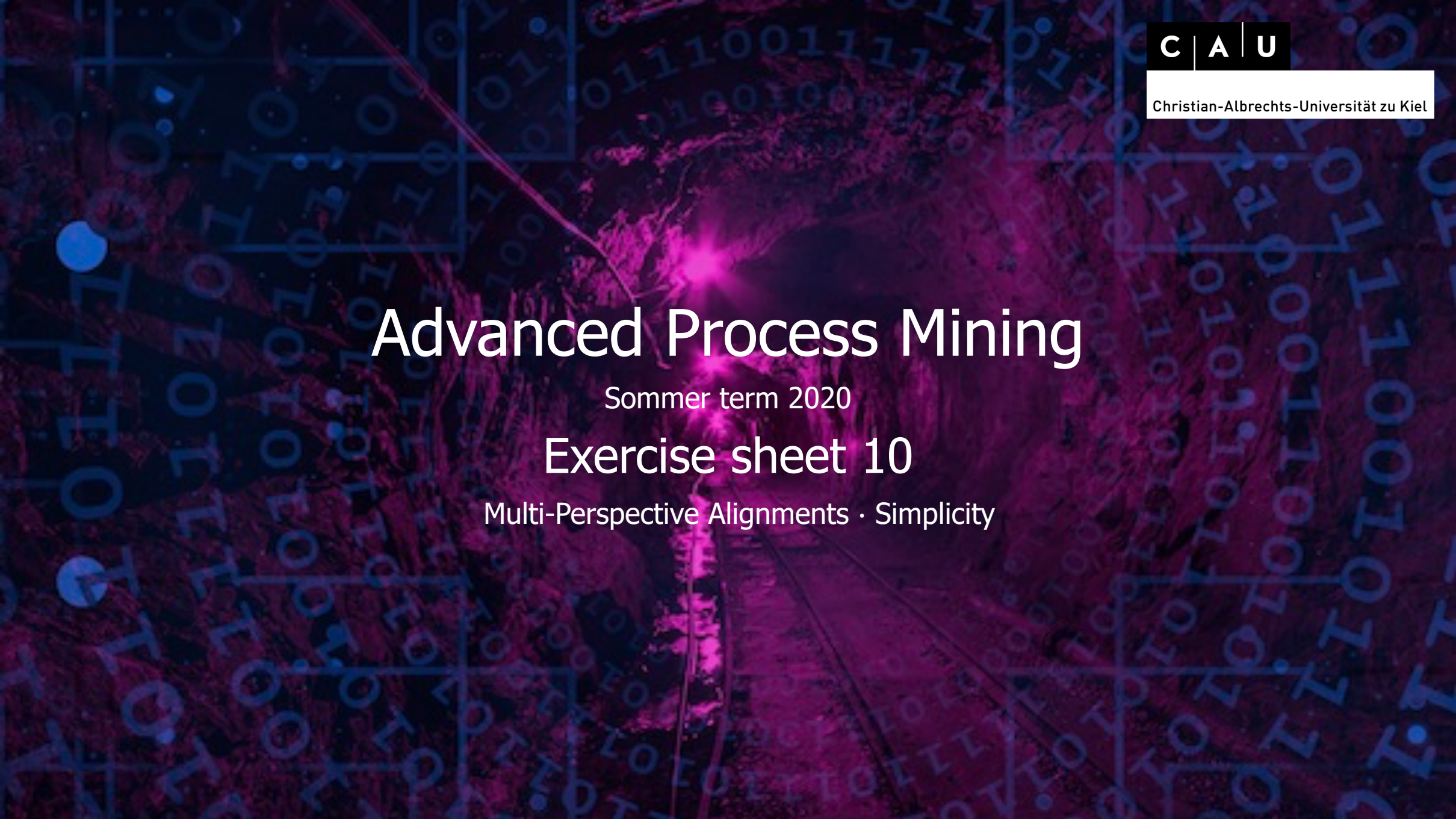


C A U

Christian-Albrechts-Universität zu Kiel

Advanced Process Mining

Sommer term 2020

Exercise sheet 10

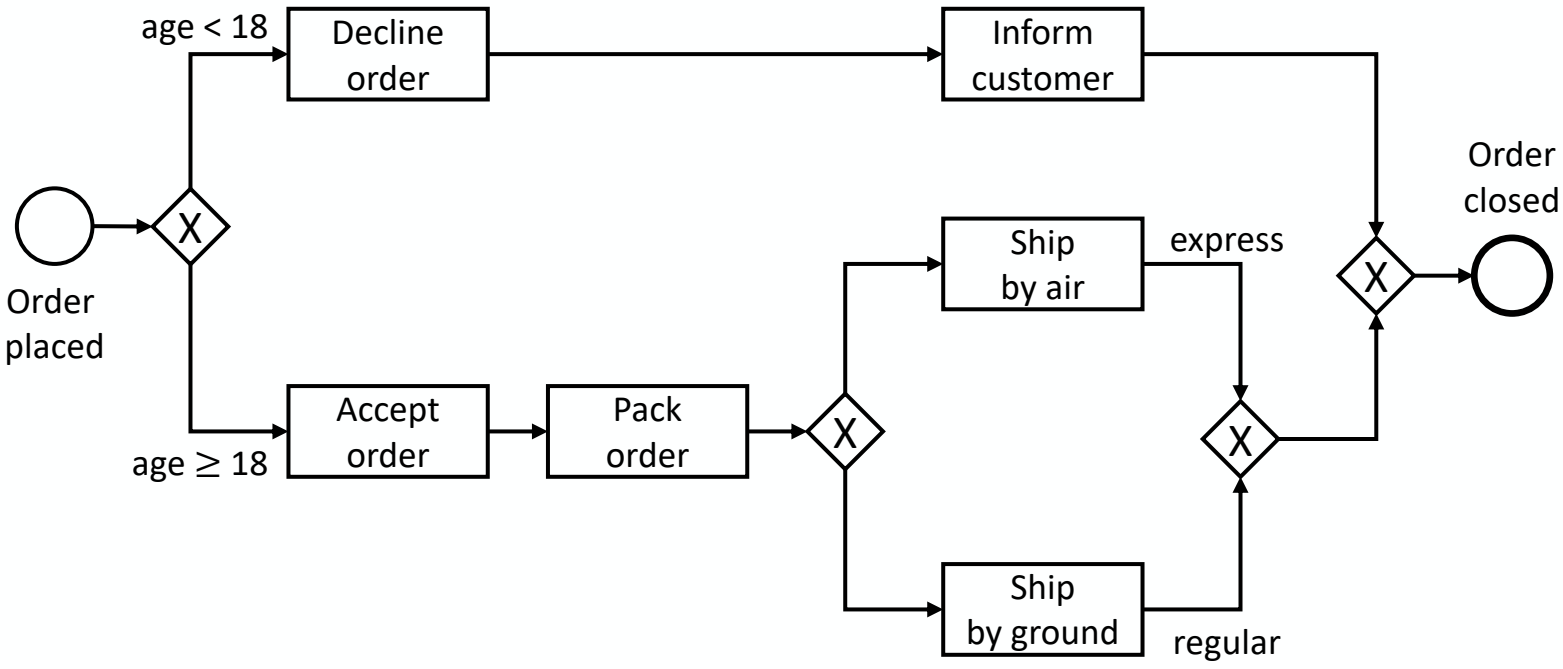
Multi-Perspective Alignments · Simplicity

Multi-Perspective Alignments

Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



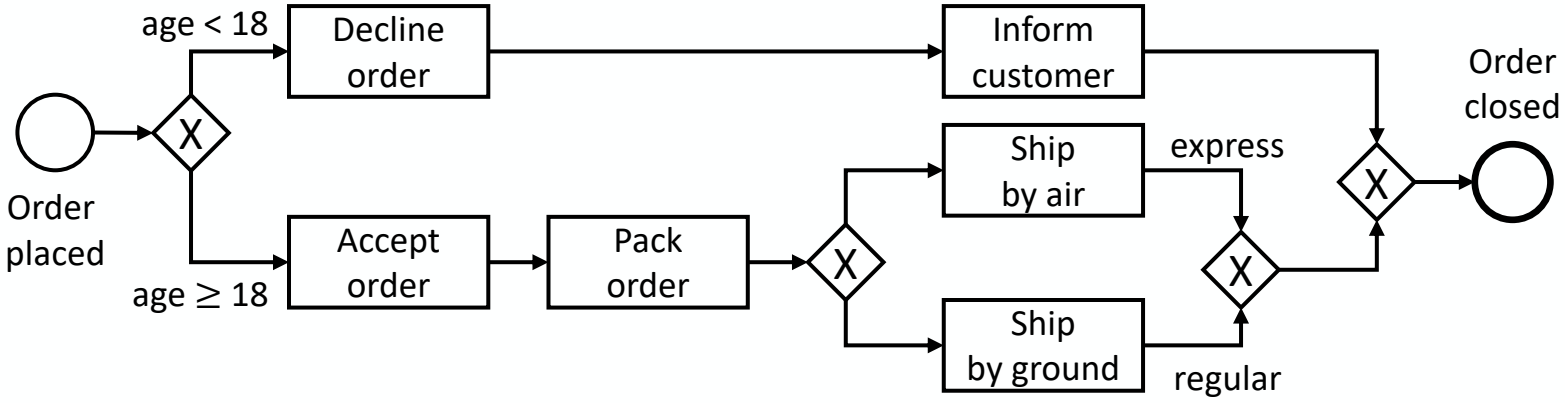
ID	Timestamp	Activity	Age	Express
1	01.06.20 - 18:00:42	order placed	16	true
2	01.06.20 - 18:00:50	order placed	32	false
1	01.06.20 - 18:01:02	decline order	16	true
1	01.06.20 - 18:01:12	inform customer	16	true
1	01.06.20 - 18:01:12	order closed	16	true
2	01.06.20 - 18:02:20	accept order	32	false
3	01.06.20 - 22:01:27	order placed	15	true
3	01.06.20 - 22:33:57	accept order	15	true
2	02.06.20 - 08:43:44	pack order	32	false
2	02.06.20 - 13:22:53	Ship by air	32	false
3	02.06.20 - 13:41:02	Ship by ground	15	true
2	02.06.20 - 13:50:00	Order closed	32	false
3	02.06.20 - 13:50:01	Order closed	15	true

Multi-Perspective Alignments

Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



ID	Timestamp	Activity	Age	Express
1	01.06.20 - 18:00:42	order placed	16	true
1	01.06.20 - 18:01:02	decline order	16	true
1	01.06.20 - 18:01:12	inform customer	16	true
1	01.06.20 - 18:01:12	order closed	16	true

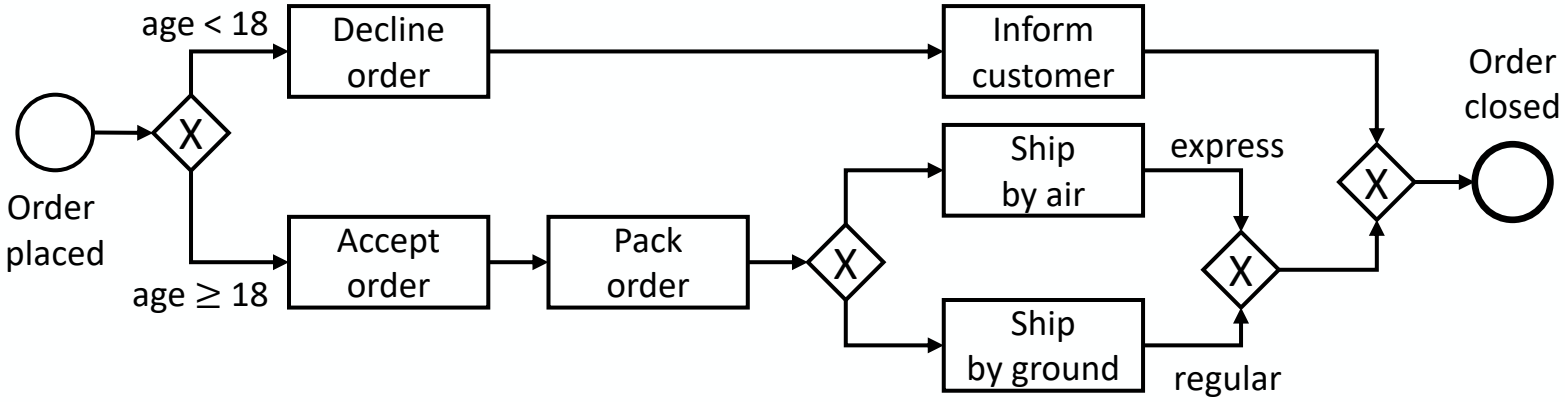
σ_1	order placed	decline order *age = 16*	inform customer	order closed
N				

Multi-Perspective Alignments

Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



ID	Timestamp	Activity	Age	Express
1	01.06.20 - 18:00:42	order placed	16	true
1	01.06.20 - 18:01:02	decline order	16	true
1	01.06.20 - 18:01:12	inform customer	16	true
1	01.06.20 - 18:01:12	order closed	16	true

σ_1	order placed	decline order *age = 16*	inform customer	order closed
N	order placed	decline order *age < 18*	inform customer	order closed

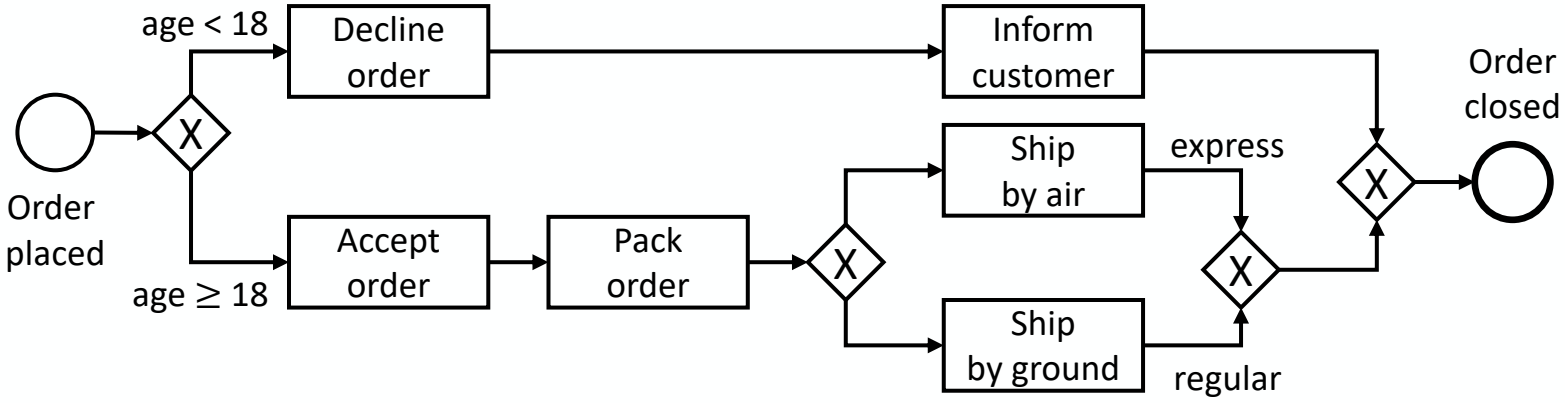
$$C_1 = 0$$

Multi-Perspective Alignments

Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



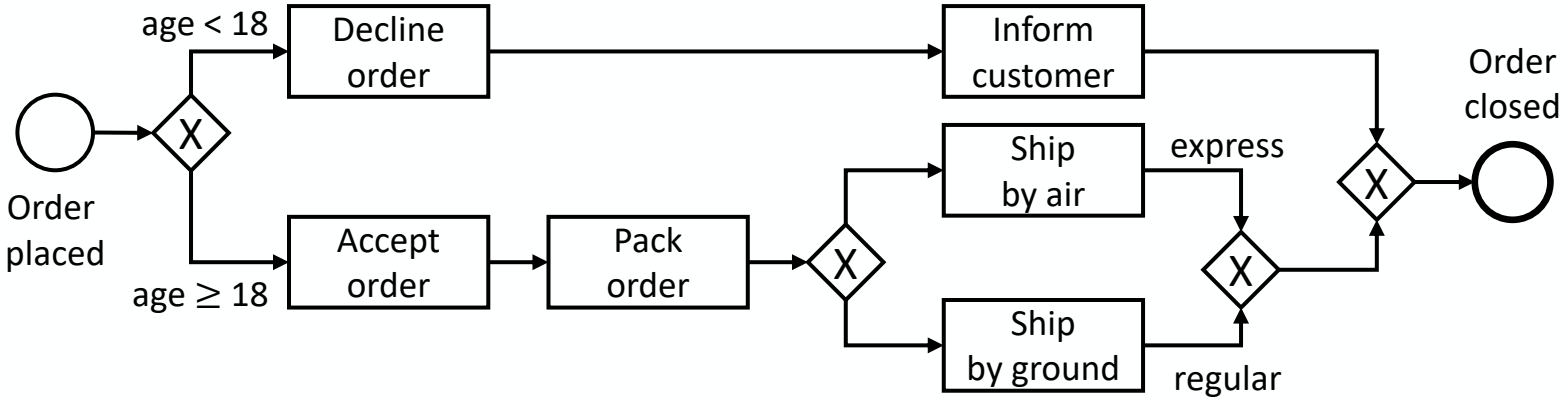
ID	Timestamp	Activity	Age	Express
2	01.06.20 - 18:00:50	order placed	32	false
2	01.06.20 - 18:02:20	accept order	32	false
2	02.06.20 - 08:43:44	pack order	32	false
2	02.06.20 - 13:22:53	Ship by air	32	false
2	02.06.20 - 13:50:00	Order closed	32	false

σ_2					
N					

Multi-Perspective Alignments
Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



ID	Timestamp	Activity	Age	Express
2	01.06.20 - 18:00:50	order placed	32	false
2	01.06.20 - 18:02:20	accept order	32	false
2	02.06.20 - 08:43:44	pack order	32	false
2	02.06.20 - 13:22:53	Ship by air	32	false
2	02.06.20 - 13:50:00	Order closed	32	false

σ_2	order placed	accept order *age = 32*	pack order	ship by air *express = false*	order closed
N	order placed	accept order *age ≥ 18*	pack order	ship by air *express = true*	order closed

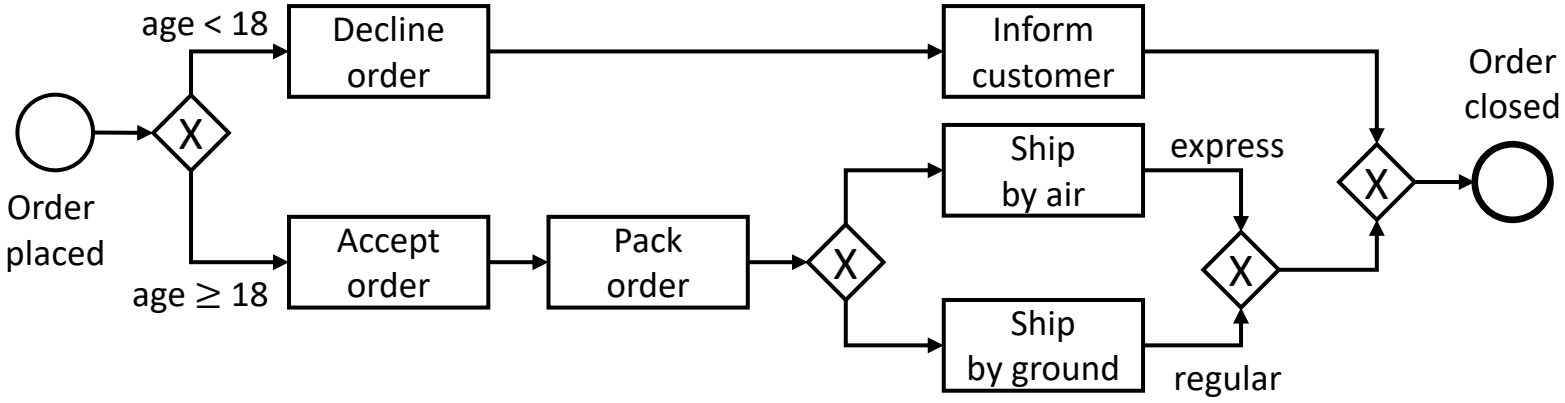
$C_2 = 1$

Multi-Perspective Alignments

Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



ID	Timestamp	Activity	Age	Express
3	01.06.20 - 22:01:27	order placed	15	true
3	01.06.20 - 22:33:57	accept order	15	true
3	02.06.20 - 13:41:02	Ship by ground	15	true
3	02.06.20 - 13:50:01	Order closed	15	true

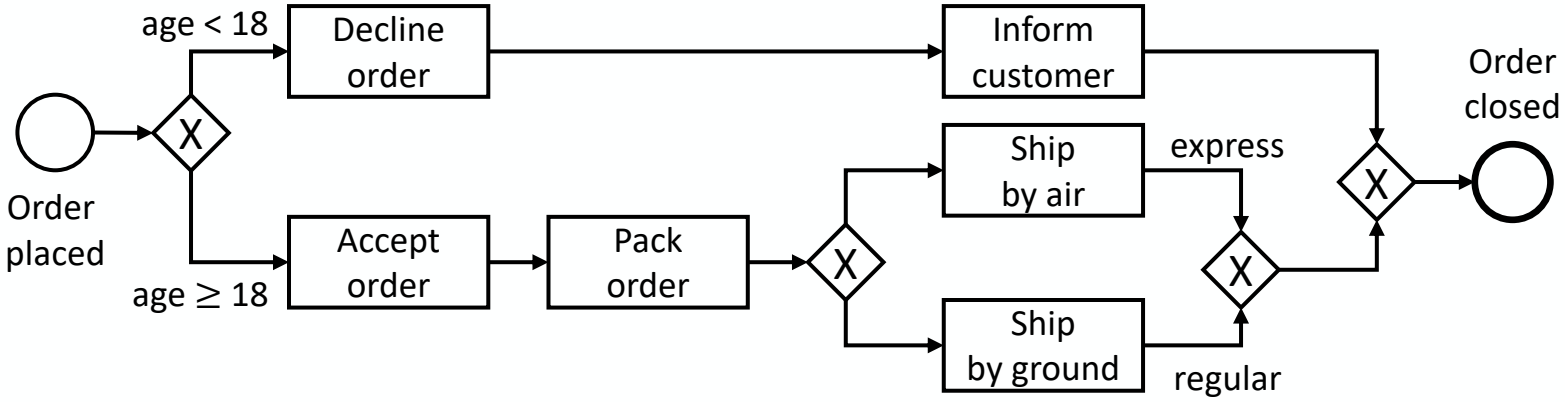
σ_3					
N					

Multi-Perspective Alignments

Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



ID	Timestamp	Activity	Age	Express
3	01.06.20 - 22:01:27	order placed	15	true
3	01.06.20 - 22:33:57	accept order	15	true
3	02.06.20 - 13:41:02	Ship by ground	15	true
3	02.06.20 - 13:50:01	Order closed	15	true

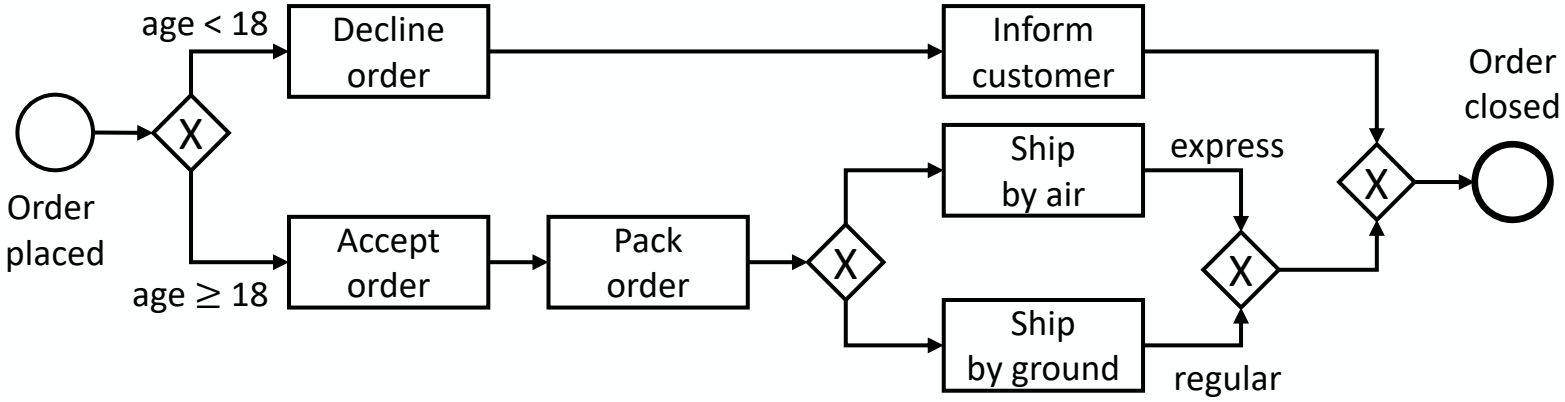
σ_3	order placed	accept order *age = 15*	≫	ship by ground *express = true*	order closed
N	order placed	accept order *age ≥ 18*	pack order	ship by air *express = true*	order closed

$$C_3 = \infty$$

Multi-Perspective Alignments
Exercise 1

Christian-Albrechts-Universität zu Kiel

Calculate the deviation cost for the traces in the following event log using multi-perspective alignments.



ID	Timestamp	Activity	Age	Express
3	01.06.20 - 22:01:27	order placed	15	true
3	01.06.20 - 22:33:57	accept order	15	true
3	02.06.20 - 13:41:02	Ship by ground	15	true
3	02.06.20 - 13:50:01	Order closed	15	true

σ_3	order placed	accept order *age = 15*	≫	ship by ground *express = true*	order closed
N	order placed	accept order *age ≥ 18*	pack order	ship by ground *express = false*	order closed

$C_3 = 3$

True or False

Exercise 2a

Christian-Albrechts-Universität zu Kiel

For a successful process mining project it is not essential to have a common understanding of the objective.

False

Before a process mining project is started, a concrete use case and a clear purpose of the project has to be identified.

True or False

Exercise 2b

Christian-Albrechts-Universität zu Kiel

The terms “Multi-Perspective Alignments” and “Multiple Trace Alignments” describe the same thing and can be used synonymous.

False

Multi-Perspective Alignment	Multiple Trace Alignment
Compared to the simple alignment method the multi-perspective alignment takes additional attributes into account.	Is not only optimising the alignment cost for a single trace but attempts to minimise the overall alignment cost.

Exercise is based on the paper:

Quality Dimensions in Process Discovery: The Importance of Fitness, Precision, Generalization and Simplicity

by:

J. C. A. M. Buijs, B. F. van Dongen and W. M. P. van der Aalst

Simplicity
Exercise 3a

Christian-Albrechts-Universität zu Kiel

What is quantified by simplicity and how is it measured?

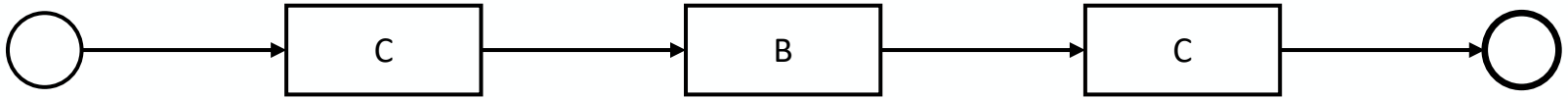
The complexity of the model is quantified by the simplicity and is measured by comparing the size of the tree with the number of activities in the log.

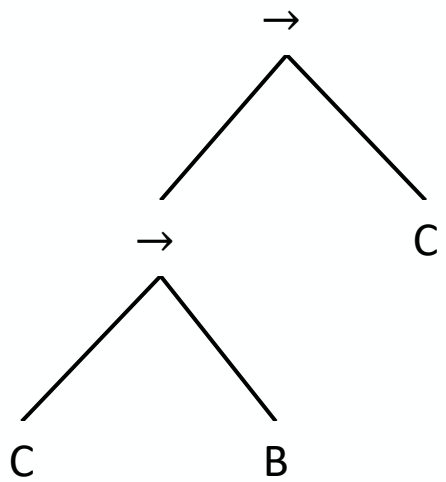
Bujis (2014)

Simplicity
Exercise 3b

Christian-Albrechts-Universität zu Kiel

Transform the process model into a process tree

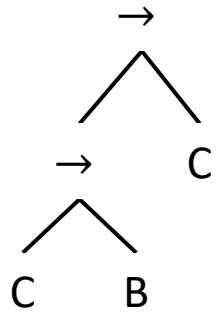




Bujis (2014)

Simplicity
Exercise 3c

Calculate the simplicity Q_S of the resulting process tree.



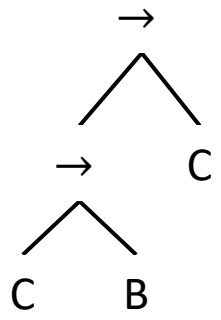
Trace	#
ABCDEG	6
ABCDFG	38
ABDCEG	12
ABDCFG	26
ABCFG	8
ACBEG	1
ADBCFG	1
ADBCEG	1
ADCBFG	4
ACDBFG	2
ACBFG	1

$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

Simplicity
Exercise 3c

Christian-Albrechts-Universität zu Kiel

Calculate the simplicity Q_S of the resulting process tree.



$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

Trace	#
ABCDEG	6
ABCDFG	38
ABDCEG	12
ABDCFG	26
ABCFG	8
ACBEG	1
ADBCFG	1
ADBCEG	1
ADCBFG	4
ACDBFG	2
ACBFG	1

"Duplication of activities is measured by counting the number of times the activity is repeated in the process model." Bujis (2014)

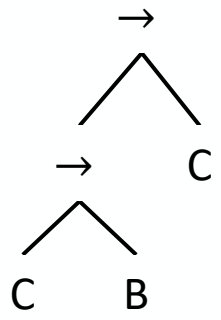
A: 1 Duplicate activities ← According to *Bujis*

B: 2 Duplicate activities

Simplicity
Exercise 3c

Christian-Albrechts-Universität zu Kiel

Calculate the simplicity Q_S of the resulting process tree.



$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

Trace	#
ABCDEG	6
ABCDFG	38
ABDCEG	12
ABDCFG	26
ABCFG	8
ACBEG	1
ADBCFG	1
ADBCEG	1
ADCBFG	4
ACDBFG	2
ACBFG	1

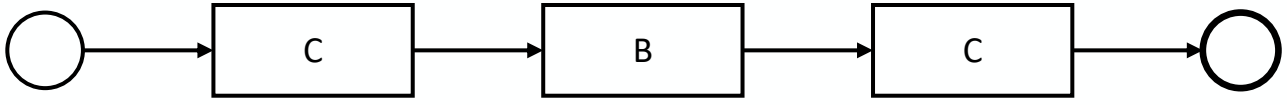
"An activity is missing from the process model if it is not included in the process model while it is present in the event log." Bujis (2014)

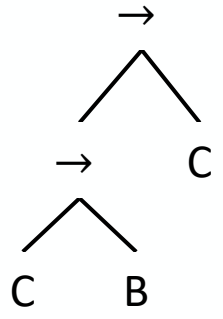
5 missing activities

Simplicity
Exercise 3c

Christian-Albrechts-Universität zu Kiel

Calculate the simplicity Q_S of the resulting process tree.





$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

"…normalized by the total number of nodes in the process tree..."

Bujis (2014)

A: 3 nodes in process tree ← According to *Bujis*

B: 4 nodes in process tree

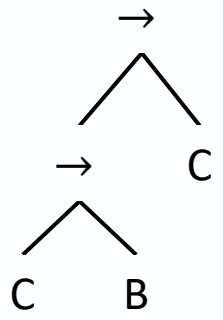
C: 5 nodes in process tree

Trace	#
ABCDEG	6
ABCDFG	38
ABDCEG	12
ABDCFG	26
ABCFG	8
ACBEG	1
ADBCFG	1
ADBCEG	1
ADCBFG	4
ACDBFG	2
ACBFG	1

Simplicity
Exercise 3c

Christian-Albrechts-Universität zu Kiel

Calculate the simplicity Q_S of the resulting process tree.



$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

"…event classes (or activities) in the event log..."

Bujis (2014)

A: 7 event classes in event log ← According to *Bujis*

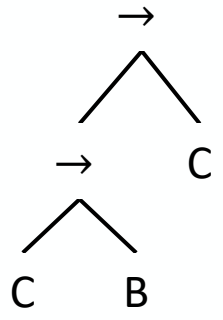
B: 11 event classes in event log

Trace	#
ABCDEG	6
ABCDFG	38
ABDCEG	12
ABDCFG	26
ABCFG	8
ACBEG	1
ADBCFG	1
ADBCEG	1
ADCBFG	4
ACDBFG	2
ACBFG	1

Simplicity
Exercise 3c

Christian-Albrechts-Universität zu Kiel

Calculate the simplicity Q_S of the resulting process tree.



$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

$$Q_S = 1 - \frac{1+5}{3+7} = 0.4$$

Trace	#
ABCDEG	6
ABCDFG	38
ABDCEG	12
ABDCFG	26
ABCFG	8
ACBEG	1
ADBCFG	1
ADBCEG	1
ADCBFG	4
ACDBFG	2
ACBFG	1

According to the approach off by *Buijs et al.*, how does a process tree look like with **Excellent simplicity**

$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

To achieve excellent simplicity, the number of duplicate activities and the number of missing activities has to be equal to zero.

Can a good simplicity still be achieved, even if the number of duplicate activities or missing activities are not zero?

Yes.
If the denominator is much greater than the numerator.

Then either the number of nodes in the process tree has to be very high,
or the event log has to consist of a high number of different activities

Simplicity
Exercise 3d

Christian-Albrechts-Universität zu Kiel

According to the approach off by *Buijs et al.*, how does a process tree look like with **Very poor simplicity**

$$Q_S = 1 - \frac{\#\ duplicate\ activties\ +\ \#\ missing\ activties}{\#\ nodes\ in\ process\ tree\ +\ \#\ event\ class\ in\ event\ log}$$

If most of the nodes in the process tree are duplicates or many activities are missing in the process tree, then the simplicity of the process tree will be very poor.