Perceived Consistency between Process Models

Matthias Weidlich[a,*], Jan Mendling[b]

[a] *Hasso-Plattner-Institute, University of Potsdam, Prof.-Dr.-Helmert-Str. 2-3, D-14482 Potsdam, Germany*
[b] *Humboldt University, Unter den Linden 6, D-10099 Berlin, Germany*

Abstract

Process-aware information systems typically involve various kinds of process stakeholders. That, in turn, leads to multiple process models that capture a common process from different perspectives and at different levels of abstraction. In order to guarantee a certain degree of uniformity, the consistency of such related process models is evaluated using formal criteria. However, it is unclear how modelling experts assess the consistency between process models, and which kind of notion they perceive to be appropriate.

In this paper, we focus on control flow aspects and investigate the adequacy of consistency notions. In particular, we report findings from an online experiment, which allows us to compare in how far trace equivalence and two notions based on behavioural profiles approximate expert perceptions on consistency. Analysing 69 expert statements from process analysts, we conclude that trace equivalence is not suited to be applied as a consistency notion, whereas the notions based on behavioural profiles approximate the perceived consistency of our subjects significantly. Therefore, our contribution is an empirically founded answer to the correlation of behaviour consistency notions and the consistency perception by experts in the field of business process modelling.

Keywords: Process Model Consistency, Consistency Perception, Behavior Equivalence, Model Refinement

---

*Corresponding author
Email addresses: matthias.weidlich@hpi.uni-potsdam.de (Matthias Weidlich), jan.mendling@wiwi.hu-berlin.de (Jan Mendling)

1. Introduction

In the last decade, the increasing awareness of the benefits of process-based management led to a broad field of application for process-aware information systems (PAIS). Such a system manages and executes operational business processes on the basis of process models [1]. On the one hand, PAIS aim at improving process-oriented organizations and managing organisational change. On the other hand, PAIS are applied for workflow automation and enterprise application integration. The diversity of application scenarios for PAIS is manifested in the multitude of dimensions along which business processes can be classified. Following on [1] and [2], business processes and, therefore also PAIS, differ with respect to the following aspects: a focus on organisational or operational aspects, an intra- or interorganisational scope, the involvement of process stakeholders and the degree of automation, and structural characteristics that exhibit a certain degree of predictability or repetition.

Naturally, the application scenarios of PAIS imply the involvement of various groups of process stakeholders. That, in turn, impacts on the way the underlying process models are designed. Depending on the concrete purpose of the process model and the concrete stakeholders, a single process of an enterprise is often modelled from different perspectives and at different levels of abstraction. This results in different related models of the same process being used to serve the information needs of different stakeholders. For instance, the processing of job applications might be captured from the angle of the Human Resources (HR) unit, such that the process model guides the work of HR employees and clarifies hand-overs. Here, the process model would not be executed directly, but specifies organisational requirements for the supporting IT infrastructure. A process model obtained for the very same process, in turn, is likely to look different from the angle of the IT department. In this case, focus lies on the interactions of HR-employees with an IT-system, whereas organisational details are largely neglected. In general, the reason for such differences between process models stems, among others, from the constraints an IT infrastructure imposes on *how* the goal of a process is achieved, whereas organisation centred modelling typically focusses on *what* needs to be done. Consequently, there will be a variety of differences between related models. Such differences might be observed with respect to the described functionality, the imposed control flow dependencies, the specification of data handling, or resource allocations [3, 4, 5].

Differences between models representing a common process, each *appropriate* in its specific context, are in the nature of process models that serve different purposes. Nevertheless, a certain consistency between these models has to be guaranteed as these related process models are used as means for communication by different stakeholders that work on the design of the same real-world process.

According to Zelewski, consistency of process models refers to a *freedom of contradictions* [6]. Still, a concrete operationalisation of this definition remains challenging. Evidently, there is a trade-off between *strictness* of a consistency notion and *appropriateness of process models* serving different purposes. A strict notion in the sense of an equivalence that requires all information of one model to be present in another model as well, will result in models that are inappropriately tailored for the purpose of a different stakeholder. On the other hand, if the act of constructing process models towards a dedicated purpose leads to models that specify highly contradicting information, effective coordination between organisational units of an enterprise is bound to fail. That leads to the question of what kind of consistency should be guaranteed between such process models. Checking such related process models for consistency is a non-trivial task that should be supported by appropriate concepts and tools. Up until now, the notion of consistency has only been discussed from a conceptual and formal point of view in process model research. It is still unclear how modellers assess the consistency between process models, and which kind of notion can best aid them in decision making.

In this paper, we focus on the control flow of process models and relate *consistency* to formal notions of behaviour consistency. We investigate the research question of *which formal notion of behavioural consistency can best approximate perceived consistency of modelling experts*. Our research design can be classified as a correlational study similar to [7, 8, 9]. In relation to formal notions, we compare the adequacy of a behavioural equivalence notion of the linear time – branching time spectrum [10] and consistency notions based on behavioural profiles [11]. The latter concept was defined to accommodate for the specifics of the consistency question. It relaxes certain properties of the aforementioned equivalence notions. So far there has been only anecdotal evidence for the proximity and remoteness of these various notions to the consistency perception. Eventually, the question of proximity can only be validated from an empirical perspective involving experts in process modelling. This paper presents the findings from an online experiment that we conducted on the perception of behaviour consistency between pairs of process models. We identified 69 expert statements from process analysts from all over the world, and we analysed how the aforementioned notions for behaviour consistency match the perceived consistency of our subjects. Therefore, our contribution is an empirically founded answer to the correlation of behaviour consistency notions and the consistency perception by experts in the field of business process modelling.

Against this background, the remainder of this paper is structured as follows. First, we discuss the issue of behaviour consistency by means of an example in Section 2. Using the example, we also discuss related work on process model relations and behaviour consistency. Based thereon, we derive a set of hypotheses on the proximity between existing consistency notions and the consistency perception of modelling experts. Section 3 elaborates on our research design and introduces the results in detail. We discuss these results in Section 4 and explicate their implications for research as well as for practice. Finally, Section 5 concludes the paper.

2. Background

We illustrate the need to consider consistency aspects between related process models by introducing two major use cases and by means of an example scenario in Section 2.1. Subsequently, we review related work. First, Section 2.2 focusses on the notion of a correspondence between process models. Based on structural properties of these correspondences, we elaborate on differences between correspondence and refinement relations. In addition, we review research on the identification of correspondences. Section 2.3 discusses related work on notions of behaviour consistency. Those can be based on a behaviour equivalence or a behavioural abstraction such as behavioural profiles. Based thereon, we derive a set of hypotheses on the proximity between these consistency notions and the consistency perception of modelling experts in Section 2.4.

2.1. Use Cases & Example Scenario

Consistency is at the very core of applying process models in order to bridge the gap between organisational management and development of the supporting technical infrastructure. In particular, two use cases are primary drivers for consistency analysis, that is, *validation* and *change management*. Business process models are often utilized as a specification of organisational requirements against which the technical workflows of an information system

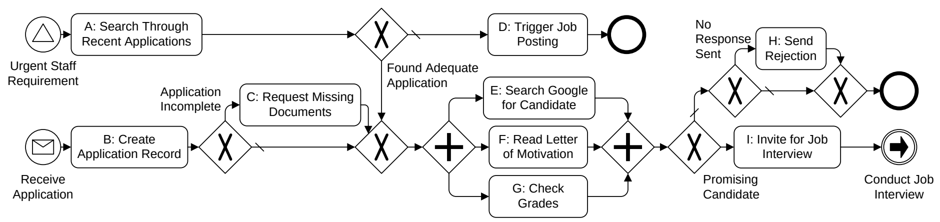


Figure 1: Application processing from the organisational perspective

are *validated*. This raises the question of consistency between the organisational and the technical process models. Even if both models are consistent at a certain point of time, there is the need to reassess it in case of changes on either level. According to Gartner, the key elements of PAIS are *'keeping the business process model in sync with process execution [and] enabling rapid iteration of processes and underlying systems for continuous process improvement and optimisation'* [12]. As changes occur frequently [13], an appropriate notion of consistency can help to identify ways of propagating changes from one model to the other for making them consistent again, cf., [14].

In order to exemplify the gap between process models focussing on different perspectives, we use a process that describes the handling of job applications as an example. For illustration purposes, we use the Business Process Modeling Notation (BPMN) throughout this paper. The usage of different modelling languages with potentially varying expressiveness might complicate any consistency assessment. Still, we explicitly exclude mismatches between different modelling languages (such as the conceptual mismatches between BPMN and BPEL [15, 16]) from our discussion.

Figure 1 shows the process from an organisational perspective, which might be used to organise the work of an HR unit. The process is triggered either by the occurrence of an urgent staff request from another department or the reception of an application in reference to a job description. In the former case, recent applications are searched for potential candidates, while a job posting might be created subsequently. In case of the reception of an application, completeness of the application is ensured first. Afterwards, the application is assessed, which leads either to an invitation for a job interview or (optional) sending of a rejection.

A different perspective of the application processing is depicted in Figure 2. Instead of the work to be done by HR employees, the process of the supporting IT is modelled. The process starts with the selection of a job profile. Subse-

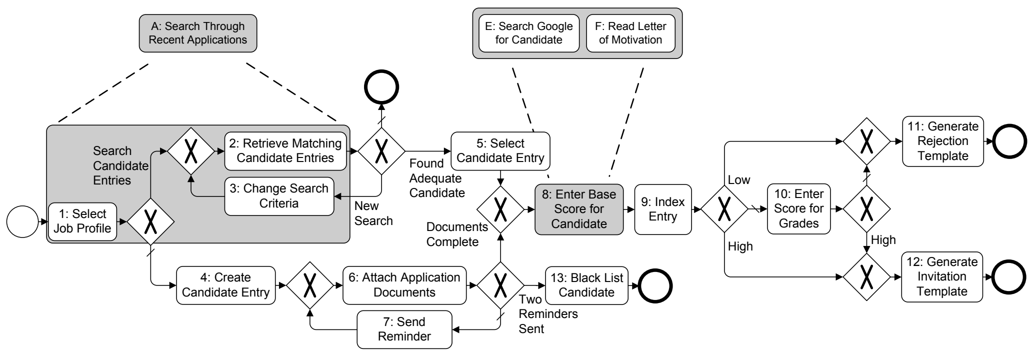


Figure 2: Application processing from the perspective of the supporting IT along with example correspondences to activities of the model in Figure 1

quently, either a search through the existing candidate entries or the creation of a new candidate entry are performed. In the former case, the search query might be refined multiple times. If a new candidate entry is created, the application documents are attached to the entry, while reminders are sent in case of missing documents. If the application is complete, the results of an application assessment are entered and depending on the outcome a template is generated, either for a rejection or an invitation for a job interview. Note that the results of an application assessment and the template generation are also executed in case an existing candidate entry is selected after a search. Thus, such an assessment is assumed to be specific to a job description.

The process models in Figure 1 and Figure 2 are highly related as both describe different perspectives on a common real-world process. Therefore, there are a couple of correspondences between activities or pairs thereof. In Figure 2, we visualised two corresponding pairs of activities. Note that these activities are often not semantically equivalent. For instance, 'Search Google for Candidate' and 'Read Letter of Motivation' obviously refer to the respective real-world activities performed by HR employees, whereas they are reflected solely by activity 'Enter Base Score for Candidate' in Figure 2. Nevertheless, these activities *correspond* to each other against the background of aligning the organisational and technical perspectives.

Both process models differ with respect to coverage of the application processing and the modelling granularity. For instance, the potential creation of a job posting modelled in Figure 1 is not part of the process model in Figure 2, i.e., the activity is not supported by the respective IT system. Moreover, the two exemplary correspondences between sets of activities of both models as depicted in Figure 2 hint at a different level of detail in both models. Another observation relates to the potential entry and exit points specified in both models. While the process in Figure 1 can be entered at two different entry points (i.e., start events), there is only a single entry for the process in Figure 2. This process, in turn, specifies other exit points (i.e., end events) as the first process. For instance, in Figure 2, the process might end after two reminders for missing documents have been sent.

2.2. Correspondences between Process Models

Correspondences between process models, as exemplified in the previous section, are needed for any assessment of behaviour consistency between process models. Therefore, this section elaborates on the notion of correspondences in detail. First, structural properties of correspondences are introduced. Second, we use these properties to discuss the relation of correspondences and common refinement operations. Third, we summarize how correspondences can be identified.

Structural Properties of Correspondences. Correspondences can be classified structurally according to their *cardinality*. An elementary 1:1 correspondence between two process models relates one activity of one model to one activity of the other model. A correspondence that is defined between sets of activities is called complex. Here, 1:n correspondences are a specific type of complex correspondences. Our scenario in Figure 2 illustrates that such 1:n correspondences are likely to be observed between models intended for managing organisational units and models that are the basis of information systems engineering.

Further on, a pair of correspondences between two process models can be assessed regarding the *direction* and *overlap*. In contrast to 1:1 or n:m correspondences, a 1:n correspondence is said to be *directed* from one activity to a set of activities. However, different 1:n correspondences do not have to coincide with respect to their direction. For instance, the two 1:n correspondences illustrated explicitly in Figure 2 have different directions. Activity 'A' corresponds to activities '1', '2', and '3', whereas activity '8' corresponds to activities 'E' and 'F'. Two correspondences might be *overlapping* in the sense that a single activity is part of multiple correspondences. For instance, the activity 'Select Job Profile' in Figure 2 is part of two correspondences, one representing the searching of recent applications (activity 'A') and one representing the creation of an application record (activity 'B') in Figure 1 (note that the second one is not visualised).

As illustrated by our scenario, correspondences between organisational models and models for information systems engineering do not necessarily show a single direction between the models. In addition, they may be overlapping. Hence, we refer to these correspondences as being *non-hierarchical*.

Refinements of Process Models. A correspondence relation between process models is close to the relation induced by model refinement operations. In general, specialisation of behavioural models might be traced back to two elementary notions, *refinement* and *extension* [17, 18, 19]. While refinement refers to the definition of an activity (or a set thereof) in more detail, extension refers to the act of adding new activities. Relating these notions to the concept of a correspondence between process models, activities that are not part of any correspondence could be seen as an extension. On the other hand, a complex correspondence could be interpreted as a refinement, if it would imply semantic equivalence of the related sets of activities. Above, we already discussed that we do not assume such a semantic equivalence to hold due to the different modelling perspectives (e.g., organisational vs. information systems). A review of literature on model refinements reveals further conceptual differences between correspondence and refinement relations. That is, refinement relations are typically equally directed, non-overlapping, and hierarchical.

Process model refinements have mainly been tackled in the context of *process views*. In order to accommodate for the multitude of potential perspectives on a real-world business process, various approaches assume the existence of a holistic *core model*. Such a model might be adapted towards a dedicated purpose and group of stakeholders by deriving customised views [20, 21, 22, 23]. In the same vein, methodologies for *integrated system design* propose to derive technical realisations from business models directly via refinements [3, 24, 25]. Although these approaches do not start with a core model, the final technical model can be seen as a core model as well. However, these approaches require the existence of hierarchical refinements between the core process and the process view. That is, the relation is assumed to be non-overlapping and equally directed. As discussed for our example and detailed in [3, 26, 27], this assumption does not hold for models that are used for managing organisational units and models used for information systems engineering, in the general case. Techniques for the integration of views of behavioural models into a core model, e.g., [28, 29, 30], aim at creating such a hierarchical relation.

Similar observations can be made for Software Engineering (SE) techniques that target the development of complex systems using multiple *viewpoints*, e.g, [31, 32, 33, 34, 35, 36]. Viewpoints realise a separation of concerns by focussing on different aspects of the system to be built. Due to this focus on a dedicated system, relations between viewpoints are hierarchical either.

Identification of Correspondences. Correspondences between parts of two process models can be identified manually by the modeller or automatically. Recently, techniques based on structural analysis and natural language processing have been introduced for identifying correspondences between single activities of business process models [37, 38]. Moreover, techniques known from the area of schema and ontology matching [39, 40] can be exploited. Here, activities might be regarded as elements of a process model schema. However, despite some notable exceptions (cf., [41, 42]), there has been a predominant focus on 1:1 correspondences. Complex 1:n or even n:m correspondences have largely been neglected. Recently, the ICoP framework [43] has been introduced, which aims at the detection of 1:1 and complex 1:n correspondences automatically. Still, identification of such complex correspondences imposes challenges to automatic matching.

2.3. Notions of Behaviour Consistency

In the previous section, we clarified the notion of a correspondence between activities of two process models. This section summarises existing work on how to judge behaviour consistency of two process models and correspondences between them.

Consistency between process models has been discussed from various angles. Above, we already mentioned a definition of consistency as the *absence of contradictions* [6]. In the same vein, consistency between reference process models and their implementation has been discussed in [44]. Based on the distinction between variant and invariant parts of a reference process model, consistency may require the absence of side effects (i.e., contradictions) once variant parts are subject to change. According to [6], the question of how to assess the *absence of contradictions* can be seen as a verification problem. In the remainder of this section, we review two elementary properties that may be verified in order to conclude on behaviour consistency. First, Section 2.3.1 focusses on behaviour equivalences and their application for consistency evaluation. Second, Section 2.3.2 discusses the concept of a behavioural profile and its usage for consistency analysis.

2.3.1. Behaviour Consistency based on Behaviour Equivalence

In order to decide whether two process models show the same behaviour, different notions of behaviour equivalence can be applied. These notions are well-investigated and classified in the linear time – branching time spectrum [10, 45, 46]. Behaviour equivalence might be based solely on the observable execution sequences of activities (alias traces). That is, *trace equivalence* between two process models means that both models coincide with respect to all sequences of activity executions. Trace equivalence is commonly seen as the lower bound of the aforementioned spectrum. There are stricter notions of behaviour equivalence that consider not only the observable traces but take the branching structure of the processes into account. Even if two models show the same observable behaviour in terms of traces, the point in time branching decisions are taken can be different. *Branching bisimulation* can be seen as the upper bound of the spectrum of process equivalences. It requires two process models to simulate each other in terms of state transitions.

Process equivalences have been used for consistency evaluations in the context of *behaviour inheritance*, *view consistency*, and *interaction consistency*.

Behaviour Inheritance. There has been a lot of work aiming at transferring the concept of inheritance known for static structures (e.g., class diagrams in UML) to the level of behavioural models. In particular, object life cycles that describe the behaviour of software or system artefacts have been investigated. In this context, behaviour consistency answers the question whether an extension or a refinement of such an object life cycle is behaviour preserving. Ebert and Engels [47] first proposed to assess the behaviour consistency object life cycles either on the *observed* behaviour or the *invocable* behaviour. Schrefl and Stumptner generalised these ideas and proposed the notions of *observation consistency* and *invocation consistency* [17, 18, 19]. Informally, two models are observation consistent, if they share all sequences of states and activities once those that are not part of any correspondence are neglected. In other words, states and activities in one model that are without counterpart in the other model are *hidden*. Two models are invocation consistent, if all sequences of activity executions valid in one model are also valid in the other model. In contrast to observation consistency, invocation consistency requires all activities that are not part of any correspondence to be *blocked* (the difference between weak and strong invocation consistency as discussed in [19] is not important in our context). Similar ideas have been presented in [48, 49]. Basten and van der Aalst distinguish two basic notions of behaviour inheritance, namely *protocol inheritance*, *projection inheritance*, and combinations thereof [49]. Protocol inheritance corresponds to the notion of invocation consistency, whereas projection inheritance resembles observation consistency, if model states are neglected.

The aforementioned work on behaviour inheritance assumes 1:1 correspondences. However, these ideas have recently been lifted to the level of complex correspondences [50]. This approach does not impose any restrictions on the cardinality or direction of correspondences. Still, it assumes non-overlapping correspondences and considers complete traces, i.e., traces from the initial to the final state of the process. Further on, there is also a large body of work on equivalence-preserving refinements for Petri nets, refer to [51] for a thorough survey. In this context, refinement is always assumed to be hierarchical, i.e., a place or transition of a Petri net is replaced by a subnet. Thus, the subnet is *embedded* into the original net [52, 53]. As a result, such a refinement leads to equally directed non-overlapping 1:n correspondences.

View Consistency. Behaviour equivalences are also related to view-based approaches. Most techniques ensure consistency between a base model and the process views derived from it by means of structural rules. Sill, the structural consistency is motivated by guaranteeing a certain degree of behaviour consistency in terms of 'activity orderings' [20, 21, 22, 23]. Albeit often not mentioned explicitly, the latter can be traced back to the notions of behaviour equivalence. This also holds for techniques that integrate multiple views of a behavioural model, such as [28, 29, 30].

Regarding multi-viewpoint engineering approaches, the authors of [32, 33] advocate the application of behaviour equivalences and partial preservation of traces as notions of behaviour consistency. Behaviour equivalences are also used to assess consistency in [36]. Other approaches favour the description of behavioural dependencies imposed by each viewpoint in terms of logic statements. Then, the conjunction of these statements is checked for satisfiability [34, 35]. Despite the rather strict consistency notions, various multi-viewpoint approaches acknowledge the inevitability of inconsistencies, cf., [54].

Interaction Consistency. Behaviour equivalences have also been applied as consistency notions in the domain of business-to-business (B2B) integration. Even though there are various different consistency requirements (see [55]), the use case of B2B integration imposes rather strict consistency requirements, in general. That is due to the goal of realising an integration following on a dedicated interaction protocol. Thereby, consistency has to be decided at different stages. For instance, public processes of interacting partners have to be consistent. In addition, consistency must be ensured between these public processes and their private implementations. Consistency of interactions has been studied extensively (e.g., [56, 57, 58, 59, 60, 61, 62, 63]) and is typically based on a behaviour equivalence.

2.3.2. Behaviour Consistency based on Behavioural Profiles

Notions of behaviour consistency can be based on a behavioural abstraction instead of a behaviour equivalence. Above, we discussed that even the weakest behaviour equivalence, trace equivalence, considers the whole observable behaviour of the processes to compare. In contrast, a behavioural abstraction defines a certain loss of information on the process behaviour. Consequently, it allows for assessing behaviour consistency of processes when neglecting dedicated differences in their behaviour.

The concept of a *behavioural profile* is such a behavioural abstraction [11] and approximates the behaviour of a process. In essence, each pair of activities of a process model is considered to be in a dedicated behavioural relation. That is, any two activities are either exclusive, in strict order, or in interleaving order to each other. Both activities never occur together in a single trace (exclusiveness), or if they occur together all occurrence of one activity will be before all occurrence of the other activity (strict order), or occurrences of two activities can be observed in any order (interleaving order). According to [11], these relations partition the Cartesian product of activities. Further on, a single activity is said to be either exclusive to itself (it can occur at most once) or in interleaving order with itself (it may occur multiple times).

The definition of behavioural profiles has been motivated by the fact that interleaving equivalences such as trace equivalence are not invariant under extensions [64]. That is, removing an activity from a process model may break behaviour equivalence. This operation is also referred to as a forgetful refinement—one activity is refined with the empty activity. Equivalence of the behavioural profiles for all pairs of corresponding activities of two process models can be used as a behaviour consistency criterion. This criterion, in turn, is weaker than criteria based on a behaviour equivalence. The profiles abstract from certain behavioural details such as causal dependencies. On the other hand, even the weakest behaviour equivalence, i.e., trace equivalence, of two process models implies the equivalence of the behavioural profiles. As illustrated in [11], consistency based on behavioural profiles can be decided even in the presence of non-hierarchical complex correspondences.

2.4. Hypotheses on Consistency Notions

Above, we outlined that notions of behaviour consistency may be either based on behaviour equivalence or behavioural abstraction such as behavioural profiles. It has been shown in prior work that trace equivalent process models also share the same behavioural profile [11], whereas the inverse implication does not hold true. Therefore, there is a spectrum of consistency notions. Regarding this spectrum, we use the following arguments and preliminary results for building our hypotheses.

- Our initial example showed that activities that are not part of any correspondence are likely to be observed when assessing the consistency of a business process model and a workflow model. It is well-known that this phenomenon (alias forgetful refinement) does not preserve any behaviour equivalence [64]. Hence, it suffices to consider the weakest behaviour equivalence, trace equivalence, as the strictest criterion of our consistency spectrum.
- When applying trace equivalence to assess behaviour consistency, activities that are not part of any correspondence might be hidden or blocked. A recent case study revealed that hiding these activities is more appropriate when assessing a reference model and its implementation [50]. That is due to the fact that activities are often inserted as part of the standard processing, such that blocking these activities would stall the process. Therefore, we consider solely activity hiding in our consistency spectrum.
- In [11], the consistency between process models of an industrial reference model has been assessed based on both, trace equivalence and behavioural profiles. Under the assumption that the models of this reference model are rather consistent, the results suggest that behaviour consistency is better approximated with the criterion based on behavioural profiles. Therefore, this criterion is also considered in our consistency spectrum.
- There is some evidence that repetitions are only modelled on a certain level of detail and ignored in high-level business process models. So-called happy path process models only capture the most frequent sequence of a process [65, 66, 67]. Furthermore, *coupled repetitive activities* are a structural pattern occurring frequently in the refinement of process models [68]. Accordingly, information on potential repetition may be neglected in a consistency assessment. Therefore, we also consider equivalence of behavioural profiles without repetition in our consistency spectrum. Technically, that means that the self-productions of two activities are skipped from the behavioural profile.

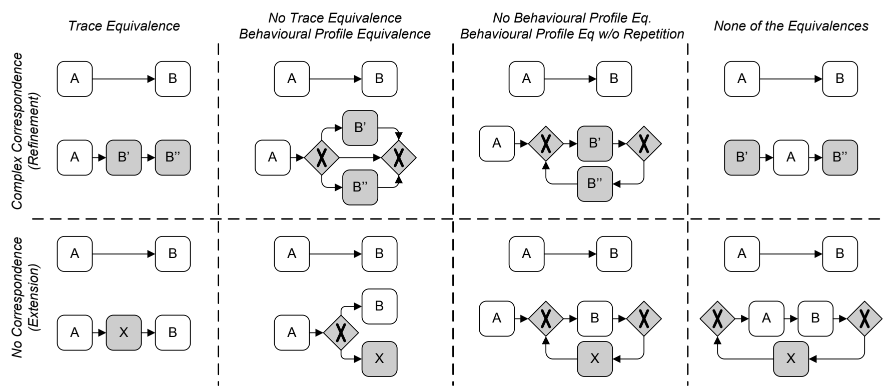


Figure 3: Overview of correspondences with respect to their behaviour preserving properties

Based on these arguments, we selected three definitions of behaviour consistency for further investigation, *trace equivalence*, *behavioural profile equivalence*, and *behavioural profile equivalence without repetition*. We illustrate the consistency notions analysed in this paper in Figure 3. The first two pairs of models on the left side of Figure 3 illustrate the presence of a complex correspondence (between B and $\{B', B''\}$ in the upper left model) and the presence of an activity that is not part of any correspondence (X in the lower left model). In the former case, the correspondence relation contains the pairs (A, A), (B, B'), and (B, B''). In the latter case, activity X has no counterpart in the upper model comprising only activities A and B in a sequence, i.e., activity X extends this model. As a consequence, the correspondence relation contains the pairs (A, A) and (B, B), while activity X is not part of any correspondence. Still, in both cases, the pairs of models are trace equivalent. For the upper pair of models, the trace (A, B) has a corresponding trace (A, B', B'') induced by the correspondence (see [50] for details on how to assess trace equivalence in the presence of complex correspondences). For the lower pair of models, traces are equal as the activity X is hidden in any trace. For both pairs of models, trace equivalence implies behavioural profile equivalence. There is a strict order relation between activities A and B (or A and B', and A and B'', respectively).

Focussing on the next column of Figure 3, we see two pairs of models that are not trace equivalent, but behavioural profile equivalent. The possibility to skip activities B' and B'' in the upper model, or B in the lower model, impacts on the equivalence of the complete traces. In the upper pair of models, the top model shows only one complete trace, i.e., (A, B). The model below it allows for the trace (A) as neither B' nor B'' is required to be executed. Hence, there are different complete traces in both models. The same holds true for the lower model pair, in which execution of activity X implies skipping of activity B. Still, all four models in the column show equal behavioural profiles. The behavioural profile of the upper most model defines a strict order relation between activities A and B – in any trace that contains both activities, A occurs before B. The same holds true for the refined model, so that there is a strict order relation between activity A and B', and A and B'', respectively. The behavioural profile abstracts from the fact that neither B' nor B'' need to be executed in this model. Instead, the profile focusses on the order dependency that is observed if these activities are executed, i.e., it refers to the order of potential occurrence. For the same reason, the strict order relation between A and B is also observed in the lower most model. The possibility to execute activity X instead of activity B does not impact on the order of potential occurrence. In any trace that contains both activities, A and B, A occurs before B.

The third column of Figure 3 illustrates the case that there is an equivalence of the behavioural relations solely if potential repetitions are neglected. Finally, the last column of Figure 3 gives examples for pairs that are inconsistent according to all presented notions.

The consideration of these three criteria can also be related to different theories of meaning that humans might implicitly or explicitly apply when judging consistency (see [8]). Trace equivalence was originally defined to check whether two computer programs have the same effect in the real world. Such a position can be traced back to early pragmatist theory, which emphasizes observable effects as the essence of meaning [69]. Behavioural profiles are rather based on the constraints that are implied by process models. This position can be related to feature theory of meaning [70]. In an experiment on schema matching, Evermann shows that people tend to give slightly more emphasis to similar features than to similar effects, yet both positions appear to be relevant [8].

We formalise our arguments in three hypotheses on the proximity between the criteria and the perception of experts on the consistency of a pair of process models.

H1: Pairs of process models that are *trace equivalent* will be perceived to be more consistent than pairs that are not.

H2: Pairs of process models that are *behavioural profile equivalent* will be perceived to be more consistent than pairs that are not.

H3: Pairs of process models that are *behavioural profile equivalent without repetition* will be perceived to be more consistent than pairs that are not.

In order to assess these hypotheses in an experimental setup with modelling experts, additional factors have to be taken into account. It is well known that personal differences have an impact on model comprehension [71]. In particular, modelling experience [72, 73] and study background [74] have been found to be relevant comprehension factors. It is important to check how these additional factors affect the results. Therefore, we aim to investigate personal characteristics of the experts including work status in academia or industry, work focus on business or IT, or time of modelling experience. Considering these factors allows us to investigate to which extent our findings are independent of these factors. We formulate the following questions to address the potential influence of personal differences.

- Does perceived consistency of academic and industry experts differ?
- Does perceived consistency of business and IT experts differ?
- Does perceived consistency of experts with longer modelling experience differ from those with shorter modelling experience?

In the following, we describe an online experiment that we conducted with modelling experts in order to assess the hypotheses.

3. Research Method

This section introduces our research approach. In the previous section, we closed with a set of hypotheses on the perception of consistency notions. Our experimental research design based on an online questionnaire to test these hypotheses is introduced in Section 3.1. Subsequently, Section 3.2 gives an overview of the demographic details of the subjects and presents the results on the consistency perception.

3.1. Research Setup

In order to introduce our research setup, we elaborate on the target audience, the selection and creation of pairs of process models used in the questionnaire along with their characteristics concerning behaviour consistency, and the experiment instrumentation.

Subjects / Target Audience. We aimed at testing the hypotheses in the most general setting. Thus, our target audience were practitioners involved in the creation, analysis, or implementation of business process models that

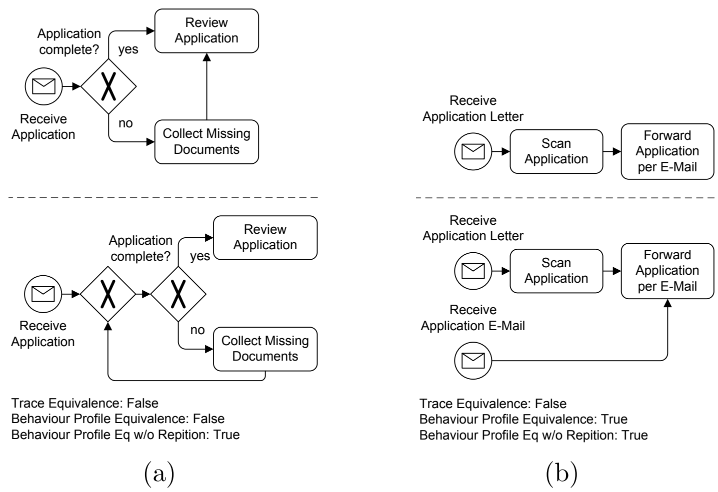


Figure 4: Two objects of our experiment illustrating a difference related to an activity (event, respectively) that is (a) part of a correspondence, (b) not part of a correspondence

have significant real-world experience in process modelling. Against this background, we decided to gather data on their consistency perception using an online questionnaire as timely access to a wide range of participants with different backgrounds was crucial for the external validity of our investigations. As it is not feasible to approximate the entire audience for the questionnaire, we could not select a representative sample and had to rely on convenient sampling and a self-recruiting questionnaire design, following on the recommendations given in [75].

Objects / Pairs of Process Models. In order to assess the perception of the subjects on process model consistency, the questionnaire comprised 11 pairs of process models, the objects of our experiment. Each pair consisted of process models in BPMN notation, which contained at most four activities and used only a small subset of BPMN modelling constructs, such as plain and message start events, activities, and XOR and AND gateways. These constructs are widely used and can be seen as the common core of process modelling languages [76]. For two models including a boundary error event and a multiple-instance activity, we included annotations explaining the respective semantics. The objects have been taken from two generally known domains to not have any domain bias, namely processing of a purchase quote and processing of a job application. However, activity labels have been chosen rather abstract like 'Create Quote', 'Submit Quote', 'Receive Application' and 'Review Application', respectively. That, in turn, should ensure that participants are not influenced by any process context.

In all 11 objects, the respective process models showed a slight difference as illustrated in Figure 3.1. The selection of differences has been guided by classifications of process model differences available in related work, e.g., [3, 4, 5]. In an earlier work, we reviewed these classifications and identified nine common differences between process models that relate to the activity or control-flow perspective, such as *activity fragmentation* or *decision distribution* [5]. As these differences might impact on the behaviour consistency, each of them was covered by at least one object in our experiment. In addition, two differences were included twice as discussions with process modelling experts in a pre-test suggested to consider these differences in two variants. Note that for six out of the 11 objects, the difference was related to activities that are part of a correspondence (similar to a refinement). In contrast, the remaining five pairs showed a difference related to activities that are not part of any correspondence (i.e., the model was extended). Thus, we ensured both types of differences are considered equally. All objects used in the experiment can be found in Appendix A.

Consistency Factors. For each object we determined, whether the equivalences referred to in Hypotheses **H1** to **H3** hold, yielding three dichotomous factors: Trace Equivalence, Behavioural Profile Equivalence, and Behavioural Profile Equivalence without Repetition. We saw that five, seven, and nine pairs of process models were trace equivalent, behavioural profile equivalent, and behavioural profile equivalent without repetition, respectively. That, in turn, illustrates that the selection of process model differences based on existing classifications impacts on all notions of behaviour consistency. The distribution of the different criteria is influenced by the implication of the different properties. Our selection of models reflects the fact that hardly any process model pairs in practice do not fulfil at least the weakest criterion [11]. On the other hand, our selection is balanced for the strictest criterion, trace equivalence, which is satisfied by nearly halve of the pairs of process models.

Instrumentation. We conducted a pre-test succeeded by a round of discussion with 8 practitioners. Their feedback led to the adaptation of some pairs of process models, an update of the introductory description, and updates on model annotations about execution semantics. We also used the feedback to choose the items to be asked for each process model pair. We considered the option that experts might judge the consistency of the process models both based on closeness and on equivalence. Therefore, it was found most appropriate to ask 'I think the processes are similar' (strong disagreement to strong agreement). This question can be related to the closeness of the features of two process models. Furthermore, we asked 'I think the processes represent the same real-world process' (strong disagreement to strong agreement), which relates to equivalent effects of the models in their relation to the real-world process. By having these two separate items, we aim to avoid a bias towards a particular position underlying the experts' perceptions of consistency.

Once these changes were implemented, we published the questionnaire online and left it available for three months. An introductory text discussed the need of having different process models for different perspectives of a common process. No further information on notions of behaviour consistency was given to the participants in order to focus on the pure consistency perception. The website of the questionnaire was advertised via channels having an affinity towards process modelling. These included the SAP Developer Network, the German BPM-Netzwerk, special interest groups on the social business platform XING, BPMN-related blogs such as BPMN.info, academic mailing lists as isworld and the German WI list, and finally via process consulting and vendor companies.

3.2. Demographics & Results

In this section, we describe demographics, distribution of perceptions, hypotheses testing, and checks for potential interactions.

Demographics. The online questionnaire of our experiment was filled out by 157 persons. We performed several steps to guarantee that the analysis data will be of high quality. We considered only answers of participants who completed the full questionnaire. Then, we excluded those entries that where obviously filled out in a rush (answer time less than 10 minutes) and those where participants got distracted by other tasks (answer time more than 100 minutes). Our pre-test with 8 practitioners revealed that it is not possible to reasonably complete the questionnaire in less than 10 minutes. Participants with more than 100 minutes obviously took a break, which might distort their performance data. Furthermore, we excluded those participants that classified themselves as students since we aim to get expert opinions. Finally, there were also cases of persons giving the same consistency assessment to all model pairs. Again, this pattern points to a thoughtless clicking through the online questionnaire, such that we excluded these answers.

The remaining data included the answers of 69 participants, originating from 27 countries from all over the world with a focus on the United States

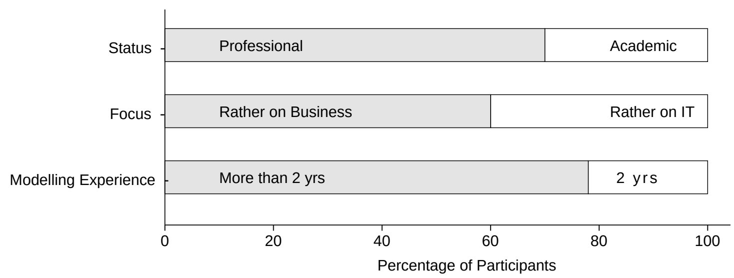


Figure 5: Demographics for the participants of our experiment (expert statements)

(15) and Germany (14). Each participant assessed the 11 pairs of process models such that we have 759 data points altogether. 20% of the participants used the *German version* of the questionnaire and 80% the *English version*. The *answer time* varied between 11 and 85 minutes with a mean of 23 and a standard deviation of 15 (only four participants took more than 50 minutes). As illustrated in Figure 5, 30% of the participants classified themselves as *academics* and 70% as *practitioners* (answers by students were excluded). 60% stated they would rather work on the *business side* of process management and 40% rather on the *IT side*. Only 22% had less than two years while 78% had more than two years of of *modelling experience*. These numbers provide us with some confidence that we actually collected the assessments of experts.

Distribution of Perceptions. In the online questionnaire, we asked for two assessments for each model pair on a 1 to 5 Likert scale ('I think the processes are similar' and 'I think the processes represent the same real-world process') in order to approximate the consistency perception. We observed that participants provided quite consistent answers to these questions. Cronbach's Alpha as a reliability check yielded a good value of 0.76 such that we summed up the two factors into one scale variable called *Perceived Consistency* ranging from 2 to 10. Its mean value was rather central with 6.44 and with a standard deviation of 2.195. We performed the Kolmogorov-Smirnov test which indicated that its values are normally distributed. This is an important criterion for making analysis of variance (ANOVA) testing applicable.

Hypotheses Testing. The major part of the online questionnaire focused on the relation between different consistency notions and the assessment of the participants. In order to investigate Hypotheses **H1** to **H3**, we first derived descriptive statistics and constructed boxplots for the *Perceived Consistency*

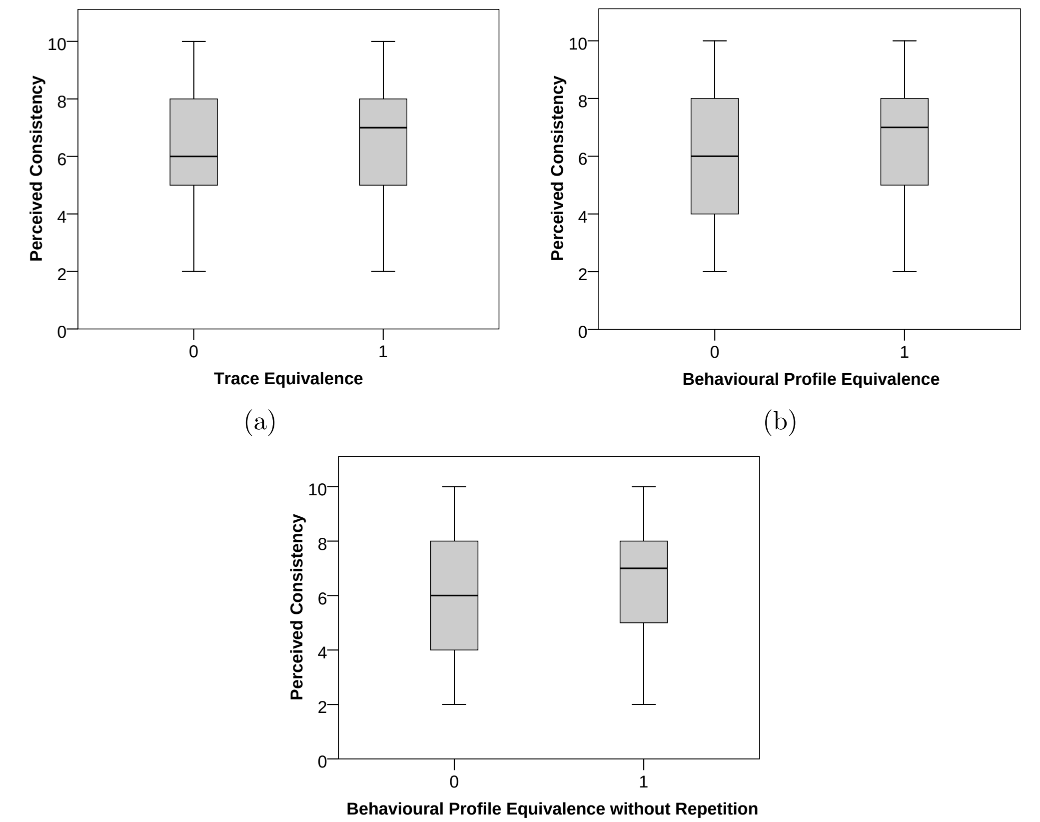
Table 1: Descriptive statistics for the perceived consistency

	Mean	Median	Standard Deviation
All Objects	6.44	6.00	2.195
Trace Equivalence	6.47	7.00	2.136
No Trace Equivalence	6.41	6.00	2.245
Behavioural Profile Eq.	6.55	7.00	2.170
No Behavioural Profile Eq.	6.23	6.00	2.228
Behavioural Profile Eq. w/o Repetition	6.62	7.00	2.167
No Behavioural Profile Eq. w/o Repetition	5.63	6.00	2.148

and each of the three consistency notions. Then, we inspected rank correlations and examined differences in variance of *Perceived Consistency* for each of the notions.

The boxplot for *Perceived Consistency* and Trace Equivalence indicates partial support for the theoretical assumption behind Hypothesis **H1** (see Figure 6(a)). When considering the median, trace equivalent model pairs are perceived to be more consistent by the participants. However, the range covered by the lower and upper quartiles is equal and the mean value is virtually equal for the models that are trace equivalent and those that are not, see also Table 1. The boxplot and the descriptive statistics for *Perceived Consistency* and Behavioural Profile Equivalence are in line with Hypothesis **H2** (see Figure 6(b)). The model pairs which have an equivalent profile are also perceived to be more consistent. A similar observation can be made for behavioural profile equivalence without repetition (see Figure 6(c) and Table 1). In contrast to the first boxplot in Figure 6(a), the plots for the criteria based on behavioural profiles indicate a bigger variance in perception if the model pair is inconsistent according to these criteria. In particular, the lower quartile of the perceived consistency is larger meaning that more participants tended to give low consistency values. This difference in the consistency perception is also reflected in the mean values given in Table 1.

As a next step, we inspected whether the observed differences are significant from a statistical point of view. We identified a minimal positive Spearman rank correlation [77] between Trace Equivalence and *Perceived Consistency* of 0.007 which is insignificant (p=.855). There is a positive correlation between Behavioural Profile Equivalence and *Perceived Consistency* of 0.068 (p=.059), and also a significant positive correlation between Behavioural



(a) (b) (c)

Figure 6: Trace Equivalence (a), Behavioural Profile Equivalence (b), and Behavioural Profile Equivalence without Repetition (c) versus Perceived Consistency

Profile Equivalence without Repetition and *Perceived Consistency* of 0.171 (p=.001) at a significance level of 99%. These facts are not in line with **H1** and only partially with **H2**, while they comply with the assumptions of **H3**.

Additionally, we performed analysis of variance tests for each of the consistency notions and *Perceived Consistency* (see Table 2). None of the Levene statistics was significant, such that homogeneity of variance can be assumed. For Trace Equivalence and *Perceived Consistency* the difference is not significant with p=.713 and an F of 0.135. This fact points to the rejection of **H1**. For Behavioural Profile Equivalence and *Perceived Consistency* the significance of the difference is below the 95% level with p=.051 and an F of 3.816. This does not support **H2**. For Behavioural Profile Equivalence

Table 2: Support for Hypotheses **H1** to **H3**

	H1	**H2**	**H3**
F-Statistic	0.135	3.816	23.469
Significance	.713	.051	.000
Support	no support	no support	strong support

without Repetition and *Perceived Consistency* the difference is significant with p=.000 and an F of 23.469. This fact supports **H3** strongly.

Potential Interactions. Finally, we considered the set of moderating factors and checked potential interactions between equivalence notions and personal variables *status* (academic or practitioner), *focus* (business or IT), and *modelling experience* (less than two years or more than two years). For this purpose, we calculated an ANOVA for *Perceived Consistency* using these three personal variables and the three notions as input. Table 3 shows the most important variables of this ANOVA model and their interactions. It can be seen that Behavioural Profile Equivalence without Repetition is the most significant factor with more than 99% significance. Focus is also significant as a factor. That stems from the fact that IT persons gave slightly higher consistency perceptions than business participants. There is no significant interaction effect of any of these variables on the relation between each of the three consistency notions and *Perceived Consistency* except for Trace Equivalence and focus. We inspected the data in further detail and found that business people showed a slightly positive correlation between consistency perceptions and trace equivalence, while there was no connection for IT people (neither positive nor negative). Both correlations were insignificant.

While only one of the background factors is significant and there is only one significant interaction, our results show that consistency perception is not completely independent of the personal background. As mentioned above, business people tend to judge the consistency of our experiment objects on a lower level. Against this background, it is remarkable that Hypothesis **H3** is still supported with more than 99% significance. The correlation between *Perceived Consistency* and Behavioural Profile Equivalence without Repetition is very strong, such that differences with respect to *Perceived Consistency* between business and IT people are of no consequence. The significant interaction between Trace Equivalence and work focus can be expected to stem from the same observation. There are less objects that show

Table 3: ANOVA model for *Perceived Consistency*

			Trace Equivalence	Behavioural Profile Eq.	Behavioural Profile Eq. w/o Repetition
	F-Stat.		1.997	0.003	6.804
	Sign.		.158	.957	.009
Status	F-Stat.	0.001	1.062	0.612	0.100
	Sign.	.976	.303	.434	.752
Focus	F-Stat.	2.891	2.641	2.112	0.605
	Sign.	.035	.048	.097	.612
Modelling Experience	F-Stat.	0.504	0.321	0.050	1.313
	Sign.	.680	.811	.985	.269

Trace Equivalence than objects that show the other two consistency criteria. Hence, the fact that business people gave lower consistency values yields a correlation with Trace Equivalence, which is still not significant.

4. Implications

This section discusses the implications of our research. After a general discussion of our results in Section 4.1, we elaborate on potential threats to validity in Section 4.2. Subsequently, we draw implications for research as well as for practice in Section 4.3 and Section 4.4, respectively.

4.1. Discussion of Results

Our results on testing Hypothesis **H1** on consistency perception suggests that the trace equivalence criterion not suited to be used as a consistency criterion for process models that capture different perspectives of a common process. Even though trace equivalence is the most relaxed criterion of the aforementioned linear time – branching time spectrum, it turns out to be too strict in this context. The results presented in Section 3.2 suggests that the trace equivalence criterion and the consistency perception are not correlated. The rejection of Hypothesis **H1** shows that trace equivalence does not necessarily imply consistency. Hence, any analysis of the validity of a business process support system that is based on this criterion might lead to results that are not intuitive for process stakeholders. This observation is not surprising in the light of the context that led to the definition of the trace equivalence criterion. Hoare introduced equivalence of traces for Communicating Sequential Processes (CSP), a formal model for the description of concurrent, distributed, and non-deterministic computations [78]. Still, correctness of computer programs and investigations on the design of programming languages have been primary concerns for the creation of CSP and the trace equivalence criterion. These use cases impose different requirements than the consistency analysis of conceptual behavioural models that is investigated in this paper. While the equivalence of computer programs requires all execution paths to the available in both equivalent programs, it appears that process models often focus at least partially on the so-called happy path [65, 66, 67]. Accordingly, as modelling experts are aware that not all execution options are explicitly captured, the perception of consistency seems to closer correlate with the weaker notions introduced in this paper.

The tests for Hypothesis **H3** illustrate that the notions of equivalence based on behavioural profiles are more suited as they better approximate the consistency perception of process modelling experts. While the tests for **H2** are not significant at the 95% level, the behavioural profile equivalence that neglects potential repetition (**H3**) led to a highly significant correlation with the perceived consistency. These notions are less strict than the trace equivalence criterion. However, it is important to see that consistency requirements are relaxed in a well-motivated manner. That is, *freedom of contradictions* between the behaviour of two process models is interpreted in a way that the *order* of potential activity occurrences has to be preserved, whereas optionality of activity execution and causal dependencies are considered to be negligible. Any conclusion in the sense 'the more relaxed a notion of consistency, the closer to expert judgement', therefore, must not be drawn.

Another important aspect of our experiment are the findings regarding the personal background of the subjects, that is, the participants of the questionnaire. The status (academia or industry) does not have an influence on the consistency perception. The argument that people from academia would be stricter in their consistency assessment potentially due to a formal background and less experience with real-world settings is not supported by our data. In addition, modelling experience did not turn out to have an effect for our results. Our results are twofold with respect to the fact, whether a participant classifies their work to reside on the business side or the IT side of process management. While this difference is often called to account for failing process management efforts, its impact on the consistency perception of process models illustrating these different perspectives is reflected solely in the absolute level of consistency values. Hence, there seems to be a common understanding with respect to the inevitability of certain differences between such process models. Nevertheless, business people tend to give lower consistency values for our objects in general. We consider this to be a remarkable, yet unexpected, observation. IT-people are typically concerned with concrete implementations of business processes, whereas business people stay on the rather conceptual side. Hence, the latter could be expected to show a more relaxed understanding of consistency.

4.2. Potential Threats to Validity

There are some threats to validity of our research. We want to highlight self-recruitment and the selection of experiment objects.

We already mentioned that we did not try to apply a random sampling, but rather relied on self-recruitment. This implies a certain threat to representativeness. Our strategy towards good external validity was to make sure that the sample of participants that we considered in our statistical evaluation can actually be regarded as modelling experts with some confidence. Our assumption in this context is that modelling experts in general share a common understanding of what consistency is. If there were different schools of thought that led to diverging perceptions on consistency, there might be a bias in our data if these schools were not covered in a representative way. However, we do not possess any evidence that such differences exist. In our various discussions with modelling experts we observed a strong agreement among them on consistency matters.

The selection of model pairs could introduce a potential threat to validity. It might be the case that perceptions on consistency vary along with the size and complexity of process models. The pairs that we choose were rather compact in this dimension. The choice was made for different reasons. First, a pair of process models should always show solely one difference as classified in related work [3, 4, 5]. In this way, we ensured that any obligations of a participant towards a dedicated difference would have a minor impact on the result. In such a case, the perceived consistency would be affected for solely one object instead of for multiple objects. Second, we would have run into complex interactions if we had varied the size and complexity of the pairs. It is well known from research on process model metrics that size and complexity have a negative effect on process model comprehension [79]. This implies that larger models would likely introduce a substantial level of noise in the data, as the chance of bad comprehension (and therefore inadequate consistency assessment) rises with complexity. As a consequence of this fact, we selected rather small model pairs to limit the cognitive effort of the participants and reduce the risk of participants leaving the questionnaire without completion.

4.3. Implications for Research

We want to highlight four major implications for research, related to extensions of the linear time – branching time spectrum, adaptations of similarity metrics for consistency analysis, theories of meaning for conceptual models, and the importance of utility considerations for formal criteria in process management.

Above, we discussed that our experiments revealed that consistency based on behavioural profiles better approximates experts' perceptions on process model consistency than trace equivalence. Thus, process models capturing different perspectives of a common process impose other consistency requirements than the widely studied areas of behaviour inheritance or business-to-business integration. As a consequence, the results obtained for these research areas cannot be transferred to our setting in a straight-forward manner. This clearly shows the need to consider behavioural abstraction beyond the behaviour equivalences of the linear time – branching time spectrum as the grounding for consistency analysis. Behavioural profiles are such an abstraction, but there may be further notions to be investigated on the downwards side in terms of strictness.

Behavioural profile equivalence is a boolean criterion as much as trace equivalence. There is considerable potential in terms of refining it. Further adaptations might be inspired by the work on similarity metrics of process models such as [80, 81, 82, 83, 84, 86, 85, 87, 88, 89]. In particular, process similarity metrics that exploit the process behaviour may be leveraged to define a notion of behaviour consistency for related process models. To this end, approaches might use behavioural abstractions of the process [84], relate similar (sub-) traces of two models to each other [81, 82], or quantify the degree of state-based simulation [83]. It seems reasonable to adapt these approaches to measure behaviour consistency in a metric way such that potential deviations in the process behaviour can be quantified. The proximity of these metrics and consistency perception can again be challenged in an experimental way.

All the three criteria used in this study have a relation to different theories of meaning. Evermann has investigated the impact of these theoretical positions on perceived similarity of data models [8]. While the underlying theories of meaning have not been a focal point of our study, we foresee potential to investigate this aspect in more detail in future research. Similarity of effect as emphasized by early pragmatists and incorporated in trace equivalence seem to be less important in our setting than in the schema matching setting as investigated by Evermann. There is some evidence that process model features in terms of behavioural profile constraints are well suited to approximate expert judgements on consistency. This support for feature theory of meaning requires further research.

Further on, our research highlights the importance of empirical research on formal correctness criteria for process management. This observation is not limited to scenarios where consistency matters. Many formal correctness criteria such as *soundness* or *relaxed soundness* [90, 91] have been defined with a clear motivation of utility. This utility can be evaluated to a great share using empirical research methods. Up until now, no experiments have been conducted to prove their utility. We foresee that a feedback loop between research on formal properties and the evaluation of their usefulness in a certain context might be of great benefit to advance the field of process management. Moreover, our results provide a concrete starting point for further empirical research in this domain. The observation of business people evaluating the consistency of process model pairs on a lower level than people with an IT background clearly needs further investigations.

4.4. Implications for Practice

Current process modelling tools rarely support the alignment of models representing different perspectives of a common process and if so, there are typically no means of consistency analysis. The latter is no surprise given the fact that established notions like trace equivalence seem to be too strict for this purpose. Consequently, our findings can be seen as a stimulus for the realisation of consistency analysis in commercial process modelling tools based on behavioural profiles. The notion of behaviour consistency is also crucial for change propagation support between two process models. Our work can be regarded as a foundation for the implementation of such support in process modelling tools.

5. Conclusion

In this paper, we took the variety of application scenarios for process-aware information systems as a starting point and motivated the need for consistency analysis of the underlying process models. In this regard, validation of IT infrastructure supporting a business process and the assessment and propagation of process changes have to be seen as major drivers. We argued that, once correspondences are in place, any difference between process models capturing different perspectives of a common process can be evaluated whether it preserves a certain notion of behaviour consistency. While there is no doubt about the importance of behaviour consistency, investigations on how process modelling experts actually assess behaviour consistency are missing.

Addressing the need for empirically founded insights into the correlation of behaviour consistency notions and the consistency perception by process modelling experts, we reported on the results of an experiment to compare different behavioural consistency notions. In particular, we used on online questionnaire to investigate how well trace equivalence, which is often applied as a consistency criterion in different contexts, and two alternative criteria based on behavioural profiles match the consistency perception. Our findings demonstrate that trace equivalence is less suited as a consistency notion as it is less correlated with the consistency perception. In contrast, the criterion based on behavioural profiles that neglects repetition of single activities appeared to approximate the consistency perception best. Moreover, we discussed our results in the light of the modelling expert's background.

Our research provides empirically sound insights into the perception of behaviour consistency notions. Yet, we have to reflect on some limitations. First, for our investigations on behaviour consistency, we assumed behaviour consistency to be a symmetric concept. That is, we reject the prospect of any consistency notion that might consider one model to be consistent towards a second model, but not vice versa. Although the discussions with process modelling experts as part of the pre-test suggested this assumption, we have not tested it as part of our research in order to reduce complexity of the setup. Note that the question whether behaviour consistency is a symmetric concept is independent of the direction of correspondences as discussed in Section 2.2.

Second, we approached the general question of consistency from the angle of process behaviour. Most probably, this is only one factor influencing the perception of process model consistency. Other dimensions might be the labelling of process model elements or structural factors; see the work in the field of process model similarity [80]. For instance, different structural patterns realising the very same execution semantics in two process models might impact on the perceived consistency. Note that we eliminated effects due to these aspects in our experiments by using equal labels for corresponding elements and limiting structural deviations to a minimum throughout all pairs of process models.

Besides the fact that behaviour equivalences such as trace equivalence do not approximate consistency perception in our context, they are computationally hard [10, 45, 46]. In prior work, we showed the efficient calculation of behavioural profiles [11] for a certain subclass of process models that do not exhibit behavioural anomalies and are based on free-choice constructs. In future work, we aim to develop techniques for alleviating the state explosion problem when computing behavioural profiles in the general case.

We also assumed the different perspectives of a business process to be captured in process models. However, it is a common observation that certain, often rather technical perspectives, are captured in a data centric way. That is, instead of specifying the process flow, valid state transitions of business objects are specified, such that the processing logic can be deduced [92]. While there has been initial work on consistency aspects between process models and object life-cycle models [93], the insights gained from our experiments can guide further research in this area.

Acknowledgements

We would like to thank the anonymous reviewers for constructive comments and valuable feedback. Their suggestions have helped us to improve the quality of this paper.

[1] M. Dumas, W. M. van der Aalst, A. H. ter Hofstede, Process-aware information systems: bridging people and software through process technology, John Wiley & Sons, Inc., New York, NY, USA, 2005.

[2] M. Weske, Business Process Management: Concepts, Languages, Architectures, Springer-Verlag, Berlin, Heidelberg, Germany, 2007.

[3] M. Henkel, J. Zdravkovic, P. Johannesson, Service-based processes: Design for business and technology, in: M. Aiello, M. Aoyama, F. Curbera, M. P. Papazoglou (Eds.), ICSOC, ACM, 2004, pp. 21–29.

[4] R. M. Dijkman, A Classification of Differences between Similar Business Processes, in: EDOC, IEEE Computer Society, 2007, pp. 37–50.

[5] M. Weidlich, A. Barros, J. Mendling, M. Weske, Vertical Alignment of Process Models - How can we get there?, in: S. Nurcan, R. Schmidt,

P. Soffer, R. Ukor (Eds.), CAiSE 2009 Workshop Proceedings - 10th Workshop on Business Process Modeling, Development, and Support (BPMDS), no. 29 in LNBIP, Springer, 2009, pp. 71–84.

[6] S. Zelewski, Petrinetzbasierte Modellierung komplexer Produktionssysteme - Eine Untersuchung des Beitrags von Petrinetzen zur Prozeßkoordinierung in komplexen Produktionssystemen, insbesondere Flexiblen Fertigungssystemen, Band 9: Beurteilung des Petrinetz-Konzepts, Tech. rep., University of Leipzig, (in German) (1995).

[7] V. Basili, L. Briand, W. Melo, A validation of object-oriented design metrics as quality indicators, IEEE Transactions on Software Engineering 22 (10) (1996) 751–761.

[8] J. Evermann, Theories of meaning in schema matching: An exploratory study, Inf. Syst. 34 (1) (2009) 28–44.

[9] J. Mendling, H. A. Reijers, J. Recker, Activity labeling in process modeling: Empirical insights and recommendations, Inf. Syst. 35 (4) (2010) 467–482.

[10] R. van Glabbeek, Handbook of Process Algebra, Elsevier, 2001, Ch. The linear time - branching time spectrum I. The semantics of concrete, sequential processes, pp. 3–99.

[11] M. Weidlich, J. Mendling, M. Weske, Efficient consistency measurement based on behavioural profiles of process models, IEEE Transactions on Software Engineering. To appear.

[12] J. B. Hill, M. Cantara, M. Kerremans, D. C. Plummer, Magic Quadrant for Business Process Management Suites, Gartner Research G00164485.

[13] B. Mutschler, M. Reichert, J. Bumiller, Unleashing the effectiveness of process-oriented information systems: Problem analysis, critical success factors, and implications, IEEE Transactions on Systems, Man, and Cybernetics, Part C 38 (3) (2008) 280–291.

[14] M. Weidlich, M. Weske, J. Mendling, Change propagation in process models using behavioural profiles, in: IEEE SCC, IEEE Computer Society, 2009, pp. 33–40.

[15] J. Recker, J. Mendling:, On the Translation between BPMN and BPEL: Conceptual Mismatch between Process Modeling Languages, in: T. Latour, M. Petit (Eds.), CAiSE 2006 Workshop Proceedings - Eleventh International Workshop on Exploring Modeling Methods in Systems Analysis and Design (EMMSAD 2006), 2006, pp. 521–532.

[16] M. Weidlich, G. Decker, A. Großkopf, M. Weske, Bpel to bpmn: The myth of a straight-forward mapping, in: Meersman and Tari [94], pp. 265–282.

[17] M. Schrefl, M. Stumptner, Behavior consistent extension of object life cycles, in: M. P. Papazoglou (Ed.), OOER, Vol. 1021 of Lecture Notes in Computer Science, Springer, 1995, pp. 133–145.

[18] M. Schrefl, M. Stumptner, Behavior consistent refinement of object life cycles, in: D. W. Embley, R. C. Goldstein (Eds.), ER, Vol. 1331 of Lecture Notes in Computer Science, Springer, 1997, pp. 155–168.

[19] M. Schrefl, M. Stumptner, Behavior-consistent specialization of object life cycles, ACM Trans. Softw. Eng. Methodol. 11 (1) (2002) 92–148.

[20] D.-R. Liu, M. Shen, Workflow Modeling for Virtual Processes: an Order-Preserving Process-View Approach, Information Systems 28 (6) (2003) 505–532.

[21] R. Eshuis, P. W. P. J. Grefen, Constructing Customized Process Views, Data & Knowledge Engineering 64 (2) (2008) 419–438.

[22] X. Zhao, C. Liu, W. Sadiq, M. Kowalkiewicz, Process view derivation and composition in a dynamic collaboration environment, in: Meersman and Tari [94], pp. 82–99.

[23] R. Bobrik, M. Reichert, T. Bauer, View-Based Process Visualization, in: Alonso et al. [95], pp. 88–95.

[24] B. Andersson, M. Bergholtz, A. Edirisuriya, T. Ilayperuma, P. Johannesson, A Declarative Foundation of Process Models, in: O. Pastor, J. F. e Cunha (Eds.), CAiSE, Vol. 3520 of Lecture Notes in Computer Science, Springer, 2005, pp. 233–247.

[25] J. Koehler, R. Hauser, J. M. Küster, K. Ryndina, J. Vanhatalo, M. Wahler, The Role of Visual Modeling and Model Transformations in Business-driven Development, Electr. Notes Theor. Comput. Sci. 211 (2008) 5–15.

[26] V. Grover, K. Fiedler, J. Teng, Exploring the Success of Information Technology Enabled Business Process Reengineering, IEEE Transactions on Engineering Management 41 (3) (1994) 276–284.

[27] C. Rolland, N. Prakash, Bridging the Gap Between Organisational Needs and ERP Functionality, Requirements Engineering 5 (3) (2000) 180–193.

[28] G. Preuner, S. Conrad, M. Schrefl, View integration of behavior in object-oriented databases., Data & Knowledge Engineering 36 (2) (2001) 153–183.

[29] V. Pankratius, W. Stucky, A formal foundation for workflow composition, workflow view definition, and workflow normalization based on petri nets., in: S. Hartmann, M. Stumptner (Eds.), Conceptual Modelling 2005, Second Asia-Pacific Conference on Conceptual Modelling (APCCM2005), Newcastle, NSW, Australia, January/February 2005, Vol. 43 of CRPIT, Australian Computer Society, 2005, pp. 79–88.

[30] J. Mendling, C. Simon, Business Process Design by View Integration, in: J. Eder, S. Dustdar (Eds.), Proceedings of BPM Workshops 2006, Vol. 4103 of Lecture Notes in Computer Science, Springer-Verlag, Vienna, Austria, 2006, pp. 55–64.

[31] J. A. Garda, J. A. Casal, R. G. Vázquez, J. Pazos, S. R. Yánez, A. Silva, A methodological framework for viewpoint-oriented conceptual modeling, IEEE Trans. Software Eng. 30 (5) (2004) 282–294.

[32] E. A. Boiten, H. Bowman, J. Derrick, P. F. Linington, M. Steen, Viewpoint consistency in odp, Computer Networks 34 (3) (2000) 503–537.

[33] H. Bowman, M. Steen, E. A. Boiten, J. Derrick, A formal framework for viewpoint consistency, Formal Methods in System Design 21 (2) (2002) 111–166.

[34] B. Nuseibeh, J. Kramer, A. Finkelstein, A framework for expressing the relationships between multiple views in requirements specification, IEEE Trans. Software Eng. 20 (10) (1994) 760–773.

[35] A. Finkelstein, D. M. Gabbay, A. Hunter, J. Kramer, B. Nuseibeh, Inconsistency Handling in Multperspective Specifications, IEEE Transactions on Software Engineering 20 (8) (1994) 569–578.

[36] R. M. Dijkman, D. A. C. Quartel, M. van Sinderen, Consistency in multi-viewpoint design of enterprise information systems, Information & Software Technology 50 (7-8) (2008) 737–752.

[37] S. Nejati, M. Sabetzadeh, M. Chechik, S. M. Easterbrook, P. Zave, Matching and merging of statecharts specifications, in: ICSE, IEEE Computer Society, 2007, pp. 54–64.

[38] R. Dijkman, M. Dumas, L. García-Bauelos, R. Kääriky, Aligning business process models, in: 13th International IEEE Enterprise Distributed Object Computing Conference (ECOC), 2009, pp. 45–53.

[39] E. Rahm, P. A. Bernstein, A survey of approaches to automatic schema matching, VLDB Journal 10 (4) (2001) 334–350.

[40] J. Euzenat, P. Shvaiko, Ontology Matching, Springer-Verlag, 2007.

[41] R. Dhamankar, Y. Lee, A. Doana, A. Halevy, P. Domingos, imap: Discovering complex semantic matches between database schemas, in: Proceedings of the International Conference on Management of Data (SIGMOD), 2004, pp. 383–394.

[42] L. Xu, D. W. Embley, Discovering direct and indirect matches for schema elements, in: DASFAA, IEEE Computer Society, 2003, pp. 39–46.

[43] M. Weidlich, R. Dijkman, J. Mendling, The icop framework: Identification of correspondences between process models, in: Proceedings of the 22nd International Conference on Advanced Information Systems Engineering (CAiSE'10), Vol. 6051 of LNCS, Springer-Verlag, 2010, pp. 483–498.

[44] U. Frank, Reference Modeling for Business Systems Analysis, Idea Group, 2007, Ch. Evaluation of Reference Models, pp. 118–139.

[45] L. Pomello, G. Rozenberg, C. Simone, A survey of equivalence notions for net based systems, in: G. Rozenberg (Ed.), Advances in Petri Nets: The DEMON Project, Vol. 609 of Lecture Notes in Computer Science, Springer, 1992, pp. 410–472.

[46] J. Hidders, M. Dumas, W. M. P. van der Aalst, A. H. M. ter Hofstede, J. Verelst, When are two Workflows the Same?, in: M. D. Atkinson, F. K. H. A. Dehne (Eds.), CATS, Vol. 41 of CRPIT, Australian Computer Society, 2005, pp. 3–11.

[47] J. Ebert, G. Engels, Observable or Invocable Behaviour - You Have to Choose, Technical Report 94-38, Department of Computer Science, Leiden University (December 1994).

[48] W. M. P. van der Aalst, T. Basten, Life-cycle inheritance: A petri-net-based approach, in: Azéma and Balbo [96], pp. 62–81.

[49] T. Basten, W. M. P. van der Aalst, Inheritance of Behavior, Journal of Logic and Algebraic Programming (JLAP) 47 (2) (2001) 47–145.

[50] M. Weidlich, R. Dijkman, M. Weske, Deciding behaviour compatibility of complex correspondences between process models, in: Proceedings of the 8th International Conference on Business Process Management (BPM), Vol. 6336 of LNCS, Springer-Verlag, 2010, pp. 78–94.

[51] W. Brauer, R. Gold, W. Vogler, A survey of behaviour and equivalence preserving refinements of petri nets, in: G. Rozenberg (Ed.), Applications and Theory of Petri Nets, Vol. 483 of Lecture Notes in Computer Science, Springer, 1989, pp. 1–46.

[52] J. L. Peterson, Petri nets, ACM Comput. Surv. 9 (3) (1977) 223–252.

[53] T. Murata, Petri nets: Properties, analysis and applications, Proceedings of the IEEE 77 (4) (1989) 541–580.

[54] B. Nuseibeh, J. Kramer, A. Finkelstein, Viewpoints: meaningful relationships are difficult!, in: ICSE, IEEE Computer Society, 2003, pp. 676–683.

[55] A. Schönberger, G. Wirtz, Taxonomy on consistency requirements in the business process integration context, in: SEKE, Knowledge Systems Institute Graduate School, 2008, pp. 593–598.

[56] J. M. Zaha, M. Dumas, A. H. M. ter Hofstede, A. P. Barros, G. Decker, Bridging global and local models of service-oriented systems, IEEE Transactions on Systems, Man, and Cybernetics, Part C 38 (3) (2008) 302–318.

[57] B. Benatallah, F. Casati, F. Toumani, Representing, analysing and managing web service protocols, Data Knowl. Eng. 58 (3) (2006) 327–357.

[58] A. Martens, Consistency between executable and abstract processes, in: EEE, IEEE Computer Society, 2005, pp. 60–67.

[59] G. Decker, M. Weske, Behavioral consistency for b2b process integration, in: J. Krogstie, A. L. Opdahl, G. Sindre (Eds.), CAiSE, Vol. 4495 of Lecture Notes in Computer Science, Springer, 2007, pp. 81–95.

[60] C. Canal, E. Pimentel, J. M. Troya, Compatibility and inheritance in software architectures, Sci. Comput. Program. 41 (2) (2001) 105–138.

[61] X. Zhao, H. Yang, Z. Qiu, Towards the formal model and verification of web service choreography description language, in: M. Bravetti, M. Núñez, G. Zavattaro (Eds.), WS-FM, Vol. 4184 of Lecture Notes in Computer Science, Springer, 2006, pp. 273–287.

[62] W. M. P. van der Aalst, N. Lohmann, P. Massuthe, C. Stahl, K. Wolf, From public views to private views - correctness-by-design for services, in: M. Dumas, R. Heckel (Eds.), WS-FM, Vol. 4937 of Lecture Notes in Computer Science, Springer, 2007, pp. 139–153.

[63] N. Lohmann, P. Massuthe, C. Stahl, D. Weinberg, Analyzing interacting bpel processes, in: S. Dustdar, J. L. Fiadeiro, A. P. Sheth (Eds.), Business Process Management, Vol. 4102 of Lecture Notes in Computer Science, Springer, 2006, pp. 17–32.

[64] R. J. van Glabbeek, U. Goltz, Refinement of actions and equivalence notions for concurrent systems, Acta Inf. 37 (4/5) (2001) 229–327.

[65] K. Lunn, A. Sixsmith, A. Lindsay, M. Vaarama, Traceability in requirements through process modelling, applied to social care applications, Information and Software Technology 45 (15) (2003) 1045–1052.

[66] B. Andersson, M. Bergholtz, A. Edirisuriya, T. Ilayperuma, P. Johannesson, A declarative foundation of process models, in: O. Pastor, J. F. e Cunha (Eds.), Advanced Information Systems Engineering, 17th International Conference, CAiSE 2005, Porto, Portugal, June 13-17, 2005, Proceedings, Vol. 3520 of Lecture Notes in Computer Science, Springer, 2005, pp. 233–247.

[67] D. Miers, Best practice (BPM), Queue 4 (2) (2006) 40–48.

[68] J. Koehler, G. Tirenni, S. Kumaran, From business process model to consistent implementation: A case for formal verification methods, in: EDOC, IEEE Computer Society, 2002, pp. 96–106.

[69] L. Wittgenstein, Philosophical Investigations, Prentice Hall, Englewood Cliffs, 1953.

[70] G. Frege, Sense and reference, The Philosophical Review (1948) 209–230.

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

[72] M. Petre, Why looking isn't always seeing: Readership skills and graphical programming, Commun. ACM 38 (6) (1995) 33–44.

[73] J. Recker, A. Dreiling, Does it matter which process modelling language we teach or use? an experimental study on understanding process modelling languages without formal education, in: 18th Australasian Conference on Information Systems, 2007, pp. 356–366.

[74] J. Mendling, H. A. Reijers, J. Cardoso, What makes process models understandable?, in: Alonso et al. [95], pp. 48–63.

[75] M. Ciolkowski, O. Laitenberger, S. Vegas, S. Biffl, Practical Experiences in the Design and Conduct of Surveys in Empirical Software Engineering, in: R. Conradi, A. I. Wang (Eds.), ESERNET, Vol. 2765 of Lecture Notes in Computer Science, Springer, 2003, pp. 104–128.

[76] M. zur Muehlen, J. Recker, How much language is enough? theoretical and practical use of the business process modeling notation, in: Bellahsene and Léonard [97], pp. 465–479.

[77] C. Spearman, The proof and measurement of association between two things, The American journal of psychology 15 (1) (1904) 72–101.

[78] C. A. R. Hoare, Communicating Sequential Processes, Prentice-Hall, 1985.

[79] J. Mendling, Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness, Vol. 6 of Lecture Notes in Business Information Processing, Springer, 2008.

[80] M. Dumas, L. García-Bañuelos, R. M. Dijkman, Similarity search of business process models, IEEE Data Eng. Bull. 32 (3) (2009) 23–28.

[81] A. Wombacher, Evaluation of technical measures for workflow similarity based on a pilot study, in: Meersman and Tari [98], pp. 255–272.

[82] A. K. A. de Medeiros, W. M. P. van der Aalst, A. J. M. M. Weijters, Quantifying process equivalence based on observed behavior, Data Knowl. Eng. 64 (1) (2008) 55–74.

[83] O. Sokolsky, S. Kannan, I. Lee, Simulation-based graph similarity, in: H. Hermanns, J. Palsberg (Eds.), TACAS, Vol. 3920 of Lecture Notes in Computer Science, Springer, 2006, pp. 426–440.

[84] B. F. van Dongen, R. M. Dijkman, J. Mendling, Measuring Similarity between Business Process Models, in: Bellahsene and Léonard [97], pp. 450–464.

[85] R. M. Dijkman, M. Dumas, L. García-Bañuelos, Graph matching algorithms for business process model similarity search, in: U. Dayal, J. Eder, J. Koehler, H. A. Reijers (Eds.), BPM, Vol. 5701 of Lecture Notes in Computer Science, Springer, 2009, pp. 48–63.

[86] R. M. Dijkman, M. Dumas, B. F. van Dongen, R. Käärik, J. Mendling, Similarity of Business Process Models: Metrics and Evaluation, Inf. Syst. 36 (2) (2010) 498–516.

[87] J. C. Corrales, D. Grigori, M. Bouzeghoub, Bpel processes matchmaking for service discovery, in: Meersman and Tari [98], pp. 237–254.

[88] R. Eshuis, P. W. P. J. Grefen, Structural matching of bpel processes, in: ECOWS, IEEE Computer Society, 2007, pp. 171–180.

[89] C. Li, M. Reichert, A. Wombacher, On measuring process model similarity based on high-level change operations, in: Q. Li, S. Spaccapietra, E. S. K. Yu, A. Olivé (Eds.), ER, Vol. 5231 of Lecture Notes in Computer Science, Springer, 2008, pp. 248–264.

[90] W. M. P. van der Aalst, Verification of workflow nets, in: Azéma and Balbo [96], pp. 407–426.

[91] J. Dehnert, P. Rittgen, Relaxed soundness of business processes, in: K. R. Dittrich, A. Geppert, M. C. Norrie (Eds.), CAiSE, Vol. 2068 of Lecture Notes in Computer Science, Springer, 2001, pp. 157–170.

[92] J. M. Küster, K. Ryndina, H. Gall, Generation of business process models for object life cycle compliance, in: Alonso et al. [95], pp. 165–181.

[93] K. Ryndina, J. M. Küster, H. Gall, Consistency of business process models and object life cycles, in: T. Kühne (Ed.), MoDELS Workshops, Vol. 4364 of Lecture Notes in Computer Science, Springer, 2006, pp. 80–90.

[94] R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems: OTM 2008, OTM 2008 Confederated International Conferences, CoopIS, DOA, GADA, IS, and ODBASE 2008, Monterrey, Mexico, November 9-14, 2008, Proceedings, Part I, Vol. 5331 of Lecture Notes in Computer Science, Springer, 2008.

[95] G. Alonso, P. Dadam, M. Rosemann (Eds.), Business Process Management, 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings, Vol. 4714 of Lecture Notes in Computer Science, Springer, 2007.

[96] P. Azéma, G. Balbo (Eds.), Application and Theory of Petri Nets 1997, 18th International Conference, ICATPN '97, Toulouse, France, June 23-27, 1997, Proceedings, Vol. 1248 of Lecture Notes in Computer Science, Springer, 1997.

[97] Z. Bellahsene, M. Léonard (Eds.), Advanced Information Systems Engineering, 20th International Conference, CAiSE 2008, Montpellier, France, June 16-20, 2008, Proceedings, Vol. 5074 of Lecture Notes in Computer Science, Springer, 2008.

[98] R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2006: CoopIS, DOA, GADA, and ODBASE, OTM Confederated International Conferences, CoopIS, DOA, GADA, and ODBASE 2006, Montpellier, France, October 29 - November 3, 2006. Proceedings, Part I, Vol. 4275 of Lecture Notes in Computer Science, Springer, 2006.

Appendix A. Objects of the Experiments

Figure A.7 and Figure A.8 show all objects, i.e., pairs of process models, used in our experiment.

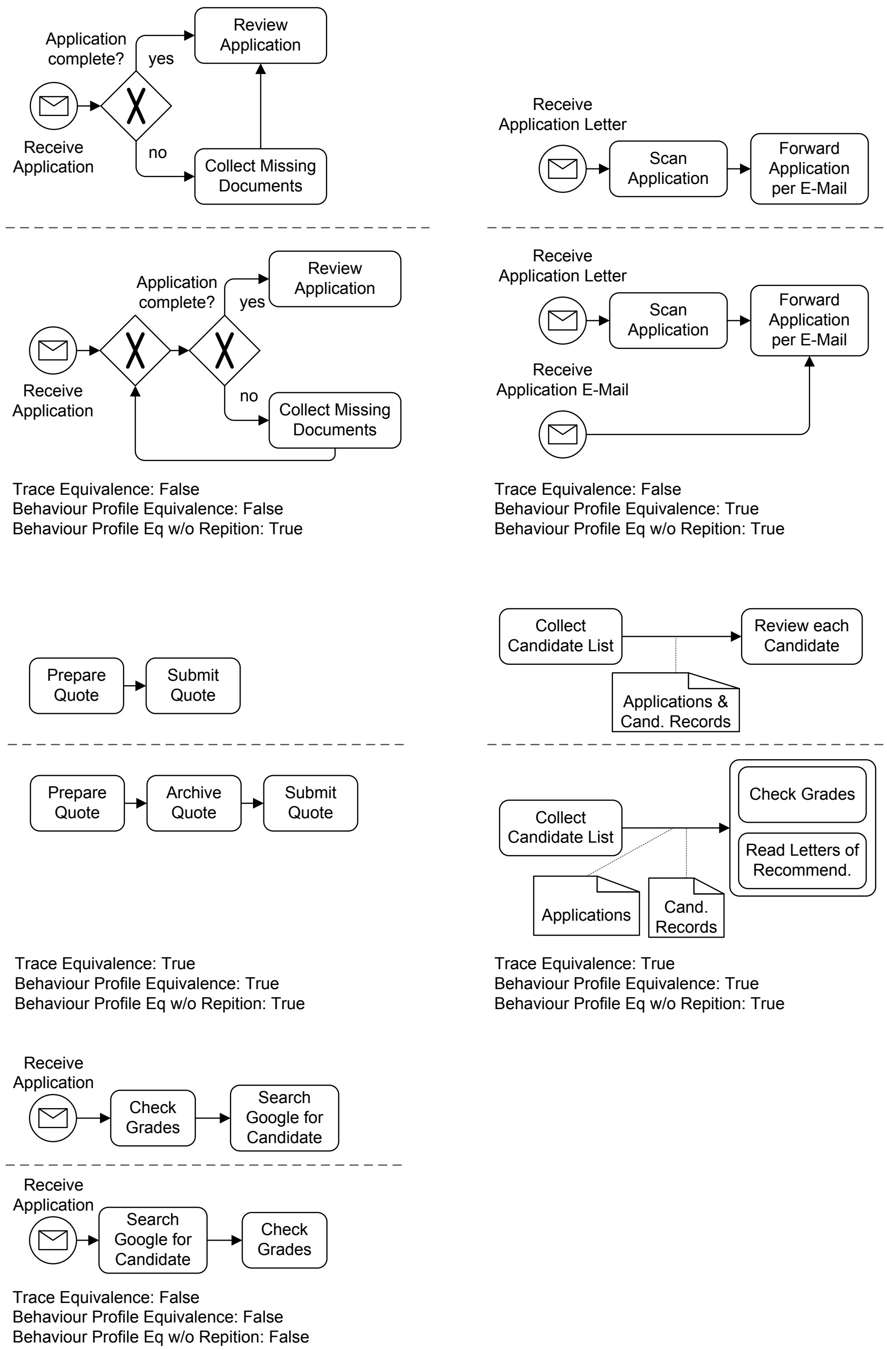


Figure A.7: Objects used in our experiment

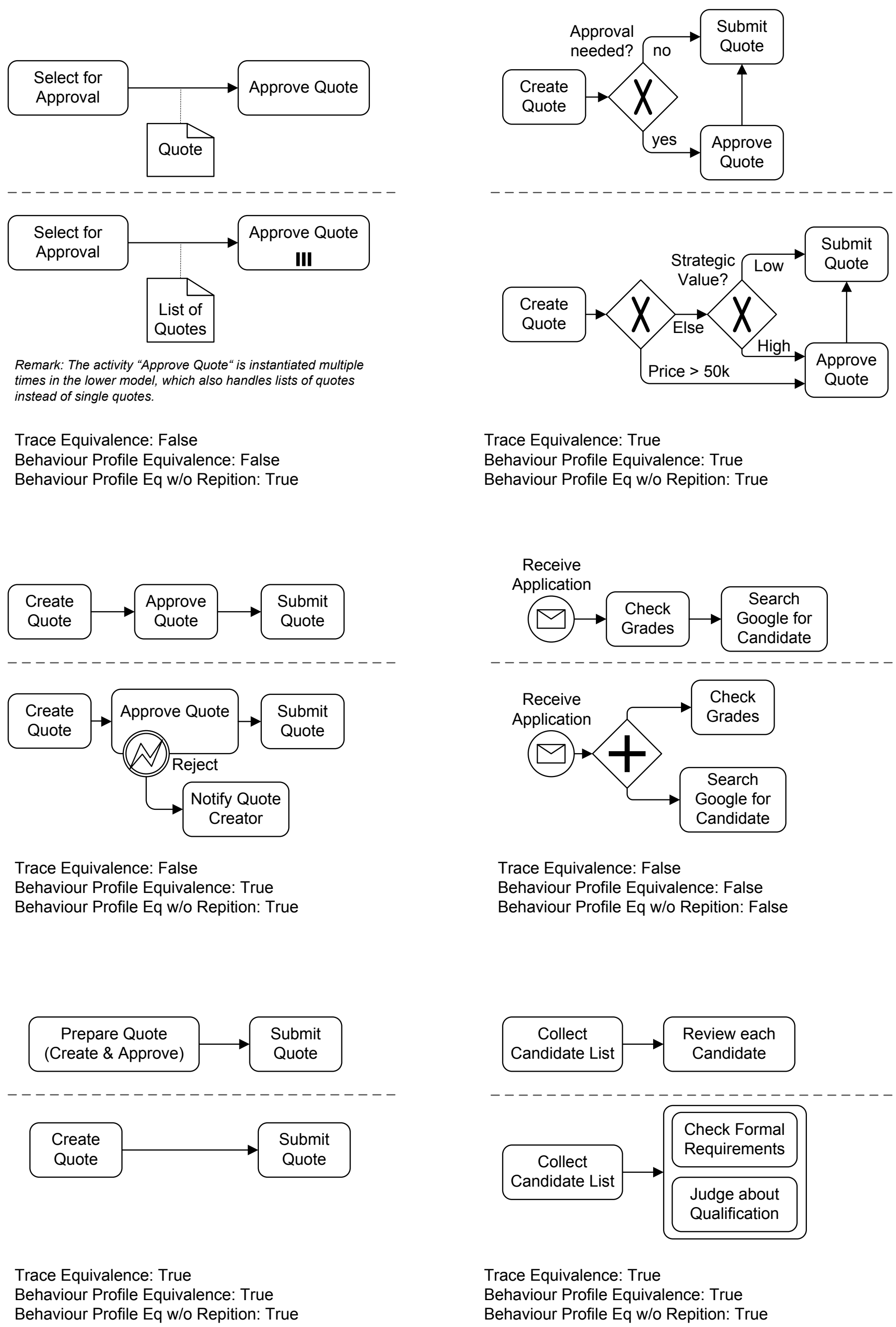


Figure A.8: Objects used in our experiment (Cont.)