A Cost-Benefit Method for Business Rules Normalization

Koen Smit
Digital Smart Services
HU University of Applied Sciences Utrecht
Utrecht, the Netherlands
koen.smit@hu.nl

Martijn Zoet
Optimizing Knowledge-Intensive Business Processes
Zuyd University of Applied Sciences
Sittard, the Netherlands
martijn.zoet@zuyd.nl

***Abstract*— This paper presents a cost/benefit analysis method for the normalization of business rules. To determine the economic benefit of business rules normalization three variables are addressed: 1) the number of anomalies a rule set endures, 2) the storage space a rule set requires and the 3) deterioration of rules in response time. The approach is evaluated by means of an experiment, based on mortgage data of an international bank. Results show that the method is useful for determining when to normalize business rule sets; the method enables business rules analysts to produce more cost-effective business rules architectures. In this paper, we re-address and - present our earlier work [1], yet we extended the previous research with more detailed descriptions of the related literature, findings, and results, which provides a grounded basis from which further research on business rules normalization can be conducted.**

Keywords-Business Rules; Decision Management; Normalization; Cost-Benefit Analysis

I. Introduction

Good decision making is a key denominator for a corporation's competitiveness [2]. Therefore, organizations are increasingly urged to make fast and accurate decisions. At the same time, decisions are becoming more and more complex, affecting maintainability and transparency. Decisions can be formulated by means of business rules [3][4]. A business rule is defined by Morgan [5] as: "*a statement that defines or constrains some aspects of the business intending to assert business structure or to control the behavior of the business.*" To realize changes within an organization's decision-making process, an organization should be able to maintain the aforementioned asserts and it should be able to adapt its business rules efficiently and effectively to realize changes within its decision-making process [6]. In order to realize this, information systems, such as expert systems, knowledge management systems, case based reasoning systems, fuzzy expert systems and business rules management systems have been built for and adopted by organizations [7].

Research on the management of business rules has been conducted since the mid-1960's [7][8]. Distinct research streams have emerged, focusing on the following three subjects: 1) subject transformation, 2) platform transformation, and 3) business rule model transformation [9]. Subject transformation research focuses on processes, methods and information systems used for mining and cleansing decision sources, such as regulations, organizational policies, laws, documents and databases. Platform transformation research focuses on the use of information technology for the deployment, execution and monitoring of business rules. Important research topics in this stream are: 1) algorithms for faster and easier execution, 2) business rules architectures, and 3) business rules engines [10][11][12]. Business rule model transformation research focuses on verification, validation and improvement of existing business rules. To verify business rules, a formal grammar notation and/or a set of constructs is applied. A grammar notation describes how a business rule should be constructed or formulated. An example of a standardized business rules grammar is the Semantics of Business Vocabulary and Business Rules [13].

Despite the accumulation of literature, there is a surprisingly scarce amount of research that examines methods and processes to factor business rules [3]. Factoring entails the process of dividing business rules, and therefore decisions, in more comprehensible structural elements to increase maintainability and transparency [14]. Research that has focused on this subject is "single language oriented" [9][3][15]. Since a relatively high number of business rules modelling languages exist within scientific and professional literature, a factoring procedure per language is not desired from the viewpoint of the authors. Furthermore, current research does not provide guidelines to financially quantify the value of factoring business rules [9]. As far as the authors are aware, no method exists that is business rules modelling language-independent in combination with quantifying the financial benefits of factoring business rules. An example is the work of [15], which solely focuses on achieving the third normal form while factoring business rules, without investigating whether this is financially optimal. Given the fact that organizations invest large amounts of money for implicitly managing business rules, a valid question is whether and when an explicit factoring procedure is economically beneficial. For example, a business rule set, which only changes or is executed twice a year might, from an economic perspective, be better off in an un-factored form. Taken previous statements into account, the following research question arose: "*How can business rules be factored such that economic beneficial manageability is realized?*" Following Van Thienen and Snoeck's [16] research on factoring decision tables and Zoet's [3] research on factoring business rules, relational theory is adopted to factor business rules.

The current study extends previous research by developing a factoring method that incorporates mainstream rule modeling languages and guidelines to determine the cost and revenue of (re-)factored business rules. A factoring method is developed and validated by means of an experiment based on case study data at a large international bank. The results showed that our method is effective in determining the economic costs and benefits.

In section two, a discussion is provided on the theoretical foundations of factoring business rules in terms of relational theory, normalization and economic factors. This is followed by the construction of the method in section three. In section four, a demonstration of the application of the method on mortgage decision making at a large international bank is provided. The paper is concluded, in section five, with the study's core findings, contributions as well as its limitations.

II. Background and Related Work

There are few methods available to (re-)factor business rules [3]. Currently, two different methods are described: one by Van Thienen and Snoeck [16] and one by Zoet et al. [9]. Van Thienen and Snoeck's [16] method has two underlying assumptions; 1) business rules are specified in decision tables and 2) relational theory is the basis for normalizing business rules. Guidelines are proposed to factor decision tables, thereby improving maintainability. However, instead of formulating one common procedure they proposed multiple exceptions to the normal form. These exceptions are an implicit result of the foundation of their research, namely the use of decision tables. The second method proposed by Zoet et al. [9] also takes relational theory into account. Moreover, this method distinguishes itself by applying one common procedure, which can be used for several languages. Similar to previous studies, this paper also applies relational theory as underlying foundation.

The definition of the term relational as used in this paper is adopted from the mathematical domain, more specifically from the relational algebra theory [17]. Relational algebra theory has received a lot of attention during the last four decades, since it is popularized by Codd [17] for database normalization. The basic idea of the relational algebra theory involves that a relationship (R) can exist of a given set of elements (Sn), visualized as follows: R = (S1, S2, ..., Sn) [17]. The elements (Sn) can be condition- or conclusion-facts. Most authors [17][18] represent element sets by applying two-dimensional arrays. In order to apply relational theory on business rules, one must be able to translate business rules to sets of relationships. Previous research has answered the question [9] whether current business rule modelling languages can be translated to unified views by applying relational algebra theory. Based on representational difference analysis, the authors show that the six most common business rules languages can be transformed to sets of relations [19][20]. The six languages, that were examined during this study are: If-Then business rules [21], Decision Tables [22][16], Decision Trees [23], Score Cards [24], Event, Condition & Action Business Rules [25], and Event Condition Action Alternative Business Rules [26]. By translating business rules to relations between specific sets of elements, normalization is made possible. Normalization is the process of removing partial dependencies and transitive dependencies [17][18].

III. Method Construction

A detailed explanation of the business rules normalization procedure can be found in [9]. However, to ground our research, a summary of the normalization procedure is provided in sub-section A. Subsequently, in sub-section B, the cost reduction analysis method for business rules normalization is described.

A. *Business Rules Normalization Procedure*

The process for business rules normalization consists of three activities. The results of these activities are 1) the transformation of business rules to the proper relational structure, and 2) the removal of partial and 3) the removal of transitive dependencies. The latter is realized by applying the third normal form, while the second normal form deals with partial dependencies and the 1st normal form deals with achieving the proper structure for business rules.

The first normal form is realized by duplicating the original business rules equally often as the amount of conclusion-facts that exist. In other words, all of the duplicated rules exist of all condition- and conclusion-fields. The difference between the original and new tables is that only one of the original conclusion-fields is now still a conclusion-field while the others are condition-fields. In order for a relation to be in the second normal form, all condition-facts must be functionally dependent on a conclusion-fact and adhere to the first normal form. Condition-facts, which are not fully dependent on the conclusion-fact must be deleted or added to another relationship. The second normal form reveals whether condition-facts are included that actually do not contribute to a conclusion. To realize the third normal form in business rule sets, condition-facts that are not fully dependent on the conclusion-fact (but on another condition fact) should be removed and added to a new relation. The new relation contains the removed condition-facts, as well as the conclusion-fact to which they are related. A relationship is established between two sets of relations by means of a secondary decision. After applying the third normal form, all specified relations do not contain any repeating groups, partial dependencies and transitive dependencies anymore.

To visualize the normalization procedure a decision tree can be used [27]. A decision tree consists of two types of nodes: 1) normalization decision nodes (squares) and 2) end nodes (circles), for example see Fig. 1. A normalization decision node represents the decision to further normalize the relationship. From a normalization decision node, two types of branches can emerge: 1) a stop branch, and 2) a normalization branch. A stop branch emerges when further normalization is not needed, consequently leading to an end node. When further normalization is needed, two or more normalization branches emerge from the decision node. These branches lead to other decision nodes representing the newly normalized relationships.

End nodes do not have further identification information, whereas normalization decision nodes do. Each node starts with the capital letter R, which is an abbreviation for relationship. The digit before the decimal point shows the number of the relationship. In case two digits are included before the comma, it designates a relationship resulting from another relationship. Furthermore, the digit after the decimal point indicates in what normalization form the relationship resides. In our example (see Fig. 1), the node R1,2 means that relationship 1 is in the second normal form. Moreover, the nodes R11,3 and R12,3 are both in the third normal form and are a relationship resulting from R1,2.

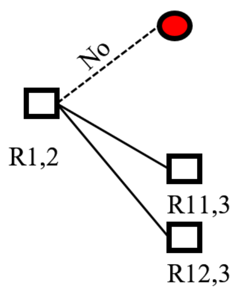

Figure 1. Decision Tree for Normalization

To demonstrate the business rules normalization procedure, a decision is normalized, which is based on the case study material that was collected, along the lines of the earlier work on the normalization of business rules by Zoet et al. [9]. In this example a process is considered in which the eligibility of a mortgage request is determined by the bank, see Fig. 2. The first step of the procedure is to determine the scope of the decision to normalize. During the process, visualized in Fig. 2, two analytical tasks are executed; 1) determine mortgage request eligibility and 2) discuss mortgage details with mortgage advisor. In this section, the focus is on the first analytical task. During this activity the eligibility of the mortgage request will be determined based on multiple criteria with regards to the personal situation of the applicant, the financial situation of the applicant, and the employment situation of the applicant. Based on the values of these criteria the mortgage request is either approved or rejected. When the mortgage request is approved, the applicant is invited to discuss mortgage details with a mortgage advisor from the bank.

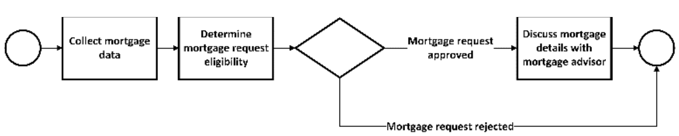

Figure 2. Determination process for the eligibility of mortgage requests

Now that the scope of the normalization procedure for this case is determined, the next step comprises the elicitation of the facts and their relationships used to determine the eligibility of the mortgage request.

To ground the elaboration of the normalization procedure, the end results of the normalization procedure are presented first along with the third normal form decision tables and their relationships (see Fig. 3), after which the procedure is explained step-by-step. In our examples, conditions and conclusions are shown without instantiated values. This is due to the fact that the case has to be reported in an anonymous way.

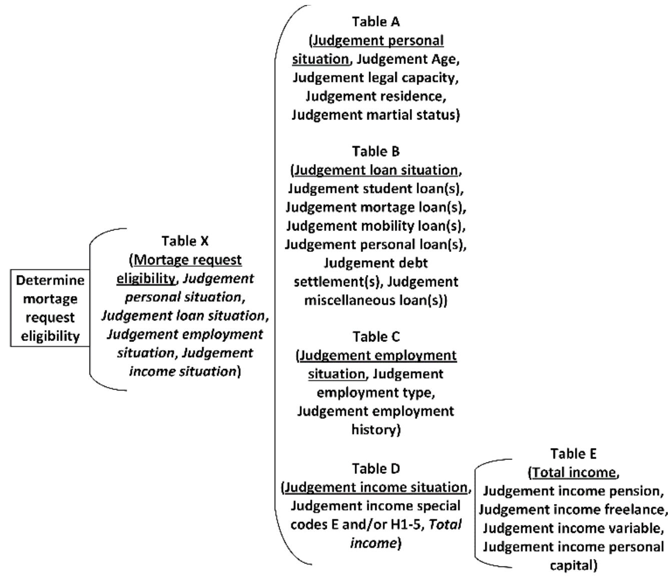

Figure 3. Overview of the normalized decision tables (derivation structure)

The elicitation of the facts and their relationships used to determine the eligibility of the mortgage request can be done in several ways. First, if the organization has already made the conditions (facts) explicit, written down in text or in a specific representation (i.e., decision trees, decision tables, domain models), this can serve as a starting point. When this is not the case, backward chaining can be applied to elicitate the facts and their relationships. With regards to our sample case, three decision tables were already present, see Fig. 4.

Table A: Judgement personal situation

Conditions				Conclusions
Judgement age	Judgement legal capacity	Judgement residence	Judgement marital status	Judgement personal situation
Value	Value	Value	Value	Value

Table B: Judgement loan situation

Conditions			Conclusions
Judgement student loan(s)	Judgement mortgage loan(s)	Judgement mobility loan(s)	Judgement loan situation
Value	Value	Value	Value
Judgement personal loan(s)	Judgement debt settlement(s)	Judgement miscellaneous loan(s)	
Value	Value	Value	

Table C: Judgement employment situation

Conditions				Conclusions	
Judgement income pension	Judgement income freelance	Judgement income variable	Judgement income Personal capital	Judgement income situation	Judgement employment situation
Value	Value	Value	Value	Value	Value
Judgement income special codes E and/or H1-5	Total income	Judgement employment type	Judgement employment history		
Value	Value	Value	Value		

Figure 4. Original and un-normalized decision tables

The next step of the procedure focuses on establishing the first normal form. The first normal form requires that every relation only contains one conclusion fact. In our case study this means that Table A and B comply with the requirement of the first normal form. Table C contains multiple conclusions and therefore needs to be transformed to comply with the first normal form. The transformation comprises the creation of two identical copies of Table C with the two different conclusion facts separated, one per table in Fig. 5.

Table C: Judgement employment situation

Condition				Conclusion
Judgement income pension	Judgement income freelance	Judgement income variable	Judgement income Personal capital	Judgement income situation
Value	Value	Value	Value	Value
Judgement employment history	Judgement employment type	Total income	Judgement income special codes E and/or H1-5	
Value	Value	Value	Value	

Condition				Conclusion
Judgement income pension	Judgement income freelance	Judgement income variable	Judgement income Personal capital	Judgement employment situation
Value	Value	Value	Value	Value
Judgement employment history	Judgement employment type	Total income	Judgement income special codes E and/or H1-5	
Value	Value	Value	Value	

Figure 5. First normal form decision tables

Table C: Judgement employment situation

Conditions		Conclusion
Judgement employment type	Judgement employment history	Judgement employment situation
Value	Value	Value

Table D: Judgement income situation

Conditions			Conclusion
Judgement income pension	Judgement income freelance	Judgement income variable	Judgement income situation
Value	Value	Value	Value
Judgement income personal capital	Judgement income special codes E and/or H1-5	Total income	
Value	Value	Value	

Figure 6. Second normal form decision tables

The second normal form is established when all relationships comply with the requirement for the first normal form, and additionally, when all conditions that are not fully dependent on the conclusion fact are removed. This is achieved by determining which of the conditions are irrelevant when formulating the conclusion and deleting these. In our case study, this affects all conditions labeled with the '*income*' statement that contribute to the '*Judgement income situation*' conclusion. The same holds for both the conditions '*Judgement employment type*' and '*Judgement employment history*', which contribute to the '*Judgement employment situation*' conclusion. This results into the following relationships, shown in Fig. 6.

The third normal form requires that all conditions only lead to the conclusion and do not derive another condition present in the decision. All conditions that are not fully dependent on the conclusion must be removed and added to a new decision. This is done by determining what conditions are not a determinant of the conclusion, but actually from another conclusion. In our case, the conditions '*Judgement income pension*', '*Judgement income freelance*', '*Judgement income variable*' and '*Judgement income personal capital*' from Table D are determining another conclusion, namely '*Total income*' and are therefore removed and defined as a separate decision (calculation), see Fig. 7, table E. All relationships are now specified in the third normal form, see Fig. 7.

Table C: Judgement employment situation

Conditions		Conclusion
Judgement employment type	Judgement employment history	Judgement employment situation
Value	Value	Value

Table D: Judgement income situation

Conditions		Conclusion
Judgement income special codes E and/or H1-5	Total income	Judgement income situation
Value	Value	Value

Table E: Calculate Total income

Conditions				Conclusion
Judgement income pension	Judgement income freelance	Judgement income variable	Judgement income personal capital	Total income
Value	Value	Value	Value	Value

Figure 7. Third normal form decision tables

B. *Cost Reduction Analysis Method for Business Rules Normalization*

Currently, in most normalization procedures the decision to normalize is generally based on intuitive flair. It remains uncertain whether the normalization effort is economically beneficial. For example, from an economic perspective, a business rule set, which only changes twice a year may not be beneficial to normalize.

Lee [28] and Westland [27] have conducted research towards the cost reduction of database normalization. Cost reductions realized by database normalization are 1) decreased machine time and 2) decreased data-inconsistencies (avoiding loss of business). The three main drivers of cost reduction are a) reduced anomalies, b) reduced storage requirements, and c) deteriorated response time. Anomalies that occur to data are: update-anomalies, insert-anomalies and deletion-anomalies

[17]. Previous research has shown that database normalization principles can be applied to business rule sets [9]. Taken previous statement into account, the following question arose: "*Can the cost reduction model for database normalization be adopted for business rules normalization?"*

Before adopting and adapting the model for business rules normalization, first the fit between the database determinants and business rules determinants has to be investigated. First, both business rules and data are updated, deleted, and inserted. Second, previous research [28] has shown that business rules normalization can also lead to fewer storage requirements, such as the case is with database normalization. Thirdly, deteriorated response time is an important issue since decision making in organizations is becoming increasingly complex with, for example, predictive analytics. As such, the formulas proposed by Lee [28] are adopted. However, before the formulas can be applied, the variables need to be adapted towards business rules. The remainder of this section will discuss the formulas provided by Lee altered towards business rules.

The cost reduction realized by normalization is calculated in four phases; 1) cost reduction due to reduced anomalies, 2) cost reduction due to reduced storage space, 3) cost increase due to increased join processing, and 4) comparing cost reduction due to reduced anomalies and cost reduction due to reduced storage space with the cost increase due to increased join processing.

Let ϕ be the cost reduction due to reduced anomalies, see also (1). ϕ is defined as:

$$\phi = \sum_{M=1}^{Nu} \alpha_M^U \, \lambda_M^U \, \acute{\omega}_M^U + \sum_{M=1}^{Ni} \alpha_M^I \, \lambda_M^I \, \acute{\omega}_M^I + \sum_{M=1}^{Nd} \alpha_M^D \, \lambda_M^D \, \acute{\omega}_M^D \quad (1)$$

Where Nu, Ni, and Nd are the number of updates, number of inserts and number of deletions, respectively, λ_M^U, λ_M^I and λ_M^D denote the frequency of the m'th update, the m'th insertion and the m'th deletion. The average number of business rules affected by the update, insertion and deletion are denoted by $\acute{\omega}_M^U$, $\acute{\omega}_M^I$ and $\acute{\omega}_M^D$. Furthermore, α_M^U, α_M^I and α_M^D denote the cost for each insert, update and deletion.

Let ψ be the cost reduction due to reduced storage space, see also (2). ψ is defined as:

$$\psi = B\acute{\omega} - B_x \, \acute{\omega}_x - B_y \acute{\omega}_y \quad (2)$$

Where B represents the storage cost per business rule in the current normalized situation. B_x and B_y denote the storage cost per business rule in the normalized situation + 1. The number of business rules stored in the current normalization situation is depicted by $\acute{\omega}$, while the normalized situation + 1 is depicted by $\acute{\omega}_x$ and $\acute{\omega}_y$.

Let Ω be the cost increase due to increased join processing, see also (3). Ω is defined as:

$$\Omega = \sum_{\substack{M=1 \\ x,y \in o^m}}^{\emptyset} \ddot{\Upsilon}_m \, \mu_m \, \acute{\omega}_x \, \acute{\omega}_y \quad (3)$$

Where Ø is the number of joins required to determine the conclusion of a specific decision. $\ddot{\Upsilon}_m$ denotes the cost per execution per business rule for join M. Moreover, μ_m represents the frequency of join M. The number of business rules in the business rule sets that are joined are expressed by $\acute{\omega}_x$ and $\acute{\omega}_y$. The business rule sets (x and y) between which the join M is realized, is denoted by x, y, ϵ o^m. Let O be the cost reduction from normalization form R (R1,2) to normalization form R+1 (R11,3). Summarizing, O = ϕ + $\psi \geq \Omega$. O can be either positive or negative [3], [16]. If O is positive, then normalization should be applied.

IV. Experiment setup

In our validation, an experiment on case study data is applied. This allows us to use data from an actual case while fully controlling the execution of the method and input variables. The method is applied to a mortgage decision of an Anonymous International Bank (AIB). Our choice to select this case study setting was based on two theoretical criteria. Firstly, the case had to provide a proper amount of business rules used to take a decision. The mortgage decision at AIB consisted of 1479 facts (conditions and conclusions), and 665 individual business rules. Secondly, the organization had to be willing to provide the financial details needed to perform the calculations. AIB agreed to this, however, with two conditions. The first condition implied that their name and financial data were altered when it would be published. The second condition entailed that the applied business rule sets were not completely published.

Figure 8. Photo impression 1 of normalized business rules

The evaluation, by means of conducting an experiment, was divided into three phases. Phase one was used to make the researchers familiar with the case parameters, by analyzing 133-pages with descriptions of decisions for completeness and accuracy. This phase resulted in the identification of multiple gaps. With the help of additional documentation and experts these gaps have been fixed. During the second phase, the business rules have been normalized according to our method. This normalization was done on paper, after which the results were presented on a big wall (see Fig. 8 and Fig. 9). During the normalization, additional gaps were identified. These gaps have been marked with "post-its", see Fig. 8 and Fig. 9. Again, with the help of additional documentation and experts, these gaps have been resolved.

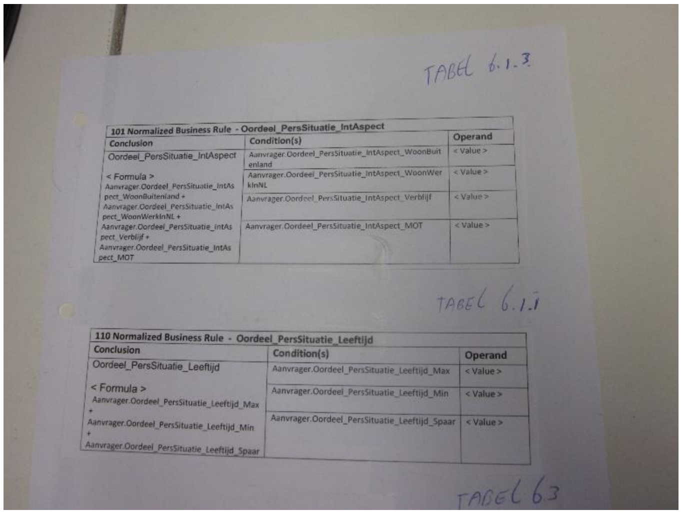

Figure 9. Photo impression 2 of normalized business rules

V. Application of the Method

To ground our method, three example scenarios are presented and elaborated. First, the determination of the cost reduction from normalization form R to normalization form R+1 for the business rule set "*personal situation of applicant*" is explained. It is part of the case described in the previous section. The data used in these examples are derived from the collected case data.

A. *Example one – small part of the mortgage case (positive benefits)*

The business rule set exists of 10 facts, 1 conclusion fact and 9 condition facts; see the left side of Fig. 10. The question that needs to be answered before normalizing this business rule set is: "*Does normalizing the business rule set from R to R+1 realize a cost reduction?*"

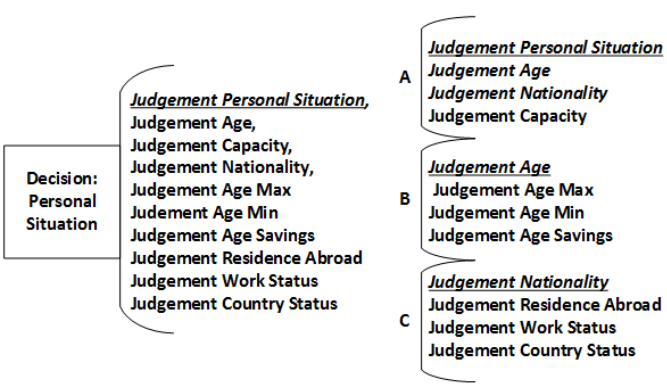

Figure 10. Normalization from the second to the third normal form

The decision personal situation is only affected by update and insert anomalies. For example, the facts "*judgment age*" and "*judgment age savings*" are updated regularly. Insert anomalies occur when new type of rules for age determination are inserted. The application of the method exist out of four phases 1) determine benefits in terms of reduced anomalies, 2) determine savings of storage requirements and 3) determine effect on response time, and 4) comparing cost reduction due to reduced anomalies and cost reduction due to reduced storage space with the cost increase due to increased join processing.

During phase one, three steps can be distinguished. *Step one:* determine the amount of update, insert and deletion operations on a specific business rule set. In our case, "update judgment age rules" and "insert age determination rules". For each identified operation type, it should be determined if the operation is affected by anomalies. If anomalies do not occur, normalization is not needed at all. If anomalies do occur, the frequency of each operation type and the number of business rules that are affected should be determined, this corresponds to step *two*. In this specific case λ_1^{U} = 7 (/per 2 weeks), and λ_2^{U} = 6 (/ per 2 weeks). Additionally, the number of business rules affected by each update needs to be determined. In this specific case $\acute{\omega}_1^{U}$ = 2 and $\acute{\omega}_2^{U}$ = 1.5. During step *three*, the cost of an anomaly should be determined. In this case, the cost of a person that adjusts the specific business rules α_1^{U} = €35.00 per instance and α_2^{U} = €52.50 per instance, see also (4). So, the total benefit due to reduced number of anomalies is:

$$\phi = (35 * 7 * 2) + (52.5 * 6 * 1.5) = €962.50 \qquad (4)$$

The first step of phase two is to determine the results of the transformation in terms of business rule sets. In this case, one business rule set (personal situation) is divided into three business rule sets, namely: 1) judgment personal situation, 2) judgment age, and 3) judgment nationality. The results of the normalization are shown in Fig. 10. For each business rule set, the number of business rules must also be determined, in this case, respectively, $\acute{\omega}$ = 20, $\acute{\omega}_x$= 2, $\acute{\omega}_y$= 3, $\acute{\omega}_z$= 6. During the second step, the cost per stored business rule must be determined. This needs to be determined for the current situation as well as for the post normalization situation. This information was retrieved from the information technology department, in this case, respectively, B =€4, $B_X = €0.5$, $B_y = €0.5$ and B_z = €0.5. Duplications are removed, thereby decreasing the number of individual business rules, see also (5). The total benefit due to reduced storage space is:

$$\psi = 20 * 4 - 2 * 0.5 - 3 * 0.5 - 6 * 0.5 = €74.50 \qquad (5)$$

To form a decision, two joins are required in the new situation, so $\emptyset$ =2. The cost for each join $\ddot{Y}_m$= 0.015. The execution frequency of the join is 4000 per two weeks (μ_m), see also (6). The additional cost due to additional join operations (Ω) is therefore:

$$\Omega = 0.015 * 4000 * (2 + 3 + 6) = €660.00 \qquad (6)$$

In conclusion, further normalization for the decision personal situation is recommended since:

$$(\phi = €962.50 + \psi = €74.50) > \Omega = €660.00$$

Assume a situation where λ_1^U = 7(/per 2 weeks), λ_2^U= 6 (/ per 2 weeks) are decreased to λ_1^U = 2(/per 2 weeks), λ_2^U = 2 (/ per 2 weeks). Applying these changes reduces ϕ from €962.50 to €446.25, which changes O from (ϕ = €962.50 + ψ = €73.00) > Ω = €660.00 to (ϕ = €446.25 + ψ = €73.00) < Ω = €660.00, in which case further normalization would not realize a cost reduction.

The above example has shown a situation in which normalization leads to cost reduction and therefore the normalization should occur. By changing two parameters, it is demonstrated that normalization would lead to a negative cost reduction, and therefore an increase in cost. Based on this, normalization should not be performed. To demonstrate such a situation another example is discussed in the next paragraph.

B. Example two – (negative benefits)

In this example, slightly altered case characteristics are defined as input for the formulas used to determine economic fit of further normalization of the decision "*Determine nationality code*". This decision is utilized to determine the code of the nationality of the applicant, for all products and services that the AIB Bank offer. Again, applying the method starts with determining the amount of update, insert and deletion operations on a specific business rule set, which is λ_1^U = 1 (/per 6 months), and λ_2^U = 2 (/ per 6 months), and λ_3^U = 1 (/per 6 months). These low amount of update, insertion, and deletion modifications are explained by the fact that the list of recognized countries by the UN, which the AIB bank adheres to, is rarely altered. Furthermore, the number of business rules affected by each update need to be determined, which is $\acute{\omega}_1^U$ = 2, $\acute{\omega}_2^U$ = 2, and $\acute{\omega}_3^U$ = 2.

Lastly, the cost of an anomaly, which represents the total cost to repair the business rule set to a state without the anomaly, should be determined. In this case, the cost of a person that adjusts the specific business rules α_1^U = €1.00 per instance. α_2^U = €2.00 per instance and α_3^U = €1.00 per instance. The costs are relatively low as modifying the list of country codes is a simple task for the AIB bank information specialist. The costs for insertion type modifications are doubled as a new country needs to be registered with several meta-data fields, which is not the case with regards to update and deletion type modifications.

With regards to storage space, the number of business rules must be determined, which is in this case $\acute{\omega}$ = 25, $\acute{\omega}_x$= 10, and $\acute{\omega}_y$= 10. Normalization of the decision to the first normal form leads to the removal of five redundant business rules. Additionally, the cost per stored business rule must be determined, in this case B = €0.75, B_X = €0.5, B_y = €0.5 and B_z = €0.5. In this particular case, the storage costs are €0.75 in the un-normalized form.

With regards to additional costs due an increase of join operations, the amount of joins to form a decision need to be determined, which is $\emptyset$ = 1. The cost for each join $\ddot{\Upsilon}_m$ = 0.00015. The execution frequency of the join is 500.000 per month (μ_m). The execution frequency is relatively high, which is caused by the generality of the decision and underlying business rules. Therefore, this particular decision is utilized in many other services by the AIB (i.e., different financial products like insurances and investment options) to determine the country of origin of an applicant.

To normalize from the un-normalized form to the first normal form, again the 1) total benefit due to reduced anomalies is calculated, see (7):

$$\phi = (1 * 1 * 2) + (2 * 2 * 2) + (1 * 1 * 2) = €12.00 \quad (7)$$

2) the total benefit due to reduced storage space, see (8):

$$\psi = 250 * 0,75 - 10 * 0.5 - 10 * 0.5 = €87.50 \quad (8)$$

3) the additional cost due to additional join operations, see (9):

$$\Omega = 0.00015 * 500.000 * (10 + 10) = €1.500,00 \quad (9)$$

Further normalization for the decision 'Determine nationality code' is not recommended since:

$$(\phi = €12 + \psi = €87.50) < \Omega = €1.500,00$$

C. Example three – normalization of the decision "Judgement Age"

Lastly, the method is demonstrated by evaluating the normalization steps with regards to the decision "*Judgement Age*". To ground the normalization, first the transformations from the 0th to 1st normal form, from the 1st to the 2nd normal form, and from the 2nd to the 3rd normal form are shown in Fig. 11.

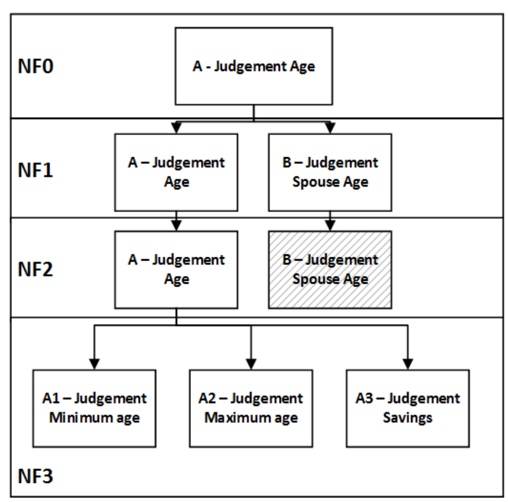

Figure 11. Normalization mapping of the decision "*Judgement Age*" of the mortgage case

First, the variables with regards to anomalies, storage space, and join operations need to be determined. In this specific case λ_1^U = 70 (/per 2 weeks), λ_2^U = 70 (/ per 2 weeks), and λ_3^U = 70 (/ per 2 weeks). Additionally, the number of business rules affected by each operation are $\acute{\omega}_1^U$ = 6, $\acute{\omega}_2^U$ = 6, and $\acute{\omega}_3^U$ = 6. In this case, the cost of a person that adjusts the specific business rules α_1^U = €2.50 per instance, α_2^U = €2.50 per instance, and α_3^U = €2.50 per instance.

With regards to storage space, the number of business rules must be determined, which in this case is $\acute{\omega}$ = 54, $\acute{\omega}_x$= 54, and $\acute{\omega}_y$= 54, where the latter two represent the tables that are copied as part of the normalization into the first normal form. Additionally, the cost per stored business rule must be determined, in this case B = €0.75, B_X = €0.5, B_y = €0.5 and B_z = €0.5. In this particular case, the storage costs are €0.75 in the un-normalized form, and €0.5 in the first, second and third normal form.

With regards to additional costs due an increase of join operations the amount of joins to form a decision need to be determined, which is $\emptyset$ = 1 in case of the first normal form. The cost for each join $\ddot{\Upsilon}_m$= 0.00015. The execution frequency of the join is 5.000 per month (μ_m). Applying the method starts from the un-normalized decision table "*Judgement Age*" that includes all fact types in one single table towards the first normal form, see (10), (11) and (12):

$$\phi = (2.5 * 70 * 6) + (2.5 * 70 * 6) + (2.5 * 70 * 6) = \text{€ } 3.150{,}00 \quad (10)$$

$$\psi = 54 * 0.75 - 54 * 0.5 - 54 * 0.5 = \text{€ - } 13.50 \quad (11)$$

$$\Omega = 0.00015 * 10000 * (54 + 54) = \text{€ } 162.00 \quad (12)$$

Further normalization for the decision "*Judgement Age*" is recommended since:

$$(\phi = \text{€}3.150.00 + \psi = \text{€-}13.50) > \Omega = \text{€}162.00$$

Normalization from the un-normalized decision table to the first normal form results in the creation of two separate decision tables, A "*Judgement Age*" and B "*Judgement Spouse*". The normalization of the case to the second normal form is continued. In this situation, only one join ($\emptyset$ = 1) is required to derive a conclusion for the decision "*Judgement Age*". Furthermore, normalizing table A results in the removal of 18 business rules. Normalizing table B results in the removal of 52 business rules. The fact that 18 and 52 business rules were removed had the following reason. Analysis of the case showed that only six condition facts and one conclusion fact are related to table A. Furthermore, only two conditions and one conclusion fact are related to table B. Again, the same formulas are applied, see (13), (14) and (15):

$$\phi A = (2.5 * 70 * 6) + (2.5 * 70 * 6) + (2.5 * 70 * 6) = \text{€}3.150{,}00$$
$$\phi B = (2.5 * 70 * 6) + (2.5 * 70 * 6) + (2.5 * 70 * 6) = \text{€}3.150{,}00 \quad (13)$$

$$\psi = (54 + 54) * 0.5 - 36 * 0.5 = \text{€}36.00 \quad (14)$$

$$\Omega = 0.00015 * 10000 * (36 + 2) = \text{€}57.00 \quad (15)$$

Further normalization of both decisions are recommended since:

$$(\phi = \text{€}6.300.00 + \psi = \text{€}37{,}00) > \Omega = \text{€}56.00$$

The case is normalized to the second normal form since the benefits of the normalization are greater than the costs (($\phi + \psi$) > Ω). Normalizing tables A and B, which are in the second normal form, towards the third normal form results in the creation of an additional three tables. Table A will split into table A1 "*Judgement Minimum Age*", A2 "*Judgement Maximum Age*", and A3 "*Judgement Savings*". In the third normal form table A1 will contain 2 business rules, table A2 will contain 3 business rules and table A3 will contain 3 business rules as well. Table B will remain as is. To derive a conclusion for the decision "*Judgement Age*", two joins ($\emptyset$ = 2) are required in the third normal form. Furthermore, normalization towards the third normal form results in a decrease of possible anomalies per type of operation. Also, the average number of affected business rules are reduced due to the reduced number of business rules in the tables due to normalization. Normalization from the second to the third normal form results in the following situation, see (16), (17) and (18):

$$\phi = (2.5 * 30 * 4) + (2.5 * 30 * 4) + (2.5 * 30 * 4) = \text{€}900.00 \quad (16)$$

$$\psi = (36 + 2) * 0.5 - (2 * 0.5 - 3 * 0.5 - 3 * 0.5) = \text{€}15.00 \quad (17)$$

$$\Omega = 0.00015 * 15000 * (2 + 3 + 3) = \text{€}18.00 \quad (18)$$

When transforming table B from the first to the second normal form it automatically adheres to the third normal form. Therefore, no further normalization can be applied. Further normalization for table A is recommended since:

$$(\phi = \text{€}900.00 + \psi = \text{€}15.00) > \Omega = \text{€}18.00$$

The resulting derivation structure for the decision "*Judgement Age*" in the context of its parent decision "*Judgement Personal Situation*" is depicted in Fig. 12.

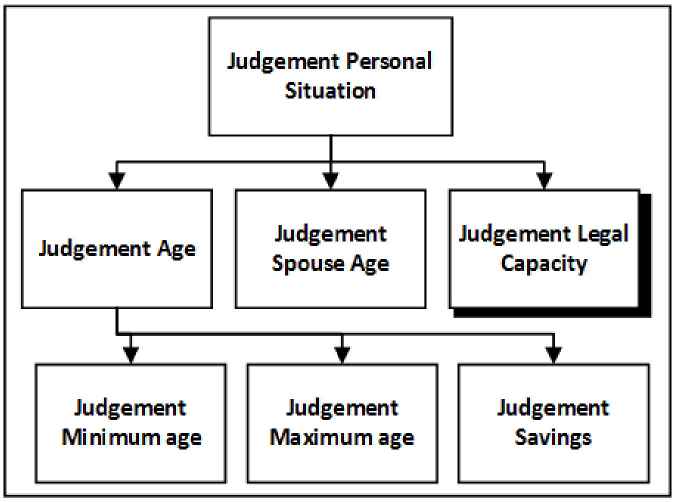

Figure 12. Normalization mapping of the decision "*Judgement Age*" of the mortgage case

VI. EXPERIMENT VALIDITY

Internal validity threats, when conducting controlled experiments, can be classified into eight categories: 1) selection, 2) history, 3) maturation, 4) regression, 5) attrition, 6) testing, 7) instrumentation, and 8) additive and interactive effect of threats to internal validity [29]. We can ensure that the learning effect was not present during our case. Given the fact that all four subjects who have participated in the experiment already had executed the business normalization procedure before. Furthermore, the economical beneficially calculation itself was made explicit in Excel and required the respondents only to enter the variables. We cannot exclude learning during the transformation of the case information to the relational representation. Selection, history, maturation, attrition, instrumentation and additive and interactive effects of threats to internal validity are mitigated due to the application of a controlled experiment.

Outcomes of an experiment can vary when subjects, tasks or the environment changes. External validity is concerned with the extension of variations on such changes [29]. Our results were obtained from one decision: a mortgage decision. Therefore, we cannot claim that our conclusions are generally applicable. However, the answer to the research question itself is not influenced by the fact that only one case has been analyzed. Our experiment has been applied outside the project life cycle of AIB. We do not consider this as a threat to environmental validity since the entire procedure can be repeated during normal project life cycles.

VII. CONCLUSION, DISCUSSION AND FUTURE RESEARCH

Business rules are a key denominator for a corporation's competitiveness. Thereby, the management of such business rules is increasingly becoming more important. However, business rules are becoming more and more complex affecting maintainability and transparency. In order to properly structure business rules, normalization is applied. Normalization increases control over insertion, update and deletion anomalies affecting storage requirements and response time. Currently, the normalization procedure does not take the costs and benefits of normalization into account but is based on intuitive flair. Therefore, we defined the research question: *How can business rules guiding decisions be factored such that economic beneficial manageability is realized?*

We presented a cost/benefit formula, which output provides guidelines for normalizing business rules. To determine the normalization business case, three variables were addressed 1) the number of anomalies a business rule set endures, 2) the storage space a business rule set requires, and the 3) deterioration in response time due to an increased amount of joins. By means of an experiment based on case study data from an international bank, we have shown the applicability of the model. Results show the importance of properly normalized decisions and what role the cost and benefit analysis plays in this. On the one hand, modelers should attempt to properly factor business rules. To achieve this factoring, the three normalization forms can be applied. On the other hand, practitioners should take cost and benefits of the organization into account when applying such normalizations forms. Currently, the transformation of the business rules is performed manually. However, in future research we aim to develop an approach that applies an algorithm to re-write (transform) business rules for applying the method presented in this paper. Furthermore, future research should also focus on further validating the method presented in this paper using more cases, and ideally, cases from different industries in various sizes to improve its generalizability.

From a practical perspective, our study provides product engineers, business rules modelers and (business) decision modelers with a method that can be used to normalize business rules based on an economic rationale. This rationale comprises the fit between storage space utilization, anomaly management and execution costs. The method will enable organizations to guard, on the one hand, execution costs and, on the other hand, performance of business rules.

The factor of speed to decide to normalize business rules is becoming increasingly important. Un-normalized business rules have a higher execution speed then normalized business rules, due to the additional join that have to be created. For a mortgage decision the reduction in speed is not that important since a few seconds extra doesn't affect the outcome for clients. For different decisions in banking, these few extra seconds (or mili- or microseconds) are a challenge. An example of an area where such decisions can be found is High Frequency Trading (HFT). Therefore, business rules in this area are usually not normalized. An interesting direction for future research would be to identify types or patterns of decisions that have an economic benefit for normalization and those that do not.

In this paper we focus on the normalization of decisions based on economic incentive. However, economic incentive is not the only factor that can affect the choice whether to normalize or not to normalize. For example, with the arrival of the GDPR regulation in May 2018 [30], transparency becomes an important factor for banks and other organizations. Normalization has an effect on the understandability (and therefore transparency) of the decision [31]. Therefore, the formula to normalize a decision and underlying business rules should be extended by a factor that takes into account transparency.

The current normalization procedure is based on relational algebra theory. However it focusses on highly explicit business rules. This means that the business rules are known upfront and can be fully written down. Other forms of decision logic that are based on relational algebra are Machine Learning algorithms and Neural networks. The method might apply machine learning and neural networks. However, the formulae needs to be properly adapted for such changes. This will be part of future research as well as taking into account the privacy factor for neural networks.

References

[1] M. Zoet and K. Smit, "An Economic Approach to Business Rules Normalization," in *Proceedings of the Ninth International Conference on Information, Process, and Knowledge Management*, 2017, pp. 1–6.

[2] M. W. Blenko, M. C. Mankins, and P. Rogers, "The Decision-Driven Organization," *Harv. Bus. Rev.*, vol. 88, no. 6, pp. 54–62, Jun. 2010.

[3] M. Zoet, *Methods and Concepts for Business Rules Management*, 1st ed. Utrecht: Hogeschool Utrecht, 2014.

[4] R. G. Ross, "The business rule approach," *Computer (Long. Beach. Calif).*, vol. 36, no. 5, pp. 85–87, 2003.

[5] T. Morgan, *Business rules and information systems: aligning IT with business goals*. Addison-Wesley Professional, 2002.

[6] M. L. Nelson, R. L. Rariden, and R. Sen, "A lifecycle approach towards business rules management," in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, 2008, pp. 113–123.

[7] S. Liao, "Expert system methodologies and applications—a decade review from 1995 to 2004," *Expert Syst. Appl.*, vol. 28, no. 1, pp. 93–103, Jan. 2004.

[8] A. Kovacic, "Business renovation: business rules (still) the missing link," *Bus. Process Manag. J.*, vol. 10, no. 2, pp. 158–170, 2004.

[9] M. Zoet, P. Ravesteyn, and J. Versendaal, "A structured analysis of business rules representation languages : Defining a normalisation form," *Proc. 22nd Aust. Conf. Inf. Syst.*, p. Paper 20, 2011.

[10] D. Arnott and G. Pervan, "A critical analysis of decision support systems research," *J. Inf. Technol.*, vol. 20, no. 2, pp. 67–87, 2005.

[11] I. Graham, *Business rules management and service oriented architecture: a pattern language*. John wiley & sons, 2007.

[12] M. L. Nelson, J. Peterson, R. L. Rariden, and R. Sen, "Transitioning to a business rule management service model: Case studies from the property and casualty insurance industry," *Inf. Manag.*, vol. 47, no. 1, pp. 30–41, Jan. 2010.

[13] Object Management Group, "Semantics of Business Vocabulary and Business Rules (SBVR) - V1.0," 2008.

[14] E. Eessaar, "The database normalization theory and the theory of normalized systems: finding a common ground," *Balt. J. Mod. Comput.*, vol. 4, no. 1, pp. 5–33, 2016.

[15] B. Von Halle and L. Goldberg, *The Decision Model: A Business Logic Framework Linking Business and Technology*. CRC Press, 2009.

[16] J. Vanthienen and M. Snoeck, "Knowledge factoring using normalisation theory," *Int. Symp. Manag. Ind. Corp. Knowl.*, pp. 27–28, 1993.

[17] E. F. Codd, "A relational model of data for large shared data banks," *Commun. ACM*, vol. 13, no. 6, pp. 377–387, 1970.

[18] W. Kent, "A Simple Guide to Five Normal Forms in Relational Database Theory," *Commun. ACM*, vol. 6, no. 2, pp. 120–125, 1983.

[19] M. Hubank and D. Schatz, "Identifying Differences in Mrna Expression by Representational Difference Analysis of Cdna," *Nucleic Acids Res.*, vol. 22, no. 5, pp. 5640–5648, 1994.

[20] M. zur Muehlen and M. Indulska, "Modeling languages for business processes and business rules: A representational analysis," *Inf. Syst.*, vol. 35, no. 4, pp. 379–390, Jun. 2010.

[21] R. Rivest, "Learning Decision Lists," *Mach. Learn.*, vol. 2, no. 3, pp. 229–246, 1987.

[22] R. Kohavi, "The Power of Decision Tables," in *Proceedings of the 8th European Conference on Machine Learning*, 1995, pp. 174–189.

[23] J. Boyer and H. Mili, *Agile business rule development: Process, Architecture and JRules Examples.* Springer Berlin Heidelberg, 2011.

[24] D. A. Morrow, E. M. Antman, A. Charlesworth, R. Cairns, S. A. Murphy, and E. de Lemos, J. A. Braunwald, "Timi Risk Score for St-Elevation Myocardial Infarction: A Convenient, Bedside, Clinical Score for Risk Assessment at Presentation," *Circulation*, vol. 10, no. 2, pp. 2031–2037, 2000.

[25] U. Dayal, A. P. Buchmann, and D. R. McCarthy, "Rules are objects too: a knowledge model for an active, object-oriented database system," in *Proceedings of the International Workshop on Object-Oriented Database Systems*, 1988, pp. 129–143.

[26] T. Heimrich and S. Günther, "Enhancing Eca Rules for Distributed Active Database Systems," in *NODe 2002 Web- and Database-Related Workshops*, 2003, pp. 199–205.

[27] C. Westland, "Economic Incentives for Database Normalization," *Inf. Process. Manag.*, vol. 28, no. 5, pp. 647–662, 1992.

[28] H. Lee, "Justifying Database Normalization: A Cost/Benefit Model," *Inf. Process. Manag.*, vol. 31, no. 1, pp. 59–67, 1995.

[29] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning, 2002.

[30] European Commission, "Protection of personal data - GDPR," 2017. [Online]. Available: http://ec.europa.eu/justice/data-protection/ [Accessed: 14-Aug-2017].

[31] B. Goodman and S. Flaxman, "EU regulations on algorithmic decision-making and a 'right to explanation,'" in *ICML workshop on human interpretability in machine learning (WHI)*, 2016.