Process Mining and the Black Swan: An Empirical Analysis of the Influence of Unobserved Behavior on the Quality of Mined Process Models

Jana-Rebecca Rehse, Peter Fettke, and Peter Loos

Institute for Information Systems (IWi) at the German Center for Artificial Intelligence (DFKI GmbH) and Saarland University
Campus D3 2
66123 Saarbruecken
Germany
{Jana-Rebecca.Rehse,Peter.Fettke,Peter.Loos}@iwi.dfki.de

Abstract. In this paper, we present the epistomological problem of induction, illustrated by the metaphor of the black swan, and its relevance for Process Mining. The quality of mined models is typically measured in terms of four dimensions, namely fitness, precision, simplicity, and generalization. Both precision and generalization rely on the definition of "unobserved behavior", i.e. traces not contained in the log. This paper is intended to analyze the influence of unobserved behavior, the potential black swan, has on the quality of mined models. We conduct an empirical analysis to investigate the relation between a system, its observed and unobserved behavior and the mined models. The results show that the unobserved behavior, mainly determined by the nature of the unknown system, can have a significant impact on the quality assessment of mined models, hence eliciting the need to explicate and discuss the assumptions underlying the notions of unobserved behavior in more depth.

Key words: Process Mining, Process Discovery, Evaluation Metrics, Process Model Quality

1 Introduction

The goal of process discovery is to describe the behavior of an information system in form of a business process model, based on the observed behavior of said information system, represented as an event log [1]. Traditionally, the quality of process discovery algorithms is assessed in terms of four dimensions [1, 2]. *Fitness* measures how much of the observed behavior is captured by the model. *Simplicity* measures the complexity of the model. *Generalization* measures how well the model explains unobserved behavior. *Precision* measures how much unobserved behavior exists in the model.

Several metrics exist for each dimension, quantifying the quality of a model with respect to a given log. Process discovery is targeted towards achieving a satisfactory balance among them [2]. A model should be able to reproduce the behavior in the log, resulting in a high fitness, while being as simple as possible. Also, a model should be precise, i.e. not allowing for behavior greatly different from the log, but also generalizing, i.e. not only allowing for the behavior seen in the log.

Current quality measures typically evaluate the discovered model against the given log. However, an event log is just a limited sample of the behavior of an unknown process system [3]. Since the system is unknown, its behavior is usually divided into "observed behavior", i.e. behavior contained in the given log, and "unobserved behavior", i.e. behavior that is *not* contained in the given log. Both notions are necessary for the quality dimensions above, however, since the unobserved behavior is by definition unknown, it is epistemologically problematic to make universal statements about it without explicating further assumptions. This relates to the epistemological problem of induction, illustrated by the metaphor of the black swan.

The objective of this paper is to demonstrate the relevance the problem of induction has for Process Mining by means of an experiment, analyzing the potential impact unobserved behavior has on the measured quality of mined models. Therefore, we introduce the metaphor of the black swan in Sect. 2. Definitions and preliminary considerations are given in Sect. 3. In Sect. 4, we present our experiment and discuss its impact in Sect. 5. We report on Related Work in Sect. 6, before concluding the paper in Sect. 7.

2 The Black Swan - A Metaphor

In logic, an inference allows the derivation of new statements, called conclusions, from existing ones, called premises. We differentiate deductive inferences from inductive ones. A deductive inference concludes a singular statement from a universal one, whereas an inductive inferences concludes a universal statement from one or several singular ones. From a philosophical point of view, inductive inferences are highly problematic. No matter how many singular statements are considered, any conclusion drawn in this manner may turn out to be false [4].

The "Black Swan" is a metaphor introduced by philosopher Karl Popper to illustrate the difficulties arising with inductive inferencing. In his *Logic of Scientific Discovery*, he states that "no matter how many instances of white swans we may have observed, this does not justify the conclusion that all swans are white" [4]. Independent from the observed number of white swans, the next swan can always be black. The question under which circumstances inductive inferences are justified is what Popper calls the "problem of induction" [4].

This problem also applies to empirical sciences, i.e. those disciplines that try to formulate theories and hypotheses based on experience and observation [4]. Process mining in general and process discovery in particular can be considered empirical science disciplines. Their objective is to gather empirical knowledge in the form of an event log and use it to draw conclusions about the nature of the unknown system underlying this event log. Discovering a model may be interpreted as stating a hypothesis about the design of said system. Assessing its quality then corresponds to testing the validity of the hypothesis.

Given this interpretation of process discovery, each observed trace corresponds to a singular statement, while the discovered model is a universal statement intended to conform with as many of the singular statements as possible. The unobserved behavior is the potential Black Swan. No matter which behavior has been observed so far, neither can it be justified that this is all the observable behavior, nor that the unobserved behavior will follow the same patterns.

So far, this aspect is mainly implicitly addressed by the four dimensions used for assessing the quality of discovered models. In their informal definition above, both precision and generalization rely on the notion of "unobserved behavior". Addressing generalization is particularly complicated, since typically no assumptions can be made as to the nature of the unobserved behavior. However, such assumptions are necessary to make justifiable statements regarding a model's ability to generalize, since different assumptions may lead to completely different understandings of the unobserved behavior and a high generalization depends on correctly assessing this behavior.

3 On Logs, Models, and Systems

3.1 Definitions

Our experiments rely on several definitions and considerations. According to [3], we define traces, logs, models, and systems as follows.

Definition 1 (Traces [3]). *Let $\mathcal{A}$ be the activity alphabet. A trace $\sigma \in \mathcal{A}$ is a finite sequence of activities, where $\mathcal{A}^*$ the set of all finite sequences over $\mathcal{A}$, i.e. the universe of traces.*

Definition 2 (Event Log [3]). *Let $\mathcal{A}^*$ be the universe of traces. An event log $L \in \mathbb{B}(\mathcal{A}^*)$ is a multiset of traces, where $\mathbb{B}(\mathcal{A}^*)$ is the set of all multisets of $\mathcal{A}^*$.*

Definition 3 (Model and System [3]). *Let $\mathcal{A}^*$ be the universe of traces. A model $M \in \mathbb{P}(\mathcal{A}^*)$ and a system $S \in \mathbb{P}(\mathcal{A}^*)$ are subsets of the universe of traces, where $\mathbb{P}(\mathcal{A}^*)$ is the power set of $\mathcal{A}^*$.*

3.2 Observed and Unobserved Behavior

Definition 4 (Process Discovery). *Let L be an event log. A process discovery technique is a function $D : \mathbb{B}(\mathcal{A}^*) \rightarrow \mathbb{P}(\mathcal{A}^*)$ which assigns a discovered model $M \in \mathbb{P}(\mathcal{A}^*)$ to a given event log $L \in \mathbb{B}(\mathcal{A}^*)$.*

In a typical process mining setting, the system is unknown. The objective of process discovery is to find a model, which is as close to the system as possible. The input for this task is a log, i.e. a multiset of traces that is generated by the system. From an epistemological point of view, the log can be seen as empirical

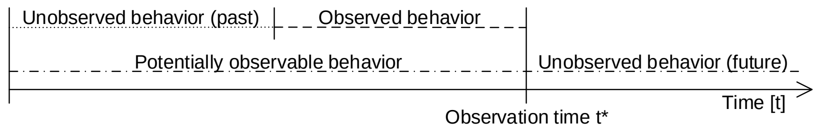

Fig. 1. Temporal characterization of observed behavior

evidence, gathered by means of observing the running system. This means that the system behavior S is divided into two subsets, determined by the point of observation t^*. All system behavior before t^* is potentially observable behavior. This behavior again consist of two subsets. Observed behavior is the system behavior that was witnessed by the observer, i.e. the behavior contained in the log. Unobserved behavior is the system behavior that was not observed, i.e. behavior that happened before t^*, but that is not included in the log. All system behavior after t^* lies in the future and is by definition unobserved. These relations are illustrated in Fig. 1.

3.3 Perspectives on Measuring Model Quality

Fitness, precision, and generalization measures all relate a model M to a log L and an underlying system S. According to the above definitions, S,M,L are all subsets of the same trace universe $\mathcal{A}^*$. Their relations are shown as a Venn diagram in Fig. 2 [5]. Focusing on different interjections in this diagram leads to three different perspectives on measuring the quality of a mined model, each with a different focus and objective.

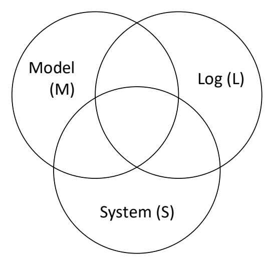

Fig. 2. Perspectives on process behavior [5]

The *Model-Log Perspective* relates the model with the log without considering the system. As systems are typically unknown when discovering a process, this perspective is usually chosen when assessing the quality of discovery approaches [3]. The *Model-System Perspective* inspects the relation between the model and the system, i.e. the model's ability to replicate the unknown system behavior. This is especially relevant when measuring generalization, as this dimension measures a model's ability to reproduce behavior outside of the considered log. The *Log-System Perspective*, relating the log to the generating system, is often not considered in Process Mining, because it does not directly relate to the process model [5]. However, this perspective plays an important role in process discovery, as the relationship between log and system has a direct influence on the ability of a mined model to replicate the system.

4 Demonstrating the Influence of Unobserved Behavior

4.1 Experimental Outline and Goals

The objective of this paper is to analyze the influence unobserved behavior can have on the quality assessment of discovered models. Therefore, we perform an experiment comparing the Log-System Perspective with the Model-Log Perspective. Normally, the logs used for process discovery are neither complete nor accurate. Hence, dealing with the influence of unobserved behavior is a common occurrence in process discovery. Depending on the size of the log and the frequency distribution among the traces, the discovered model may neglect or wrongly emphasize certain system behavior.

To analyze this effect, eight systems over the same activity universe are chosen and different logs are generated for each of them. Each log L is then randomly divided into 10 samples of equal size, $L_0, ..., L_9$. For each sample L_i, we discover a model M_i. These sample-based models are compared against the system S, considering the complete log L. To estimate whether M_i is a good approximation of S, we measure the differences in fitness, precision, and generalization:

$$\Delta F(S, M_i, L) = F(S, L) - F(M_i, L) \tag{1}$$

$$\Delta P(S, M_i, L) = P(S, L) - P(M_i, L) \tag{2}$$

$$\Delta G(S, M_i, L) = G(S, L) - G(M_i, L) \tag{3}$$

If $\Delta F, \Delta P, \Delta G$ are positive, the model M_i underestimates the respective value of S. If they are negative, M_i overestimates S. The smaller the values are, the better model M_i estimates S and the smaller is the influence of unobserved behavior. Please note that we do not evaluate the fourth dimension, simplicity, as it does not directly relate system and log.

4.2 Set-Up

The experiment was designed according to the evaluation framework for comparing process mining algorithms described in [6]. The CoBeFra Benchmarking Framework for Conformance Checking [7] was used for all evaluations.

1. Choose systems: Eight systems, representing different semantics over the same activity universe and shown in Fig. 3 [1], are the basis of our experiment. The models were specifically designed to illustrate the impact of fitness, precision, and generalization, which is why we adopted them here.
2. Generate logs and log samples: For each system, we generate nine logs, varying in size (100, 200, 500 traces) and frequency distribution (uniform, binomial, poisson distribution). Each system is separately used as input for the log generator [8]. Wherever possible, a complete log is generated, otherwise a limit is set on the maximum number of loop iterations. If the possible behavior exceeds the intended log size, traces are picked randomly. Hence, each system log contains only behavior that can be executed on that system,

but some behavior is contained in multiple logs. Frequencies are assigned by computing a pre-defined distribution for the intended log size and assigning them randomly. Finally, each generated log is randomly divided into 10 samples of equal size, resulting in 792 different logs.

3. Discover models: Models are discovered for each sample. To mitigate side effects of individual algorithms, three different miners are used (Heuristics Miner [9], Inductive Miner [10], ILP [2, 11]), resulting in 2,376 total models.
4. Evaluate systems and models: The discovered models are evaluated by comparing the differences in fitness, precision, and generalization between the models and the systems, with regard to the complete log. To mitigate side effects of individual quality metrics, multiple metrics are used. Fitness is measured using Alignment-based Fitness [12], Negative event recall [13], Token-based Fitness [14]. Precision measures are Alignment-based Precision [12], Best-align-etc Precision [15], Negative event Precision [13], One-align-etc Precision [15]. Alignment-based Generalization [12], Negative event Generalization [13] are used to assess generalization.

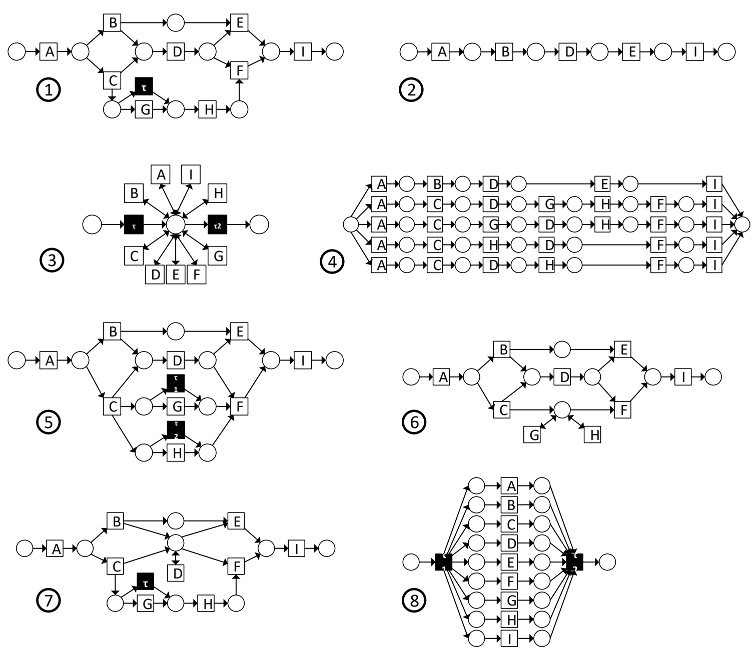

Fig. 3. Systems used in the experiments [1]

4.3 Results

The experiment results are shown in Fig. 4 to 6. Figure 4 shows in the top left corner the distribution of $\Delta F, \Delta P, \Delta G$ over all systems. The remaining plots refer to each system separately, from S1 in the top center to S8 in the bottom right, going from left to right and top to bottom. The distributions of $\Delta F, \Delta P, \Delta G$ for the different frequency distributions are shown in Fig. 5 and Fig. 6 shows how $\Delta F, \Delta P, \Delta G$ are distributed for the different log sizes.

Inspecting the distribution of $\Delta F, \Delta P, \Delta G$ across all systems, the values span a large range, but are fairly similar within a system. This range is therefore caused by the differences between the systems. This also explains the outliers, particularly for the generalization dimension. The system is the main discriminator regarding $\Delta F, \Delta P, \Delta G$, as neither the log sizes nor the frequency distributions exhibit large deviations. Size 500 covering a slightly smaller range is due to the larger samples providing a better estimation of system behavior. The different frequency distributions in the logs have virtually no effect on the measures. This is remarkable, as the frequency assignment of the binomial and poisson distributions caused many traces originally contained in the log to be left out. Apparently, the remaining traces are frequent enough to represent typical process behavior in each sample. The uniform distribution with the most traces does not yield better quality measures, probably caused by the fairly high amount of singular traces, which could be considered as noise.

As shown in Fig. 5, the distribution of $\Delta F, \Delta P, \Delta G$ differs by system characteristics. All values are considerably different from 0, which would indicate a perfect system approximation. The only exception is S2, but since S2 contains only one trace, this is not surprising. In general, we see that the more behavior a system allows, the larger $\Delta F, \Delta P, \Delta G$ are. Very generalizing systems (S3 and S8) cause a high discrepancy in both fitness and generalization, as none of the logs can be considered a representative sample of system behavior. S4 does not generalize at all, causing a high overestimation by the discovered models. Even for small systems like S1 or S5, the values are fairly high, indicating a considerable deviation between the discovered model and the underlying system.

5 Analyzing the Impact of Unobserved Behavior

The experiment in Sect. 4 demonstrates that unobserved behavior can influence the quality of discovered models. Thus, the problem of induction, as described in Sect. 2 also applies to process discovery as an empirical research discipline. This does not mean that process discovery as a whole is meaningless. Its practical applicability and usefulness, as demonstrated by numerous projects and contributions, is beyond all question. However, with this paper, we want to establish the awareness among researchers and practitioners that not only does this problem exist, but it also has an impact on their everyday work.

As the previous sections have shown, a model's quality mainly depends on the (unknown) system character. Hence, assumptions on the unobserved are nec-

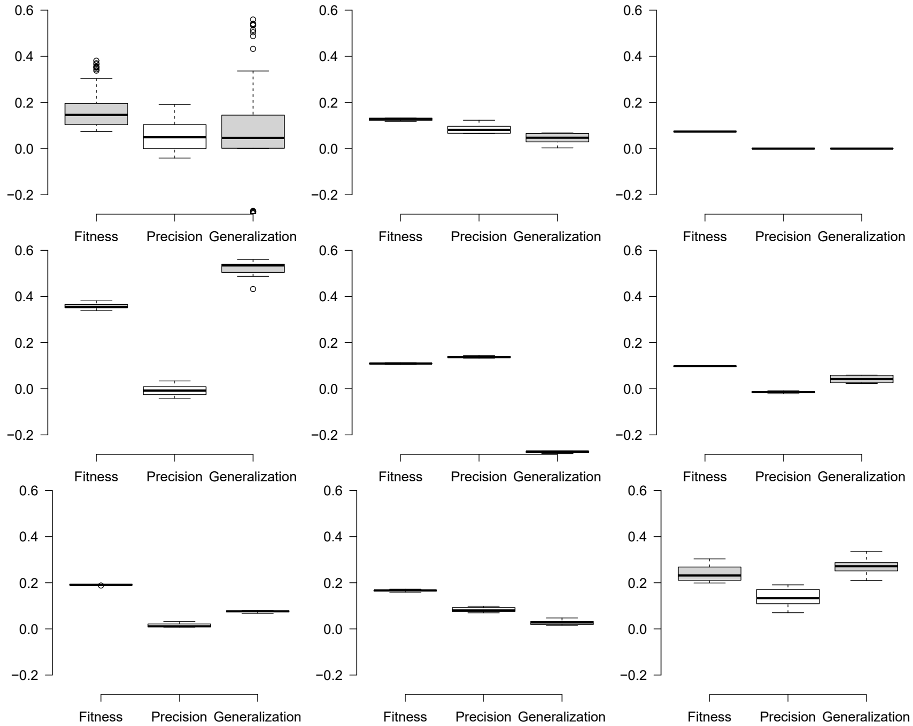

Fig. 4. Distribution of $\Delta F, \Delta P, \Delta G$ for all systems (upper left) and system 1 (upper center) to system 8 (lower right)

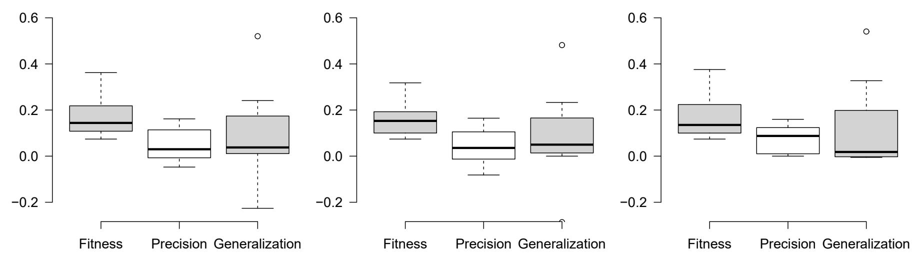

Fig. 5. Distribution of $\Delta F, \Delta P, \Delta G$ for Binomial (left), Poisson (center), Uniform (right) frequency distribution

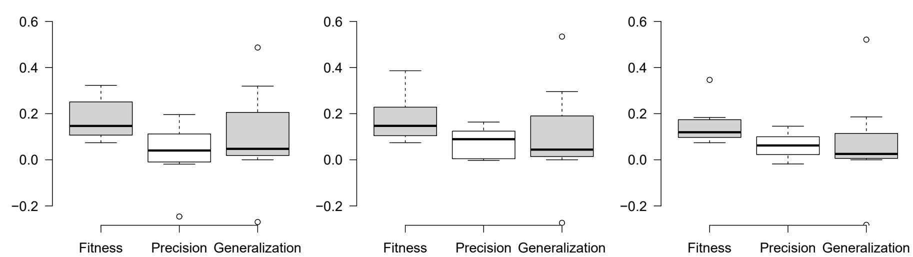

Fig. 6. Distribution of $\Delta F, \Delta P, \Delta G$ for log size 100 (left), 200 (center), 500 (right)

essary for correctly interpreting a quality measure. In the following, we explicate two possible assumptions, relating the unobserved (past) behavior with the unobservable (future) behavior. The first dimension concerns model continuity. We can assume a system to be *continuous*, i.e. expect the unobserved behavior of the future to repeat the unobserved behavior of the past, or *non-continuous*, i.e. expect the future behavior to differ from the past. Second, we distinguish between *descriptive* models intended to represent the past and *prescriptive* models intended to draw conclusions about the future. Combining these two assumptions leads to four different scenarios, as outlined in Fig. 7.

Models of **Type I** are assumed to be continuous and descriptive. The future behavior is not expected to deviate from the past, while the model is primarily used to describe the observed behavior. These models are retrospective, focusing solely on the known past. Assessing their quality should be focused on high fitness and precision values. Generalization can be neglected, as the unobserved behavior is not considered relevant to the model quality. In such a case, an enumerating model such as S4 is seen as ideal.

Models of **Type II** are assumed to be non-continuous and descriptive. While the model is focused on representing the past, the future behavior is expected to contain new patterns. This leads to a contradiction; should the unobserved behavior be considered in the model or not? If it is considered, the model cannot claim to be strictly descriptive anymore, as it includes behavior from an assumed future. However, if it is not considered, the model deliberately excludes behavior that is assumed to be possible, purposely reducing the model quality.

Models of **Type III** are assumed to be continuous and prescriptive. The model will be used to reason about a future, which is not expected to considerably deviate from the past. Observed behavior equates to unobserved, making it easy to reason about the latter based on the former. With regard to quality assessment, high fitness and precision and medium generalization values should be pursued to achieve a balanced model with a limited degree of generalization.

Models of **Type IV** are assumed to be non-continuous and prescriptive. The future behavior, which the model reasons about, will contain unseen patterns. This implies drawing conclusions about unknown model behavior without any

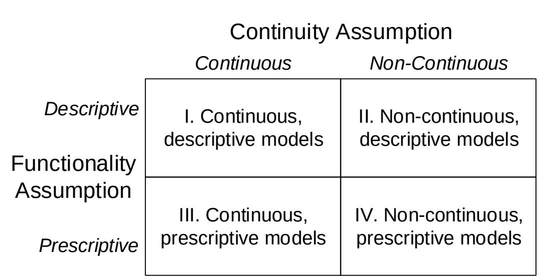

Fig. 7. Four different models types for quality assessment

additional knowledge. We must expect to see all different kinds of behavior, eliciting a need for maximum generalization. Fitness and precision are less relevant, as the unobserved behavior can be expected to be large.

All four types are generally possible, however, the continuity assumptions necessitates strong evidence from the system context in order to be valid. Continuity might be given in a highly automated process, but the more degrees of freedom a process allows, the less realistic this is. In any application of process discovery, explicating these assumptions will simplify correctly interpreting the resulting model.

6 Related Work

With many technical issues in process discovery now solved, new discovery techniques supposed to outperform existing ones rely on commonly agreed and scientifically sound instruments for comparing their performance. A first attempt towards the development of comparative frameworks has recently been suggested [7]. It includes a variety of previously suggested evaluation metrics, such as the ones used in this paper.

The need for an explicit differentiation between system, log, and model in process discovery is originally discussed in [5]. The author considers the system as either a concrete information system implementation, or, more likely, as the context of a process, i.e. its organization, rules, economy, etc. Such a definition is rather intangible and does not allow for measuring concrete process behavior. The author introduces and shortly discusses the notion of an "unknown system", but neglects the Log-System perspective. To analyze the system in a more formal way, our paper follows [3], defining a system as a set of traces.

The strict distinction between system, log, and model in process discovery has not yet been widely adopted. In [16], the author states that "both process discovery and conformance checking aim to tell something about the unknown real process rather than the example traces in the event log". As pointed out in [3], this suggests that "the one and only goal of process discovery would be to represent the true underlying process". Typically used quality metrics, however, do not take on the Log-System perspective, but evaluate the discovered model with regard to the event log. Little empirical work exists on the impact of this assumption. Rogge-Solte et al. illustrate the difficulties of the Model-Log perspective in their Generalized Conformance Checking Framework [17]. Janssenswillen et al. provide a first analysis of the Log-System perspective by measuring the effect that noise and incompleteness of event logs may have on discovered models [3]. However, their main objective is to analyze the dependability of certain fitness and precision metrics in terms of bias and ranking, whereas our contribution is focused on the influence the unobserved behavior has on the generic quality assessment.

7 Discussion and Conclusion

In this paper, we present the well-known epistemological problem of induction [4] and discuss its consequences for the field of Process Mining. By means of an empirical analysis, we demonstrate that the quality of discovered models can be significantly influenced by unobserved system behavior. These findings elicit a need for Process Miners to not only explicate the assumptions they make on the nature of their underlying system, but also to develop these assumptions on a more detailed and context-driven level. They are closely related to the choice of log as well as the intended objective of the task at hand. Altogether, these factors can have a significant impact on the success of any Process Mining project.

Our work is not intended as a full analysis or solution. Instead, we want to raise awareness on how existing epistemological discussions may relate to Process Mining and spark a discussion on how they influence the day-to-day work of researchers and professionals. Our experiment is intended to demonstrate potential influences under realistic circumstances. Therefore, we need to make several assumptions to its design. The systems we use for our experiment are small and artificial. They are able to demonstrate the potential effects of unobserved behavior, however, in how far the results are generalizable remains to be seen. Also, although we try to mitigate side effects of individual miners or quality measures by working with average values, they cannot be completely eliminated.

We also do not consider log completeness, but focus on size and frequency distribution instead. This is done for matters of comparability; the effect of unobserved behavior is influenced by the amount of unobserved and unobservable behavior, which depends on the system itself. Simplicity was not included in our experiment, since it does not directly relate a model to a log. Another shortcoming is that we do not consider the presence of noise in neither our experiment nor the framework in Sect. 5. Noise can play an important role, as it may influence both the discovered model and its quality assessment. We are aware of these influences, however, we decided to focus on the system perspective here.

This contribution can be the starting point for a number of future activities. The influence of log completeness and noise should be investigated, using larger and more realistic systems. The framework in Sect. 5 should be elaborated and enriched with additional dimensions. Its validity could also be investigated by means of a larger case study, where we want to demonstrate the differences between the model types. There is also a need to deepen the epistemological discussion, and also take log choices and goals of Process Mining into account.

One can also argue that the metaphor of the Black Swan (and thus the necessity to falsify a hypothesis) is of little significance to the reality of Process Mining, as it can also be interpreted as positivist research, eliciting a need to verify a theory based on empirical data. These differences, although subtle, are highly relevant when it comes to the interpretation of scientific results, including Process Mining as a discipline. As this is ultimately a philosophical and epistemological discussion, we are not favoring one interpretation over the other. However, we want to encourage researchers and professionals to think about the implications of this discussion on their everyday work.

References

1. van Dongen, B., Carmona, J., Chatain, T.: A unified approach for measuring precision and generalization based on anti-alignments. In La Rosa, M., Loos, P., Pastor, O., eds.: Business Process Management: BPM 2016, Springer (2016) 39–56
2. Buijs, J., van Dongen, B., van der Aalst, W.: Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity. International Journal of Cooperative Information Systems **23**(01) (2014) 1440001
3. Janssenswillen, G., Jouck, T., Creemers, M., Depaire, B.: Measuring the quality of models with respect to the underlying system: An empirical study. In La Rosa, M., Loos, P., Pastor, O., eds.: Business Process Management. BPM 2016, Springer (2016) 73–89
4. Popper, K.: The logic of scientific discovery. Routledge (2005)
5. Buijs, J.: Flexible evolutionary algorithms for mining structured process models. PhD thesis, Technische Universiteit Eindhoven (2014)
6. Weber, P., Bordbar, B., Tiňo, P., Majeed, B.: A framework for comparing process mining algorithms. In: GCC Conference and Exhibition (GCC), IEEE (2011) 625–628
7. vanden Broucke, S., De Weerdt, J., Vanthienen, J., Baesens, B.: A comprehensive benchmarking framework (cobefra) for conformance analysis between procedural process models and event logs in prom. In: Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2013), IEEE (2013) 254–261
8. vanden Broucke, S., De Weerdt, J., Vanthienen, J., Baesens, B.: An improved process event log artificial negative event generator. Feb research report, KU Leuven Faculty of Economics and Business, Leuven (Belgium) (2012)
9. Weijters, A., van Der Aalst, W., De Medeiros, A.: Process mining with the heuristics miner-algorithm. Technical Report 166, TU Eindhoven (2006)
10. Leemans, S., Fahland, D., van der Aalst, W.: Discovering block-structured process models from event logs containing infrequent behaviour. In Lohmann, N., Song, M., Wohed, P., eds.: BPM Workshops 2013, Springer (2014) 66–78
11. van der Werf, J., van Dongen, B., Hurkens, C., Serebrenik, A.: Process discovery using integer linear programming. In van Hee, K., Valk, R., eds.: International Conference on Applications and Theory of Petri Nets, Springer (2008) 368–387
12. van der Aalst, W., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **2**(2) (2012) 182–192
13. vanden Broucke, S., De Weerdt, J., Vanthienen, J., Baesens, B.: Determining process model precision and generalization with weighted artificial negative events. IEEE Transactions on Knowledge and Data Engineering **26**(8) (2014) 1877–1889
14. Rozinat, A., van der Aalst, W.: Conformance checking of processes based on monitoring real behavior. Information Systems **33**(1) (2008) 64–95
15. Adriansyah, A., Munoz-Gama, J., Carmona, J., van Dongen, B., van der Aalst, W.: Alignment based precision checking. In La Rosa, M., Soffer, P., eds.: Business Process Management Workshops. BPM 2012, Springer (2013) 137–149
16. van der Aalst, W.: Mediating between modeled and observed behavior: the quest for the right process. In: IEEE International Conference on Research Challenges in Information Science (RCIS 2013). (2013) 31–43
17. Rogge-Solti, A., Senderovich, A., Weidlich, M., Mendling, J., Gal, A.: In log and model we trust? a generalized conformance checking framework. In La Rosa, M., Loos, P., Pastor, O., eds.: Business Process Management. BPM 2016, Springer (2016) 73–89