QUT Centre for Data Science

Process Voids

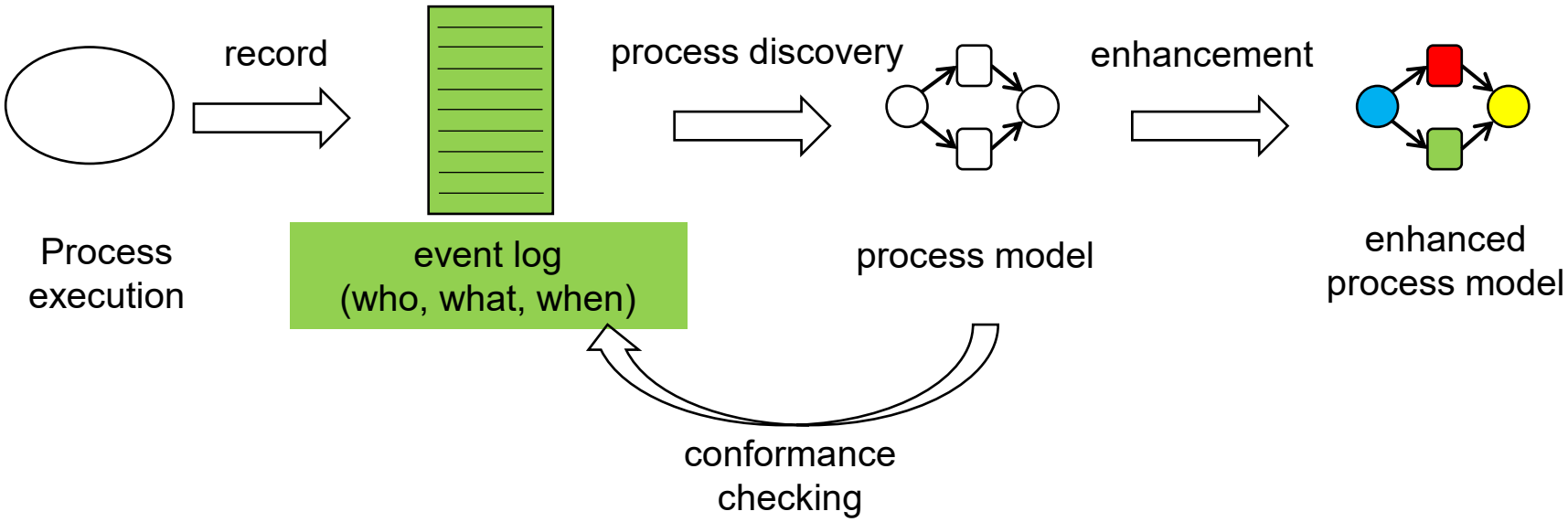
Data Science Without Data

November 2025

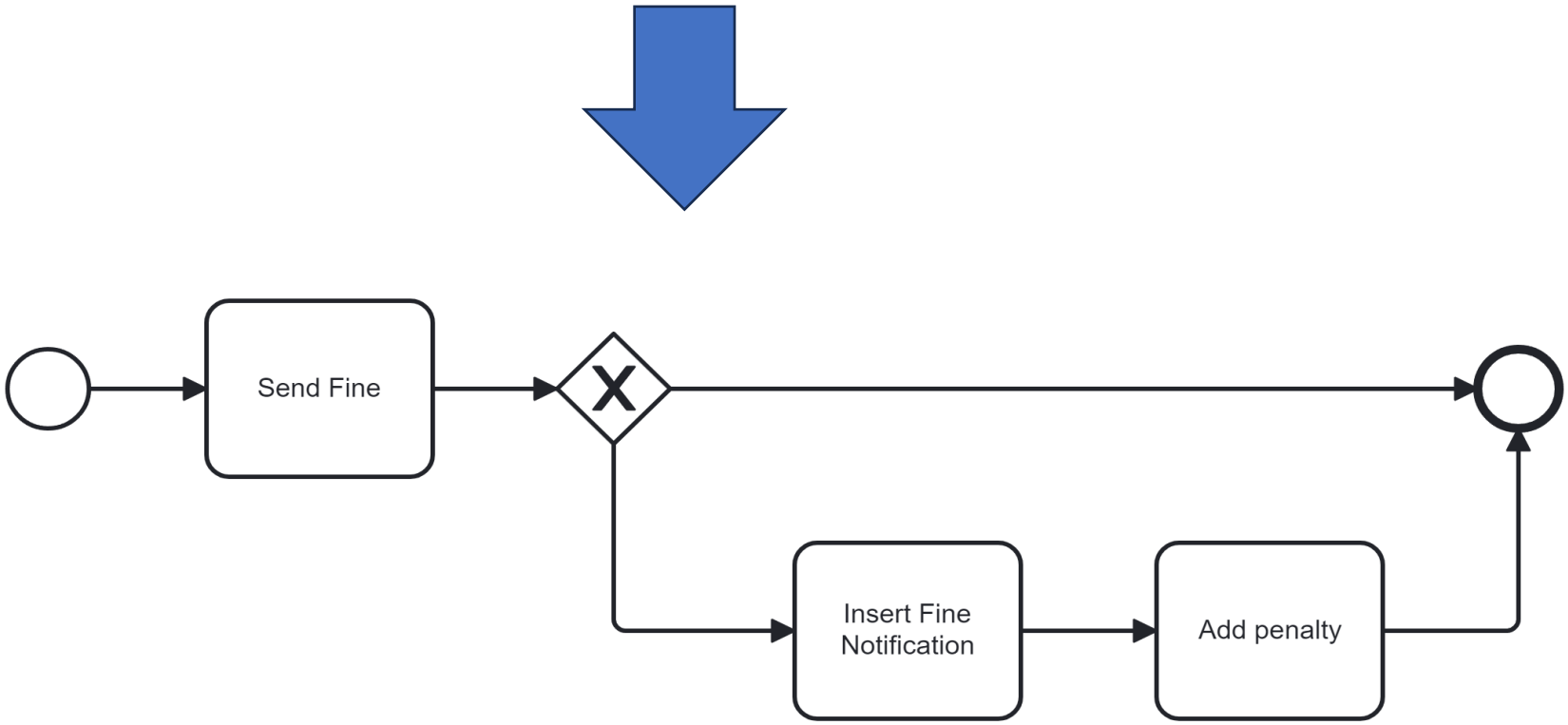
QUT the university for the real world

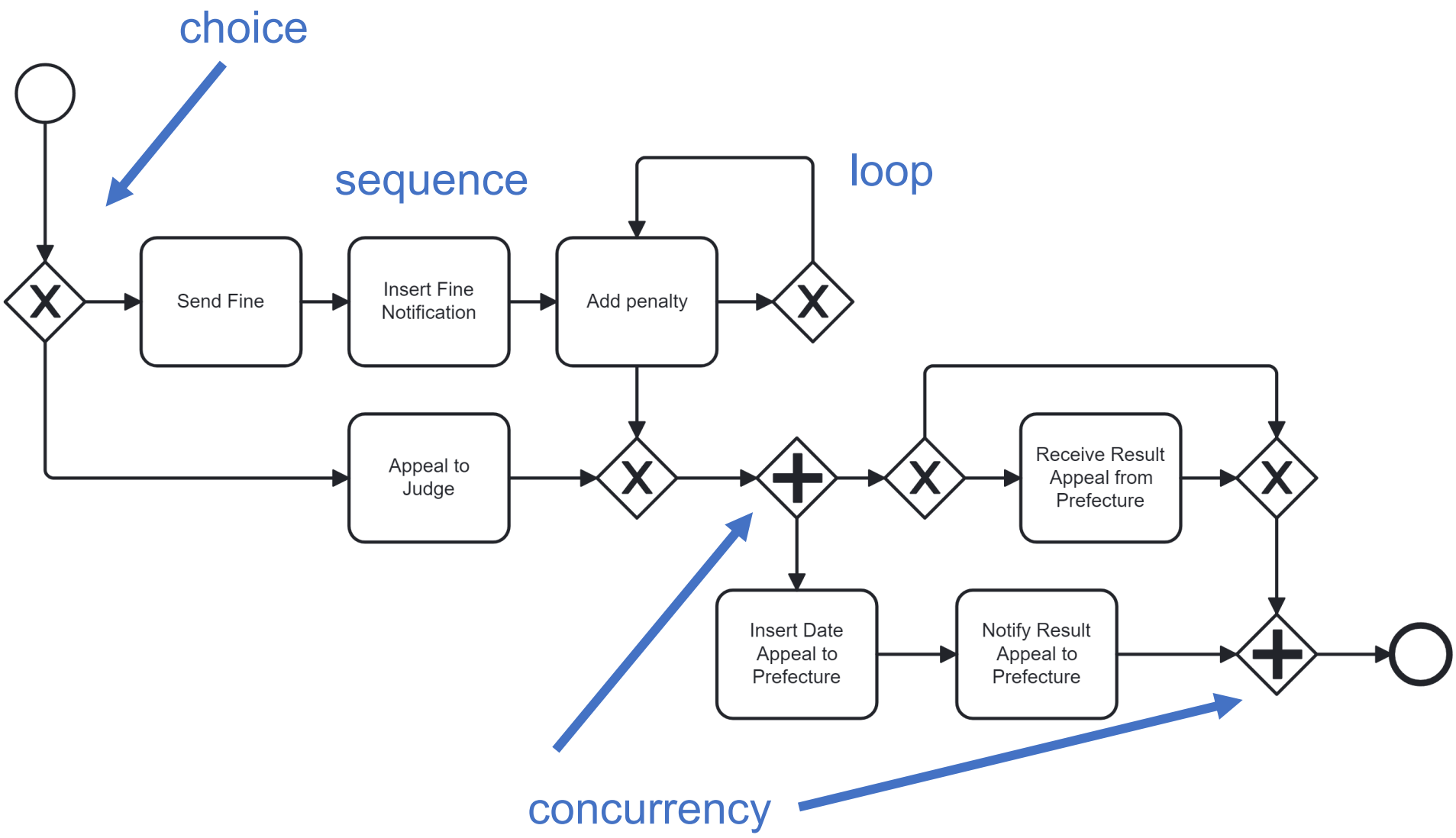
Adam Burke
Moe Thandar Wynn
Process Science

Process Mining = Data mining + Process Mgt + Visualisation

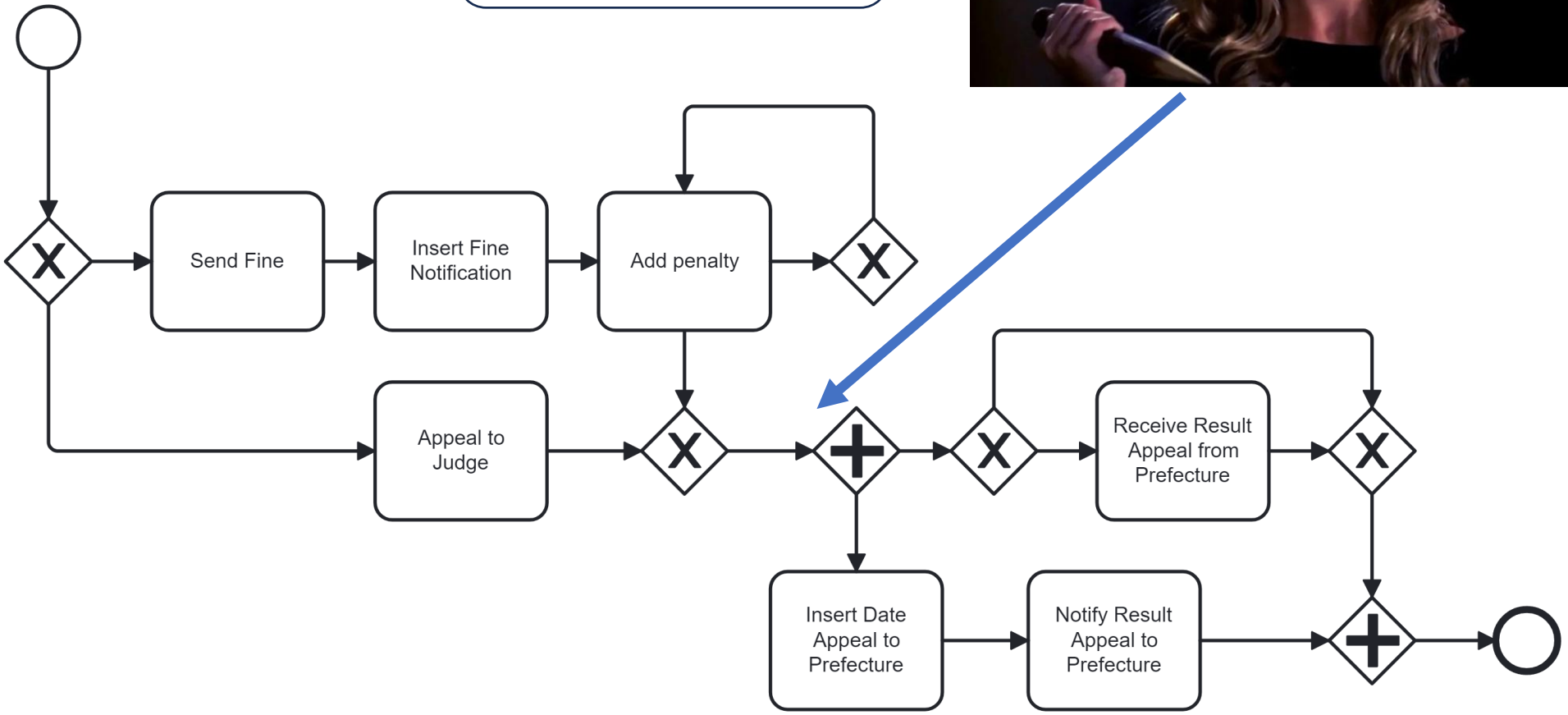


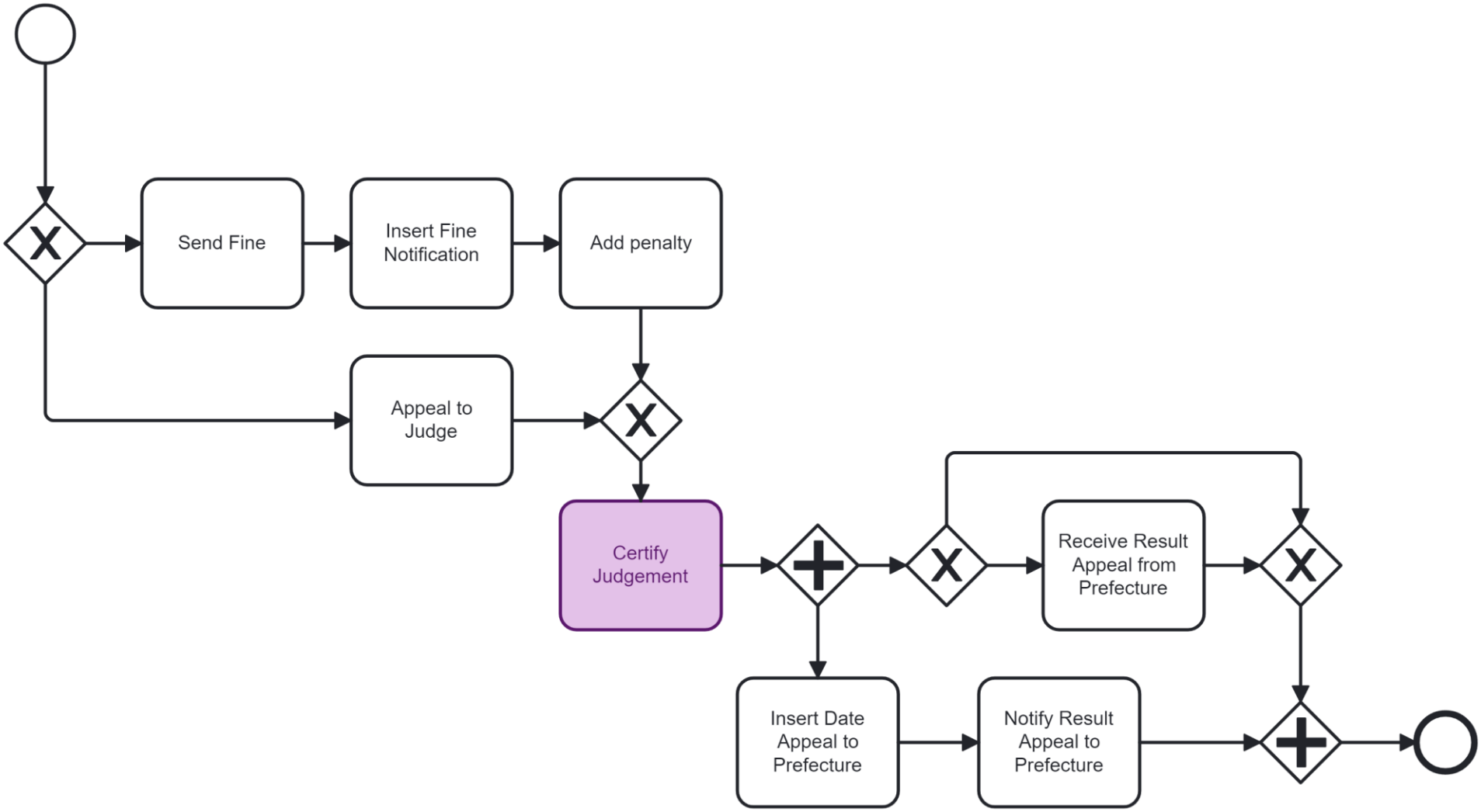
Case ID	Activity	Timestamp	Expense	Amount
A10008	**Send Fine**	**2007-07-17 08:00:00**	13.0	
A10008	**Insert Fine Notification**	**2007-08-02 08:00:00**		
A10008	**Add penalty**	**2007-10-01 08:00:00**		74.0
A11668	**Send Fine**	**2007-08-02 08:00:00**	13.0	



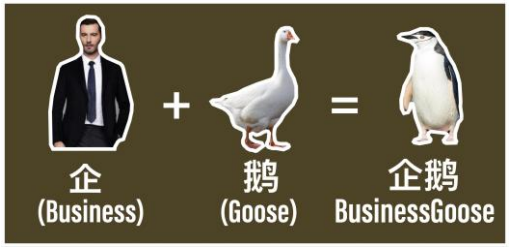

choice
sequence
loop
Send Fine
Insert Fine Notification
Add penalty
Appeal to Judge
Receive Result Appeal from Prefecture
Insert Date Appeal to Prefecture
Notify Result Appeal to Prefecture
concurrency

missing!


Send Fine
Insert Fine Notification
Add penalty
Appeal to Judge
Receive Result Appeal from Prefecture
Insert Date Appeal to Prefecture
Notify Result Appeal to Prefecture

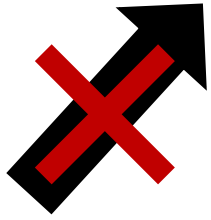

Send Fine
Insert Fine Notification
Add penalty
Appeal to Judge
Certify Judgement
Receive Result Appeal from Prefecture
Insert Date Appeal to Prefecture
Notify Result Appeal to Prefecture

Data Linkages



Lorem Ipsum
- **Dolor**
- sit amet
- consectetur adispiscing elit

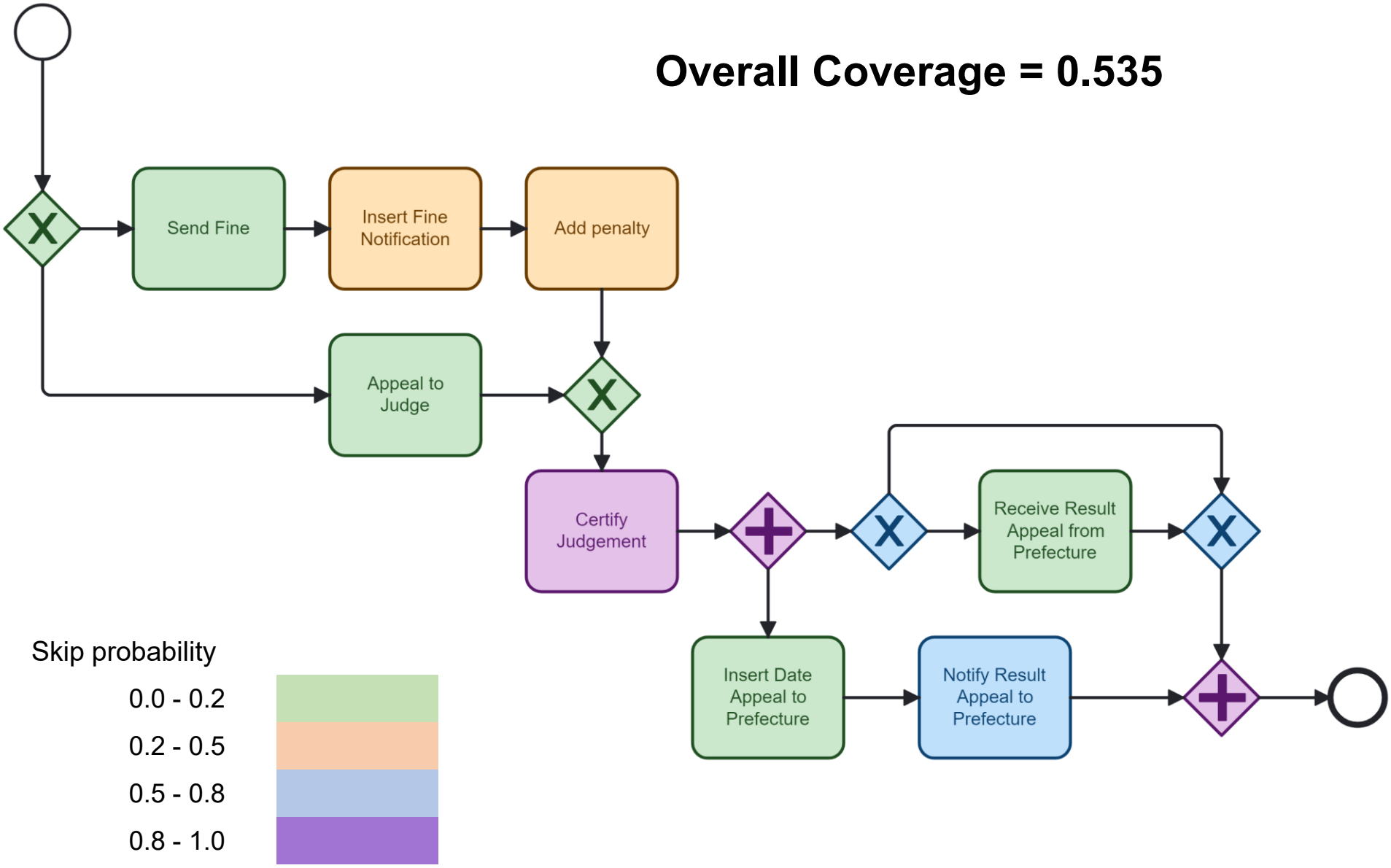
Sed
- **Do Eismod ID**
- tempor incidunt
- ut labore et dolore



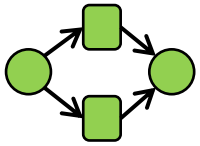
Magna
- **Aliqua ID**
- ut enim
- ad minim veniam

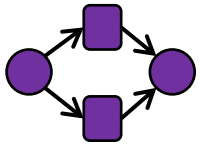
Quis nostrud
- **Exerciation ID**
- ullamco laboris
- nisi ut aliquip

Overall Coverage = 0.535



- How to handle this lack of data? …
  - Or perhaps we could say, this partially contradictory multi-modal data?
  - Reconcile?
  - Quantify?

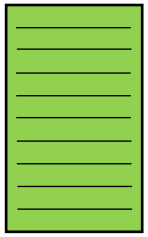




process modelling

process discovery

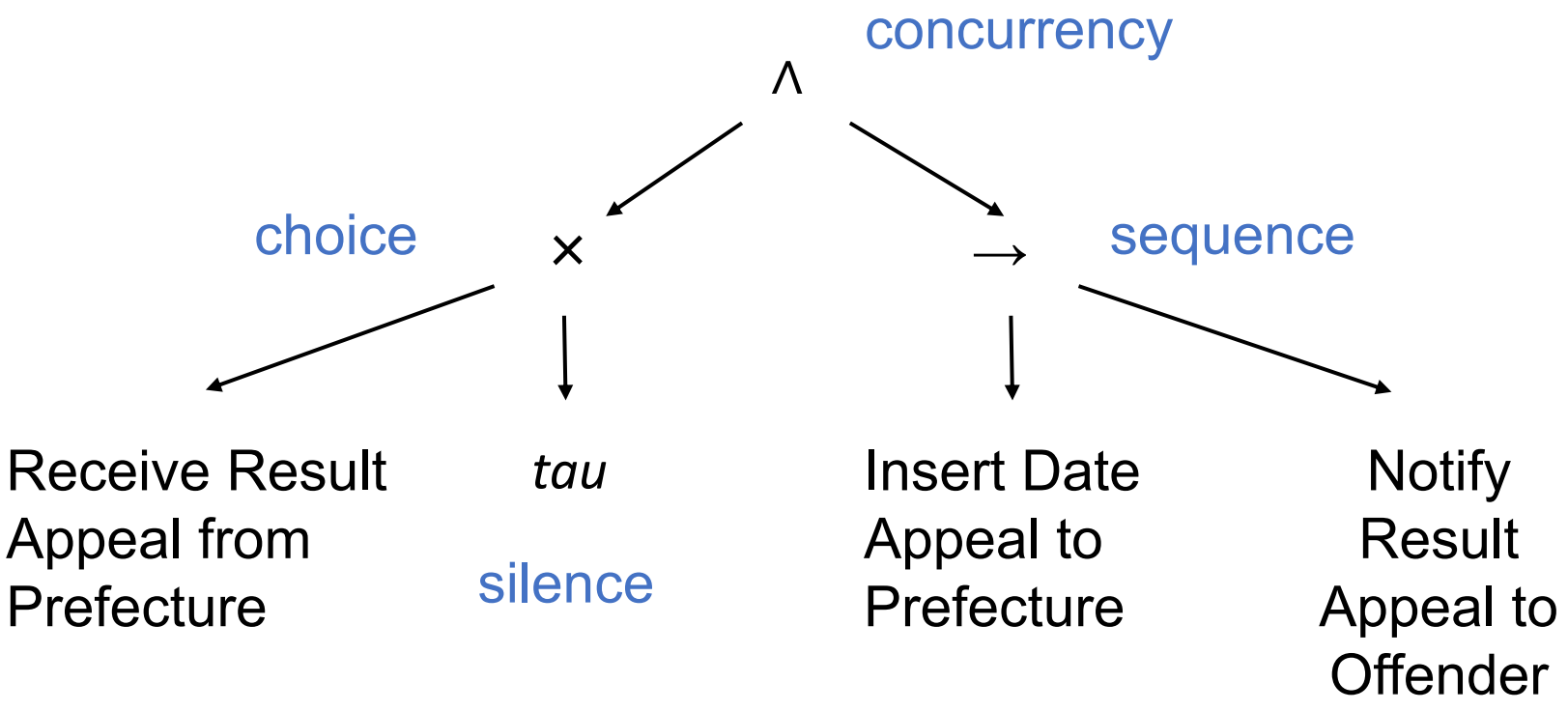
process model 1 | process model 2



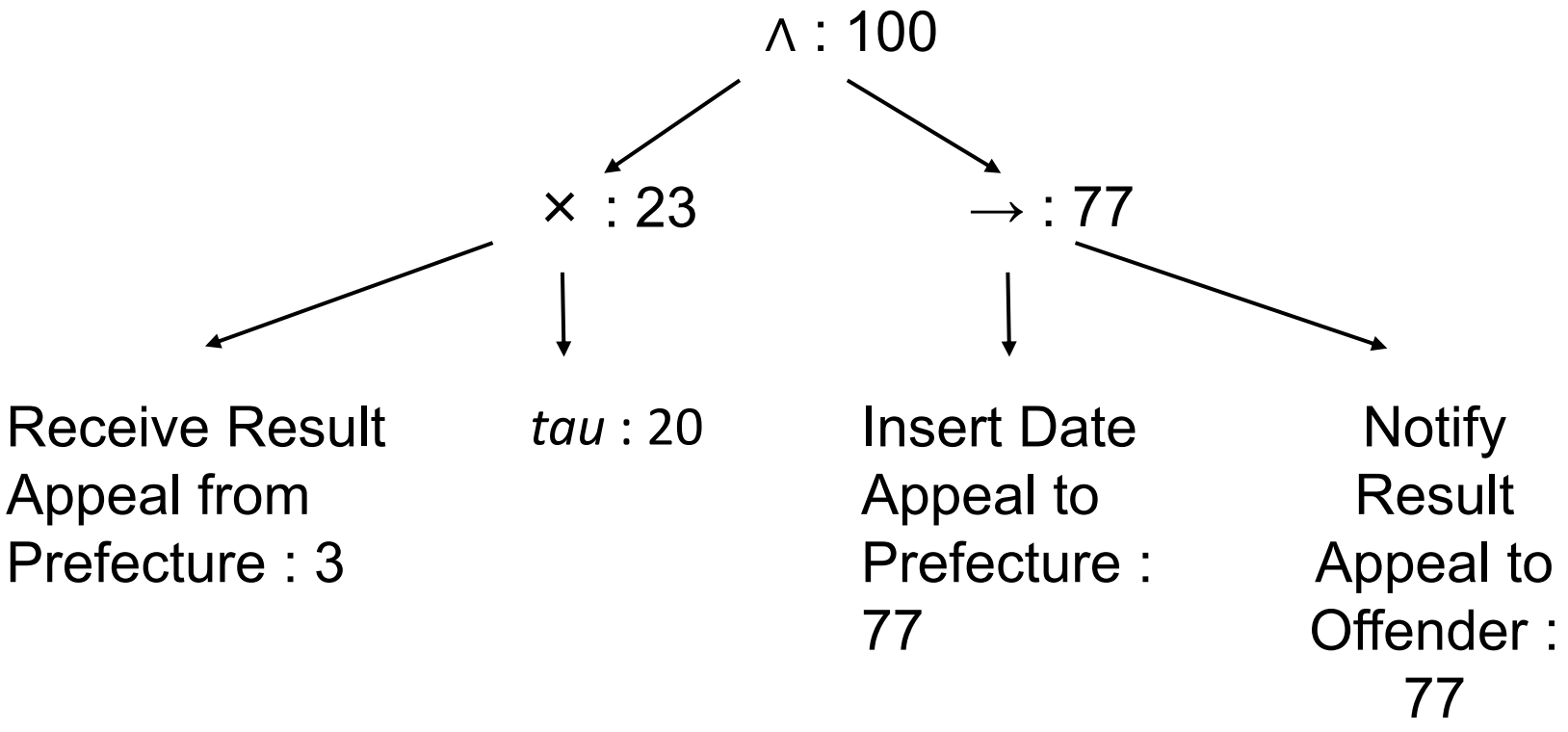
event log
(who, what, when)

Process Trees
+ Stochastics
+ Skip Alignments

Process Trees



Process Trees with Stochastics



Stochastic Languages

- Stochastic models can yield a distribution of traces called a *stochastic language*

  <a,b,c> : 0.5 <a,c> : 0.3 <a,b> :2

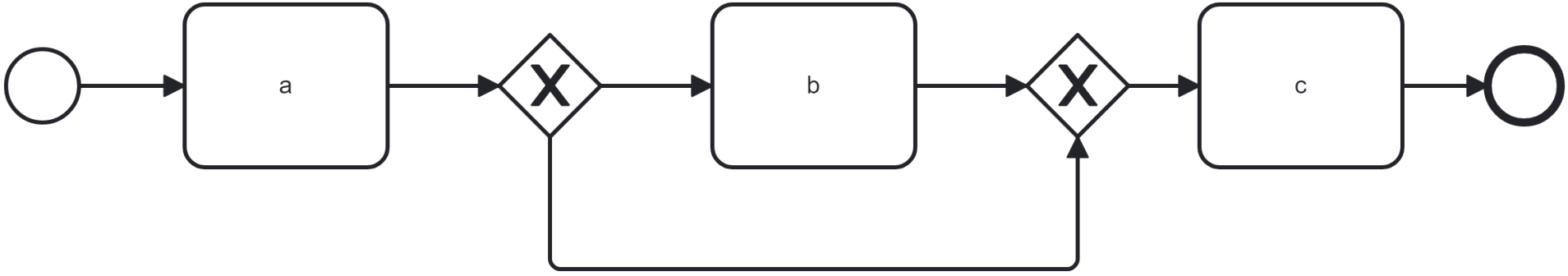
- Event logs can *already* be straightforwardly transformed into a stochastic language
- Stochastic process mining
  - *Discovering* stochastic process models
  - *Comparing* logs and models
- Metrics
  - Earth-movers stochastic conformance measure
  - Entropic relevance
  - Alpha precision
  - …

<a,***void***,c> : 3

With a process void, ***may be no intersection between log and model stochastic languages***

Alignments

Alignments describe matches and mismatches between
- *traces in an event log* and
- *paths through a process model*



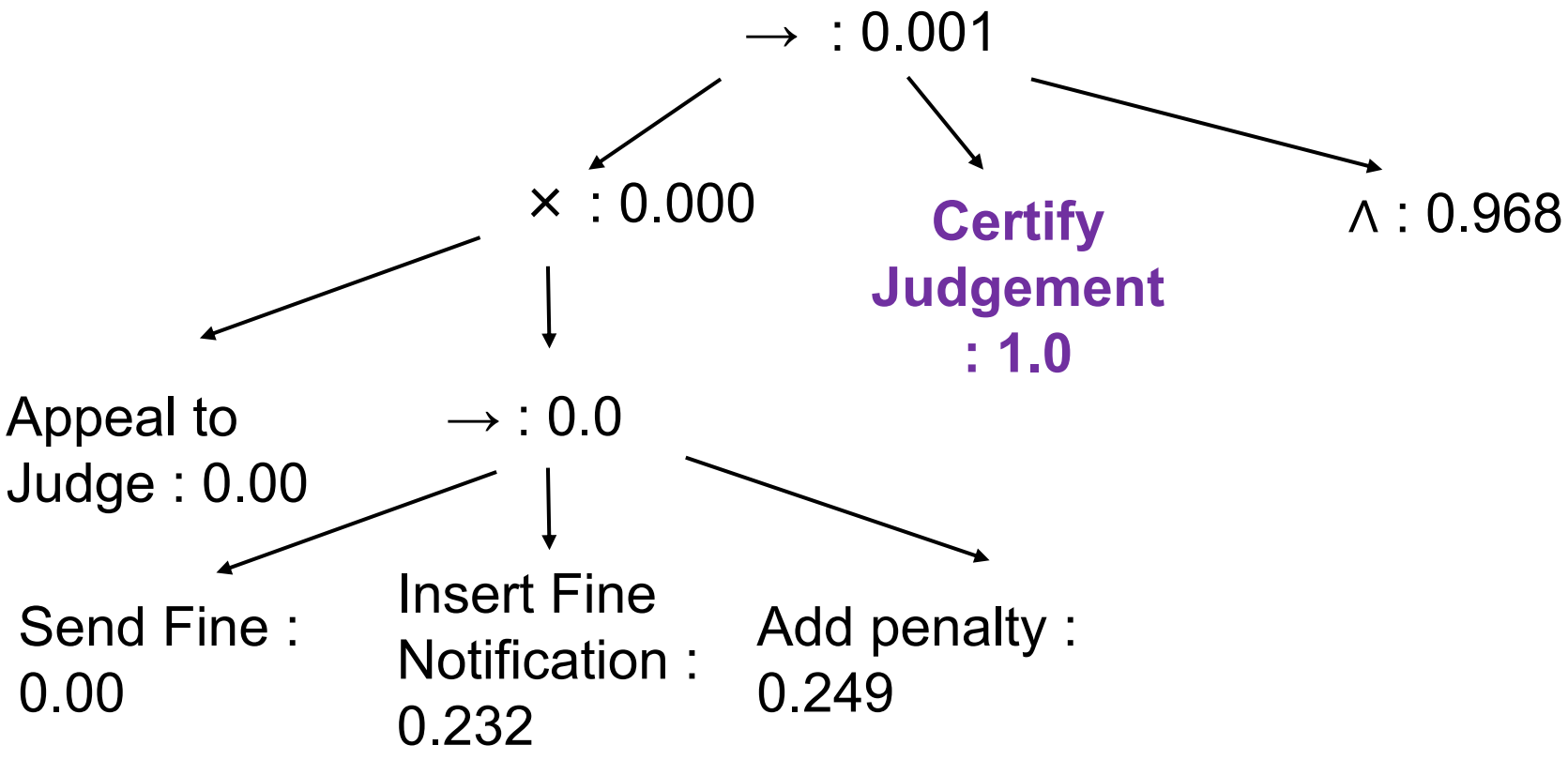
Log:
<a,b,c> : 10 <a,c> : 3 **<a,b> :5**

log trace	a	b	> >
model path	a	b	c

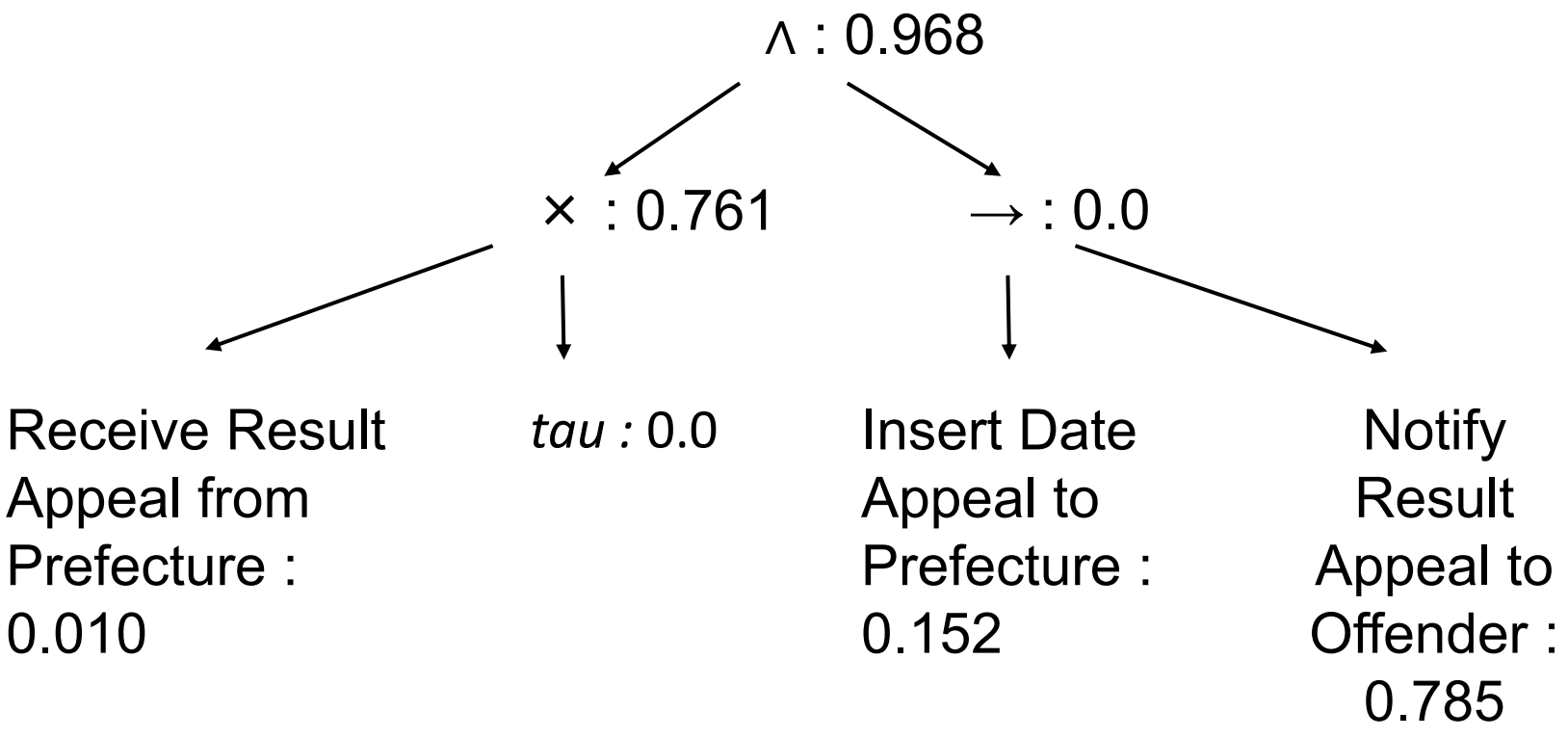
Skip Alignments

- Many possible alignments for non-trivial logs and models
  - Possibly infinite
- Computationally expensive to compute optimal alignments (*A** graph search algorithm)
- Summarise using a tree data structure: *skip alignments* (Bär et al, 2025)
- Combined with stochastic languages of log and model, can estimate *skip probabilities*
  - Probability that a particular activity or subprocess (subtree) is skipped

Skip Probabilities (1)



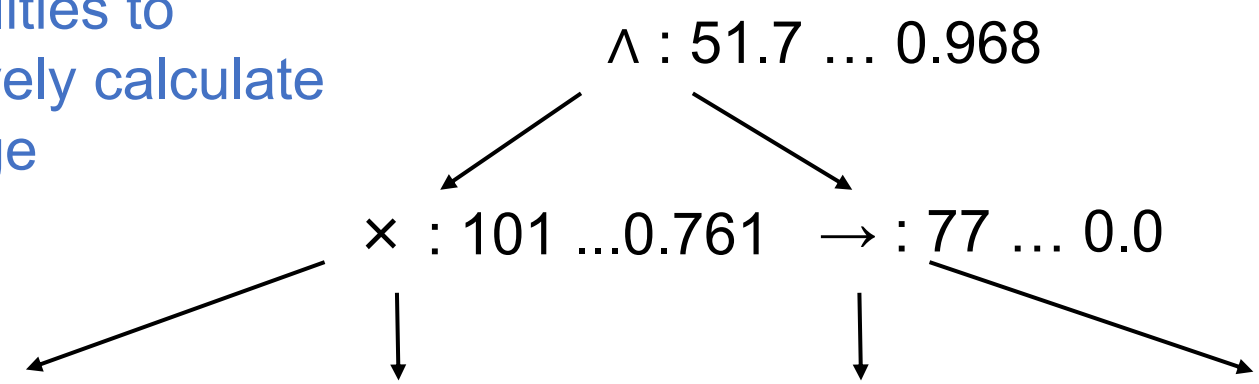
Skip Probabilities (2)



Coverage

Combine weights and probabilities to recursively calculate coverage

Overall Coverage = 0.535



Receive Result Appeal from Prefecture : 0.96 … 0.010

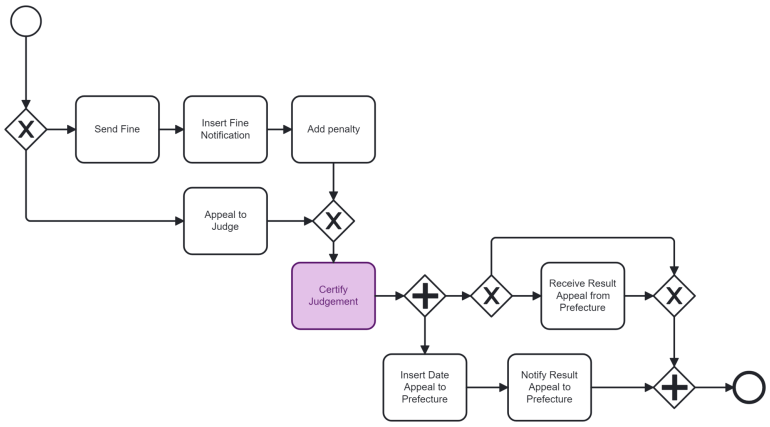
tau : 100 … 0.0

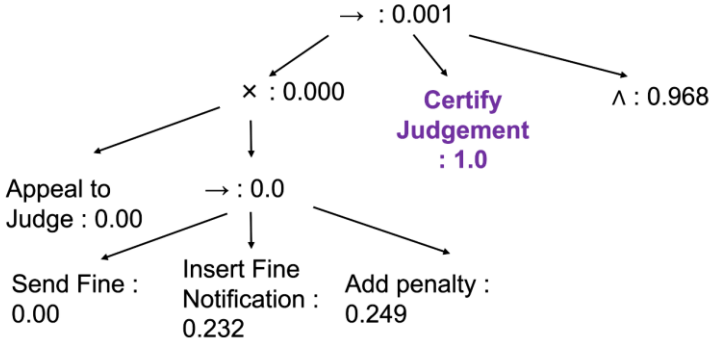
Insert Date Appeal to Prefecture : 4.02 … 0.152

Notify Result Appeal to Offender : 0.861 … 0.785

Summary

- Data may often be missing on end-to-end process paths – *process voids*
- We can reason about these regions of missing data using
  - Domain expert input, captured as process models
  - Stochastic process models
  - Skip alignments
- This allows end-to-end process analytics while still precisely describing real data gaps





Process Voids

Data Science Without Data

November 2025

QUT Centre for Data Science

Adam Burke
Moe Thandar Wynn
QUT Process Science

References

Background

- van der Aalst, W. M. P. (2016). *Process Mining: Data Science in Action* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/978-3-662-49851-4
- van der Aalst, W. M., & Carmona, J. (2022). *Process mining handbook*. Springer Nature.

Skip Alignments

- Bär, P., Wynn, M. T., & Leemans, S. J. (2025). *A full picture in conformance checking: Efficiently summarizing all optimal alignments*. International Conference on Business Process Management, 69–87.
- Bär, P., Burke, A. T., Wynn, M. T., & Leemans, S. J. (2025). *Skip Probabilities for Subprocesses*. International Conference on Process Mining (ICPM).