding).

1995) applied concepts from all three models, resulting in overlap in Table 2.

Figure 1 shows RT's application domains over time. It was applied initially to data modeling. Several somewhat isolated applications occurred in other areas—for example, facet modeling (Opdahl, 1998; Opdahl & Sindre, 1997) and data quality (Wand & Wang, 1996). OO modeling also received attention, initially through isolated applications, and then with an increased focus from 1999 onwards (peak occurring in 2005). The introduction of the ontology application domain, which generally focused on the development or comparison of ontologies, started in 2001. This domain has had an ongoing focus (peak occurring in 2009). Process modeling saw several early applications, peaked in 2007, and has continued to be a focus. More recently, the application of RT to other domains has increased.

Table 2. RM, GDM, and STM Application Areas

	RM	**STM**	**GDM**
Data modeling	24	1	1
Ontology	27	0	0
OO modeling	28	0	2
Process modeling	53	0	1
Other	68	3	4
Grand Total	**200**	**4**	**8**

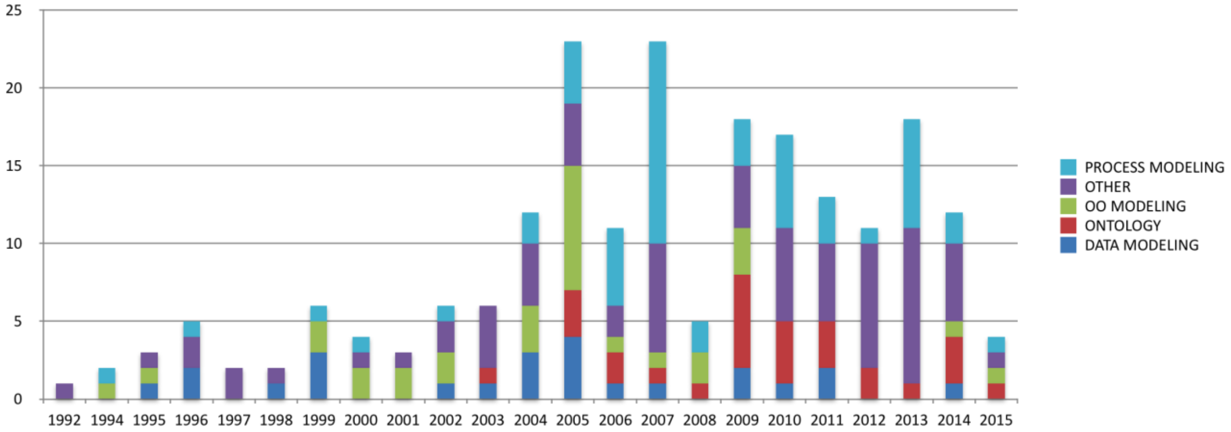


Figure 1. Frequency and Type of Application by Year

4.1 Applications of the Representation Model

The early focus on data modeling followed Wand and Weber's (1993) use of the RM to evaluate the ontological completeness of the ER modeling grammar. For example, Weber and Zhang (1996) used the RM to evaluate the NIAM grammar, and Gregersen and Jensen (1999) used the RM to evaluate three ER grammar extensions (ERT, TERC+, and TimeER). Following this initial emphasis on evaluating the ontological completeness of modeling grammars as a whole, subsequent work focused on specific features of the grammars. Some examples are studies examining the effect of representing relationships with attributes (Burton-Jones & Weber, 1999), use of mandatory versus optional properties (Bodart, Patel, Sim, & Weber, 2001; Gemino & Wand, 2005), alternative representation of things and properties (Shanks, Nuredini, Moody, Tobin, & Weber, 2003; Shanks, Nuredini, & Weber, 2005), and differences between state-based and event-based representations of real-world phenomena (Allen & March, 2006b). More recent work has explored the extent to which domain knowledge moderates the effects of ontological clarity on users' understanding of EER models (Bera, Burton-Jones, & Wand, 2014).

Soon after the RM was proposed, a sustained focus on its use to evaluate OO modeling commenced (Figure 1). Some examples are evaluations of OO grammars overall, such as LOOPN++ (Keen & Lakos, 1994, 1996) and OML and UML (Opdahl & Henderson-Sellers, 1999, 2001, 2002), and analyses of particular features of OO grammars, such as part-whole relationships (Opdahl, Henderson-Sellers, & Barbier, 2001). More recent research has used the RM to study modeling of part-whole relations in UML class diagrams (Shanks, Tansley, Nuredini, Tobin, & Weber, 2008), conceptual-modeling rules for UML grammars (Evermann & Wand, 2005), the ontological premises that underlie and the effects of using UML association classes (Bera & Evermann, 2014), and how UML's Statechart notation might be extended to better cover security phenomena (El-Attar, Luqman, Kárpáti, Sindre, & Opdahl, 2015).

Almost a decade elapsed after the RM was introduced before it was used to evaluate process-modeling phenomena. Whereas only a few data-modeling grammars had been examined by the end of the 1990s, during the 2000s many process-modeling grammars were studied (e.g., Green & Rosemann, 2000; Green, Rosemann, & Indulska, 2005; Recker & Indulska, 2007; Recker, Indulska, Rosemann, & Green, 2006a). The RM also allowed reasoning about the relative complexity of process-modeling grammars (e.g., Recker, zur Muehlen, Siau, Erickson, & Indulska, 2009; zur Muehlen, Recker, & Indulska, 2007). While this research was mainly conceptual, significant empirical work was also done (e.g., Davies, Rosemann, & Green, 2004; Recker, Indulska et al., 2006; Recker, Indulska, Rosemann, & Green, 2010).

The purely analytical application of the RM for the purposes of understanding the strengths and weaknesses of process-modeling grammars was extended subsequently to the analysis of reference models (Fettke & Loos, 2005), the conceptualization of goals in process models (Soffer & Wand, 2004, 2005, 2007), and reasoning about the complementarity of process and business-rule modeling grammars (zur Muehlen & Indulska, 2010). Potential transformations between different process-modeling grammars were also examined (Indulska, Recker, Green, & Rosemann, 2007; Meertens, Iacob, & Eckarts, 2010).

A spate of research applied the RM to develop or extend process-modeling grammars. For example: Heidari, Loucopoulos, Brazier, and Barjis (2013) considered several existing modeling notations and created an abstraction of these notations to provide a metamodel that can be used to compare process-modeling notations and to develop additional notations; Singer (2014) argued an evaluation of a grammar's ontological completeness, together with design principles for cognitively effective visual notations, are the building blocks of business process-modeling notations; and Altuhhova, Matulevičius, and Ahmed (2013) proposed an extension to business process model and notation (BPMN) to enable modeling of IS security risk management (ISSRM) phenomena.

A decade after the RM was proposed, Figure 2 shows it was used to develop and evaluate ontologies. Davies, Green, Milton, and Rosemann (2003, 2005) were the first to compare ontologies (specifically, the RM and Chisholm's ontology) via their underlying metamodels. Subsequently, Kruchten, Woo, Monu, and Sotoodeh (2008) evaluated an ontology they had proposed for disasters, Goumopoulos and Kameas (2008) developed an ontology for ambient ecologies, Colomb and Ahmad (2010) developed an ontology for interlocking institutional worlds, Opdahl and colleagues (2011, 2012) developed the unified enterprise modeling ontology (UEMO), Tegarden, Schaupp, and Dull (2013) evaluated the resource-event-agent (REA) enterprise ontology, Avédissian, Valverde, and Barrad (2015) evaluated the agent language lab ontology, and Ahmad and Odeh (2013) evaluated the EIAOnt (enterprise information architecture ontology).

While most applications of the RM have occurred in data modeling, process modeling, OO modeling, and ontology development, the RM has also been applied elsewhere (Figure 3). Of the 68 papers that applied the RM outside the four main areas, 11 focused on other modeling approaches. For example, Matulevicius, Heymans, and Opdahl (2007) compared and evaluated two goal-oriented modeling grammars (GRL and KAOS), Fettke and Loos (2007) evaluated Scheer's reference model for production planning and control systems, and Kwon (2011) developed a method for representing decision makers' knowledge of causality.

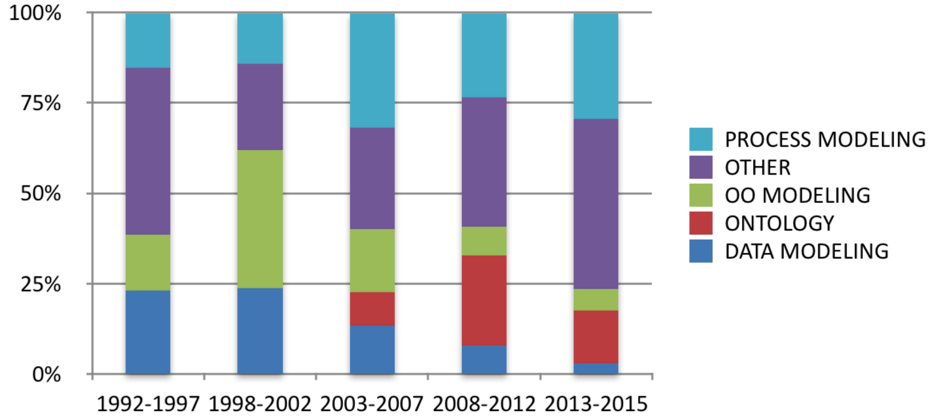


Figure 2. Types of Applications per Time Period

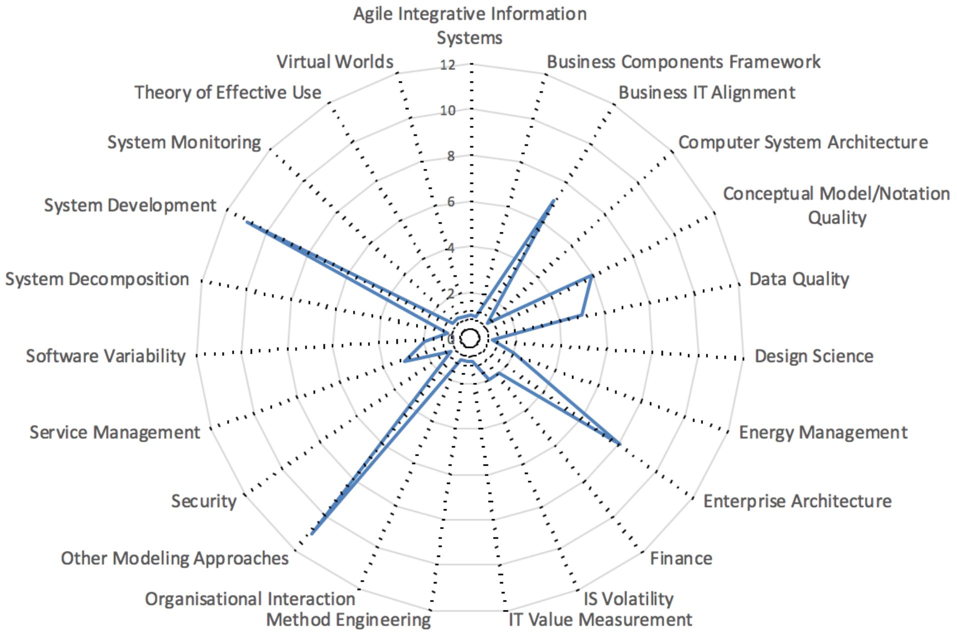


Figure 3. Areas of RM Application Outside the Main Areas

Some researchers have used the RM to develop guidelines for assessing the quality of models and grammars. For example: Rockwell and Bajaj (2004) developed a framework for evaluating the effectiveness, efficiency, and readability of conceptual models; Nelson, Poels, Genero, and Piattini (2012) combined the RM and Lindland, Sindre, and Sølvberg's (1994) framework to assimilate modeling quality evaluations from product and process perspectives; and Krogstie (2012) developed guidelines for evaluating the quality of modeling notations.

In the domain of systems development, the RM has been deployed in several ways. For example: Rohde (1995) evaluated Jackson's (1983) methodology; Karow, Gehlert, Becker, and Esswein (2006) transformed models of real-world perceptions to software designs; Rittgen (2006) mapped the dynamic essential modeling of organizations (DEMO) notation to UML; Bernaert (2010) analyzed requirements engineering phases, with a specific focus on the *event* construct; and Reinhartz-Berger, Sturm, and Wand (2013) formalized the concept of systems behavior.

By using the RM in the domain of business-IT alignment, Etien and Rolland (2005) developed metrics of alignment between business and system models, Rosemann, Vessey, and Weber (2004) proposed the notion of ontological distance between organizational requirements and existing system capabilities, Singh and Woo (2009) developed a goal-based framework for business-IT alignment, and Strong and Volkoff (2010) extended Wand and Weber's conceptualization of IS structure to rethink the nature of the IT artifact.

The RM has also provided a foundation for research on data quality. For example: Wand and Wang (1996) examined data-quality dimensions from an ontological perspective; Parsons and Wand (2003) reconciled differences at the attribute level when data are supplied from separate sources; Ram and Liu (2007) analyzed the semantics of provenance; and Lukyanenko, Parsons, and Wiersma (2014b) studied the relationship between conceptual models and information quality.

Less frequently, the RM has been applied in domains such as service-delivery architecture definitions (O'Brien & Burmeister, 2003; Tziallas & Theodoulidis, 2003), services description in services management (Kazemzadeh & Milton, 2015), modeling of autonomic computing systems (Tziallas & Theodoulidis, 2003), and monitoring of engineering phenomena (Allmark, Grosvenor, Byrne, Anayi, & Prickett, 2013). Wand and Weber's distinction between an IS's surface structure, deep structure, and physical structure has underpinned studies on IS volatility (Heales, 2002) and virtual worlds (Chaturvedi, Dolk, & Drnevich, 2011).

4.2 Applications of the State-Tracking Model

We identified only four applications of the STM in the literature we reviewed. Wand and Weber (1995) evaluated the ER grammar. They concluded ER modelers would have difficulty satisfying the STM's four conditions (Appendix A3) because the ER grammar lacks constructs that allow high-fidelity scripts to be generated. Thomas and Dhillon (2012) used the STM, in concert with the RM and GDM, to analyze the deep structure of IS security. Their focus was to ensure the completeness and fidelity of an IS security model they developed. Relative to the RM and GDM, Tollington and Spinelli (2012) concluded the STM is most useful in the financial-reporting domain because it can be employed to track transaction modifications. Reinhartz-Berger et al. (2013) proposed an RT-based approach with STM constructs to compare software systems functionality. For novice software developers, they found empirically the RT-based approach was significantly faster than a competing approach.

4.3 Applications of the Good-Decomposition Model

We identified six applications of the GDM in the literature we reviewed. Paulson and Wand (1992) argued the GDM, as originally proposed, lacked a means of operationalization. They provided heuristic rules to govern the search for good candidate decompositions, proposed a measure of complexity that allows candidate decompositions to be ranked, and described a method to automate the process of decomposing a system. To test their ideas, they developed and evaluated prototype software to undertake the decomposition process.

In the only application of the GDM to data modeling, Wand and Weber (1995) evaluated the ER grammar. Because the grammar inadequately represents dynamics, they concluded well-defined events cannot be distinguished from poorly defined events—a distinction required for generating good decompositions.

Burton-Jones and Meso (2002) focused on OO modeling. They conducted experiments with models that comply with and do not comply with the GDM's good-decomposition conditions. Consistent with the GDM's predictions, they concluded UML models that comply with the GDM's good-decomposition conditions are easier to understand. Nonetheless, they found compliance with these conditions had no effect on perceived ease of use of the UML models. Burton-Jones and Meso (2006, 2008) subsequently replicated these results.

Reijers, Mendling, and Dijkman (2011) used the GDM to study process model modularization. Their empirical research focused on the usefulness of decompositions in general, however, rather than undertaking a test of the GDM's five good-decomposition conditions (Appendix A4).

Thomas and Dhillon (2012) applied the GDM to the development of an IS security framework. They argued that good decompositions facilitate focused analyses of the security implications of each subsystem in a system and the identification of external events. They contend that better decompositions result in higher-quality security management approaches.

Tollington and Spinelli (2012) applied GDM guidelines to analyze the GDM's applicability in a financial context—in particular, financial reporting systems. When systems are structured based upon the five financial accounting elements of asset, liability, expense, income, and capital, they argued that all decompositions can be defined at the outset. Nonetheless, good decompositions are sometimes undermined because some financial terms are ambiguous.

4.4 Some Reflections

Many RT applications have been in the conceptual-modeling domain (e.g., data modeling, process modeling, OO modeling, and conceptual-modeling quality). This outcome is not surprising because conceptual models are representations of reality upon which IS are often understood and built. Nonetheless, researchers have extended RT's applications beyond these traditional domains (e.g., ontology development and business-IT alignment) (Figure 3). The breadth of these new domains indicates RT can be fruitful outside of traditional conceptual-modeling domains, especially if abstractions of real-world phenomena are important.

5 Empirical Support for RT

A second goal we had in reviewing and appraising RT was to determine whether it proved robust when subjected to empirical tests (Godfrey-Smith, 2003). If the answer is positive, researchers are likely to have more interest in RT. Conversely, if the answer is negative, researchers are likely to have less interest in RT. We undertook our evaluation of empirical support for RT from three perspectives. First, we examined the extent to which RT had undergone different *types* of *empirical tests*. Second, we examined the *types of participants* in the empirical tests. Third, we evaluated the *fit* of the evidence with theoretical predictions generated by RT. In the first three subsections below, we provide our findings in relation to each of these three perspectives. In the fourth subsection, we present some reflections.

5.1 Types of Empirical Tests

Seventy-three of the set of 365 papers (20%) reported some form of empirical work on RT. We classified these papers based on the research method they used:

- *Qualitative tests* (23 papers), consisting of *case studies* (e.g., Evermann & Wand, 2005; Strong & Volkoff, 2010), *expert panels* (Reinhartz-Berger, Itzik, & Wand, 2014), and *interviews* (e.g., Hadar & Soffer, 2006).
- *Quantitative tests* (45 papers), consisting of *surveys* (e.g., Recker, Rosemann, Green, & Indulska, 2011), *laboratory experiments* (e.g., Khatri, Vessey, Ramesh, Clay, & Sung-Jin, 2006), and *field experiments* (Lukyanenko, Parsons, & Wiersma, 2014a).
- *Hybrid tests* (five papers), consisting of *field studies* that combine qualitative and quantitative evidence (e.g., Chidamber & Kemerer, 1994) and designated *mixed-method studies* that combine different qualitative and/or quantitative research methods (e.g., Green, Rosemann, Indulska, & Recker, 2011).

Table B2 shows the types and number of empirical tests reported in papers published between 1994 and 2016.[4] Most used laboratory experiments (38 of 73 studies), followed by interviews and case studies (both 10). Four used field studies. Six used mixed-method designs: (1) survey and interviews (3 papers), (2) action research with experimentation and structured interviews (Moody & Shanks, 2003), (3) laboratory experiment with field experiment (Moody, 2003), and (4) survey and laboratory experiment (El-Attar et al., 2015).

5.2 Types of Participants

For two reasons, we examined the types of participants used in empirical tests of RT. First, when empirical tests draw samples from student populations, concerns sometimes arise about the extent to which knowledge claims can be generalized (Compeau, Marcolin, Kelley, & Higgins, 2012). Second, many existing RT-based studies examined how differences among participants impacted development and use of representations. For instance, some examined whether differences in prior domain and modeling knowledge among novices and experts or students and practitioners moderated the effects of a representation on users' perceptions of the representation (Burton-Jones & Meso, 2008; Gemino, 2004; Khatri et al., 2006).

Table B3 shows the types of participant cohorts used in empirical tests of RT and the mean sample sizes for these cohorts. Three types of participant cohorts have been used:

- *Practitioners*: Participants from industry who were systems analysts, consultants, or expert modelers (e.g., Milton, Rajapakse, & Weber, 2012; Recker et al., 2011).
- *Students*: Participants who were undergraduate/postgraduate students, often enrolled in business or IT degrees (Burton-Jones and Meso, 2008; Parsons, 2011).
- *Mixed*: Participants who were students and practitioners (e.g., Shanks, Moody, Nuredini, Tobin, & Weber, 2010; van Kleef, Noltes, & van der Spoel, 2010) or students and academics (e.g., Genero, Poels, & Piattini, 2008).

Table B3 shows that 27 of 45 quantitative studies (60%) used student participants, whereas 18 of 23 qualitative studies (78%) used practitioner participants. Seventy-six percent of all laboratory experiments (29 of 38) used students only. In contrast, all interview-based studies (10 in total) involved practitioners, either exclusively (six) or with students

---

[4] No empirical test of the theory was reported prior to 1994.

(four). Overall, just under 50% of empirical studies (36 of 73) involved practitioners.

The mean number of participants in case studies was 19.8. If Strong and Volkoff's (2010) large field study involving 72 interviewees is not considered, however, the average number of participants in case studies was small (6.8). The mean sample size for interviews was moderately large (31.7), but again this reflects one outlier (Soja & Paliwoda-Pękosz, 2013) involving 164 practitioner interviews. On average, the other nine interviews had 16.4 respondents. Mean sample sizes for surveys (401.7) and laboratory experiments (86.1) were reasonably large. Two studies classified as field studies (Chidamber & Kemerer, 1994; Nelson & Monarchi, 2007) did not report sample sizes, but the other two had a mean sample size of 31.5. The mixed-method studies had a mean sample size of 102.5 participants.

5.3 Fit of the Evidence

To determine how well the findings from empirical tests supported RT's predictions (as interpreted by the respective research teams), we identified 44 papers that offered *explicit* propositions (qualitative tests) or hypotheses (quantitative tests). In most of the 29 empirical papers that offered no explicit predictions, a qualitative test of the theory was reported (e.g., Patel, Sim, & Weber, 1998; Regev & Wegmann, 2004; Rittgen, 2006)—however, this situation was also the case for some quantitative tests (e.g., Moody, Sindre, Brasethvik, & Sølvberg, 2003; Reijers & Mendling, 2008).

To analyze the 44 papers with explicit predictions, we used the classification in Table 3 to assign the strength of the reported empirical support for the propositions or hypotheses.

Table 3. Coding Scheme to Evaluate Fit of the Evidence by Test Type

Qualitative Tests	**Quantitative Tests**
Proposition with apparent support (PS): Authors concluded a proposition received “good”, “apparent”, or “strong support”.	*Hypothesis supported (HS):* Tests showed correct effects directionality. Effect sizes were statistically significant.
Proposition with partial support (PP): Authors concluded a proposition received “some”, “limited”, or “partial support”.	*Hypotheses not supported (HN):* Effect sizes were not statistically significant.
Proposition with no support (PN): Authors concluded a proposition received no support or was refuted.	*Hypotheses refuted (HR):* Tests showed incorrect effects directionality. Effect sizes were statistically significant.
Proposition support inconclusive (PI): Impossible to determine from the paper how the proposition should be evaluated.	

The 44 papers contained 183 explicit predictions in total, with 42 propositions across five qualitative tests (mean = 8.4 propositions), 132 hypotheses across 37 quantitative tests (mean = 3.6 hypotheses), and nine propositions across two mixed-method tests. Across the quantitative tests, 78 of 132 hypotheses received support (59.1%), 17 received no support (12.9%), and 36 were refuted based on the data (27.3%). Across the qualitative tests, 19 of 42 propositions received apparent support (45.2%), 10 received partial support (23.8%), and 13 received no support (31.0%). In two papers (Recker, Indulska et al., 2006, 2010), one proposition each was impossible to evaluate. The tests in the two mixed-method studies resulted in support for five and no support for four propositions.

5.4 Some Reflections

One criticism of RT has been lack of empirical tests of its predictions (e.g., Allen & March, 2006a, p. 3). Our review shows otherwise. We found 35 of 60 (58%) of the empirical studies in our data set were published before 2007. We also found RT has been used to generate many empirically testable predictions—183 propositions or hypotheses across 44 papers. We further found RT has been robust. Our analysis of the fit of evidence shows more than half the propositions or hypotheses tested were supported (102 of 183). An additional 27 received partial support via qualitative data or were not refuted (but received no significant support) from quantitative tests. Fifty-three predictions (29.0%) were refuted.

Our review of empirical studies also shows researchers have been innovative in their approach to testing RT and its three constituent models. This situation is reflected in the variety of types of tests, from qualitative to quantitative to hybrid designs, and the breadth of reported evidence, from cohorts including students, practitioners, experts, and academics.

With these strong points in mind, we have four concerns about the state of empirical evaluations of

RT. First, the base of evidence is not extensive. In 20 years, 73 empirical tests of the RM, STM, or GDM have been reported. In contrast, meta-analyses and reviews of other research programs (e.g., technology acceptance or IS success) over a similar timeframe include hundreds of empirical studies (King & He, 2006; Petter, DeLone, & McLean, 2013).

Second, the outcomes of empirical tests of the RM, STM, and GDM remain somewhat inconclusive. On the one hand, most studies have used quantitative methods (primarily laboratory experiments). They often have stronger internal validity and statistical conclusion validity than qualitative studies. On the other hand, most quantitative studies have used student participants. Few have collected data from practitioners. Moreover, we identified only one that used experienced practitioners (expert data modelers, Milton et al., 2012). Thus, the external validity of the evidence obtained so far is limited (Compeau et al., 2012).

Third, as we noted in Subsection 5.3, 39.8% (29 of 73) of empirical studies done to test RT lacked explicit propositions or hypotheses. As the RM, STM, and GDM were refined and applied more widely over time, we expected more empirical research would have confirmed or falsified theory-based predictions and not been undertaken without a priori expectations. This outcome has not ensued; it is also mirrored by lack of a steady increase of empirical work over time (Table B2). Opportunities exist, therefore, for more rigorous, ongoing tests of RT.

Fourth, we have concerns about the *research designs* of some empirical tests reported. Table B4 summarizes these concerns, the resulting threats to the validity of results, and our recommendations for improving future empirical tests. For instance, we identified few articles that addressed instrument validity—an exception is Recker and Rosemann (2010). Without valid instruments, achieving high levels of internal and statistical conclusion validity has little purpose (Straub, Boudreau, & Gefen, 2004). Similarly, Table B4 shows we found only a few empirical studies with a high level of external validity. These were three cross-sectional surveys of modeling practitioners, each having a sample size of more than 100 respondents (Green et al., 2011; Recker et al., 2011; Soja & Paliwoda-Pękosz, 2013). We also noted that few replications of empirical work have occurred—an exception is Burton-Jones and Meso (2006).

6 Theoretical Advances to RT

Our third goal was to ascertain whether RT had been developed further since its original formulation. Thus, we sought to find papers that somehow advanced RT. In particular, we wanted to determine whether a paper proposed a substantive change of RT's components and/or one or more of its three constituent models.

6.1 Types of Theoretical Advances

We identified six papers providing theoretical advances to RT. Five relate primarily to the RM; one relates primarily to the GDM. We found none providing a theoretical advance to the STM.

6.1.1 Theoretical Advances to the Representation Model

Burton-Jones and Grange (2013) used RT to propose a model of effective use of IS. They redefined some of RT's constructs, proposed new constructs and associations, and expanded the boundary of phenomena covered by RT. For instance, they redefined *representation fidelity* "in terms of what users obtain from the system when using it" rather than "a property of the system alone". They argued an IS will not be deemed useful if users are unable to extract the representations it enacts because, for instance, they lack knowledge. Similarly, they defined the construct of *transparent interaction*, which is "the extent to which a user is accessing the system's representations unimpeded by its surface and physical structures". Users will not be able to understand the deep structure (meaning) of a representation if they cannot access it easily. In short, Burton-Jones and Grange (2013) proposed a more nuanced notion of representational fidelity than RT's initial notion.

Clarke, Burton-Jones, and Weber (2013) studied semantic quality in conceptual-modeling (CM) grammars. They argued a grammar's semantic quality is critical because it underpins how scripts are generated and how modeling rules and methods are devised. They contended that a complete assessment of the quality of a CM grammar's semantics must consider (1) its vocabulary, and (2) its production rules. Furthermore, because ontological analysis focused only on mappings between a grammar's constructs and a reference ontology, they argued, it achieved only a partial evaluation. They extended RT by using aspects of logical quality to show how information loss during the construction of scripts can be avoided.

Kiwelekar and Joshi (2013) used ontological categories to assign meaning to OO programming abstractions (specifically, the *class* construct). They showed how RM constructs could be reconciled with and could help clarify constructs in an OO programming language. In particular, they constructed classification rules that explicated and extended four constructs from the RM's ontology—thing, property, event, and process. In formulating the classification rules, they extended RT by incorporating the syntactic and implementation features of OO programming abstractions needed for the ontological interpretation. Similar to Clarke et al. (2013), they were concerned with the meaning of script elements when they are implemented.

Green (1997), Green, Rosemann, Indulska, and Manning (2007), and Green et al. (2011) extended the RM to cover evaluations of multiple grammars rather than a single grammar. They introduced two new theoretical constructs: maximum ontological completeness (MOC) and minimum ontological overlap (MOO). Their motivation was their observation that a single modeling grammar rarely, if ever, provides all constructs needed to model a domain. Thus, they predicted stakeholders will choose a *set* of grammars to represent a domain—those that, in combination, cover to the extent possible all constructs needed to model the domain (MOC) but also minimize the number of constructs that overlap across the grammars (MOO). They found empirical support for their predictions.

Strong and Volkoff (2010) undertook a three-year grounded-theory study of requirements misfits that arose when their case-study organization implemented an enterprise system (ES). While they found RT's three-structure conceptualization of an IT artifact (surface, deep, and physical structures) was useful in classifying the misfits they encountered, they concluded a fourth structure was needed. Specifically, they proposed a "latent-structure" construct to capture changes in organizational culture, controls, and roles that "emerge from and depend on" users' engagement with the other three structures (Strong & Volkoff, 2010, p. 752).

Recker, Rosemann, Green, and Indulska (2006) extended the RM by linking some of its constructs to the technology acceptance model (TAM) (Davis, 1986). Specifically, they argued that improvements in the ontological *clarity* of a conceptual-modeling grammar would lead to it being perceived as easier to use and improvements in the ontological *completeness* of a conceptual-modeling grammar would lead to it being perceived as more useful. Following TAM, they predicted that improvements in perceptions of ease of use and usefulness would lead to a higher intention to continue to use the grammar. They found empirical support for their predictions.

Rosemann, Vessey et al. (2004) proposed the construct of "ontological distance" as a measure of the fit between an ES's capabilities and a user organization's needs. User needs can be mapped to ontological constructs; similarly, the models embedded in an ES to support its capabilities can be mapped to ontological constructs. Ontological distance reflects how well instances of ontological constructs that underpin user needs can be mapped to instances of ontological constructs that underpin the models embedded in an ES. Different distance weights were assigned to different ontological constructs. For instance, mismatches between instances of "things" received higher weights than mismatches between instances of "intrinsic properties". Rosemann, Vessey et al. (2004, p. 446) argued the level of ontological distance can be used to predict the seriousness of problems that organizations will encounter when they implement an ES.

6.1.2 Theoretical Advances to the Good-Decomposition Model

Based on Bunge (1977, 1979) and Wand and Weber (1990b), Yang and Marquardt (2009) propose an ontology to facilitate multiscale systems modeling. Such models allow relationships between the values of properties of different things at different levels (and also the same level) in the level structure of a system to be examined. An important outcome of Yang and Marquardt's (2009) work is a deeper understanding of how the values of an emergent property of a thing relate to the values of the properties of the thing's components. In some ways, the focus of their work is on theoretical extensions to the reference ontology rather than theoretical extensions to the GDM. Nonetheless, we decided to classify their paper as a theoretical extension to the GDM because we suspect their ontological extensions will prove essential to future theoretical refinements of the GDM.

6.2 Some Reflections

We believe the theoretical advances made to RT are rich, innovative, and interesting. Some establish useful links between RT and other theories that are important in the IS field (e.g., TAM). Some enhance RT in ways that provide deep insights into phenomena that have been a major focus of IS research (e.g., effective use of IS and the fit between ES capabilities and user needs). Thus, the theoretical advances suggest RT has substantial external validity and can provide rich insights about IS phenomena. Nonetheless, in a research program spanning thirty years and hundreds of publications, we found only six papers that advanced RT in some way. As with our findings about applications and tests of RT, the theoretical advances focused primarily on the RM. None addressed the STM, and only one addressed the GDM.

Perhaps the lack of theoretical advances to RT indicates its relative robustness to scrutiny. Unless researchers systematically refine RT's high-level premises and the explanations offered by the RM, STM, and GDM, however, we suspect exploration of new application areas and designing and executing tests of RT to assess its validity will be inhibited.

7 Methodological Improvements in Enacting and Testing RT

Another of our goals was to determine whether methodological problems had surfaced as researchers attempted to enact and test the RM, STM, and GDM. If so, we sought to find out whether they had been mitigated. We discovered that substantive methodological improvements have occurred almost exclusively in the ways the RM has been enacted and tested. Little work had been done in relation to the STM and GDM. This finding mirrors our observations about the imbalances in applying, testing, and advancing RT.

In the subsections below, we first examine problems that motivated improvements in the ways the RM has been enacted and tested and some consequential methodological improvements that have been made. We then provide some reflections on these improvements.

7.1 Problems with Enacting and Testing the Representation Model and Consequential Methodological Improvements

The RM has often been *enacted* via a technique called “ontological analysis”, which involves (1) a *representation mapping* of all constructs in a benchmark ontology to constructs in a target grammar, and (2) an *interpretation mapping* from each construct in the target grammar to constructs in the benchmark ontology (e.g., Green & Rosemann, 2000; Rosemann, Recker, Green, & Indulska, 2009; Wand & Weber, 1989, 1993; Weber & Zhang, 1996). The representation mapping pinpoints instances of construct deficit. The interpretation mapping pinpoints instances of construct redundancy, overload, and excess. The mappings are sometimes difficult to undertake, however, and their predicted outcomes are sometimes difficult to test. In the subsections below, we examine the nature of the problems that researchers have encountered and the ways they have sought to overcome them.

7.1.1 Use of a Focused Ontology in Ontological Analyses

Rosemann and Green (2000) were concerned that some results they obtained via ontological analyses lacked importance and relevance. Thus, they suggested use of a “focused” ontology for the representation and interpretation mappings. Ontological analysis often occurs within the context of a *perspective*, which reflects the types of and purposes of users who undertake conceptual modeling. For example, analysts who use grammars to model executable workflows might not be interested in representing the ontological constructs of *thing, class, kind, and level structure* (Rosemann & Green, 2000). This concept of *perspective* was developed and used in subsequent studies to pinpoint defects in a grammar that most likely concerned its users (Recker, Rosemann, & Krogstie, 2007; Rosemann et al., 2009).

7.1.2 Enhancing Comprehensibility and Comparability of Ontological Analyses

Some researchers who undertook ontological analyses reported they found various constructs in the benchmark ontology and target grammar difficult to understand (Opdahl & Henderson-Sellers, 2004; Rosemann, Green, & Indulska, 2004a). As a result, they argued that comparing the results of different ontological analyses was problematic because different researchers might have interpreted the ontological and grammatical constructs differently.

To address these concerns, Rosemann and Green (2002) proposed use of a metamodel for both the benchmark ontology and target grammar. They argued the metamodel should be expressed in a commonly used grammar such as the extended entity-relationship grammar (eER). They predicted metamodels would enhance the understandability of constructs in a benchmark ontology and target grammar. Similarly, Opdahl and Henderson-Sellers (2004) proposed a template to define benchmark ontological constructs in UML. If the template was then used to define constructs in a target grammar, they argued that problems of understandability and comparability would be mitigated. In the same way, Harzallah, Berio, and Opdahl (2012) proposed a unified enterprise modeling language (UEML) based on a benchmark ontology that used constructs from the CASE tool, IDEF3. By defining rules for applying UEML to a target grammar, they argued that ontological analyses would be more understandable and comparable.

Based on a manual analysis of 250 classes in an OO application, Kiwelekar and Joshi (2013) derived syntactic and implementation features of OO programming elements. They then used these elements to develop classification rules based on four ontological constructs—thing, property, event, and process. The rules were implemented in Ontoclassifier, an automated classifier system. Subsequently, they applied Ontoclassifier to two different OO applications. A human analyst validated the resulting classification. The overall agreement levels were 56% and 60%. They argued their rule-based approach mitigated subjectivity issues associated with prior approaches to ontological analyses.

7.1.3 Improving the Validity and Reliability of Mappings During Ontological Analyses

Because Rosemann et al. (2004a; 2004b) found both the representation and interpretation mappings to be sometimes uncertain, they suggested a three-step, dual-coder process to improve the validity and reliability of the mappings. They argued their process improves the construct validity of ontological analyses. It has been used by Green et al. (2007) to demonstrate construct validity in their analysis of four candidate Web service interoperability standards and Recker et al. (2010, 2009) in their mapping analysis of BPMN. Tegarden et al. (2013) used a similar process to improve the validity and reliability of mappings between Bunge's (1977) ontology and McCarthy's (1982) REA enterprise ontology.

Kazemzadeh and Milton (2015) compared two grammars used to visualize service delivery processes—service blueprinting and process-chain networks. They employed a mapping process proposed by Milton and Kazmierczak (2004) and showed how it could be implemented when the reference ontology was not Bunge's (1977). They assessed the level of overlap in the semantic meaning of constructs in the two grammars as total, partial, or none. Similar to Rosemann et al.'s (2004a, 2004b) mapping process, theirs also required researchers to discuss and agree upon the mappings.

7.1.4 Improving the Validity and Reliability of Measures of User Perceptions of Construct Redundancy, Overload, Excess, and Deficit

Recker and Rosemann (2010) were concerned with empirical testing of the RM's predictions about the deleterious effects of construct deficit, redundancy, overload, and excess on the users of a grammar. They required a way to measure how a grammar's users would perceive such deficiencies. They describe a rigorous process for the development of the items they used to measure users' perceptions of construct deficit, redundancy, overload, and excess in a grammar. They worked with theory experts and practitioner panels to identify, rank, select, and revise items for the final measures in their survey instrument.

7.1.5 Ensuring Ontological Constructs and Grammatical Constructs are at the Same Level of Abstraction

Several researchers have noted that instances of construct redundancy and overload sometimes mean the level of abstractions used by the ontological benchmark and the modeling grammar to classify real-world phenomena are not aligned (e.g., Fickinger & Recker, 2013; Opdahl & Henderson-Sellers, 2002, pp. 60-62; Weber, 1997, p. 99). As a result, an interpretation mapping indicates only prima facie cases of redundancy and overload. These then need to be investigated further to determine whether they are substantive.

For instance, Tegarden et al. (2013) found multiple constructs in McCarthy's (1982) REA enterprise ontology map to Bunge's (1977) singular "coupling" construct (a prima facie case of ontological redundancy). They point out, however, that constructs in the REA enterprise ontology are based on a more detailed, problem-domain ontology of "couplings" than the construct of "coupling" used in Bunge's upper-level ontology. As a result, they concluded, "ontological redundancies…are not an issue from an ontological completeness or clarity perspective" (Tegarden et al., 2013, p. 118).

Similar considerations apply to prima facie instances of construct overload. For instance, Tegarden et al. (2013) found that several constructs in McCarthy's (1982) REA enterprise ontology each map to Bunge's constructs of "class" and "kind". Within the ontological benchmark, however, "kinds" are a particular type of "class". Thus, the overloaded constructs in McCarthy's (1982) REA enterprise ontology arise because Bunge (1977) benchmark ontology uses a more finely grained classification of real-world phenomena.

7.2 Some Reflections

All reported methodological improvements are valuable. Nonetheless, only two have had some acceptance among researchers—namely, use of dual coders to increase the validity and reliability of the representation and interpretation mappings, and adaptation of the process to derive a more appropriate ontological benchmark than the RM for a focal domain. Moreover, we are aware of no studies that have defined a process for applying the STM to the scripts that underlie an IS. Similarly, we are aware of only a few studies that have attempted to apply the GDM to system decompositions to evaluate whether users' understanding of real-world phenomena is impacted (Burton-Jones & Meso, 2006; Reijers et al., 2011; Tollington & Spinelli, 2012). Much scope exists, therefore, to adopt existing methodological improvements that enable more rigorous testing of the RM and to develop methodological improvements that facilitate enactment and testing of the STM and GDM.

8 Critiques of RT and its Models

We found six categories of critiques of RT made by researchers. Each addresses a different aspect of the RM, STM, and GDM:

1. *Critiques of the representation assumption*—that an IS provides a representation of users' perception of some real-world domain.
2. *Critiques of the ontological assumption*—that ontological theories can give insights into the nature of the real world and thus users' perceptions of the real world.
3. *Critiques of the use of Bunge's ontology*—that Bunge's ontology provides a suitable benchmark for evaluating conceptual-modeling grammars.

4. *Critiques of the validity and reliability of representation and interpretation mappings*—that the mappings provide a valid and reliable basis for drawing implications about the strengths and weaknesses of conceptual-modeling grammars.
5. *Critiques of the validity and reliability of empirical results*—that the empirical results obtained so far support the theory.
6. *Critiques of implications for practice*—that the results of the research offer useful implications for practice.

Table B5 outlines the six categories and the critiques and responses in the associated literature. Interestingly, to date we found one aspect of the research program has escaped criticism—namely, the mapping principle, which states the quality of a conceptual-modeling grammar can be assessed by determining whether its constructs have a one-to-one mapping with the set of constructs in a benchmark ontology. Researchers seem to accept that a mapping between ontological constructs and grammatical constructs can be useful even if they do not agree on the ontological benchmark to be employed and how to use it.

In our view, the criticisms and responses shown in Table B5 illustrate both the benefits and costs of the discourse that has occurred so far. The primary benefit is that researchers have been forced to understand relevant issues more deeply. For instance, to use Bunge's ontological benchmark in the mapping process (Element 3 in Table B5), most researchers in the early years of research on RT relied on the list of Bunge's (1977) ontological constructs articulated by Wand and Weber (1993). Although their list remains influential, researchers now frequently go back to Bunge (1977) to understand his work more clearly or to identify new constructs worthy of attention. A case in point is the ontological concept of "precedence" (the fact that one property, such as being female, is required for another property, such as bearing a child). Wand and Weber (1993) did not include "precedence" in their list of constructs. Parsons (2011) critiqued prior work for not considering precedence and showed how representing it explicitly in conceptual models could be useful.

The primary cost of the debate has been that it has slowed research. This outcome applies most to debates about Elements 1, 2, and 3 in Table B5, which involve difficult-to-resolve philosophical issues. Opdahl (2006) argues that often forays about such issues are unlikely to be productive because IS researchers are ill-trained to engage in them. Instead, he recommends that IS researchers focus on the practical utility of the outcomes obtained from RT-based research. If it is low, RT's philosophical merits are secondary. If it is high, RT's philosophical underpinnings might then be scrutinized.

We agree with Opdahl (2006). Debates about the representation and ontological assumptions and use of Bunge's ontology (Elements 1-3 in Table B5) remind researchers of the value of adopting different perspectives (Locke, 1998). Resolving such philosophical debates is difficult, however, because the protagonists must try to convince others through rhetorical force. Moreover, these types of debates differ from *substantive* debates about whether research based on particular assumptions or ontologies leads to useful outcomes. The advantage of concentrating on substantive issues is that we learn lessons that apply, irrespective of one's philosophical perspective. For instance, whether one adopts Bunge's (1977, 1979) ontology or Searle's (1995, 2006, 2010) ontology, one needs to design empirical tests that offer useful insights for practice.

These reflections lead to our second major observation regarding Table B5—the most salient criticism made of the research to date is lack of evidence about its practical usefulness. This outcome most likely reflects the research program's history. The early years involved substantial theoretical and analytical work (and debates about philosophical issues). Attention then moved to operationalizing the theory and conducting empirical tests. Only recently have researchers actively engaged with practitioners to design grammars using RT principles (Recker, Indulska, & Green, 2007) and to design practice interventions (Wand, Woo, & Wand, 2008). We believe research on RT would also benefit significantly from inductive field work, but such work remains rare overall (Patel et al., 1998; Strong & Volkoff, 2010).

9 Future Research Using RT

Based on our review of how RT has been applied, advanced, enacted, and tested to date, we contend that research on RT should continue. Thus, our final goal in writing this paper was to identify some ways in which future research on RT might be conducted. Rather than list an agenda with open research questions, in this section we outline two examples (Appendices C1 and C3 provide more detail) that we believe show that RT provides a rich basis for future research.

9.1 Large Data Sets, Data Mining, and the Good-Decomposition Model

One opportunity we see to deploy RT in a novel way is to use the GDM to help interpret relationships identified via data-mining operations on large data sets. Often these operations show that many statistically significant relationships exist (Lin, Lucas Jr., & Shmueli, 2013). Initially, a major focus is to identify the *material* relationships via the size of an effect and the variance explained (George, Haas, &

Pentland, 2014, p. 323). Once such relationships have been identified, often a major challenge is to interpret their *meaning* (are they spurious or substantive?). A good test of the GDM would be to see whether it might assist with this task.

Appendix C1 describes a case study of a large hospital that has endeavored to use data-mining activities to improve outcomes for patients who are transferred between its emergency department and inpatient wards. To date, these activities have proved fruitless, primarily because the relationships identified are many, their meaning is often unclear, and obtaining a coherent, overall interpretation of them has been difficult. In Appendix C2, we show in some detail how the GDM might be used in a bottom-up way to build a level structure (Bunge, 1979, pp. 13-14) of systems and subsystems on top of the relationships. We predict that use of level structures will enable stakeholders to better interpret the meaning of the relationships.

9.2 Effective Use of Information Systems

As IS evolve (often enabled by technological improvements), a major challenge is to ensure they are used effectively. In this regard, Zuboff (1998, p. 70) notes: “technological change [is] an occasion for developing a new set of skills—skills that are able to exploit the informating capacity of the technology.… We first have to understand the nature of these new skills”. In Appendix C3, we show how the RM, STM, and GDM can be used to inform research that explores how IS can be used more effectively. We argue the three models can be employed to foster innovative ideas about *affordances* (Volkoff & Strong, 2018) that IS might offer. Through creative elaboration of and analysis of such affordances, we predict that users can employ an IS more effectively. The new skills that users need are those that enable them to apply the three models in ways that tease out the innovative affordances that IS might offer.

10 Some Broader Implications

In this paper, we reviewed the literature published on RT with five specific goals in mind. The analyses we conducted in our review led us to conclude that research on RT should continue, and thus we developed two specific proposals for future research on RT. While we hope both the analyses and proposals have merit on their own, we also see two broader implications of the current and previous review work that has been done.

The first implication becomes apparent when comparing and contrasting our current review with the previous two reviews of RT (Burton-Jones et al., 2017; Saghafi & Wand, 2014). Even though these previous reviews use subsets of the publications covered in the current review, Table 4 shows that in combination we have a more comprehensive picture of the cumulative knowledge generated by scholars who have used RT than any of the three reviews can provide in isolation. In particular, the three reviews indicate:

- Support for the RM is strong, but work on and a verdict about the merits of the STM and GDM are lacking.
- While diverse research methods have been used to study RT, improved methods are required, and need to be adopted more broadly, to increase the trustworthiness and credibility of the results obtained.
- Some early research suggests that RT has potential to account for phenomena in a number of new domains, but the work so far across diverse domains has been limited.
- A better understanding of both the explicit and implicit assumptions that underpin RT is needed, but the focus of work should then be on testing the merits of these assumptions rather than engaging in philosophical debates about them.

The second implication of our review arises from the way we designed our analyses of the literature. As we note in Burton-Jones et al. (2017), some notable exceptions aside (e.g., Berthon, Pitt, Ewing, & Carr, 2002; Gray & Cooper, 2010), few frameworks have been developed to assist with the evaluation of entire research programs about a single theory. Our approach to coding the literature addresses topics that we believe are relevant to most research programs that focus on a particular theory (e.g., application domains, empirical support, theoretical advances, methodological advances, and critiques), which implies our approach might prove useful when evaluating other theoretically focused research programs, including those competing with RT (e.g., Allen & March, 2006b; March & Allen, 2014).

11 Limitations

Our findings should be considered in the context of the way we undertook the literature review. First, we were selective in our reading of literature that cited the papers we deemed seminal. For those papers published during or subsequent to 2011, we read only those that had been cited at least once. For those papers published prior to 2011, we read only those that had been cited at least 10 times. Nonetheless, we reviewed all potential exclusion candidate papers to establish whether to consider them (17 papers). We may still have missed papers that have used RT in rich, innovative ways and that in due course will have an impact on scholars or practitioners.

Second, our coding scheme was derived after substantial reading of and reflection on the literature.

Nonetheless, it imposes a particular "lens" on the literature—a lens that enables us to "see" certain aspects of a paper but blinds us to other aspects. Other lenses may lead to interpretations of the literature that differ from ours. Potentially these interpretations might be richer and more insightful.

Third, we are mindful of the fact that any literature review is an *interpretive* act. From a hermeneutic perspective, we read "texts" and tried to make sense of them. As we read each text, we endeavored to form an overall view of the themes conveyed in the literature. As this overall view changed, we sometimes reread papers because we had a new understanding of their narratives (a form of the hermeneutic circle).

Fourth, written texts have the properties of *distantiation* (detachment from their authors) and *autonomy* (the meaning ascribed to them may not match the author's meaning) (Ricoeur, 1975). Thus, as readers of the literature on RT, we may have *appropriated* meaning in ways the authors neither intended nor envisaged. Indeed, we saw these outcomes present in a number of publications that cited our own work. At times, we found our ideas had been used in surprising ways. We anticipate, therefore, that in some cases we may have also interpreted texts about RT in ways that are not congruent with the meaning intended by their authors.

Table 4. Summary of Key Findings from the Three Reviews of RT

Findings	**Saghafi & Wand (2014)**	**Burton-Jones et al. (2017)**	**This Paper**
Support for the representation model (RM)	Adhering to the RM's predictions when designing conceptual models improves users' understanding of them.	The RM appears to be a robust, parsimonious theory.	For the most part, empirical support for the premises of the RM is substantial.
Support for the state-tracking model (STM) and good-decomposition model (GDM)	*Not the focus of this paper*	The overall success or failure of the STM and GDM is uncertain.	The relative merits of the STM and GDM are uncertain. Their uptake has been too limited to evaluate their premises.
Methodological issues	*Not the focus of this paper*	Researchers have studied various independent, moderating, and dependent variables, but they have not consistently used comparable measures.	Methodological variety is evident, but methodological advances are few. Some designs choices used in some empirical studies are problematic.
RT's scope	*Not the focus of this paper*	RT accounts for a wider variety of phenomena than first conceived. Even so, RT's empirical implications in some newer domains where it has been applied remain uncertain.	Over time, RT's application scope has broadened and is now decisively varied.
RT's assumptions	*Not the focus of this paper*	RT may have some problematic assumptions, but whether they are in fact problematic has been underexplored.	Some of RT's assumptions have been debated, while others have not yet entered the discourse.
Philosophical critiques of RT	*Not the focus of this paper*	*Not the focus of this paper*	Critiques of philosophical issues surrounding RT may have slowed down research but have also motivated studies to resolve deep issues.

12 Conclusions

In this paper, we reviewed the literature associated with RT—a theory that conceives IS as representations of real-world phenomena. In our review of 365 papers referencing RT, we found it has motivated a large amount of research in diverse areas. It has also been used in many ways that were not envisioned at the time it was developed. These outcomes attest to its richness. Moreover, while empirical tests of RT have produced mixed results, they have been sufficiently encouraging to motivate its ongoing use.

RT and its applications have been subjected to a number of criticisms. These relate to its fundamental assumptions, the validity and reliability of results obtained from empirical work done to test it, and the practical usefulness of the results obtained. While we believe some criticisms are substantive, we believe others are misplaced in light of the collected evidence. We also believe some criticisms are better canvassed by scholars in other disciplines.

Even though RT has been applied in increasingly diverse ways, its main premises and constructs have remained relatively stable. Few theoretical advances were evident in the literature. Some good progress has been made, however, in terms of how RT might be better tested empirically. Unfortunately, we found little evidence of take-up in the literature of measures developed to improve the validity, reliability, and credibility of empirical tests of RT.

For many years, we have been active researchers on and contributors to the theory that is the focus of our review. Our long-term engagement with RT reflects our fundamental belief in its merits. While we are circumspect about its strengths and limitations, our review indicates that much of RT's potential has yet to be realized.

Acknowledgments

The preparation of this paper was supported in part by three Australian Research Council Discovery Grants (DP110104386 to Andrew Burton-Jones and Ron Weber, DP130102454 to Marta Indulska, Jan Recker, and Peter Green, and DP140101815 to Peter Green and Andrew Burton-Jones). We thank the senior editor and the three reviewers for their helpful comments.

References

Ågerfalk, P. J. (2010). Getting pragmatic. *European Journal of Information Systems, 19*(3), 251-256.

Ahmad, M., & Odeh, M. (2013). Blueprint of a semantic business process-aware enterprise information architecture: The EIAOnt ontology. In S. Hammoudi, J. Cordeiro, L. A. Maciaszek, & J. Filipe (Eds.), *Enterprise information systems: ICEIS 2013 proceedings* (pp. 520-539). New York, NY: Springer.

Allen, G. N., & March, S. T. (2006a, December). *A critical assessment of the Bunge-Wand-Weber ontology of conceptual modeling*. Paper presented at the 16th Workshop on Information Technology and Systems, Milwaukee, WI.

Allen, G. N., & March, S. T. (2006b). The effects of state-based and event-based data representation on user performance in query formulation tasks. *MIS Quarterly, 30*(2), 269-290.

Allen, G. N., & March, S. T. (2012). A research note on representing part-whole relations in conceptual modeling. *MIS Quarterly, 36*, 945-964.

Allmark, M., Grosvenor, R., Byrne, C., Anayi, F., & Prickett, P. (2013, September). *Condition monitoring of a tidal stream turbine: Development of an experimental methodology*. Paper presented at the 10th European Wave and Tidal Energy Conference, Aalborg, Denmark.

Altuhhova, O., Matulevičius, R., & Ahmed, N. (2013). An extension of business process model and notation for security risk management. *International Journal of Information System Modeling and Design, 4*(4), 93-113.

Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32.

Avédissian, A., Valverde, R., & Barrad, S. (2015). An extension proposition for the agent-based language modeling ontology for the representation of supply chain integrated business processes. *WSEAS Transactions on Business and Economics, 12*, 211-228.

Bera, P., Burton-Jones, A., & Wand, Y. (2011). Guidelines for designing visual ontologies to support knowledge identification. *MIS Quarterly, 35*(4), 883-908.

Bera, P., Burton-Jones, A., & Wand, Y. (2014). How semantics and pragmatics interact in understanding conceptual models. *Information Systems Research, 25*(2), 401-419.

Bera, P., & Evermann, J. (2014). Guidelines for using UML association classes and their effect on domain understanding in requirements engineering. *Requirements Engineering, 19*(1), 63-80.

Bernaert, M. (2010, June). *Integrating the semantics of events, processes and tasks across requirements engineering layers*. Paper presented at the The CAiSE Doctoral Consortium 2010, Hammamet, Tunisia.

Berthon, P., Pitt, L., Ewing, M., & Carr, C. L. (2002). Potential research space in MIS: A Framework for envisioning and evaluating research replication, extension and generation. *Information Systems Research, 13*(4), 416-427.

Bhatt, C. A., & Kankanhalli, M. S. (2011). Multimedia data mining: State of the art and challenges. *Multimedia Tools and Applications, 51*(1), 35-76.

Bjeković, M., Proper, E., & Sottet, J.-S. (2014). Embracing Pragmatics. In E. S. K. Yu, G. Dobbie, M. Jarke & S. Purao (Eds.), *Conceptual modeling: ER 2014 proceedings* (pp. 431-444). New York, NY: Springer.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*(1-2), 245-271.

Bodart, F., Patel, A., Sim, M., & Weber, R. (2001). Should optional properties be used in conceptual modelling? A theory and three empirical tests. *Information Systems Research, 12*(4), 384-405.

Bunge, M. A. (1977). *Treatise on basic philosophy Volume 3: Ontology I: The furniture of the world*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Bunge, M. A. (1979). *Treatise on basic philosophy: Volume 4: Ontology II: A world of systems*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Burton-Jones, A., & Grange, C. (2013). From use to effective use: A representation theory Perspective. *Information Systems Research, 24*(3), 632-658.

Burton-Jones, A., & Meso, P. (2002, December). *How good are these UML Diagrams? An empirical test of the Wand and Weber good decomposition model.* Paper presented at the 23rd International Conference on Information Systems, Barcelona, Spain.

Burton-Jones, A., & Meso, P. (2006). Conceptualizing systems for understanding: An empirical test of decomposition principles in object-oriented analysis. *Information Systems Research, 17*(1), 38-60.

Burton-Jones, A., & Meso, P. (2008). The effects of decomposition quality and multiple forms of information on novices' understanding of a domain from a conceptual model. *Journal of the Association for Information Systems, 9*(12), 784-802.

Burton-Jones, A., Recker, J., Indulska, M., Green, P., & Weber, R. (2017). Assessing representation theory with a framework for pursuing success and failure. *MIS Quarterly, 41*(4), 1307-1333.

Burton-Jones, A., & Volkoff, O. (2017). How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. *Information Systems Research, 28*(3), 468-489.

Burton-Jones, A., Wand, Y., & Weber, R. (2009). Guidelines for empirical evaluations of conceptual modeling grammars. *Journal of the Association for Information Systems, 10*(6), 495-532.

Burton-Jones, A., & Weber, R. (1999, December). *Understanding relationships with attributes in entity-relationship diagrams.* Paper presented at the 20th International Conference on Information Systems, Charlotte, NC.

Cerrito, P. B. (2006). A Process to Define Sequential Treatment Episodes for Patient Care. In B. Kitts & G. Melli (Eds.), *Data mining case studies* (pp. 65-76). New York, NY: ACM.

Chaturvedi, A. R., Dolk, D. R., & Drnevich, P. L. (2011). Design principles for virtual worlds. *MIS Quarterly, 35*(3), 673-684.

Chen, W. S., & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal, 14*(3), 197-235.

Chen, Z., & Pooley, R. (2009a, April). *Domain modeling for enterprise information systems-formalizing and extending Zachman framework using BWW ontology*. Paper presented at the 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles, CA.

Chen, Z., & Pooley, R. (2009b, July). *Rediscovering Zachman framework using ontology from a requirement engineering perspective.* Paper presented at the 33rd Annual IEEE International Computer Software and Applications Conference, Seattle, WA.

Chidamber, S. R., & Kemerer, C. F. (1994). A metrics suite for object oriented design. *IEEE Transactions on Software Engineering, 20*(6), 476-493.

Chowdry, S. M., Mishuris, R. G., & Mann, D. (2017). Problem-oriented charting: A review. *International Journal of Medical Informatics, 103*(July), 95-102.

Clarke, R., Burton-Jones, A., & Weber, R. (2013, December). *Improving the semantics of conceptual-modeling grammars: A new perspective on an old problem*. Paper presented at the 34th International Conference on Information Systems, Milan, Italy.

Claus, P. L., Carpenter, P. C., Chute, C. G., Mohr, D. N., & Gibbons, P. S. (1997, October). *Clinical care management and workflow by episodes*. Paper presented at the AMIA Annual Fall Symposium, Nashville, TN.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and psychological measurement, 20*(1), 37-46.

Colomb, R. M., & Ahmad, M. N. (2010). A perdurant ontology for interoperating information systems based on interlocking institutional worlds. *Applied Ontology, 5*(1), 47-77.

Compeau, D. R., Marcolin, B. L., Kelley, H., & Higgins, C. A. (2012). Generalizability of information systems research using student subjects: A reflection on our practices and recommendations for future research. *Information Systems Research, 23*(4), 1093-1109.

Davies, I., Green, P., Milton, S. K., & Rosemann, M. (2003, June). *Using meta models for the comparison of ontologies*. Paper presented at the 8th CAiSE/IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design, Velden, Austria.

Davies, I., Green, P., Milton, S. K., & Rosemann, M. (2005). Analysing and comparing ontologies with meta models. In J. Krogstie, T. Halpin & K. Siau (Eds.), *Information modeling methods and methodologies* (pp. 1-16). Hershey, PA: Idea Group.

Davies, I., Rosemann, M., & Green, P. (2004, December). *Exploring proposed ontological issues of ARIS with different categories of modellers*. Paper presented at the 15th Australasian Conference on Information Systems, Hobart, Australia.

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results* (Unpublished doctoral dissertation) Massachusetts Institute of Technology, Boston, MA.

Dennis, A. R., & Valacich, J. S. (2015). A replication manifesto. *AIS Transactions on Replication Research, 1*(1), 1-4.

El-Attar, M., Luqman, H., Kárpáti, P., Sindre, G., & Opdahl, A. L. (2015). Extending the UML statecharts notation to model security aspects. *IEEE Transactions on Software Engineering, 41*(7), 661-690.

El-Tawy, N., & Tollington, T. (2013). Some thoughts on the recognition of assets, notably in respect of intangible assets. *Accounting Forum, 37*(1), 67-80.

Etien, A., & Rolland, C. (2005). Measuring the fitness relationship. *Requirements Engineering, 10*(3), 184-197.

Evermann, J., & Wand, Y. (2005). Ontology based object-oriented domain modelling: Fundamental concepts. *Requirements Engineering, 10*(2), 146-160.

Fettke, P., & Loos, P. (2005). Ontological analysis of reference models. In P. Green & M. Rosemann (Eds.), *Business systems analysis with ontologies* (pp. 56-81). Hershey, PA: Idea Group.

Fettke, P., & Loos, P. (2007). Ontological evaluation of Scheer's reference model for production planning and control systems. *Journal of Interoperability in Business Information Systems, 2*(1), 9-28.

Fickinger, T., & Recker, J. (2013, June). *Construct redundancy in process modelling grammars: Improving the explanatory power of ontological analysis*. Paper presented at the 21st European Conference on Information Systems, Utrecht, Netherlands.

Frank, U. (2013). Domain-specific modeling languages: Requirements Analysis and design guidelines. In I. Reinhartz-Berger, A. Sturm, T. Clark, S. Cohen & J. Bettin (Eds.), *Domain engineering: Product lines, languages, and conceptual models* (pp. 133-157). Berlin: Springer.

Galliers, R. D., & Whitley, E. A. (2007). Vive les Differences? Developing a profile of european information systems research as a basis for international comparisons. *European Journal of Information Systems, 16*(1), 20-35.

Gammon, D., Berntsen, G. K. R., Koricho, A. T., Sygna, K., & Ruland, C. M. (2015). The Chronic care model and technological research and innovation: A scoping review at the crossroads. *Journal of Medical Internet Research, 17*(2), e25.

Gehlert, A., & Esswein, W. (2005). Ontological analysis reconsidered: A formal approach. In M.-A. Sicilia, S. Sánchez-Alonso, E. García-Barriocanal & J. J. Cuadrado-Gallego (Eds.), *Ontology, conceptualizations and epistemology for software and systems engineering 2005: Proceedings of the ONTOSE 2005 International Workshop* (Vol. 163).

Gemino, A. (1999). *Empirical methods for comparing system analysis techniques* (Unpublished doctoral dissertation). University of British Columbia, Vancouver, Canada.

Gemino, A. (2004). Empirical comparisons of animation and narration in requirements validation. *Requirements Engineering, 9*(3), 153-168.

Gemino, A., & Wand, Y. (2005). Complexity and clarity in conceptual modeling: Comparison of mandatory and optional properties. *Data & Knowledge Engineering, 55*(3), 301-326.

Genero, M., Poels, G., & Piattini, M. (2008). Defining and validating metrics for assessing the understandability of entity-relationship diagrams. *Data & Knowledge Engineering, 64*(3), 534-557.

George, G., Haas, M. R., & Pentland, A. (2014). From the editors: Big data and management. *Academy of Management Journal, 57*(2), 321-326.

Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science*. Chicago, IL: University of Chicago Press.

Goumopoulos, C., & Kameas, A. (2008). Achieving Co-operation and developing smart behavior in collections of context-aware artifacts. In K.

Delaney (Ed.), *Ambient intelligence with microsystems: Augmented materials and smart objects* (pp. 205-237). New York, NY: Springer.

GRADE Working Group. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal, 328*(7454), 1490-1494.

Gray, P. H., & Cooper, W. H. (2010). Pursuing failure. *Organizational Research Methods, 13*(4), 620-643.

Green, P. (1997, June). *Use of information systems analysis and design (ISAD) grammars in combination in upper case tools: An ontological evaluation*. Paper presented at the 2nd CAiSE/ IFIP8.1 International Workshop on Evaluation of Modeling Methods in Systems Analysis and Design, Barcelona, Spain.

Green, P., & Rosemann, M. (2000). Integrated process modeling. an ontological evaluation. *Information Systems, 25*(2), 73-87.

Green, P., Rosemann, M., & Indulska, M. (2005). Ontological evaluation of enterprise systems interoperability using ebXML. *IEEE Transactions on Knowledge and Data Engineering, 17*(5), 713-725.

Green, P., Rosemann, M., Indulska, M., & Manning, C. (2007). Candidate interoperability standards: an ontological overlap analysis. *Data & Knowledge Engineering, 62*(2), 274-291.

Green, P., Rosemann, M., Indulska, M., & Recker, J. (2006, December). *Improving representational analysis: An example from the enterprise interoperability domain*. Paper presented at the 17th Australasian Conference on Information Systems, Adelaide, Australia.

Green, P., Rosemann, M., Indulska, M., & Recker, J. (2011). Complementary use of modeling grammars. *Scandinavian Journal of Information Systems, 23*(1), 59-86.

Gregersen, H., & Jensen, C. S. (1999). On the ontological expressiveness of temporal extensions to the entity-relationship model. In P. P. Chen, D. W. Embley, J. Kouloumdjian, S. W. Liddle & J. F. Roddick (Eds.), *Advances in conceptual modeling: Proceedings of ER '99 Workshops on Evolution and Change in Data Management, Reverse Engineering in Information Systems, and the World Wide Web and Conceptual Modeling* (pp. 110-121). New York, NY: Springer.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611-642.

Hadar, I., & Soffer, P. (2006). Variations in conceptual modeling: classification and ontological analysis. *Journal of the Association for Information Systems, 7*(8), 568-592.

Harzallah, M., Berio, G., & Opdahl, A. L. (2012). New Perspectives in Ontological Analysis: guidelines and rules for incorporating modelling languages into UEML. *Information Systems, 37*(5), 484-507.

Harzing, A.-W. (2010). *The publish or perish book: Your guide to effective and responsible citation analysis*. Melbourne, Australia: Tarma Software Research.

Heales, J. (2002). A model of factors affecting an information system's change in state. *Journal of Software Maintenance and Evolution: Research and Practice, 14*(6), 409-427.

Heidari, F., Loucopoulos, P., Brazier, F., & Barjis, J. (2013). A unified view of business process modelling languages *Internal University Report*. Delft, Netherlands: Delft University of Technology.

Herrera, S. I., Pallioto, D., Tkachuk, G., & Luna, P. A. (2005). Ontological modelling of information systems from Bunge's contributions. In J. Castro & E. Teniente (Eds.), *Proceedings of the CAiSE'05 Workshops* (Vol. 2, pp. 571-582).

Hovorka, D. S., Johnston, R., & Riemer, K. (2014, October). *Reclaiming meaning in IS theory: Theory for why, not what*. Paper presented at the 2014 Information Systems Foundations Workshop: The Value of Information Systems Canberra, Australia.

Indulska, M., Recker, J., Green, P., & Rosemann, M. (2007, August). *Are we there yet? Seamless mapping of BPMN to BPEL4WS*. Paper presented at the 13th Americas Conference on Information Systems, Keystone, CO.

Jackson, M. A. (1983). *System development*. Englewood Cliffs, NJ: Prentice Hall.

Johnston, R., & Milton, S. K. (2002). The foundation role for theories of agency in understanding information systems design. *Australasian Journal of Information Systems, 10*(1), 40-49.

Karow, M., Gehlert, A., Becker, J., & Esswein, W. (2006, August). *On the transition from computation independent to platform independent models*. Paper presented at the 12th Americas Conference on Information Systems, Acapulco, Mexico.

Kazemzadeh, Y., & Milton, S. K. (2015). Service blueprinting and process-chain-network: An ontological comparison *International Journal of Qualitative Research in Services, 2*(1), 1-12.

Keen, C. D., & Lakos, C. (1994, September). *Information systems modelling using LOOPN++, an Object Petri Net scheme.* Paper presented at the 4th International Working Conference on Dynamic Modelling and Information Systems, Noordwijkerhout, Netherlands.

Keen, C. D., & Lakos, C. (1996, January). *Analysis of the design constructs required in process modelling.* Paper presented at the International Conference on Software Engineering: Education and Practice, Dunedin, New Zealand.

Khatri, V., Vessey, I., Ramesh, V., Clay, P., & Sung-Jin, P. (2006). Understanding conceptual schemas: exploring the role of application and IS domain knowledge. *Information Systems Research, 17*(1), 81-99.

King, W. R., & He, J. (2006). A Meta-analysis of the technology acceptance model. *Information & Management, 43*(6), 740-755.

Kiwelekar, A. W., & Joshi, R. K. (2013, September). *Ontological interpretation of object-oriented programming abstractions*. Paper presented at the 13th International Working Conference on Source Code Analysis and Manipulation, Eindhoven, Netherlands.

Klein, H. K., & Hirschheim, R. (2006). Further reflections on the IS discipline: Climbing the Tower of Babel. In J. L. King & K. Lyytinen (Eds.), *Information systems: The state of the field* (pp. 307-323). Chichester, UK: Wiley.

Kolodny, A., Courtwright, D. T., Hwang, C. S., Kreiner, P., Eadie, J. L., Clark, T. W., & Alexander, G. C. (2015). The prescription opioid and heroin crisis: A public health approach to an epidemic of addiction. *Annual Review of Public Health, 36*, 559-574.

Krogstie, J. (2012). *Model-based development and evolution of information systems*. London: Springer.

Kruchten, P., Woo, C. C., Monu, K., & Sotoodeh, M. (2008). A conceptual model of disasters encompassing multiple stakeholder domains. *International Journal of Emergency Management, 5*(1), 25-56.

Kwon, S. J. (2011). Conceptual modeling of causal map: object oriented causal map. *Expert Systems with Applications, 38*(1), 360-370.

Lemieux, V., & Limonad, L. (2011). What “good” looks like: Understanding records ontologically in the context of the global financial crisis. *Journal of Information Science, 37*(1), 29-39.

Lin, M., Lucas Jr., H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research, 24*(4), 906-917.

Lindland, O. I., Sindre, G., & Solvberg, A. (1994). Understanding quality in conceptual modeling. *IEEE Software, 11*(2), 42-49.

Locke, J. (1998). Essay concerning human understanding [1689 reprint]. In J. Perry & M. E. Bratman (Eds.), *Introduction to philosophy: Classical and contemporary readings* (pp. 139-144). New York, NY: Oxford University Press.

Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2014a, December). *The impact of conceptual modeling on dataset completeness: A field experiment*. Paper presented at the 35th International Conference on Information Systems, Auckland, New Zealand.

Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2014b). The IQ of the crowd: Understanding and improving information quality in structured user-generated content. *Information Systems Research, 25*(4), 669-689.

Lyytinen, K. (2006). “Ontological foundations of conceptual modeling” by Boris Wyssusek: A critical response. *Scandinavian Journal of Information Systems*, *18*(1), 81-84.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293-334.

March, S. T., & Allen, G. N. (2014). Toward a social ontology for conceptual modeling. *Communications of the Association for Information Systems, 34*(70), 1347-1358.

Matulevicius, R., Heymans, P., & Opdahl, A. L. (2007). Comparing GRL and KAOS using the UEML approach. In R. Jardim-Gonçalves, J. P. Müller, K. Mertins & M. Zelm (Eds.), *Enterprise interoperability II: IESA 2007 proceedings* (pp. 77-88). New York, NY: Springer.

McCarthy, W. E. (1982). The REA accounting model: A generalized framework for accounting systems in a shared data environment. *Accounting Review, 57*(3), 554-578.

Meertens, L. O., Iacob, M.-E., & Eckarts, S. M. (2010). Feasibility of EPC to BPEL model transformations based on ontology and patterns. In S. Rinderle-Ma, S. Sadiq & F. Leymann (Eds.), *Business process management workshops: Proceedings of BPM 2009 International Workshops* (pp. 347-358). New York, NY: Springer.

Milton, S. K. (2007). Ontological foundations of representational information systems. an australian perspective. *Scandinavian Journal of Information Systems, 19*(1), 109-134.

Milton, S. K., & Kazmierczak, E. (2004). An ontology of data modelling languages: A study using a common-sense realistic ontology. *Journal of Database Management, 15*(2), 19-38.

Milton, S. K., Rajapakse, J., & Weber, R. (2012). Ontological clarity, cognitive engagement, and conceptual model quality evaluation: An experimental investigation. *Journal of the Association for Information Systems, 13*(9), 657-694.

Moody, D. L. (2003, June). *The method evaluation model: A theoretical model for validating information systems design methods.* Paper presented at the 11th European Conference on Information Systems, Naples, Italy.

Moody, D. L., & Shanks, G. (2003). Improving the quality of data models: Empirical validation of a quality management framework. *Information Systems, 28*(6), 619-650.

Moody, D. L., Sindre, G., Brasethvik, T., & Sølvberg, A. (2003, May). *Evaluating the quality of information models: Empirical testing of a conceptual model quality framework.* Paper presented at the 25th International Conference on Software Engineering, Portland, OR.

Nagykaldi, Z., Tange, H., & De Maeseneer, J. (2018). Moving from problem-oriented to goal-directed health records. *Annals of Family Medicine, 16*(2), 155-159.

Nelson, H. J., & Monarchi, D. E. (2007). Ensuring the quality of conceptual representations. *Software Quality Journal, 15*(2), 213-233.

Nelson, H. J., Poels, G., Genero, M., & Piattini, M. (2012). A conceptual modeling quality framework. *Software Quality Journal, 20*(1), 201-228.

Nurcombe, B. (1989). Goal-directed treatment planning and the principles of brief hospitalization. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(1), 26-30.

O'Brien, P., & Burmeister, J. (2003). Ubiquitous travel service delivery. *Information Technology and Tourism, 5*(4), 221-233.

Opdahl, A. L. (1998, October). *Multi-perspective modelling of requirements: A case study using facet models*. Paper presented at the 3rd Australian Conference on Requirements Engineering, Geelong, Australia.

Opdahl, A. L. (2006). Response to Wyssusek's "On ontological foundations of conceptual modelling". *Scandinavian Journal of Information Systems, 18*(1), 95-106.

Opdahl, A. L. (2011). Anatomy of the unified enterprise modelling ontology. In M. van Sinderen & P. Johnson (Eds.), *Enterprise interoperability: IWEI 2011 proceedings* (pp. 163-176). New York, NY: Springer.

Opdahl, A. L., Berio, G., Harzallah, M., & Matulevicius, R. (2012). An ontology for enterprise and information systems modelling. *Applied Ontology, 7*(1), 49-92.

Opdahl, A. L., & Henderson-Sellers, B. (1999, January). *Evaluating and improving OO modelling languages using the BWW-Model.* Paper presented at the Information Systems Foundations Workshop: Ontology, Semiotics and Practice, Sydney, Australia.

Opdahl, A. L., & Henderson-Sellers, B. (2001). Grounding the OML Metamodel in Ontology. *Journal of Systems and Software, 57*(2), 119-143.

Opdahl, A. L., & Henderson-Sellers, B. (2002). Ontological evaluation of the UML using the Bunge-Wand-Weber model. *Software and Systems Modeling, 1*(1), 43-67.

Opdahl, A. L., & Henderson-Sellers, B. (2004). A template for defining enterprise modelling constructs. *Journal of Database Management, 15*(2), 39-73.

Opdahl, A. L., Henderson-Sellers, B., & Barbier, F. (2001). Ontological analysis of whole-part relationships in OO-models. *Information and Software Technology, 43*(6), 387-399.

Opdahl, A. L., & Sindre, G. (1997). Facet modelling: an approach to flexible and integrated conceptual modelling. *Information Systems, 22*(5), 291-323.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management, 52*(2), 183-199.

Parsons, J. (2011). An experimental study of the effects of representing property precedence on the comprehension of conceptual schemas. *Journal of the Association for Information Systems, 12*(6), 401-422.

Parsons, J., & Cole, L. (2005). What do the pictures mean? Guidelines for experimental evaluation of representation fidelity in diagrammatical conceptual modeling techniques. *Data & Knowledge Engineering, 55*(3), 327–342.

Parsons, J., & Wand, Y. (2003). Attribute-based semantic reconciliation of multiple data sources. *Journal on Data Semantics I, 1*, 21-47.

Patel, A., Sim, M., & Weber, R. (1998, December). *Stakeholder experiences with conceptual modeling: An empirical investigation*. Paper presented at the 19th International Conference on Information Systems, Helsinki, Finland.

Paulson, D., & Wand, Y. (1992). An automated approach to information systems decomposition. *IEEE Transactions on Software Engineering, 18*(3), 174-189.

Perepletchikov, M., Ryan, C., & Zahir, T. (2013). An analytical framework for evaluating service-oriented software development methodologies. *Journal of Software, 8*(7), 1642-1659.

Petter, S., DeLone, W. H., & McLean, E. R. (2013). Information systems success: The quest for the independent variables. *Journal of Management Information Systems, 29*(4), 7-62.

Ram, S., & Liu, J. (2007). Understanding the semantics of data provenance to support active conceptual modeling. In P. P. Chen & L. Y. Wong (Eds.), *Active conceptual modeling of learning* (pp. 17-29). Berlin: Springer.

Recker, J. (2005, August). *Evaluation of conceptual modeling languages: An epistemological discussion*. Paper presented at the 11th Americas Conference on Information Systems, Omaha, NE.

Recker, J. (2011). *Evaluations of process modeling grammars: Ontological, qualitative and quantitative analyses using the example of BPMN*. Berlin: Springer.

Recker, J., & Indulska, M. (2007). An ontology-based evaluation of process modeling with Petri Nets. *Journal of Interoperability in Business Information Systems, 2*(1), 45-64.

Recker, J., Indulska, M., & Green, P. (2007). Extending representational analysis: BPMN User and developer perspectives. In G. Alonso, P. Dadam & M. Rosemann (Eds.), *Business process management: BPM 2007 proceedings* (pp. 384-399). New York, NY: Springer.

Recker, J., Indulska, M., Rosemann, M., & Green, P. (2006, June). *How good is BPMN really? Insights from theory and practice*. Paper presented at the 14th European Conference on Information Systems, Goeteborg, Sweden.

Recker, J., & Rosemann, M. (2010). The measurement of perceived ontological deficiencies of conceptual modeling grammars. *Data & Knowledge Engineering, 69*(5), 516-532.

Recker, J., Rosemann, M., Green, P., & Indulska, M. (2006, September). *Extending the scope of representation theory: A review and a proposed research model*. Paper presented at the 3rd Biannual ANU Workshop on Information Systems Foundations, Canberra, Australia.

Recker, J., Rosemann, M., Indulska, M., & Green, P. (2009). Business process modeling: A comparative analysis. *Journal of the Association for Information Systems, 10*(4), 333-363.

Recker, J., Indulska, M., Rosemann, M., & Green, P. (2010). The ontological deficiencies of process modeling in practice. *European Journal of Information Systems, 19*(5), 501-525.

Recker, J., & Niehaves, B. (2008). Epistemological perspectives on ontology-based theories for conceptual modeling. *Applied Ontology, 3*(1-2), 111-130.

Recker, J., Rosemann, M., Green, P., & Indulska, M. (2011). Do ontological deficiencies in modeling grammars matter? *MIS Quarterly, 35*(1), 57-79.

Recker, J., Rosemann, M., & Krogstie, J. (2007). Ontology- versus Pattern-based evaluation of process modeling languages: A comparison. *Communications of the Association for Information Systems, 20*(48), 774-799.

Recker, J., Zur Muehlen, M., Siau, K., Erickson, J., & Indulska, M. (2009, August). *Measuring method complexity: UML versus BPMN*. Paper presented at the 15th Americas Conference on Information Systems, San Francisco, CA.

Regev, G., & Wegmann, A. (2004, September). *Defining early IT system requirements with regulation principles: The lightswitch approach*. Paper presented at the 12th IEEE International Requirements Engineering Conference, Kyoto, Japan.

Reijers, H. A., & Mendling, J. (2008). Modularity in process models: Review and effects. In M. Dumas, M. Reichert & M.-C. Shan (Eds.), *Business process management: BPM 2008 proceedings* (pp. 20-35). New York, NY: Springer.

Reijers, H. A., Mendling, J., & Dijkman, R. M. (2011). Human and automatic modularizations of process models to enhance their comprehension. *Information Systems, 36*(5), 881-897.

Reinhartz-Berger, I., Itzik, N., & Wand, Y. (2014). Analyzing variability of software product lines using semantic and ontological considerations. In M. Jarke, J. Mylopoulos, C. Quix, C. Rolland, Y. Manolopoulos, H. Mouratidis & J. Horkoff (Eds.), *Advanced information systems engineering: CAiSE 2014 proceedings* (pp. 150-164). New York, NY: Springer.

Reinhartz-Berger, I., Sturm, A., & Wand, Y. (2013). Comparing functionality of software systems: An ontological approach. *Data & Knowledge Engineering, 87*, 320-338.

Ricoeur, P. (1975). Biblical hermeneutics. *Semeia, 4*, 27-148.

Riemer, K., Johnson, R. B., Hovorka, D. S., & Indulska, M. (2013, December). *Challenging the philosophical foundations of modeling organizational reality: the case of process modeling*. Paper presented at the 34th International Conference on Information Systems, Milan, Italy.

Rittgen, P. (2006). A language-mapping approach to action-oriented development of information systems. *European Journal of Information Systems, 15*(1), 70-81.

Rockwell, S., & Bajaj, A. (2004). COGEVAL: Applying cognitive theories to evaluate conceptual models. In K. Siau (Ed.), *Advanced Topics in Database Research* (Vol. 4, pp. 255-282). Hershey, PA: Idea Group.

Rohde, F. (1995). An ontological evaluation of jackson's system development model. *Australasian Journal of Information Systems, 2*(2), 77-87.

Rosemann, M., & Green, P. (2000, December). *Integrating multi-perspective views into ontological analysis*. Paper presented at the 21st International Conference on Information Systems, Brisbane, Australia.

Rosemann, M., & Green, P. (2002). Developing a meta model for the Bunge-Wand-Weber ontological constructs. *Information Systems, 27*(2), 75-91.

Rosemann, M., Green, P., & Indulska, M. (2004a). A reference methodology for conducting ontological analyses. In H. Lu, W. Chu, P. Atzeni, S. Zhou & T. W. Ling (Eds.), *Conceptual modeling: ER 2004 proceedings* (pp. 110-121). New York, NY: Springer.

Rosemann, M., Green, P., & Indulska, M. (2004b, June). *Towards an enhanced methodology for ontological analyses*. Paper presented at the 16th International Conference on Advanced Information Systems Engineering Forum, Riga, Latvia.

Rosemann, M., Recker, J., Green, P., & Indulska, M. (2009). Using Ontology for the representational analysis of process modeling techniques. *International Journal of Business Process Integration and Management, 4*(4), 251-265.

Rosemann, M., Vessey, I., & Weber, R. (2004, December). *Alignment in enterprise systems implementations: The role of ontological distance*. Paper presented at the 25th International Conference on Information Systems, Washington DC.

Rosemann, M., & Wyssusek, B. (2005, August). *Enhancing the expressiveness of the Bunge–Wand–Weber ontology*. Paper presented at the 11th Americas Conference on Information Systems, Omaha, NE.

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations. *European Journal of Information Systems, 23*(3), 241-255.

Saghafi, A., & Wand, Y. (2014, January). *Do ontological guidelines improve understandability of conceptual models? A meta-analysis of empirical work*. Paper presented at the 47th Hawaii International Conference on System Sciences, Waikoloa, HI.

Searle, J. R. (1995). *The construction of social reality*. New York, NY: Free Press.

Searle, J. R. (2006). Social ontology: Some basic principles. *Anthropological Theory, 6*(1), 12-29.

Searle, J. R. (2010). *Making the social world: The structure of human civilization*. Oxford, UK: Oxford University Press.

Shanks, G., Moody, D. L., Nuredini, J., Tobin, D., & Weber, R. (2010). Representing classes of things and properties in general in conceptual modelling: An empirical evaluation. *Journal of Database Management, 21*(2), 1-25.

Shanks, G., Nuredini, J., Moody, D. L., Tobin, D., & Weber, R. (2003, June). *Representing things and properties in conceptual modelling: An empirical evaluation*. Paper presented at the 11th European Conference on Information Systems, Naples, Italy.

Shanks, G., Nuredini, J., & Weber, R. (2005). Evaluating conceptual modeling practices: composites, things, properties. In P. Green & M. Rosemann (Eds.), *Business systems analysis with ontologies* (pp. 28-55). Hershey, PA: IGI Global.

Shanks, G., Tansley, E., Nuredini, J., Tobin, D., & Weber, R. (2008). Representing part–whole relations in conceptual modeling: An empirical evaluation. *MIS Quarterly, 32*(3), 553-573.

Shanks, G., & Weber, R. (2012). The hole in the whole: A response to Allen and March. *MIS Quarterly, 36*(3), 965-980.

Sia, S. K., & Soh, C. (2007). An assessment of package-organisation misalignment: Institutional and ontological structures. *European Journal of Information Systems, 16*(5), 568-583.

Siau, K. (2002). The psychology of information modeling. In K. Siau (Ed.), *Advanced topics in database research.* (Vol. 1, pp. 106-118). Hershey, PA: Idea Group.

Singer, R. (2014). User centered development of agent-based business process models and notations. Retrieved from https://arxiv.org/abs/1404.2737

Singh, S. N., & Woo, C. C. (2009). Investigating business-IT alignment through multi-disciplinary goal concepts. *Requirements Engineering, 14*(3), 177-207.

Smith, S. W., & Koppel, R. (2014). Healthcare information technology's relativity problems: A typology of how patients' physical reality, clinicians' mental models, and healthcare information technology differ. *Journal of the American Medical Informatics Association, 21*(1), 117-131.

Soffer, P., & Wand, Y. (2004, June). *Goal-driven analysis of process model validity.* Paper presented at the 16th International Conference on Advanced Information Systems Engineering, Riga, Latvia.

Soffer, P., & Wand, Y. (2005). On the notion of soft-goals in business process modeling. *Business Process Management Journal, 11*(6), 663-679.

Soffer, P., & Wand, Y. (2007). Goal-driven multi-process analysis. *Journal of the Association for Information Systems, 8*(3), 175-202.

Soja, P., & Paliwoda-Pękosz, G. (2013). Comparing benefits from enterprise system adoption in transition and developed economies: an ontology-based approach. *Information Systems Management, 30*(3), 198-217.

Staib, A., Sullivan, C. M., Prins, J. B., Burton-Jones, A., Fitzgerald, G., & Scott, I. (2017). Uniting emergency and inpatient clinicians across the ED-inpatient interface: The last frontier. *Emergency Medicine Australasia, 29*(6), 740-745.

Stanford. (2017). Stanford medicine 2017 health trends report: harnessing the power of data in health. Stanford, CA: Stanford University Press.

Straub, D. W. (2012). Editor's comments: Does MIS have native theories? *MIS Quarterly, 36*(2), iii-xii.

Straub, D. W., Boudreau, M.-C., & Gefen, D. (2004). Validation guidelines for is positivist research. *Communications of the Association for Information Systems, 13*(24), 380-427.

Strohmeier, S., & Röhrs, F. (2017). Conceptual modeling in human resource management: A design research approach. *AIS Transactions on Human-Computer Interaction, 9*(1), 34-58.

Strong, D. M., & Volkoff, O. (2010). Understanding Organization-enterprise system fit: A path to theorizing the information technology artifact. *MIS Quarterly, 34*(4), 731-756.

Sullivan, C. M., Staib, A., Khanna, S., Good, N. M., Boyle, J., Cattell, R., . . . Scott, I. (2016). The national emergency access target (NEAT) and the 4-hour rule: Time to review the target. *Medical Journal of Australia, 204*(9), 354e351-354e355.

Tegarden, D. P., Schaupp, L. C., & Dull, R. (2013). Identifying ontological modifications to the resource-event-agent (REA) enterprise ontology using a Bunge-Wand-Weber ontological evaluation. *Journal of Information Systems, 27*(1), 105-128.

Templier, M., & Paré, G. (2015). A framework for guiding and evaluating literature reviews. *Communications of the Association for Information Systems, 37*(6), 112-137.

Thomas, M., & Dhillon, G. (2012). Interpreting deep structures of information systems security. *The Computer Journal, 55*(10), 1148-1156.

Tollington, T., & Spinelli, G. (2012). Applying Wand and Weber's surface and deep structure approaches to financial reporting systems. *Abacus, 48*(4), 502-517.

Tziallas, G., & Theodoulidis, B. (2003, September). *Building autonomic computing systems based on ontological component models and a controller synthesis algorithm*. Paper presented at the 14th International Workshop on Database and Expert Systems Applications, Prague, Czech Republic.

van Kleef, N., Noltes, J., & van der Spoel, S. (2010). Success factors for augmented reality business models *Study Tour Pixel 2010* (pp. 1-36). Twente, Netherlands: The University of Twente.

Vera, A. H., & Simon, H. A. (1993a). Situated action: A symbolic interpretation. *Cognitive Science, 17*(1), 17-48.

Vera, A. H., & Simon, H. A. (1993b). Situated action: Reply to reviewers. *Cognitive Science, 17*(1), 77-86.

Verdonck, M., Gailly, F., de Cesare, S., & Poels, G. (2015). Ontology-driven conceptual modeling: A systematic literature mapping and review. *Applied Ontology, 10*(3-4), 197-227.

Veres, C., & Hitchman, S. (2002, June). *Using psychology to understand conceptual modelling*. Paper presented at the 10th European Conference on Information Systems, Gdansk, Poland.

Veres, C., & Mansson, G. (2005). Cognition and modeling: Foundations for research and practice. *Journal of Information Technology Theory and Application, 7*(1), 93-104.

Vessey, I., Ramesh, V., & Glass, R. L. (2002). Research in information systems: An Empirical study of diversity in the discipline and its journals. *Journal of Management Information Systems, 19*(2), 129-174.

Volkoff, O., & Strong , D. M. (2018). Affordance theory and how to use it in IS research. In R. Galliers & M.-K. Stein (Eds.), *The Routledge companion to management information systems* (pp. 232-246). New York, NY: Routledge.

Volkow, N. D., & McLellan, A. T. (2016). Opioid abuse in chronic pain: Misconceptions and mitigation strategies. *The New England Journal of Medicine, 374*, 1252-1263.

vom Brocke, J., Braccini, A. M., Sonnenberg, C., & Spagnoletti, P. (2014). Living IT infrastructures: An ontology based approach to align IT infrastructure capacity and business needs. *International Journal of Accounting Information Systems, 15*(3), 246-274.

Wand, Y., Monarchi, D. E., Parsons, J., & Woo, C. C. (1995). Theoretical foundations for conceptual modelling in information systems development. *Decision Support Systems, 15*(4), 285-304.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86-95.

Wand, Y., & Weber, R. (1988, Japan). *An ontological analysis of some fundamental information systems concepts*. Paper presented at the 9th International Conference on Information Systems, Minneapolis, MN.

Wand, Y., & Weber, R. (1989). An ontological evaluation of systems analysis and design methods. In E. D. Falkenberg & P. Lindgreen (Eds.), *Information system concepts: An in-depth analysis: Proceedings of the IFIP TC 8/WG 8.1 Working Conference on Information System Concepts* (pp. 79-107). Amsterdam: Elsevier.

Wand, Y., & Weber, R. (1990a). Mario Bunge's ontology as a formal foundation for information systems concepts. In P. Weingartner & G. J. W. Dorn (Eds.), *Studies on Mario Bunge's treatise* (pp. 123-149). Amsterdam: Rodopi.

Wand, Y., & Weber, R. (1990b). An ontological model of an information system. *IEEE Transactions on Software Engineering, 16*(11), 1282-1292.

Wand, Y., & Weber, R. (1990c, December). *Towards a theory of the deep structure of information systems*. Paper presented at the 11th

International Conference on Information Systems, Copenhagen, Denmark.

Wand, Y., & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Journal of Information Systems, 3*(4), 217-237.

Wand, Y., & Weber, R. (1995). On the deep structure of information systems. *Information Systems Journal, 5*(3), 203-223.

Wand, Y., & Weber, R. (2002). Research commentary: Information systems and conceptual modeling—A research agenda. *Information Systems Research, 13*(4), 363-376.

Wand, Y., & Weber, R. (2006). On ontological foundations of conceptual modeling: A response to Wyssusek. *Scandinavian Journal of Information Systems, 18*(1), 127-138.

Wand, Y., Woo, C. C., & Wand, O. (2008). Role and request based conceptual modeling: A methodology and a CASE tool. In Q. Li, S. Spaccapietra, E. S. K. Yu & A. Olivé (Eds.), *Conceptual modeling: ER 2008 proceedings* (pp. 540-541). New York, NY: Springer.

Weber, R. (1987). Toward a theory of artifacts: A paradigmatic basis for information systems research. *Journal of Information Systems, 1*(2), 3-19.

Weber, R. (1997). *Ontological foundations of information systems*. Melbourne, Australia: Coopers & Lybrand and the Accounting Association of Australia and New Zealand.

Weber, R. (2003). Conceptual modelling and ontology: possibilities and pitfalls. *Journal of Database Management, 14*(3), 1-20.

Weber, R., & Zhang, Y. (1996). An analytical evaluation of NIAM's grammar for conceptual schema diagrams. *Information Systems Journal, 6*(2), 147-170.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii-xxiii.

Weed, L. L. (1968). Medical records that guide and teach. *New England Journal of Medicine, 278*(11), 593-600.

Whitt, N., Harvey, R., McLeod, G., & Child, S. (2007). How many health professionals does a Patient see during an average hospital stay? *New Zealand Medical Journal, 120*(1253), 1-8.

Wright, A., Maloney, F. L., & Feblowitz, J. C. (2011). Clinician attitude toward and use of electronic problem lists: A thematic analysis. *BMC Medical Informatics and Decision Making, 11*(36), 1-10.

Wyssusek, B. (2004, August). *Ontology and Ontologies in Information Systems Analysis and Design: A Critique*. Paper presented at the 10th Americas Conference on Information Systems, New York, NY.

Wyssusek, B. (2006). On ontological foundations of conceptual modelling. *Scandinavian Journal of Information Systems, 18*(1), 63-80.

Wyssusek, B., & Klaus, H. (2005). On the foundation of the ontological foundation of conceptual modeling grammars: The construction of the Bunge-Wand-Weber ontology. In J. Castro & E. Teniente (Eds.), *Proceedings of CAiSE'05 Workshops* (Vol. 2, pp. 583-594). Porto, Portugal: Faculdade de Engenharia da Universade do Porto.

Yang, A., & Marquardt, W. (2009). An ontological conceptualization of multiscale models. *Computers and Chemical Engineering, 33*(4), 822-837.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artifical Intelligence, 17*(5-6), 375-381.

Zuboff, S. (1998). *In the age of the smart machine: The future of work and power*. New York, NY: Basic Books.

zur Muehlen, M., & Indulska, M. (2010). Modeling languages for business processes and business rules: A representational analysis. *Information Systems, 35*(4), 379-390.

zur Muehlen, M., Recker, J., & Indulska, M. (2007, October). *Sometimes less is more: Are process modeling languages overly complex?* Paper presented at the 11th International IEEE EDOC Conference Workshop, Annapolis, MD.

Appendix A. Main Concepts of Representation Theory

Appendix A1. Fundamental Concepts and Premises

According to RT, the essence of an IS is that it provides a *representation* of other real-world phenomena (the *focal real-world phenomena*). To the extent that IS provide more faithful representations of the focal real-world phenomena, they will be deemed more useful (Figure A1). Humans can then employ them to obtain knowledge about real-world phenomena without having to observe the phenomena directly. Thus, RT predicts that IS will be built and used whenever they provide more cost-effective ways of obtaining knowledge about real-world phenomena than observing the phenomena directly. For instance, an organization might conclude a more cost-effective way to obtain information about customers' needs (the focal real-world phenomena) is to build an online order system that customers can use to indicate their needs rather than to have salespeople regularly visit customers to determine their needs.

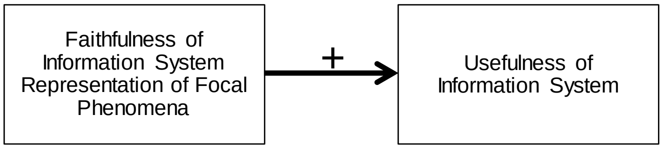


Figure A1. Impact of Representational Faithfulness on the Usefulness of an IS

RT focuses on the "deep structure" of an IS rather than its "surface structure" and "physical structure" (Wand & Weber, 1995). The *deep structure* is defined as the characteristics of the IS that manifest the *meaning* of its focal real-world phenomena (as perceived by stakeholders). The *surface structure* is defined as the characteristics of the IS that manifest the ways in which users can engage with the deep structure. The *physical structure* is defined as the characteristics of the IS that manifest the hardware/software platform used to implement the deep and surface structures (Figure A2).

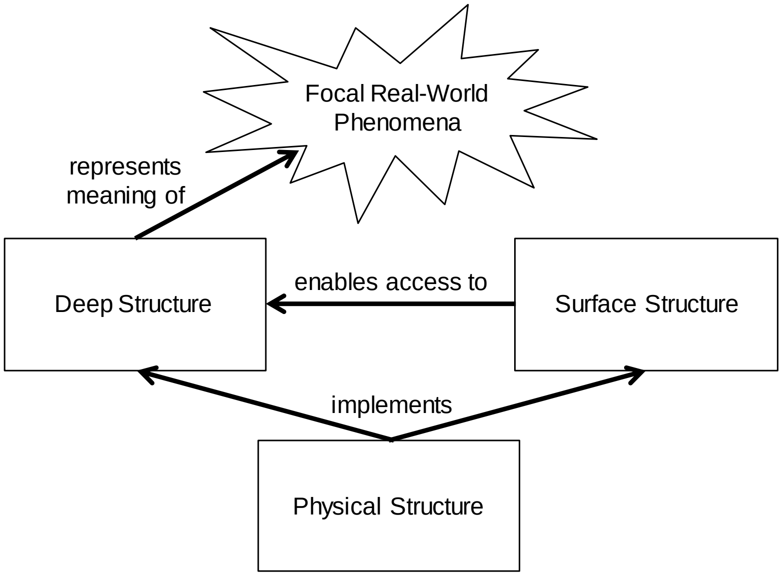


Figure A2. Three Structures of an IS

RT is concerned with *the extent to which the deep structure of an IS remains a faithful representation of someone's or some group's perception of the meaning of the focal real-world phenomena*. The perception of the meaning of the focal real-world phenomena is taken as *given*. RT recognizes that eliciting this meaning from the stakeholders in an IS can be a complex, difficult task—one that continues to unfold throughout the system development process, one that often requires negotiations among the stakeholders in an IS, and one that might never reach a clear resolution. Nonetheless, phenomena associated with this task lies outside the boundary of RT.

Appendix A2. The Representation Model (RM)

The representation model (RM) conceptualizes IS as manifested in the form of *scripts* generated via grammars. At different stages in the IS development process, different scripts are generated and then transformed to move the scripts progressively toward one that can be read and enacted by a machine (Figure A3). In practice, this script transformation process might be iterative rather than purely sequential. Nonetheless, *the fundamental requirement of IS development according to the RM is that the meaning (deep structure) of the focal real-world phenomena be preserved across the different forms of scripts*. Otherwise, the RM predicts that any loss of deep-structure representational fidelity in a script will lead to the implemented IS being deemed less useful.

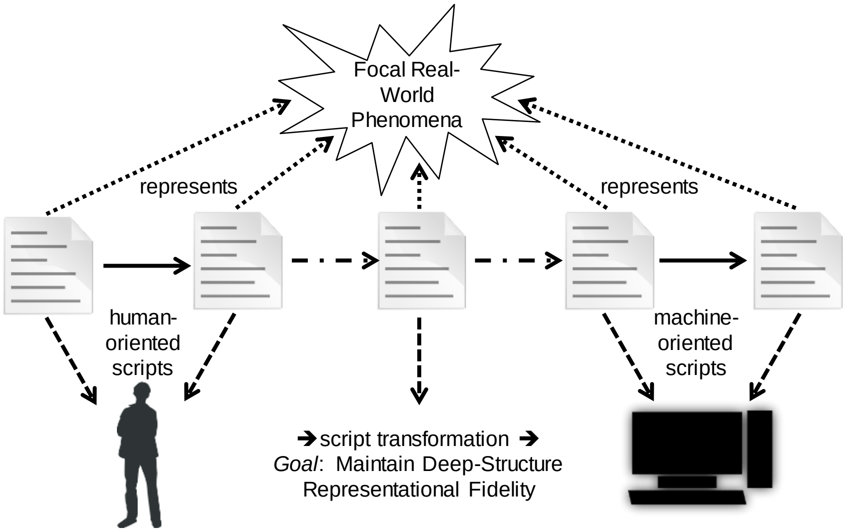


Figure A3. Transforming Human-Oriented Representations of Focal Real-World Phenomena to Machine-Oriented Representations

For instance, in the early stages of the system development process, modeling grammars such as UML and BPMN are often used to generate scripts that provide representations of the focal real-world phenomena that can be understood readily by humans. At later stages in the IS development process, programming grammars such as Java or BPEL might be used to generate scripts that can be read and transformed by a compiler or interpreter into machine-readable code. Whatever the form of a script, the RM requires that it maintain representation fidelity in relation to the deep structure of the focal real-world phenomena.

The RM-motivated efforts to identify the properties of IS scripts that preserve *deep-structure representational fidelity* (e.g., Wand & Weber, 1995; Weber, 1997). An initial focus was the *grammars* employed to generate human-oriented representations of a person’s or a group’s perception of the semantics of a focal real-world domain (Wand & Weber, 1993). A primary concern was the ability of these grammars to generate scripts that could represent the perceived domain semantics clearly and completely. In part, the motivation to focus on these grammars arose from frustrations occurring among researchers concerned with modeling grammars about the frequency with which new grammars appeared and the inadequate means they had to evaluate these new grammars against existing grammars—the so-called YAMA (yet another modeling approach) problem (Siau, 2002).

For such an evaluation to be possible, however, some type of benchmark was needed against which a grammar and the scripts a grammar might generate could be compared. The approach chosen in the RM was to use a philosophical theory of ontology as a benchmark—a theory about the types of phenomena that exist in the real world (Wand & Weber, 1993). Constructs in a modeling grammar could then be mapped against constructs in a theory of ontology. When such a mapping is undertaken, the RM indicates four problematic situations can arise (Wand & Weber, 1993, pp. 228-233) (Figure A4):

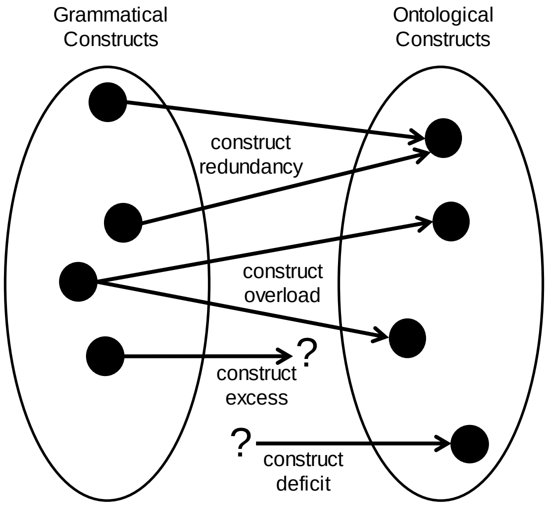


Figure A4. Types of Defects in Conceptual-Modeling Grammars

- *Construct redundancy*: Two or more grammatical constructs map to a single ontological construct. Scripts then might be produced that contain instances of the redundant constructs. As a result, users of the scripts might be confused about whether different instances of redundant constructs represent different types of real-world phenomena.
- *Construct overload*: A single grammatical construct maps to two or more ontological constructs. As a result, users of scripts that contain instances of the overloaded construct might be confused about which type of real-world phenomenon is being represented by each instance of the construct.
- *Construct excess*: A grammatical construct exists for which no ontological construct exists. As a result, users of a script that contains an instance of the excess construct might be confused about the nature of and type of real-world phenomena being represented by the construct.
- *Construct deficit*: An ontological construct exists for which no grammatical construct exists. As a result, the grammar is unable to generate a script that represents this type of real-world phenomenon if it were to occur in the focal domain.

The RM predicts that the existence of instances of construct redundancy, overload, and excess in a grammar undermine the clarity of the meaning of scripts produced using the grammar (when instances of such constructs existed in the script) (Wand & Weber, 1993, pp. 228-233). Similarly, instances of construct deficit in a grammar undermine the *completeness* of scripts produced using the grammar (when instances of the types of phenomena that could not be described using the grammar exist in the focal real-world domain) (Wand & Weber, 1993, pp. 226-228).

In short, the RM predicts that a grammar's representational fidelity depends on its ability to model the focal real-world phenomena clearly and completely (Figure A5). The RM thereby provides a means to evaluate whether the deep structure of the focal real-world phenomena can be and is preserved as the script-transformation process occurs during the IS development process.

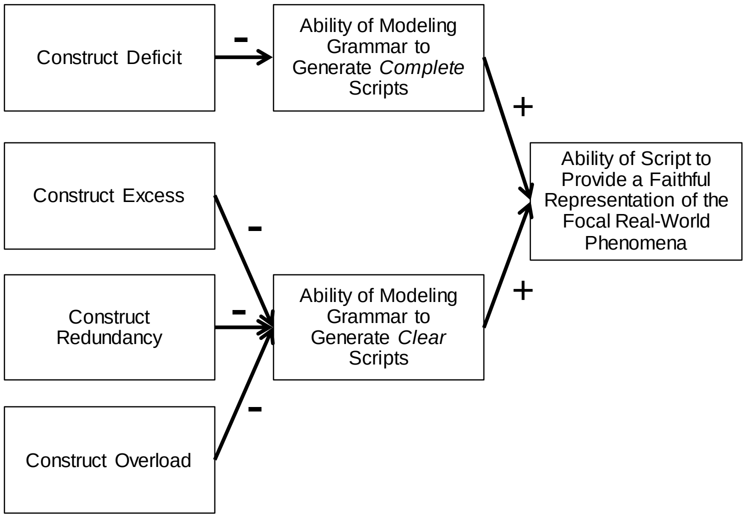


Figure A5. Impact of Construct Redundancy, Construct Overload, Construct Excess, and Construct Deficit on Grammar's Ability to Generate Clear and Complete Scripts

The fundamental ideas of construct deficit, redundancy, overload, and excess are not tied to a particular ontological benchmark. While Bunge's (1977, 1979) ontology often has been employed as a benchmark, other benchmark ontologies can be used. Indeed, the merits of a particular ontological benchmark can be evaluated in terms of whether predictions about the deleterious effects of construct redundancy, overload, excess, and deficit in any grammar that has been evaluated using the benchmark are borne out in practice. In this regard, over time some attempts have been made to use different ontological benchmarks in the evaluation of modeling grammars and scripts (e.g., Milton and Kazmierczak, 2004).

Appendix A3. The State-Tracking Model (STM)

The state-tracking model (STM) articulates four *necessary and sufficient* conditions that an IS must satisfy if it is to continue to faithfully *track* its focal real-world phenomena (Wand & Weber, 1995, pp. 211-213)—that is, maintain an accurate and complete representation of things in the focal phenomena as they change. Unless faithful state tracking occurs, the STM predicts the usefulness of the IS will decline because it no longer represents the unfolding *meaning* of its focal real-world phenomena as events occur (Figure A6).

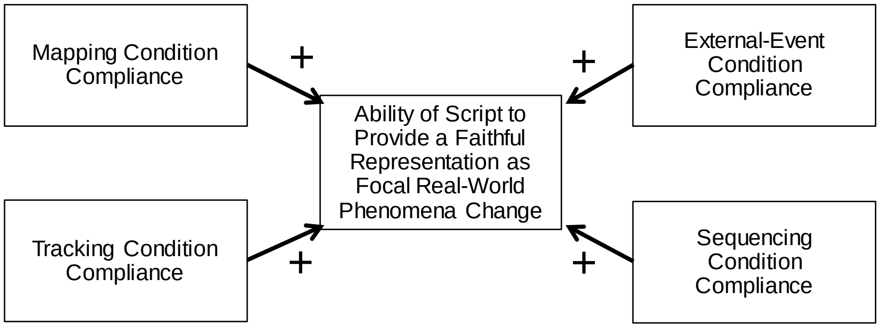


Figure A6. Faithful Representation of Unfolding Meaning of Focal Real-World Phenomena through Compliance with State-Tracking Conditions

The four state-tracking conditions are:

1. *Mapping condition*: Each state in the focal real-world phenomena must map to at least one state in the IS.

If IS are to be useful, it must be possible to tell the state of the focal real-world phenomena based on a state of the IS. *Multiple* states of the IS may map to a single state of the focal real-world phenomena. The reason is that the IS often will have state variables that do not represent focal real-world phenomena. For instance, state variables might be used to record transaction-queue lengths so the IS can operate more efficiently. Depending on queue lengths at a particular time, different states of the IS may map to the same state of the focal real-world phenomena. Nonetheless, the mapping condition requires that it must always be possible to determine the state of the focal real-world phenomena based on a state of the IS.

2. *Tracking condition*: When things in the focal real-world phenomena change states as a result of events that are *internal* to the phenomena, the IS must change from a state that faithfully represents the initial state of the thing in the focal real-world phenomena to a state that faithfully represents the subsequent state of the thing.

Events that are *internal* to the focal real-world phenomena arise as a result of transformation laws that exist within these phenomena. These laws are enacted (or fire) when the phenomena are in an *unstable state*. They transform the focal real-world phenomena to a *stable state* once again (perhaps through a series of unstable intermediate states).

The tracking condition requires that the transformation laws embedded within the IS change the state of the IS so any intermediate states and the resulting stable state in the focal real-world phenomena are faithfully represented by the IS. In short, the tracking condition is intended to ensure the IS faithfully represents *internal events* within the focal phenomena.

3. *External-event condition*: When an external event occurs in the focal real-world phenomena, an external event that is a faithful representation of the real-world external event must occur within the IS.

An external event in the focal real-world phenomena is a change of state that arises in some thing in the phenomena by virtue of the action of some thing in the environment of the phenomena. The resulting state may be stable or unstable. If it is unstable, transformation laws in the focal real-world phenomena are enacted to move the phenomena to a stable state.

If the IS is to track state changes in the things in the focal real-world phenomena faithfully, external events that occur to things in the focal real-world phenomena must be represented in the IS (these representations are often called the *input* to the IS). Somehow the occurrence of an external event in the focal real-world phenomena must be reported to the IS.

For instance, a customer may decide to place alarms in their house as a result of a robbery. The customer's change of state is an external event (motivated by the robbery). The customer reports her/his external event to an organization that sells alarms by entering an order into the organization's order entry system. This external event in the IS then triggers internal events within the IS that are intended to mirror those in the focal real-world phenomena (the customer's desire to have an alarm).

4. *Sequencing condition*: External events that occur in the IS must follow the same sequence as external events that occur in the focal real-world phenomena.

External events that occur to things in the focal real-world phenomena may or may not be reported immediately to the IS implemented to represent the phenomena. For instance, time delays may arise as records of the external events are made and transported across communication networks to the IS. As a result, the records might arrive in a sequence that differs from the sequence of external events that occurred in the focal real-world phenomena.

If the sequence of external events that occur in the IS differs from the sequence of external events that occur in the focal real-world phenomena, the sequence of internal events triggered in the IS will differ from the sequence of internal events triggered in the focal real-world phenomena. As a result, the IS will no longer provide a faithful representation of the focal real-world phenomena. The sequencing condition is intended to prevent such an outcome occurring by ensuring that the sequence of external events in the focal real-world phenomena matches the sequence of external events in the IS.

Appendix A4. The Good-Decomposition Model (GDM)

The GDM's primary focus is on how different types of decompositions of real-world phenomena into systems and their subsystems facilitate or inhibit an individual's ability to understand the meaning of real-world phenomena (Wand & Weber, 1995, pp. 213-215). Consistent with Bunge (1979, pp. 13-14), decompositions have an ontological status under the GDM—in other words, they *exist* in the real world. Some types of decompositions in the real world are fairly apparent. For instance, it is clear that a human body can be decomposed into constituent parts—arms, legs, head, and so on. Other types of decompositions that might exist in the real world are not always apparent.

Under the GDM, good systems decompositions (those that are best able to communicate the meaning of the real-world phenomena they are intended to represent) satisfy *five* conditions (Wand & Weber, 1995, pp. 213-215; Weber, 1997, pp. 152-163) (Figure A7). If IS are designed and implemented to manifest systems decompositions that satisfy these five conditions, the GDM predicts they will better convey meaning about their focal real-world phenomena. In turn, these systems will be deemed more useful.

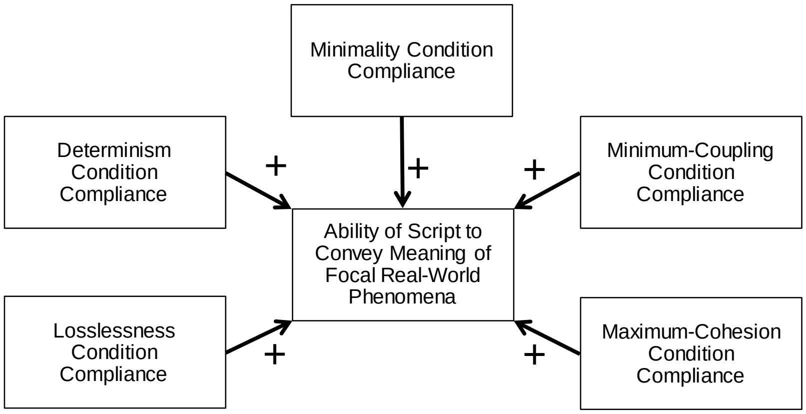


Figure A7. Ability of Representation to Convey Meaning of Focal Real-World Phenomena Through Compliance with Good-Decomposition Conditions

The *five* good-decomposition conditions are:

1. *Minimality condition*: A system decomposition is good *only if* for every subsystem at every level in the level structure of the system decomposition representing the focal real-world phenomena no redundant state variables describing the subsystem exist.

A subsystem in a system decomposition is a representation of some subset of the focal real-world phenomena. The representation is implemented via state variables that describe the subset of the focal real-world phenomena. The minimality condition requires that only the minimum number of state variables needed to represent the subset of the focal real-world phenomena be used to describe the subsystem. In other words, all state variables will be needed at some time to determine some aspect of the meaning of the focal real-world phenomena (none are redundant in term of capturing this meaning). If redundant state variables exist, users of the decomposition may become confused about the meaning of the state variables and the real-world phenomena they are intended to represent.

2. *Determinism condition*: For a given set of external events in the focal real-world phenomena, a system decomposition is good *only if* for every subsystem at every level in the level structure of the system decomposition representing the focal real-world phenomena an event is either (1) an external event, or (2) a well-defined internal event.

The subsequent state of a system that arises from an external event cannot always be predicted (the nature of the real world is such that the impact of the environment on a system cannot always be foreseen). Thus, external events are sometimes but not always well defined. In good system decompositions, however, all internal events will be well defined. If the event's prior state is known, its subsequent state can be predicted with certainty.

3. *Losslessness condition*: A system decomposition is good *only if* every hereditary property and every emergent property in the focal real-world phenomena is preserved in the system decomposition.

A hereditary property of a system is a property that is also possessed by one of its subsystems (e.g., the processor speed of a laptop computer is the same as the processor speed of its processor subsystem). An emergent property is one that is possessed by the system but not by any of its subsystems (e.g., a user's perceptions of the "power" of a laptop computer is some function of the capabilities of its various component subsystems).

The meaning of the focal real-world phenomena is manifested in the nature of and values of hereditary and emergent properties associated with things in the phenomena. Thus, under the GDM, faithful representations of focal real-world phenomena must have state variables that correspond to all hereditary and emergent properties associated with the phenomena.

4. *Minimum-coupling condition*: A system decomposition has minimum coupling if and only if the *cardinality* of the totality of input for each subsystem of the decomposition is less than or equal to the *cardinality* of the totality of input for each equivalent subsystem in the equivalent decomposition.

The minimum coupling condition involves a comparison of *equivalent decompositions* of a system—in other words, each decomposition has the same set of subsystems with the same components in each subsystem. The only difference between equivalent decompositions is their structures (the couplings that each component in a subsystem has with other components in its subsystem and other components in its environment). Unless a comparison occurs between equivalent decompositions, it is easy to create a decomposition that has less coupling than another decomposition simply by combining subsystems into one system. The combined subsystems may be less internally cohesive, however, which undermines the quality of the decomposition (see Condition 5 below).

The cardinality of the totality of input for a subsystem is the number of states that arise by virtue of the action of things in the environment of the subsystem. These are the states that occur in the subsystem by virtue of the action of external events. Because external events often are not well defined, these subsequent states cannot always be predicted based on the prior states of the subsystem. Thus, they convey less meaning (the transformation that gives rise to the event is not always clear). The motivation behind the minimum-coupling condition, therefore, is to minimize the number of external events in a subsystem.

5. *Maximum-cohesion condition*: A subsystem is maximally cohesive if the addition of another output state variable to its existing set of output state variables does not extend the set of its input state variables (those state variables on which its existing set of output state variables depend).

In essence, a subsystem is cohesive when its output state variables cannot be partitioned based on a partition of the input domain. Thus, the subsystem has no transformation that has a set of input state variables that does not overlap with the set of input state variables for at least one other transformation. As the number of independent transformations decreases, the subsystem's cohesion increases and its deep structure (meaning) becomes clearer.

Cohesion is the duality of coupling. Having a subsystem perform more functions reduces its interactions with other subsystems, which in turn reduces coupling. As a subsystem performs more functions, however, the likelihood that the functions are unrelated to each other increases. As a result, the subsystem's meaning becomes less clear.

Appendix B: Literature Review Materials

Table B1. Summary of Categorization Scheme

Category	**Selected criteria**
Focus and intent	What is the stated *research goal*? Which phenomena is the *focus of the paper*? How does the paper *refer to the theory*? • Conceptual foundation • Test of theory • Critique of theory • Extension of theory • Reference to theory • Other Does the paper report on potential *theoretical or methodological advances*?
Application of theory to conceptual modeling[5]	Which *modeling element* is prominent in the study? • Grammar • Script • Method • Context Which *modeling approach* is examined (e.g., UML, ERD, BPMN)?
Element of theory	Which *theoretical premise* is examined primarily? • Ontological completeness • Ontological clarity • Ontological overlap • Good decomposition • State-tracking quality • Other
Research method	Which *research method or approach* has been used? • Representational analysis • Survey • Laboratory experiment • Field study • Case study • Interviews • Design science • Other
Empirical evidence	What is the *quantity*, *quality*, and *results* of the reported evidence? Students • Practitioners • Case data • Expert panels • Other

---

[5] This category was coded only if the focus of the paper was on conceptual-modeling phenomena. We examined this category further because most research using RT falls within the conceptual-modeling domain.

Table B2. Number of Theory Tests by Year and Type

	Quantitative tests			**Qualitative tests**			**Hybrid tests**	
Year	Survey	Laboratory experiment	Field experiment	Interviews	Case Study	Expert Panel	Field study	Mixed method
1994							1	
1995								
1996		1						
1997								
1998					1			
1999		1						
2000								1
2001								
2002		2		1				
2003		3						2
2004		7		1	1			
2005		4		1	1			1
2006		3		2	1			
2007		2		1			1	
2008		4			1			
2009		2			3			
2010	2	2		1	1			
2011	1	1						1
2012		1		3	1			
2013		2					2	
2014		3	1					
2015						1		1
Total	**3**	**38**	**1**	**10**	**10**	**1**	**4**	**6**

Table B3. Empirical Evidence by Data Source and Test of Type

Data source		**Practitioners**		**Students**		**Mixed**		**Total**	
Type of test		Mean sample size	Number of studies	Mean sample size	Number of studies	Mean sample size	Number of studies	Mean sample size	Number of studies
Qualitative	Case study	22.8	9	8	1			19.8	10
	Interviews	41.5	6			15.8	4	31.7	10
	Expert Panel	5	1					5	1
Quantitative	Survey	578	2			49	1	401.7	3
	Laboratory experiment	33.5	4	96.6	29	57	5	86.1	38
	Field experiment	81	1					81	1
Hybrid	Field study	31.5	4					31.5	4
	Mixed method	154.5	2	112	2	41	2	102.5	6
Total		**133.8**	**29**	**64.5**	**32**	**40.7**	**12**	**93**	**73**

Table B4. Concerns about Empirical Tests and Recommendations for Future Research

Concern	Resulting threat	Recommendation	Examples
Lack of work on instrumentation and measurement	Instrumentation validity	Develop a repository of experimental design materials. Conduct research on construct development using appropriate methodological guidelines (e.g., MacKenzie, Podsakoff, & Podsakoff, 2011).	Bodart et al. (2001) used and adapted experimental materials from prior work by Gemino (1999). Recker and Rosemann (2010) report on the development of a perception measurement instrument.
Lack of replications of studies and experiments	External validity	Replicate existing study designs. Identify outlets that welcome replication studies.	Burton-Jones and Meso (2006) report on a replication. The *AIS Transactions on Replication Research* publishes replication studies (Dennis & Valacich, 2015).
Lack of cross-sectional studies	External validity	Develop and examine predictions about behaviors across use contexts.	Green et al. (2011) studied systems development tool users.
Lack of tests of causality and temporality	Statistical conclusion validity and predictive validity	Increase the number of studies using longitudinal and repeated-measures designs.	
Selection bias: dominant use of student samples	Internal validity and external validity	Increase the number of studies involving practitioners and experts.	Milton et al. (2012) specifically engaged experts in their experiment.
Lack of multi- and cross-level analyses	Measurement validity	Increase number of studies focusing on social processes (e.g., through agent-based modeling).	
Lack of formalized hypotheses Large share of papers without a priori expectations—many exploratory studies	Statistical conclusion validity and nomological validity	Increase expectations in the community to offer logically developed a priori expectations in the form of hypotheses and/or propositions.	Burton-Jones and Meso (2008) offer a set of formalized as well as visualized expectations about good-decompositions. Burton-Jones, Wand, and Weber (2009) offer a framework that identifies research spaces to generate further hypotheses about the semantics, syntax, and pragmatics of modeling grammars.
Lack of alternative hypotheses testing Few tests of alternative hypotheses, even in experimental designs	Statistical conclusion validity	Develop study designs that consider alternative hypotheses from comparable rival theories such as cognition or semiotics.	
Lack of studies that examine theory in relation to immediate conceptual frameworks Comparisons with previous magnitude measures missing (e.g., comparison of alternative path coefficients)	Nomological validity and factorial validity	More studies that involve advanced statistical analysis.	Recker et al. (2011) use SEM analysis to examine perceived usefulness and ease of use of a modeling grammar.

Table B5. Elements of the RT Research Program and Criticisms in the Literature

Element of the research program	Criticisms in the literature	Responses in the literature
1. *The representation assumption:* An IS provides a representation of users' perception of some real-world domain.	• It is not possible to represent the world accurately and completely.	• The representation assumption still applies.
	• Much of what we might want to represent in an IS is embodied and tacit (Hovorka, Johnston, & Riemer, 2014; Riemer, Johnson, Hovorka, & Indulska, 2013, p. 10) and a social construction (El-Tawy & Tollington, 2013), so we either cannot discover and model it, or we cannot do so in an unbiased/accurate way.	• The representation assumption does not imply omniscience. It simply implies that the role of the IS is to represent users' perceptions of the world (whether or not those perceptions are correct and unbiased) (Wand & Weber, 2006, p. 130).
	• IS and IS scripts do not just represent reality; they enact it.	• Representation is still critical for enactment.
	• Representation is only one function of a language. Assuming it is the sole purpose is known as the "descriptive fallacy" or the "representational fallacy", which has been critiqued in the artificial intelligence (AI) literature. Other functions of language and IS are to act or perform; to bring about change in the world or get something done (Ågerfalk, 2010, p. 252; Allen & March, 2006a, p. 4; Klein & Hirschheim, 2006, pp. 312-313; Riemer et al., 2013, pp. 11-12).	• An IS might well take action in the world, and a user might well take action in the world by creating or reading an IS script. Either way, the concepts encoded in the IS/script will affect the action and its outcome. Thus, irrespective of the performative or enactive nature of systems and scripts, the concepts used to represent the state of the world (as viewed by stakeholders in it) are still vital (Wand & Weber, 2006, p. 135). Vera and Simon (1993a, 1993b) provide a similar response in the AI literature.
2. *The ontological assumption:* Ontological theories can give us insights into the nature of the real world and thus users' perceptions of the real world.	• Modeling peoples' conceptions and perceptions is not the same as modeling reality.	• This criticism does not obviate the usefulness of ontology.
	• Ontology would be relevant if we were modeling reality, but we are modeling peoples' perceptions or conceptions of reality, so we should refer to theories of human conceptions/perceptions, not ontology (Wyssusek, 2006, pp. 65, 74).	• The only way we can talk about the real world is via concepts. Ontological theories are useful because they provide us with concepts that we can use to conceive and model the real world (Wand & Weber, 2006, p. 132).
	• Other theories could be studied instead of ontological theory. Many alternatives have been suggested, such as semantics (Rosemann & Wyssusek, 2005; Wyssusek, 2006), epistemology (Milton, 2007; Wyssusek, 2006), domain ontologies (Chen & Pooley, 2009a, 2009b; Wyssusek, 2004), linguistics (Bjeković, Proper, & Sottet, 2014; Recker, 2005; Wyssusek & Klaus, 2005), and cognition (Allen & March, 2006a; Veres & Hitchman, 2002; Veres & Mansson, 2005).	• Other theories can be used with ontology; it need not be an either/or situation. Alternative approaches (such as linguistics and cognition) can be used together with the ontological approach (Wand, Monarchi, Parsons, & Woo, 1995). This does not render ontology any less useful.
3. *Use of Bunge's ontology:* Bunge's ontology provides a useful ontological benchmark for the purpose of mapping.	• Insufficient justification for choosing Bunge's ontology and its particular constructs.	• The appropriate justification is empirical and context dependent (Recker & Niehaves, 2008).
	• Researchers provide little justification for their choice of ontology (Johnston & Milton, 2002, p. 41). Bunge's ontology appears to have been chosen for convenience, not philosophical appropriateness (Wyssusek, 2006, p. 71). Other ontologies could have been chosen	• Bunge's ontology was chosen because it appeared to offer a way to define and model IS constructs. Rather than have a complete philosophical justification for it or the constructs chosen from it at the outset, it was more relevant to test their usefulness in an IS context (Opdahl,

Table B5. Elements of the RT Research Program and Criticisms in the Literature

	instead (Ågerfalk, 2010, p. 253; Allen & March, 2006a, p. 5; Milton, 2007, p. 129), or in combination (such as combining Bunge's with Searle's) (Lemieux & Limonad, 2011, p. 34). Moreover, insufficient justification has been given for the particular subset of constructs chosen from Bunge's ontology (Rosemann & Wyssusek, 2005, p. 2804). • Bunge's ontology applies to material reality, not institutional reality (March & Allen, 2014). • In IS, conceptual models are used to model organizational domains. Organizations are institutional (socially constructed) entities, not material (physical) entities—they are a product of human intention and convention. Because Bunge's ontology applies only to material reality, it is inappropriate for conceptual modeling in IS (Allen & March, 2006a, p. 1; Wyssusek, 2004, p. 4304). It lacks constructs needed to represent the social world, such as intention (Allen & March, 2006a), social agency (vom Brocke, Braccini, Sonnenberg, & Spagnoletti, 2014), and culture (Herrera, Pallioto, Tkachuk, & Luna, 2005)). Even trivial examples such as "*a student attends a university*" cannot be modeled using Bunge's ontology as students and universities are part of institutional reality (Allen & March, 2006a, p. 3). • One response might be that if stakeholders in an organization perceive of something in social reality to be a real-world entity, then it can be modeled as such, but this is not consistent with Bunge's ontology (Allen & March, 2006a, p. 5). It amounts to using Bunge's ontology without its ontological commitment (Ågerfalk, 2010, p. 252), which renders Bunge's ontology just a language, not an ontology (Wyssusek, 2006, p. 73).	2006, p. 97; Wand & Weber, 2006, pp. 131-132). The ontology could be replaced, tailored, or combined with another ontology, and different constructs from it could be used or dropped over time, depending on the context (Hadar & Soffer, 2006, pp. 586-587) and the state of research (Wand & Weber, 2006, p. 133). Such a view is consistent with Bunge's own views of his ontology (Opdahl, 2006, p. 97). • Bunge's ontology can still be used to model institutional reality. • Bunge's ontology addresses natural and social science (Bunge, 1977, p. 6). Thus, it is not true that Bunge's concepts are restricted to material reality (natural science) alone. Rather, Bunge focuses on the material foundations of both natural and social reality. For instance, the psychological concept of knowledge can be modeled using Bunge's ontology in terms of the concrete actions that people take in the world to achieve their goals (Bera, Burton-Jones, & Wand, 2011). Likewise, students can be modeled as roles of people and universities as systems of learning that people (as students) can attend. • Even if Bunge's ontology was not created to describe institutional reality, it can still be used for that purpose. It does not matter if this is inconsistent with Bunge's original aims, nor even if this amounts to using Bunge's ontology as a language rather than an ontology. It can be used to describe institutional reality if it is found to be useful for that purpose—for instance, if it offers a useful way for stakeholders in a domain to understand each other's views of the domain and reach a shared understanding (Hadar & Soffer, 2006, p. 581; Wand & Weber, 2006, pp. 131-132).
4. *Results from mapping:* The results from the mapping exercise provide a valid and reliable basis for drawing implications.	• Many aspects of the mapping process seem to be subjective, including the constructs chosen from the conceptual-modeling grammar, the constructs chosen from the ontological benchmark, and the mapping itself (Gehlert & Esswein, 2005, pp. 114, 117; Herrera et al., 2005, p. 576, Perepletchikov, Ryan, & Zahir, 2013). Because of differences in opinion, different researchers can come to very different views, impeding progress (Allen	• Such limitations offer opportunities for improving the process of mapping (also known as representational analysis) (Green, Rosemann, Indulska, & Recker, 2006) and for understanding the implications of the results of mapping on dependent variables (outcomes) of interest (Recker, Rosemann et al., 2006) and the effects of pragmatics (Bera et al., 2014).

Table B5. Elements of the RT Research Program and Criticisms in the Literature

	& March, 2006a, 2012; Shanks & Weber, 2012). • The mapping is also imprecise at times. For instance, a mapping between two constructs can, in principle, be identical, distinct, or similar, but the notion of similarity (or partial satisfaction) has not been discussed sufficiently in the literature (Gehlert & Esswein, 2005, p. 118; Gregersen & Jensen, 1999, p. 120). • Finally, each deficiency identified in the mapping process is currently weighted equally, but it may be that unequal weights are more appropriate (Milton, 2007, p. 127; Krogstie, 2012, pp. 249-280) and that mappings should consider the pragmatics of the modeling context (Frank, 2013; Perepletchikov et al., 2013).	
5. *Results from empirical tests:* Empirical tests can determine the usefulness of the proposed theory.	• There is little empirical evidence to support the usefulness of Bunge's ontology in conceptual modeling (Allen & March, 2006a, p. 3); the results are mixed and sometimes weak or inconclusive (Verdonck, Gailly, de Cesare, & Poels, 2015).	• There is a fair base of empirical support for the usefulness of Bunge's ontology in conceptual modeling (Wand & Weber, 2006, p. 128). Even if the ontology itself was refuted empirically, this outcome would not invalidate the general idea of evaluating a modeling grammar by using a "benchmark ontology" (Wand & Weber, 2006, p. 135).
6. *Relevance for practice:* The results of the research offer implications for practice.	• This work has had very little interaction with industry/practice (Hadar & Soffer, 2006, p. 570, Riemer et al., 2013, pp. 3-4), and so it has little relevance for them. There has also been a tendency to adopt designs that provide "obvious" results—for instance, laboratory experiments that test "informationally inequivalent" scripts and find that a script containing more information performs better than one that contains less information. Such research, even if motivated by theory, generates little insight for practice (Parsons & Cole, 2005, p. 330).	• Neither critique is fatal; they simply offer opportunities for improved research. By working with and studying practitioners, researchers can generate new insights for developing and testing the theory (Recker et al., 2006a, Recker et al., 2011). They can also identify which aspects of informational inequivalence are obvious to practitioners and which are not. Because this research program focuses heavily on how to convey semantics, it is natural (and important) that many empirical studies use informationally inequivalent materials (Burton-Jones et al., 2009, p. 510).

Appendix C. Future Research Opportunities

Appendix C1. Hospital Case Study Showing How the GDM Might be Used to Better Understand Data-Mining Results

Patients who are admitted to the emergency department of a hospital must sometimes be transferred to a ward on an inpatient basis. This transition often proves problematic. At times, communication errors and culture clashes arise among healthcare professionals, accountabilities for patient care are unclear, and patients wait in transitory, crowded spaces where clinical equipment is not readily available (Staib et al., 2017). The dysfunctionalities that occur sometimes lead to unwanted outcomes, such as high rates of readmission and mortality among patients (Sullivan et al., 2016).

To date, the strategies and operational procedures used to improve outcomes for patients who traverse the "gray zone" between a hospital's emergency department and an inpatient ward have been difficult to identify and/or implement. One reason is that the interactions that occur among things in the domain (e.g., patients, healthcare professionals, wards, beds, equipment) are often variable, complex, and transient. For instance, in one emergency department, Whitt, Harvey, McLeod, and Child (2007) found a medical patient, on average, saw 17.8 health professionals and a surgical patient, on average, saw 26.6 health professionals. When other things in the domain are considered (e.g., beds and equipment), the systems assembled to treat patients can vary considerably in terms of their makeup and interactions. Moreover, the systems are often short-lived. As a result, we hypothesize that many escape the attention of stakeholders. Because they are not perceived or perceived only fleetingly, they are neither named nor defined, nor are their functions well managed.

One of the authors has experience through another research project with a large, public hospital that collects extensive data about phenomena associated with the transfer of patients between its emergency department and inpatient wards (to provide an audit trail for decision-making and accountability purposes). For instance, it captures demographic and health data about patients (e.g., age and symptoms), the activities of healthcare professionals (e.g., patient interventions used and handoffs), and ramping and boarding space in its emergency department and inpatient wards (e.g., size and occupancy levels). In the past, it has mined these data in the hope it could identify patterns that provide a basis for better decision-making and improved patient outcomes (e.g., reduced rates of patient mortality and readmission and reduced times for admission, transfer, and discharge from its emergency department).

To date, these data-mining projects have produced limited benefits, primarily because stakeholders have experienced difficulty distinguishing between spurious and substantive relationships and moving beyond a set of piecemeal relationships to a coherent, overall understanding of how the relationships identified fit together. The following comments by a manager at the hospital illustrate the problems that stakeholders have confronted with previous data-mining projects:

> *…he [data analyst] took our trauma data set and looked at [it and] there was 746 patients and he looked at the process for the emergency stay for all of those patients with trauma, and [the director of the Emergency Department] said to him at the start that it was a complete waste of time because they're Trauma, you're going to get 746 different processes and guess what, he did.*

In short, to address trauma needs in the emergency department of his hospital, the manager's comments imply a large number of different types of systems are formed (often created using improvisation and bricolage), the systems are often transitory, and even longer-lived systems may enact many different types of processes. The problematic data-mining results obtained by the data analyst manifest the nature of the systems used in the emergency department.

The GDM could be used to see whether it assists stakeholders to make better sense of the data-mining results obtained by the hospital. Specifically, it could be employed to construct level structures (Bunge, 1979, pp. 13-14) over the things inherently manifested in the set of relationships shown in the data-mining results. Based on the GDM, one hypothesis would be that stakeholders who use these level structures will then be better able to ascribe meaning to the data-mining results. An additional hypothesis would be that those level structures that better comply with the five good-decomposition conditions (see Appendix A4) would enable stakeholders to ascribe *more* meaning to the data-mining results. Evaluating both hypotheses would be a good way to test the GDM.

To aid in the design of such studies, Appendix C2 contains an algorithm[6] we have developed for bottom-up construction of level structures.[7] As input, the algorithm uses the set of relationships among data items detected through data-mining operations (e.g., perhaps between the number of cardiac arrests and the number of handovers among patient-care teams).[8] As output, the algorithm produces a level structure of basic things (e.g., patients), subsystems (e.g. a particular patient-care team), and systems that cover the phenomena identified via the data-mining operations (e.g., systems that employ resources from the emergency department, an inpatient ward, and the image-services department).

Once a level structure has been identified using the algorithm in Appendix C2, we propose that stakeholders should revisit each relationship identified via the data-mining operations.[9] They should assign each attribute in a relationship to the things in the level structure. When undertaking this step, they should be mindful of Condition 1 of the GDM (minimality)[10] and seek to ensure each attribute truly characterizes the thing to which it has been assigned (rather than some other thing). They should also be mindful of Condition 3 of the GDM (losslessness) and seek to ensure attributes are assigned to things at the correct level in the level structure (at first glance, it may not be clear whether an attribute represents a hereditary property or an emergent property of a thing and thus which thing in the level structure should be the assignee). For instance, the average patient-discharge time per month is an emergent property of the lowest-level patient-care team in an emergency department, but potentially it then becomes a hereditary property of some (if not all) higher-level systems of which the team is a component (e.g., that section of the emergency department that has patient-care teams to deal with pediatric emergencies and the emergency department overall). If this property is present in a relationship identified via data-mining operations, initially the component or composite to which it should be assigned may not be clear.

When stakeholders are satisfied with their assignment of attributes to things in the level structure, they should then seek to interpret the meaning of the relationship. As they undertake this step, they should be mindful of Condition 2 of the GDM (determinism) and evaluate whether the relationship is likely to lead to a deterministic outcome for any events that occur to the things that underpin the relationship. Any concerns about violations of Condition 2 (determinism) signal that the merits of the assignment of attributes to things and/or the level structure used may have to be revisited.

Use of the algorithm in Appendix C2 and subsequent interpretations of the relationships identified via data-mining operations are iterative processes. Whether they are effective in uncovering the meaning of the relationship depends somewhat on data analysts' expertise in the specific context and domain of use. Nonetheless, we hypothesize this proposed use of the algorithm in a healthcare context will provide an important test of the GDM's merits.

If use of the GDM proved to be successful, stakeholders in the hospital would then be better able to mine larger, more diverse data sets with the objective of achieving an improved understanding of phenomena associated with the emergency department-inpatient ward interface. For instance, the data set could be expanded to incorporate unstructured and multimedia data, such as social media data that patients post about their experiences in the hospital, video feeds from CCTV cameras or cameras mounted on patient beds or worn by hospital staff, and streaming data provided by smart beds about patient vitals. Once the data are cleansed (Zhang, Zhang, & Yang, 2003), machine-learning software could then be employed to extract relevant features from these data (Blum and Langley, 1997). These features could be used as input to data-mining operations (Bhatt and Kankanhalli, 2011). As before, the set of relationships identified by the operations would provide the input to the level-structure construction algorithm described in Appendix C2.

---

[6] To facilitate understanding of our algorithm and because it needs to be tested and refined, we have expressed the algorithm in natural language rather than pseudocode or a formal language.

[7] The algorithm uses a bottom-up process in the sense that it first identifies the things manifested in relationships identified via data mining and then tries to determine how these things reflect components and composites. A top-down approach to trying to understand the relationships would be to start with a putative level structure, assign attributes to each thing in the level structure, discard relationships in the data-mining results that do not include these attributes, and then try to interpret the meaning of the remaining relationships via the putative level structure.

[8] The data used in the data-mining operations could come from many sources (e.g., structured and unstructured data collected via observations from healthcare professionals and multimedia and streaming data collected from sensors on patient beds).

[9] If the data-mining operations have identified a large number of relationships (perhaps thousands), some type of sampling strategy might have to be used to obtain a subset of the relationships that is feasible to consider.

[10] Appendix A4 describes the five conditions of the GDM.

Appendix C2. Algorithm to Build a Level Structure of Systems and Subsystems Based on Data-Mining Results

Nature:

A semiformal algorithm that requires a data analyst to use judgment, heuristics, and iteration.

Input:

A set of relationships between data items identified via data-mining operations undertaken on a large data set (e.g., a data lake).

Output:

A level structure of systems and subsystems that underpins the relationships.

Objectives:

To assist data analysts to (1) separate spurious relationships from substantive relationships, and (2) make sense of the substantive relationships in the context of stakeholder objectives.

Steps:

Level structures are based on things. When devising a level structure, the attributes (properties) and relationships among attributes are considered only to the extent they help with obtaining an understanding of the level structure. At the outset, therefore, data analysts need to determine the things to which the attributes in a relationship identified via data-mining operations belong. Some of the things might be components; some might be composites. The first three steps in the algorithm therefore focus on getting clarity about the things that possess the attributes in the relationships identified via data-mining operations.

1. Construct a table, Table 1, where the rows contain the relationships identified by the data-mining operations and the columns contain the data items (attributes of things) involved in the relationship.
2. For each relationship (each row of Table 1), determine the things to which the attributes (properties) that are elements in the relationships belong. Create a new table, Table 2, by inserting the things alongside their attributes in Table 1 (add columns to Table 1 to accommodate the names of the things). Enact the minimality condition (Condition 1) of the good-decomposition model (GDM) by ensuring each attribute is assigned to its correct thing (the thing it truly characterizes) and not an extraneous thing.
3. Construct a new two-column table, Table 3. Insert each thing-attribute pair in each row of Table 2 as a row of Table 3 (thing in the first column and attribute in the second column). Sort Table 3 by “attribute” within “thing” (so that all the attributes pertaining to the same thing are in consecutive rows and the things involved in the relationships are clear).

We hypothesize that a graph of the things and their couplings will help data analysts to see possible (1) level structures over the things, and (2) groupings of the things that may reflect systems and subsystems. The first step in constructing such a graph is to draw its vertices.

4. Draw an unconnected graph, Graph 1, in which the things identified in Step 3 above are the vertices of the graph.

The next two steps of the algorithm create a table that shows the couplings that exist between things. Level structures are based on couplings—that is, relationships between attributes of different things and not relationships between attributes of the same thing. The things that are coupled with each other therefore need to be identified. Note, a single coupling may reflect one or more relationships between the things involved in the coupling. The integer n+1 in Step 6 below is a measure of the strength of the coupling between things (the number of relationships between the things).

5. Using Table 2, classify each relationship as (1) a relationship between attributes of the same thing, or (2) a relationship between attributes of different things. Create a new table, Table 4, by first deleting those rows from Table 2 that show a relationship between attributes of the same thing.
6. Create a new table, Table 5, by first deleting the attribute columns from Table 4. Then delete all but one of any rows that include the same set of things. Insert a new column next to the retained row containing the integer n+1, where n is the number of rows that have been deleted because they contain the same set of things as the retained row.

The retained rows in Step 6 represents the coupling between things. The next step of the algorithm sorts couplings relating to the same things together.

7. Sort the rows in Table 5 by the first column and then the second column within the first column, the third column within the second column, the fourth column within the third column, and so on.

We hypothesize that a graph of the things and their couplings will help analysts to see possible (1) level structures over the things, and (2) groupings of the things that may reflect systems and subsystems. The next step of the algorithm creates such a graph.

8. Using the unconnected graph created in Step 4 (Graph 1), create a connected graph, Graph 2, by using the couplings listed in Table 5 after execution of Step 7 to join the vertices (representing things) of the unconnected graph. Place each integer n+1 on its relevant coupling (edge between vertices).

The next three steps of the algorithm focus on identifying and representing couplings between composites and components. The composites and components will be at different levels of any level structure. Most likely, some couplings will exist between things within a level, but for the moment these couplings should be ignored.

9. Identify which of the couplings between different things reflects a coupling between a composite and one of its components (the coupling is underpinned by a part-of relationship between the things). The coupling might manifest a relationship between (1) an emergent property of the composite[11] and a hereditary property[12] of the component, and/or (2) hereditary properties of both the composite and the component.
10. Create a subgraph, Graph 3, of the graph in Step 8 above (Graph 2) showing only the composite-component couplings. Note that Graph 3 most likely will contain only a subset of the things and couplings identified in Step 8 and shown in Graph 2.

Because a thing belongs to a given level, the next step of the algorithm organizes both components and composites into levels. When executing this step, data analysts should be mindful of Bunge's (1979, p. 13) criterion for assigning things to levels: "A thing belongs to a given level if it is composed of things in (some or all of) the preceding levels".

11. Based on the subgraph in step 9 above (Graph 3), classify the composites and components into different levels. Create a new graph, Graph 4, where the things are organized into levels and only those couplings between things on different levels are shown.

The next two steps of the algorithm bring back into focus the intralevel and interlevel couplings between things that do not manifest composite-component couplings. These couplings and their associated things need to be considered when grouping things into systems and subsystems.

12. Take a level and reintroduce any things that appear to be at this level but were removed from Graph 2 as a result of Step 10 above (these are things that do not appear to be elements of a composite-component coupling). Then reintroduce the couplings between things within the level (also, removed from Graph 2 as a result of Step 10 above). Using Graph 4, create a new graph, Graph 5, to show all intralevel things and couplings.
13. Perform Step 12 until all levels have been covered and all things and all intralevel couplings from Graph 2 appear within the different levels shown in Graph 5. Then reintroduce any couplings between levels that do not appear to manifest a component-composite relationship (they simply manifest relationships between things that are on different levels).

After Step 13 is completed, note that Graph 5 shows the level structure of composites and components, as well as (1) couplings between things on the same level, and (2) couplings between things on different levels. All vertices (things) and edges (couplings) shown in Graph 2 should now be present in Graph 5. The edges should also show their coupling strengths (the integer n+1 that applies to the coupling). In the next two steps of the algorithm, Graph 5 provides the basis for identifying systems and subsystems in the level structure.

14. Review the intralevel couplings in Graph 5. Mindful of Condition 5 (maximum cohesion) and Condition 4 (minimum coupling) of the GDM, decide whether some things should be grouped and represented as a new subsystem and therefore whether additional levels are needed in the level structure. Coupling strengths should be used to inform decisions about how to group things into subsystems. Also, decide whether some existing couplings need to be redrawn in light of any new subsystems and levels that are created. After all intralevel couplings have been reviewed, create a new graph, Graph 6, that shows the outcome of this step.

---

[11] An emergent property is a property of the whole and not the parts, although it bears some relation (perhaps unknown) to properties of the parts.

[12] A hereditary property is a property of the parts that may or may not be also a property of the whole.

15. Review the interlevel couplings in Graph 6 that are not composite-component couplings. Mindful of Condition 5 (maximum cohesion) and Condition 4 (minimum coupling) of the GDM, decide whether some interlevel things should be grouped and represented as a new subsystem and therefore whether additional levels are needed in the level structure. Again, coupling strengths should be used to inform decisions about how to group things into subsystems. Also, decide whether some existing couplings need to be redrawn in light of any new subsystems and levels that are created. After all interlevel couplings have been reviewed, create a new graph, Graph 7, that shows the outcome of this step.

After Step 15 is complete, Graph 7 shows a level structure of components (basic things and subsystems) and composites (subsystems and systems) over the phenomena selected via the data-mining operations. One pass of the algorithm is complete. Further passes of the algorithm might be undertaken to try to identify a better level structure of systems and subsystems.

Appendix C3. Using the RM, STM, and GDM for Skills Training to Make More Effective Use of Information Systems

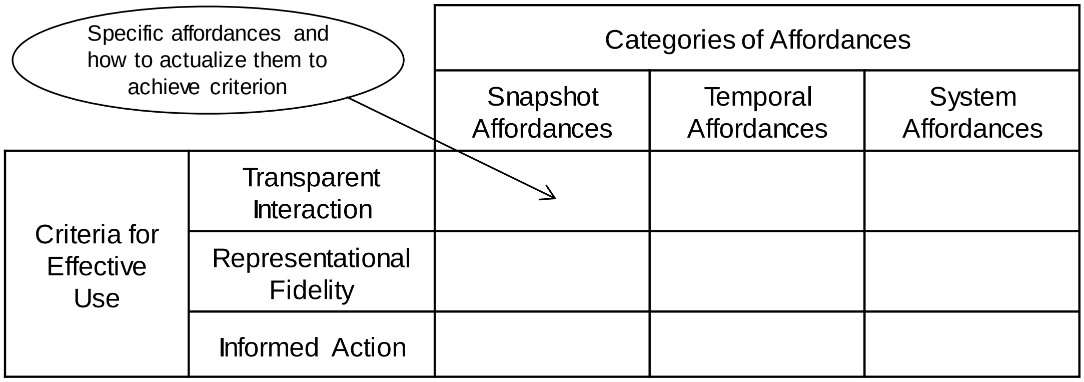
Two major approaches have been developed specifically to identify the skills that users need to leverage the capabilities provided by an IS. The first, proposed by Burton-Jones and Grange (2013), uses a top-down strategy to evaluate whether the users of an IS are able to (1) interact with the representations it offers, unimpeded by its surface and physical structure (*transparent interaction*), (2) obtain faithful data from the system (*representational fidelity*), and (3) take actions on the basis of good data (*informed action*). They indicate (p. 634) that the inspiration for using these three criteria is RT.

The second, proposed by Burton-Jones and Volkoff (2017), uses a bottom-up strategy to induce relevant skills from the *context* in which IS are deployed. When researchers apply this approach, they engage with the users of an IS who work in a particular context to (1) learn the specific affordances that users perceive the IS offers, and (2) determine how these specific affordances have been or can be actualized effectively. Volkoff and Strong (2018) provide guidelines to tease out and understand specific affordances, their actualizations, and their implications.

While RT has been used to inform the first approach, we propose it also can be used to inform the second approach. Specifically, we can glean three *general categories* of affordances from RT that researchers might consider as they work with users to learn the *specific* affordances an IS offers in a particular context and how users can actualize these affordances effectively:

- *Snapshot affordances*: Affordances related to learning relationships among phenomena manifested in the data at a given level of analysis *and* time. Researchers can use the RM for ideas about possible snapshot affordances.
- *Temporal affordances*: Affordances related to learning relationships among phenomena manifested in the data at a given level of analysis *over* time. Researchers can draw on the STM for ideas about possible temporal affordances.
- *System affordances*: Affordances related to learning relationships among phenomena manifested in the data *across* levels of analysis of a system and over time. Researchers can draw on the GDM for ideas about possible system affordances.

We propose, also, that the two approaches to studying effective system use of an IS are complementary and can benefit from being combined. Figure A7 illustrates the nature of the combined approach. The rows of the matrix are the three criteria for effective use of an IS identified by Burton-Jones and Grange (2013). The columns of the matrix are the three categories of affordances we have proposed above. Working bottom up with users, researchers could employ the categories of affordances (columns of the matrix) to facilitate elicitation of and understanding of specific affordances. Working top down with users, researchers could then employ the criteria for effective use (the rows of the matrix) to understand the actions needed to actualize the affordances effectively. For example, the top-left cell highlighted in Figure A7 (at the intersection of transparent interaction and snapshot affordances) would trigger researchers to ask users what must be done to interact transparently with static representations of a domain.



		Categories of Affordances		
		Snapshot Affordances	Temporal Affordances	System Affordances
Criteria for Effective Use	Transparent Interaction			
	Representational Fidelity			
	Informed Action			

Figure A7. A Combined Approach to Studying Effective IS Use

To illustrate how researchers could apply these ideas, consider a healthcare context where many stakeholders argue effective use of large data sets is key to a health system’s future (e.g., Stanford, 2017). We focus on the “row” for *representational fidelity* in the matrix above to show how this outcome might be achieved. We then consider an instance for each category of affordance and examine how it might be actualized to achieve high representational fidelity. To further scope the example, we also focus on the specific context of understanding how well hospitals address patients' "*problems*". We examine the notion of "problem" because a patient's "problem" is the core concept underpinning hospitals' electronic medical records (EMRs) (Weed, 1968). Nonetheless, clinicians often have difficulties using "problem-oriented" EMRs (Wright, Maloney, & Feblowitz, 2011). As a result, researchers have called for the efficacy of the "problem" concept to be studied (Chowdry, Mishuris, & Mann, 2017). Because of its focus on semantics, RT could prove informative.

Study 1: A relevant snapshot affordance offered by a hospital's EMR is understanding the *nature* of its patients' problems. How could hospitals actualize this affordance? Inspired by the RM, hospitals could allow patients to define their problems using a functional schema different from the EMR's functional schema (Smith & Koppel, 2014). Specifically, patients often define problems via their inability to meet goals (e.g., I can't walk a flight of steps), whereas EMRs define problems using formal diagnoses (Nagykaldi, Tange, & De Maeseneer, 2018; Nurcombe, 1989). We predict that more effective hospitals would circumvent this limitation to understand their patients' problems from patients' perspectives as well as diagnostic perspectives. For instance, such hospitals might ask clinicians to add data about patients' *goals* into the EMR's unstructured notes. Using natural language techniques, they could then mine the archives of unstructured notes to better understand patients' goals (and thus their patients' problems).

Study 2: A relevant temporal affordance offered by a hospital's EMR is understanding how well it *resolves* its patients' problems. To actualize this affordance, hospitals must be able to map patients' journeys through their facilities. They are often inhibited because EMRs are frequently designed for billing purposes and not clinical workflows (Cerrito, 2006; Claus, Carpenter, Chute, Mohr, & Gibbons, 1997; Gammon, Berntsen, Koricho, Sygna, & Ruland, 2015). The STM is useful in this context because it predicts hospitals will understand their performance better if they can account for outcomes due to *external* events (e.g., ambulance delays), specific *internal* events (e.g., types of clinical decisions), and specific *sequences* of internal events (e.g., deviations from care plans). We predict that hospitals that mine their EMR data from these three different perspectives will better understand how well they resolve their patients' problems.

Study 3: A relevant system affordance offered by a hospital's EMR is understanding *systemic* problems. While "problems" embedded in the semantics of EMRs are patient problems, some involve the hospital too. We predict that leading hospitals will analyze their large data sets not only in terms of patient problems but also systemic problems. For instance, consider the opioid crisis, which involves patient problems (e.g., chronic pain, addiction), clinician problems (e.g., opioid over-prescription), and hospital problems (e.g., rising demand for addiction services) (Kolodny et al., 2015; Volkow & McLellan, 2016). How could a hospital know how well it is resolving the "opioid problem"? To use the EMR to answer this question, stakeholders need appropriate system affordances. Having these affordances requires having the appropriate level structures. An EMR may have some level structures preconfigured (e.g., by having data on particular opioids, patients, clinicians, and hospital units). Most likely, however, other level structures have to be constructed in a bottom-up way to better understand the opioid problem (see Appendix C2). We predict that hospitals that engage in building and understanding these level structures will make better progress in understanding and resolving the opioid problem.

About the Authors

Jan Recker is a chaired professor of information systems and systems development at the University of Cologne and adjunct professor at Queensland University of Technology. He is presently editor in chief of *Communications of the Association for Information Systems*. His research focuses on systems analysis and design, digital innovation and entrepreneurship, and information systems for environmental sustainability.

Marta Indulska is a professor and leader of the Business Information Systems discipline at the UQ Business School, University of Queensland, Australia. She has a background in computer science, having obtained her computer science doctorate degree in 2004, and is a member of the Australian Computer Society. Marta's main research interests include conceptual modeling, business process management, and open innovation. Her research has appeared in *MIS Quarterly*, *Decision Support Systems*, *European Journal of Information Systems,* and *Journal of the Association of Information Systems*, among other journals.

Peter F. Green is a professor and head at the School of Accountancy, Queensland University of Technology. From 1999-2013, he was a professor of eCommerce at the University of Queensland. Peter has qualifications in accounting and computer science, and a PhD in commerce (information systems) from the University of Queensland. Peter's research interests focus on representation theory and its application to many different areas, including accounting information systems. His publications have appeared in such journals as *MIS Quarterly*, *European Journal of Information Systems*, *Information Systems*, and *Journal of the Association for Information Systems*, among other journals.

Andrew Burton-Jones is a professor of business information systems at the UQ Business School, University of Queensland. He has a bachelor of commerce (honours) and a master of information systems from the University of Queensland and a PhD from Georgia State University. He conducts research on systems analysis and design and the effective use of information systems. Prior to his academic career, he was a senior consultant in a big-4 accounting/consulting firm.

Ron Weber is an emeritus professor at the University of Queensland and Monash University. Previously, he was dean of the Faculty of Information Technology at Monash University and then pro-vice-chancellor of Monash South Africa. His research focuses primarily on developing and testing various aspects of representation theory.