LEHRSTUHL 14
SOFTWARE ENGINEERING

Vorlesung

Methodische Grundlagen des Software-Engineering

im Sommersemester 2013

Prof. Dr. Jan Jürjens

TU Dortmund, Fakultät Informatik, Lehrstuhl XIV

Teil 2.6: Konformanzanalyse

v. 30.05.2013

2.6 Konformanzanalyse

[mit freundlicher Genehmigung basierend auf einem englischen Foliensatz von Prof. Dr. Wil van der Aalst (TU Eindhoven)]

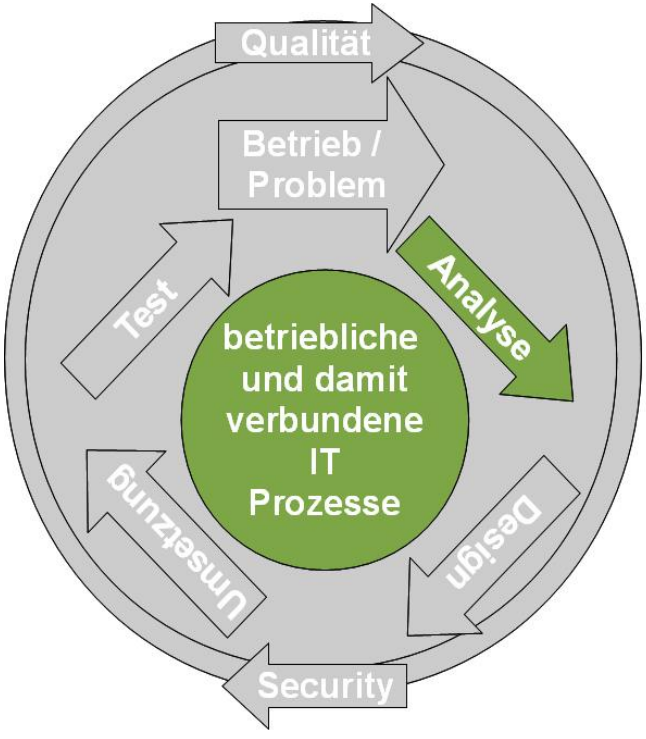
Literatur:

[vdA11] Wil van der Aalst: **Process Mining: Discovery, Conformance and Enhancement of Business Processes**, Springer-Verlag. 2011.
Unibibliothek (6 Exemplare): http://www.ub.tu-dortmund.de/katalog/titel/1332248
(Bei Engpässen kann eine **Kopiervorlage** der relevanten Ausschnitte zur Verfügung gestellt werden.)

- **Kapitel 7**

Einordnung
2.6: Konformanzanalyse

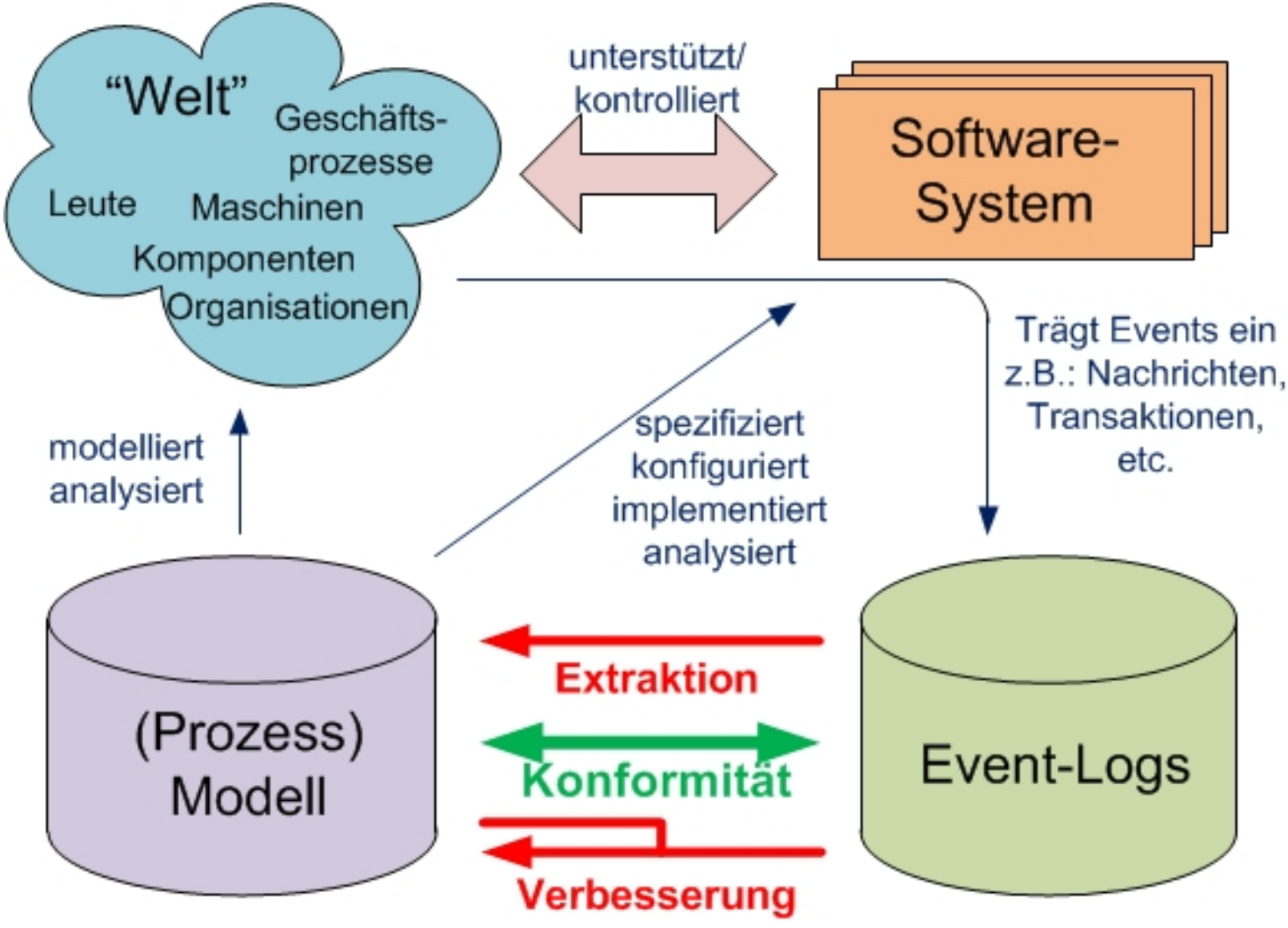
- Geschäftsprozessmodellierung
- Process-Mining
  - Einführung: Process-Mining
  - Petrinetze
  - Prozessmodellierung und Analyse
  - Data-Mining
  - Datenbeschaffung
  - Prozessextraktion
  - Konformanzanalyse
  - Mining: Zusätzliche Perspektiven
  - Betriebsunterstützung
  - Werkzeugunterstützung
  - Analysiere „Lasagne Prozesse“
  - Analysiere „Spaghetti Prozesse“
  - Kartographie und Navigation
  - Epilog
- Modellbasierte Softwareentwicklung
- Modellbasierte Entwicklung sicherer Software



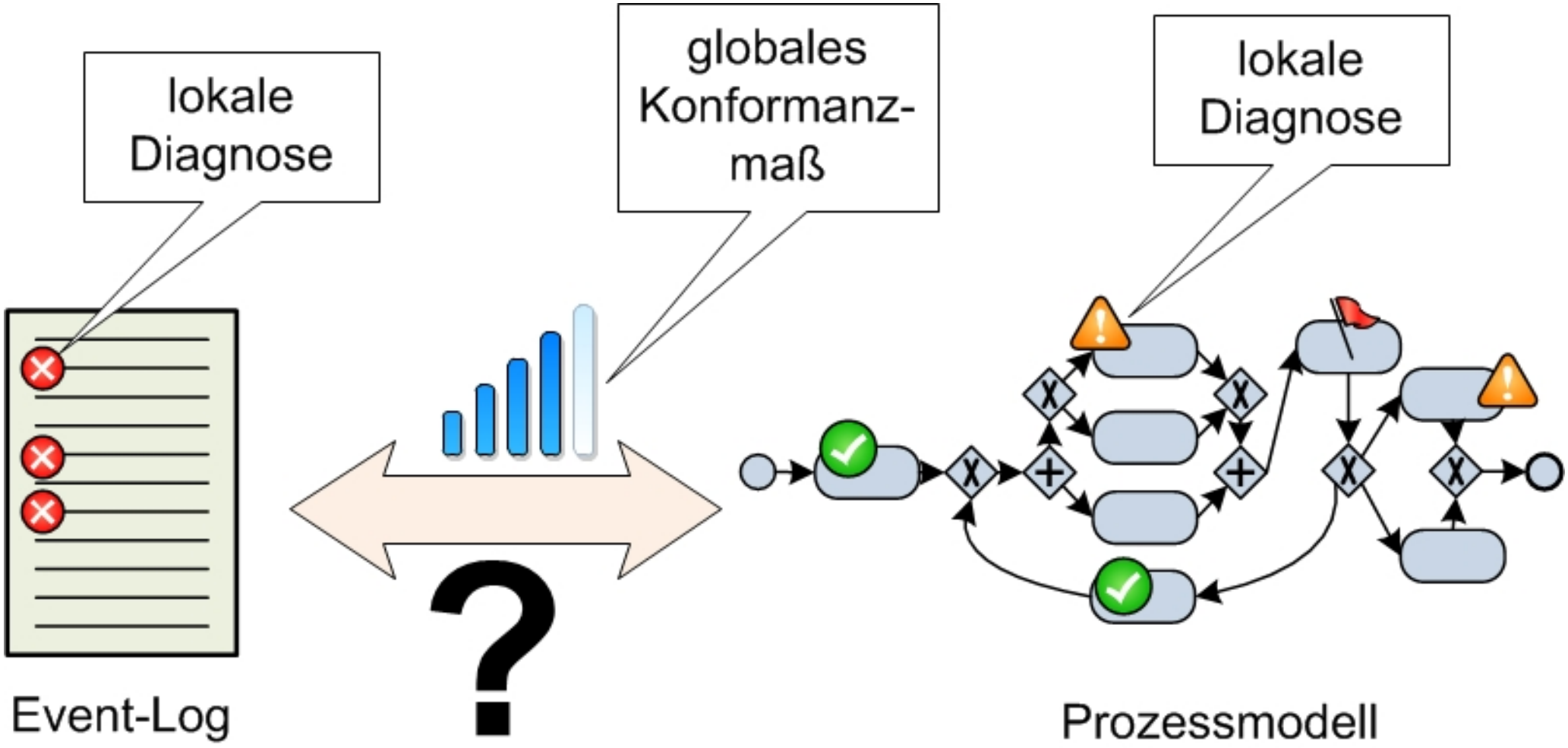
Überblick
2.6 Konformanzanalyse

- Einführung
- Konformanzanalyse basierend auf Replay
  - Motivation
  - Beispiele
- Diagnose

Konformanzanalyse



Konformanzanalyse



Kontext

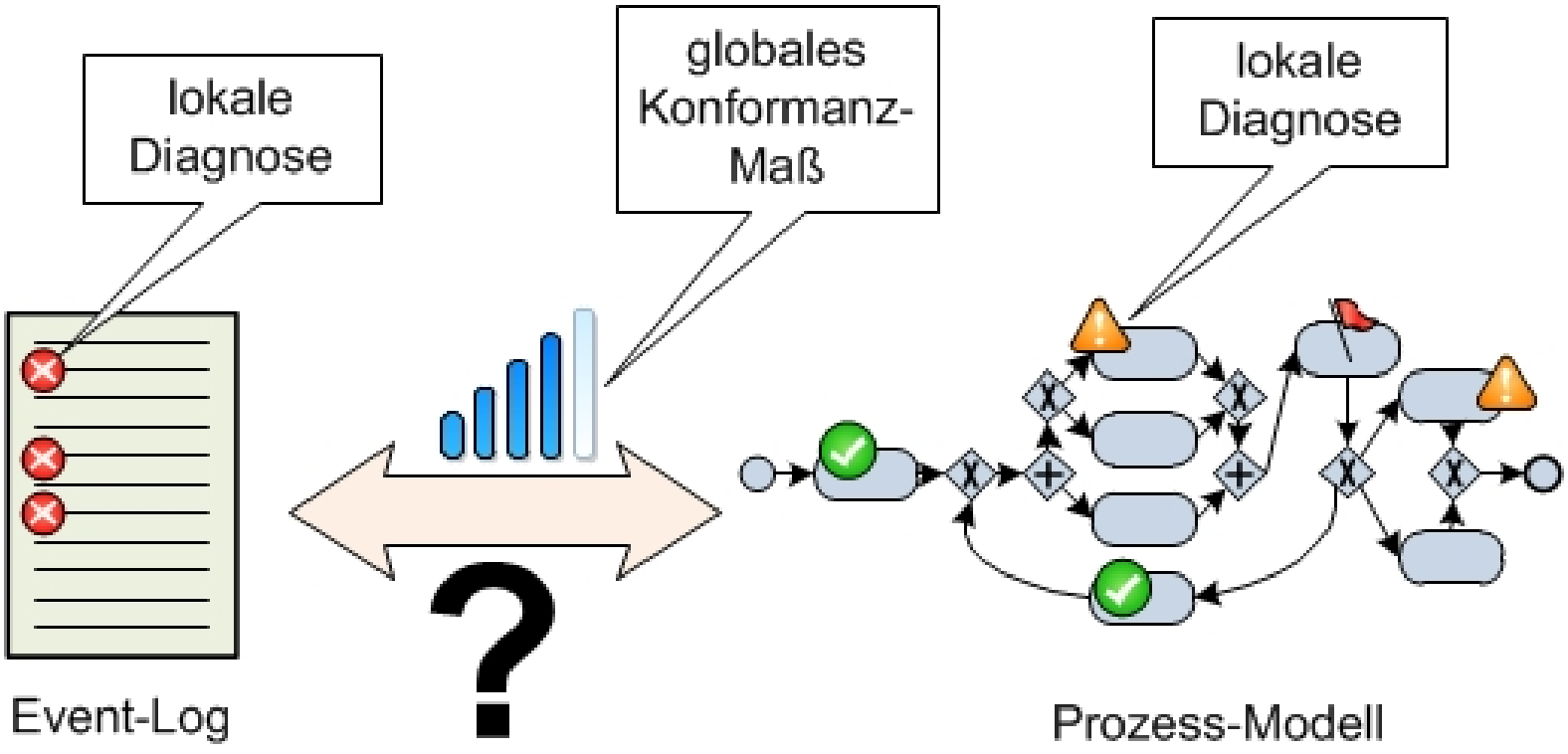
- **Corporate Governance**, **Risk**, **Compliance**, und **Gesetzgebung**
  - z.B.: Sarbanes-Oxley (US), Basel II/III (EU), J-SOX (Japan), C-SOX (Kanada), 8th EU Directive (EURO-SOX), Bilanzrechts-modernisierungsgesetz (BilMoG) (Deutschland), Markets in Financial Instruments Directive (MiFID) (EU), Law 262/05 (Italien), Code Lippens (Belgien), und Code Tabaksblat (Niederlande).
- **ISO 9001:2008:** Unternehmen müssen Betriebsprozesse modellieren.
- **Business alignment**: Sicherstellen, dass Informationssystem und realer Geschäftsprozess gut aufeinander abgestimmt sind.

Auditierung

- **Auditierung**: Evaluation von Unternehmen und ihren Prozessen.
- **Audits** stellen **Validität** und **Zuverlässigkeit** von Informationen über Unternehmen und entsprechende Prozesse sicher.
- Test, ob Geschäftsprozesse in bestimmten Grenzen ausgeführt werden (von Managern, der Politik und anderen Interessenvertretern gesetzt).
- **Process-Mining** kann bei Aufdecken von **Betrug**, **Fehlverhalten**, **Risiken** und **Ineffizienzen** helfen.
- Alle Events eines Geschäftsprozesses können evaluiert werden, auch während der Prozess noch läuft.

Abweichungen ?

- **Modell** oder **Log** “falsch” ?
- “**Gewünschte**” oder “**unerwünschte**” Abweichungen ?
- “Glas zerbrechen” könnte Leben retten !

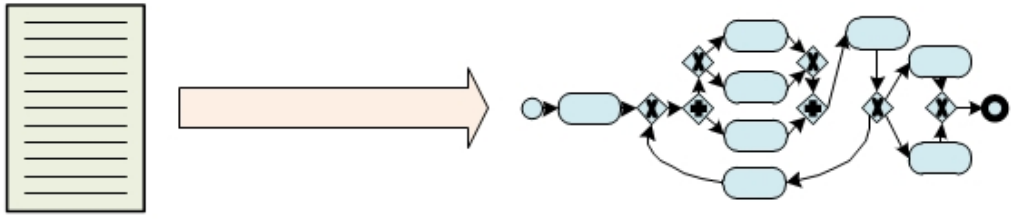


Überblick
2.6 Konformanzanalyse

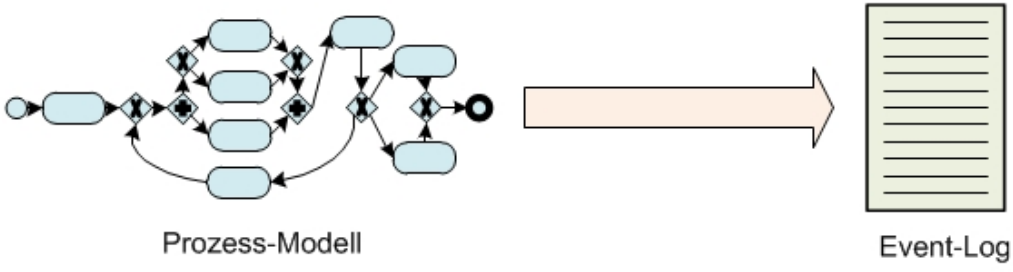
- Einführung
- Konformanzanalyse basierend auf Replay
  - Motivation
  - Beispiele
- Diagnose

Replay: Events mit Modellelementen verknüpfen

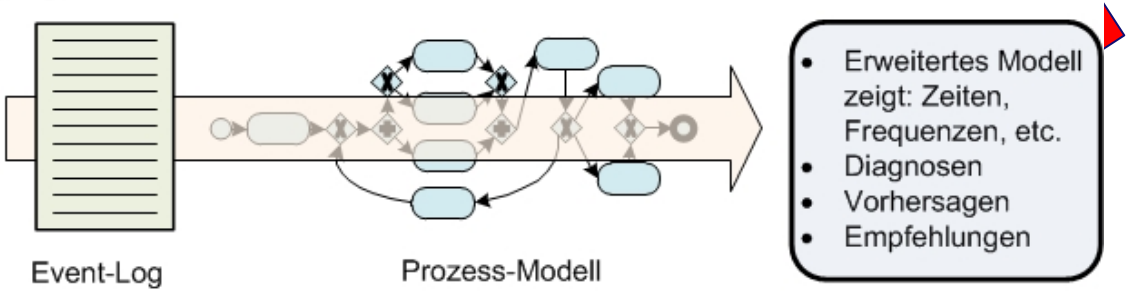
Play-In



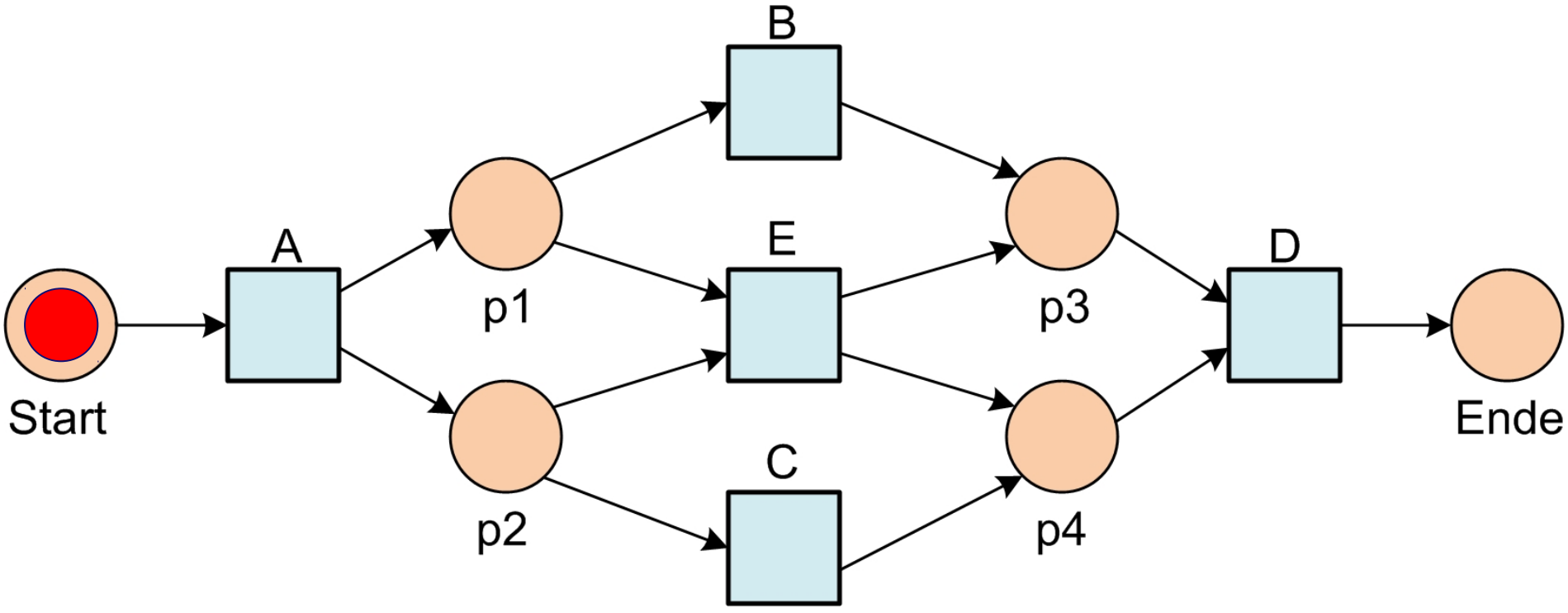
Play-Out



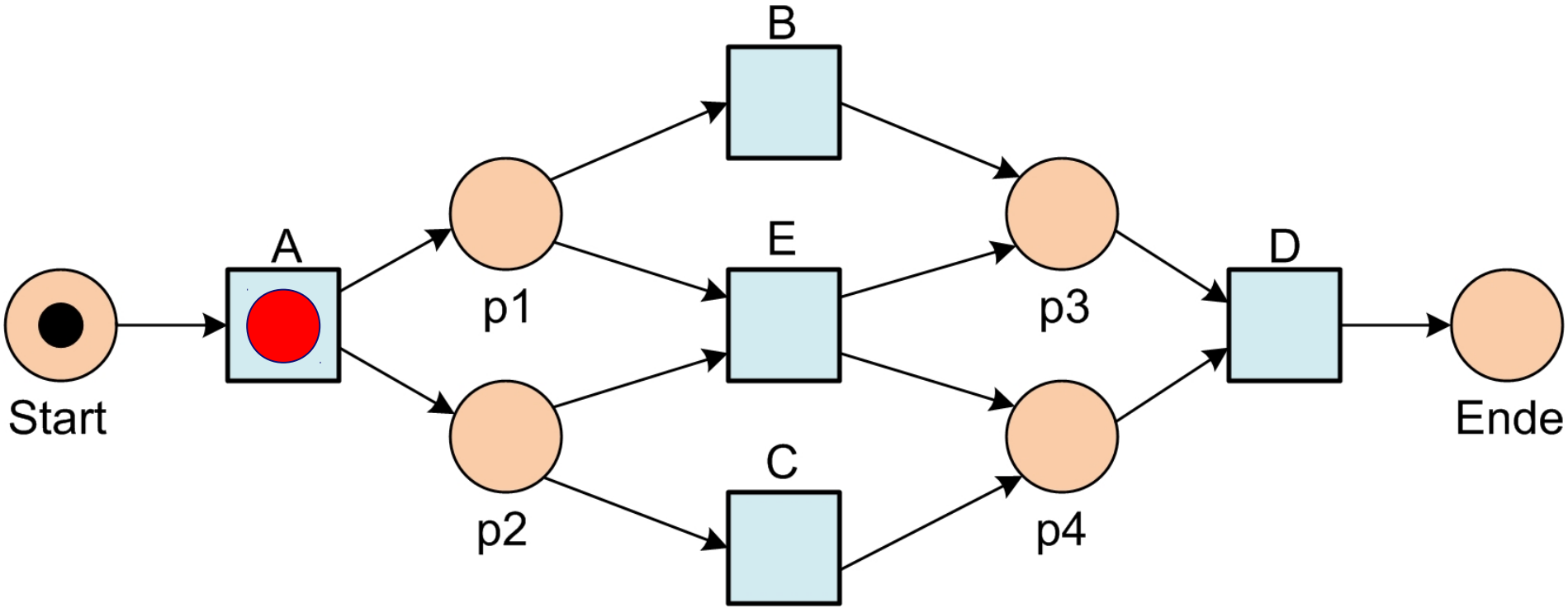
Replay



Play Out: Klassische Anwendung von Modellen

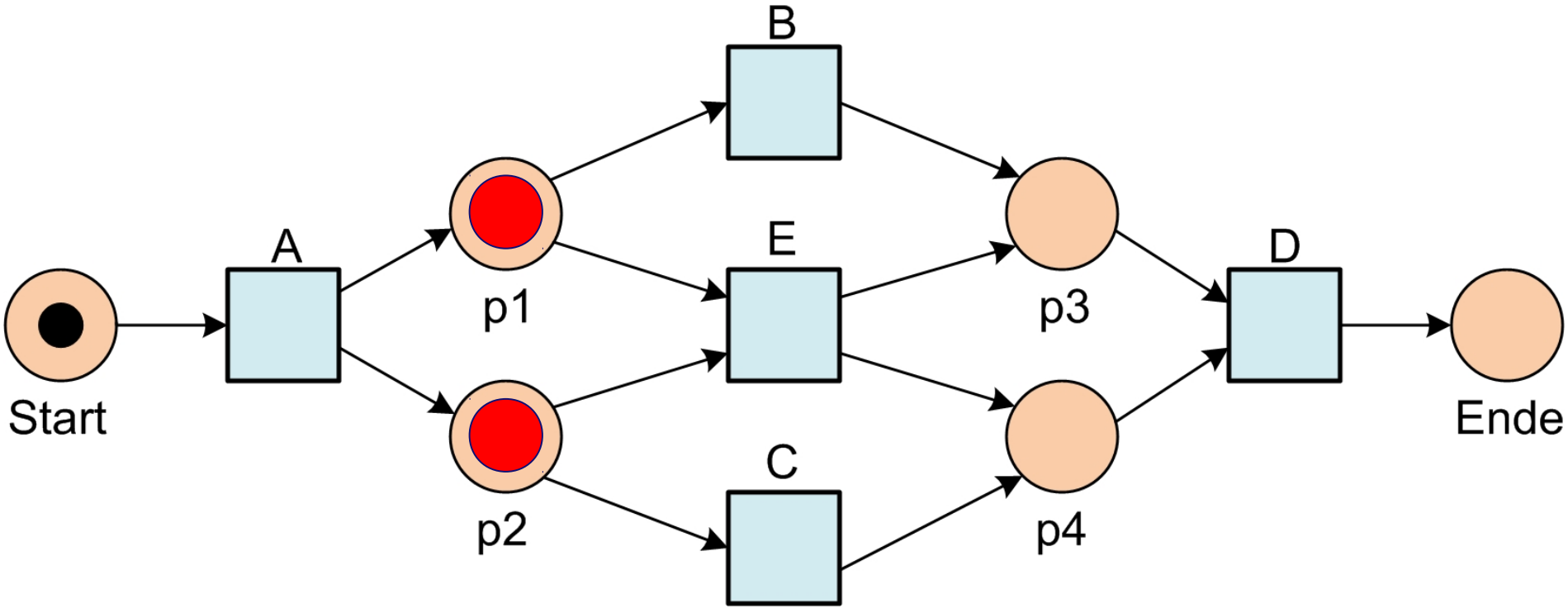


Play Out: Klassische Anwendung von Modellen



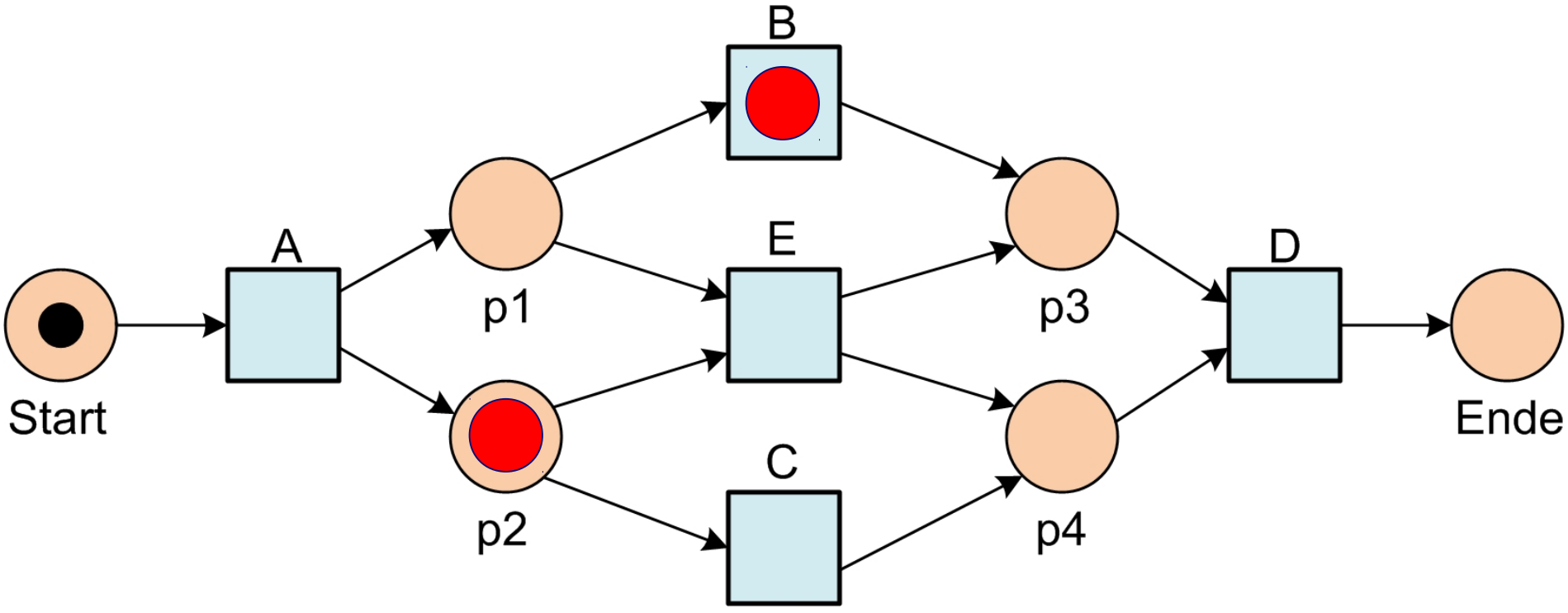
A

Play Out: Klassische Anwendung von Modellen



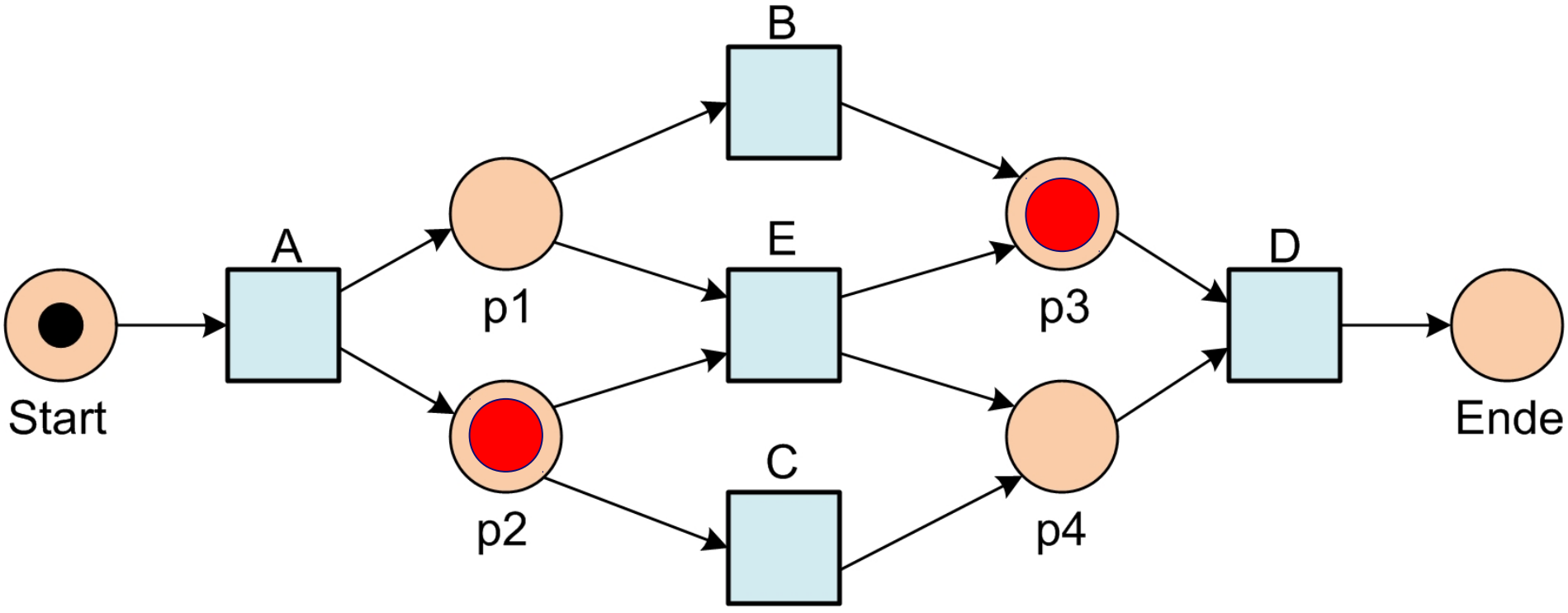
A

Play Out: Klassische Anwendung von Modellen



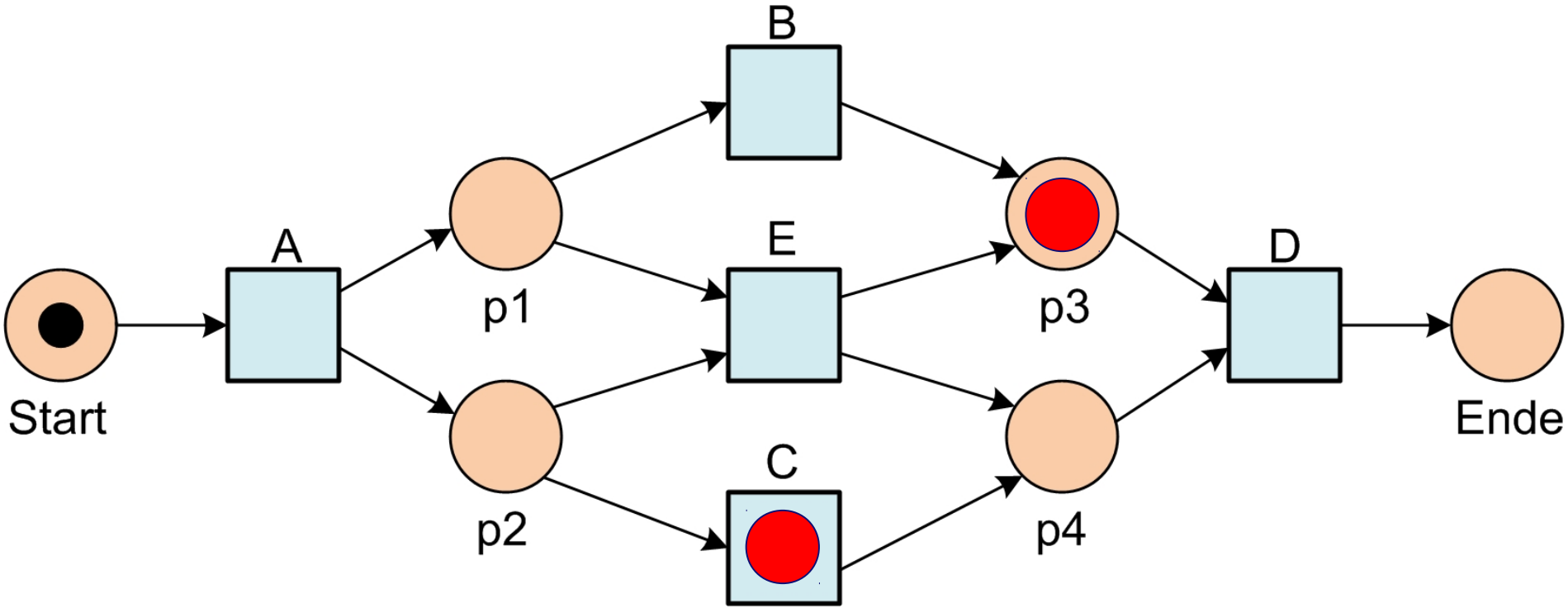
A B

Play Out: Klassische Anwendung von Modellen



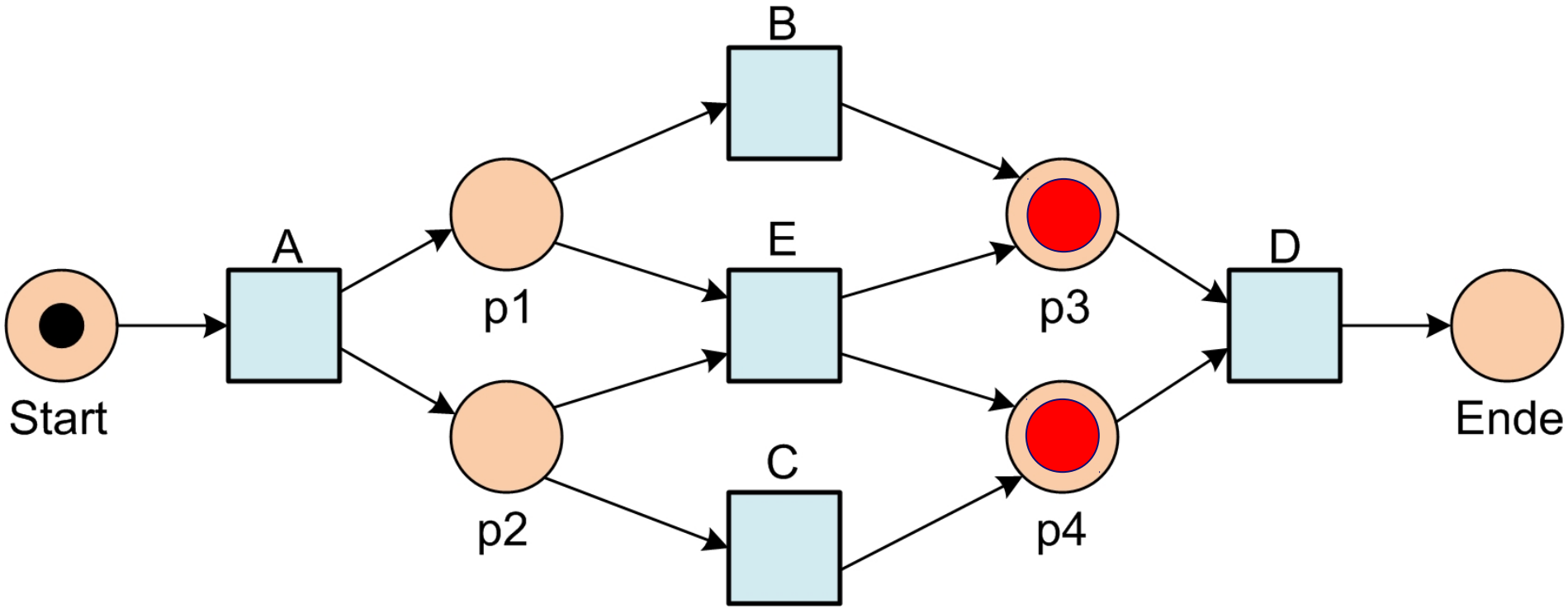
A B

Play Out: Klassische Anwendung von Modellen



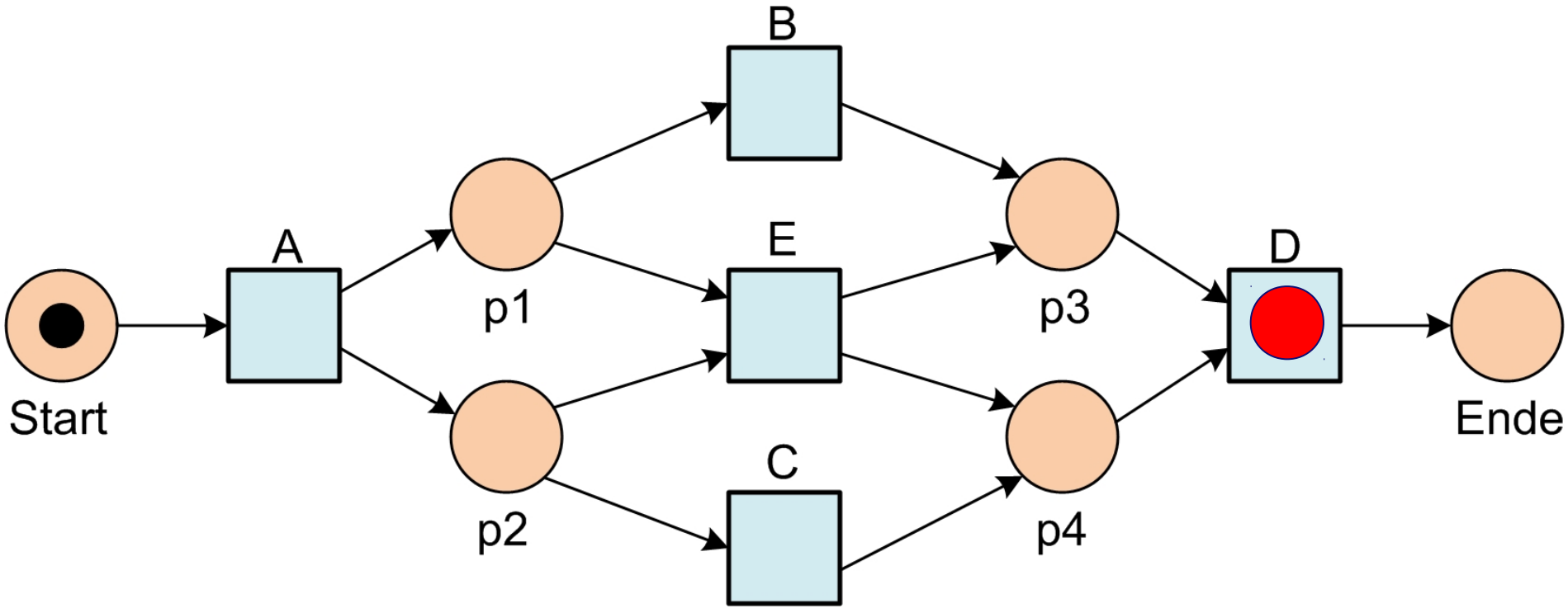
A B C

Play Out: Klassische Anwendung von Modellen



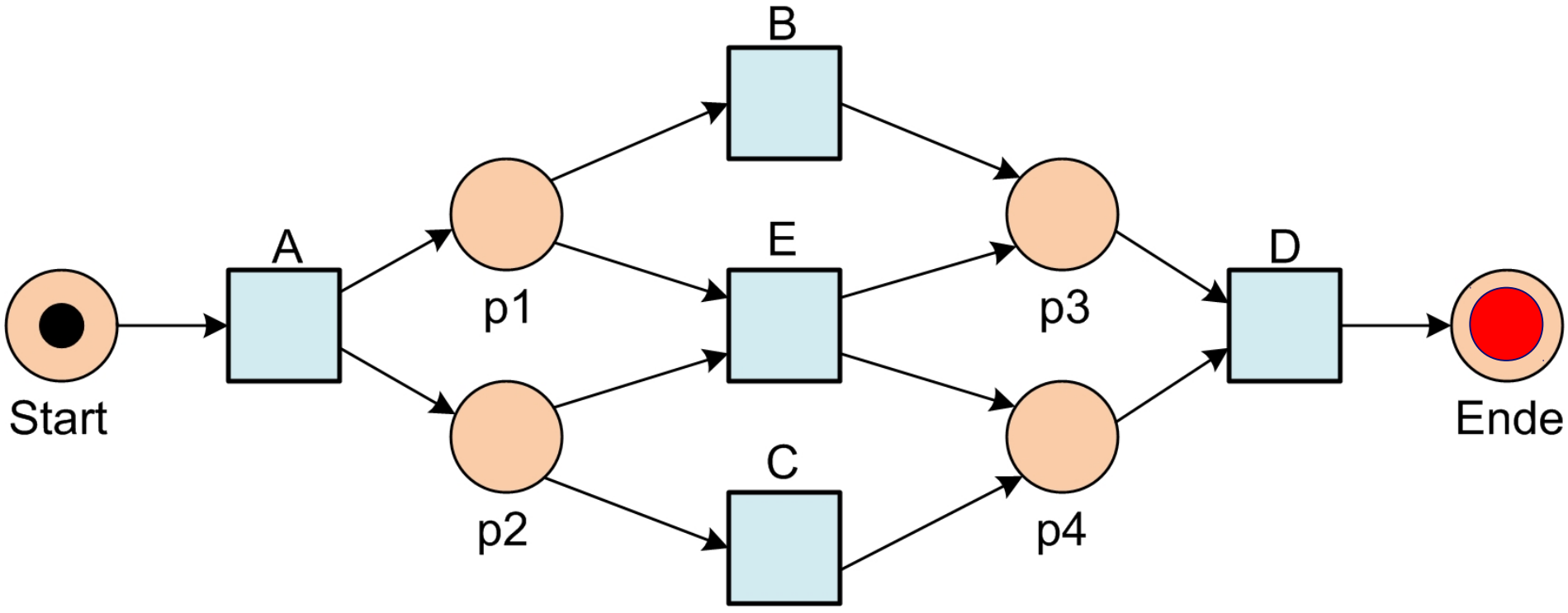
A B C

Play Out: Klassische Anwendung von Modellen



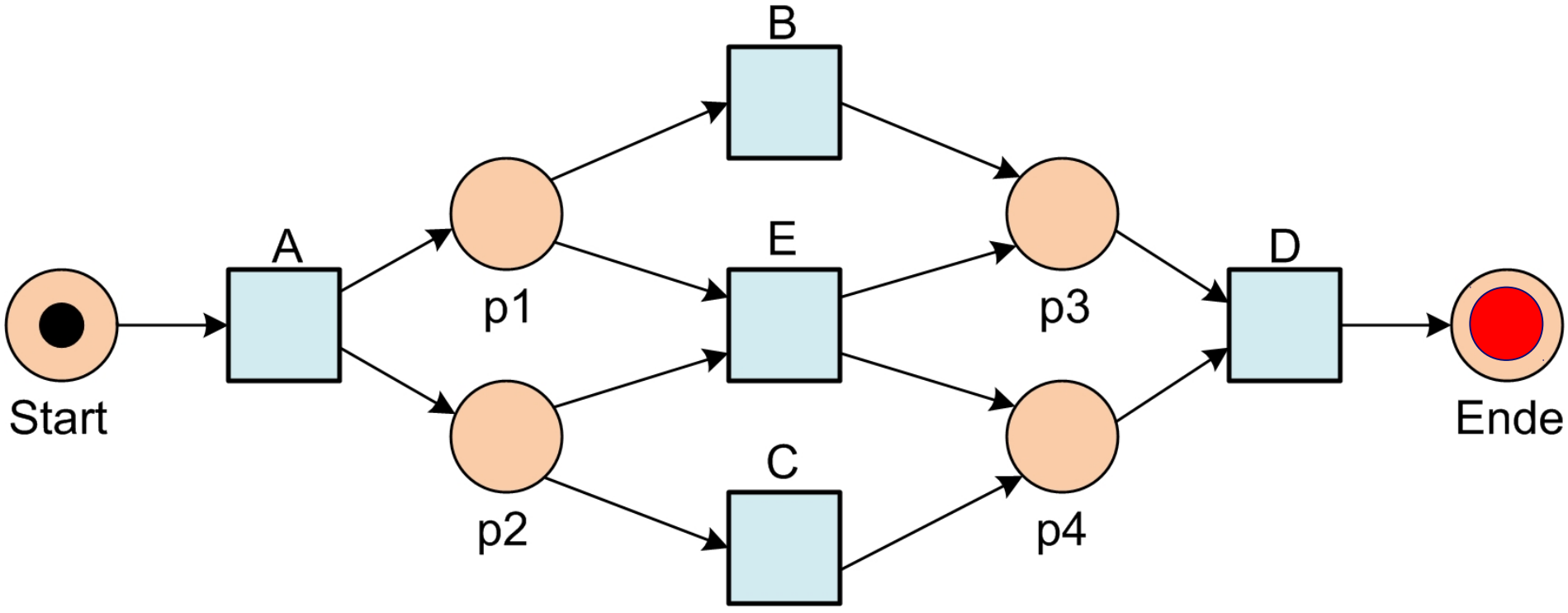
A B C D

Play Out: Klassische Anwendung von Modellen



A B C D

Play Out: Klassische Anwendung von Modellen



A B C D

A E D

A E D

A B C D

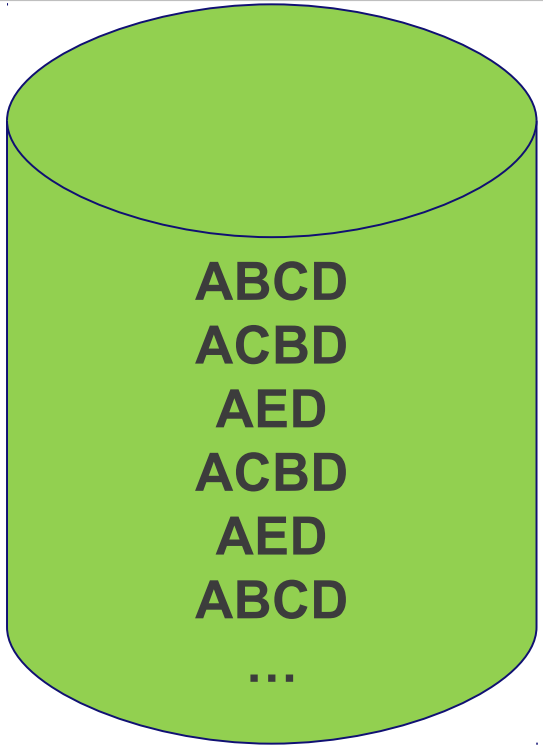
A C B D

A C B D

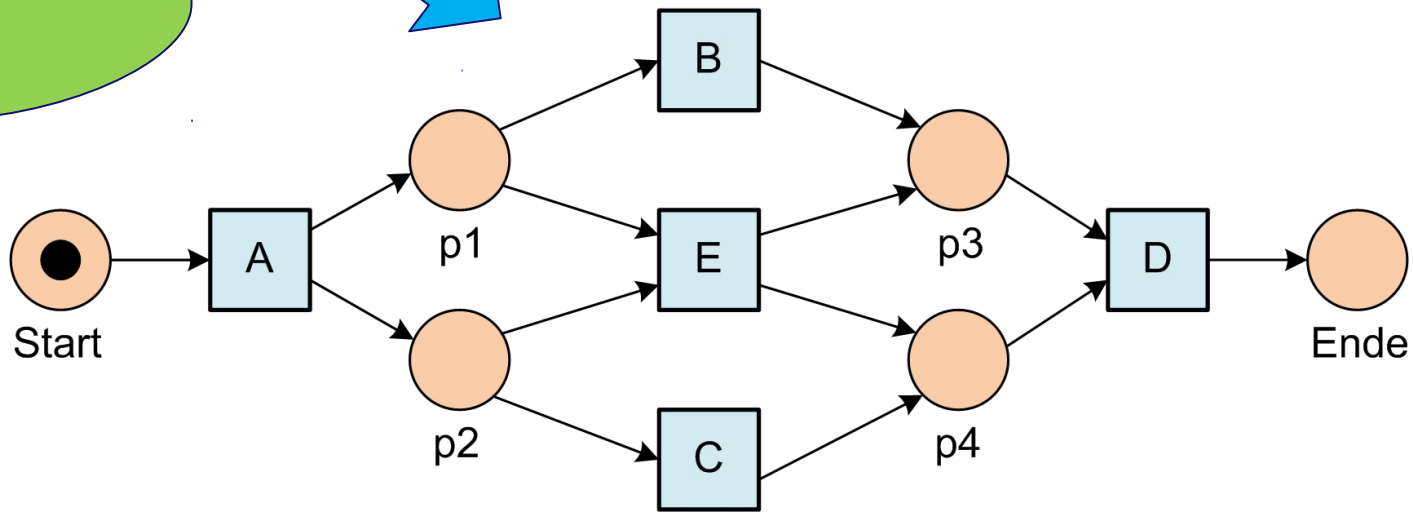
A E D

A C B D

Prozessextraktion

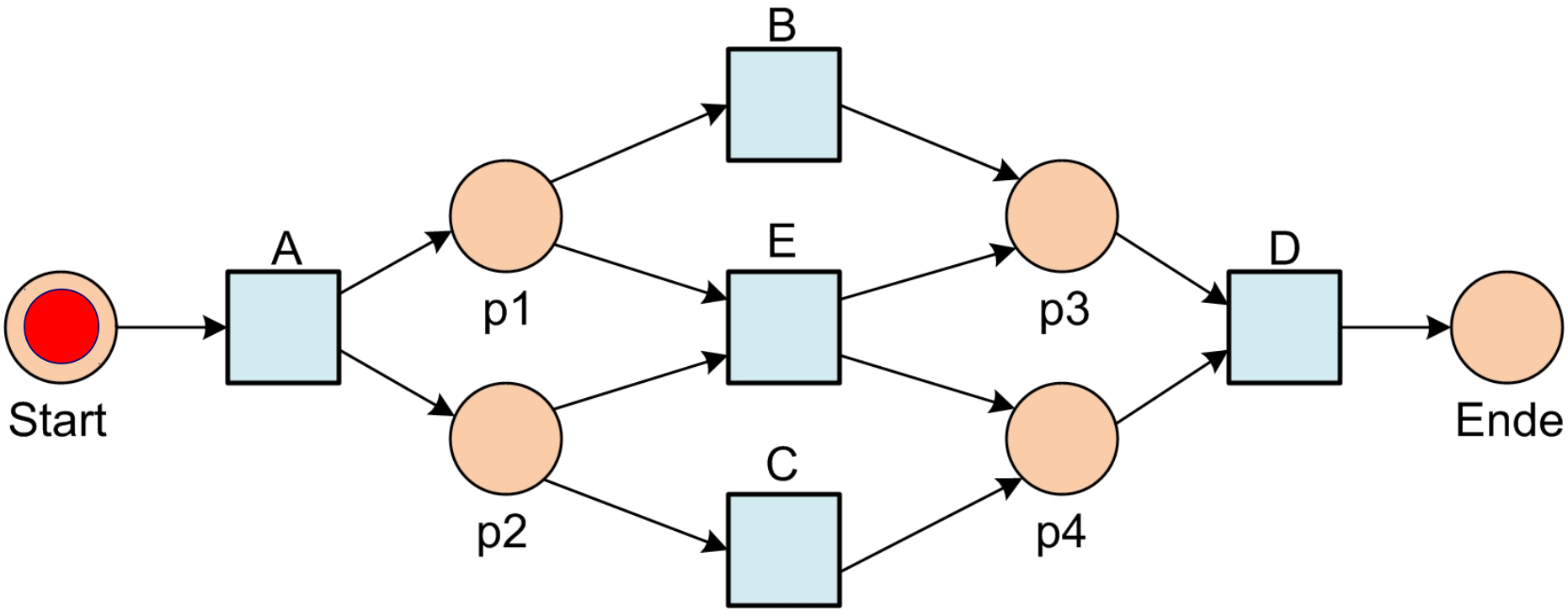


Prozessextraktions-Algorithmus wie z.B.:
α-Algorithmus



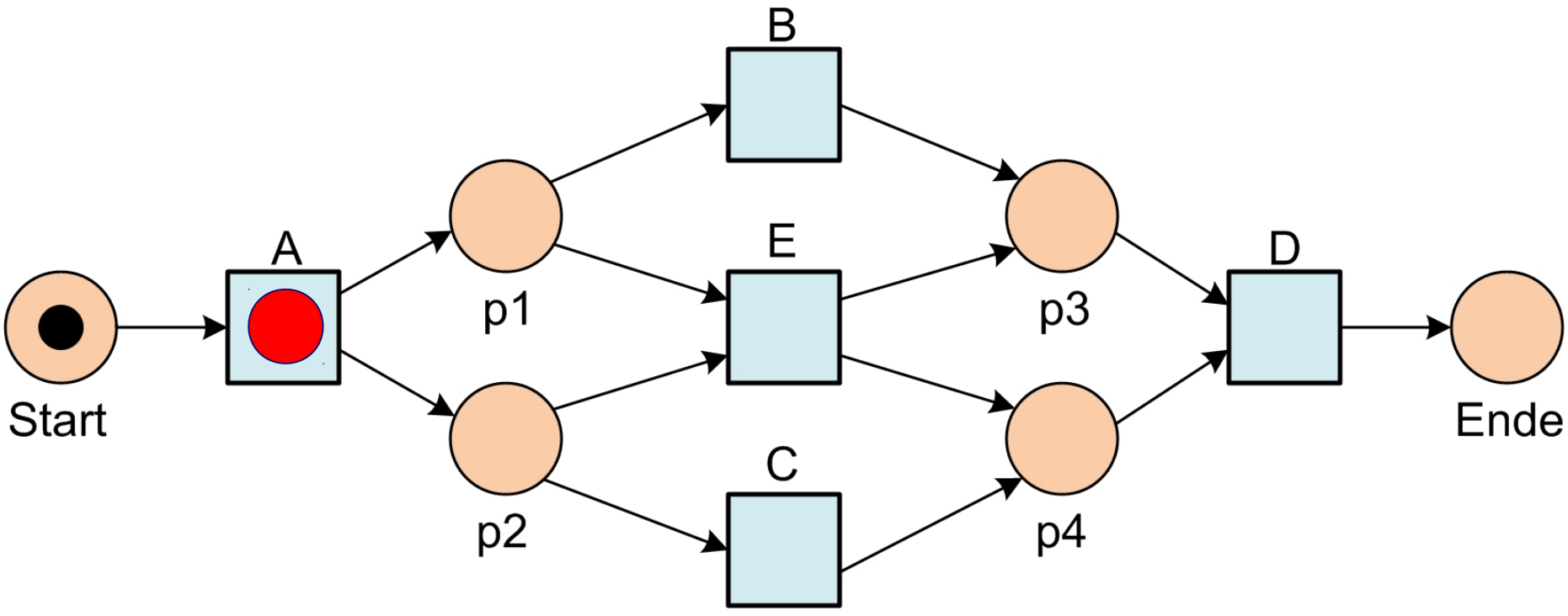
Konformanzanalyse durch Replay

A B C D



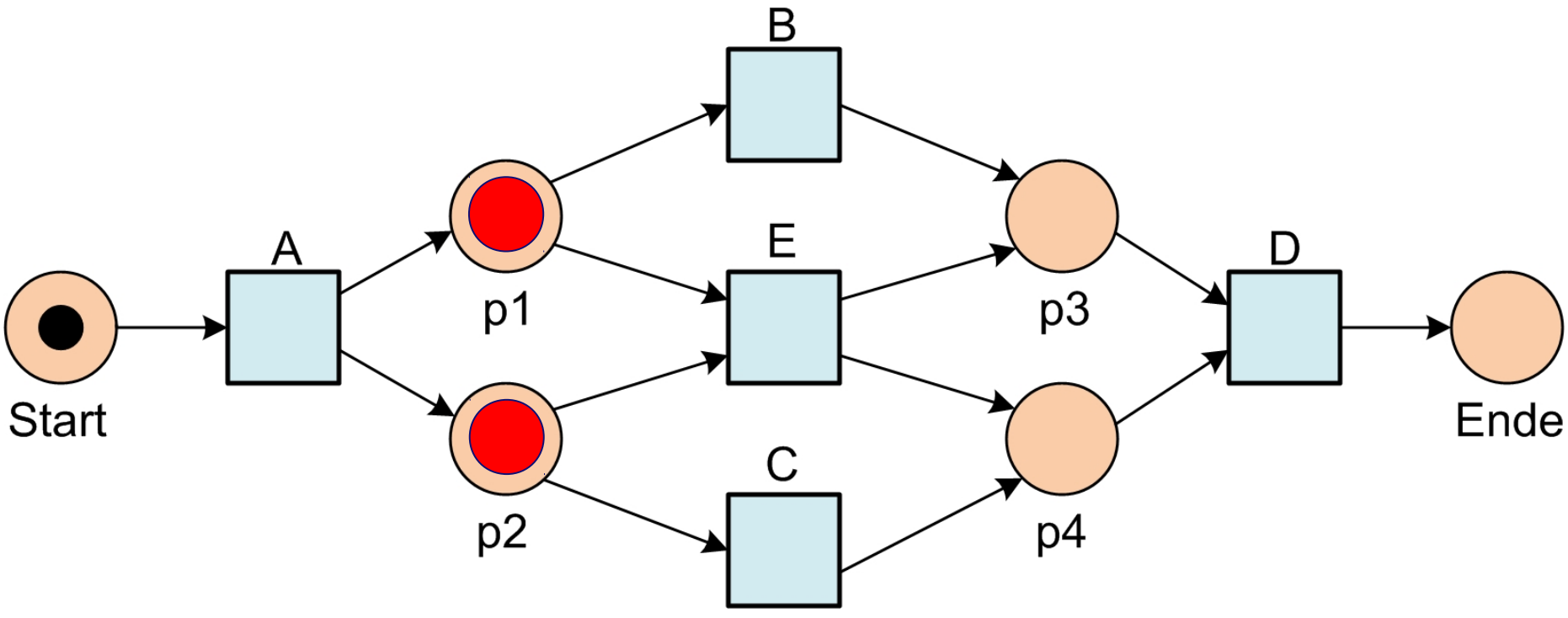
Konformanzanalyse durch Replay

B C D

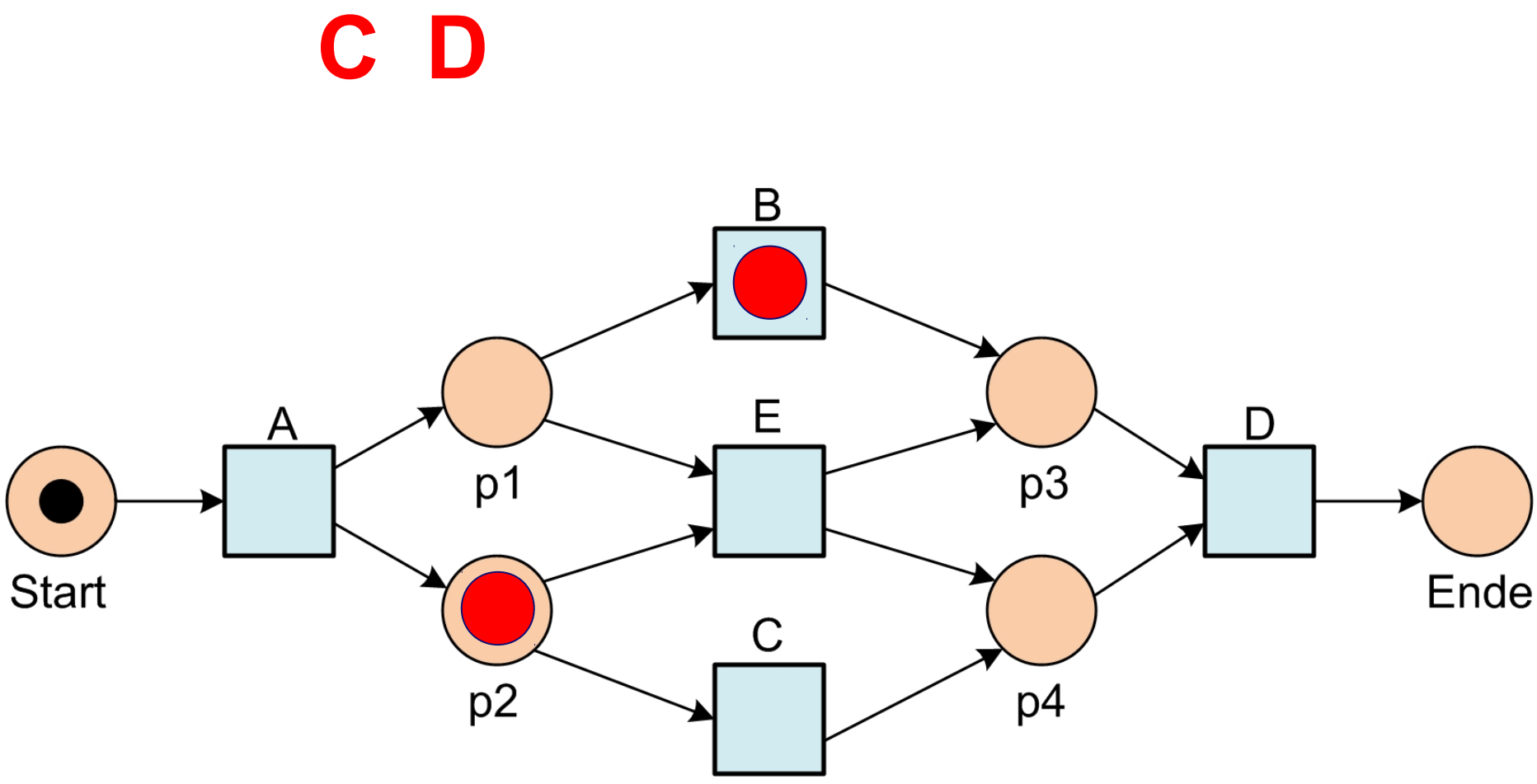


Konformanzanalyse durch Replay

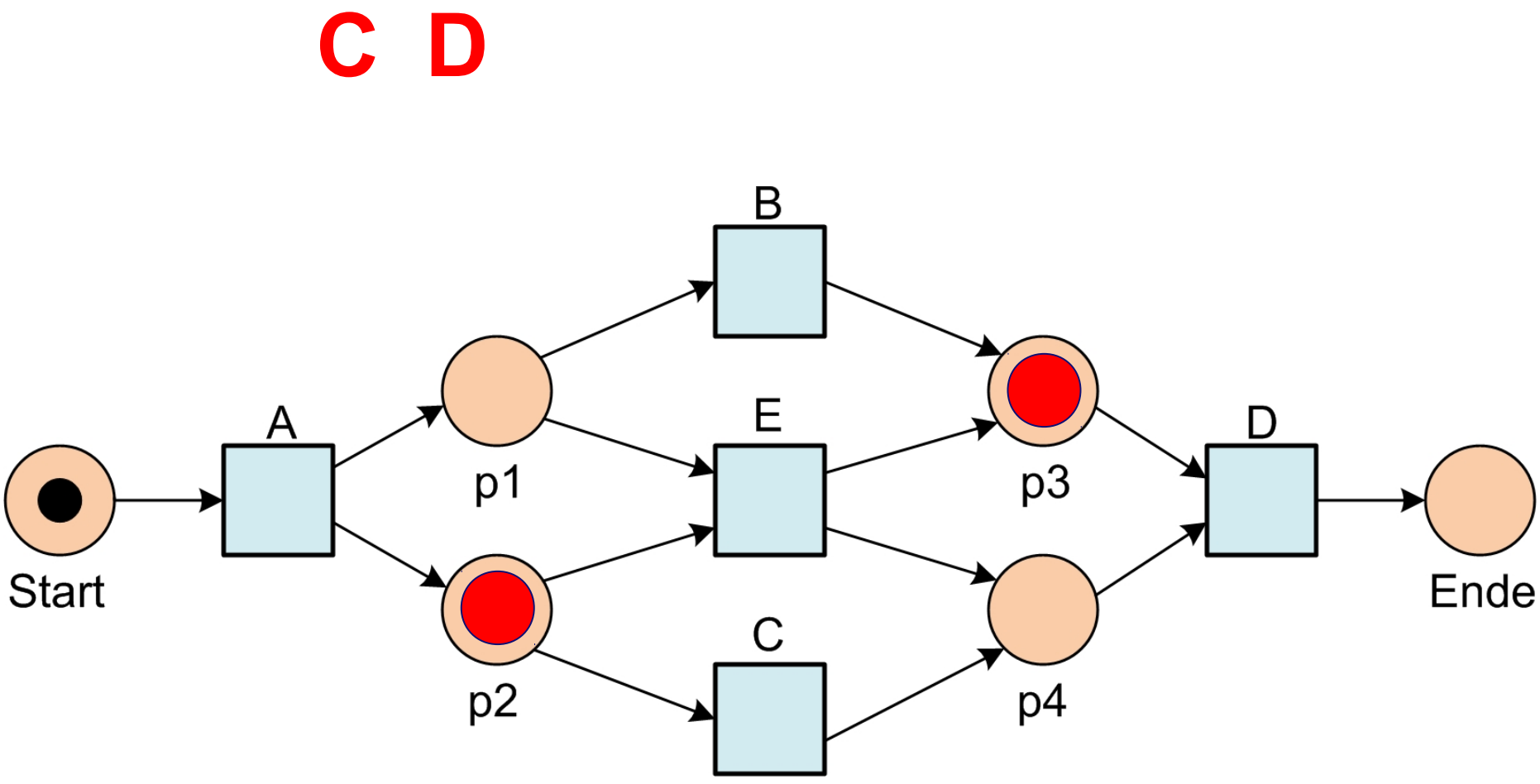
B C D



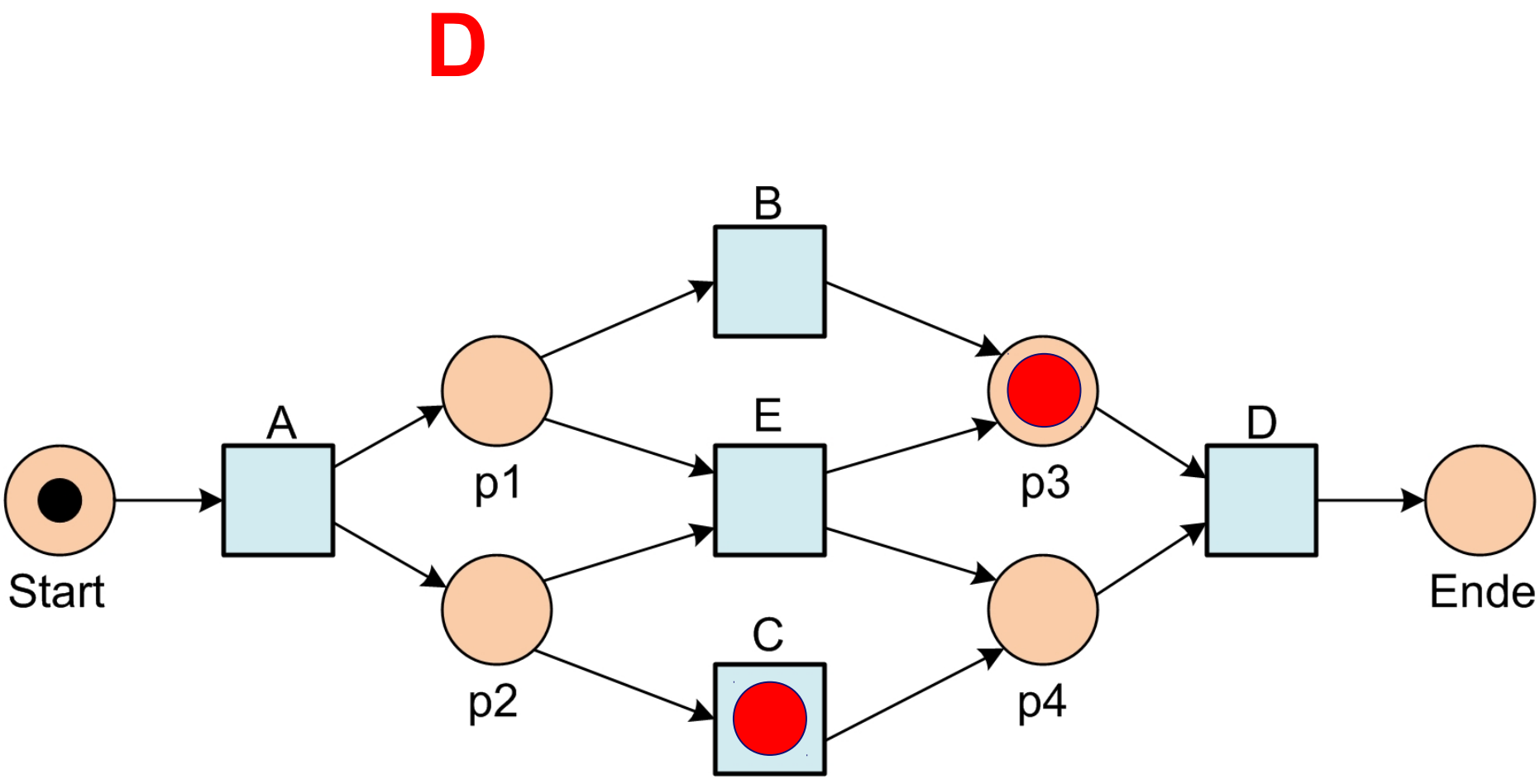
Konformanzanalyse durch Replay



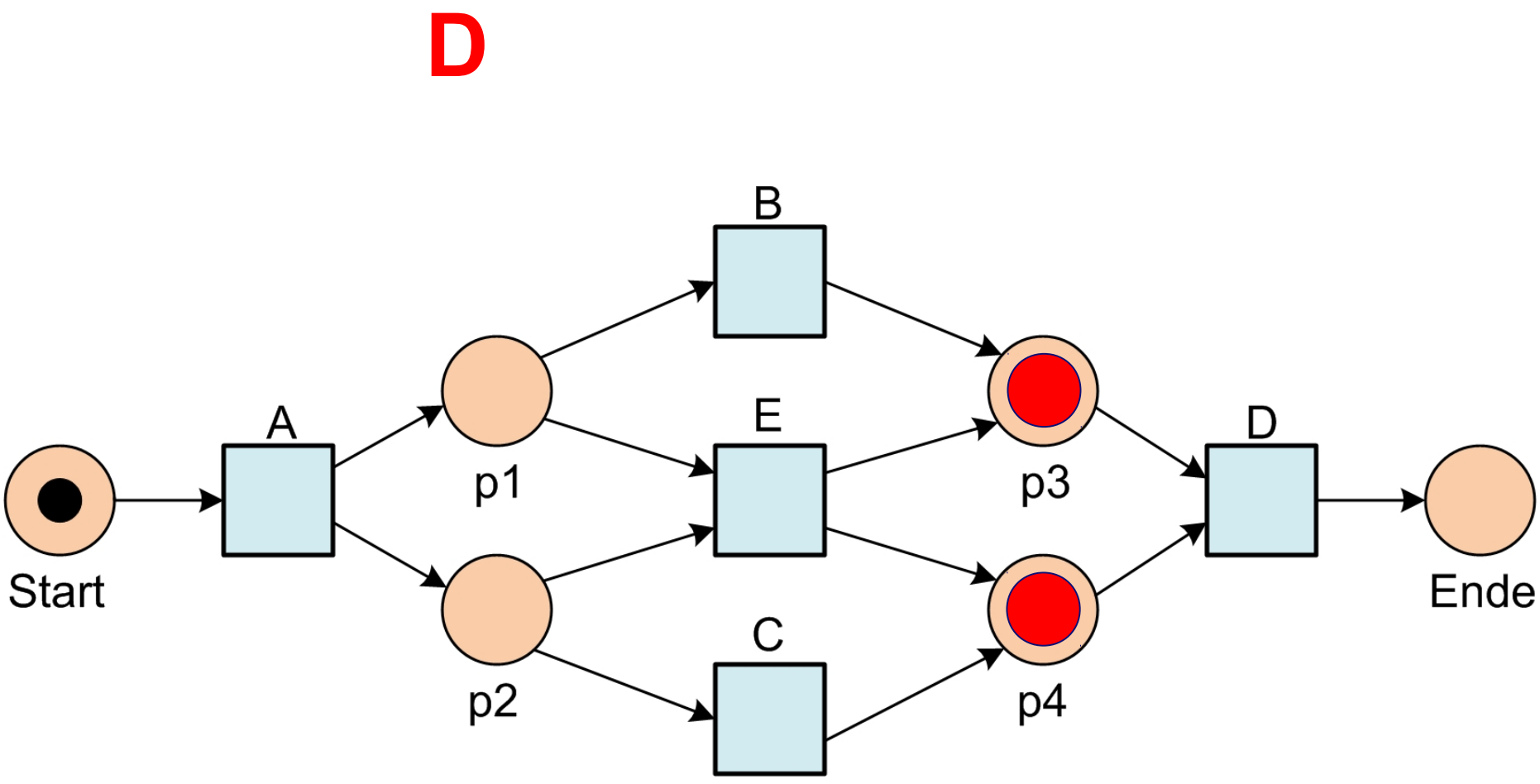
Konformanzanalyse durch Replay



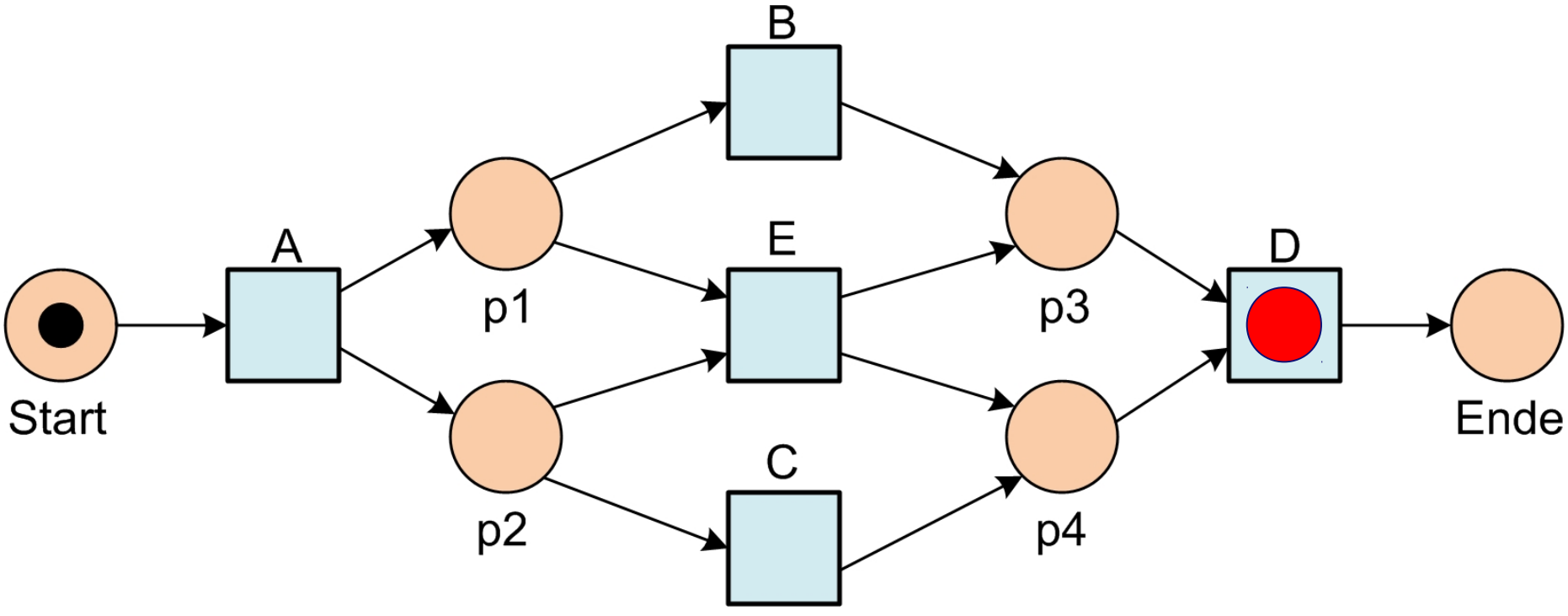
Konformanzanalyse durch Replay



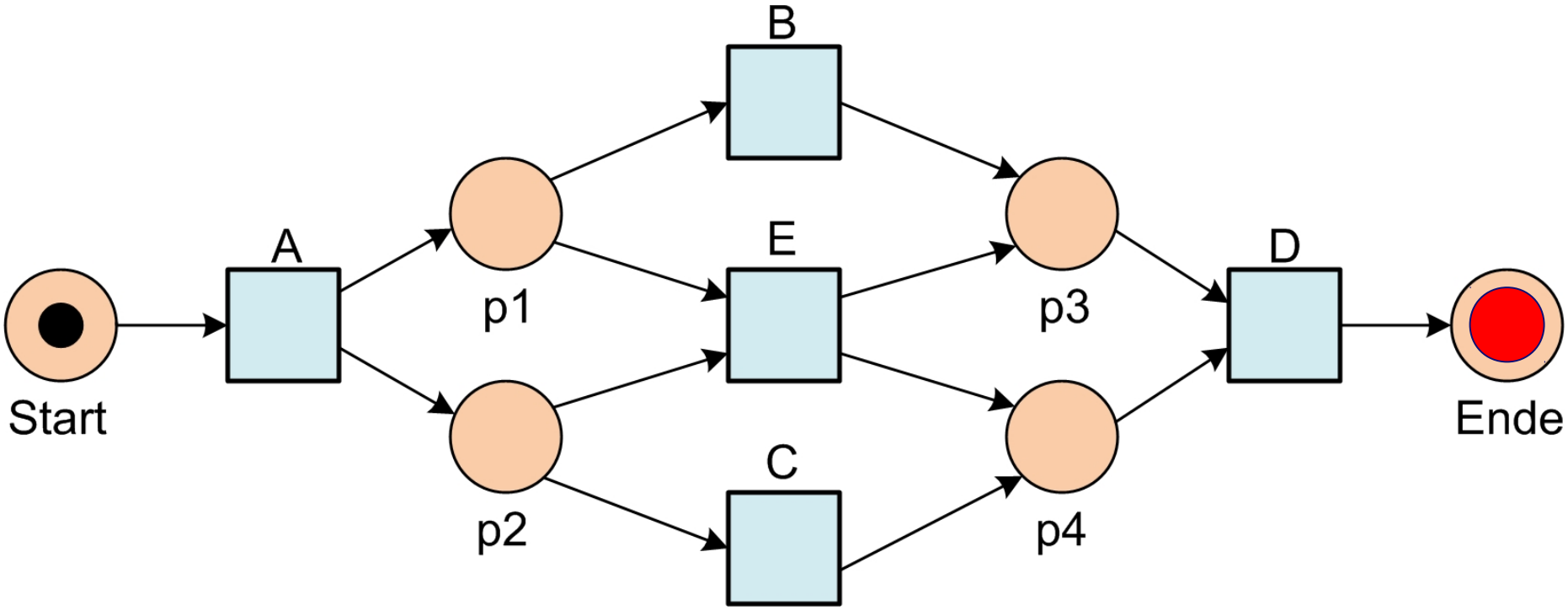
Konformanzanalyse durch Replay



Konformanzanalyse durch Replay

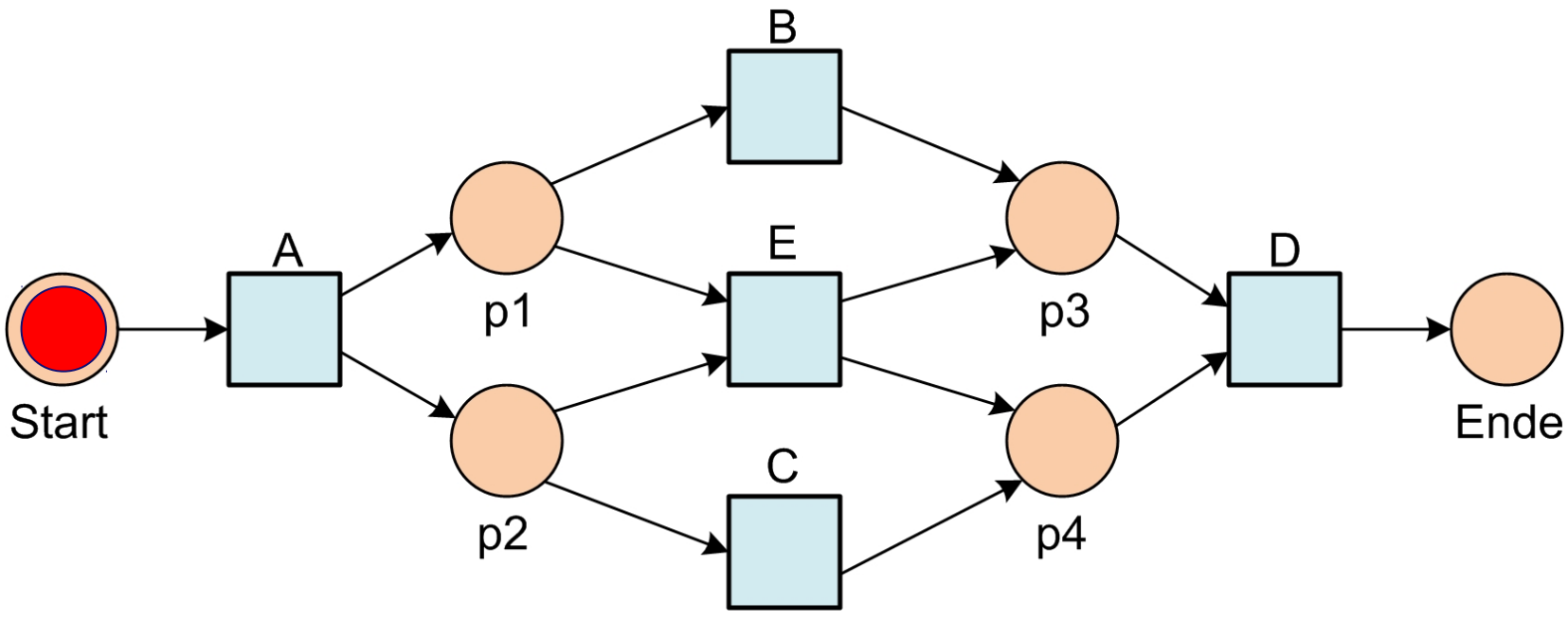


Konformanzanalyse durch Replay



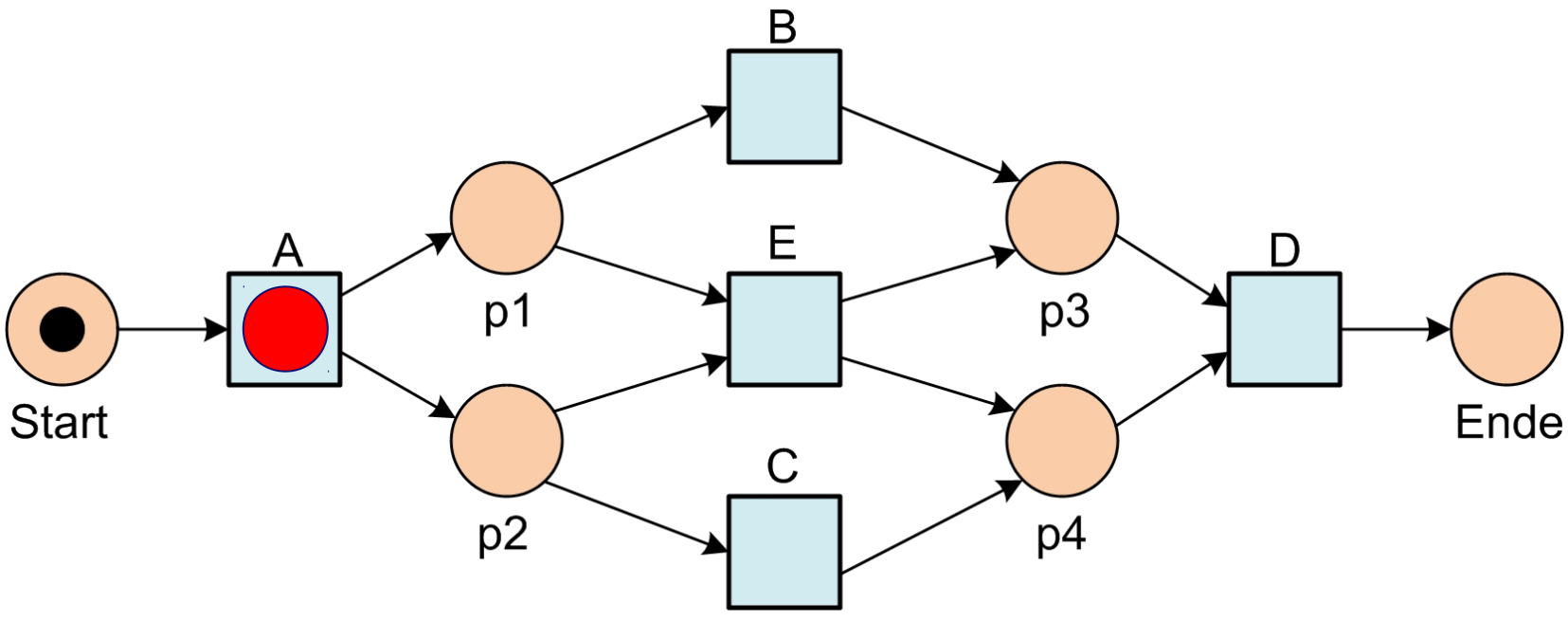
Replay: Mangelnde Konformanz aufdecken

A C D



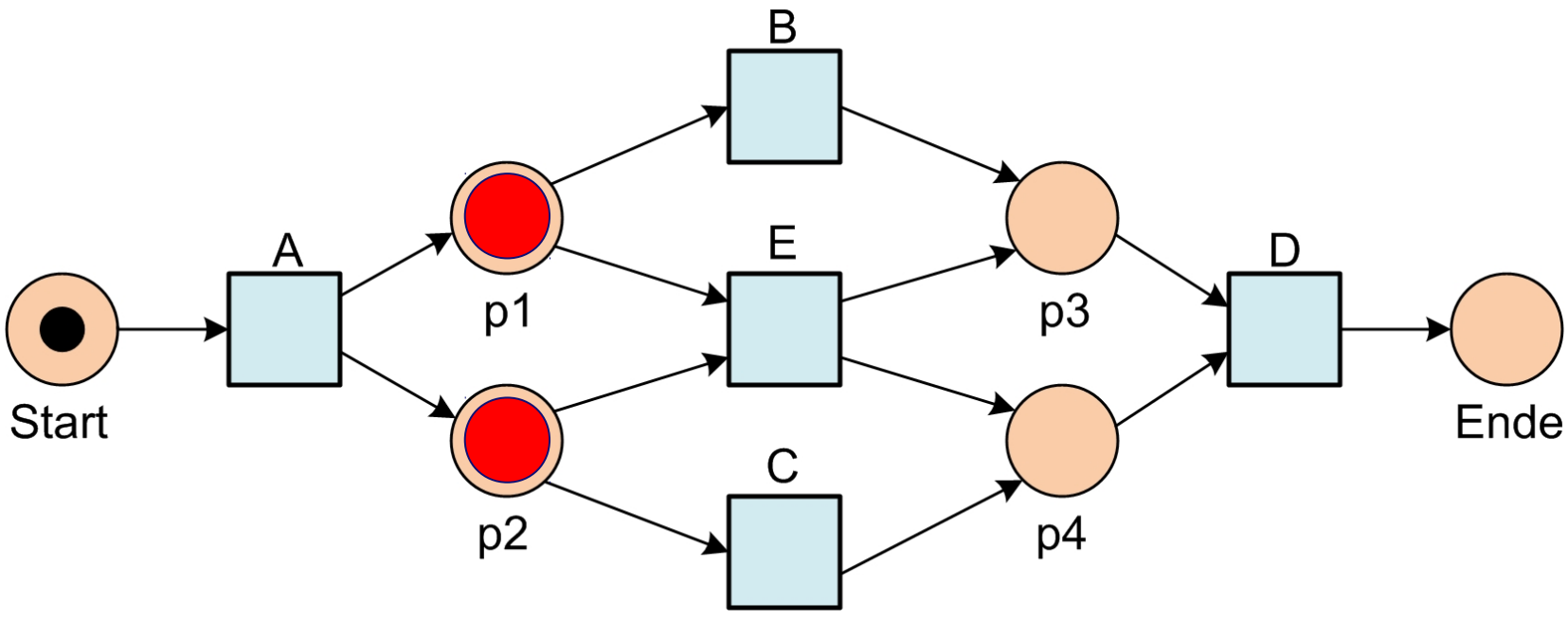
Replay: Mangelnde Konformanz aufdecken

C D



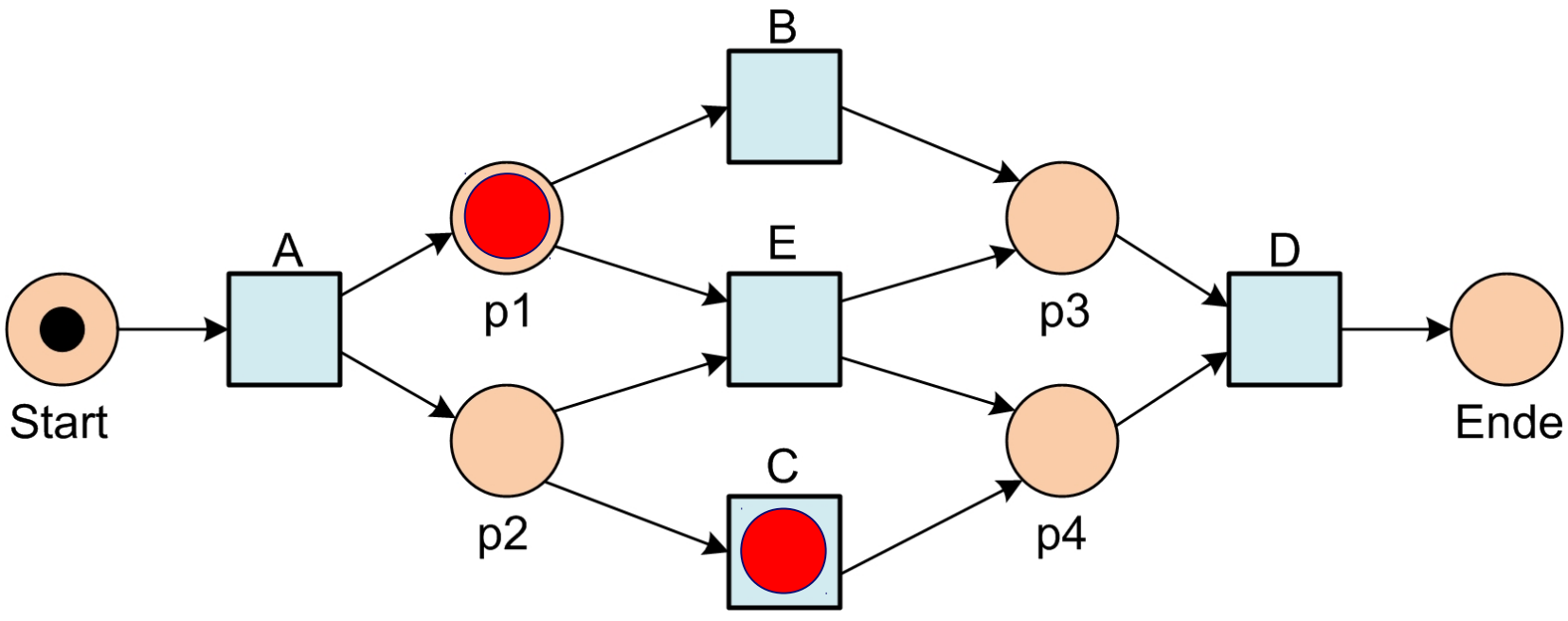
Replay: Mangelnde Konformanz aufdecken

C D



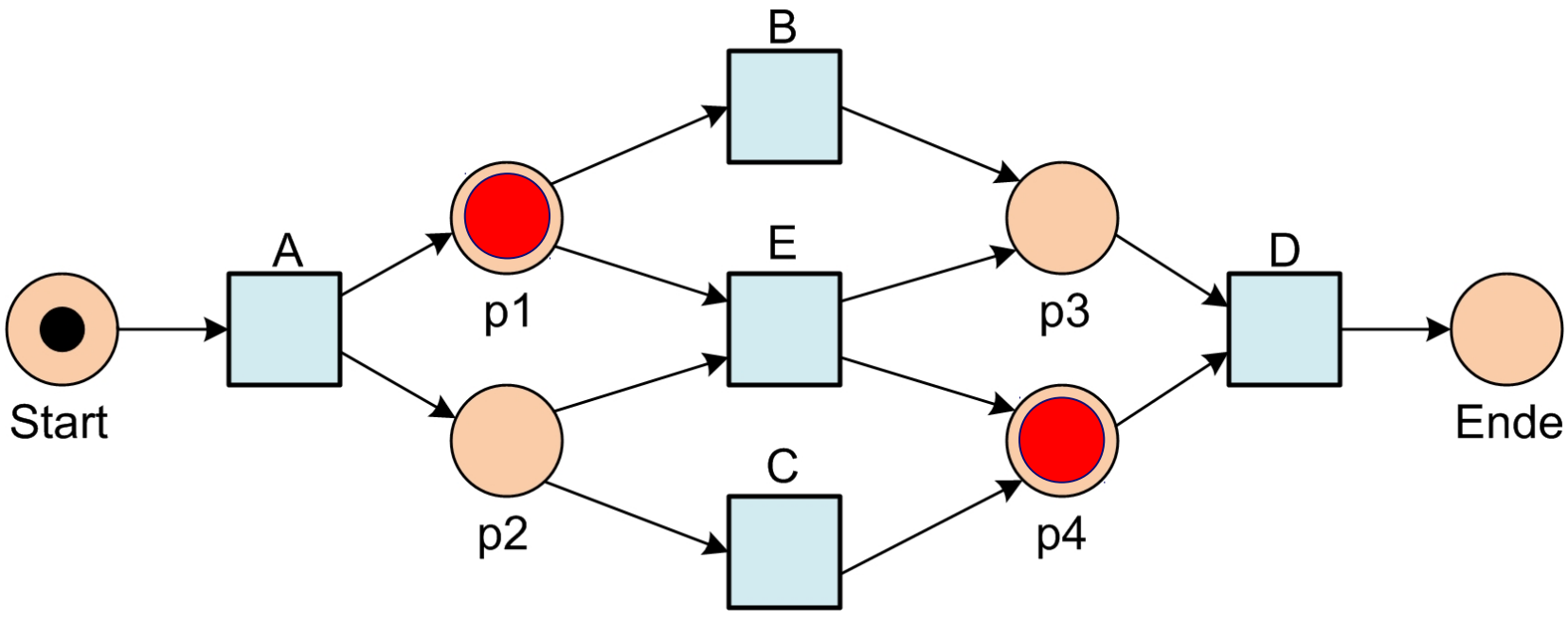
Replay: Mangelnde Konformanz aufdecken

D



Replay: Mangelnde Konformanz aufdecken

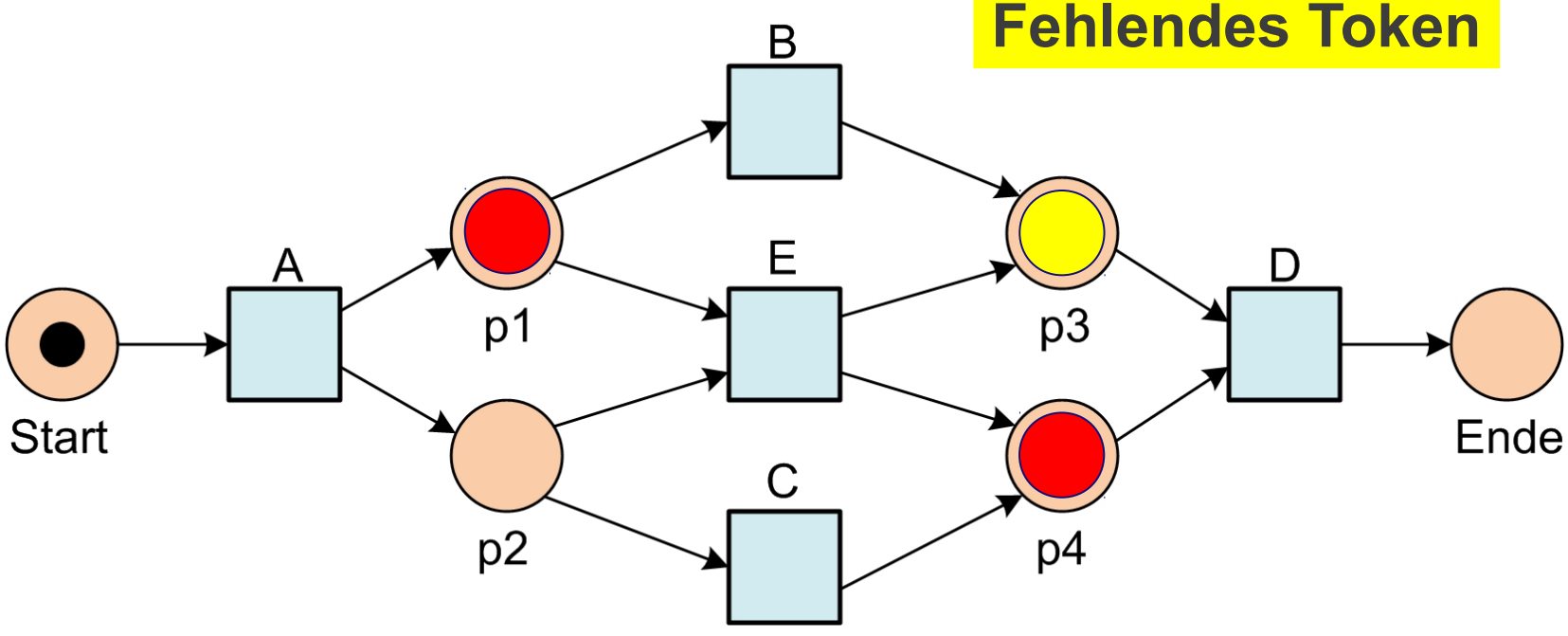
D



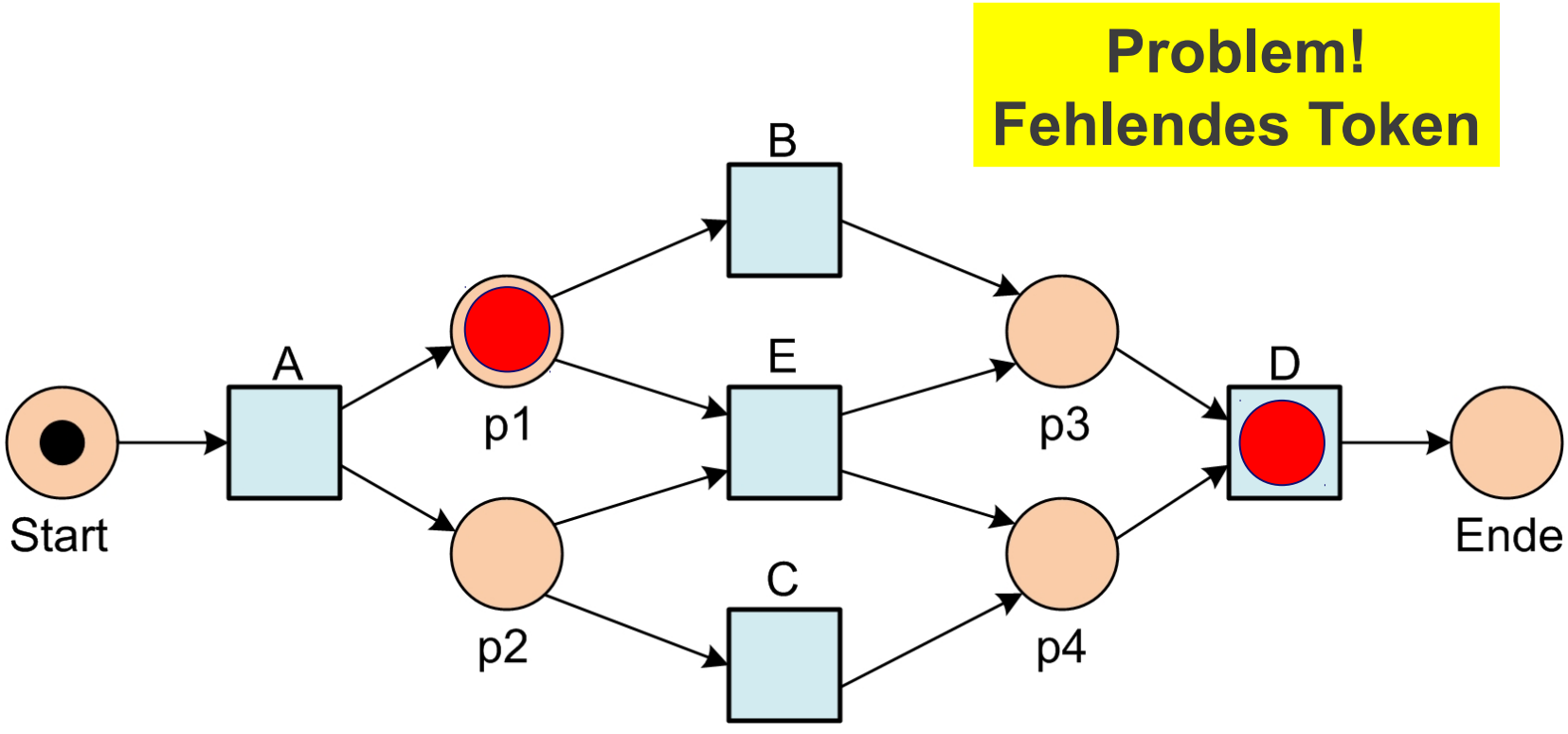
Replay: Mangelnde Konformanz aufdecken

D

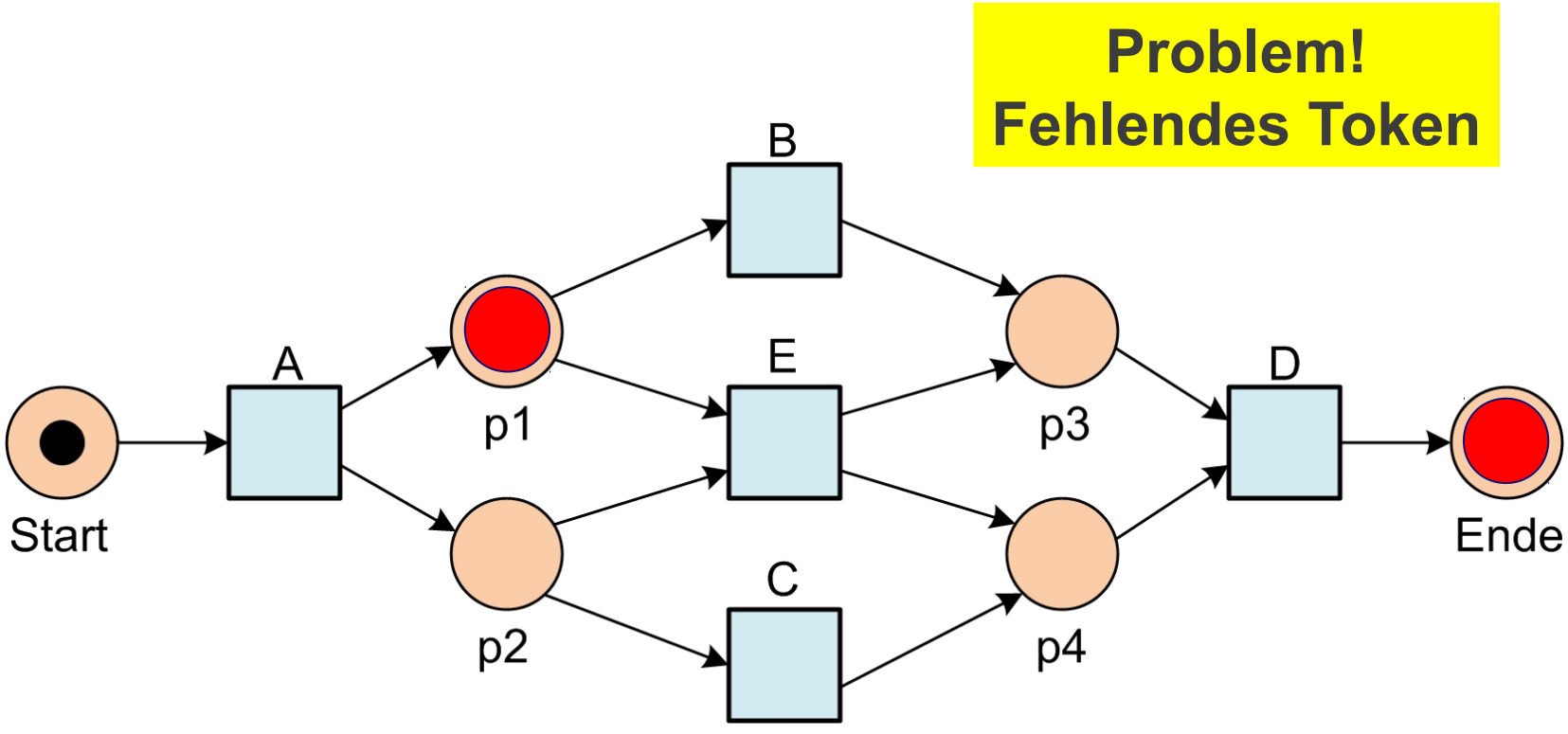
Problem!
Fehlendes Token



Replay: Mangelnde Konformanz aufdecken



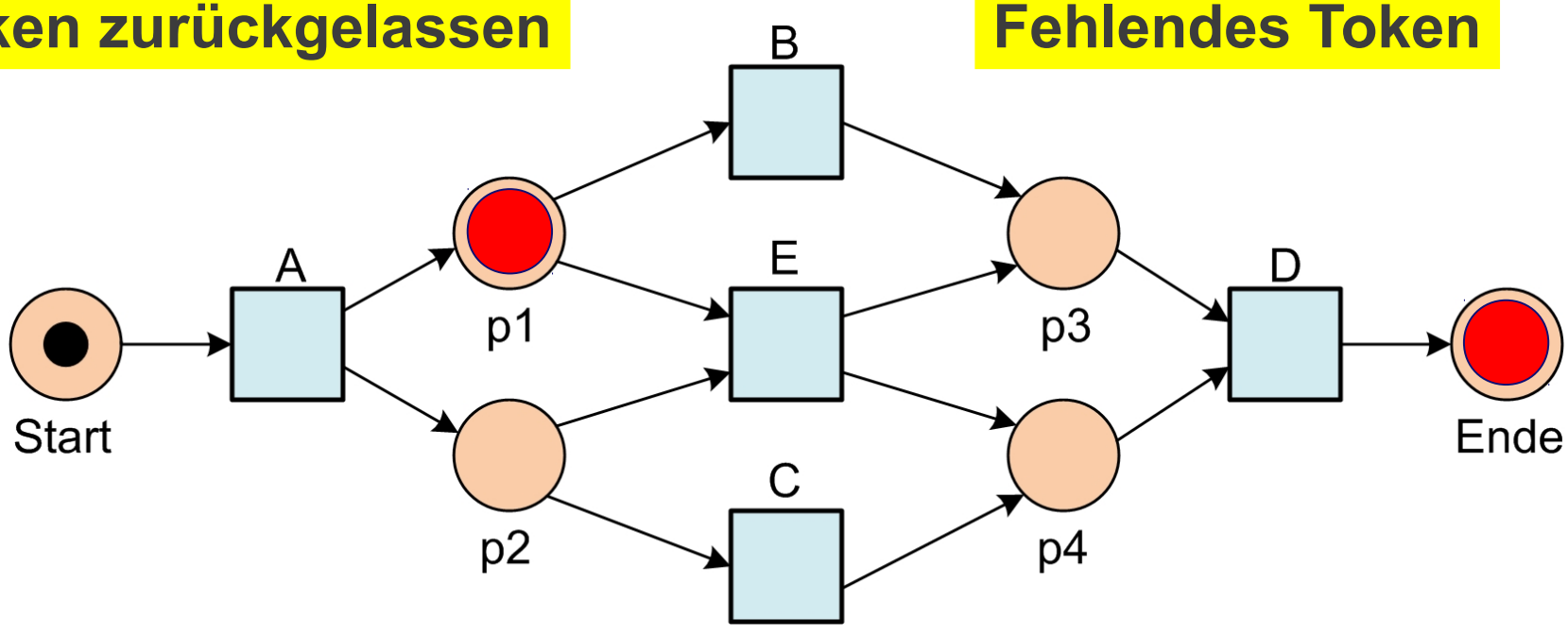
Replay: Mangelnde Konformanz aufdecken



Replay: Mangelnde Konformanz aufdecken

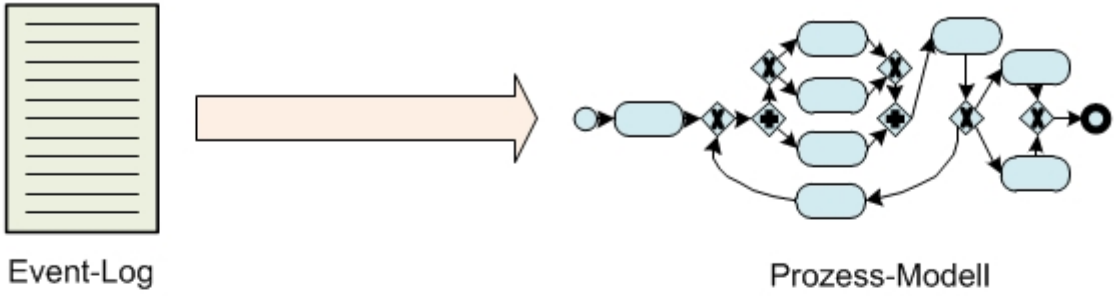
Problem!
Token zurückgelassen

Problem!
Fehlendes Token

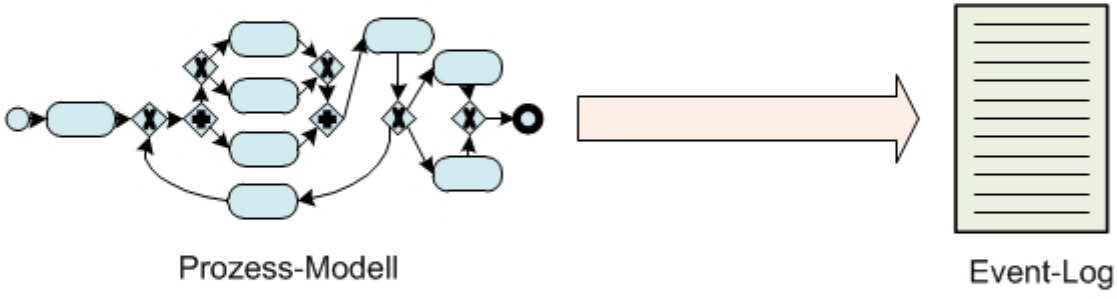


Konformanzanalyse durch Replay

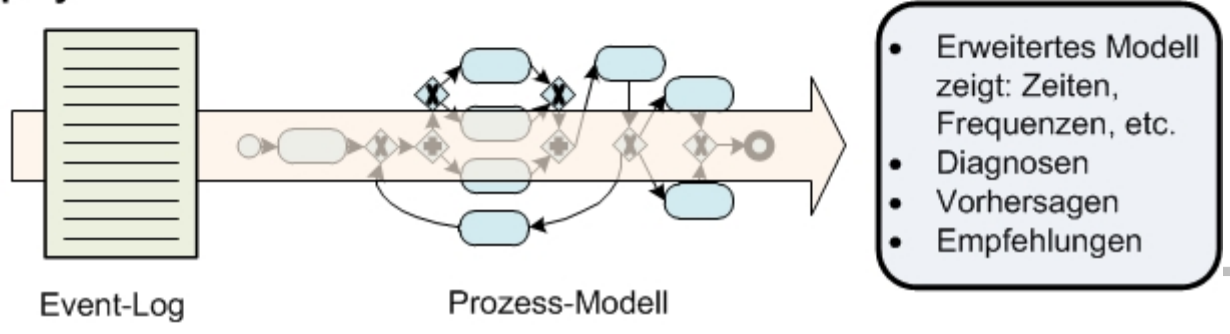
Play-In



Play-Out



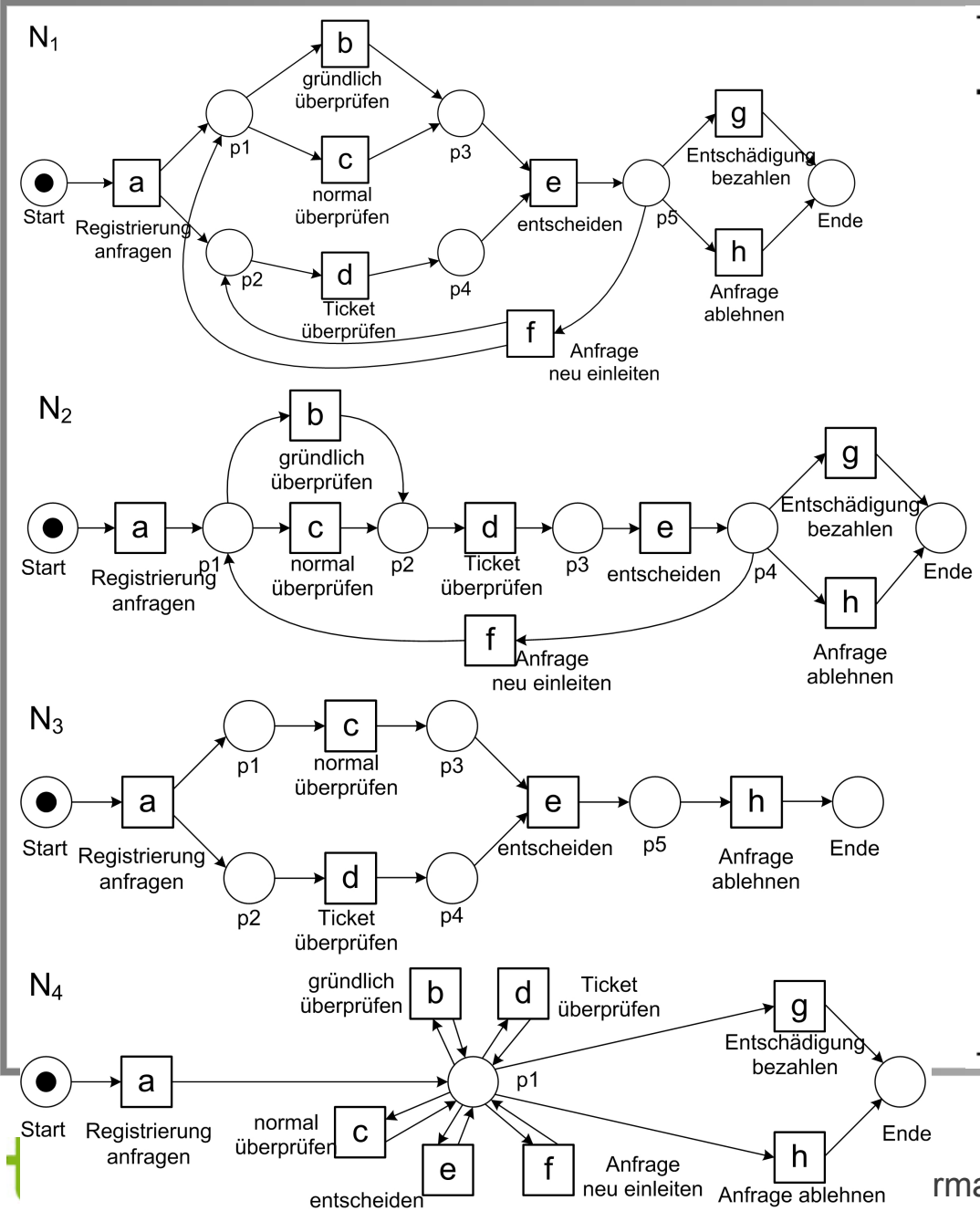
Replay



Überblick
2.6 Konformanzanalyse

- Einführung
- Konformanzanalyse basierend auf Replay
  - Motivation
  - Beispiele
- Diagnose

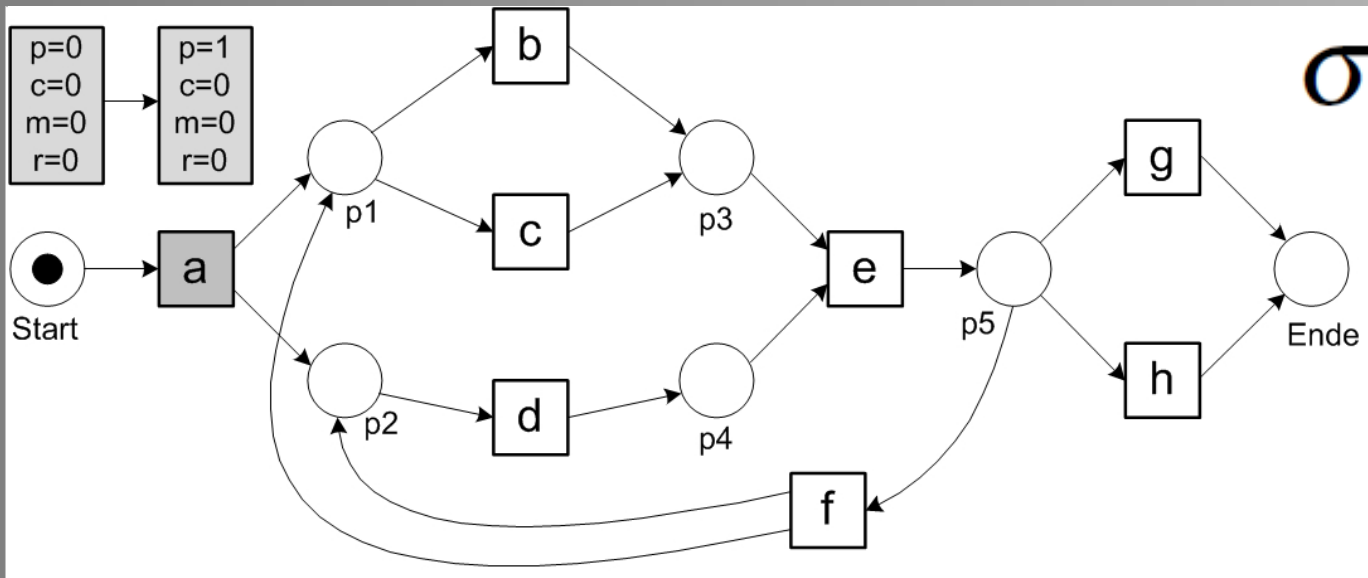
Vier Modelle, ein Log



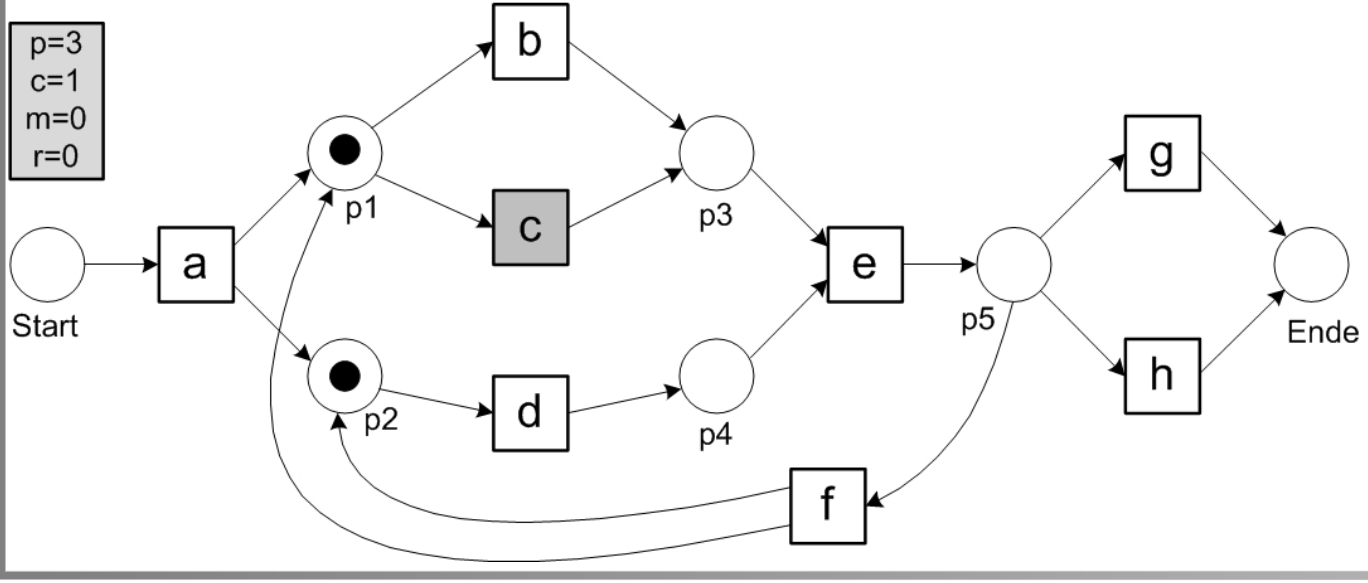
frequency	reference	trace
455	σ_1	$\langle a,c,d,e,h\rangle$
191	σ_2	$\langle a,b,d,e,g\rangle$
177	σ_3	$\langle a,d,c,e,h\rangle$
144	σ_4	$\langle a,b,d,e,h\rangle$
111	σ_5	$\langle a,c,d,e,g\rangle$
82	σ_6	$\langle a,d,c,e,g\rangle$
56	σ_7	$\langle a,d,b,e,h\rangle$
47	σ_8	$\langle a,c,d,e,f,d,b,e,h\rangle$
38	σ_9	$\langle a,d,b,e,g\rangle$
33	σ_{10}	$\langle a,c,d,e,f,b,d,e,h\rangle$
14	σ_{11}	$\langle a,c,d,e,f,b,d,e,g\rangle$
11	σ_{12}	$\langle a,c,d,e,f,d,b,e,g\rangle$
9	σ_{13}	$\langle a,d,c,e,f,c,d,e,h\rangle$
8	σ_{14}	$\langle a,d,c,e,f,d,b,e,h\rangle$
5	σ_{15}	$\langle a,d,c,e,f,b,d,e,g\rangle$
3	σ_{16}	$\langle a,c,d,e,f,b,d,e,f,d,b,e,g\rangle$
2	σ_{17}	$\langle a,d,c,e,f,d,b,e,g\rangle$
2	σ_{18}	$\langle a,d,c,e,f,b,d,e,f,b,d,e,g\rangle$
1	σ_{19}	$\langle a,d,c,e,f,d,b,e,f,b,d,e,h\rangle$
1	σ_{20}	$\langle a,d,b,e,f,b,d,e,f,d,b,e,g\rangle$
1	σ_{21}	$\langle a,d,c,e,f,d,b,e,f,c,d,e,f,d,b,e,g\rangle$

Replaying σ_1 auf N_1 (1/3)

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



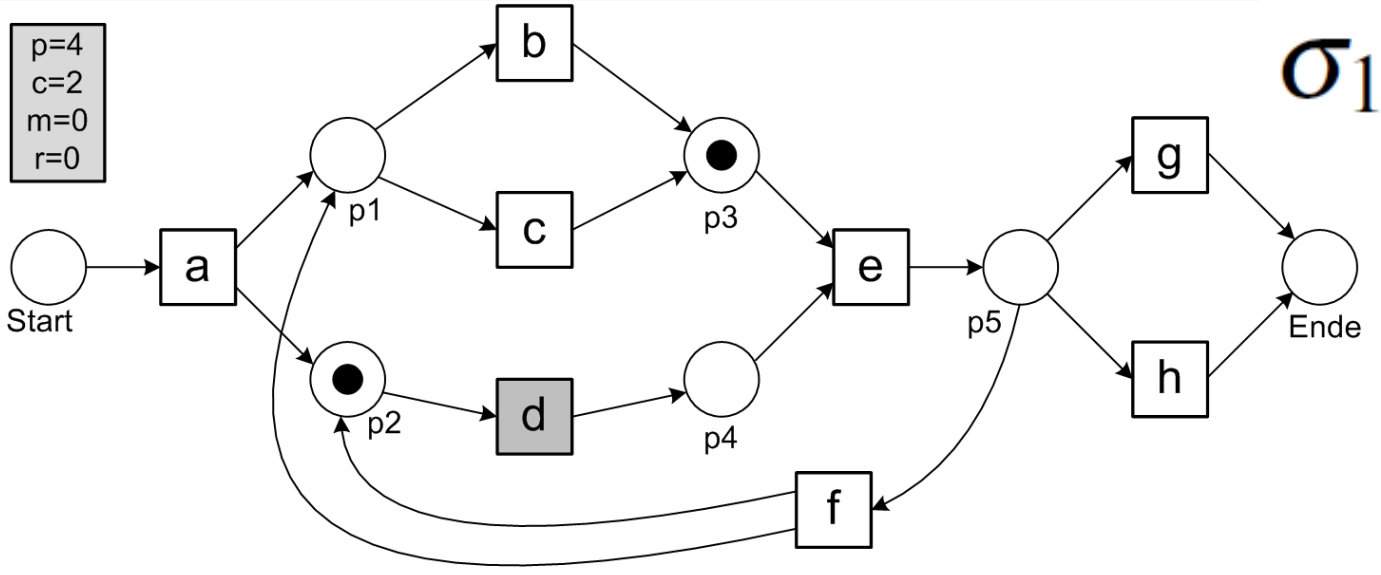
- 1 Token **p**roduziert (Start)
- 0 Token **c**onsumiert, **m**issing (fehlt) oder **r**emaining (übrig)
- a aktiv gemäß Log

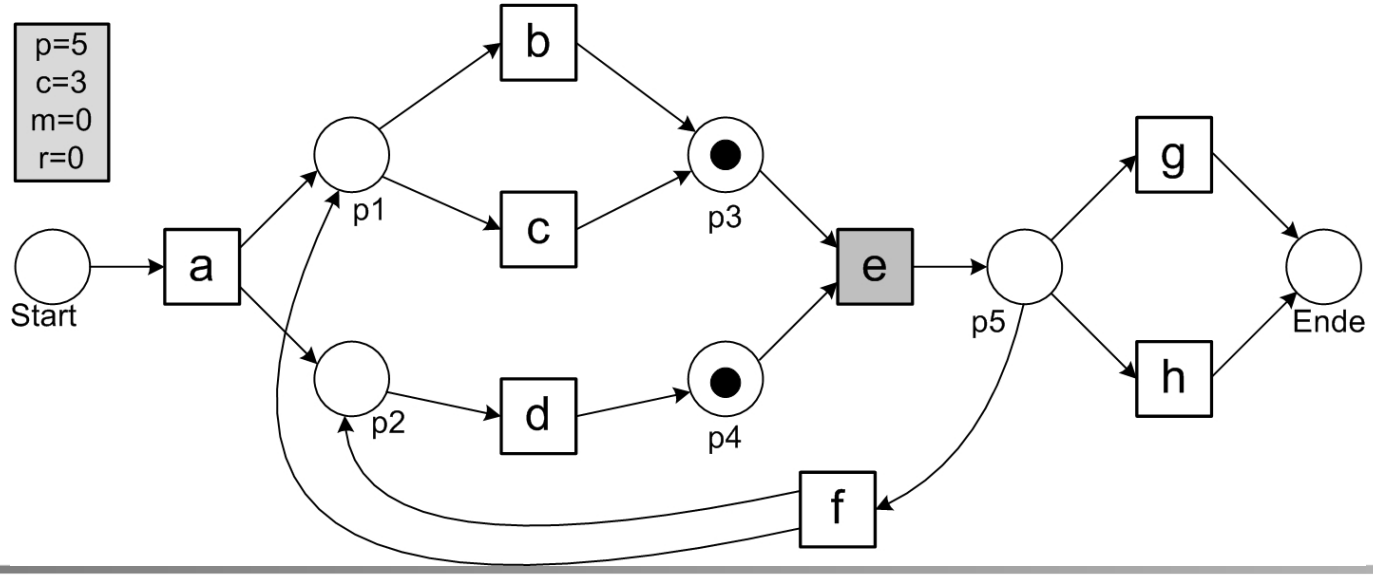


- 3 Token **p**roduziert (Start, p1, p2)
- 1 Token **c**onsumiert (Start)
- 0 Token **m**issing (fehlt) oder **r**emaining (übrig)
- c aktiv gemäß Log

Replaying σ_1 auf N_1 (2/3)

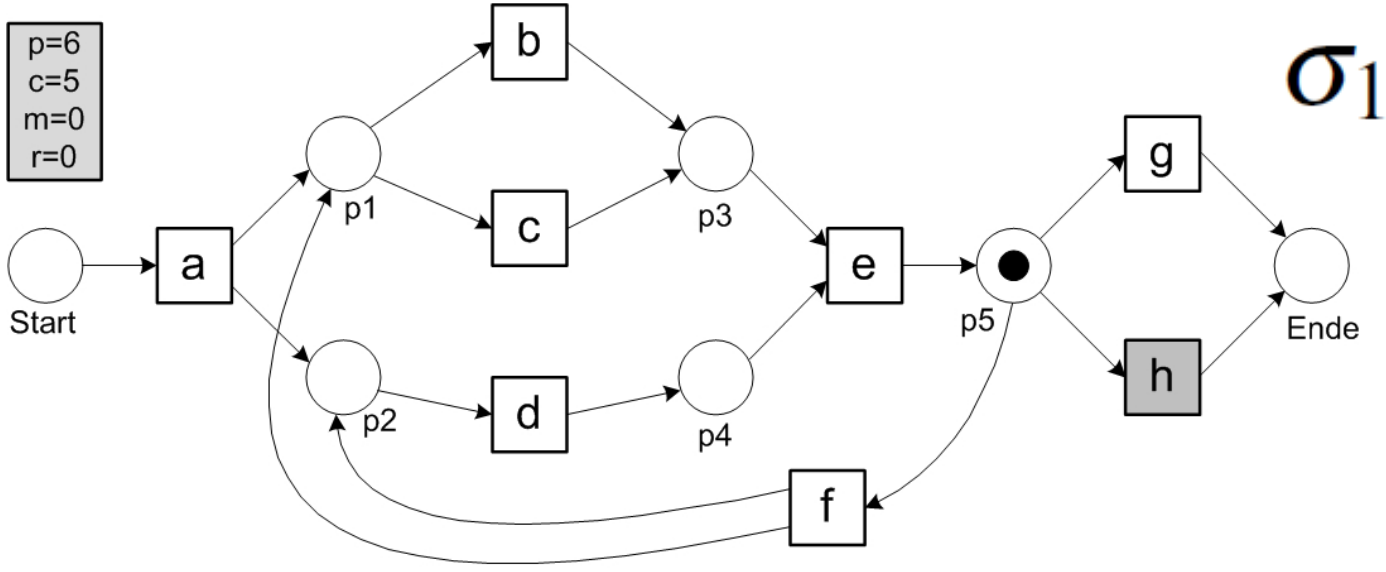
$$\sigma_1 = \langle a, c, d, e, h \rangle$$

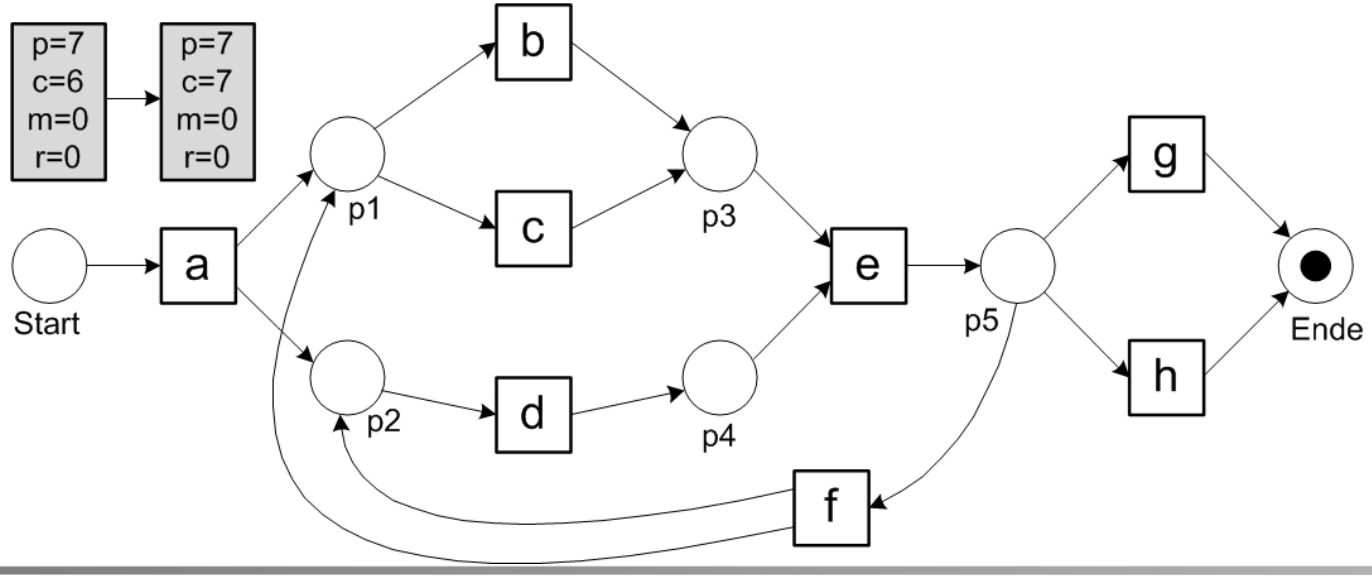




Replaying σ_1 auf N_1 (3/3)

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

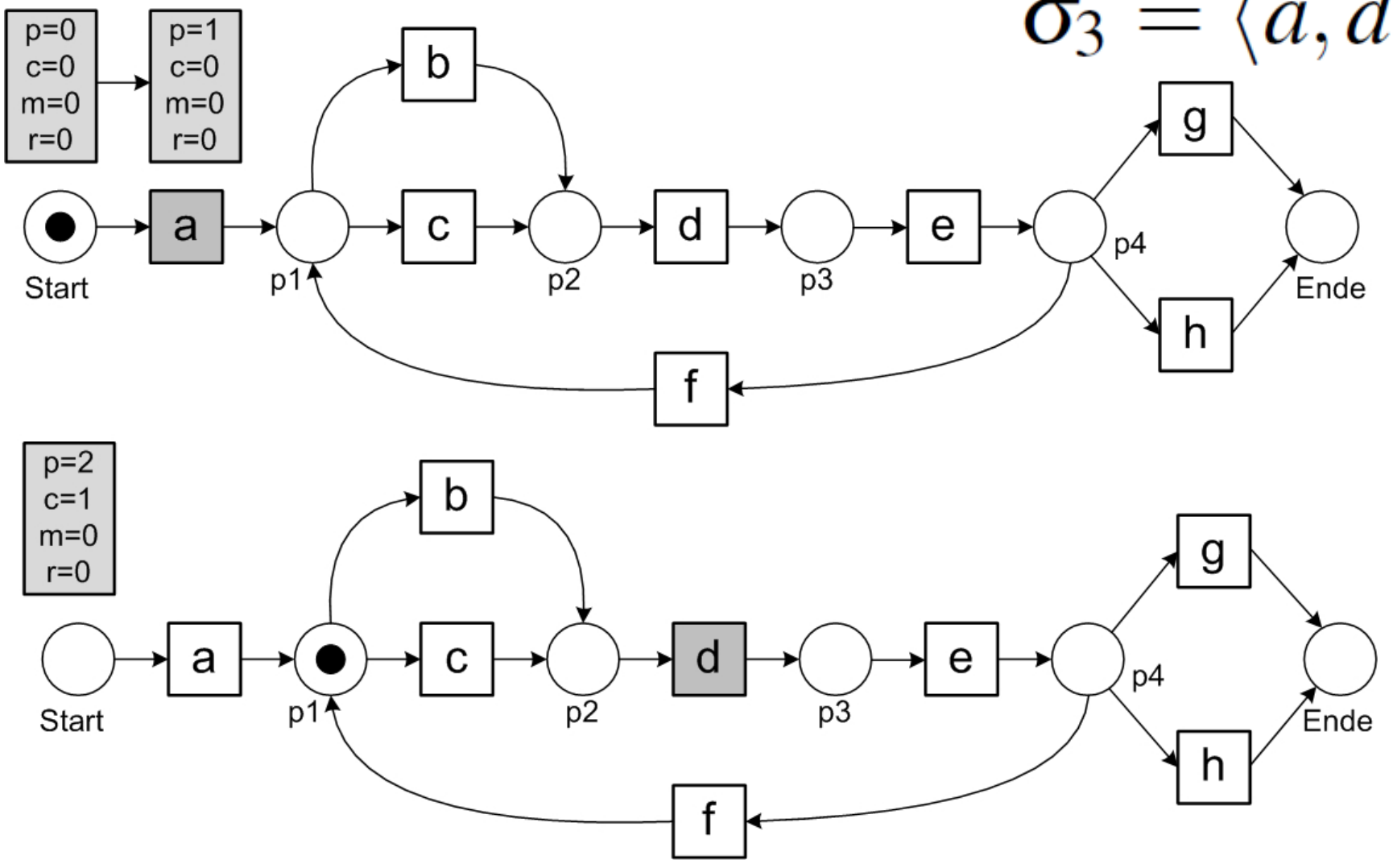




Keine Probleme gefunden

Replaying σ_3 auf N_2 (1/3)

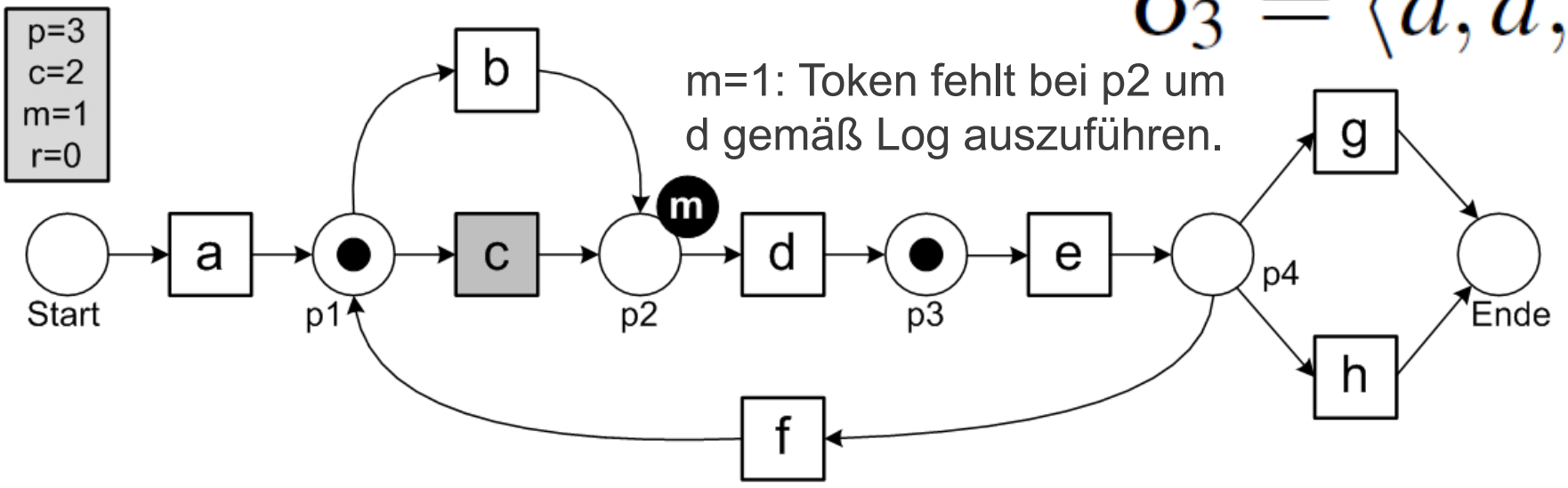
$$\sigma_3 = \langle a, d, c, e, h \rangle$$

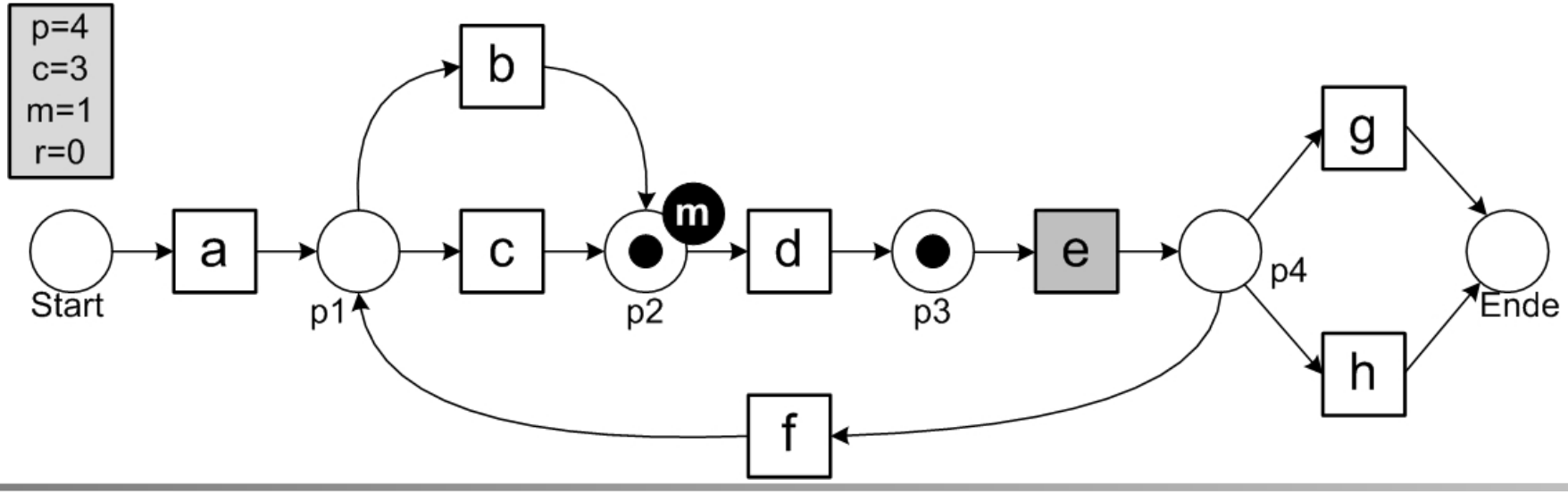


Replaying σ_3 auf N_2 (2/3)

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

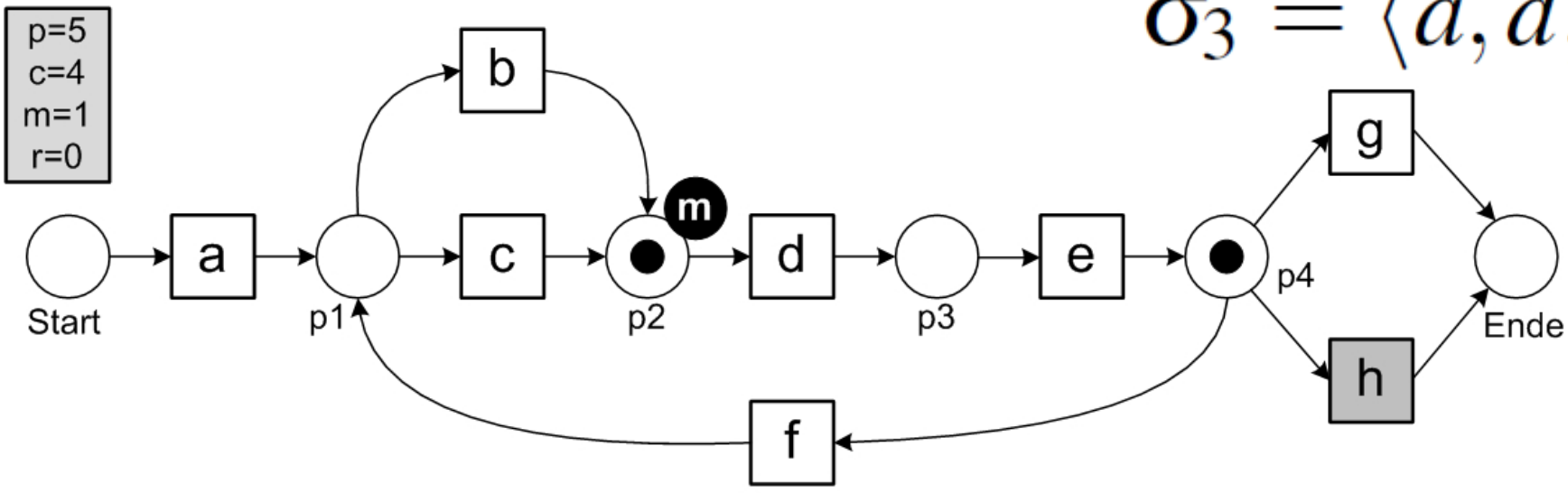
m=1: Token fehlt bei p2 um d gemäß Log auszuführen.

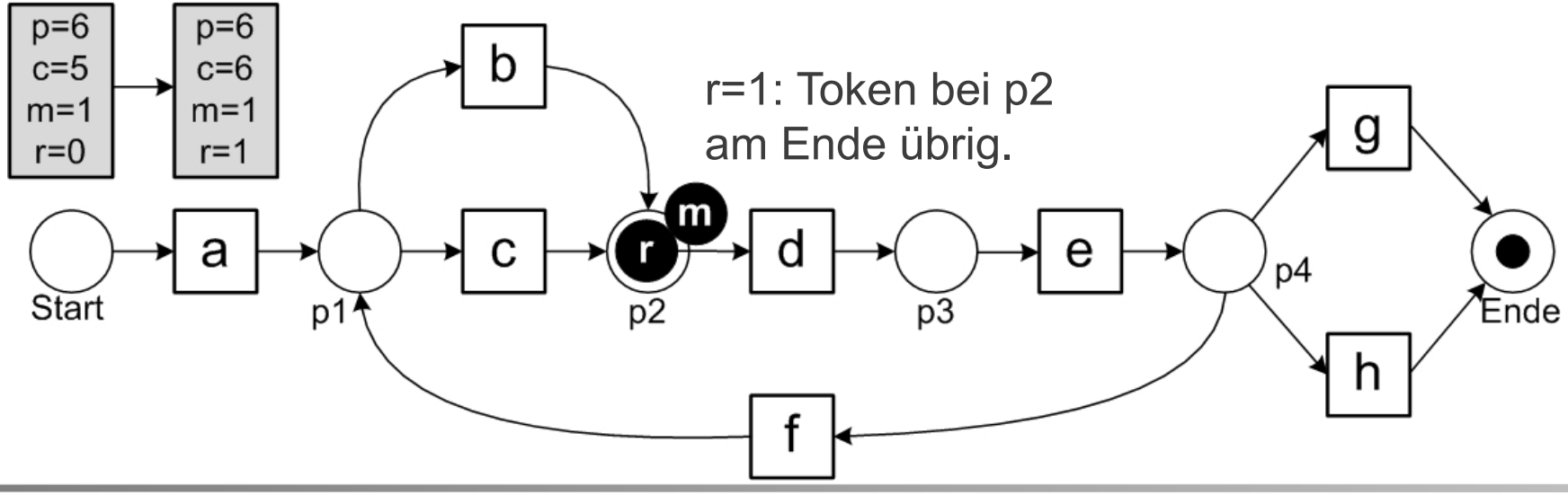




Replaying σ_3 auf N_2 (3/3)

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

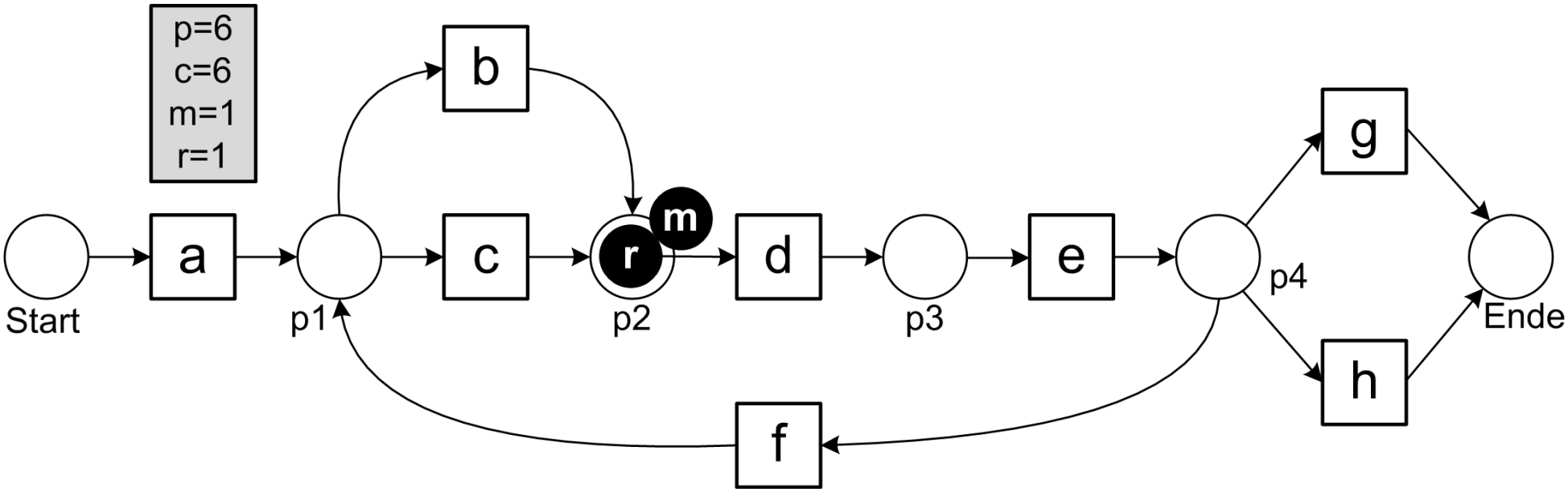




r=1: Token bei p2 am Ende übrig.

Probleme beim Replay von σ_3 auf N_2

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

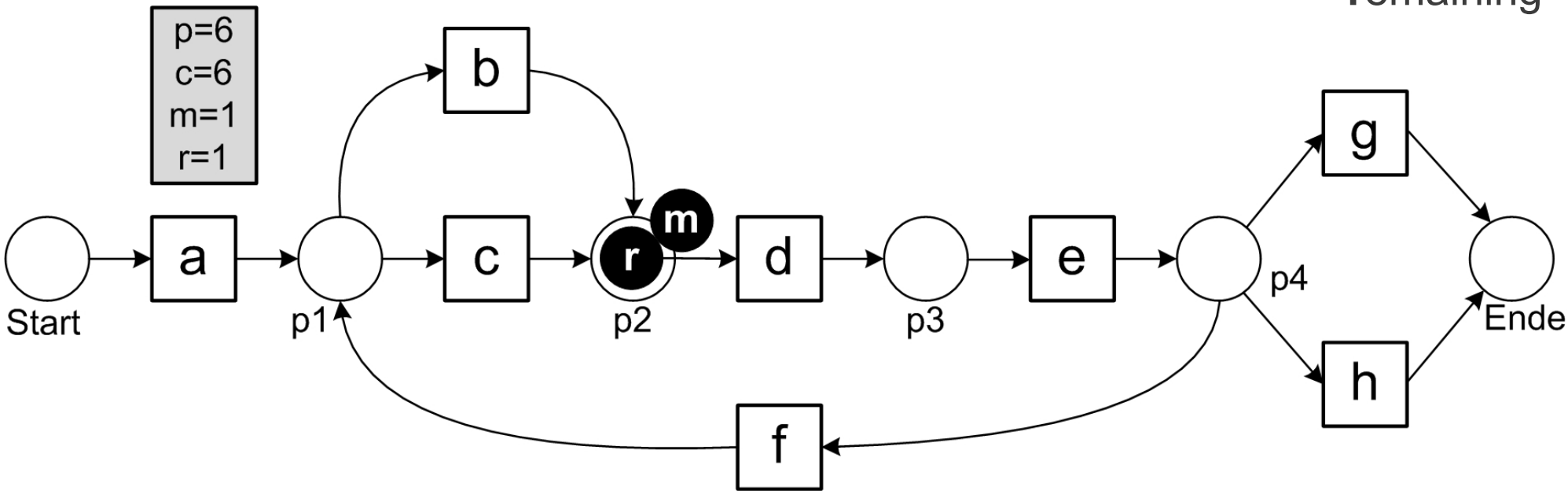


- Ein fehlender Token (von 6 aufgenommenen Token)
- Ein Token bleibt übrig (von 6 erzeugten Token)

Angemessenheit (Fitness) des Modelles zu Log-Folge

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

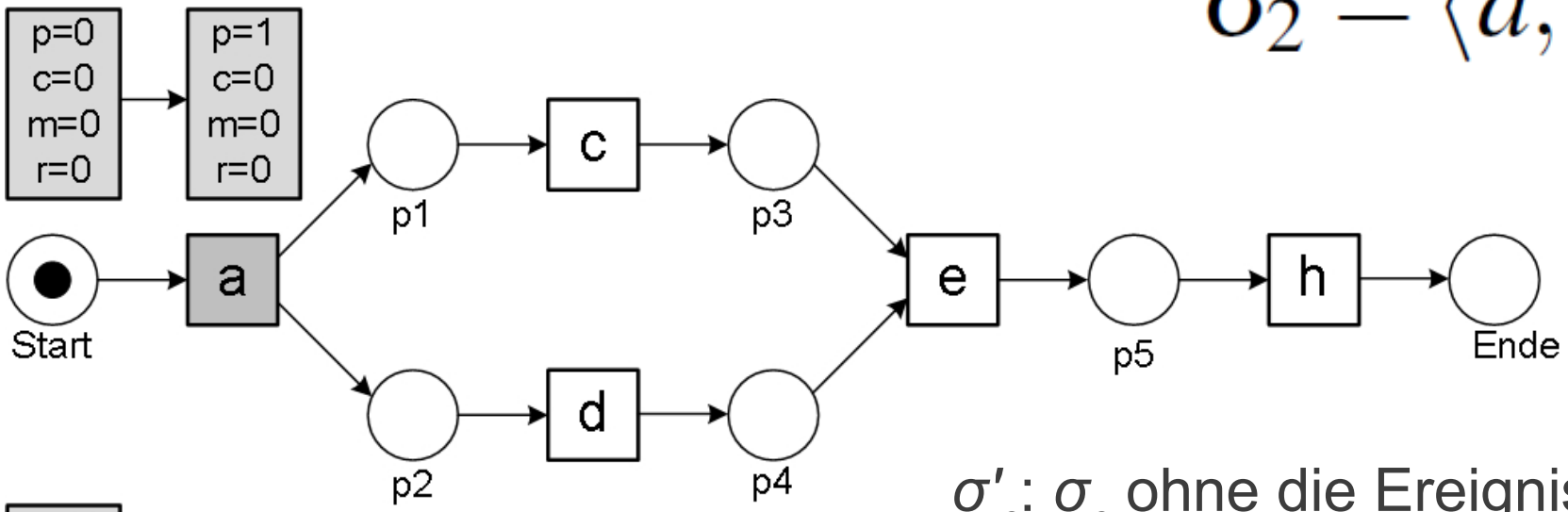
produced,
consumed,
missing,
remaining



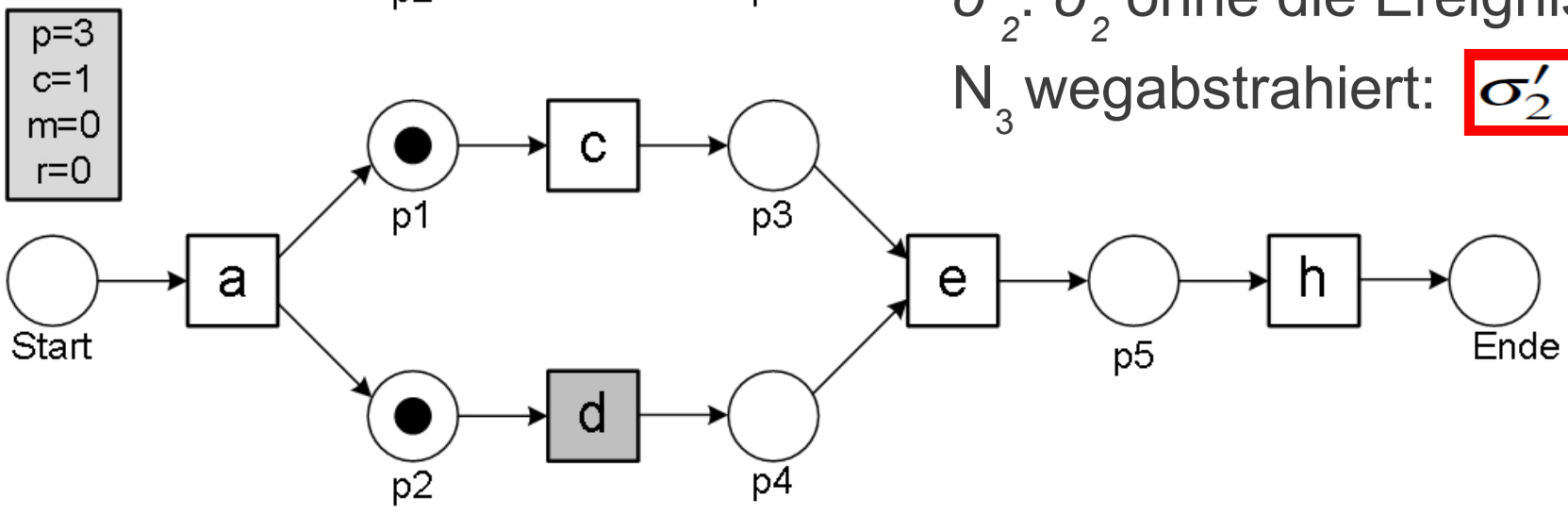
$$fitness(\sigma_3, N_2) = \frac{1}{2}\left(1 - \frac{1}{6}\right) + \frac{1}{2}\left(1 - \frac{1}{6}\right) = 0.8333$$

Replaying σ_2 auf N_3 (1/3)

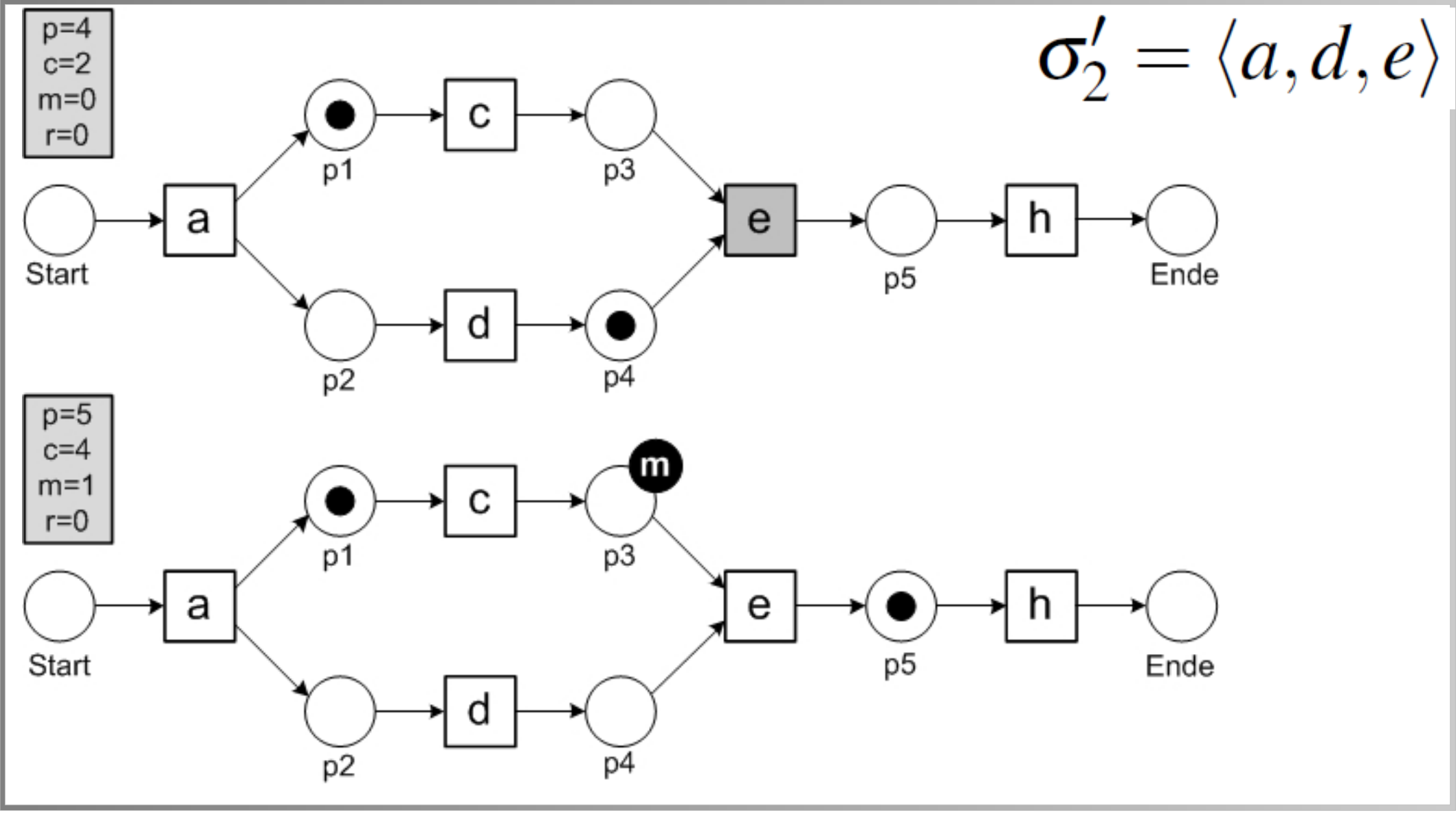
$$\sigma_2 = \langle a, b, d, e, g \rangle$$



σ'_2: σ_2 ohne die Ereignisse, die in N_3 wegabstrahiert: $\boxed{\sigma'_2 = \langle a, d, e \rangle}$

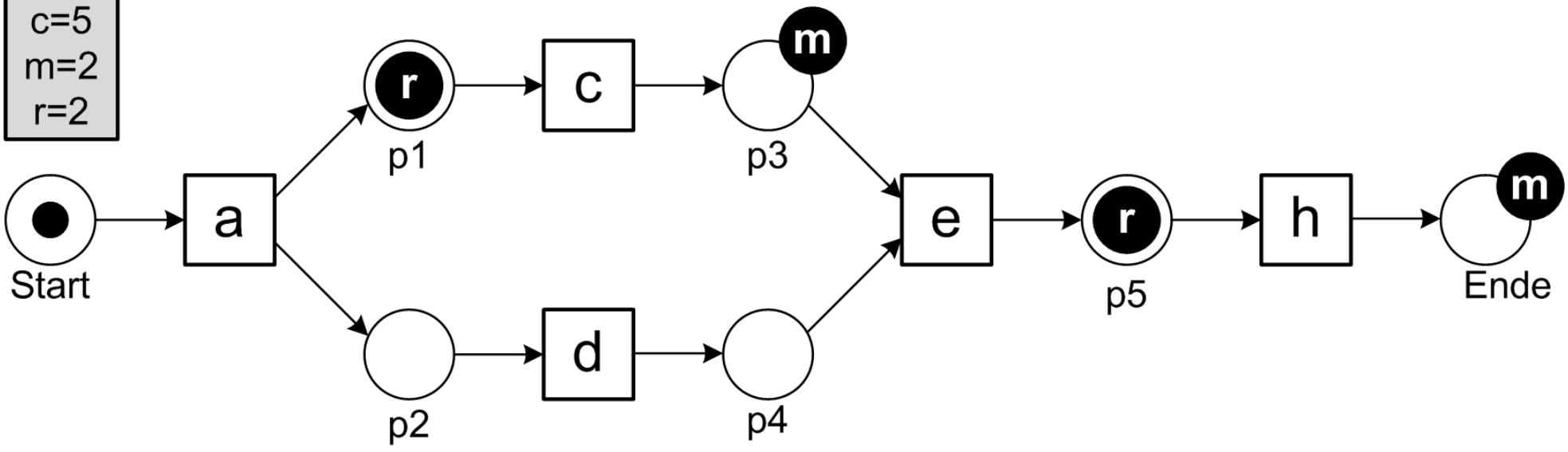


Replaying σ_2 auf N_3 (2/3)



Replaying σ_2 auf N_3 (3/3)

$$\sigma_2' = \langle a, d, e \rangle$$



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

Angemessenheit (Fitness) zu Menge von Log-Folgen

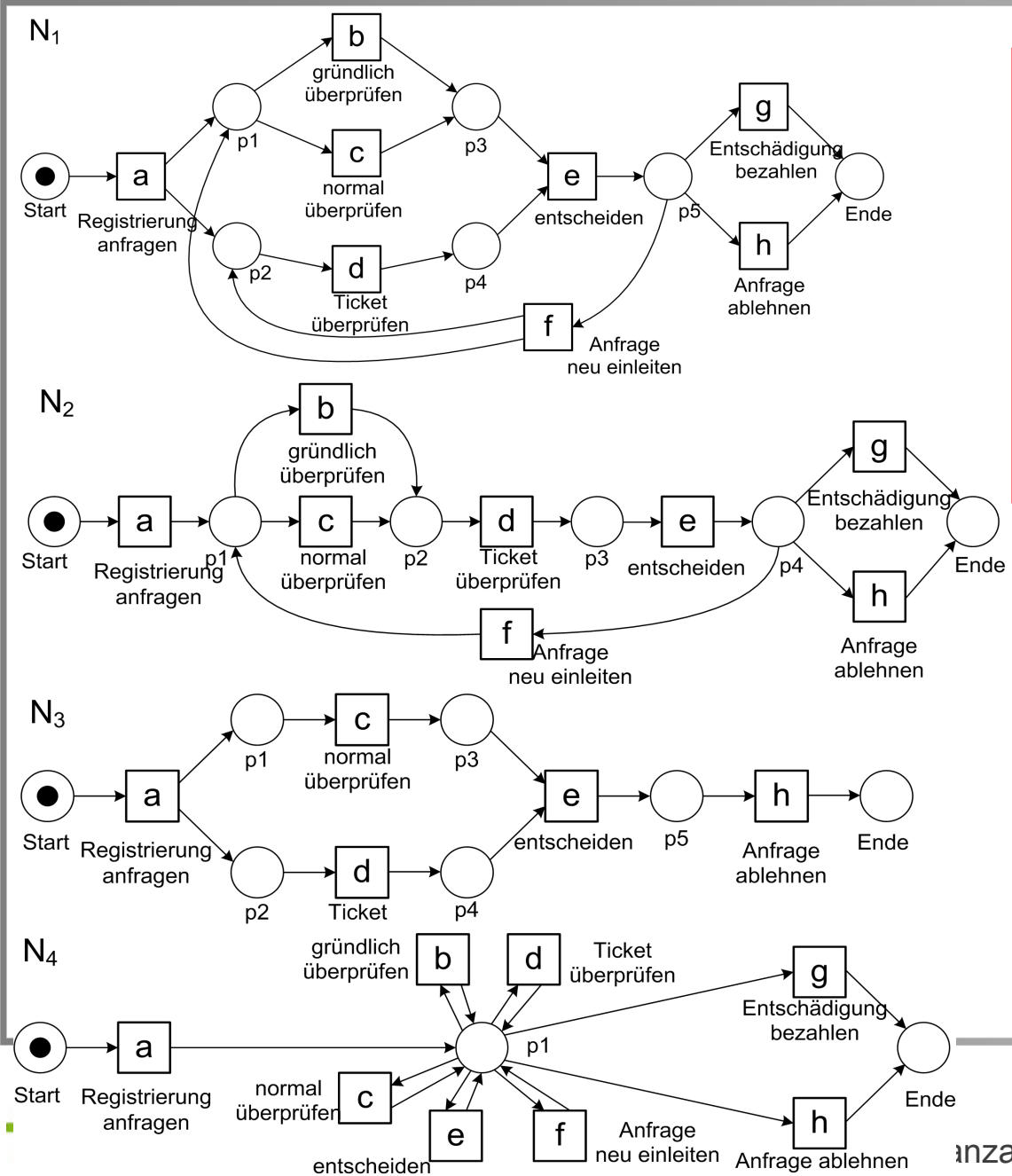
$$fitness(L,N) = \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}}\right) + \frac{1}{2}\left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}}\right)$$

N: Modell

$L(\sigma)$: Häufigkeit der Folge σ in Logdatei L.

produced, **c**onsumed, **m**issing, **r**emaining bei Replay σ von auf N.

Beispielwerte



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

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

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

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

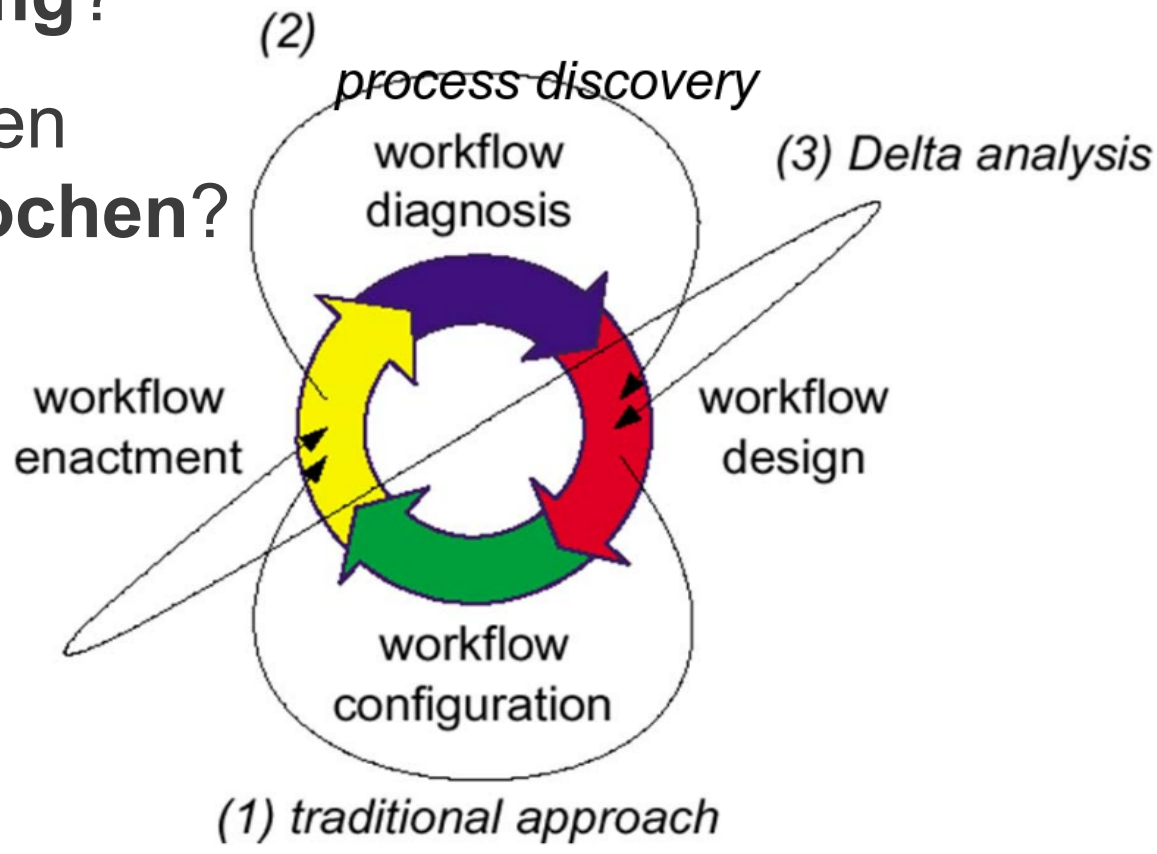
Überblick
2.6 Konformanzanalyse

- Einführung
- Konformanzanalyse basierend auf Replay
  - Motivation
  - Beispiele
- Diagnose

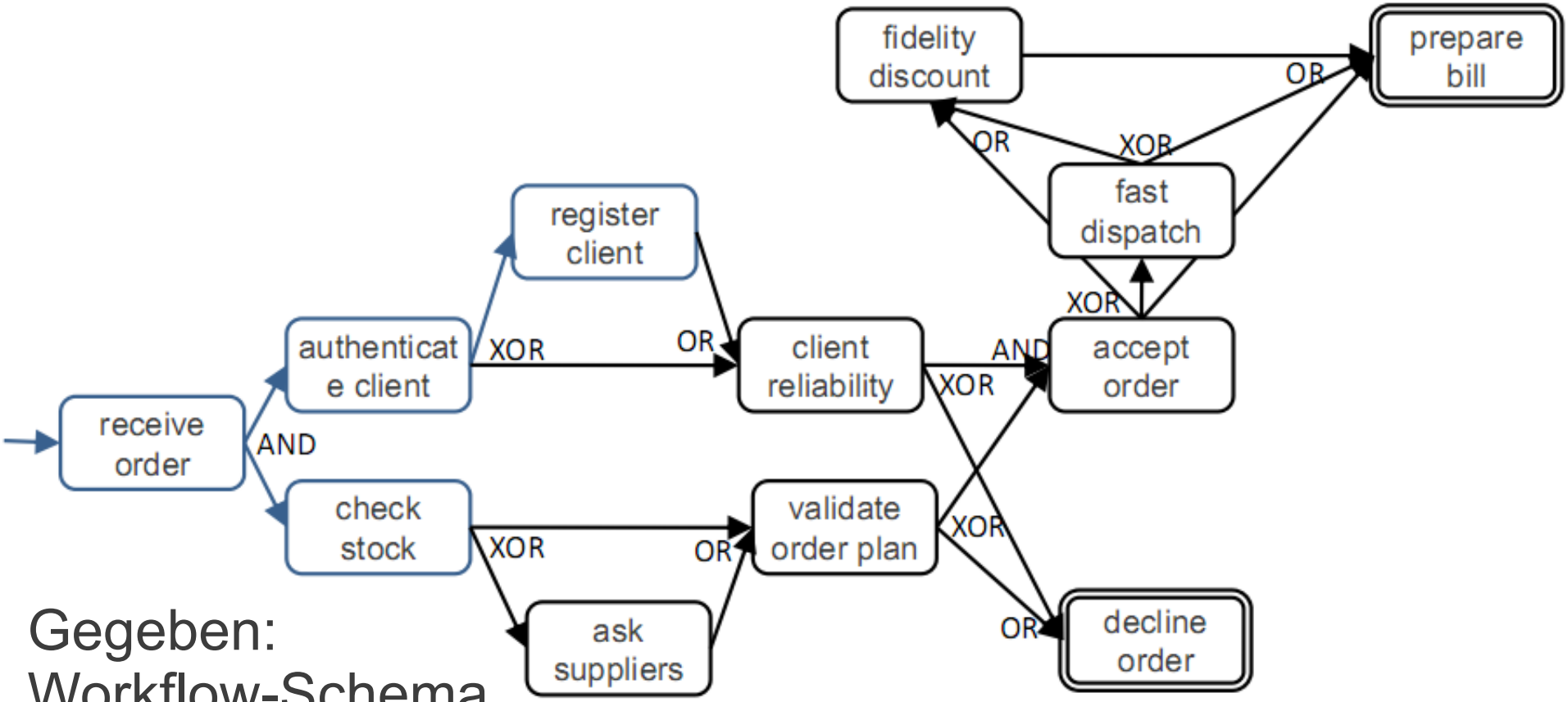
Workflow-Diagnose

Nächster Schritt („workflow diagnosis“):

- Welche Pfade sind **häufig**?
- Welche Instanzen werden wahrscheinlich **abgebrochen**?



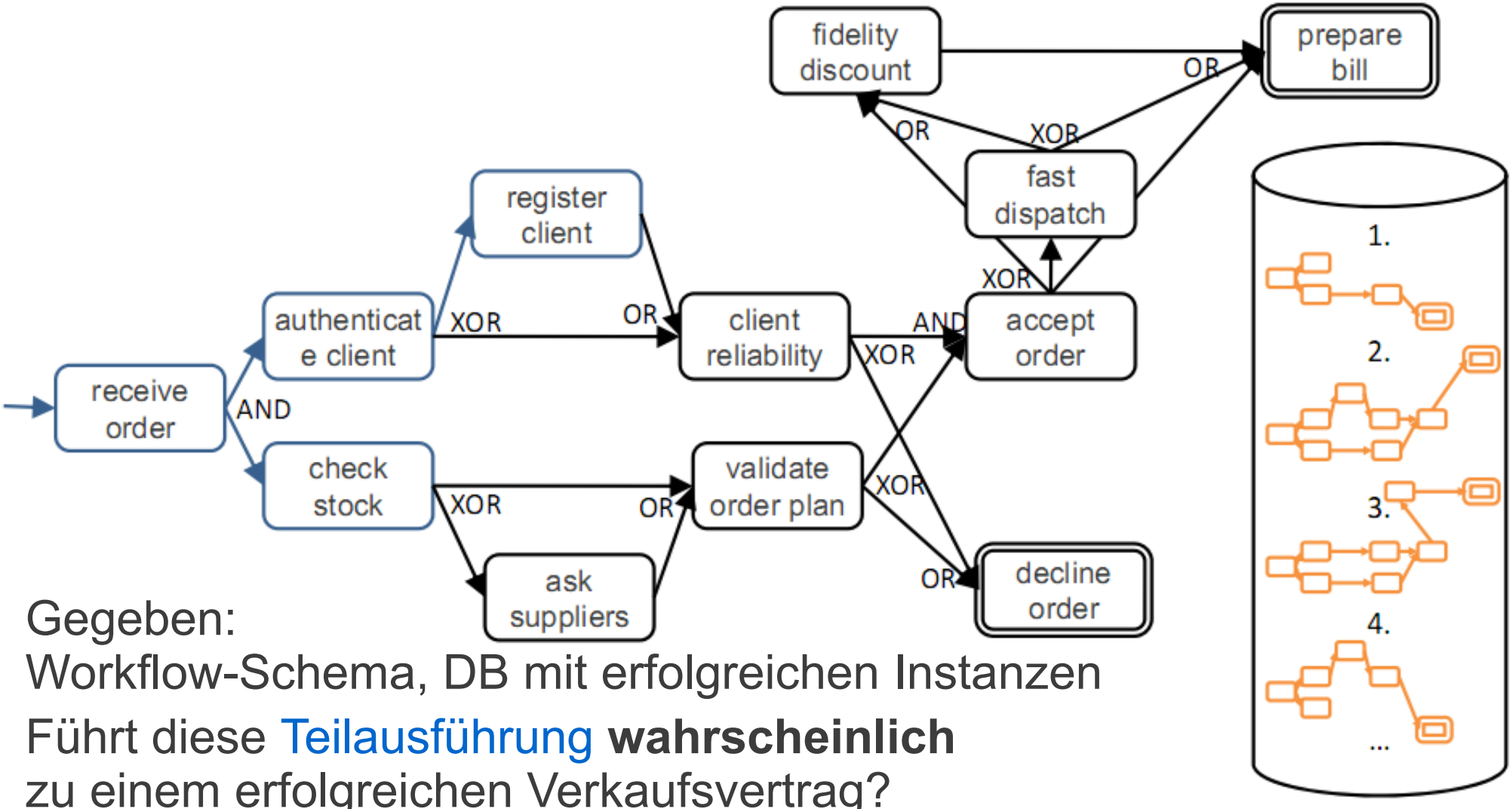
Workflow-Diagnose: Statische Analyse



Gegeben:
Workflow-Schema

Kann diese Teilausführung **möglicherweise** zu einem erfolgreichen Verkaufsvertrag führen ?

Workflow-Diagnose: Analyse der Laufzeitdaten



Gegeben:
Workflow-Schema, DB mit erfolgreichen Instanzen

Führt diese Teilausführung **wahrscheinlich** zu einem erfolgreichen Verkaufsvertrag?

Workflow-Diagnose

Weitere interessante Fragestellungen:

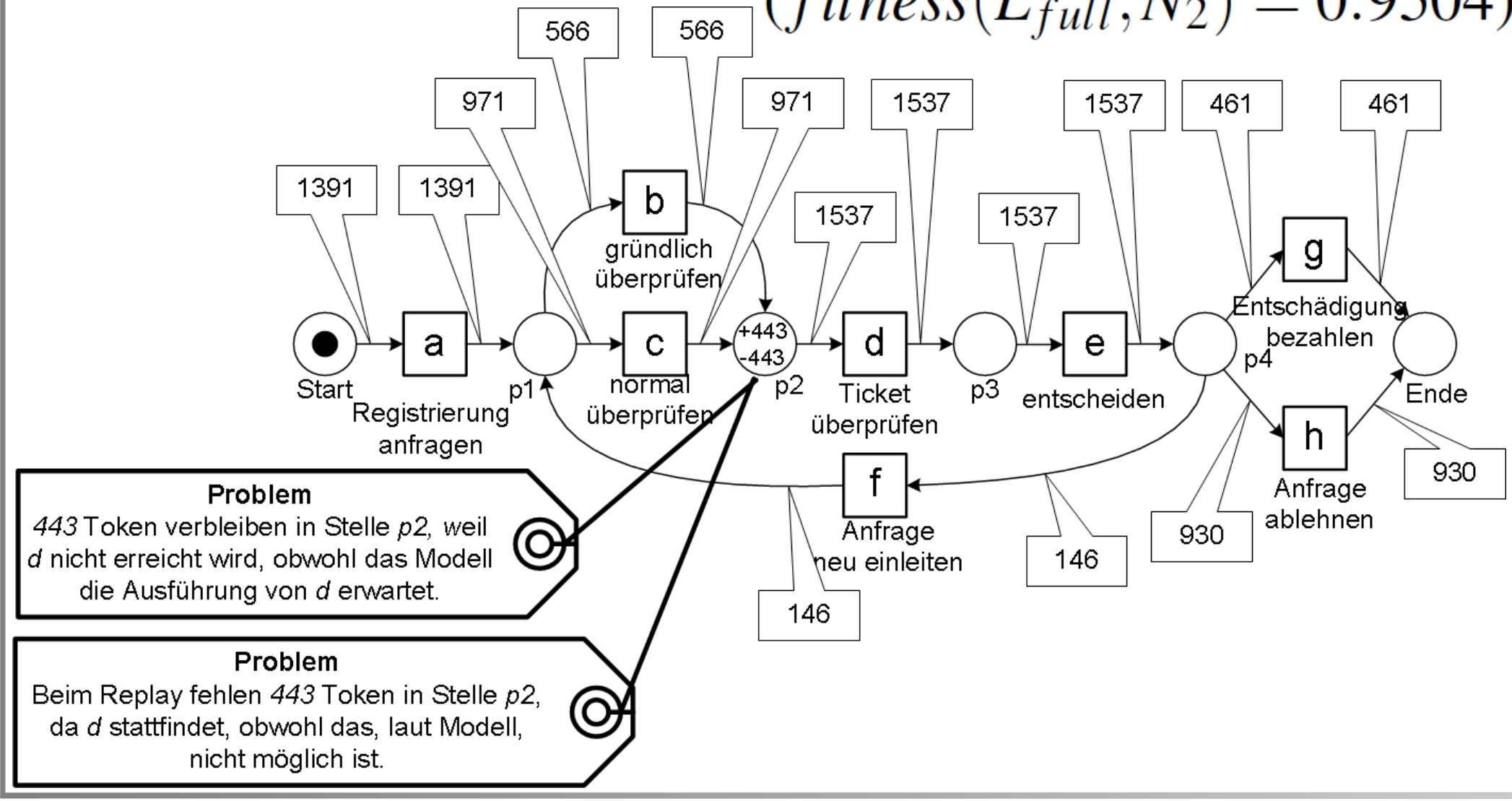
- **Fehlerfreie** Workflow-Ausführung ?
- Workflow-Ausführung mit geringem oder normalem **Ressourcenverbrauch** ?
- Identifikation **kritischer Aktivitäten** (besonders hoher Ressourcenverbrauch).
- Wie sehen **häufige / typische Ausführungen** aus?

Verwertung der Information

- Zur Laufzeit: **Scheduling**
- Information für Management / Optimierung (Business Process Reengineering (BPR), Continuous Process Improvement (CPI))

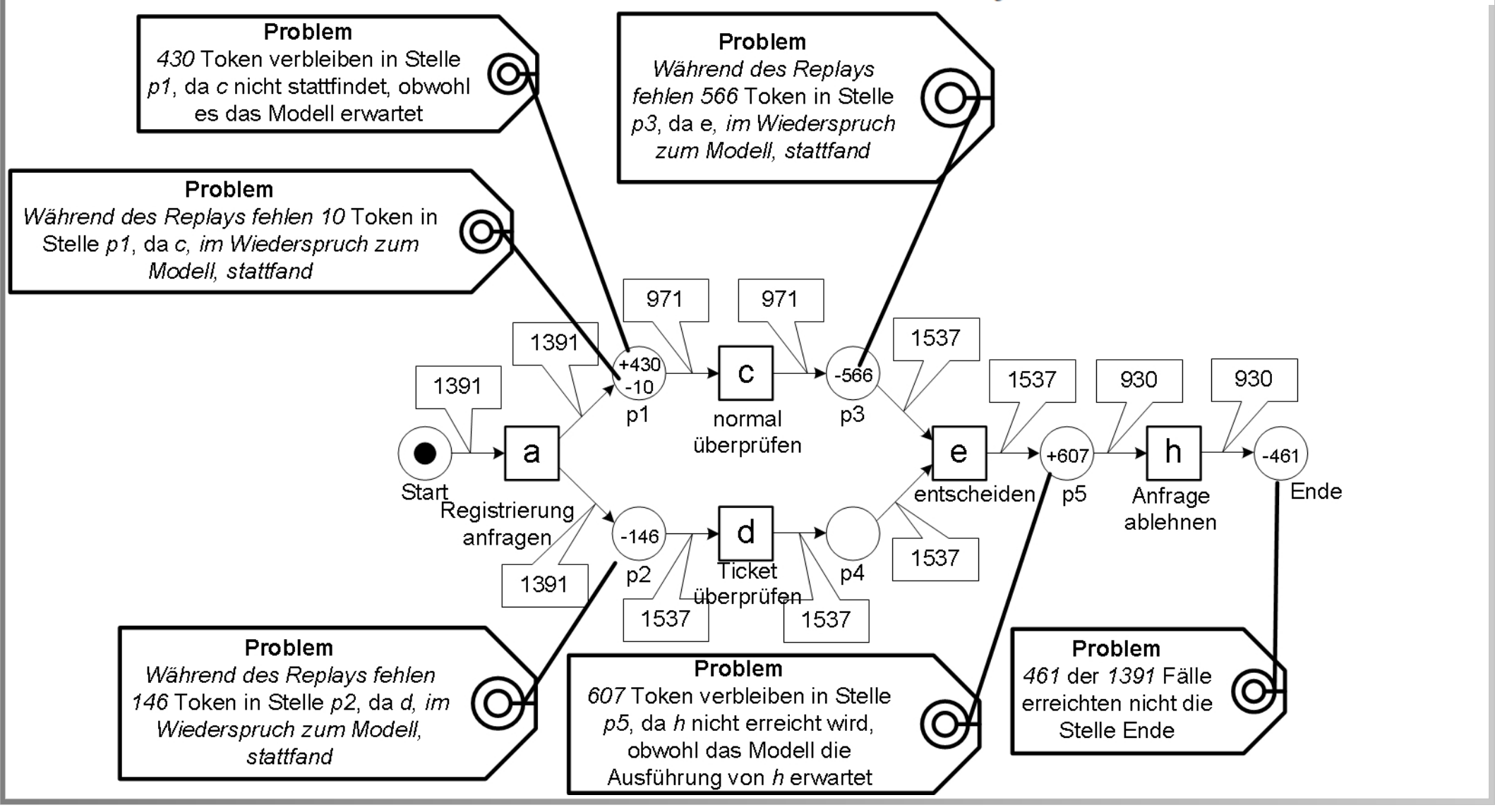
Diagnose

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

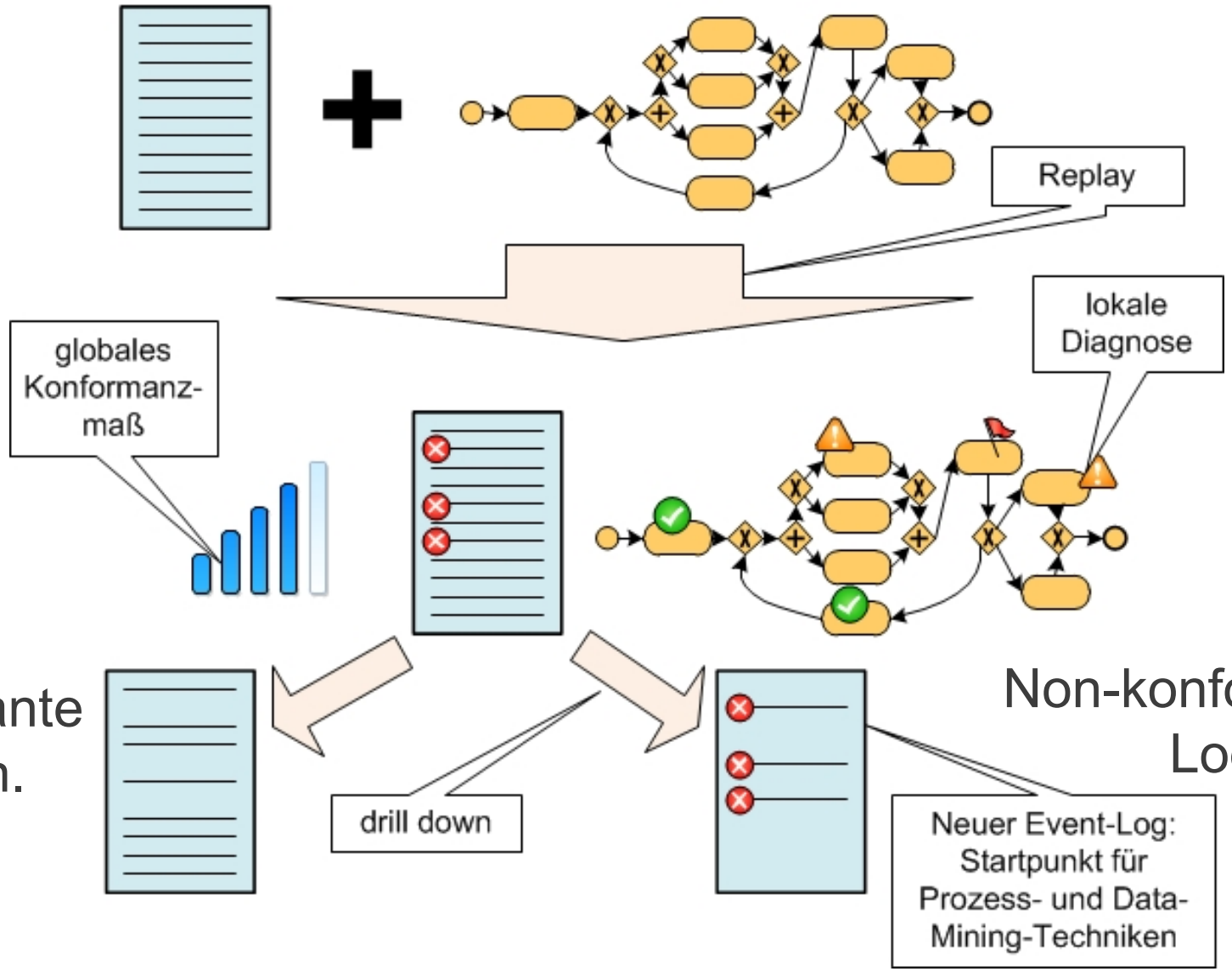


Diagnose

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



Drilling down (Verfeinerte Analyse)



Vergleiche Footprints

Vergleich der durch α-Algorithmus definierten **Relationen** (footprints), aus gegebenen **Logdaten** (z.B. L_{full}) bzw. aus **Petrinetz** (z.B. N_1) generierten Logdaten.

Für L_{full} und N_1 (identisch):

	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	#	#	#	#	$\leftarrow$	#	#	#

Vergleiche Footprints. Oben: N_2. Unten: L_{full} und N_1.

	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	$\#$	$\#$	$\#$	$\#$	$\leftarrow$	$\#$	$\#$	$\#$

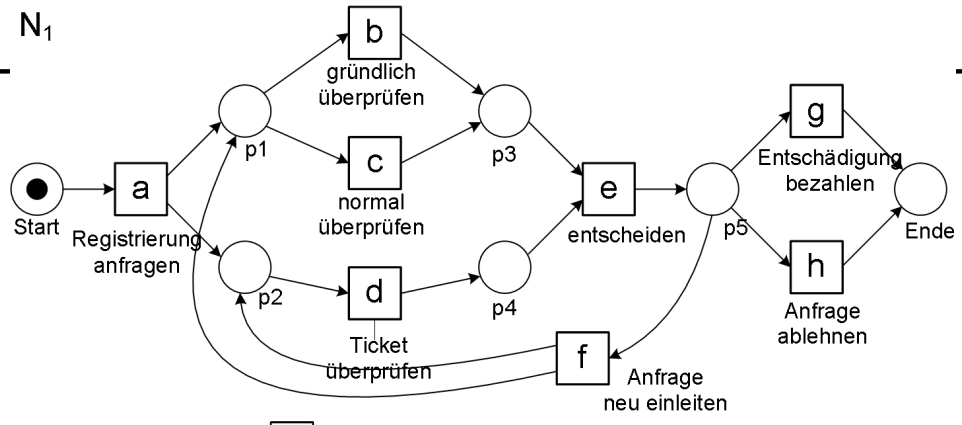
	a	b	c	d	e	f	g	h
a	$\#$	$\rightarrow$	$\rightarrow$	$\rightarrow$	$\#$	$\#$	$\#$	$\#$
b	$\leftarrow$	$\#$	$\#$	$\parallel$	$\rightarrow$	$\leftarrow$	$\#$	$\#$
c	$\leftarrow$	$\#$	$\#$	$\parallel$	$\rightarrow$	$\leftarrow$	$\#$	$\#$
d	$\leftarrow$	$\parallel$	$\parallel$	$\#$	$\rightarrow$	$\leftarrow$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

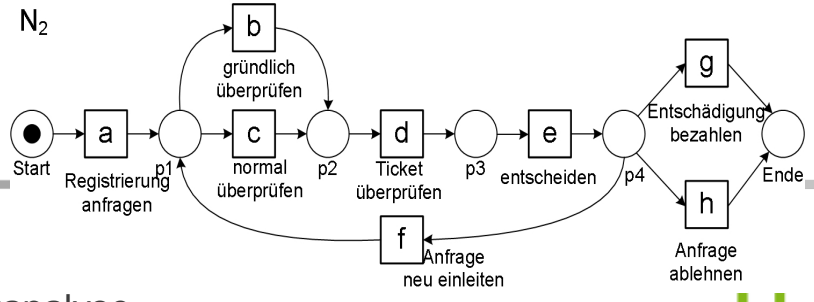
Diagnose: Unterschiede quantifiziert

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

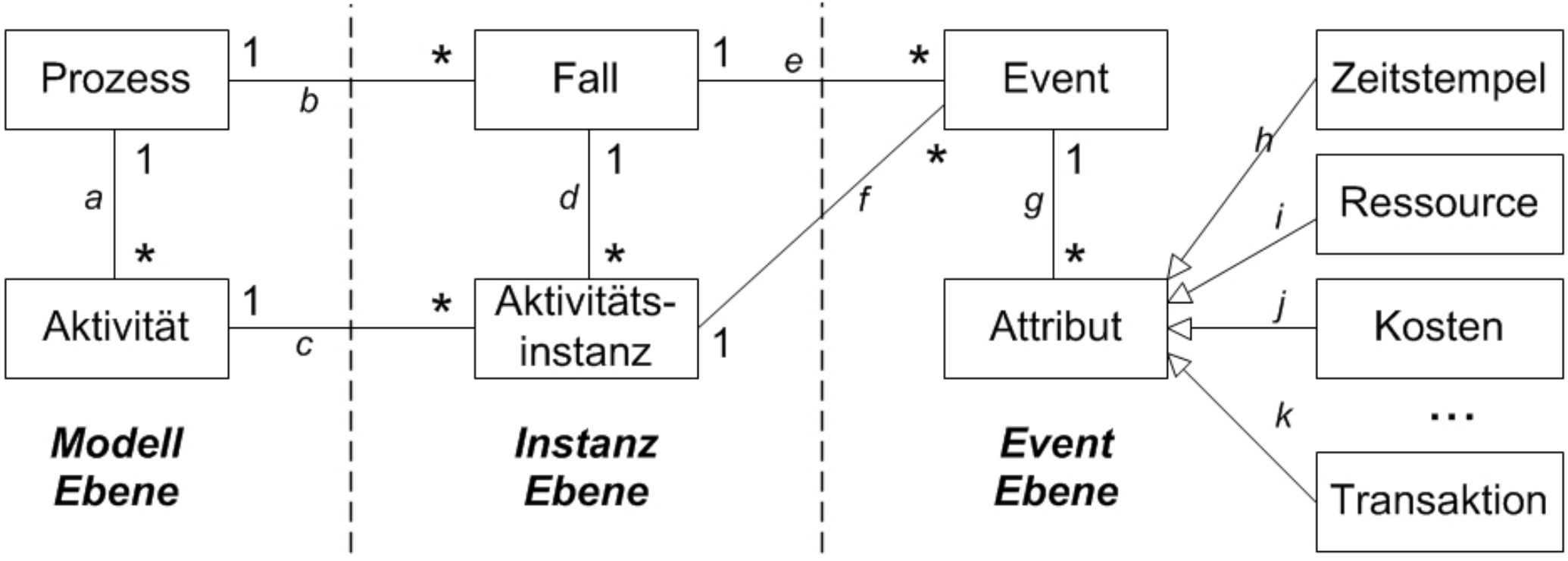
(x:y x ist in Log und y in N_2)

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





Event-Log und Modell verknüpfen



- Sehr wichtig!
- **Modell extrahiert** oder **manuell** erstellt.
- Während des **Replays** verknüpft.
- Startpunkt für andere Typen des Process-Minings!

Zusammenfassung
2.6 Konformanzanalyse

- Einführung
- Konformanzanalyse basierend auf Replay
  - Motivation
  - Beispiele
- Diagnose