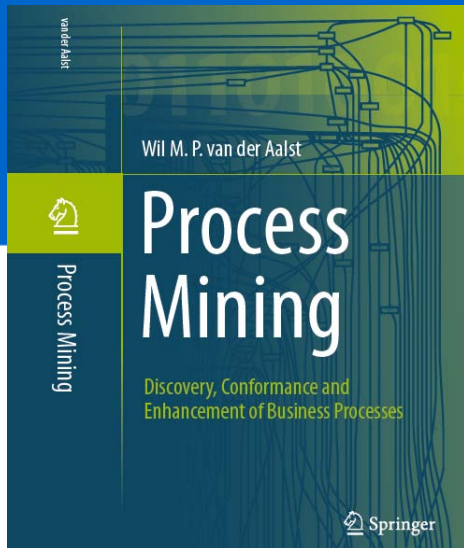


# Chapter 7

## Conformance Checking

prof.dr.ir. Wil van der Aalst  
[www.processmining.org](http://www.processmining.org)



**TU** / **e** Technische Universiteit  
**Eindhoven**  
University of Technology

Where innovation starts

# Overview

Chapter 1  
Introduction

---

## *Part I: Preliminaries*

Chapter 2  
Process Modeling and  
Analysis

Chapter 3  
Data Mining

---

## *Part II: From Event Logs to Process Models*

Chapter 4  
Getting the Data

Chapter 5  
Process Discovery: An  
Introduction

Chapter 6  
Advanced Process  
Discovery Techniques

---

## *Part III: Beyond Process Discovery*

Chapter 7  
Conformance  
Checking

Chapter 8  
Mining Additional  
Perspectives

Chapter 9  
Operational Support

---

## *Part IV: Putting Process Mining to Work*

Chapter 10  
Tool Support

Chapter 11  
Analyzing “Lasagna  
Processes”

Chapter 12  
Analyzing “Spaghetti  
Processes”

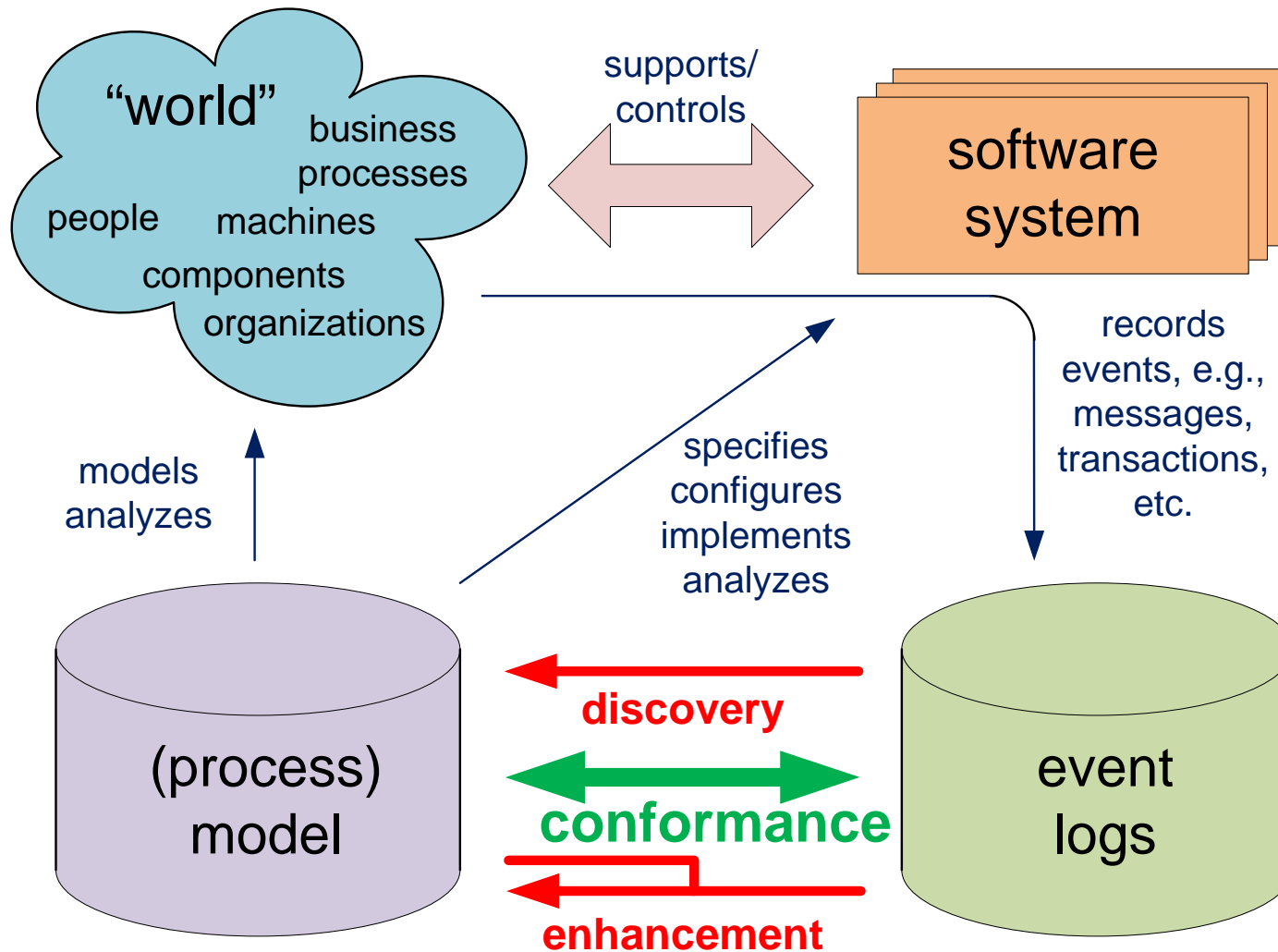
---

## *Part V: Reflection*

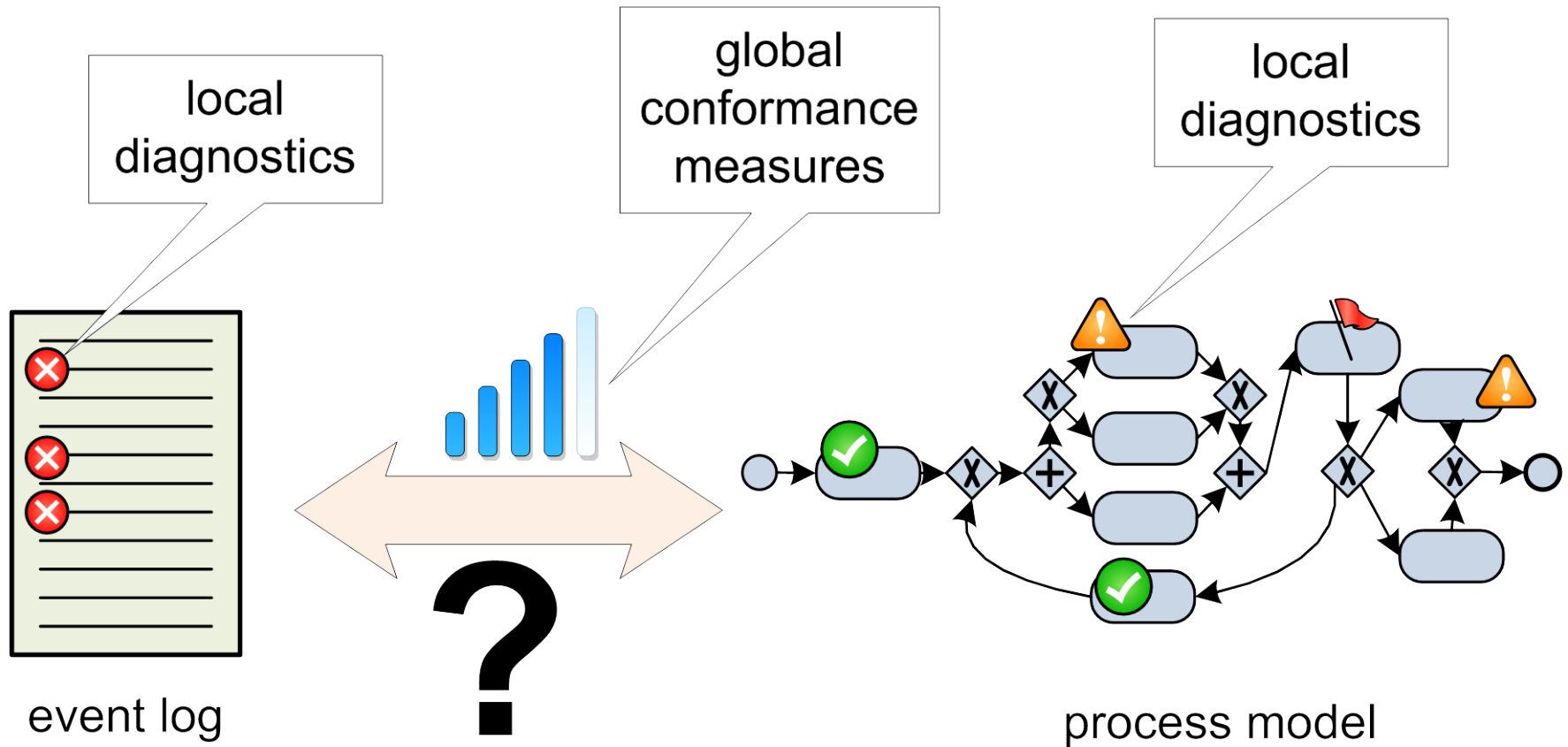
Chapter 13  
Cartography and  
Navigation

Chapter 14  
Epilogue

# Conformance checking



# Using conformance checking



# Context

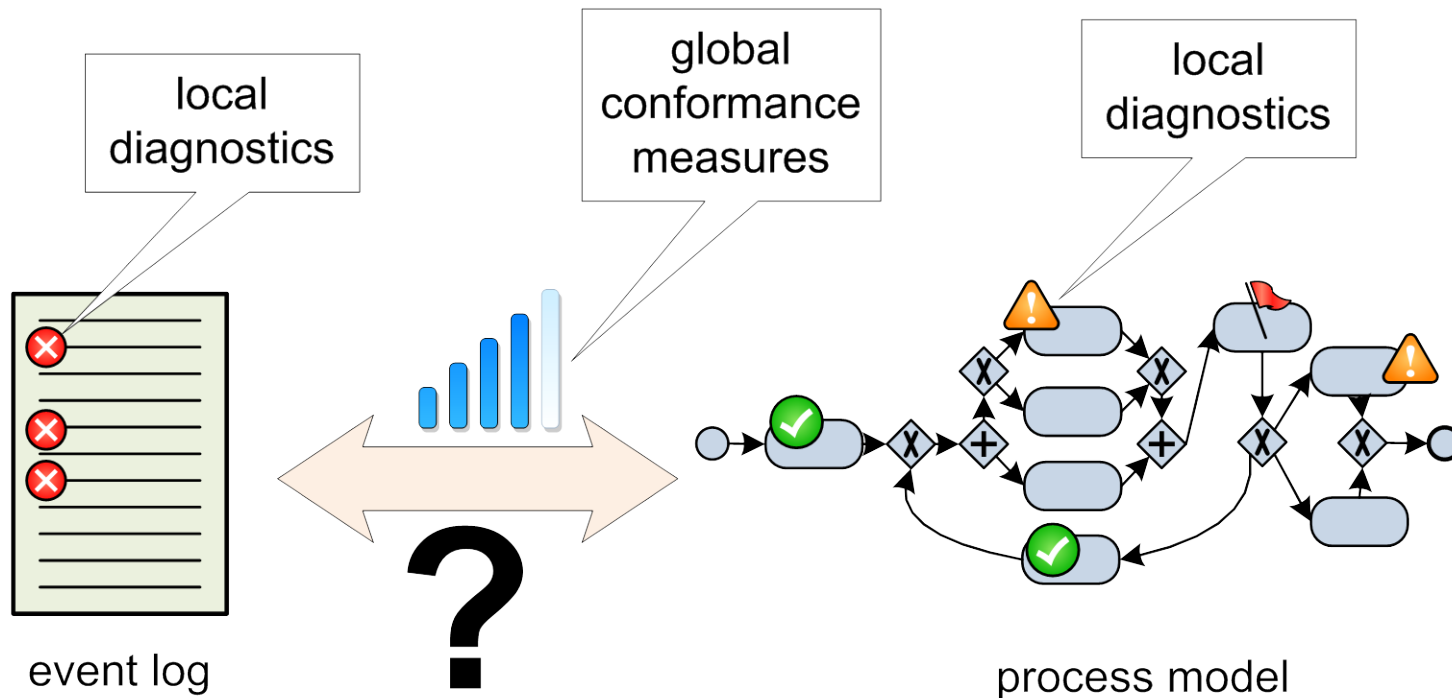
- **Corporate governance, risk, compliance, and legislation such as the Sarbanes-Oxley (US), Basel II/III (EU), J-SOX (Japan), C-SOX (Canada), 8th EU Directive (EURO-SOX), BiMoG (Germany), MiFID (EU), Law 262/05 (Italy), Code Lippens (Belgium), and Code Tabaksblad (Netherlands).**
- **ISO 9001:2008 requires organizations to model their operational processes.**
- **Business alignment: make sure that the information systems and the real business processes are well aligned.**

# Auditing

- **The term auditing refers to the evaluation of organizations and their processes.**
- **Audits are performed to ascertain the validity and reliability of information about these organizations and associated processes.**
- **This is done to check whether business processes are executed within certain boundaries set by managers, governments, and other stakeholders.**
- **Obviously, process mining can help to detect fraud, malpractice, risks, and inefficiencies.**
- **All events in a business process can be evaluated and this can be done while the process is still running.**

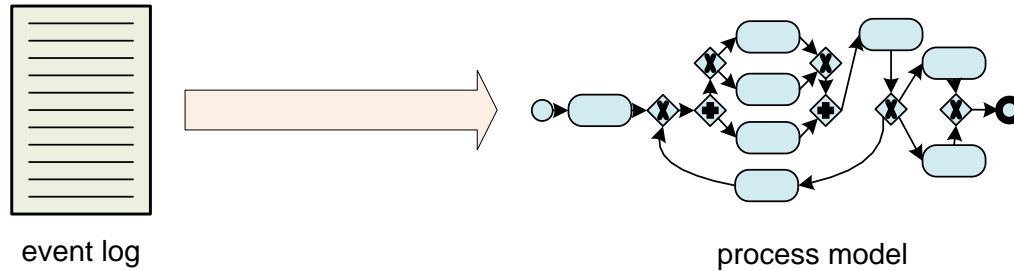
# Deviations?

- Is the model or the log “wrong”?
- “Desirable” or “undesirable” deviations?
- “Breaking the glass” may save lives!

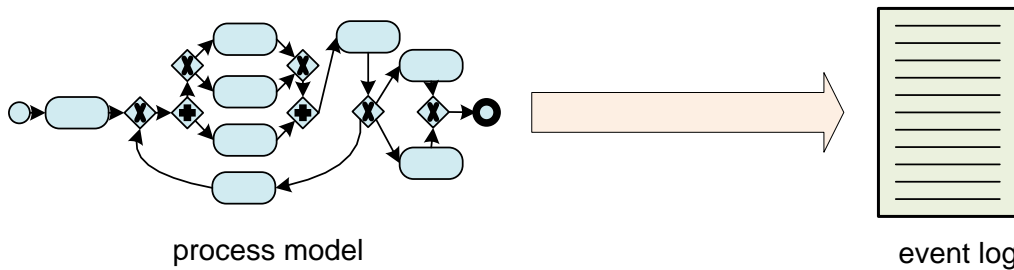


# Replay: Connecting events to model elements is essential for process mining

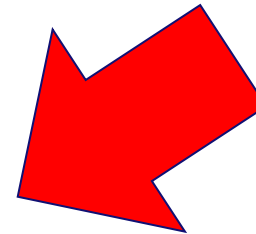
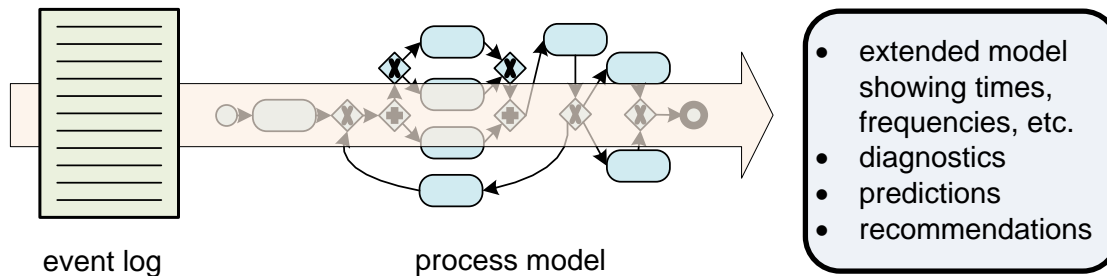
## Play-In



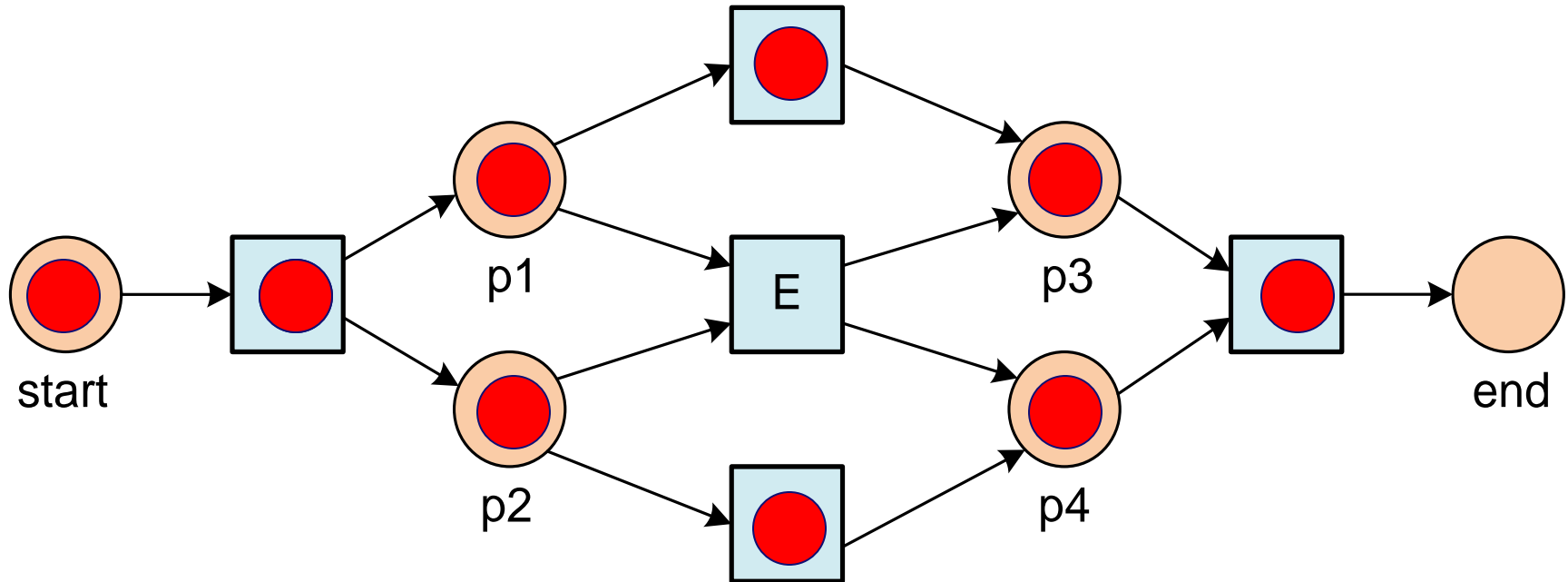
## Play-Out



## Replay



# Play Out (Classical use of models)



**A B C D**    **A E D**    **A E D**

**A B C D**

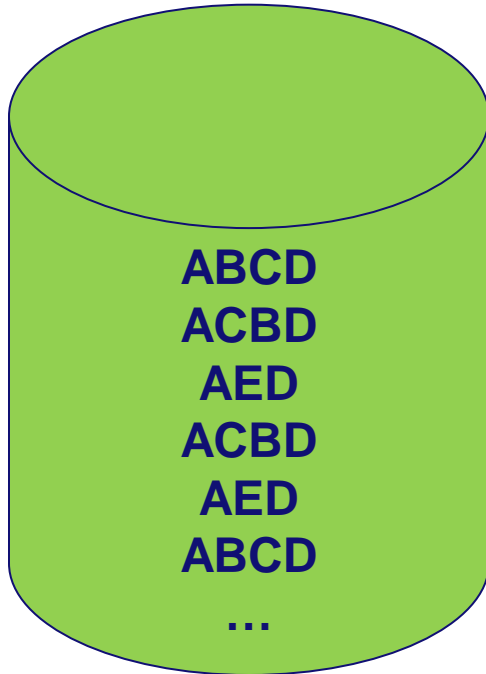
**A C B D**

**A C B D**

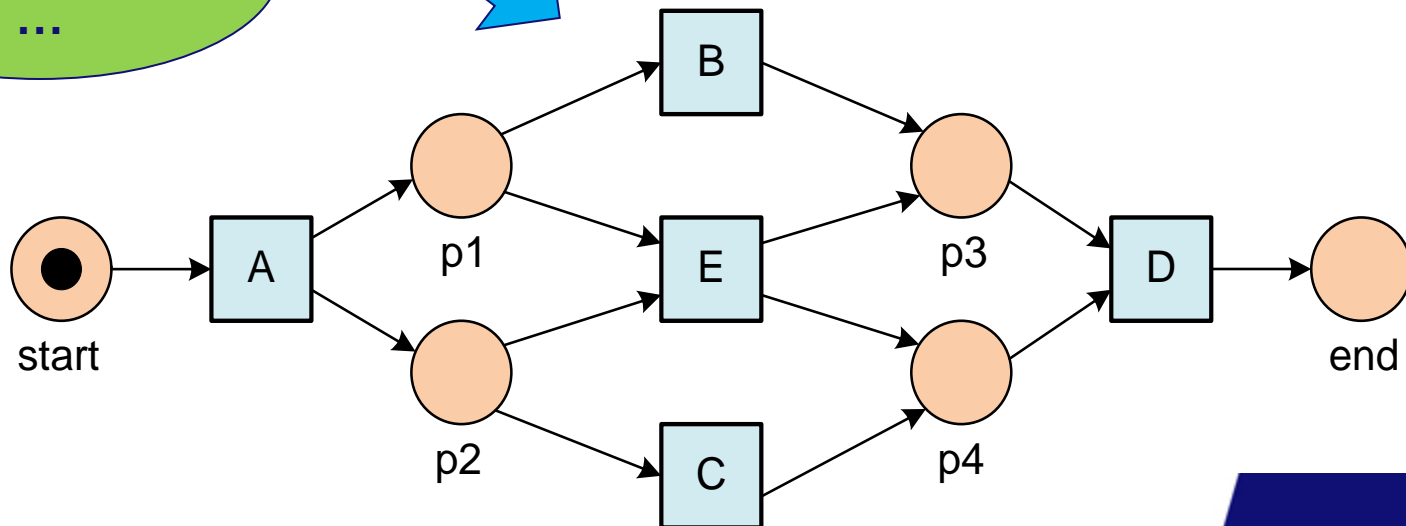
**A E D**

**A C B D**

# Play In (Process Discovery)

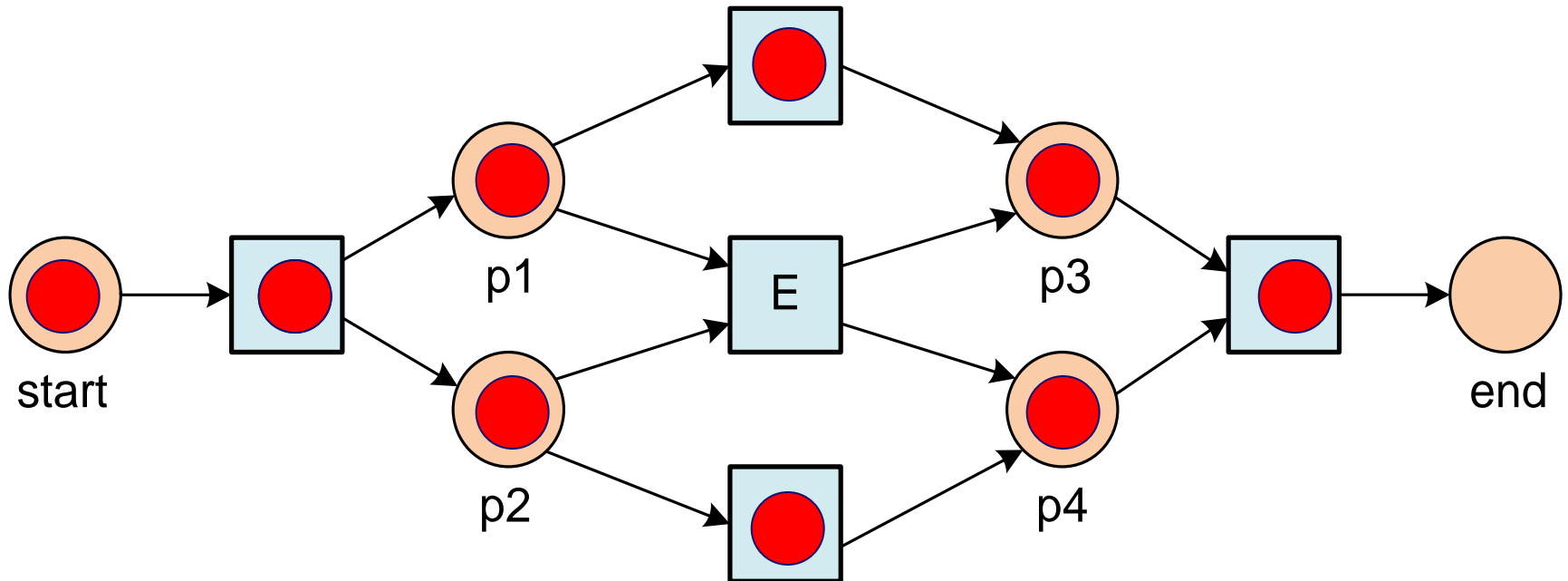


a process discovery algorithm like the  $\alpha$  algorithm



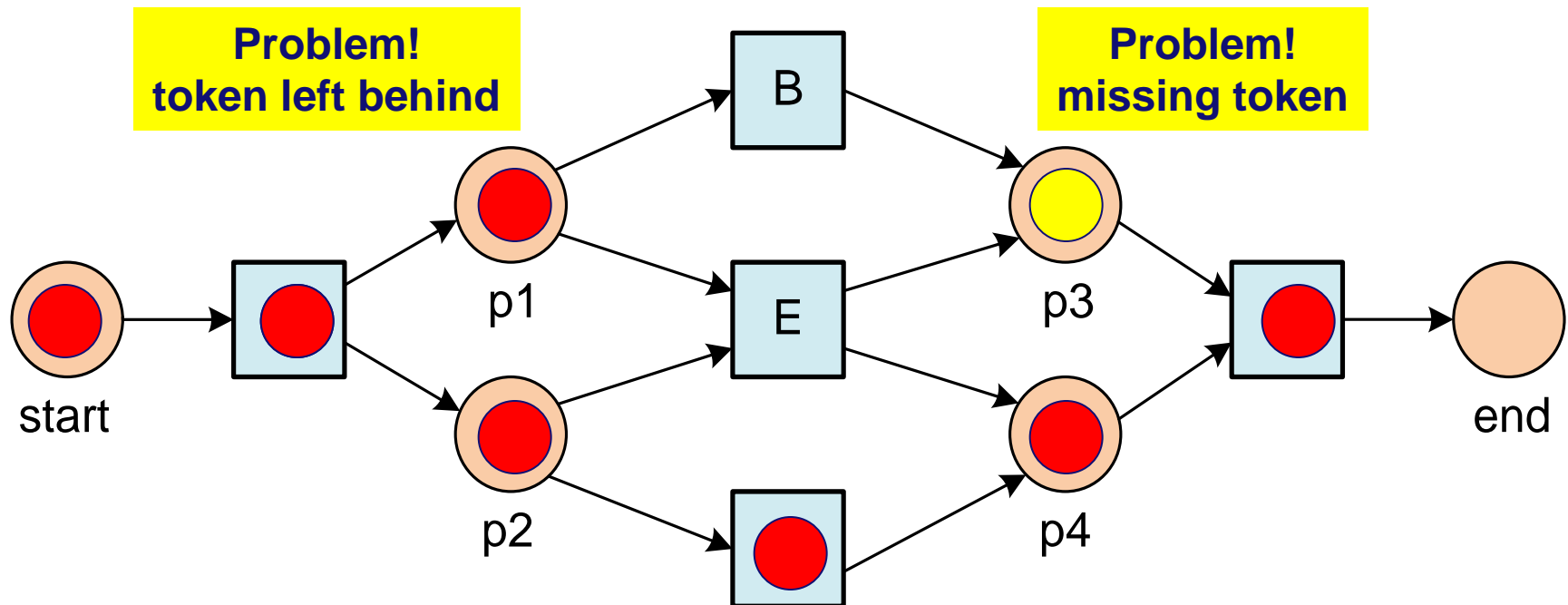
# Replay

**A B C D**



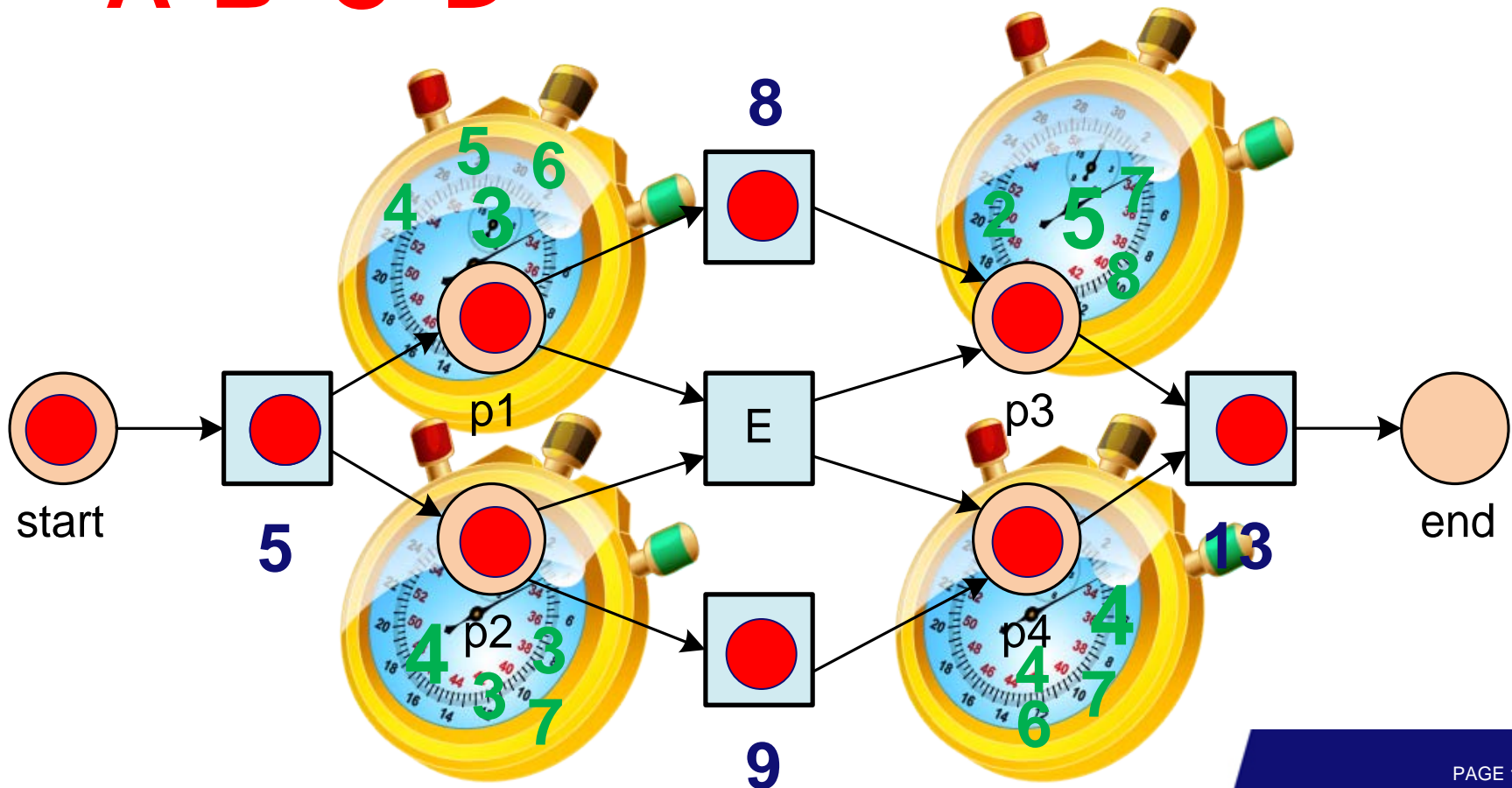
# Replay can detect problems

**ACD**



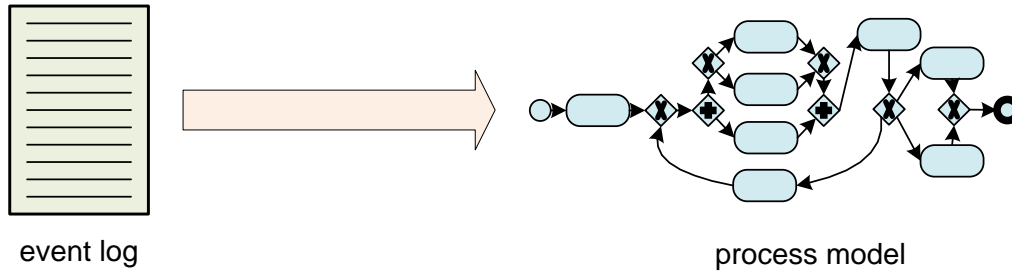
# Replay can extract timing information

**A<sup>5</sup>B<sup>8</sup>C<sup>9</sup>D<sup>13</sup>**

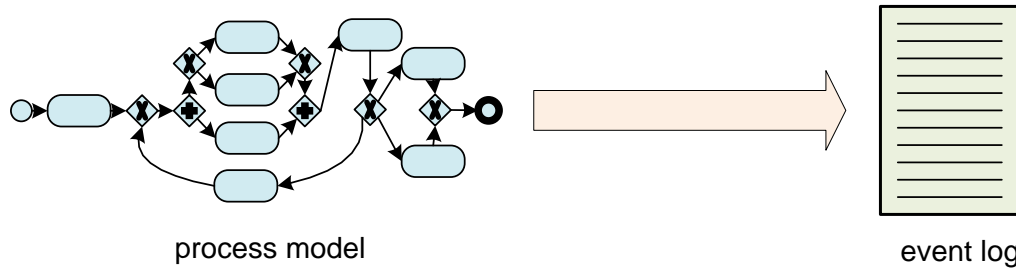


# Let us now focus on conformance checking based on Replay

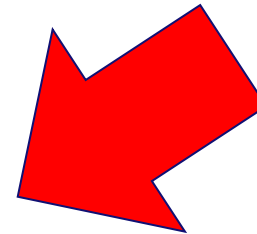
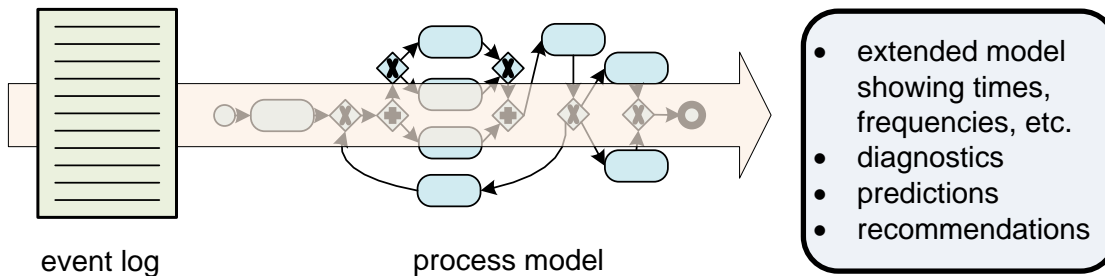
## Play-In



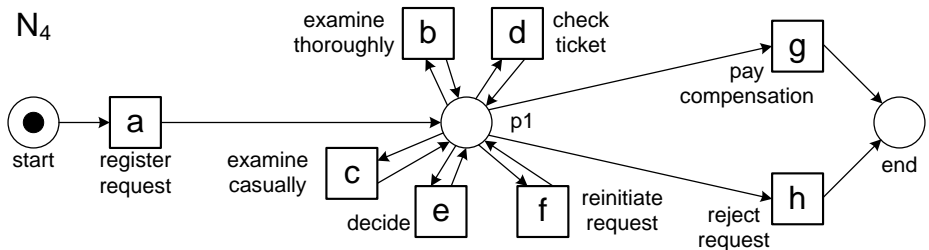
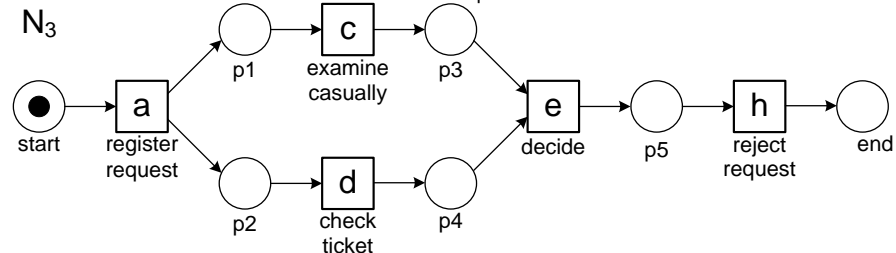
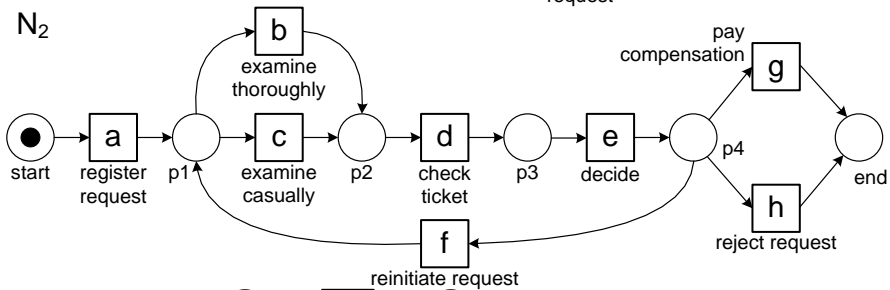
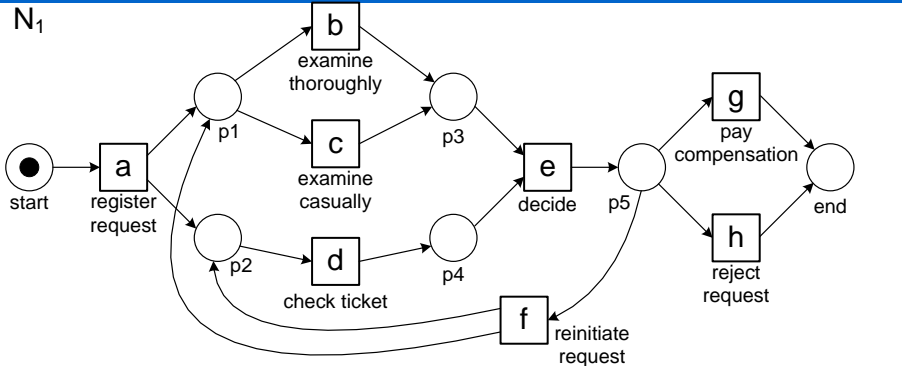
## Play-Out



## Replay



# Four models, one log



## frequency reference trace

455	$\sigma_1$	$\langle a, c, d, e, h \rangle$
191	$\sigma_2$	$\langle a, b, d, e, g \rangle$
177	$\sigma_3$	$\langle a, d, c, e, h \rangle$
144	$\sigma_4$	$\langle a, b, d, e, h \rangle$
111	$\sigma_5$	$\langle a, c, d, e, g \rangle$
82	$\sigma_6$	$\langle a, d, c, e, g \rangle$
56	$\sigma_7$	$\langle a, d, b, e, h \rangle$
47	$\sigma_8$	$\langle a, c, d, e, f, d, b, e, h \rangle$
38	$\sigma_9$	$\langle a, d, b, e, g \rangle$
33	$\sigma_{10}$	$\langle a, c, d, e, f, b, d, e, h \rangle$
14	$\sigma_{11}$	$\langle a, c, d, e, f, b, d, e, g \rangle$
11	$\sigma_{12}$	$\langle a, c, d, e, f, d, b, e, g \rangle$
9	$\sigma_{13}$	$\langle a, d, c, e, f, c, d, e, h \rangle$
8	$\sigma_{14}$	$\langle a, d, c, e, f, d, b, e, h \rangle$
5	$\sigma_{15}$	$\langle a, d, c, e, f, b, d, e, g \rangle$
3	$\sigma_{16}$	$\langle a, c, d, e, f, b, d, e, f, d, b, e, g \rangle$
2	$\sigma_{17}$	$\langle a, d, c, e, f, d, b, e, g \rangle$
2	$\sigma_{18}$	$\langle a, d, c, e, f, b, d, e, f, b, d, e, g \rangle$
1	$\sigma_{19}$	$\langle a, d, c, e, f, d, b, e, f, b, d, e, h \rangle$
1	$\sigma_{20}$	$\langle a, d, b, e, f, b, d, e, f, d, b, e, g \rangle$
1	$\sigma_{21}$	$\langle a, d, c, e, f, d, b, e, f, c, d, e, f, d, b, e, g \rangle$

---

frequency reference trace

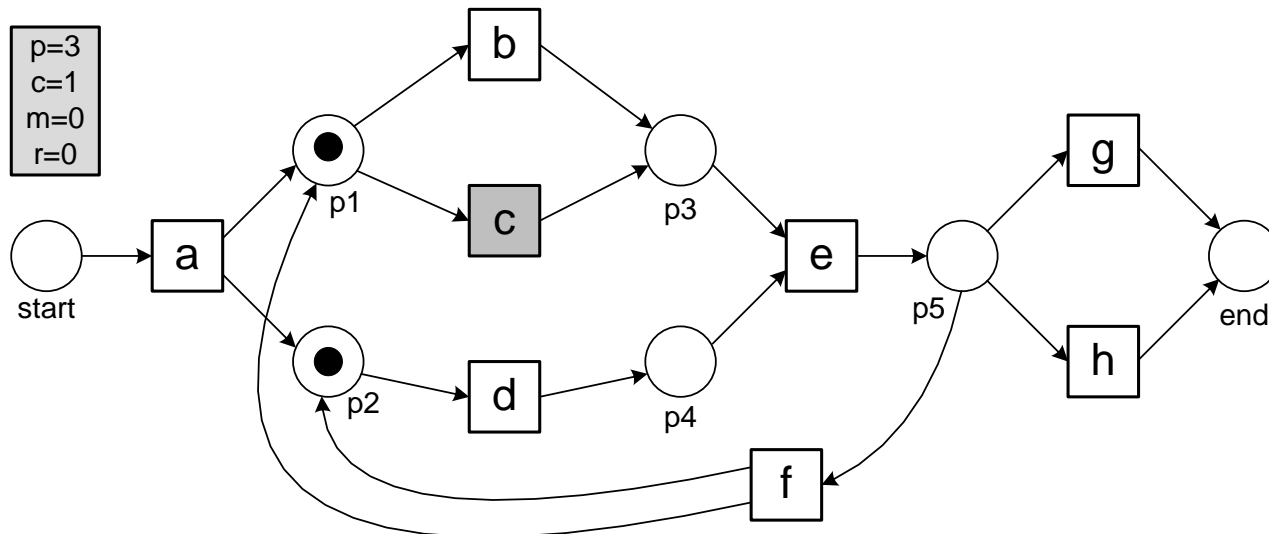
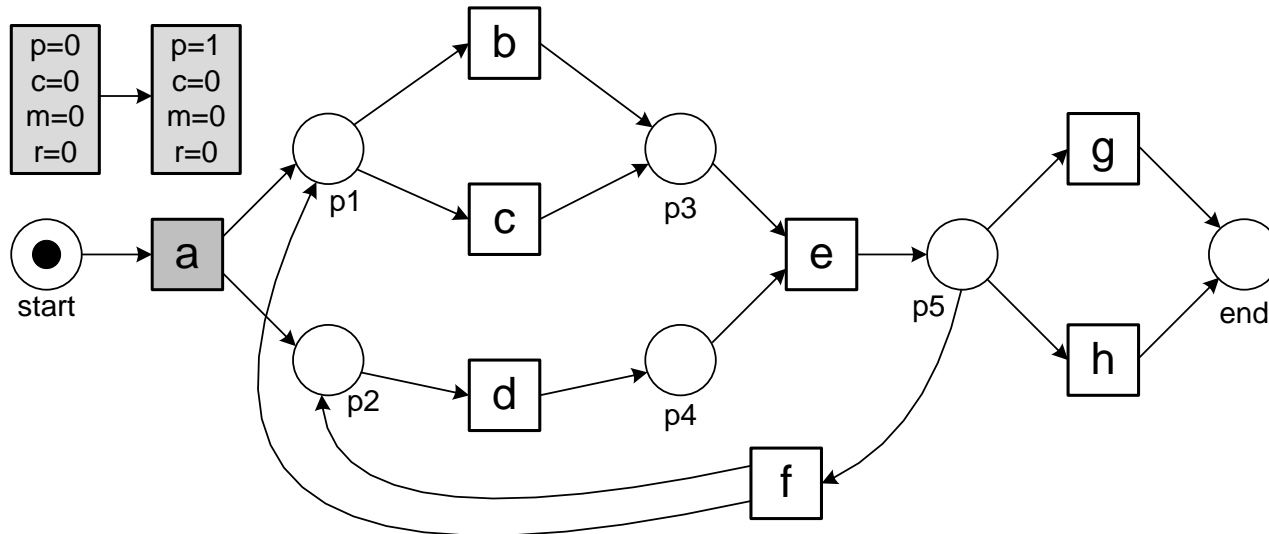
---

455	$\sigma_1$	$\langle a, c, d, e, h \rangle$
191	$\sigma_2$	$\langle a, b, d, e, g \rangle$
177	$\sigma_3$	$\langle a, d, c, e, h \rangle$
144	$\sigma_4$	$\langle a, b, d, e, h \rangle$
111	$\sigma_5$	$\langle a, c, d, e, g \rangle$
82	$\sigma_6$	$\langle a, d, c, e, g \rangle$
56	$\sigma_7$	$\langle a, d, b, e, h \rangle$
47	$\sigma_8$	$\langle a, c, d, e, f, d, b, e, h \rangle$
38	$\sigma_9$	$\langle a, d, b, e, g \rangle$
33	$\sigma_{10}$	$\langle a, c, d, e, f, b, d, e, h \rangle$
14	$\sigma_{11}$	$\langle a, c, d, e, f, b, d, e, g \rangle$
11	$\sigma_{12}$	$\langle a, c, d, e, f, d, b, e, g \rangle$
9	$\sigma_{13}$	$\langle a, d, c, e, f, c, d, e, h \rangle$
8	$\sigma_{14}$	$\langle a, d, c, e, f, d, b, e, h \rangle$
5		$\langle a, d, c, e, f, d, b, e, h \rangle$

# Replaying (1/3)

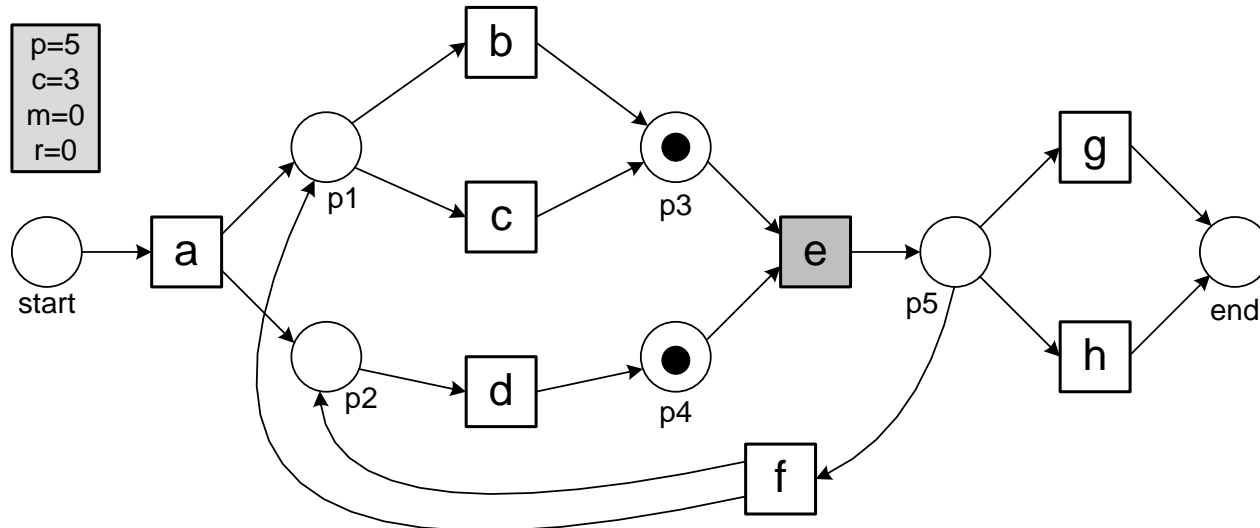
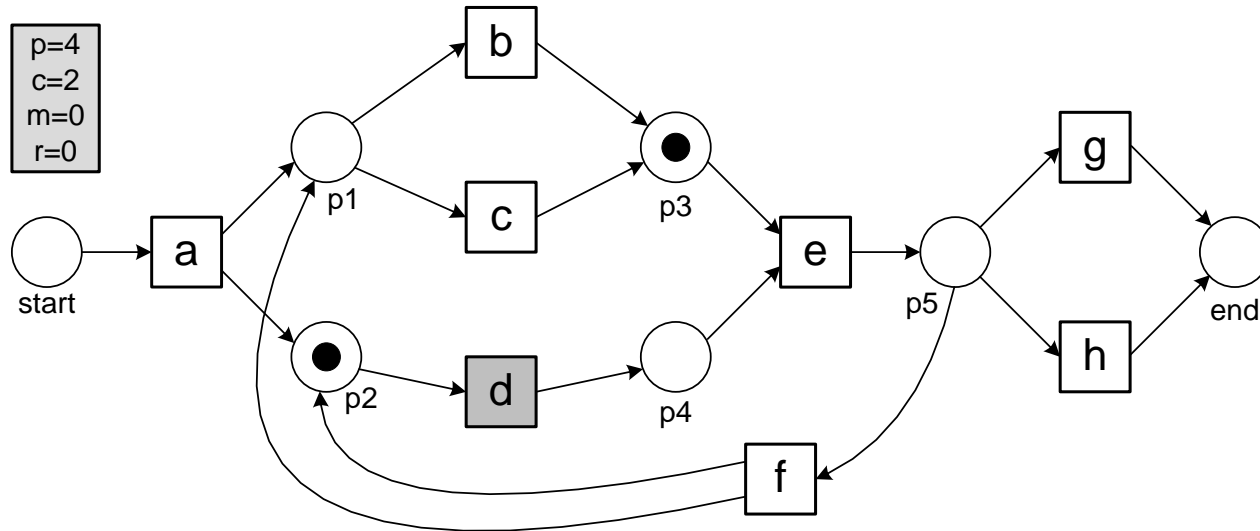
$\sigma_1$  on  $N_1$

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



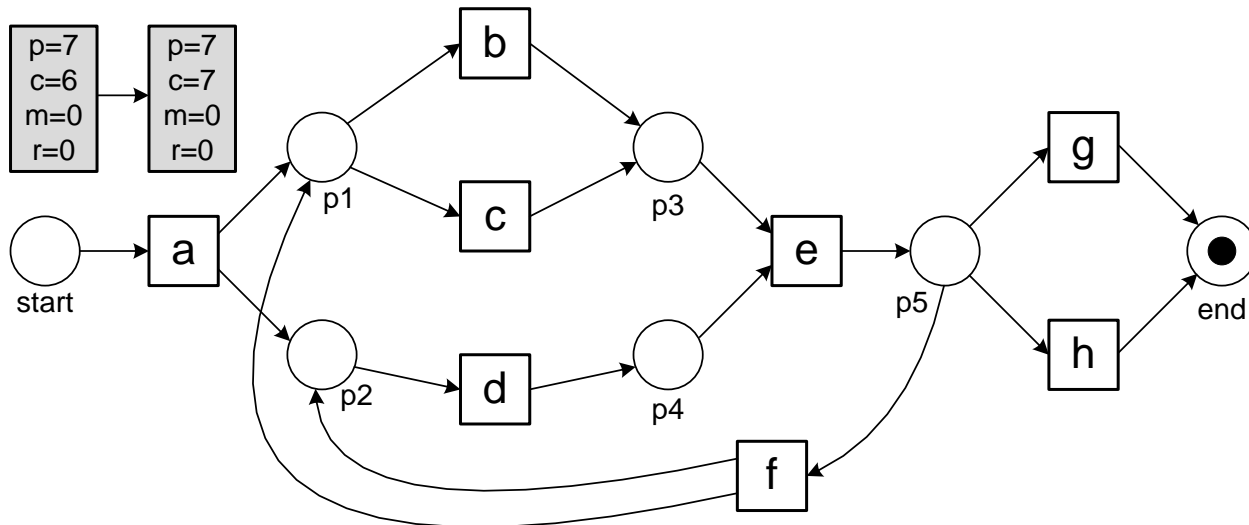
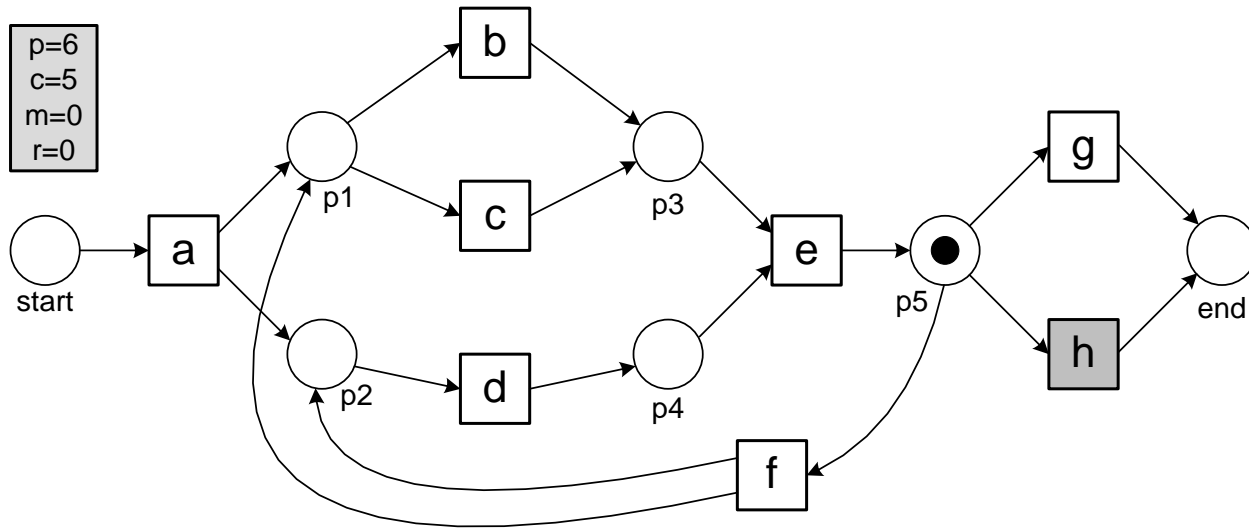
# Replaying (2/3)

$$\sigma_1 = \langle a, c, d, e, h \rangle$$



# Replaying (3/3)

$$\sigma_1 = \langle a, c, d, e, h \rangle$$

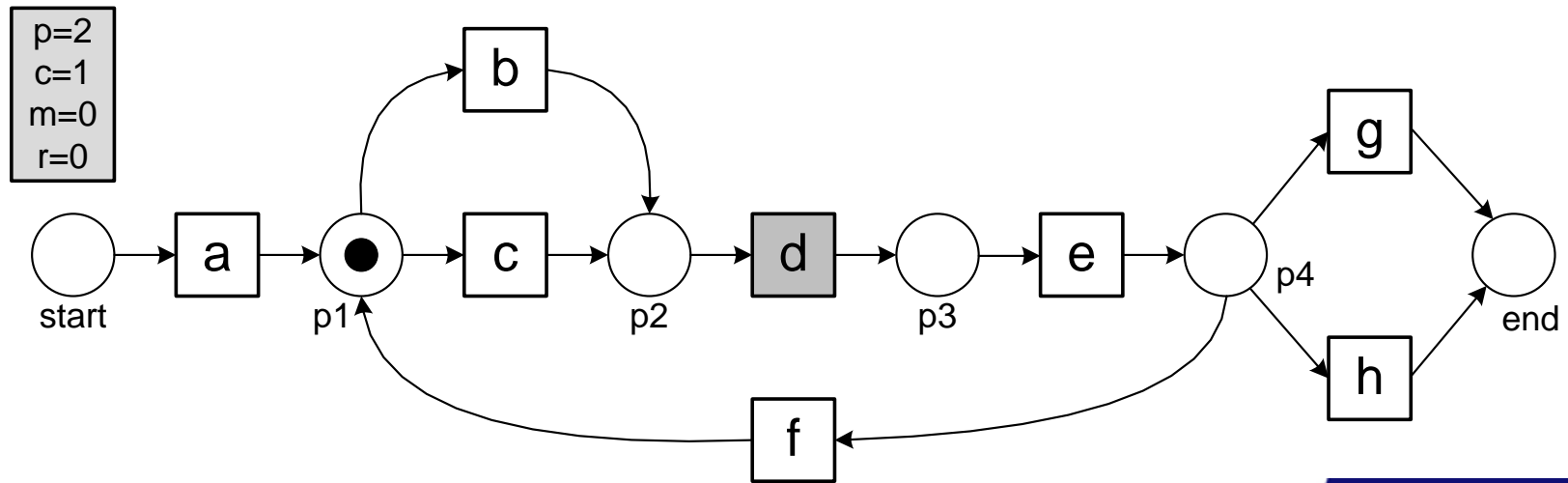
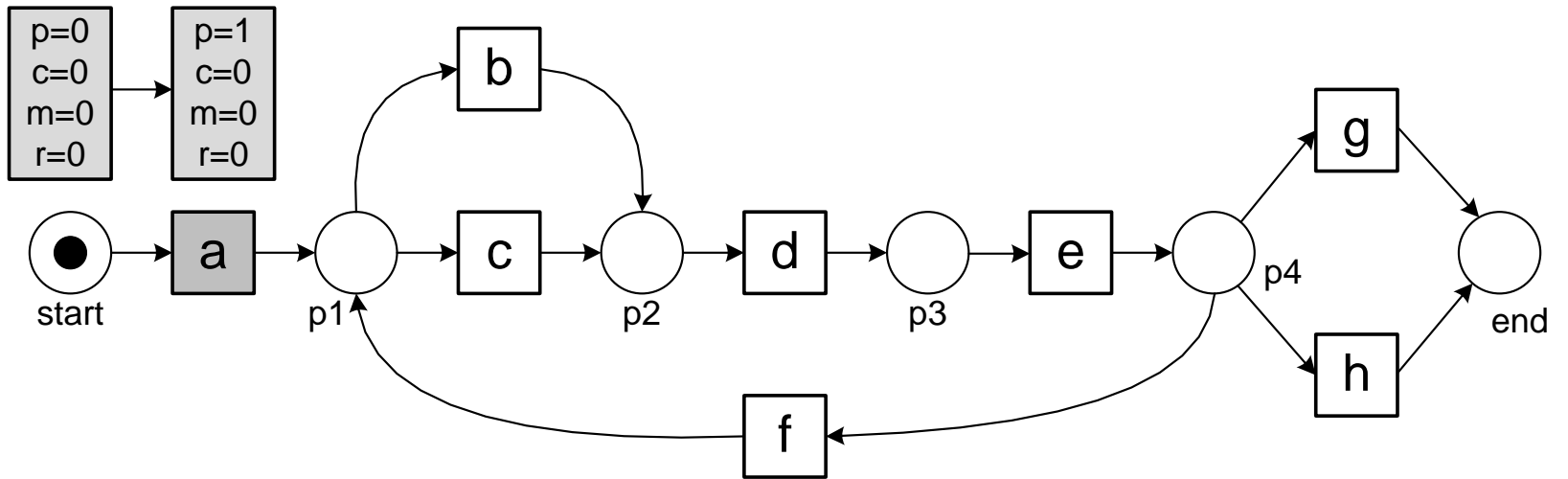


**No problems found!**

# Replaying (1/3)

$\sigma_3$  on  $N_2$

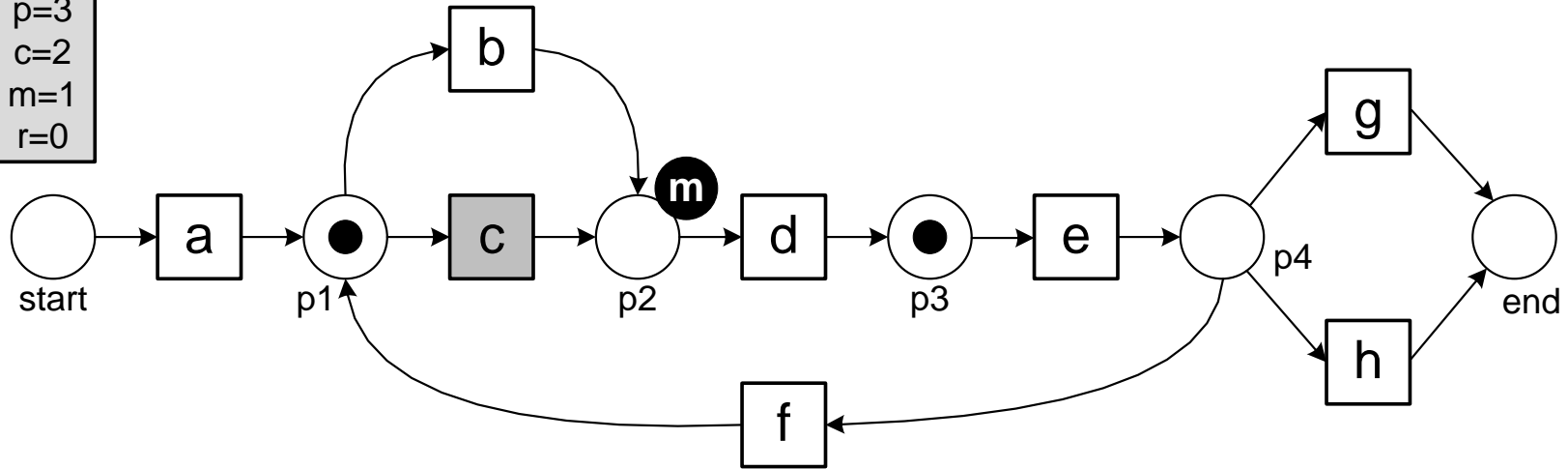
$$\sigma_3 = \langle a, d, c, e, h \rangle$$



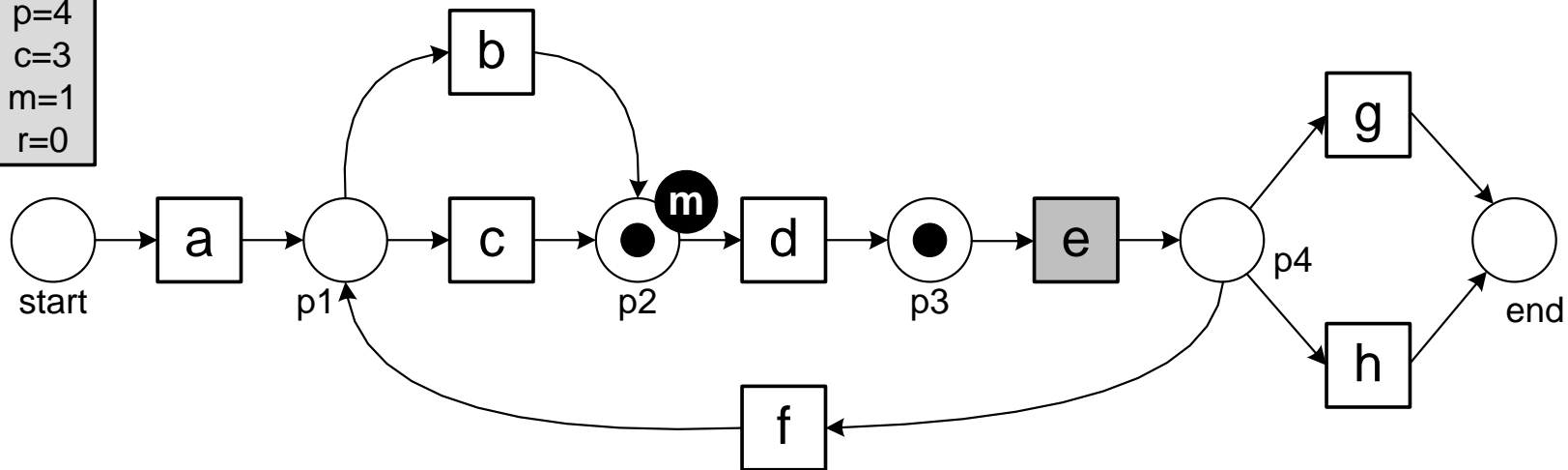
# Replaying (2/3)

$$\sigma_3 = \langle a, d, c, e, h \rangle$$

p=3  
c=2  
m=1  
r=0

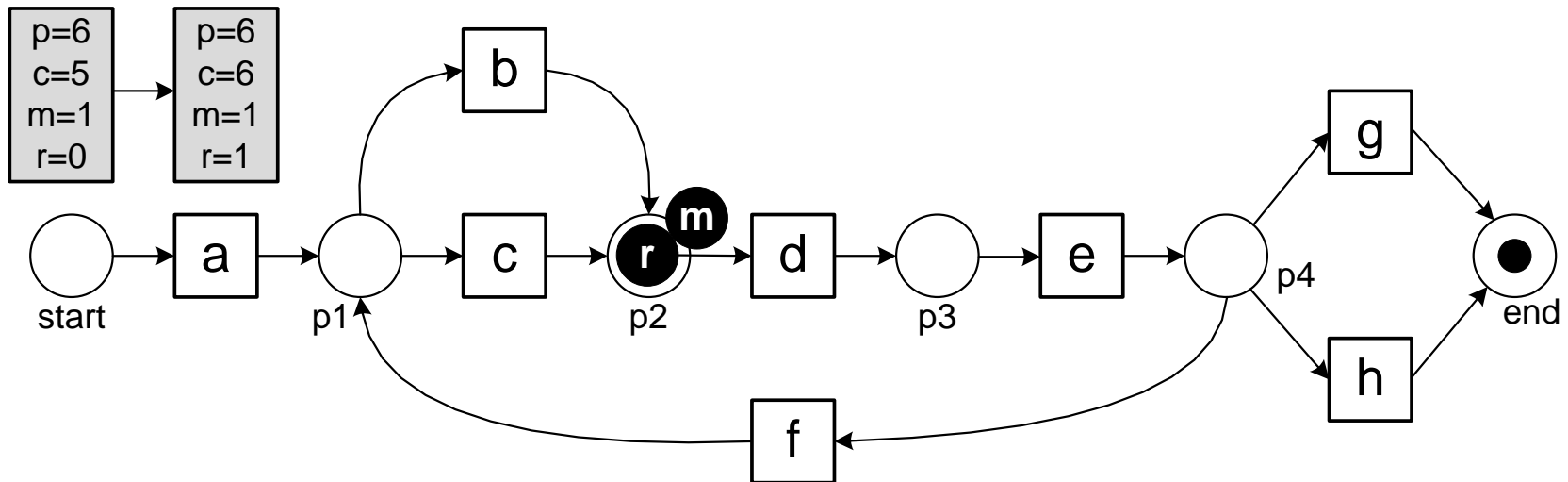
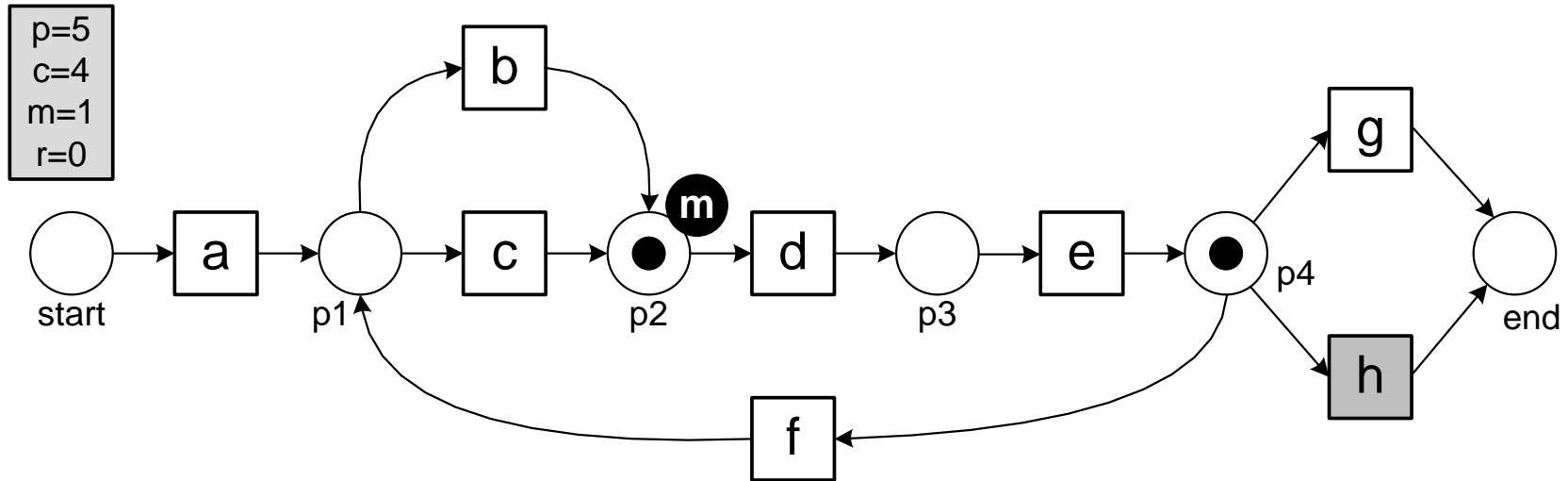


p=4  
c=3  
m=1  
r=0



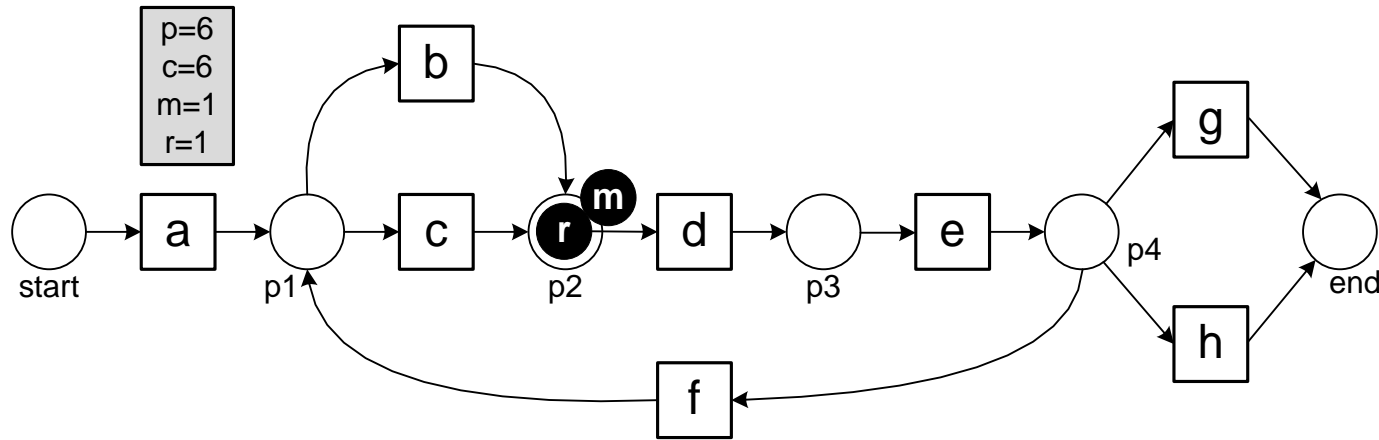
# Replaying (3/3)

$$\sigma_3 = \langle a, d, c, e, h \rangle$$



# Problems encountered when replaying $\sigma_3$ on $N_2$

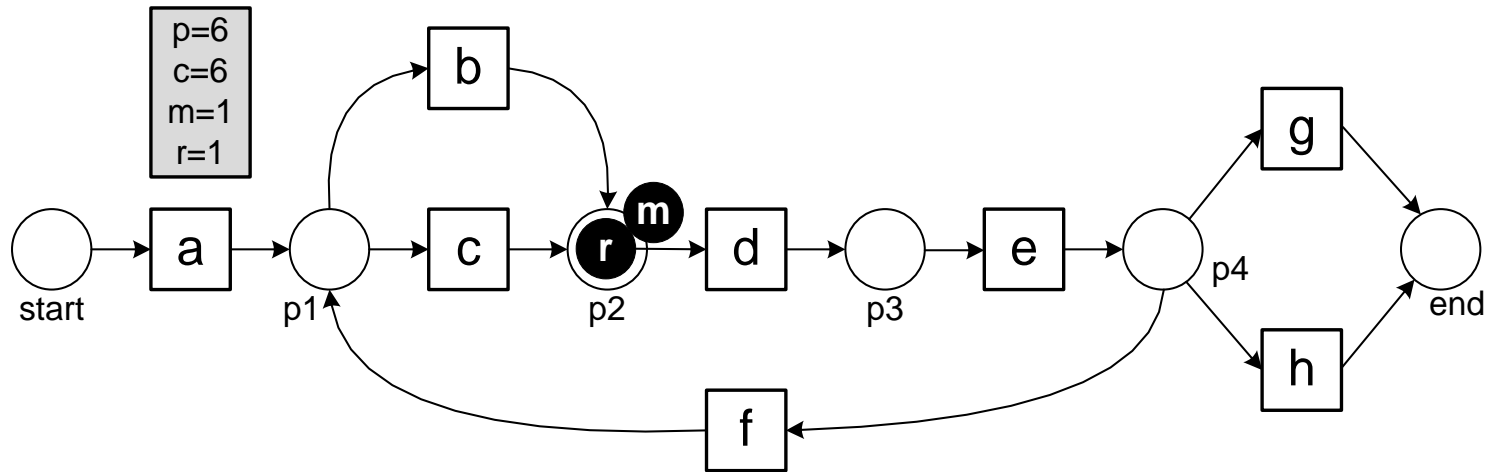
$$\sigma_3 = \langle a, d, c, e, h \rangle$$



- One missing token (of 6 consumed tokens)
- One remaining token (of 6 produced tokens)

$$fitness(\sigma, N) = \frac{1}{2} \left( 1 - \frac{m}{c} \right) + \frac{1}{2} \left( 1 - \frac{r}{p} \right)$$

# Computing fitness at trace level



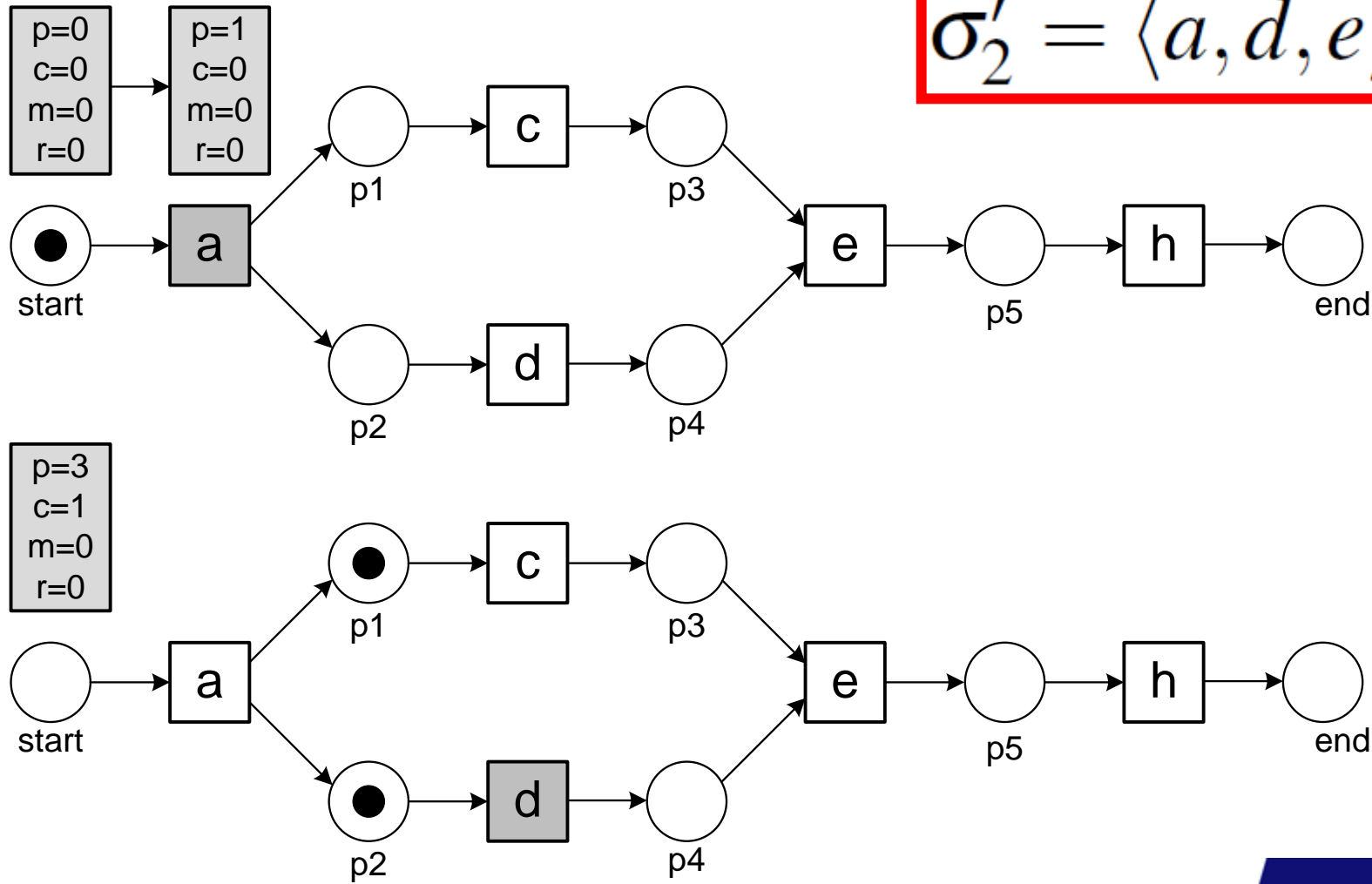
$$fitness(\sigma_3, N_2) = \frac{1}{2} \left( 1 - \frac{1}{6} \right) + \frac{1}{2} \left( 1 - \frac{1}{6} \right) = 0.8333$$

# Replaying (1/3)

## $\sigma_2$ on $N_3$

$$\sigma_2 = \langle a, b, d, e, g \rangle$$

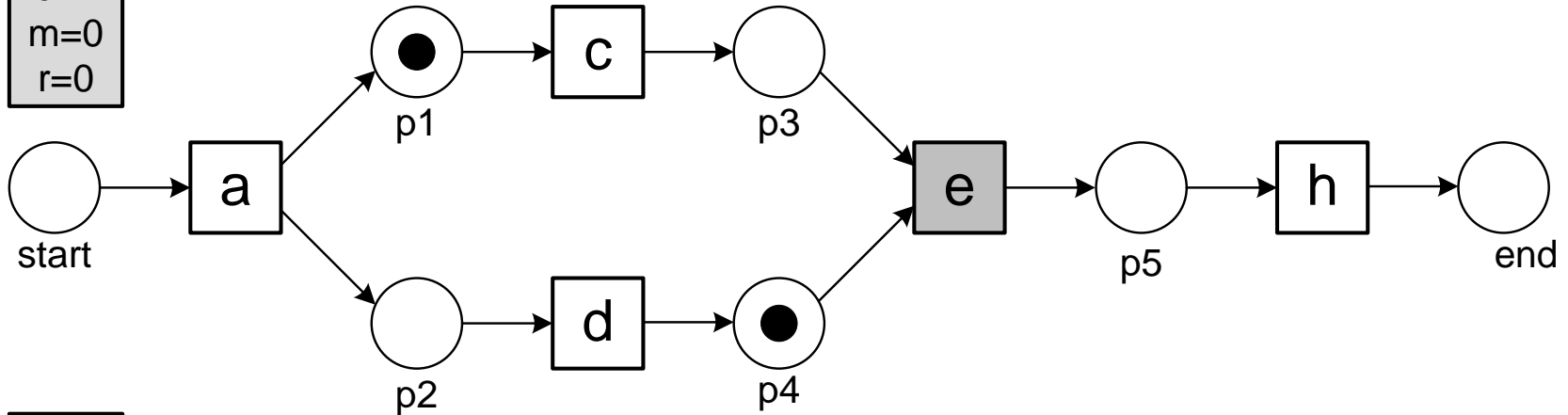
$$\sigma'_2 = \langle a, d, e \rangle$$



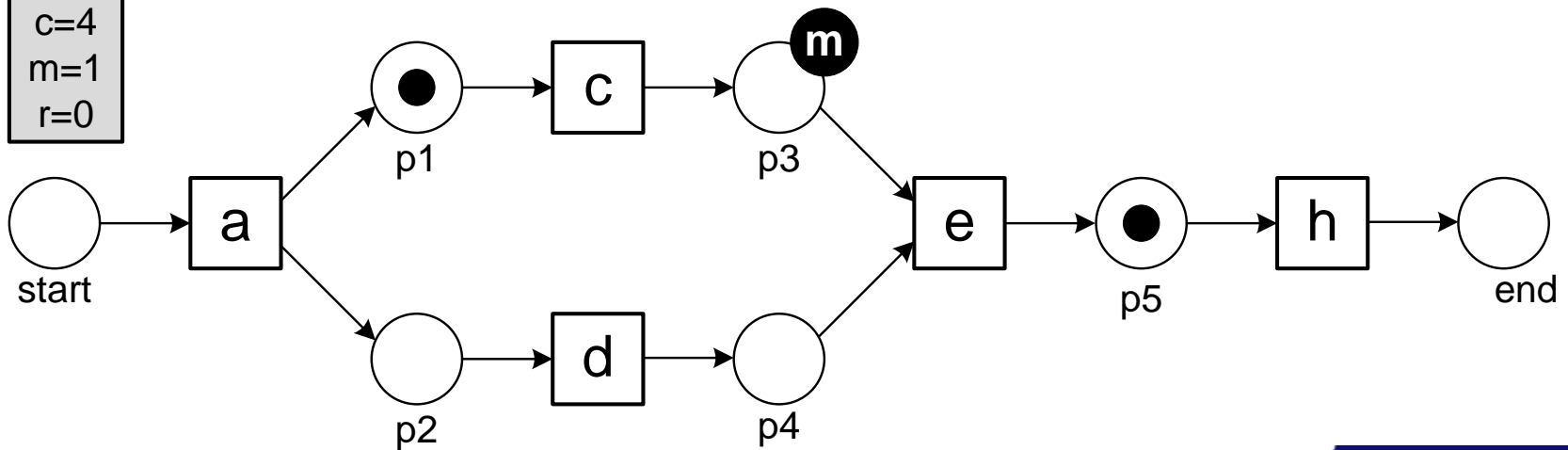
# Replaying (2/3)

$$\sigma'_2 = \langle a, d, e \rangle$$

p=4  
c=2  
m=0  
r=0

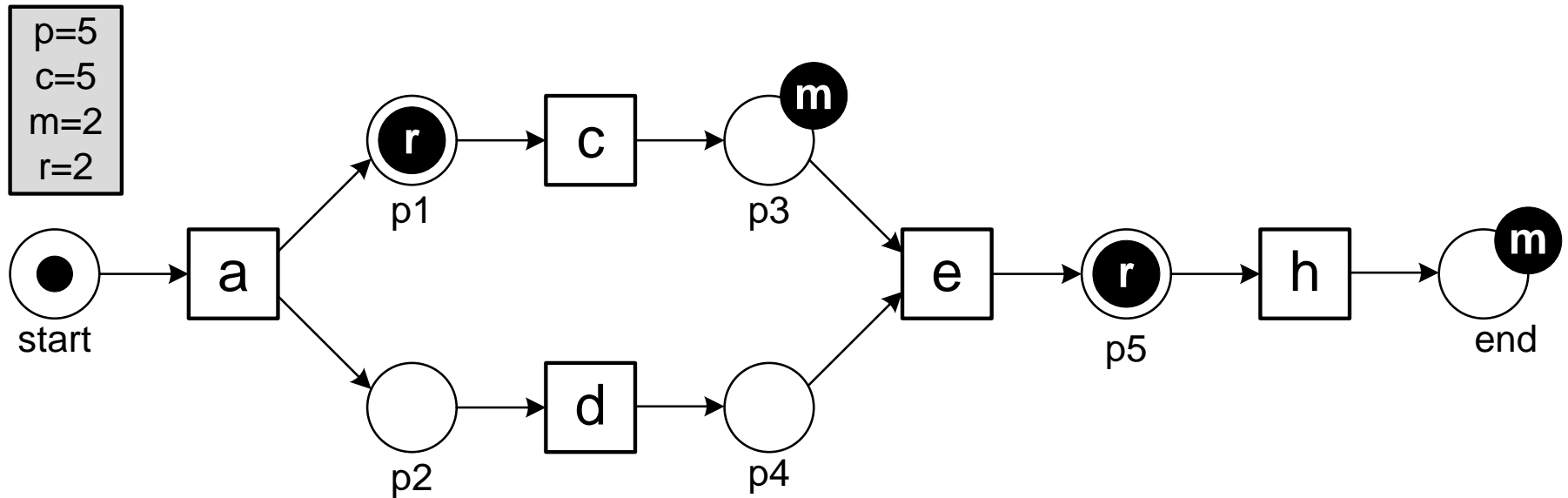


p=5  
c=4  
m=1  
r=0



# Replaying (3/3)

$$\sigma'_2 = \langle a, d, e \rangle$$



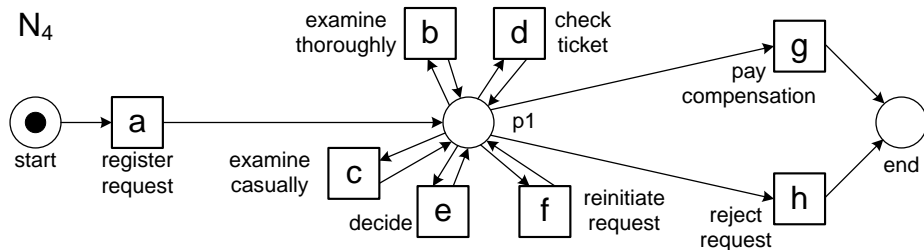
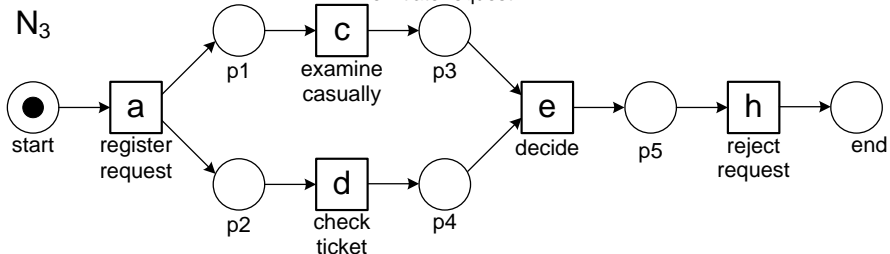
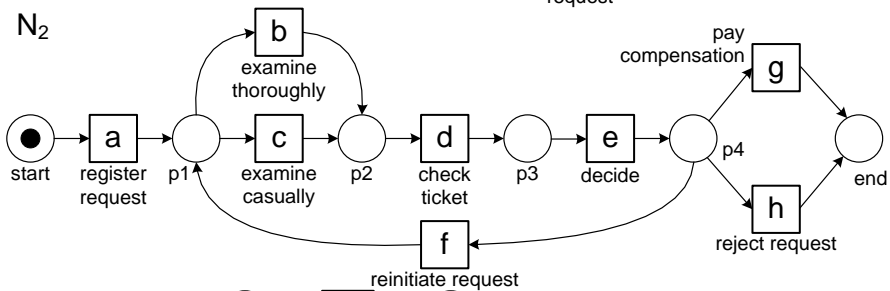
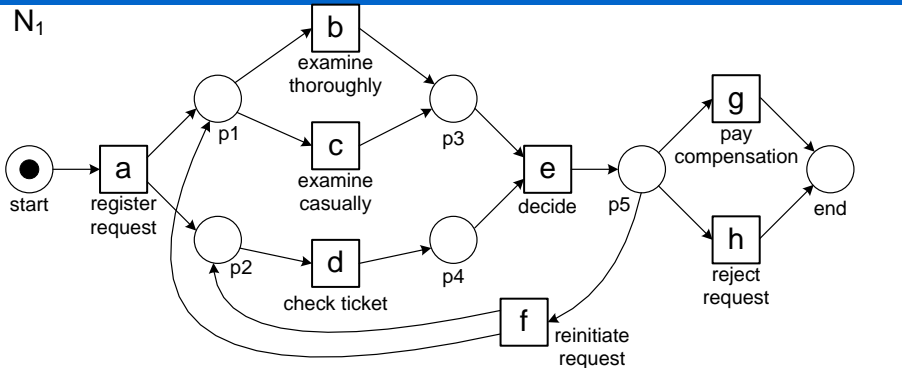
**$p = 5$ ,  $c = 5$ ,  $m = 2$ , and  $r = 2$**

$$fitness(\sigma_2, N_3) = \frac{1}{2} \left( 1 - \frac{2}{5} \right) + \frac{1}{2} \left( 1 - \frac{2}{5} \right) = 0.6$$

# Computing fitness at the log level

$$\begin{aligned} \text{fitness}(L, N) = & \frac{1}{2} \left( 1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N, \sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N, \sigma}} \right) + \\ & \frac{1}{2} \left( 1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N, \sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N, \sigma}} \right) \end{aligned}$$

# Example values



$$fitness(L_{full}, N_1) = 1$$

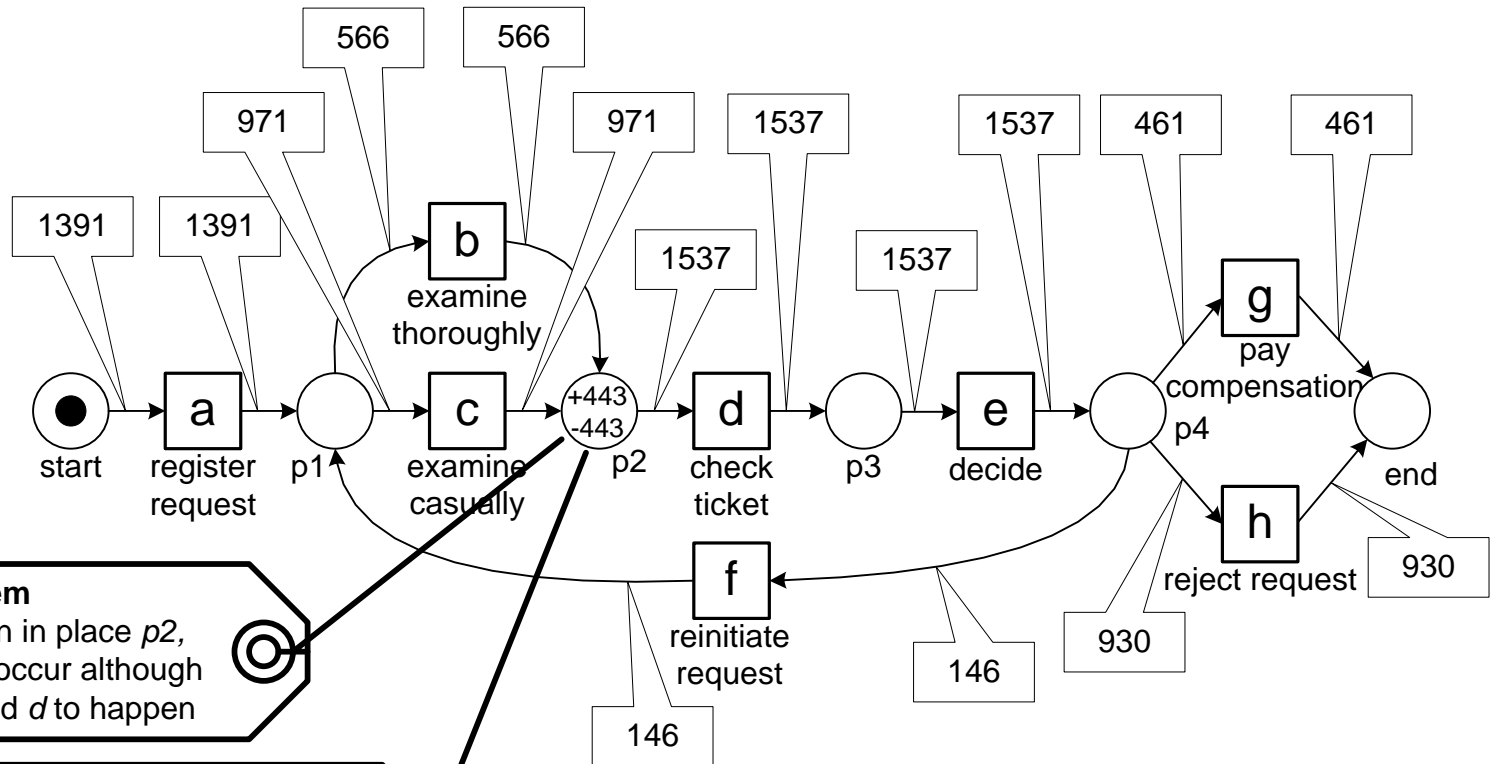
$$fitness(L_{full}, N_2) = 0.9504$$

$$fitness(L_{full}, N_3) = 0.8797$$

$$fitness(L_{full}, N_4) = 1$$

# Diagnostics

$$(fitness(L_{full}, N_2) = 0.9504)$$

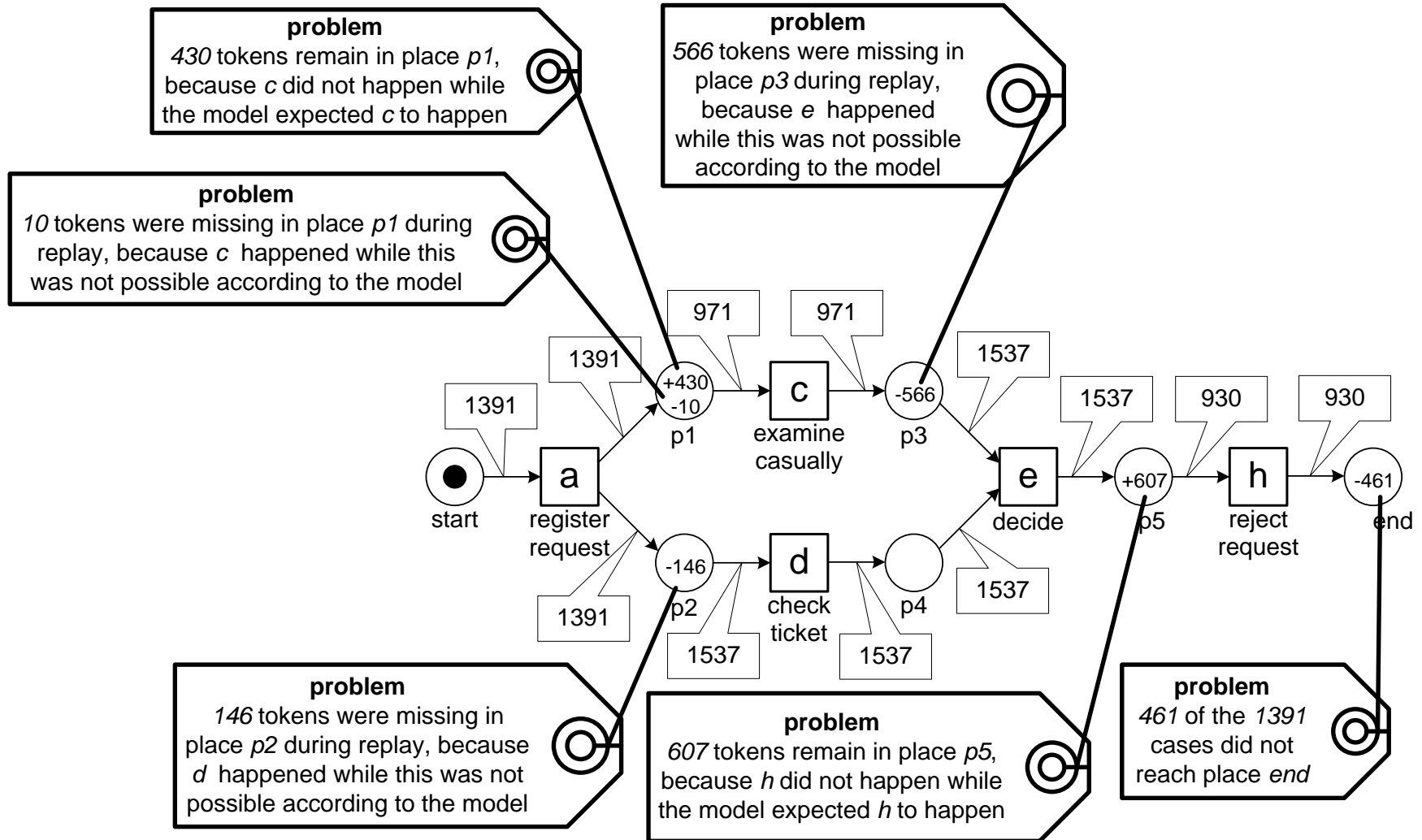


**problem**  
443 tokens remain in place  $p_2$ , because  $d$  did not occur although the model expected  $d$  to happen

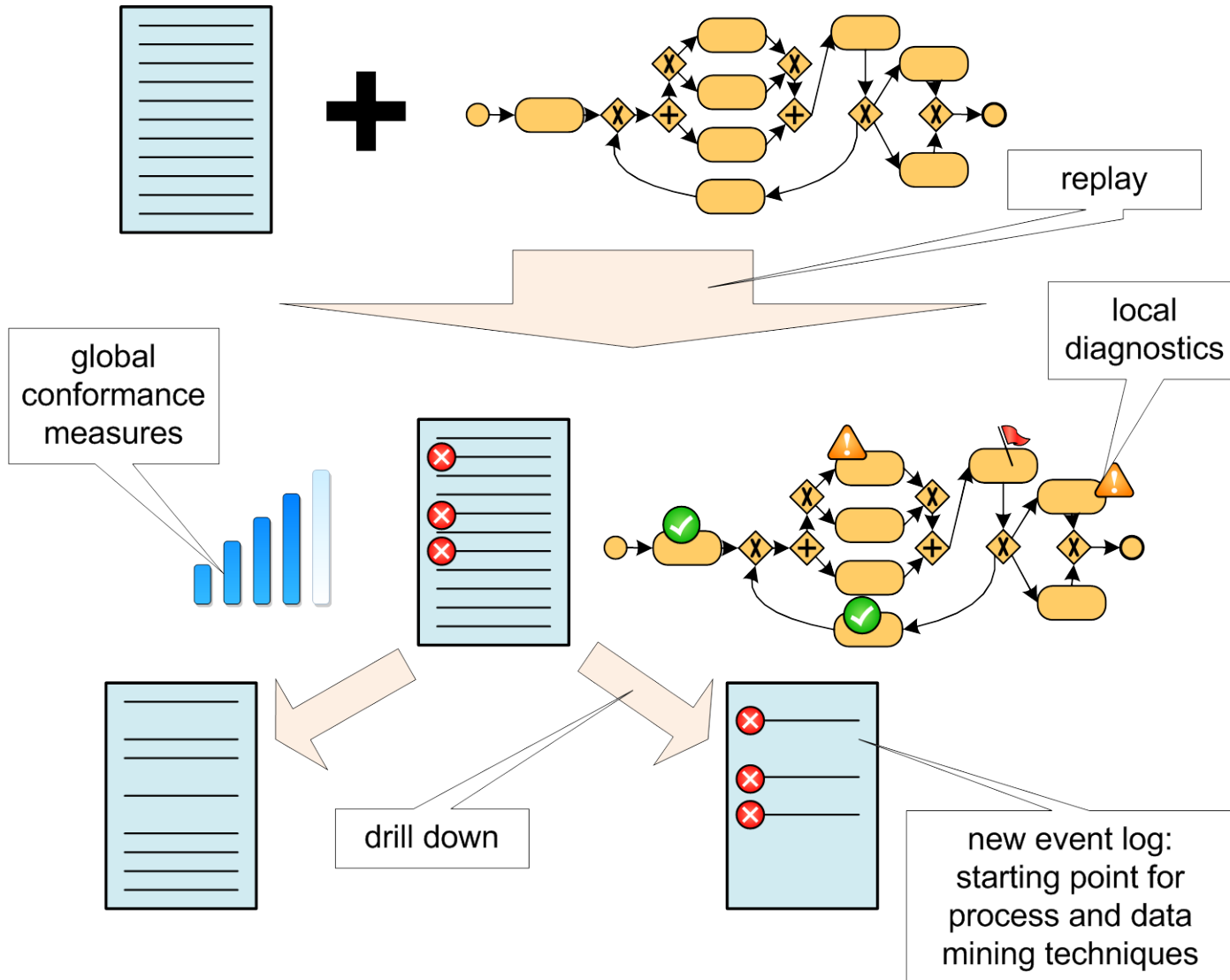
**problem**  
443 tokens were missing in place  $p_2$  during replay, because  $d$  happened even though this was not possible according to the model

# Diagnostics

$$(fitness(L_{full}, N_3) = 0.8797)$$



# Drilling down



# Comparing footprints

$L_{full}$  and  $N_1$

	$a$	$b$	$c$	$d$	$e$	$f$	$g$	$h$
$a$	#	→	→	→	#	#	#	#
$b$	←	#	#		→	←	#	#
$c$	←	#	#		→	←	#	#
$d$	←			#	→	←	#	#
$e$	#	←	←	←	#	→	→	→
$f$	#	→	→	→	←	#	#	#
$g$	#	#	#	#	←	#	#	#
$h$	#	#	#	#	←	#	#	#

	$a$	$b$	$c$	$d$	$e$	$f$	$g$	$h$
$a$	#	$\rightarrow$	$\rightarrow$	#	#	#	#	#
$b$	$\leftarrow$	#	#	$\rightarrow$	#	$\leftarrow$	#	#
$c$	$\leftarrow$	#	#	$\rightarrow$	#	$\leftarrow$	#	#
$d$	#	$\leftarrow$	$\leftarrow$	#	$\rightarrow$	#	#	#
$e$	#	#	#	$\leftarrow$	#	$\rightarrow$	$\rightarrow$	$\rightarrow$
$f$	#	$\rightarrow$	$\rightarrow$	#	$\leftarrow$	#	#	#
$g$	#	#	#	#	$\leftarrow$	#	#	#
$h$	#	#	#	#	$\leftarrow$	#	#	#

$N_2$

$L_{full}$  and  $N_1$

	$a$	$b$	$c$	$d$	$e$	$f$	$g$	$h$
$a$	#	$\rightarrow$	$\rightarrow$	$\rightarrow$	#	#	#	#
$b$	$\leftarrow$	#	#	$\parallel$	$\rightarrow$	$\leftarrow$	#	#
$c$	$\leftarrow$	#	#	$\parallel$	$\rightarrow$	$\leftarrow$	#	#
$d$	$\leftarrow$	$\parallel$	$\parallel$	#	$\rightarrow$	$\leftarrow$	#	#
$e$	#	$\leftarrow$	$\leftarrow$	$\leftarrow$	#	$\rightarrow$	$\rightarrow$	$\rightarrow$
$f$	#	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\leftarrow$	#	#	#
$g$	#	#	#	#	$\leftarrow$	#	#	#
$h$	#	#	#	#	$\leftarrow$	#	#	#

# Differences quantified

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>				→: #				
<i>b</i>				:→	→: #			
<i>c</i>				:→	→: #			
<i>d</i>	←: #	:←	:←				←: #	
<i>e</i>		←: #	←: #					
<i>f</i>				→: #				
<i>g</i>								
<i>h</i>								

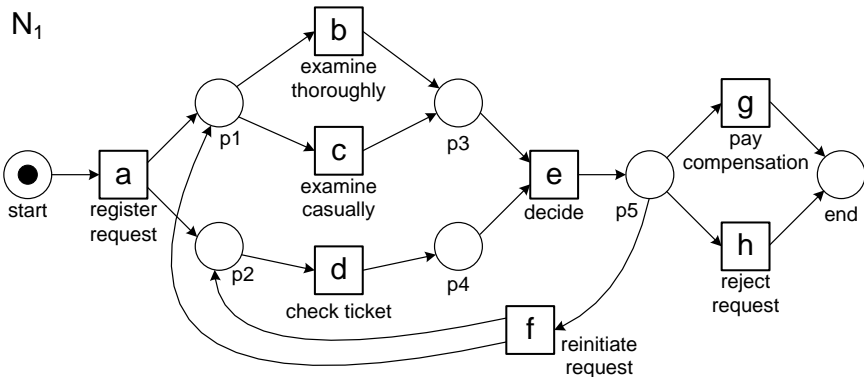
(x:y where x is in log or N<sub>1</sub> and y in N<sub>2</sub>)

$$1 - \frac{12}{64} = 0.8125$$

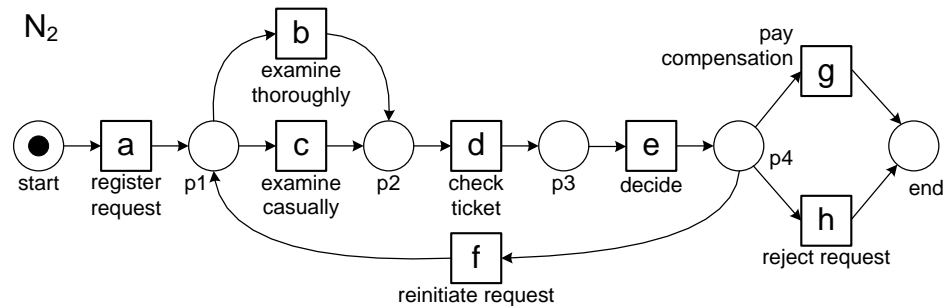
# Diagnostics

(x:y where x is in log or  $N_1$  and y in  $N_2$ )

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>				$\rightarrow: \#$				
<i>b</i>				$\parallel: \rightarrow$	$\rightarrow: \#$			
<i>c</i>				$\parallel: \rightarrow$	$\rightarrow: \#$			
<i>d</i>	$\leftarrow: \#$	$\parallel: \leftarrow$	$\parallel: \leftarrow$			$\leftarrow: \#$		
<i>e</i>		$\leftarrow: \#$	$\leftarrow: \#$					
<i>f</i>				$\rightarrow: \#$				
<i>g</i>								
<i>h</i>								

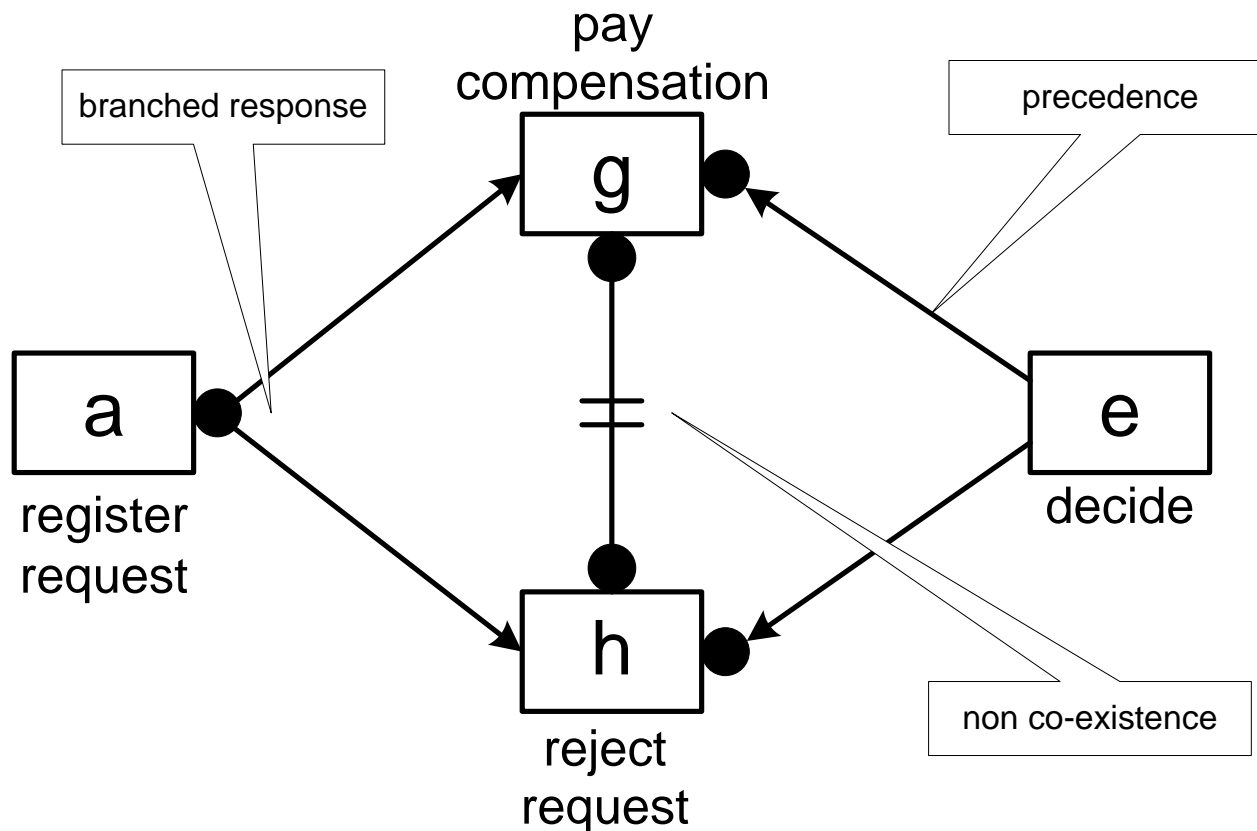


$L_{full}$  and  $N_1$



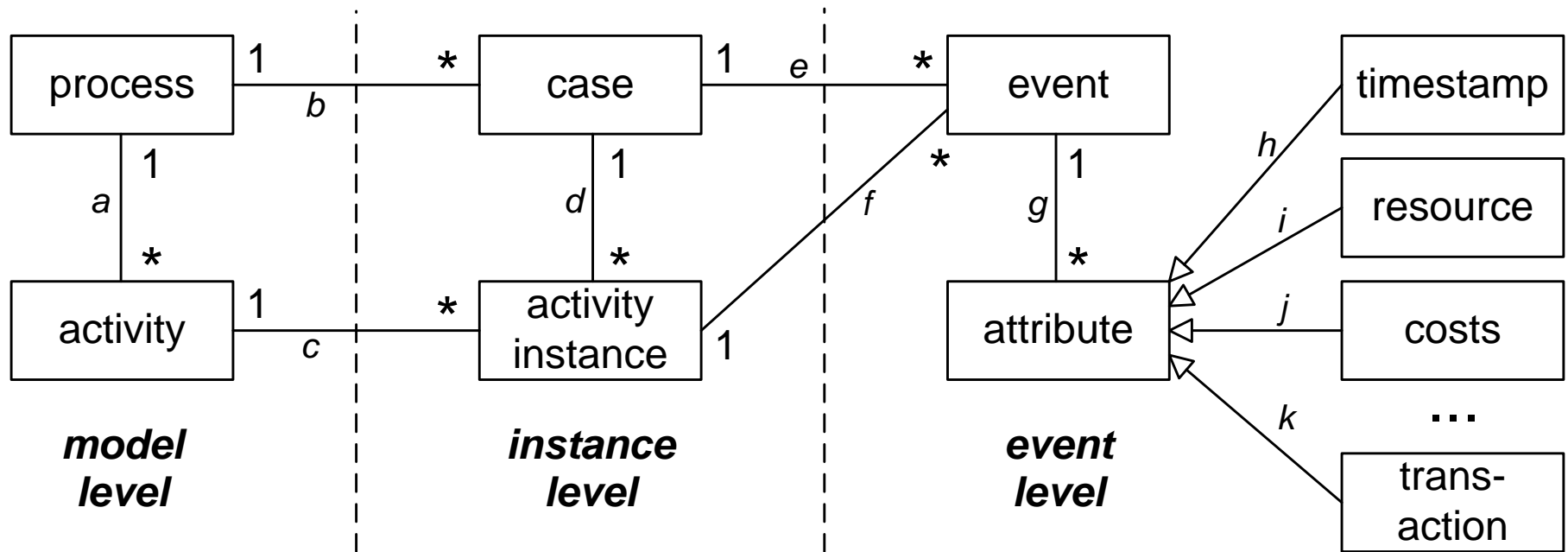
$N_2$

# Checking Declare specifications



See Declare and LTL checker in ProM

# Connecting event log and model



- **Very important!**
- **Model may be discovered or hand-made.**
- **Connected during replay.**
- **Starting point for other types of process mining!**