

# Discovering and Navigating a Collection of Process Models using Multiple Quality Dimensions

Joos Buijs (j.c.a.m.buijs@tue.nl)

Boudewijn van Dongen

Wil van der Aalst



<http://www.win.tue.nl/coselog/>

**TU/e**

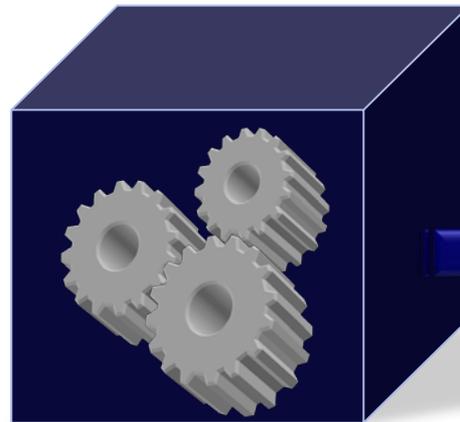
Technische Universiteit  
Eindhoven  
University of Technology

Where innovation starts

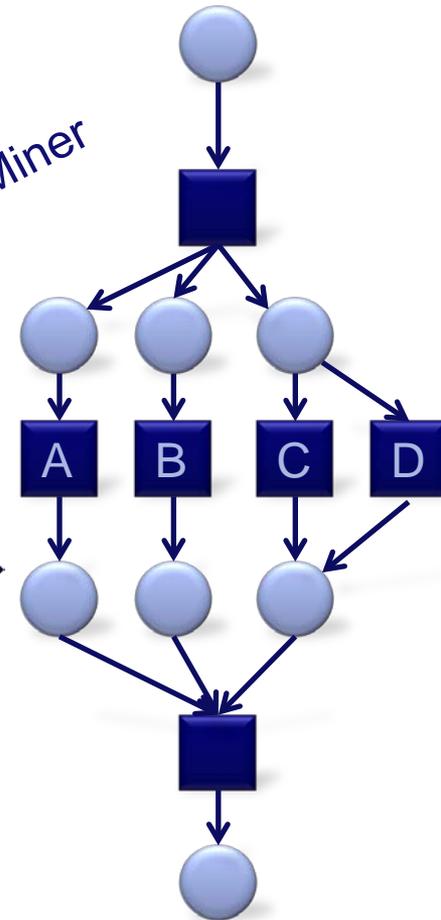
# Process Discovery

## Eventlog

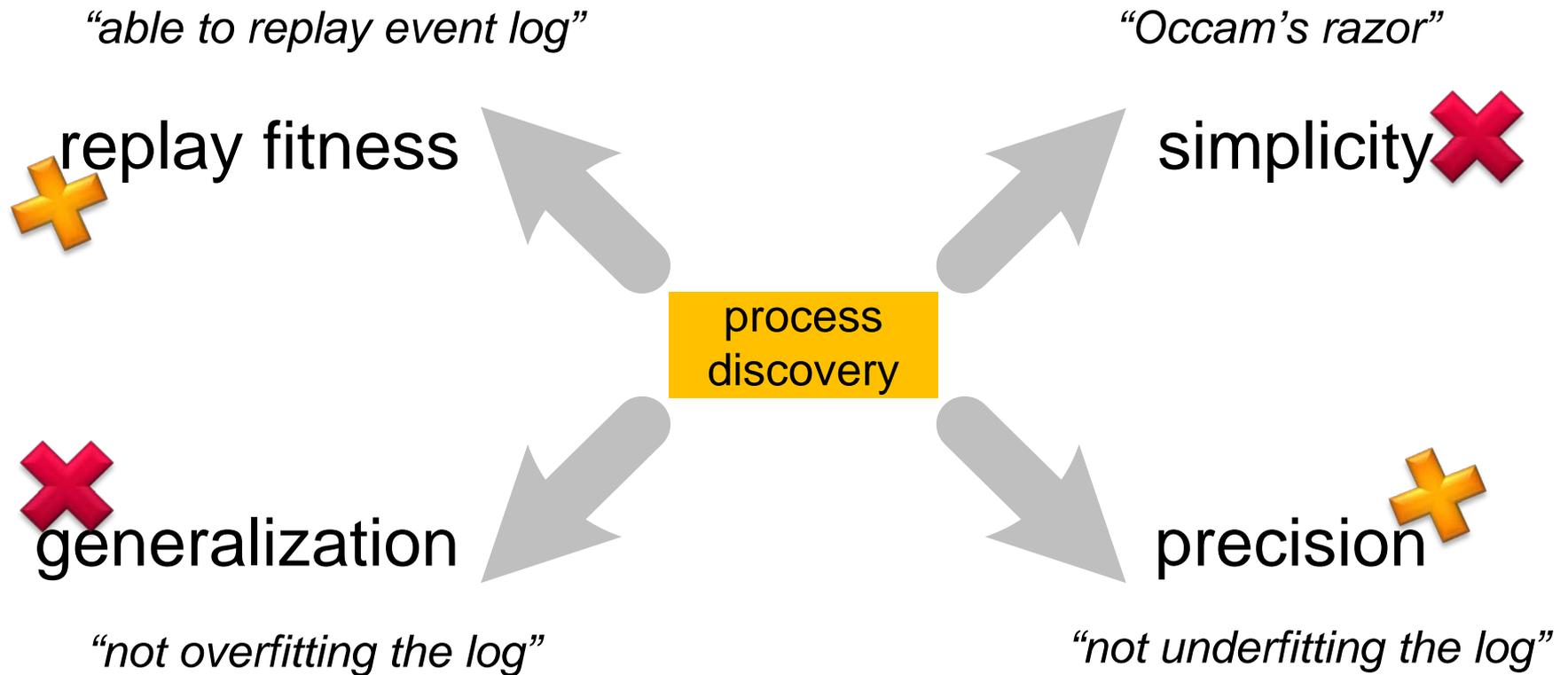
< A, B, C, D >  
< D, C, B, A >  
< A, C, D, B >  
< B, A, C, D >



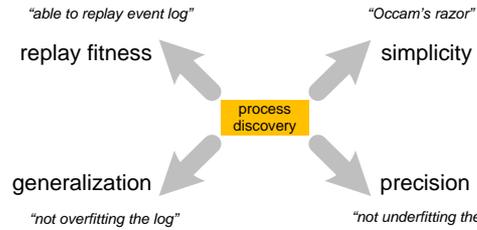
Alpha Algorithm  
ILP Miner  
Fuzzy Miner  
Heuristics Miner



# Challenge: Four Competing Quality Criteria

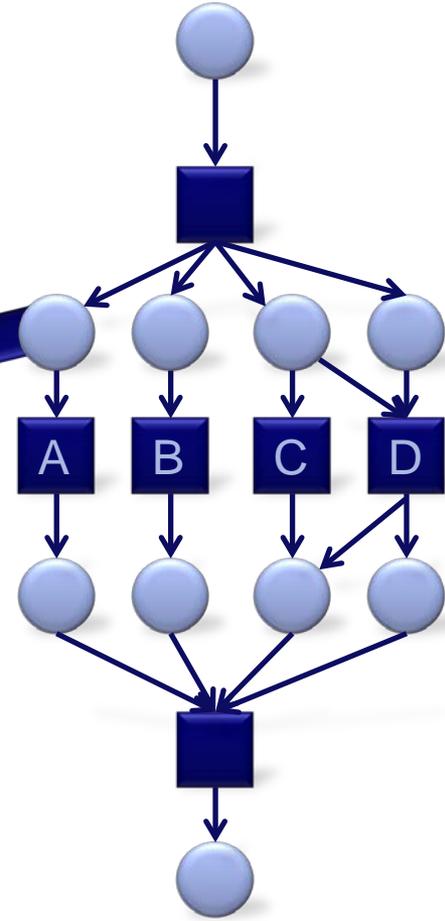


# The Evolutionary Tree Miner (ETM) Algorithm



**Eventlog**

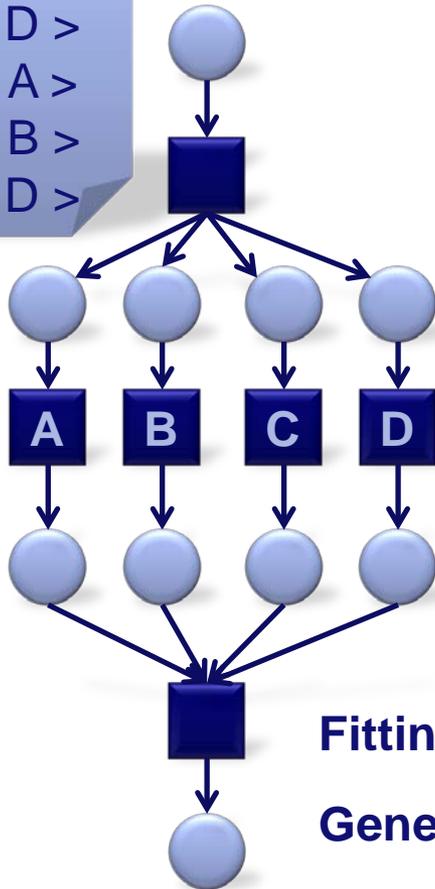
- < A, B, C, D >
- < D, C, B, A >
- < A, C, D, B >
- < B, A, C, D >



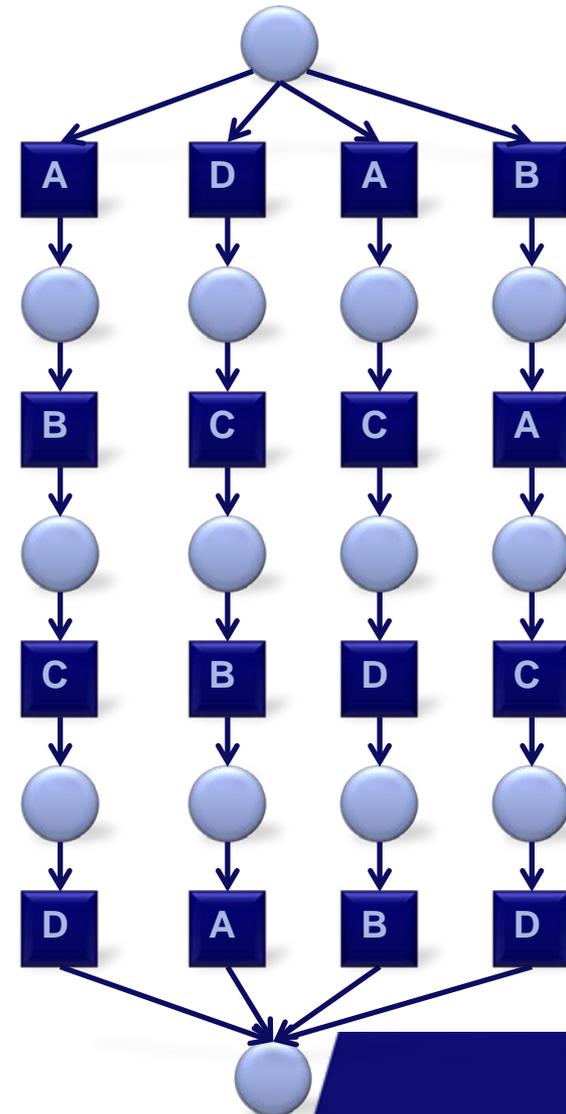
# Which Process Model is 'Best'?

## Eventlog

< A, B, C, D >  
< D, C, B, A >  
< A, C, D, B >  
< B, A, C, D >



**Fitting**  
**Generic**  
**Simple**

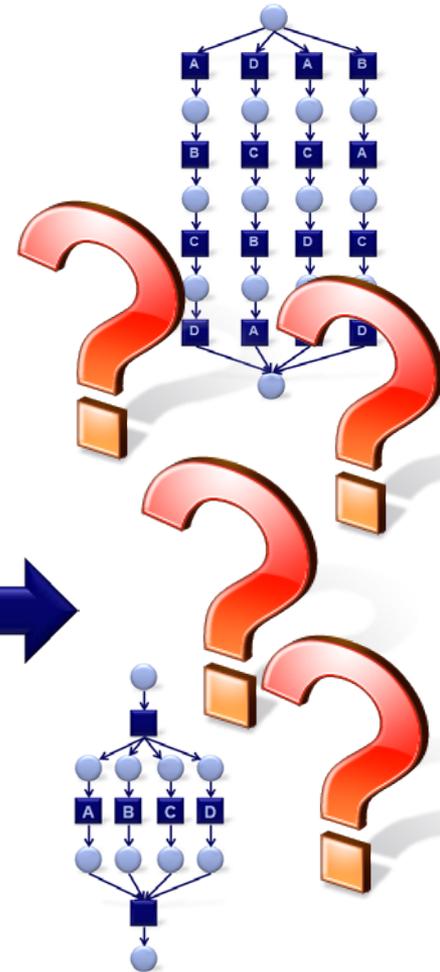


**Fitting**  
**Precise**

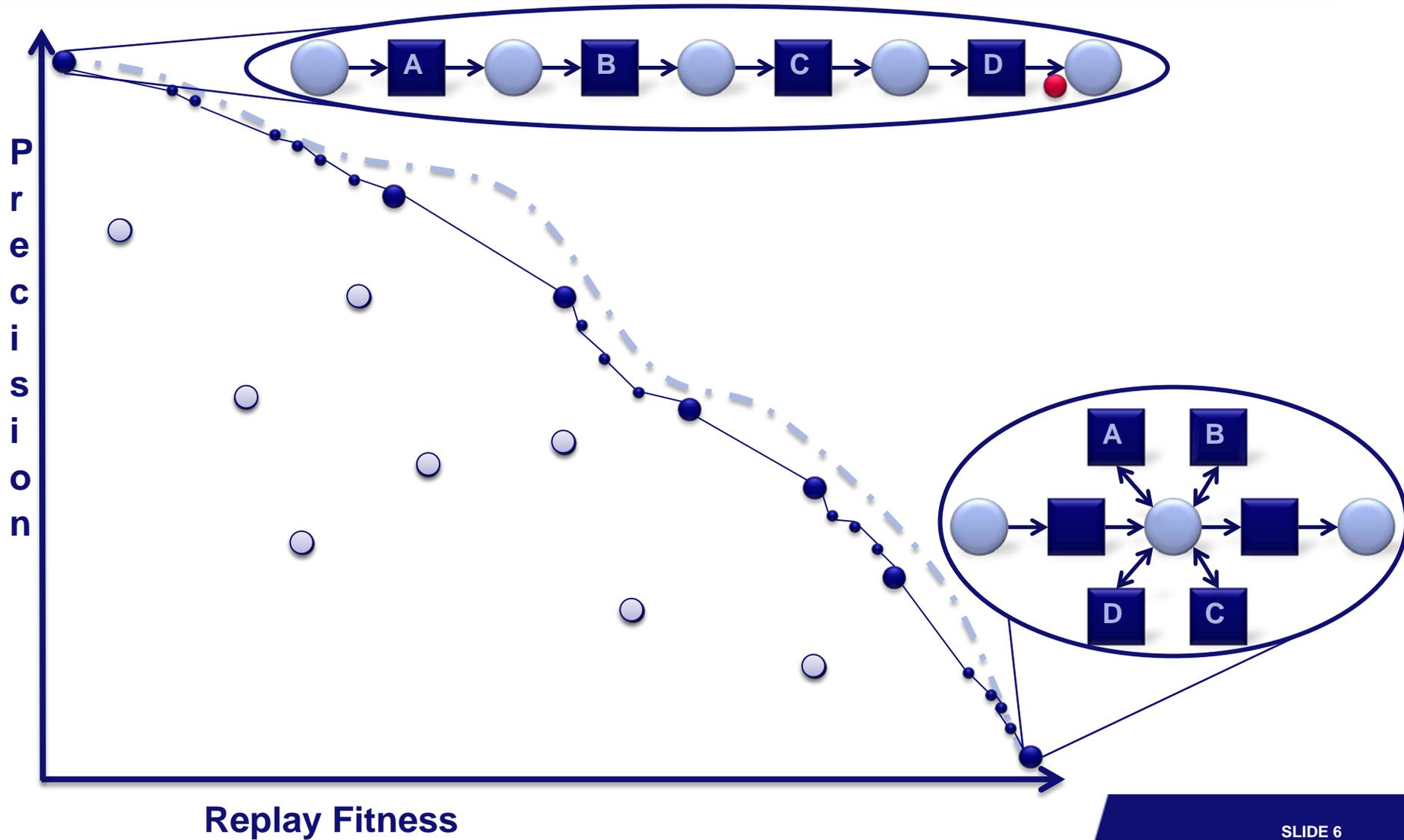
# Discovery of Multiple Process Models

## Eventlog

< A, B, C, D >  
< D, C, B, A >  
< A, C, D, B >  
< B, A, C, D >



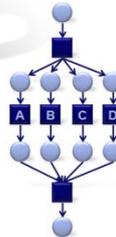
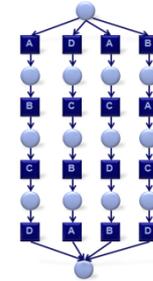
# A Pareto Front of Process Models



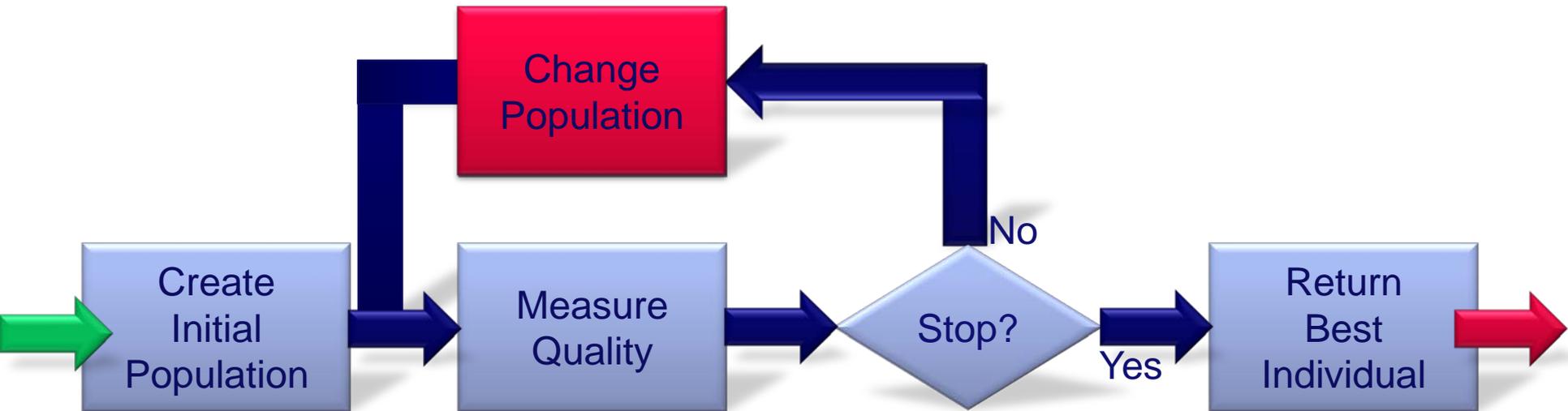
# Discovery of a Pareto Front

## Eventlog

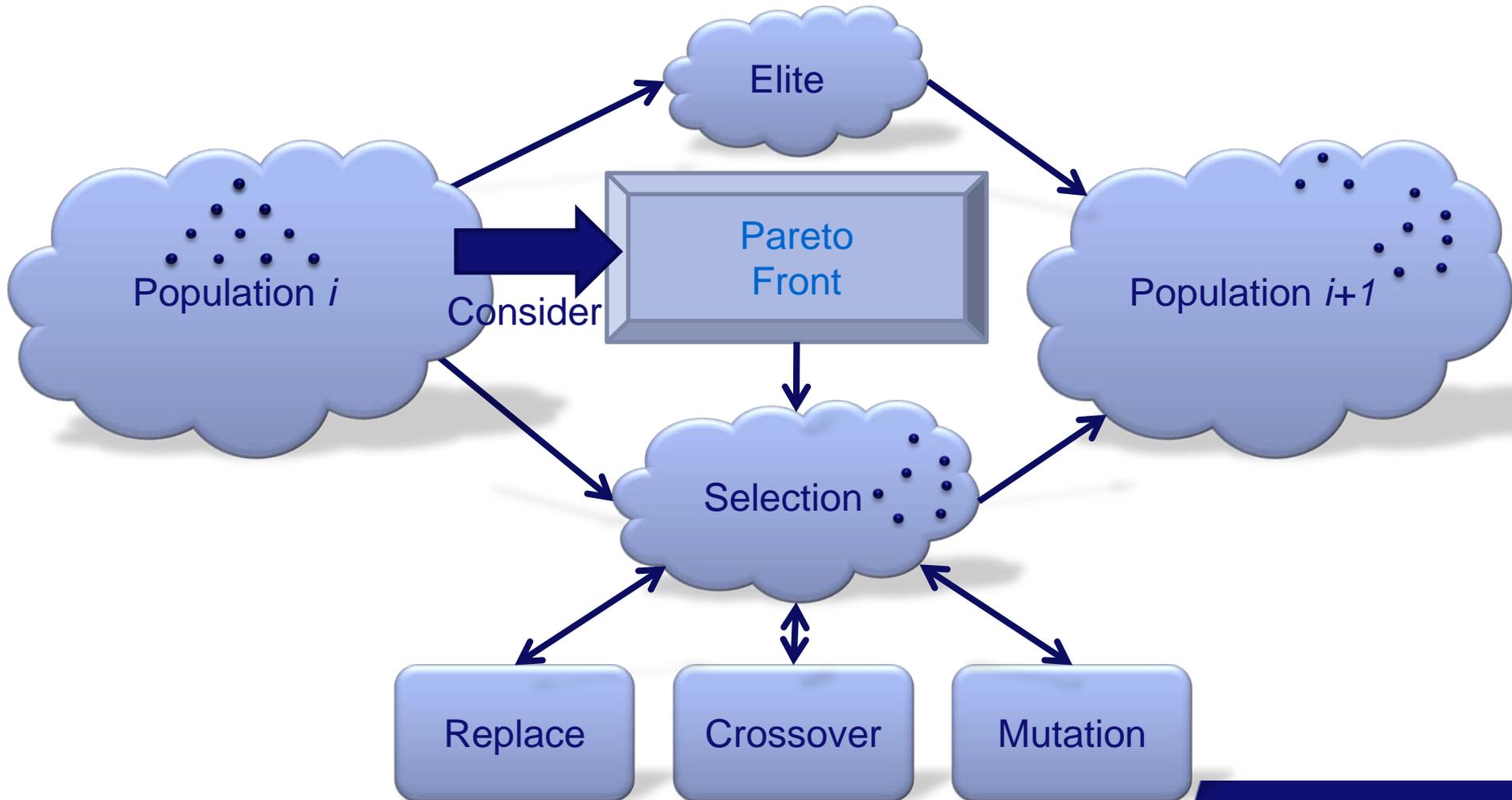
< A, B, C, D >  
< D, C, B, A >  
< A, C, D, B >  
< B, A, C, D >



# Steps of the Genetic ETM Algorithm



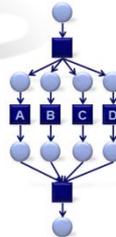
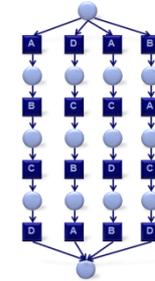
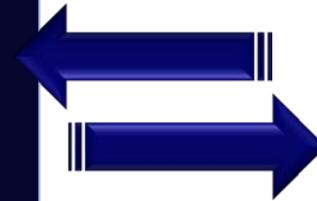
# Incorporating the Pareto Front



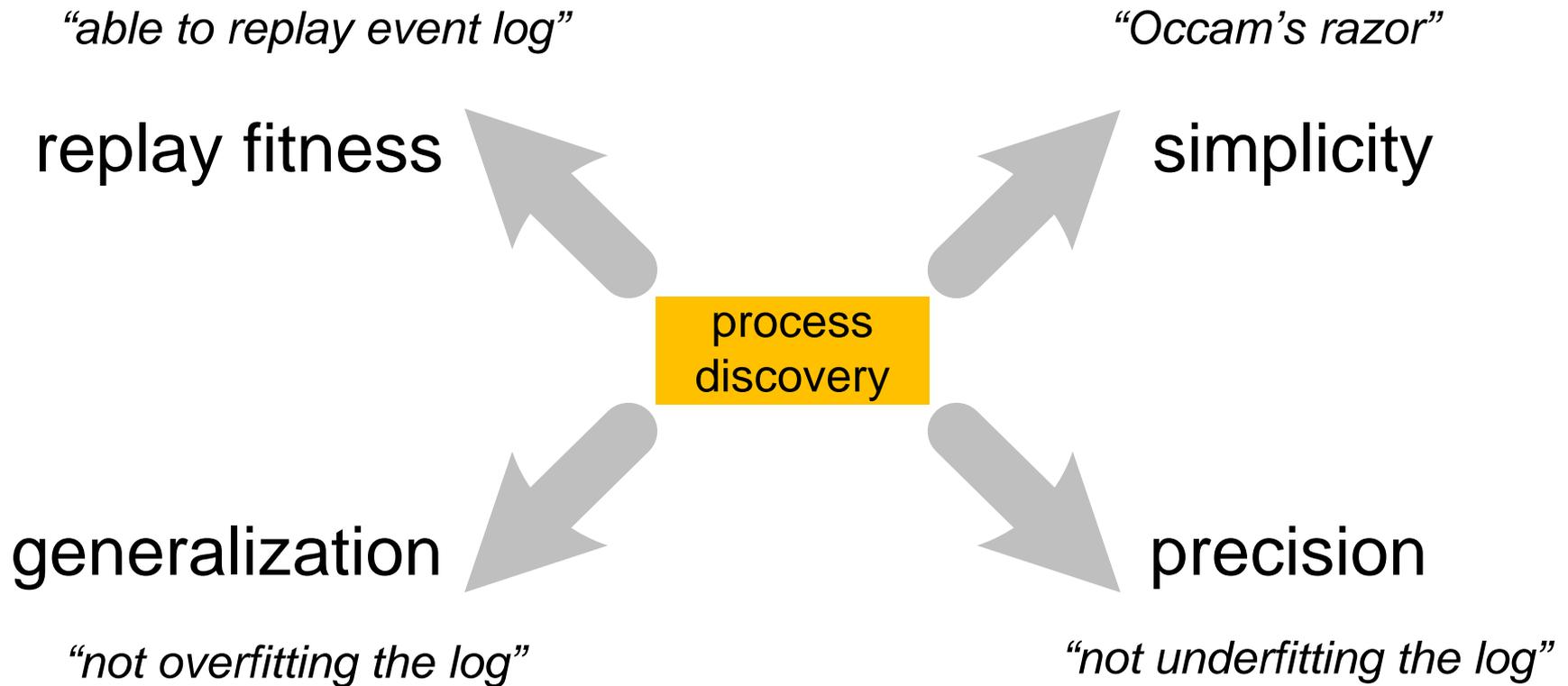
# ETM with Pareto front extension

## Eventlog

< A, B, C, D >  
< D, C, B, A >  
< A, C, D, B >  
< B, A, C, D >



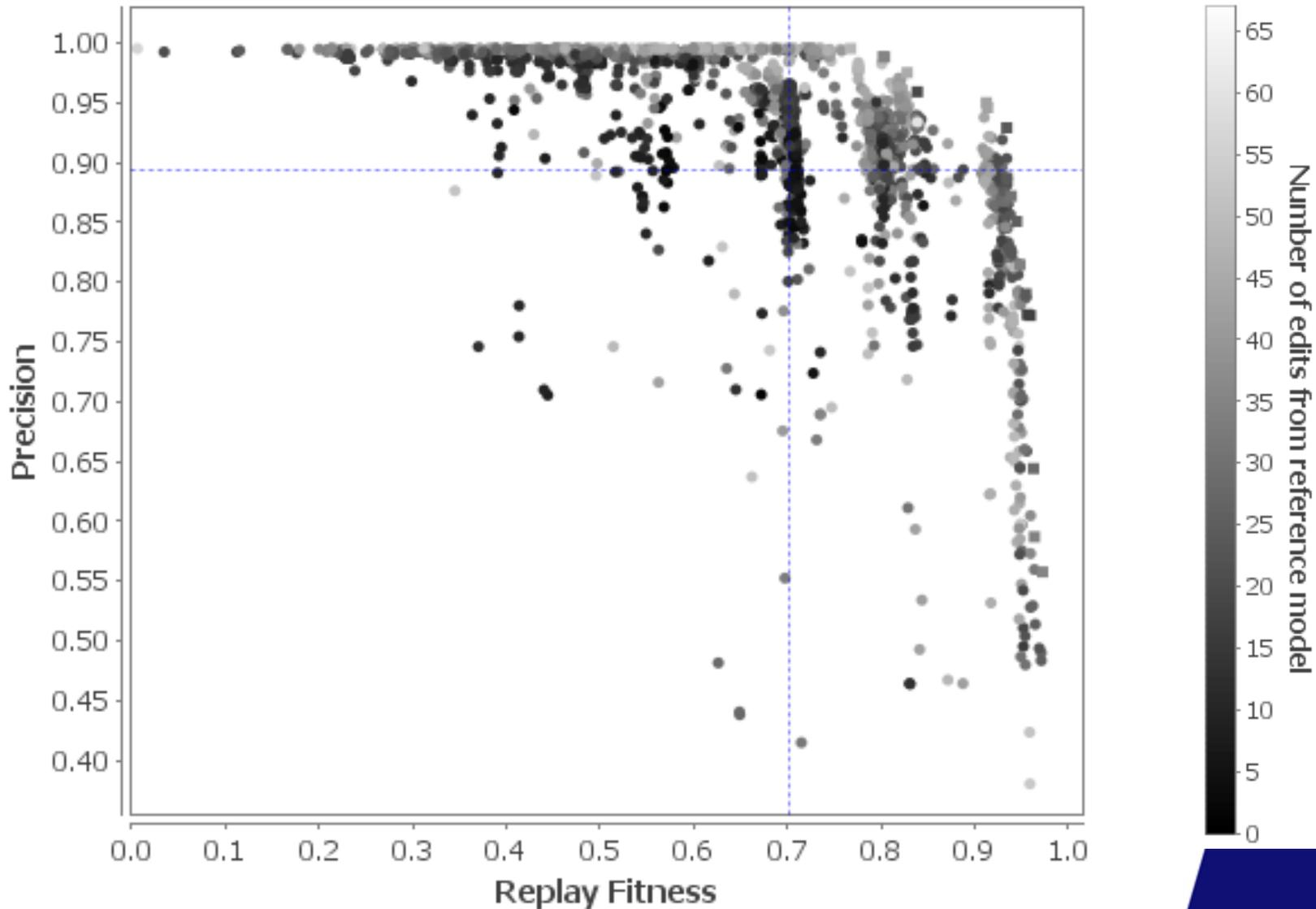
# Application on a Real Life Event Log



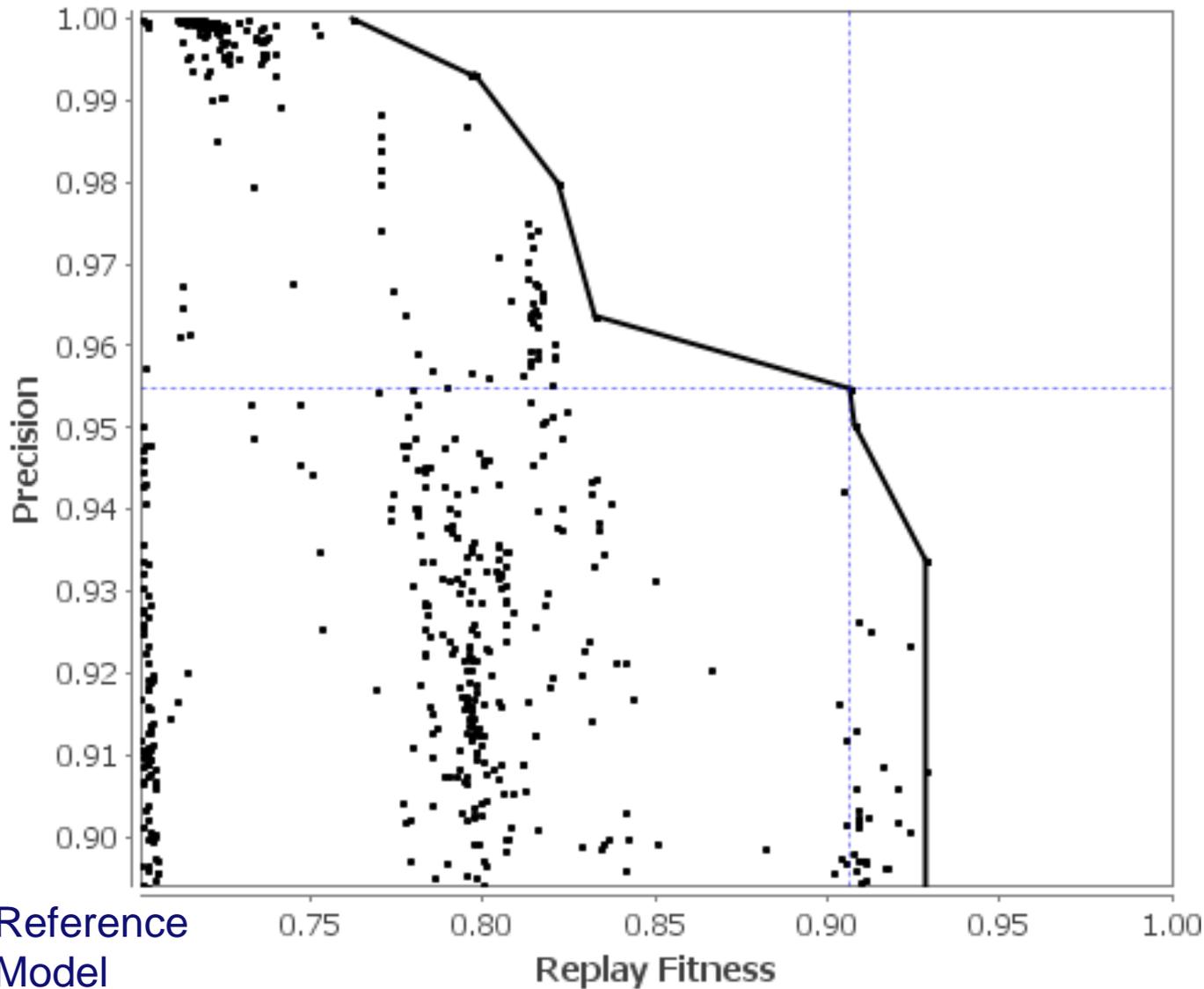
**5<sup>th</sup> Dimension: #Edits from given process model**

# Fitness vs. Precision

## Color indicates Number of Edits

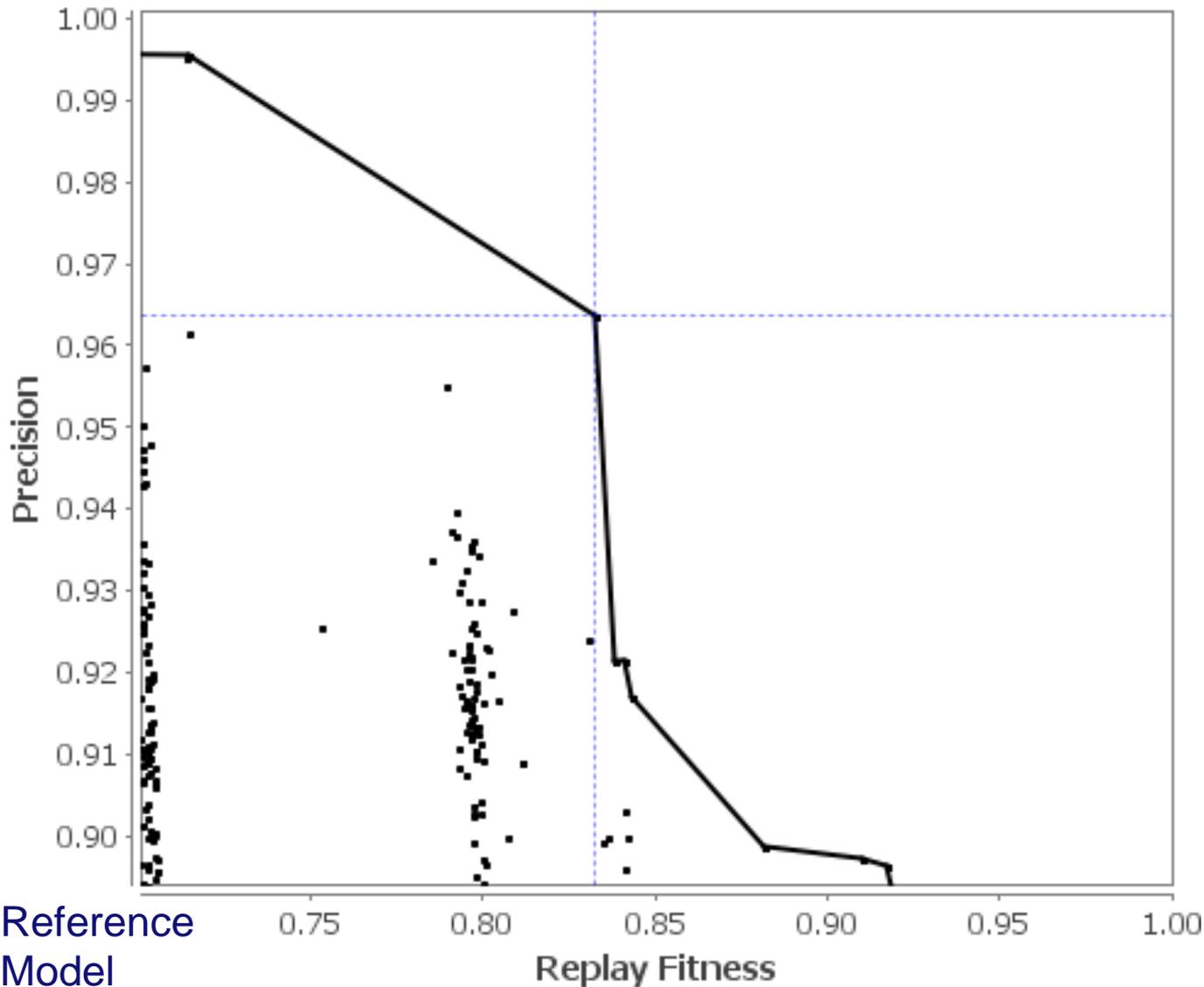


# Fitness vs. Precision Zoomed



# Fitness vs. Precision

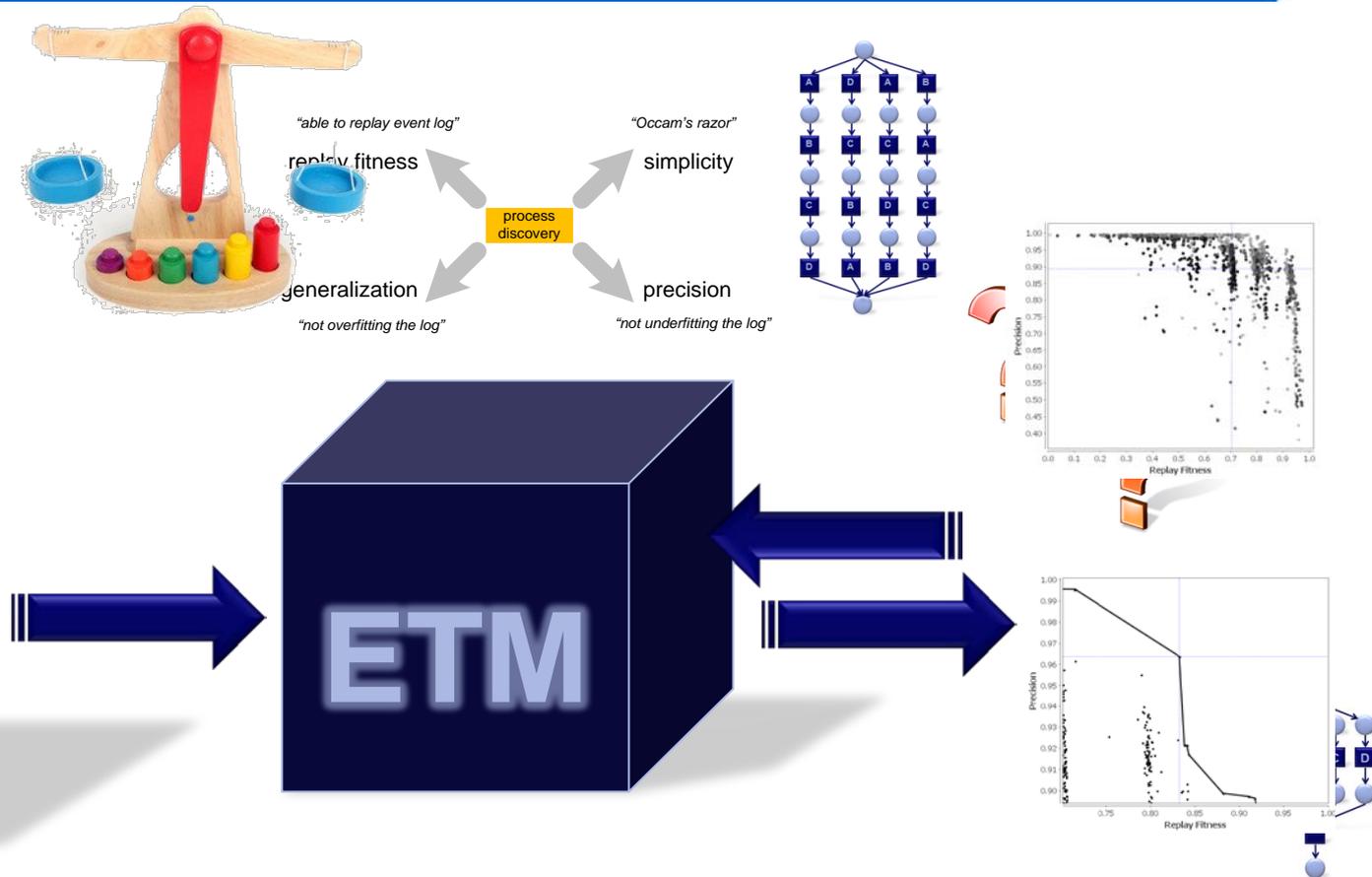
## Zoomed & Maximum of 20 Edits



# Summary

## Eventlog

< A, B, C, D >  
< D, C, B, A >  
< A, C, D, B >  
< B, A, C, D >



# Take Home Points

- **There is no single solution in process discovery**
- **A Pareto front can represent a collection of relevant process models with attached quality scores**
  - **It also allows for ‘absolute’ measures to be included**
- **Visualization of, and navigation through, a multidimensional Pareto Front is a challenge**

