Conformance Checking for Data-Aware Processes

An Axiomatic Approach

Queensland University of Technology
RWTH Aachen University
TU/e

Conformance Checking for Data-Aware Processes

An Axiomatic Approach

QUT Queensland University of Technology

Adam Banham (PhD Candidate)

QUT Queensland University of Technology

Prof. Arthur ter Hofstede

RWTH AACHEN UNIVERSITY

Prof. Sander J. J. Leemans*

TU/e

Prof Felix Mannhardt

QUT Queensland University of Technology

Dr Robert Andrews*

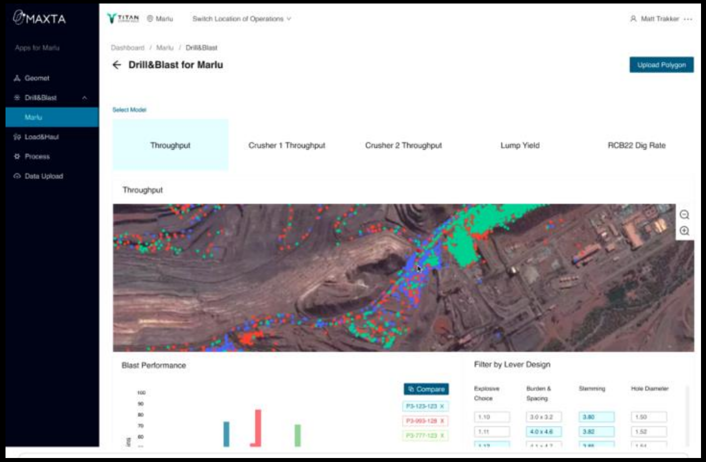
QUT Queensland University of Technology

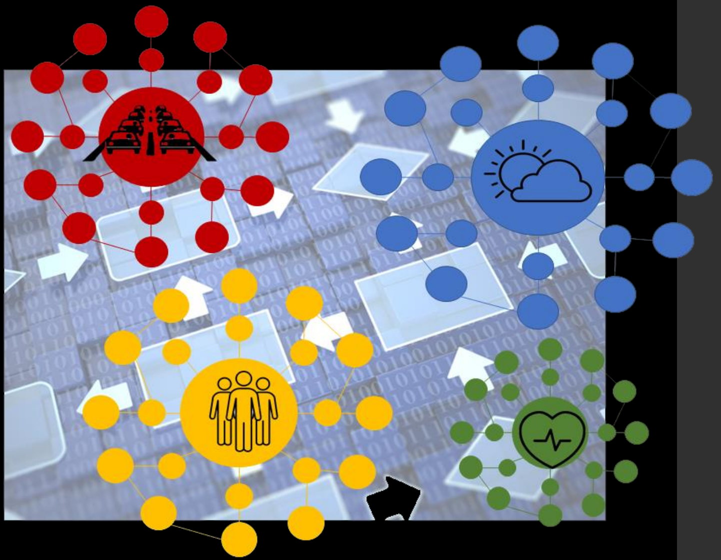
Prof. Moe T. Wynn**

A little about me…

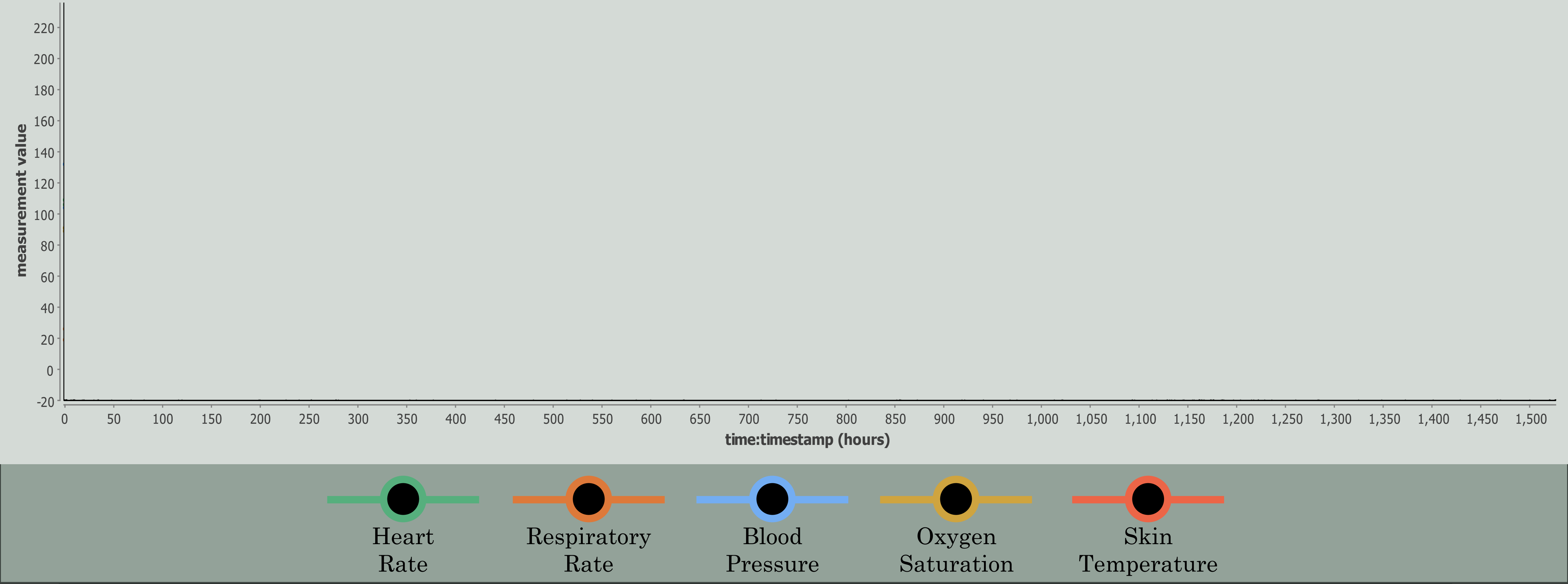
Adam Banham (PhD Candidate)

- Studied at QUT through a bachelor of information technology and transitioned into PhD
- Worked as a research assistant with Prof. Moe Wynn and Dr Robert Andrews
- My industry experience as a data scientist/ full stack developer at a tech startup around resource mining (mine to port processes)

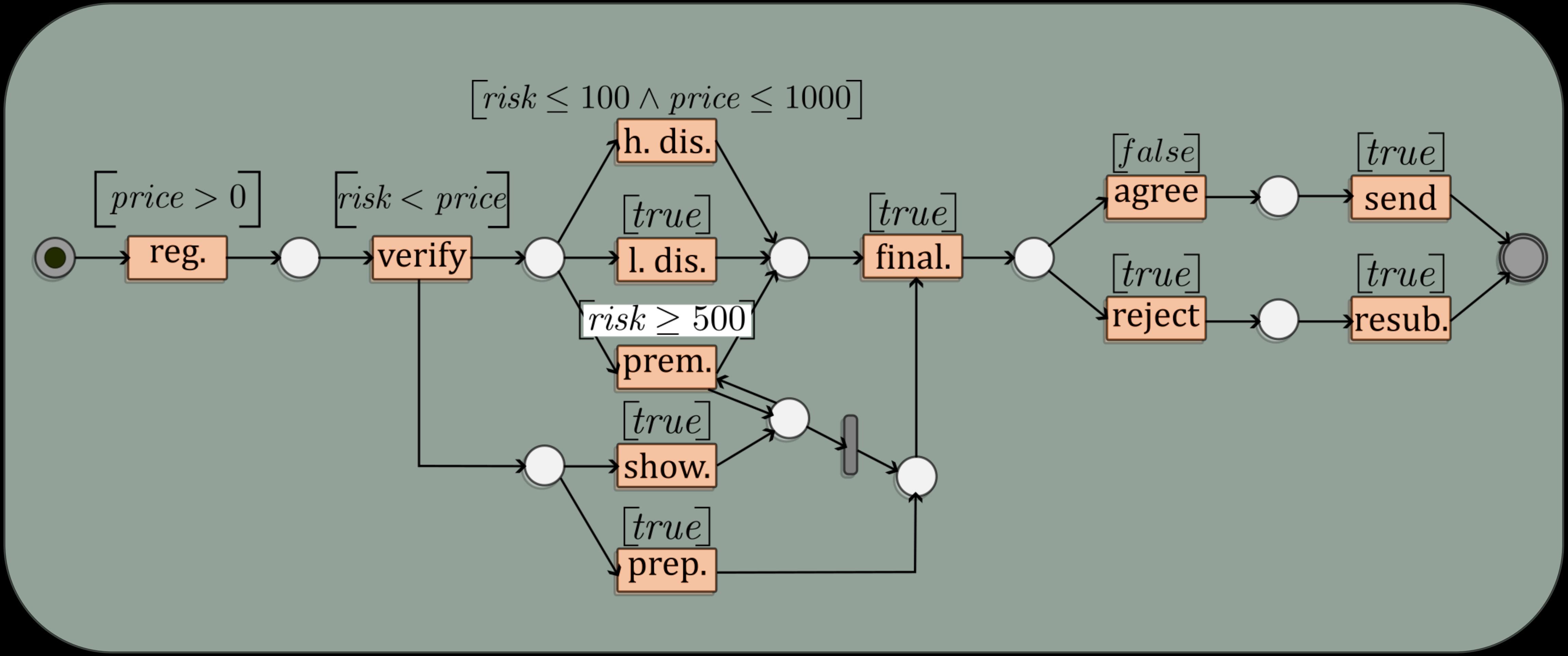




Process Mining with Exogenous Data

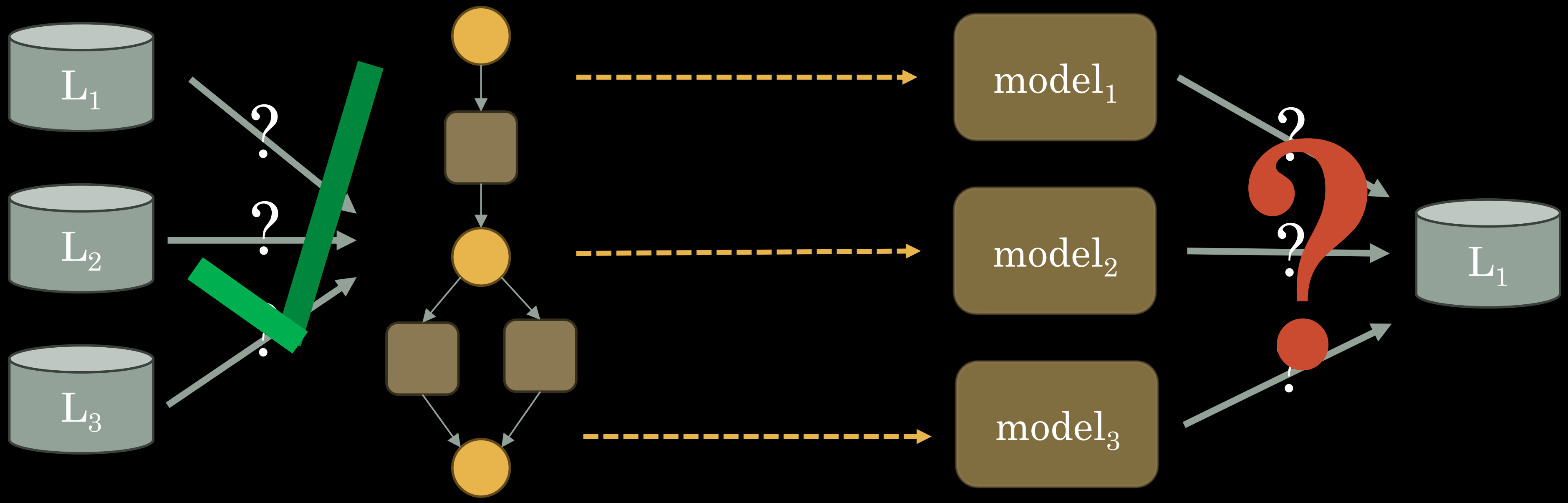


Data-aware processes



Why is conformance checking important?

Often a question that process mining considers, is if an outcome is of high quality and suitable for further analysis.



Conformance checking is critical for the development of evaluations in academia work and in our papers!

Our evaluations should be…

	Tech A	**Tech B**	**Tech C**
Log A			
Log B			
Log C			

Our evaluations should be…

Reproduceable!

	Tech A	**Tech B**	**Tech C**
Log A			
Log B			
Log C			

Our evaluations should be…

Reproduceable!

Comparable!

	Tech A	Tech B	Tech C
Log A			
Log B			
Log C			

Our evaluations should be…

Reproduceable!

Comparable!

Meaningful!

	Tech A	**Tech B**	**Tech C**
Log A			
Log B			
Log C			

Our evaluations should be…

Reproduceable!

Comparable!

Meaningful!

	Tech A	Tech B	Tech C
Log A			
Log B			
Log C			

We want to ensure that the evaluations of decision-mining techniques are *reproduceable, comparable and meaningful*!

What is an axiomatic approach?

Conformance checking techniques should consider desirable properties through a collection of axioms/propositions to ensure that we are not proposing ad-hoc measures or making design choices in an ad-hoc manner.

Niek Tax et al. “The imprecisions of precision measures in process mining”, 2018.

Wil M. P. van der Aalst. “Relating Process Models and Event Logs 21 Conformance Propositions”, 2018

Joos C. A. M. Buijs, Boudewijn F. van Dongen, and Wil M. P. van der Aalst. “Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity”. In: Int. J. Cooperative Inf. Syst. 23.1 (2014)

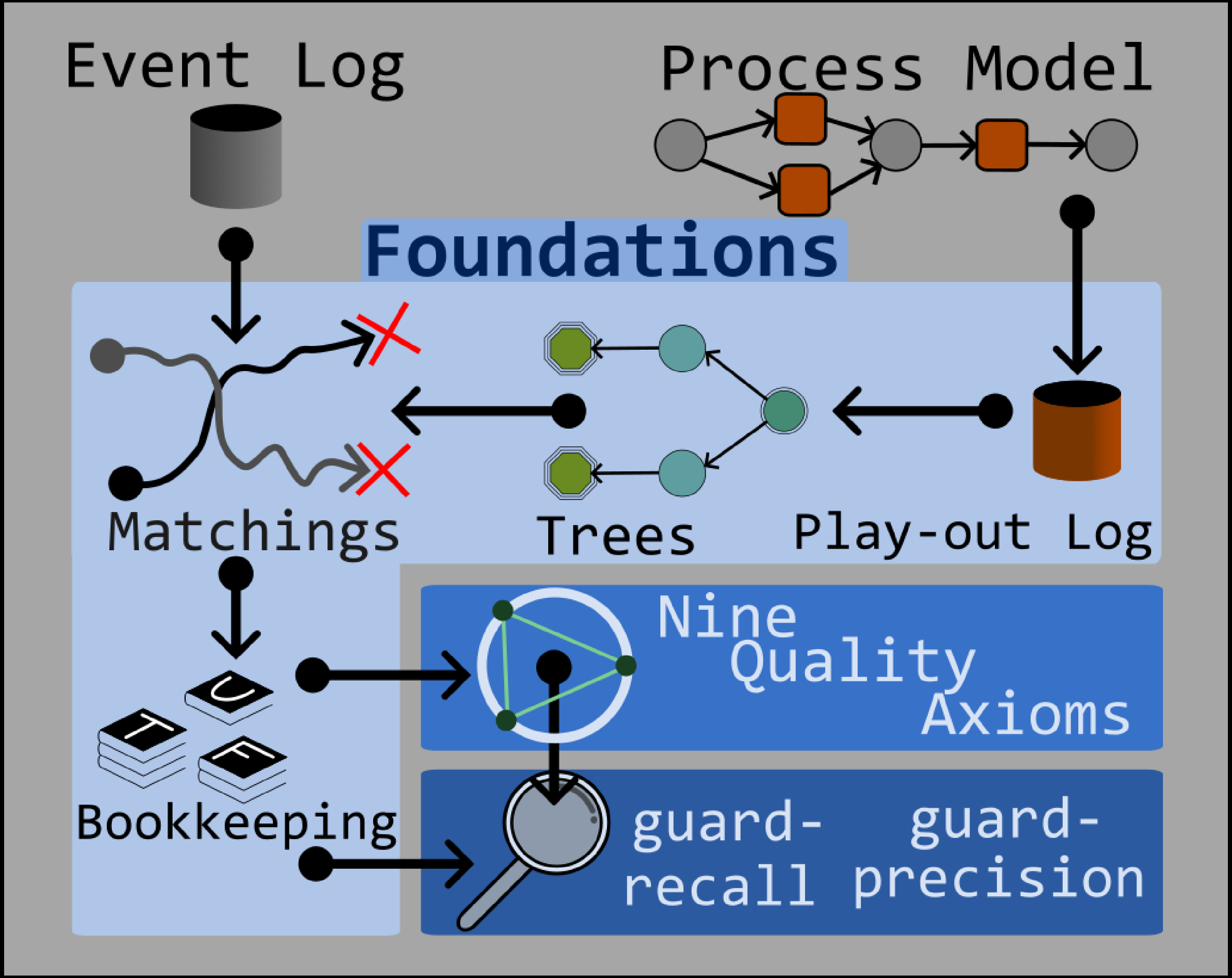
Anja F. Syring, Niek Tax, and Wil M. P. van der Aalst. “Evaluating Conformance Measures in Process Mining Using Conformance Propositions”. (2019)

What is an axiomatic approach?

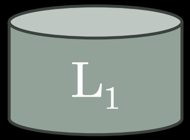
Anja F. Syring, Niek Tax, and Wil M. P. van der Aalst. “Evaluating Conformance Measures in Process Mining Using Conformance Propositions”. (2019)

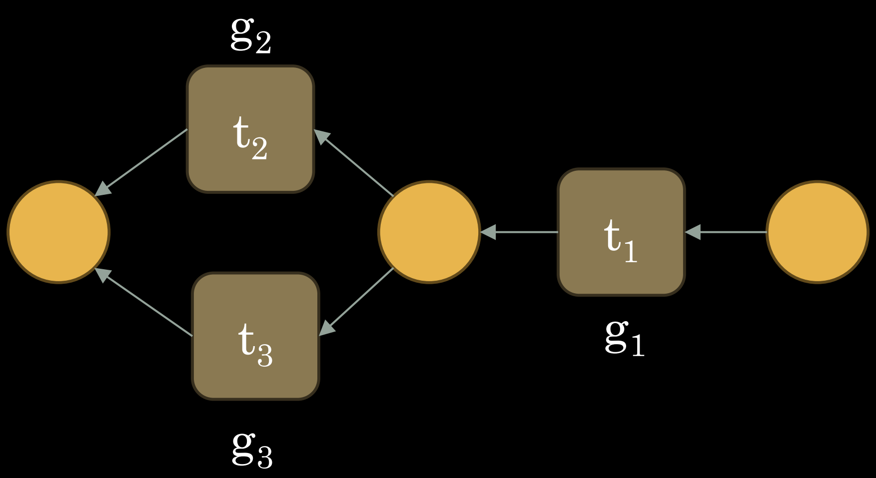
- Considered the truths or desirable properties for conformance checking in the control-flow setting,
- Presents a model agnostic quantification theory to reason about these properties,
- Recall, Precision and Generalisation are considered in terms of properties,
- Performs a large evaluation of conformance techniques using both counter examples and proofs to affirm properties being held by techniques.

This is our axiomatic approach

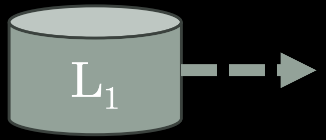


Bookkeeping

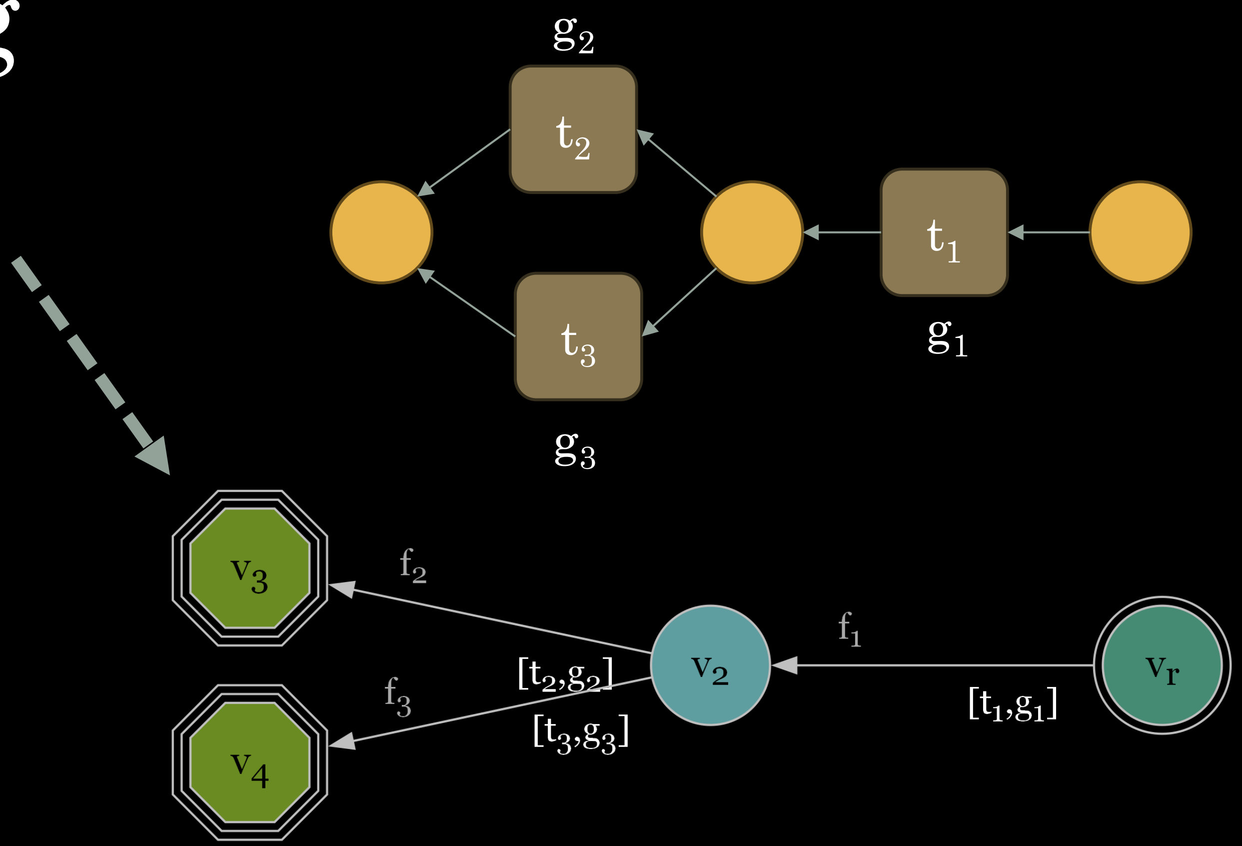




Bookkeeping

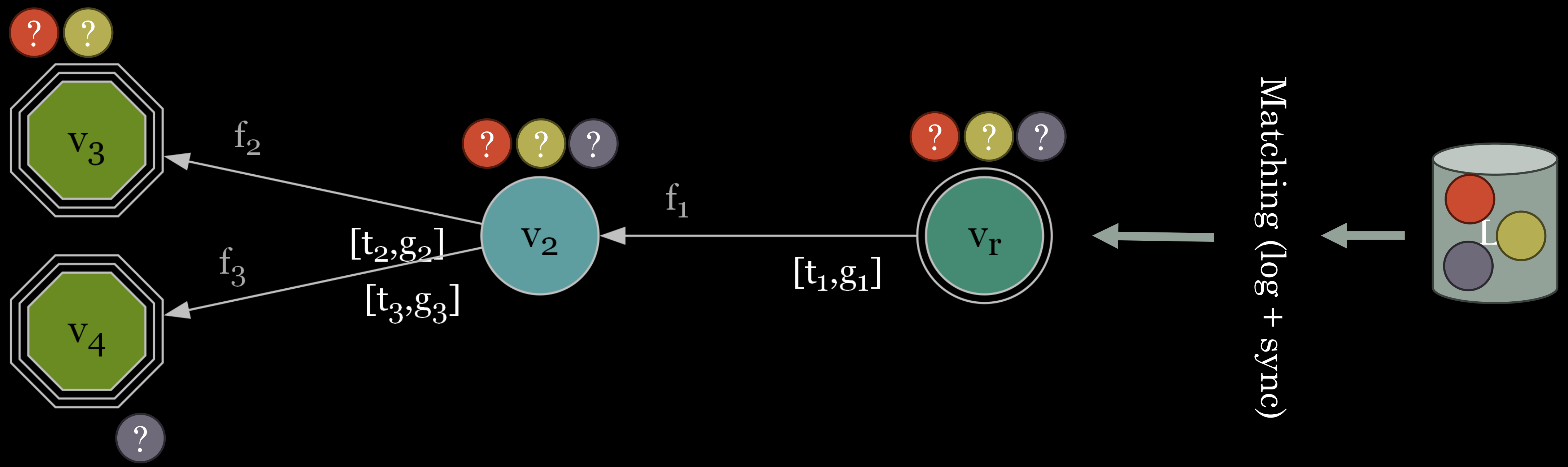


$\max |\sigma| \in L_1$



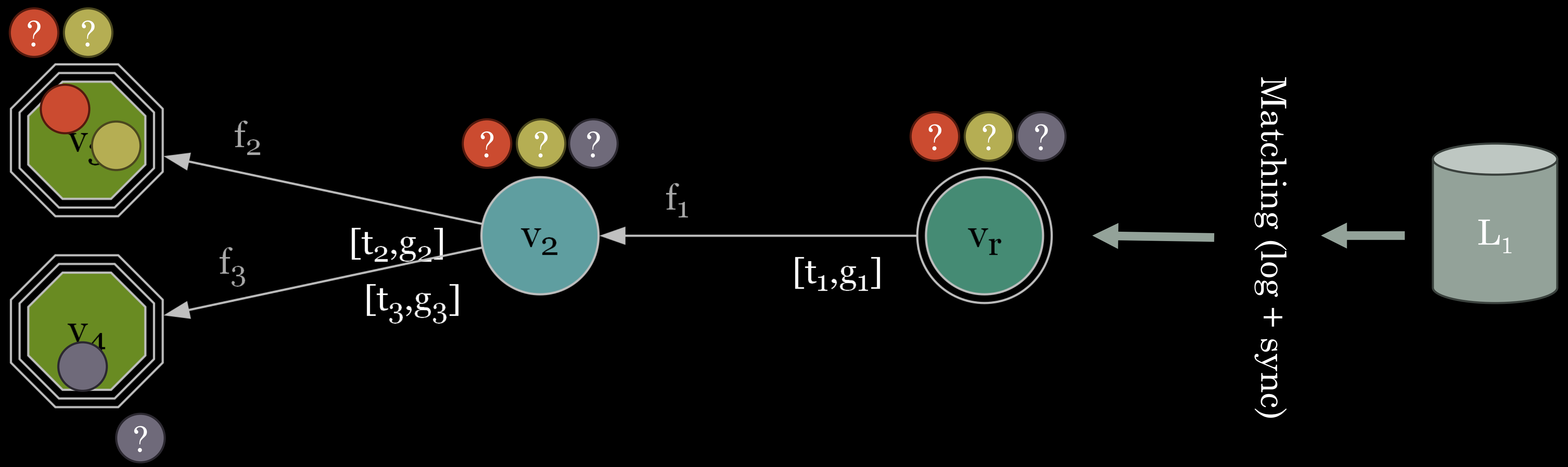
Transition trees as in “When are two Workflows the Same?” (Jan Hidders et.al., 2005)

Bookkeeping



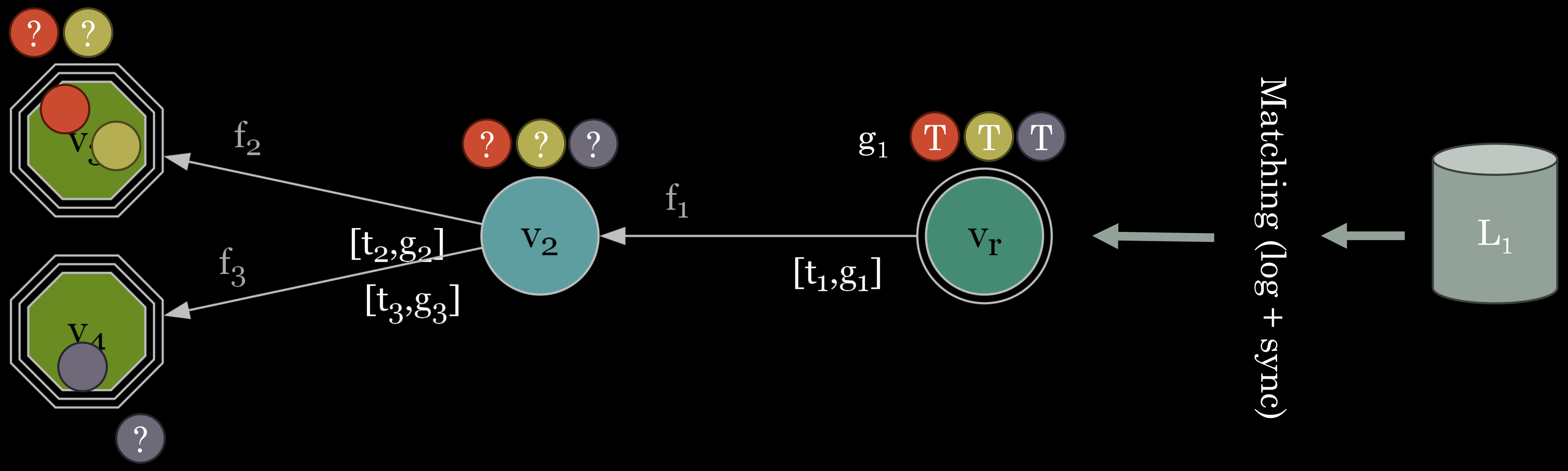
Traversal Bookkeeping (guard-recall)

Were guards evaluated to true (T) as they traversed the tree?



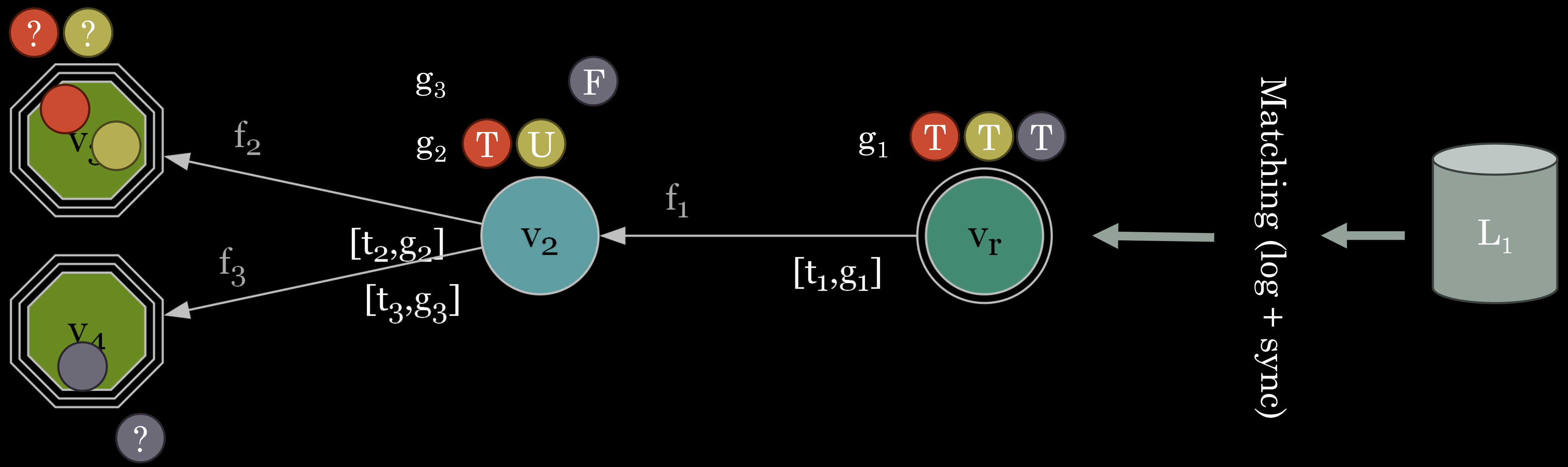
Traversal Bookkeeping (guard-recall)

Were guards evaluated to true (T) as they traversed the tree?



Traversal Bookkeeping (guard-recall)

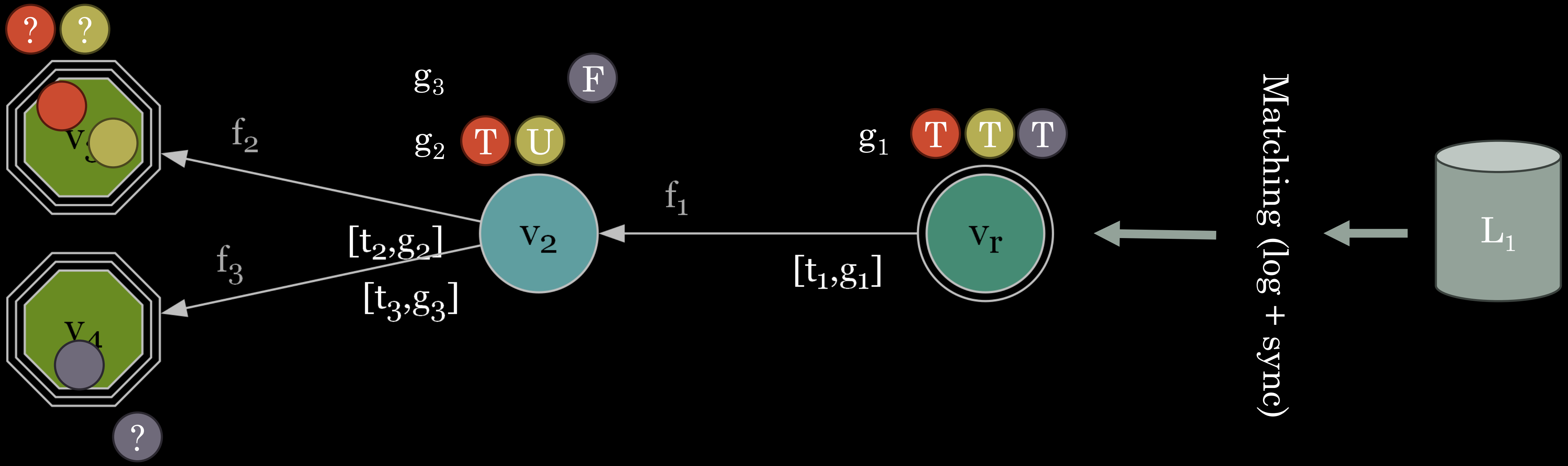
Were guards evaluated to true (T) as they traversed the tree?



$Grd_T^{\rightarrow}(f_3, \gamma, L_1) = 0$ $Grd_T^{\rightarrow}(f_2, \gamma, L_1) = 1$ $Grd_T^{\rightarrow}(f_1, \gamma, L_1) = 3$

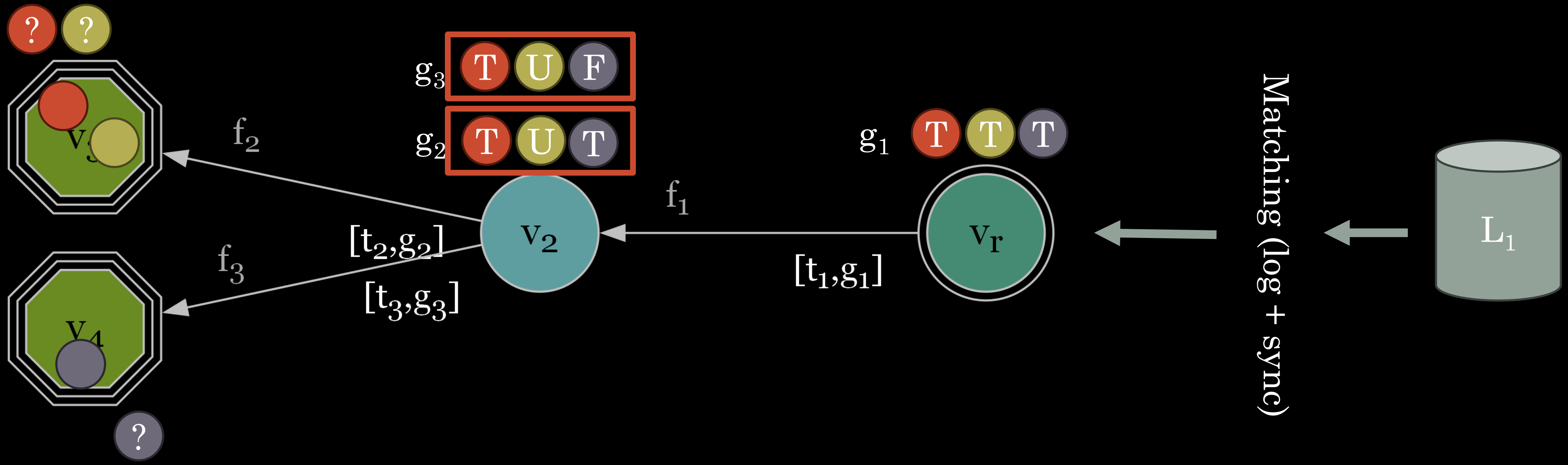
Explorative Bookkeeping (guard-precision)

As we traverse the tree, how often could have guards been evaluated to true (T)?



Explorative Bookkeeping (guard-precision)

As we traverse the tree, how often could have guards been evaluated to true (T)?



$Grd_T^{\rightarrow}(f_3, \gamma, L_1) = 0$ $\quad$ $Grd_T^{\rightarrow}(f_2, \gamma, L_1) = 1$ $\quad$ $Grd_T^{\rightarrow}(f_1, \gamma, L_1) = 3$

$Grd_T(f_3, \gamma, L_1) = 1$ $\quad$ $Grd_T(f_2, \gamma, L_1) = 2$ $\quad$ $Grd_T(f_1, \gamma, L_1) = 3$

Axiom 1 & 2

(We can always construct)

Proposition 2 (BehPro+) in [Syring19]

The form of the guards should not influence the outcome of guard-recall or guard-precision (representation does not matter!). If the behaviour of the guards between two models is the same, then their conformance should be the same.

A.1

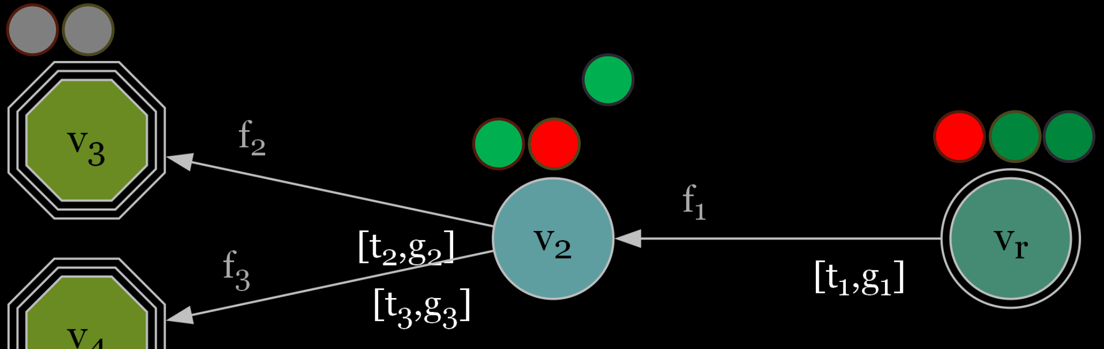
$$\forall_{f \in F}\, GrdEq(f, iso(f'), \gamma, \gamma') \implies grec(L, \gamma, M) = grec(L, \gamma', M')$$

A.2

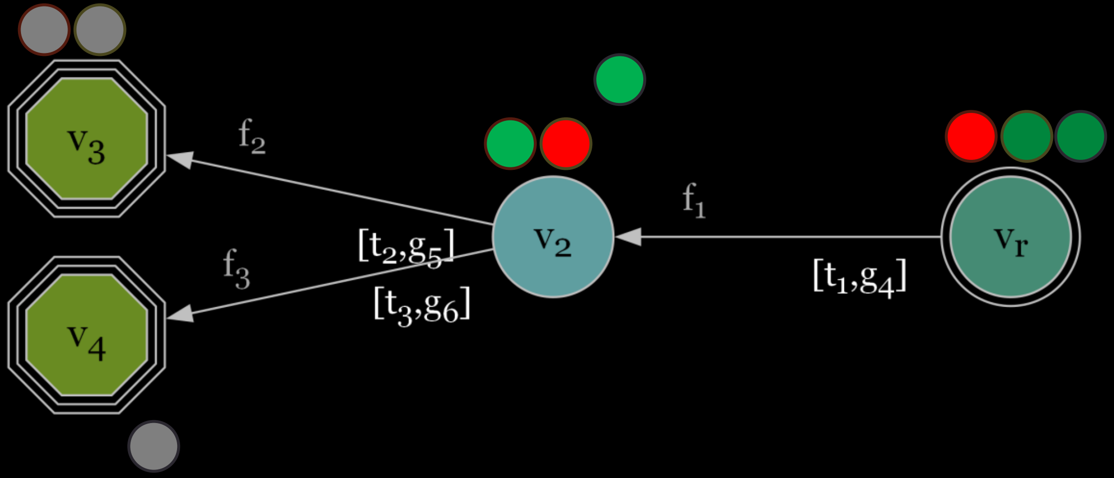
$$\forall_{f \in F}\, GrdEq(f, iso(f'), \gamma, \gamma') \implies gprec(L, \gamma, M) = gprec(L, \gamma', M')$$

Axiom 1 & 2

(We can always construct)



These are the same regardless of representation…



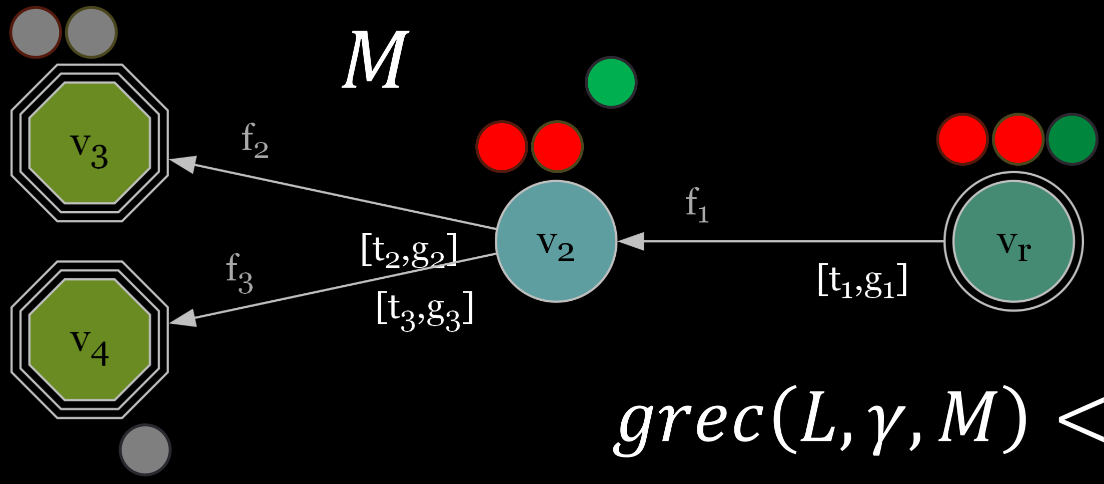
Axiom 3
(comparing)

Guard-recall quantifies if guards have been formed over the observed data and if they evaluate to true when traversing over the tree. Hence, when comparing isomorphic models with the same event log, guard-recall takes on a larger value for the model with more guard evaluations that were true

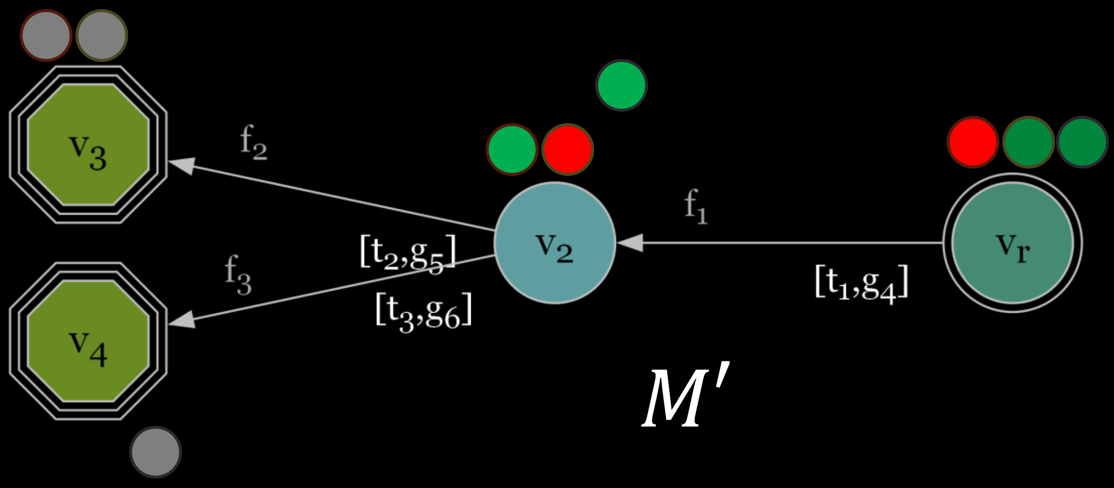
$$\sum_{f \in F} \left(Grd_T^{\rightarrow}(f, \gamma, L) - Grd_T^{\rightarrow}(iso(f), \gamma', L) \right) < 0 \Rightarrow grec(L, \gamma, M) < grec(L, \gamma', M')$$

Axiom 3

(comparing)



$$grec(L, \gamma, M) < grec(L, \gamma', M')$$



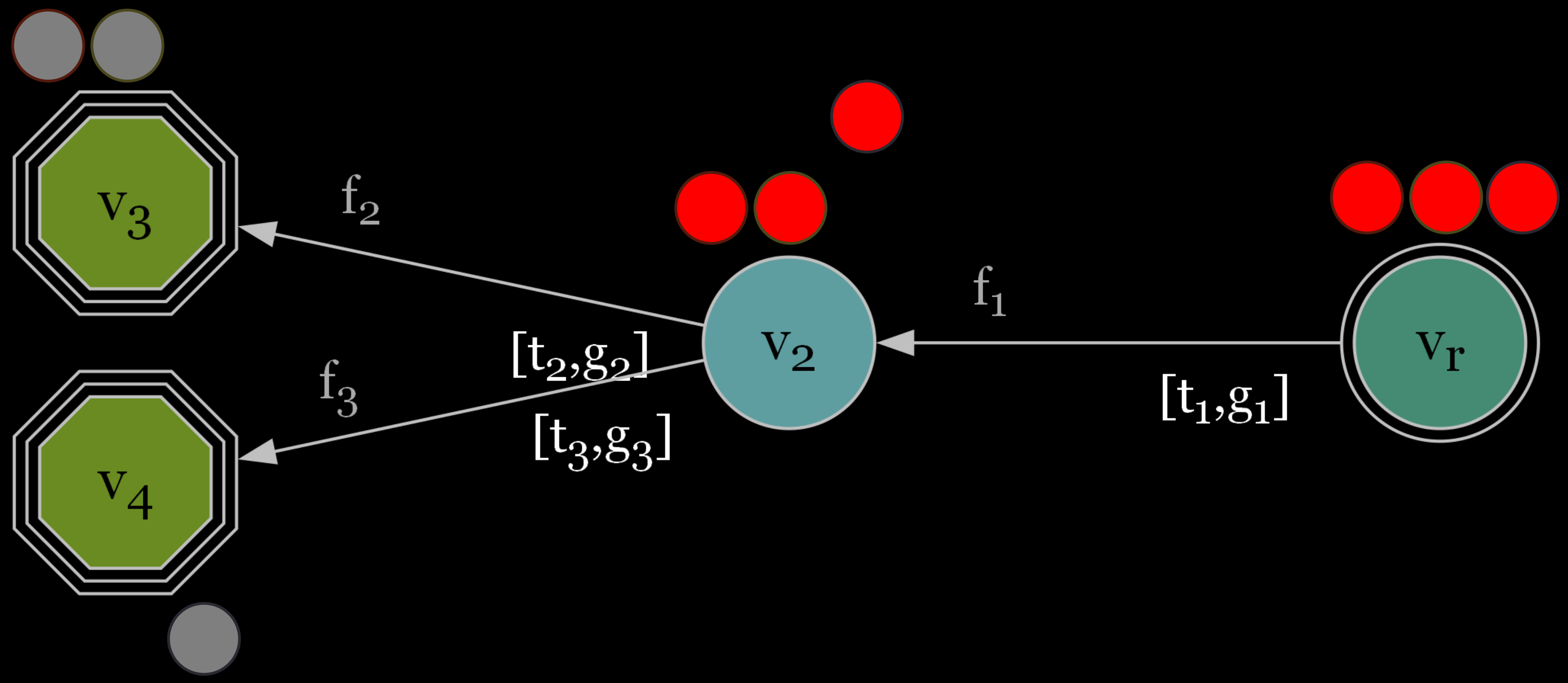
Axiom 4

(meaning)

Furthermore, if no guard evaluates to true then guard-recall takes on the minimal value (<u>min</u>).

$$\forall_{f \in F}\ [Grd_T^{\rightrightarrows}(f, \gamma, L) = 0] \iff grec(L, \gamma, M) = \underline{\mathrm{min}}$$

Axiom 4
(meaning)



$$\forall_{f\in F}\ [Grd_T^{\rightrightarrows}(f,\gamma,L)=0] \Leftrightarrow grec(L,\gamma,M) = \underline{\min}$$

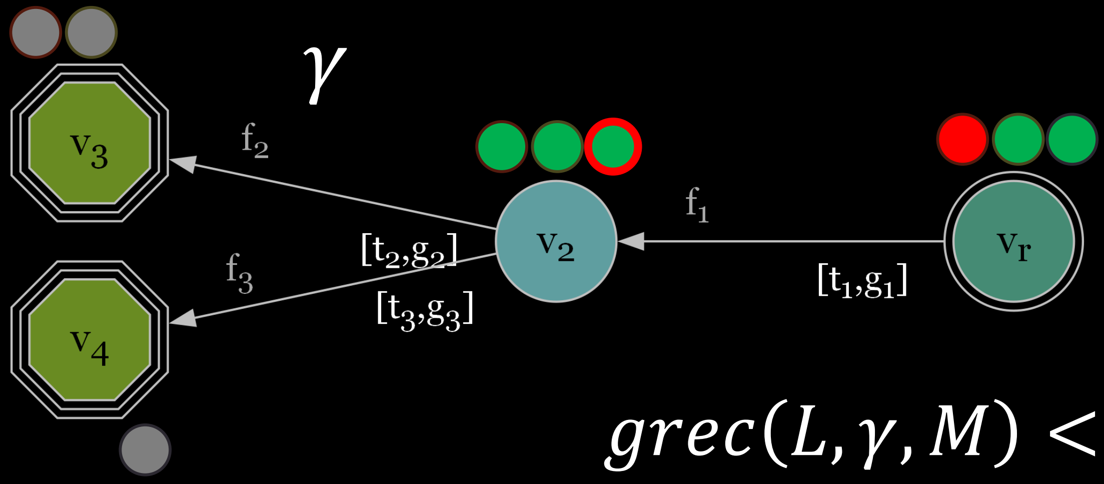
Axiom 5
(meaning)

Next, it could be the case, that the given matching is not optimal meaning that some events in the log may not have been considered in the bookkeeping functions. Hence, using a non-optimal matching, opposed to an optimal matching, even when evaluations are similar, will negatively affect guard-recall.

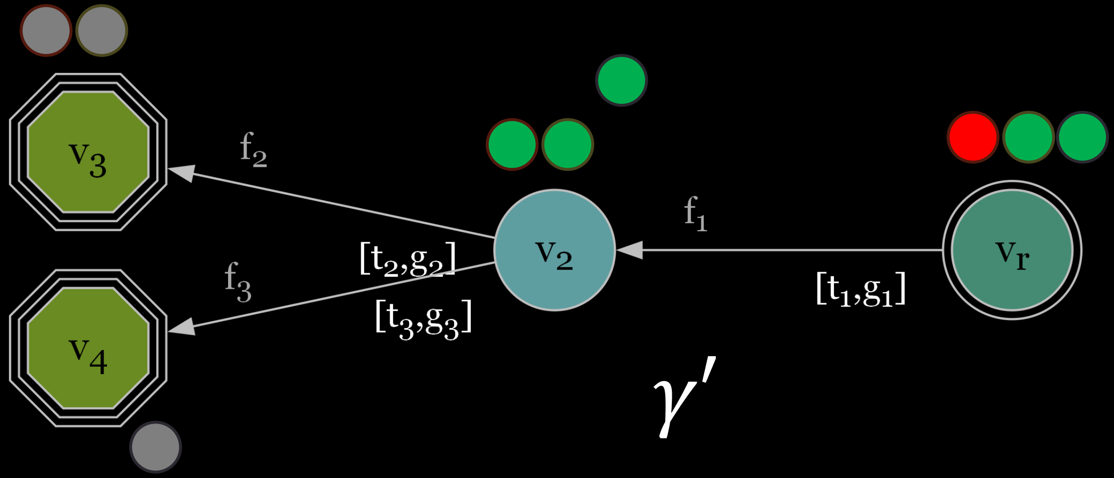
$$\sum_{f \in F} Grd_T^{\vec{}}(f, \gamma, M) \leq \sum_{f \in F} Grd_T^{\vec{}}(f, \gamma', M) \wedge \neg optimal(L, \gamma, M) \wedge optimal(L, \gamma', M)$$
$$\Longrightarrow grec(L, \gamma, M) < grec(L, \gamma', M)$$

Axiom 5

(meaning)



$$grec(L,\gamma,M) < grec(L,\gamma',M)$$



Axiom 6 & 9 (reproduceable)

Proposition 1 (DetPro+) & Proposition 6 (RecPro4) & Proposition 11 (PrecPro4)

We argue that enlarging an event log should not influence guard-recall or guard-precision.

given some $k \geq 1$

A.6

$$grec(L, \gamma, M) = grec(L^k, \gamma \quad, M \quad)$$

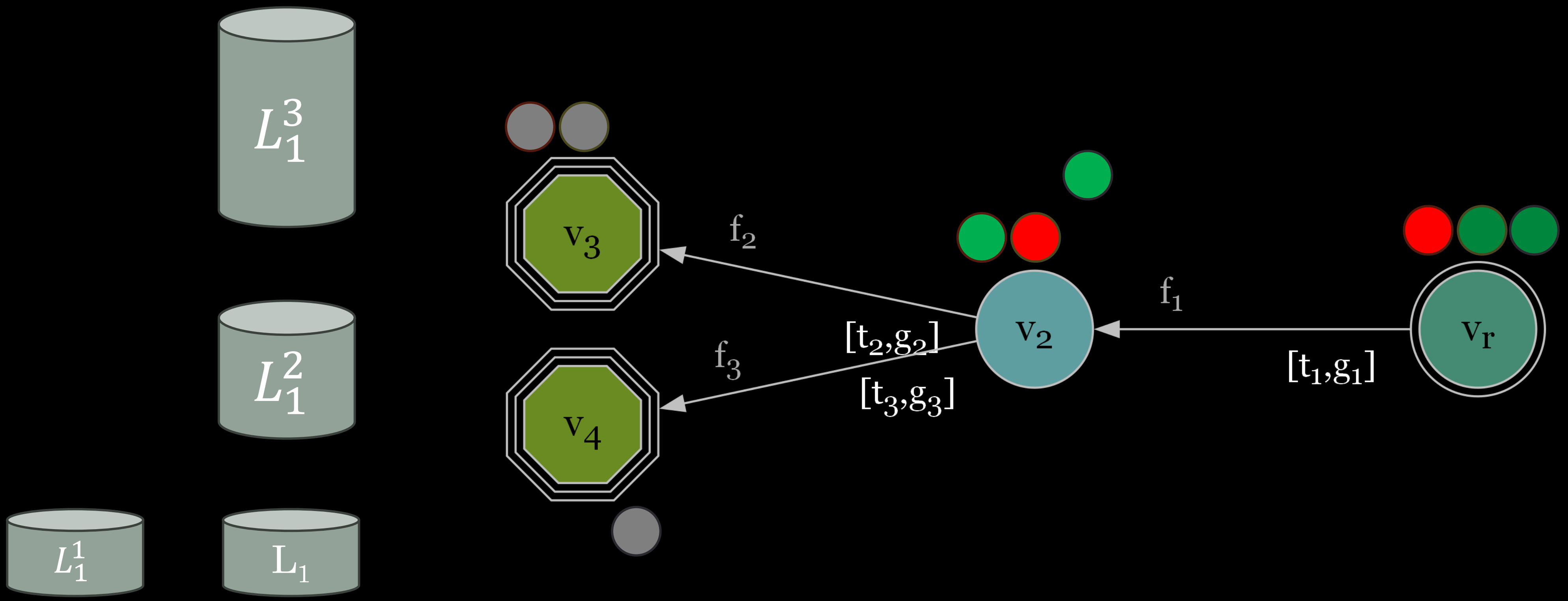
A.9

$$gprec(L, \gamma, M) = gprec(L^k, \gamma \quad, M \quad)$$

Axiom 6 & 9 (reproduceable)

Proposition 1 (DetPro+) & Proposition 6 (RecPro4) & Proposition 11 (PrecPro4)

We argue that enlarging an event log should not influence guard-recall or guard-precision.



Axiom 7

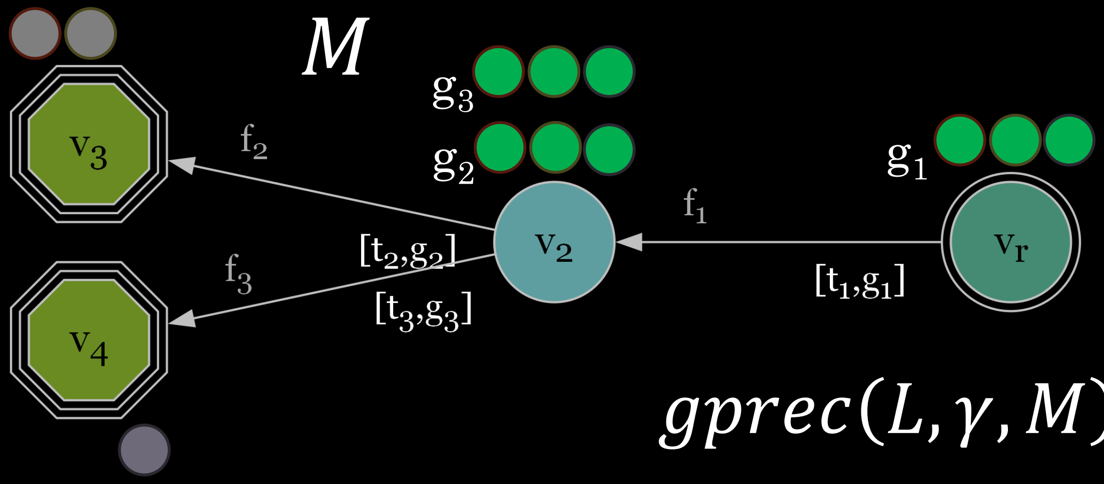
(comparing)

Ideally, flows with guards only evaluate to true when they are observed to be taken. Hence, guard-precision quantifies if guards only evaluated to true when the associated flow is observed and taken. When comparing similar models with the same event log, guard-precision takes on a larger value for the model with fewer guard evaluations that were true.

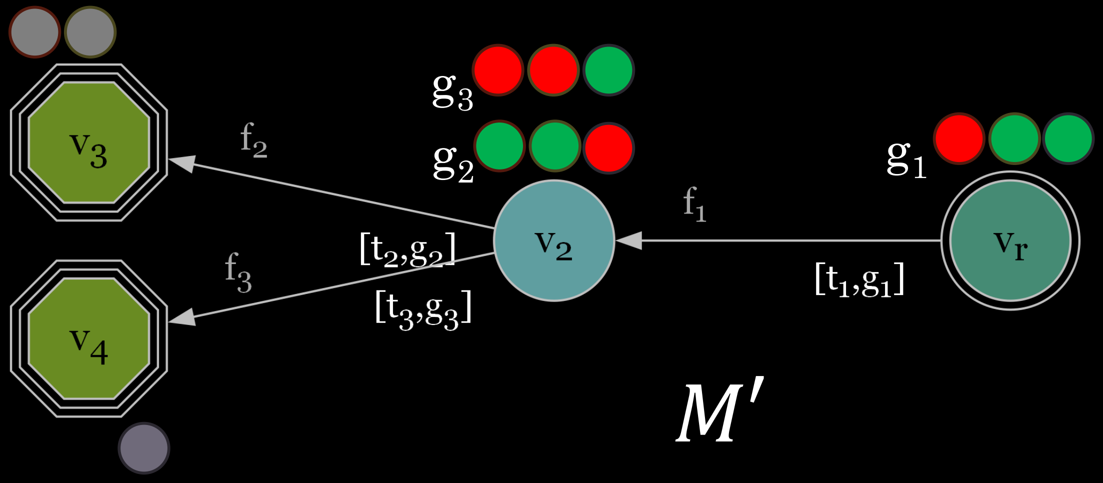
$$\sum_{f \in F} \left(Grd_T\,(f, \gamma, L) - Grd_T\,(iso(f), \gamma', L) \right) > 0 \Rightarrow gprec(L, \gamma, M) < gprec(L, \gamma', M')$$

Axiom 7

(comparing)



$$gprec(L, \gamma, M) < gprec(L, \gamma', M')$$



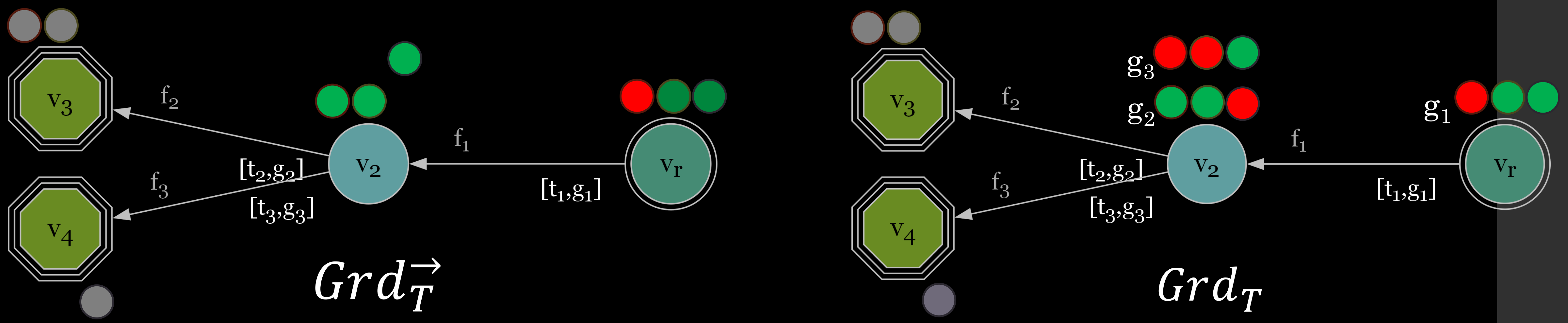
Axiom 8 (meaning)

Finally, guard-precision takes on the maximal value (<u>max</u>) if and only if all guards only evaluated to true when they are observed and taken, and no control-flow issues exist in the matching function

$$\left(optimal(L,\gamma,M) \wedge \forall_{f\in F}\left[Grd_T^{\rightrightarrows}(f,\gamma,L) = Grd_T(f,\gamma,L)\right]\right) \Leftrightarrow gprec(L,\gamma,M) = \underline{\mathrm{max}}$$

Axiom 8

(meaning)



$$\left(optimal(L,\gamma,M) \wedge \forall_{f\in F}\left[Grd_T^{\rightarrow}(f,\gamma,L) = Grd_T\,(f,\gamma,L)\right]\right) \Leftrightarrow gprec(L,\gamma,M) = \underline{\mathrm{max}}$$

Our evaluations can be…

Reproduceable!

Comparable!

Meaningful!

(guard-recall)

	Tech A	Tech B	Tech C
Log A	0.777	0.779	0.179
Log B	0.15	0.179	0.779
Log C	0.500	0	1.000

Ensuring that the evaluations of decision-mining techniques are *reproduceable, comparable and meaningful*!

Conformance Checking for Data-Aware Processes

An Axiomatic Approach

Queensland University of Technology

Adam Banham (PhD Candidate)

Queensland University of Technology

Prof. Arthur ter Hofstede

RWTH AACHEN UNIVERSITY

Prof. Sander J. J. Leemans*

TU/e

Prof Felix Mannhardt

Queensland University of Technology

Dr Robert Andrews*

Queensland University of Technology

Prof. Moe T. Wynn**

Where can we go from here…

- For the future of decision mining…
  - Future work could propose techniques that guarantee guard-recall or guard-precision
  - It would be interesting to see further work that reconsiders generalisation and simplicity for guards
- For conformance checking in this direction…
  - We need efficient ways to compute bookkeeping equivalents
  - How can SMT be used to efficiently work out guard equalities?
  - How can alignments be used to compute a distance between log and model with guard-recall or guard-precision in mind?
- For myself…
  - While this work does allow for the direct comparison of “hard” annotations, like guards, maybe “softer” annotations, like stochastics, could offer alternative options for investigations…

- [Buijs14] Joos C. A. M. Buijs, Boudewijn F. van Dongen, and Wil M. P. van der Aalst. “Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity”. In: Int. J. Cooperative Inf. Syst. 23.1 (2014). DOI: 10.1142 /S0218843014400012.
- [Tax18] Niek Tax et al. “The imprecisions of precision measures in process mining”. In: Inf. Process. Lett. 135 (2018), pp. 1–8. DOI: 10.1016/J.IPL.2018.01.013.
- [Aalst18] Wil M. P. van der Aalst. “Relating Process Models and Event Logs 21 Conformance Propositions”. In: Proceedings of the International Workshop on Algorithms & Theories for the Analysis of Event Data 2018 Satellite event of the conferences: 39th International Conference on Application and Theory of Petri Nets and Concurrency Petri Nets 2018 and 18th International Conference on Application of Concurrency to System Design ACSD 2018. Vol. 2115. CEUR Workshop Proceedings. CEUR-WS.org, 2018, pp. 56–74. URL: https://ceur- ws.org/Vol-2115/ATAED2018-56-74.pdf.
- [Syring19] Anja F. Syring, Niek Tax, and Wil M. P. van der Aalst. “Evaluating Conformance Measures in Process Mining Using Conformance Propositions”. In: Trans. Petri Nets Other Model. Concurr. 14 (2019), pp. 192–221. DOI: 10.1007/978-3-662-60651-3_8.
- [Augusto22] Adriano Augusto et al. “Measuring Fitness and Precision of Automatically Discovered Process Models: A Principled and Scalable Approach”. In: IEEE Trans. Knowl. Data Eng. 34.4 (2022), pp. 1870–1888. DOI: 10.1109/TKDE.2020.3003258.
- [Janssenwillen2017] Gert Janssenswillen et al. “A comparative study of existing quality measures for process discovery”. In: Inf. Syst. 71 (2017), pp. 1–15. DOI: 10.1016/J.IS.2017.06.002.
- [Hidders2005] Jan Hidders et al. “When are two Workflows the Same?” In: Theory of Computing 2005, Eleventh CATS 2005, Computing: The Australasian Theory Symposium, Newcastle, NSW, Australia, January/February 2005. Ed. by Mike D. Atkinson and Frank K. H. A. Dehne. Vol. 41. CRPIT. Australian Computer Society, 2005, pp. 3–11. URL: http : / / crpit . scem . westernsydney . edu . au / abstracts /CRPITV41Hidders.html.