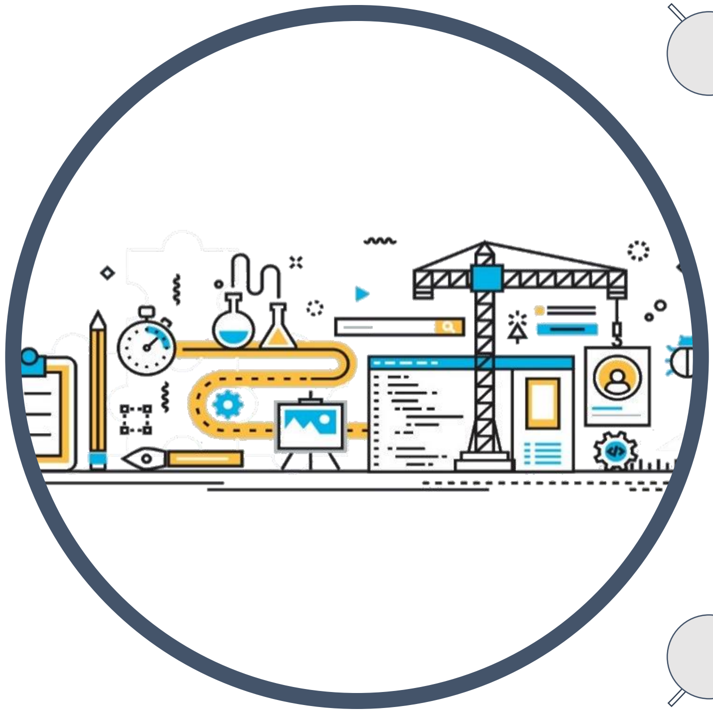
`propro`
Enhancing Discovered Process Models

Using bayesian inference & MCMC

Gert Janssenswillen, Benoît Depaire, Christel Faes

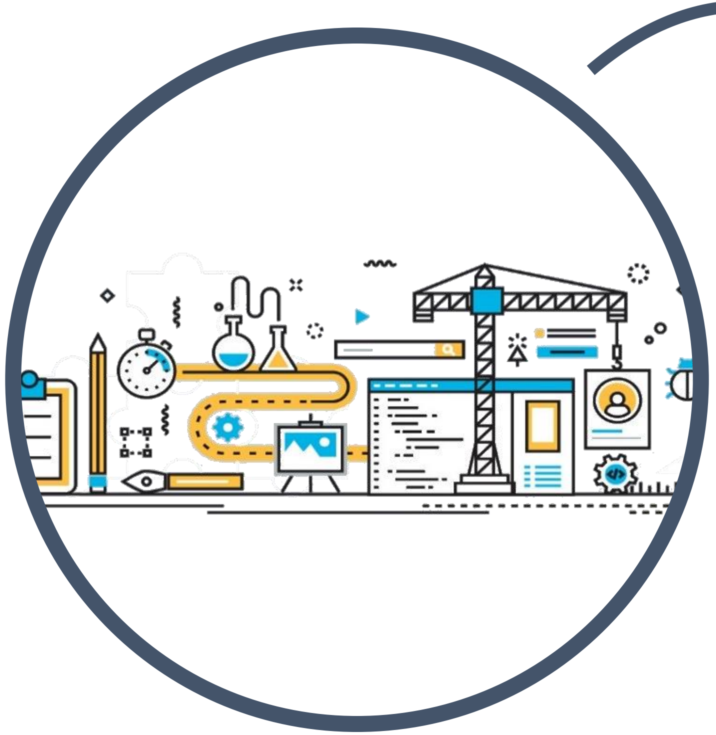
@gjanssenswillen

process mining



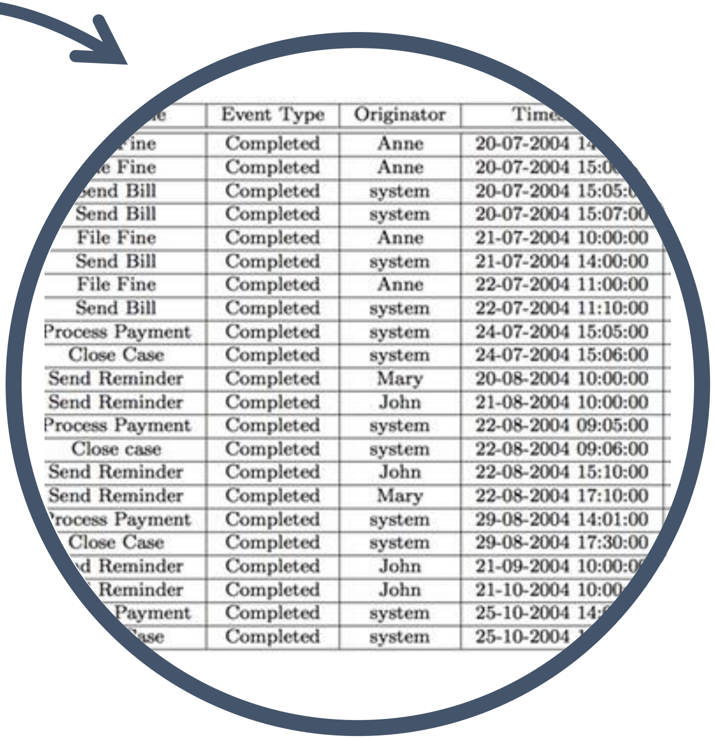
Process

- Order-to-cash
- Procurement-to-pay
- Customer journeys
- Medical procedures
- Educational trajectories
- Transportation processes
- …



Process

generate



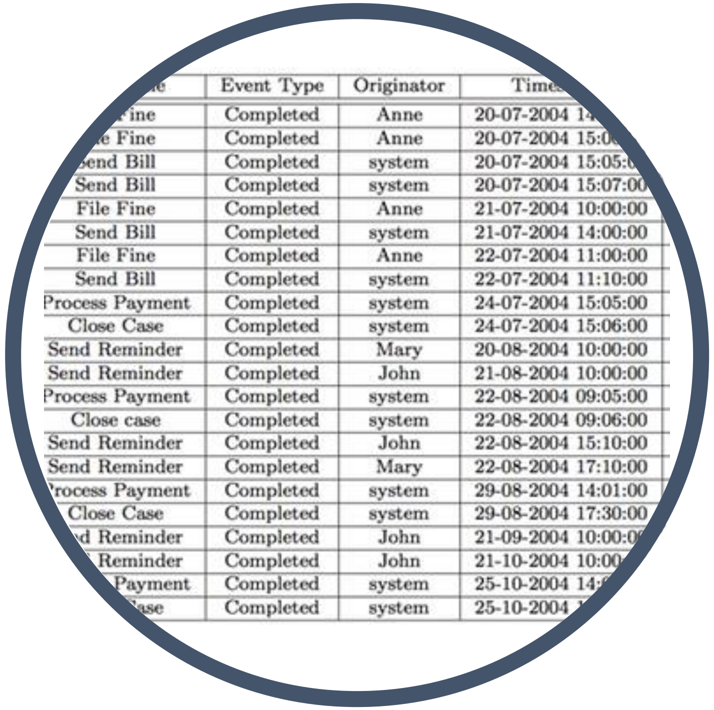
e	Event Type	Originator	Time
Fine	Completed	Anne	20-07-2004 14
e Fine	Completed	Anne	20-07-2004 15:0
end Bill	Completed	system	20-07-2004 15:05:
Send Bill	Completed	system	20-07-2004 15:07:00
File Fine	Completed	Anne	21-07-2004 10:00:00
Send Bill	Completed	system	21-07-2004 14:00:00
File Fine	Completed	Anne	22-07-2004 11:00:00
Send Bill	Completed	system	22-07-2004 11:10:00
Process Payment	Completed	system	24-07-2004 15:05:00
Close Case	Completed	system	24-07-2004 15:06:00
Send Reminder	Completed	Mary	20-08-2004 10:00:00
Send Reminder	Completed	John	21-08-2004 10:00:00
Process Payment	Completed	system	22-08-2004 09:05:00
Close case	Completed	system	22-08-2004 09:06:00
Send Reminder	Completed	John	22-08-2004 15:10:00
Send Reminder	Completed	Mary	22-08-2004 17:10:00
rocess Payment	Completed	system	29-08-2004 14:01:00
Close Case	Completed	system	29-08-2004 17:30:00
d Reminder	Completed	John	21-09-2004 10:00:0
Reminder	Completed	John	21-10-2004 10:00
Payment	Completed	system	25-10-2004 14:
ase	Completed	system	25-10-2004

Event log

Activity/Task/Event

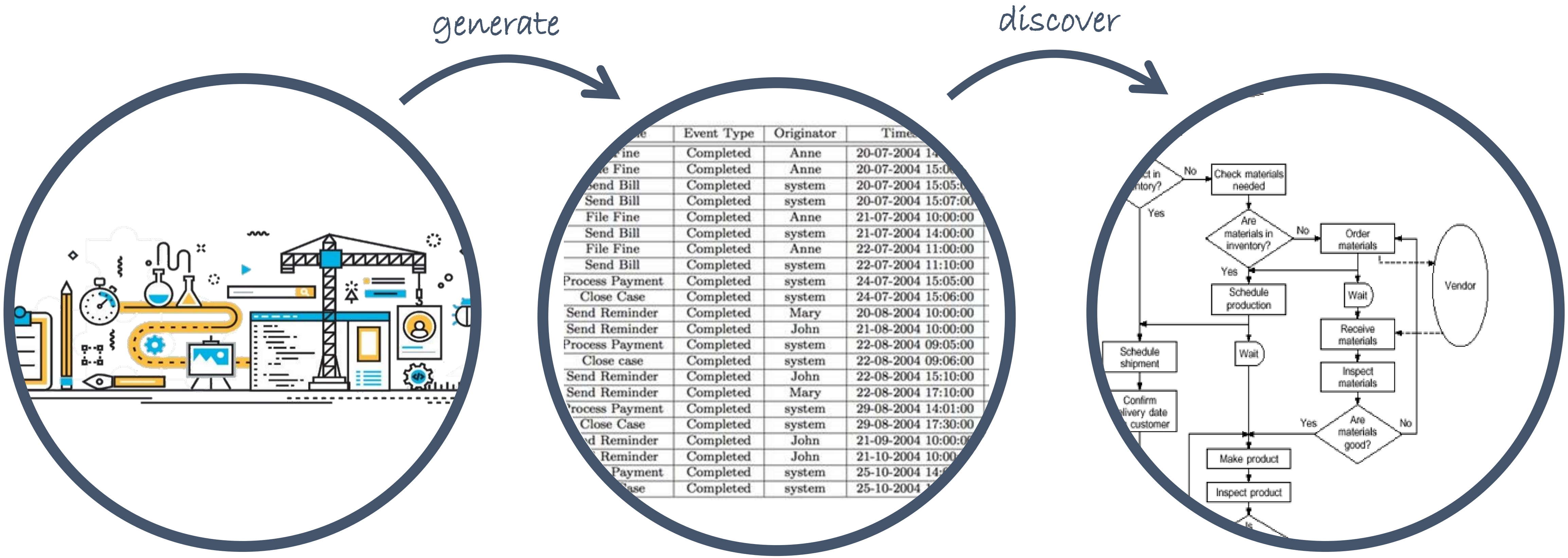
Case

Timestamp



e	Event Type	Originator	Time
Fine	Completed	Anne	20-07-2004 14
e Fine	Completed	Anne	20-07-2004 15:0
end Bill	Completed	system	20-07-2004 15:05:
Send Bill	Completed	system	20-07-2004 15:07:00
File Fine	Completed	Anne	21-07-2004 10:00:00
Send Bill	Completed	system	21-07-2004 14:00:00
File Fine	Completed	Anne	22-07-2004 11:00:00
Send Bill	Completed	system	22-07-2004 11:10:00
Process Payment	Completed	system	24-07-2004 15:05:00
Close Case	Completed	system	24-07-2004 15:06:00
Send Reminder	Completed	Mary	20-08-2004 10:00:00
Send Reminder	Completed	John	21-08-2004 10:00:00
Process Payment	Completed	system	22-08-2004 09:05:00
Close case	Completed	system	22-08-2004 09:06:00
Send Reminder	Completed	John	22-08-2004 15:10:00
Send Reminder	Completed	Mary	22-08-2004 17:10:00
rocess Payment	Completed	system	29-08-2004 14:01:00
Close Case	Completed	system	29-08-2004 17:30:00
d Reminder	Completed	John	21-09-2004 10:00:0
Reminder	Completed	John	21-10-2004 10:00
Payment	Completed	system	25-10-2004 14:
ase	Completed	system	25-10-2004

Event log



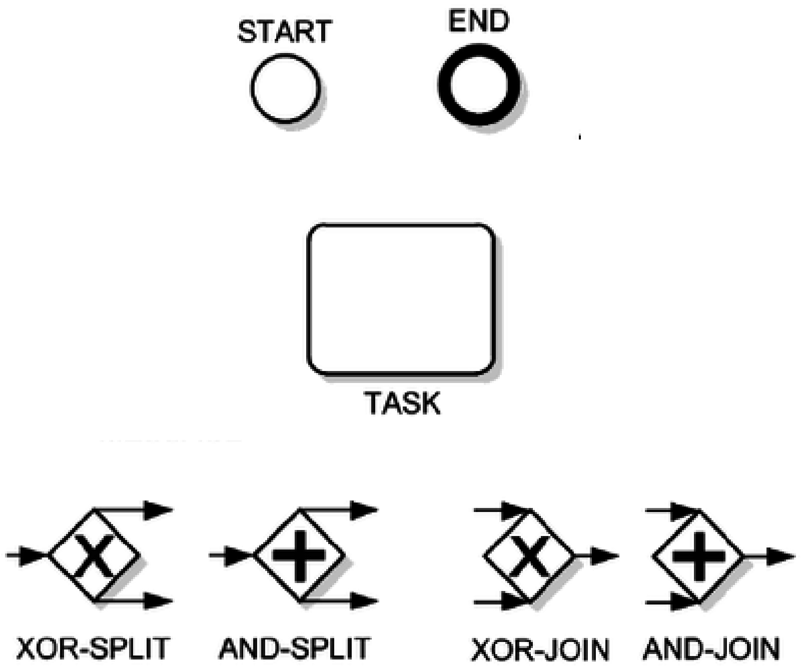
e	Event Type	Originator	Time
Fine	Completed	Anne	20-07-2004 14
e Fine	Completed	Anne	20-07-2004 15:0
end Bill	Completed	system	20-07-2004 15:05:
Send Bill	Completed	system	20-07-2004 15:07:00
File Fine	Completed	Anne	21-07-2004 10:00:00
Send Bill	Completed	system	21-07-2004 14:00:00
File Fine	Completed	Anne	22-07-2004 11:00:00
Send Bill	Completed	system	22-07-2004 11:10:00
Process Payment	Completed	system	24-07-2004 15:05:00
Close Case	Completed	system	24-07-2004 15:06:00
Send Reminder	Completed	Mary	20-08-2004 10:00:00
Send Reminder	Completed	John	21-08-2004 10:00:00
Process Payment	Completed	system	22-08-2004 09:05:00
Close case	Completed	system	22-08-2004 09:06:00
Send Reminder	Completed	John	22-08-2004 15:10:00
Send Reminder	Completed	Mary	22-08-2004 17:10:00
rocess Payment	Completed	system	29-08-2004 14:01:00
Close Case	Completed	system	29-08-2004 17:30:00
d Reminder	Completed	John	21-09-2004 10:00:0
Reminder	Completed	John	21-10-2004 10:00
Payment	Completed	system	25-10-2004 14:
ase	Completed	system	25-10-2004

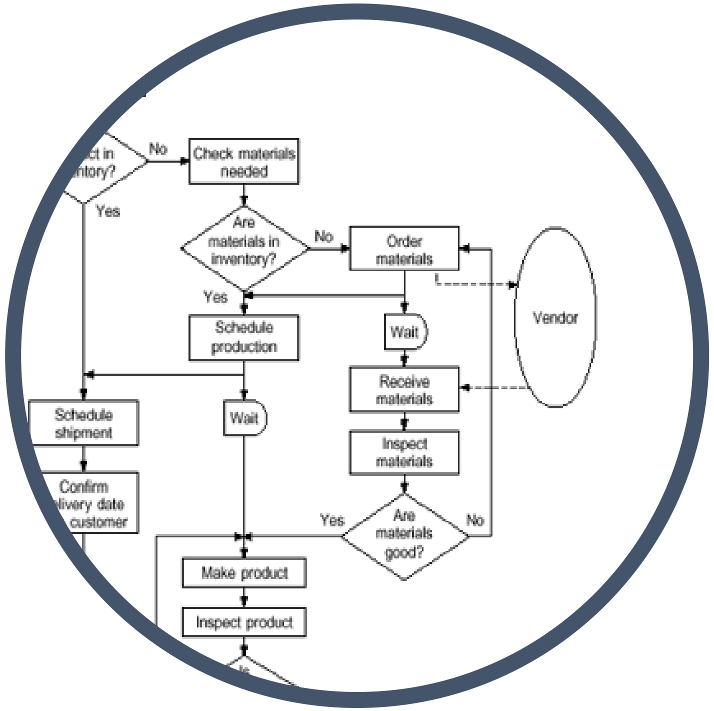
Process

Event log

Process model

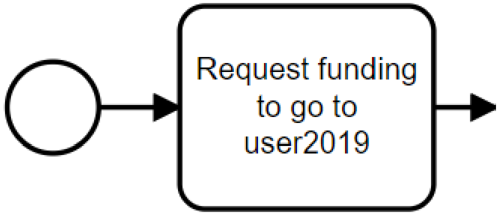
Business Process Model and Notation (BPMN) (ISO standard)



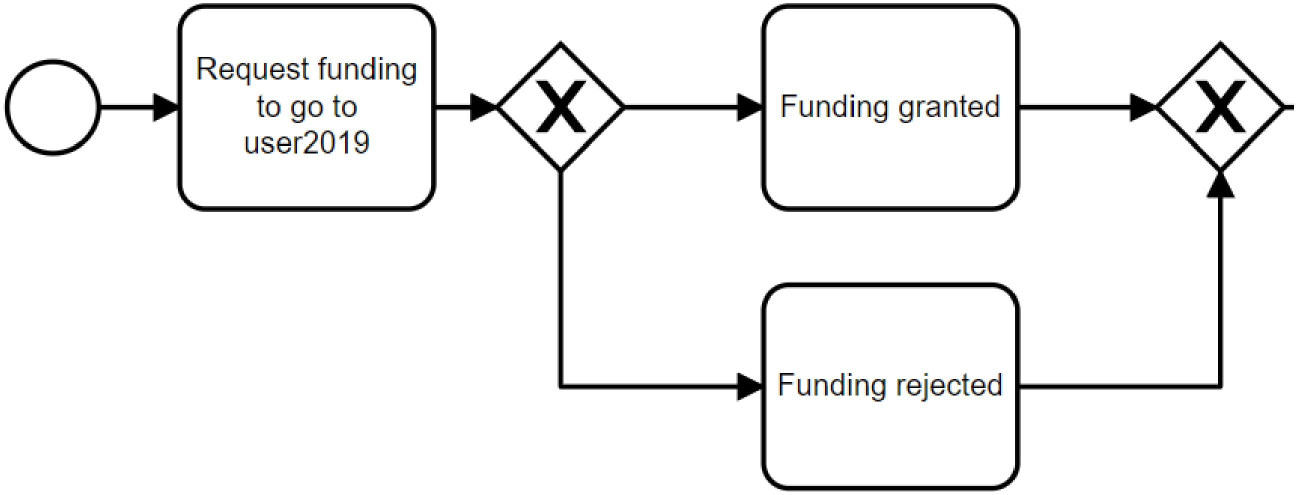


Process model

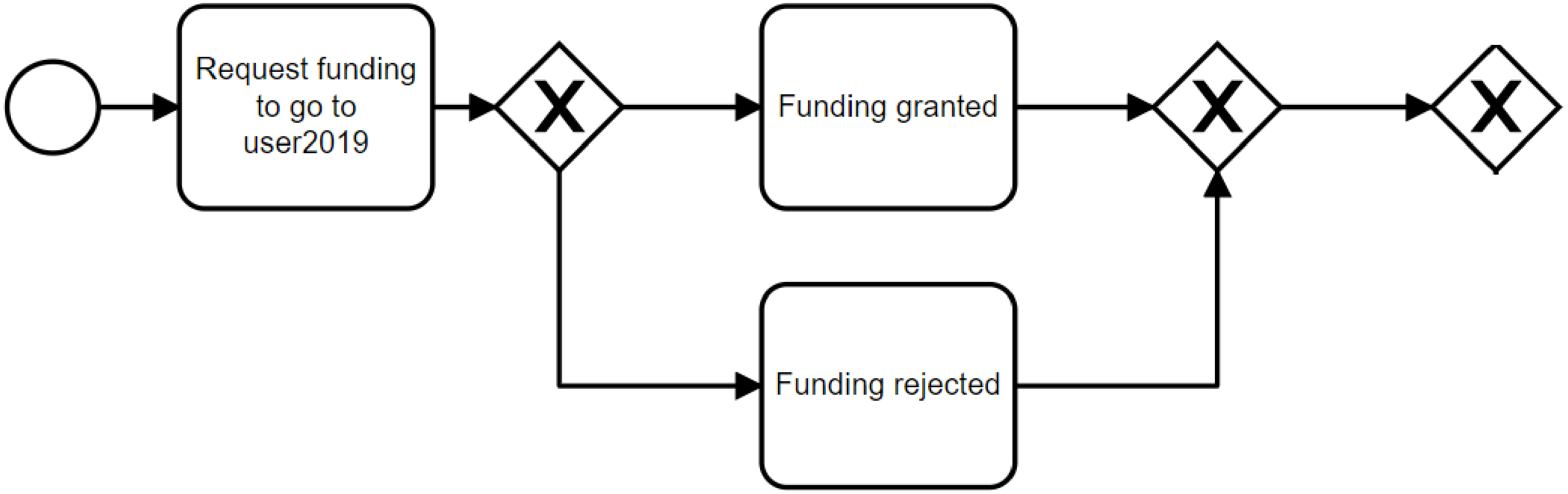
Let’s go to useR



Let's go to useR



Let’s go to useR

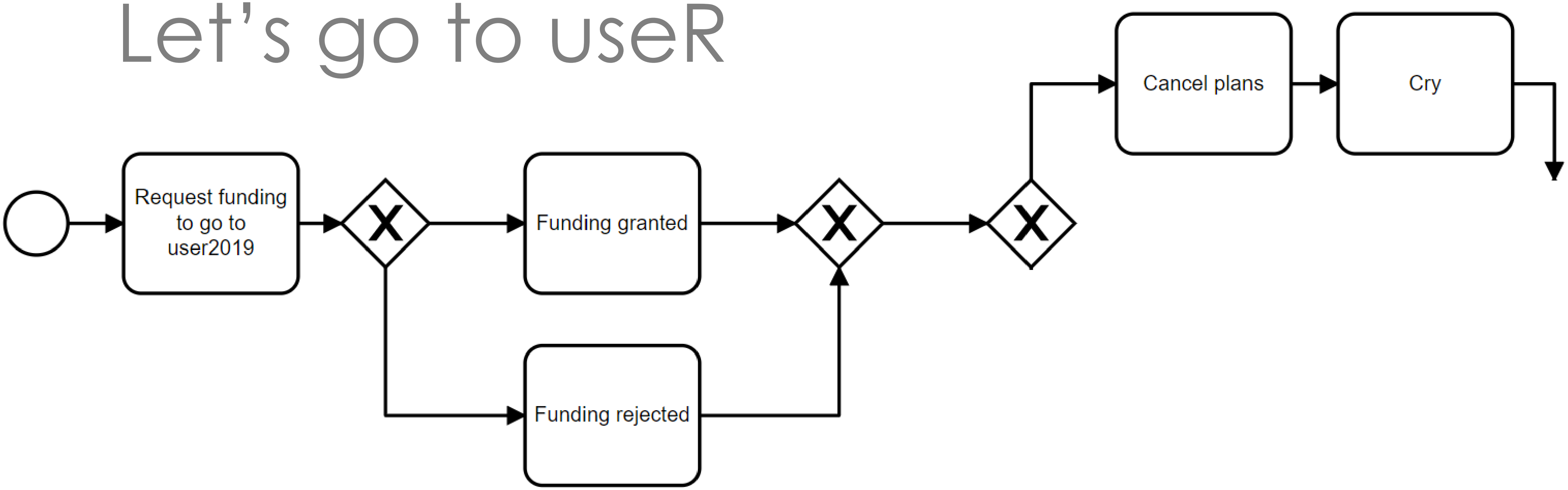


Request funding to go to user2019

Funding granted

Funding rejected

Let's go to useR



Request funding to go to user2019

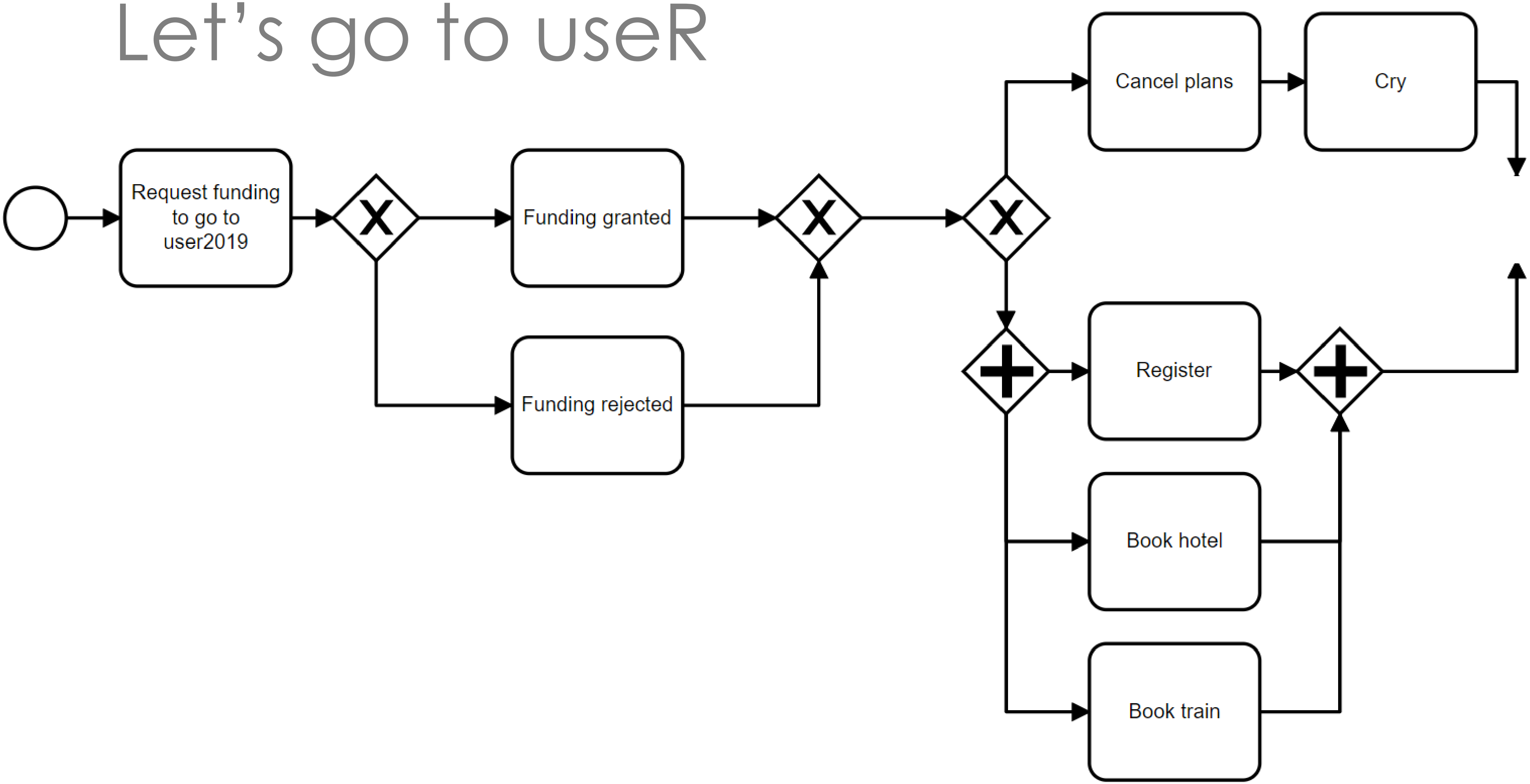
Funding granted

Funding rejected

Cancel plans

Cry

Let's go to useR



Request funding to go to user2019

Funding granted

Funding rejected

Cancel plans

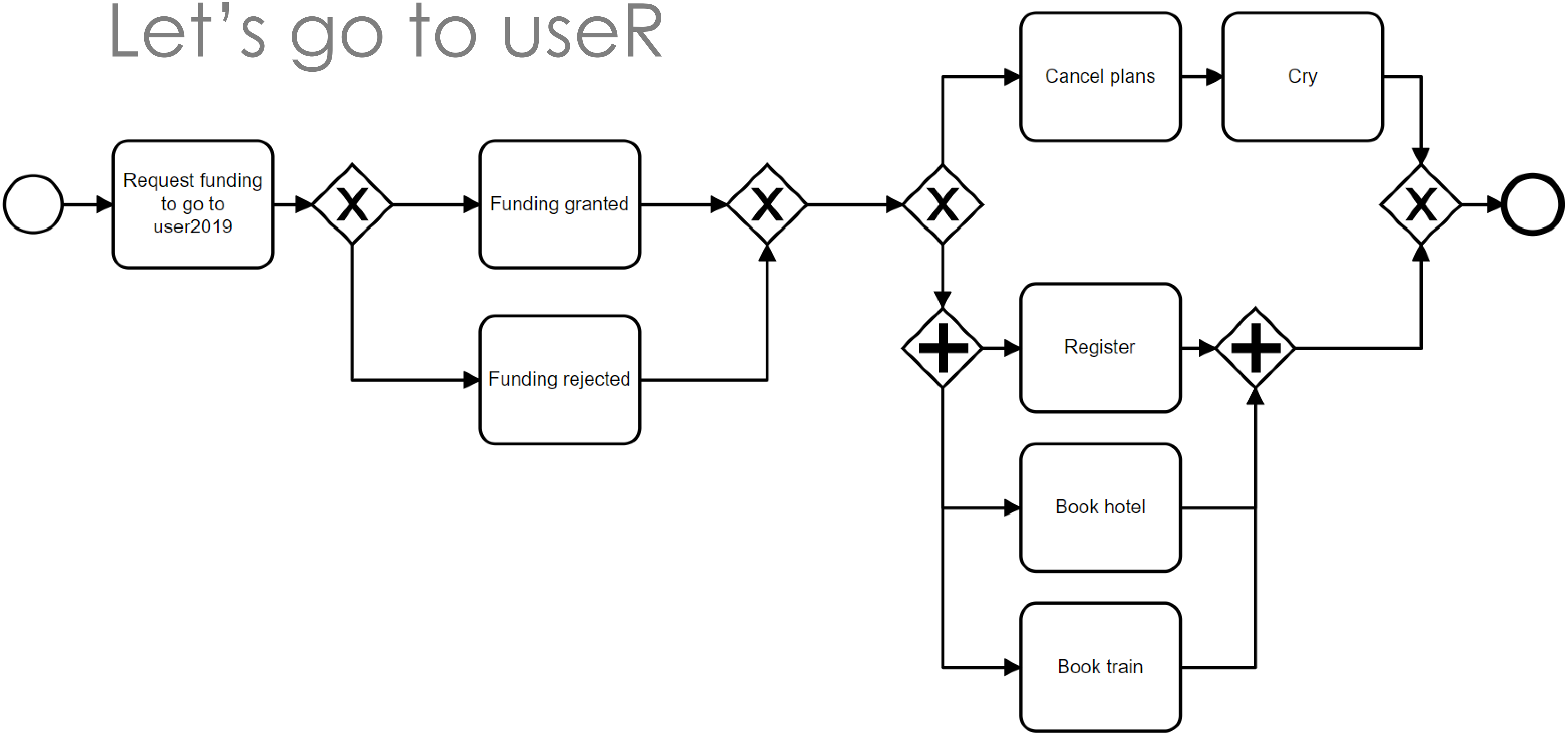
Cry

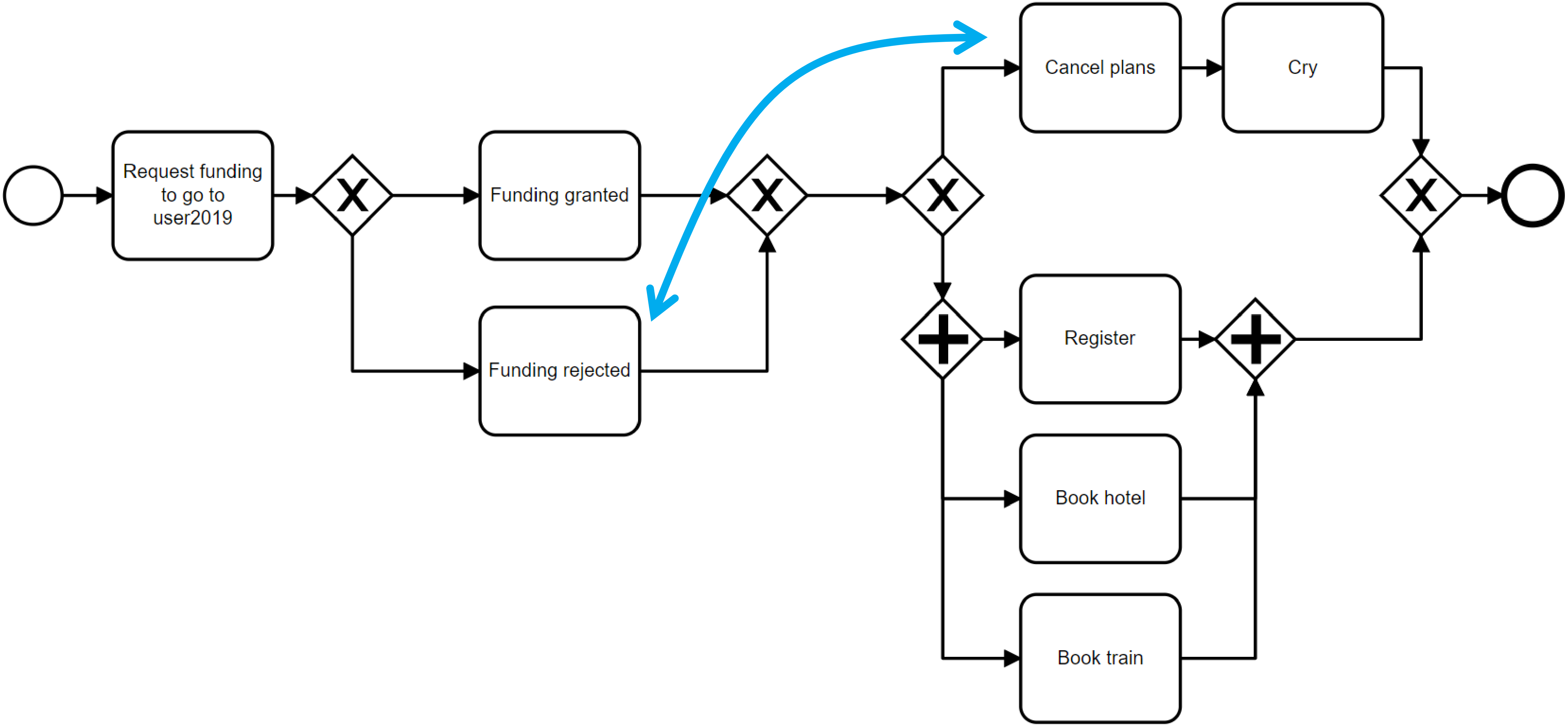
Register

Book hotel

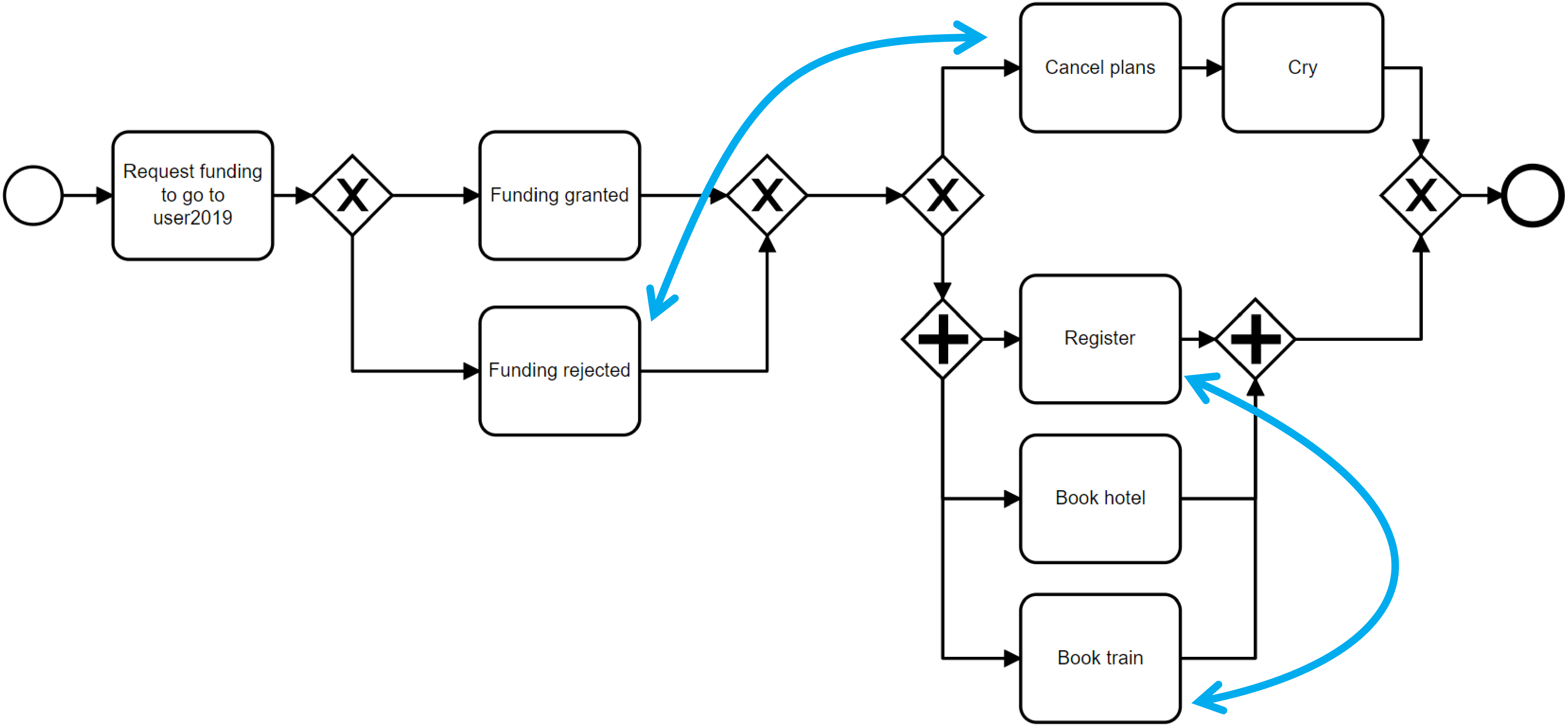
Book train

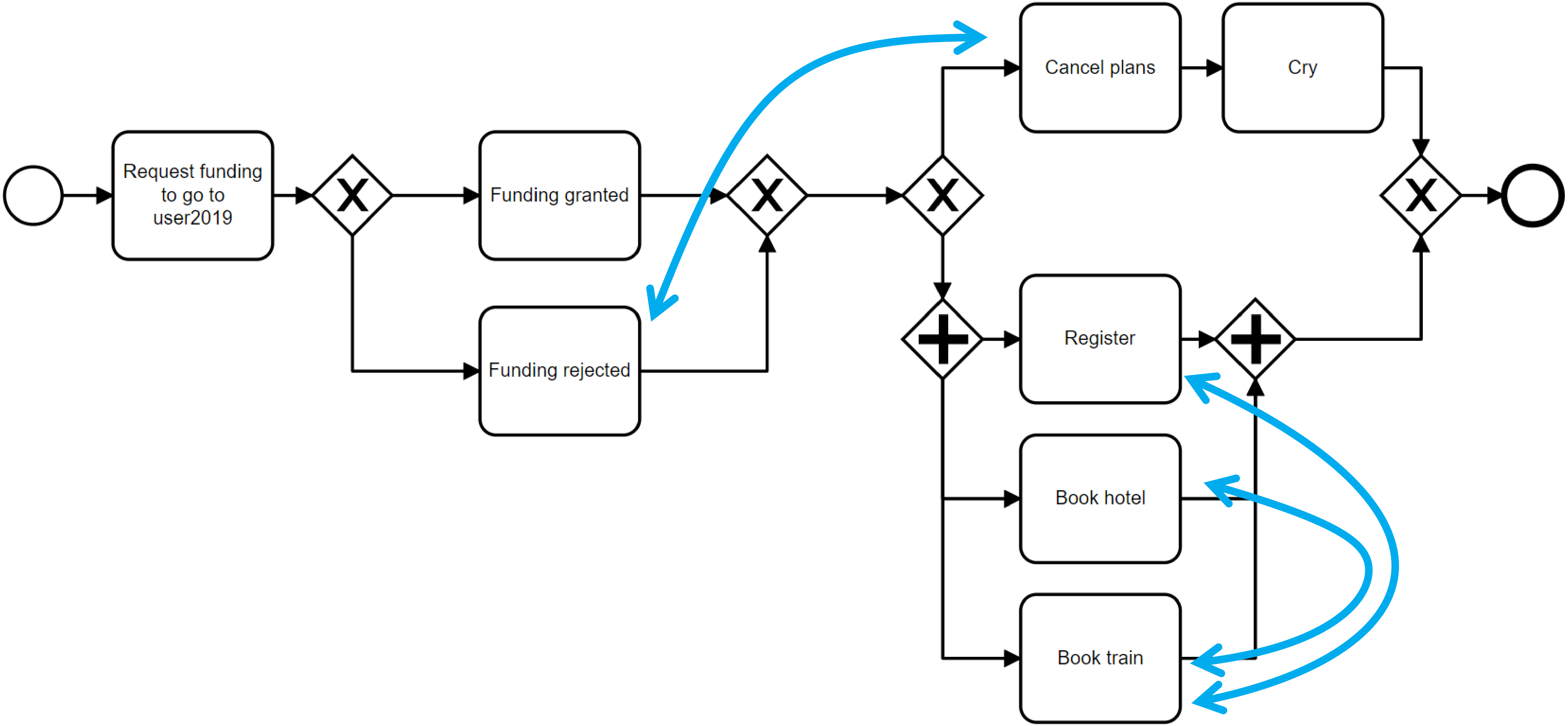
Let's go to useR





Request funding to go to user2019
Funding granted
Funding rejected
Cancel plans
Cry
Register
Book hotel
Book train


Request funding to go to user2019
X
Funding granted
Funding rejected
X
X
Cancel plans
Cry
X
Register
Book hotel
Book train


Request funding to go to user2019
X
Funding granted
Funding rejected
X
X
Cancel plans
Cry
X
Register
Book hotel
Book train
+
+

Learning probabilities

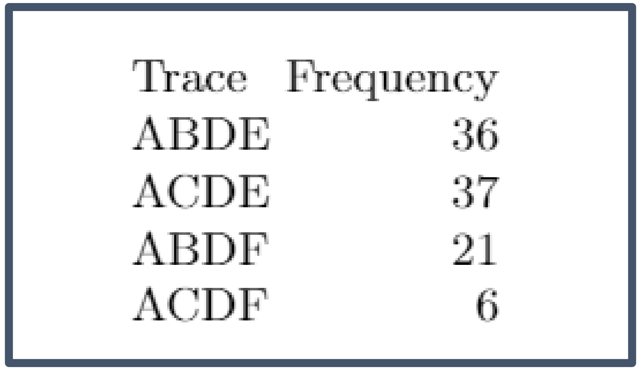
Using Bayesian inference and MCMC

Activities

A	B	C	D	E	F

Activities

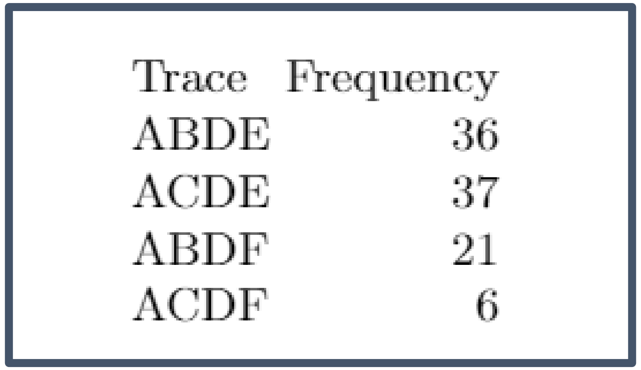
A B C D E F



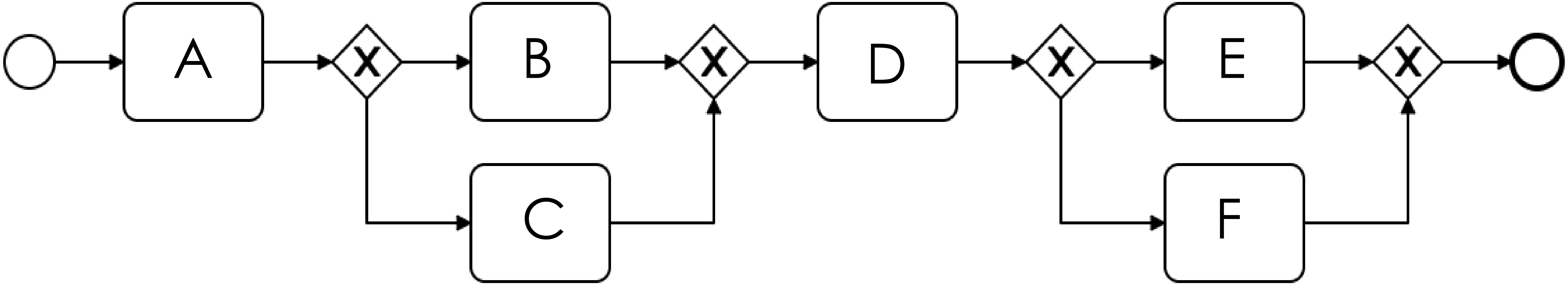
Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6

Activities

A B C D E F



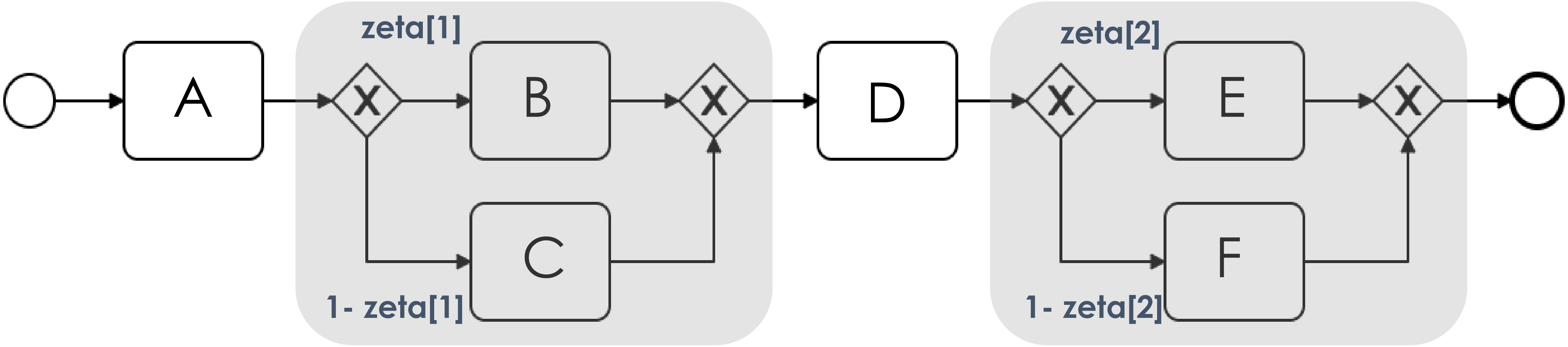
Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6

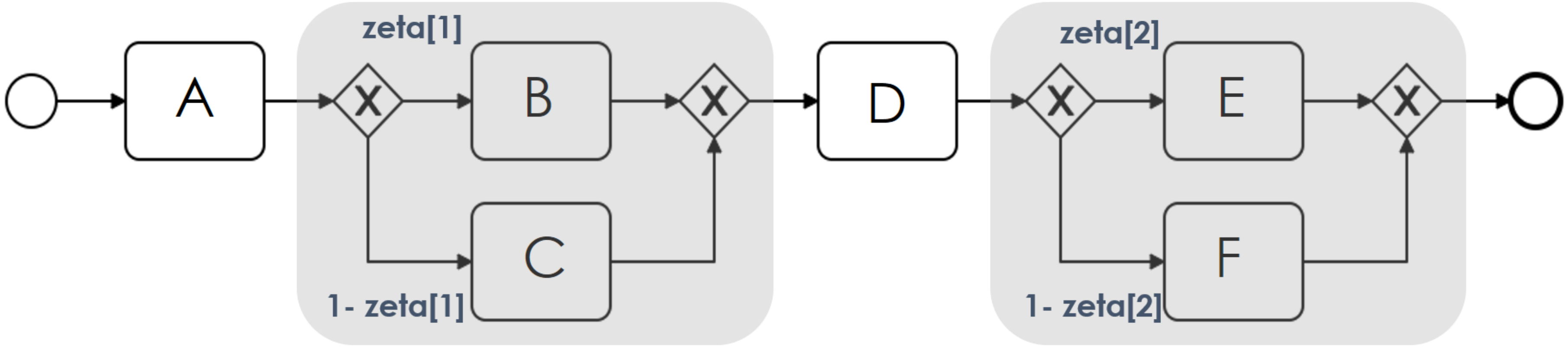


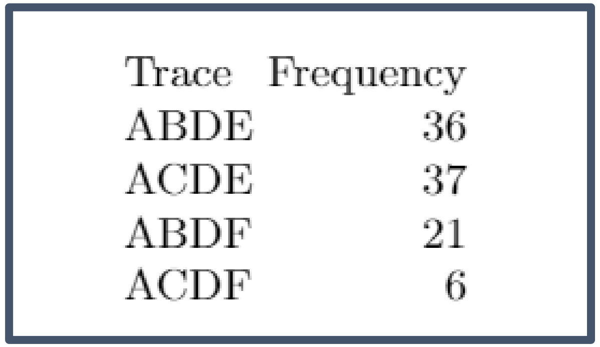
Activities

A B C D E F

Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6







Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6

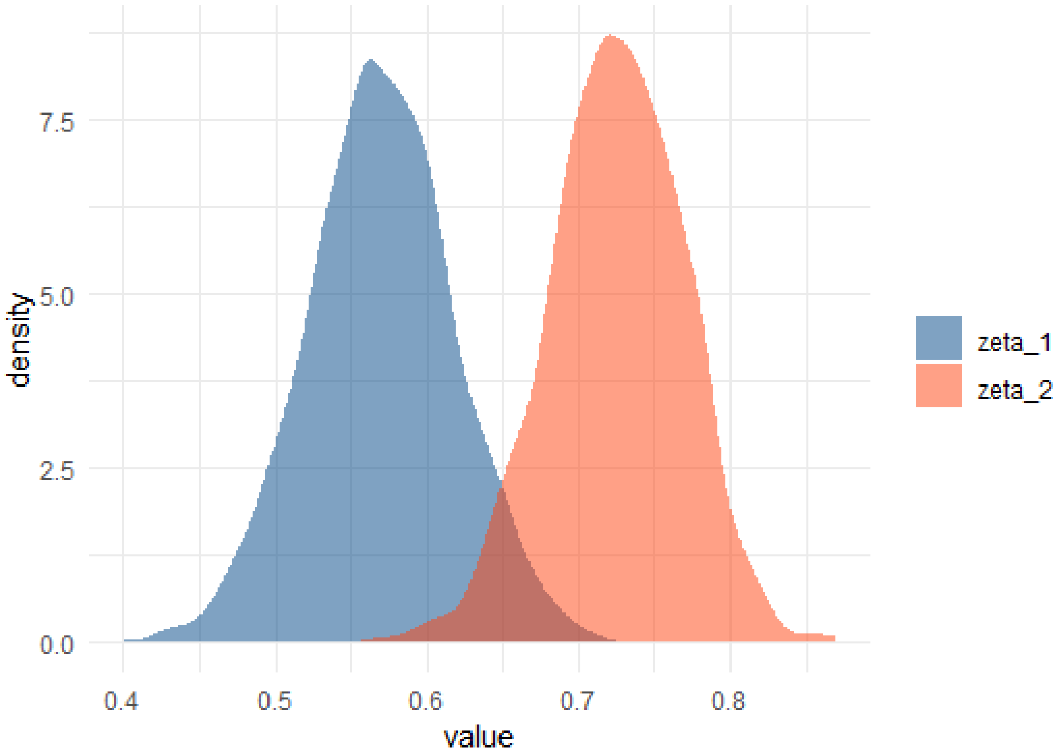
```
model{

    y[1:4] ~ dmulti(theta[1:4], N)

    theta[1] <- zeta[1]*zeta[2]
    theta[2] <- (1-zeta[1])*zeta[2]
    theta[3] <- zeta[1]*(1 - zeta[2])
    theta[4] <- (1-zeta[1])*(1 - zeta[2])


    zeta[1] ~ dbeta(2,2)
    zeta[2] ~ dbeta(2,2)

}

list(y = c(36,37,21,6), N = 100)
```

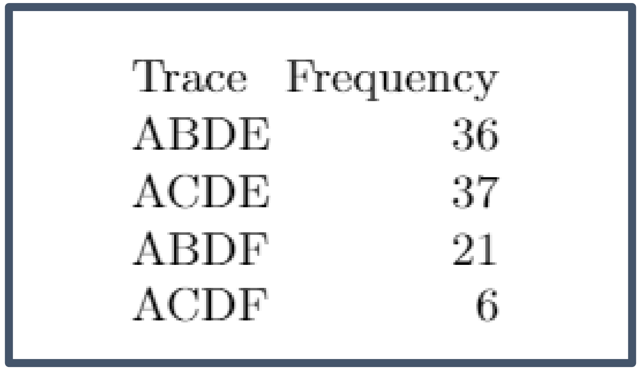
Table 3. Results Example 2 - Approach 1 - Parameters

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
theta[1]	0.4152	0.04403	4.635E-4	0.3305	0.4151	0.5026	1001	39000
theta[2]	0.3142	0.04089	4.048E-4	0.2366	0.3131	0.397	1001	39000
theta[3]	0.154	0.0288	2.883E-4	0.1033	0.1523	0.2158	1001	39000
theta[4]	0.1165	0.02347	2.475E-4	0.07534	0.1147	0.1675	1001	39000
zeta[1]	0.5693	0.04937	5.001E-4	0.4716	0.5698	0.6643	1001	39000
zeta[2]	0.7294	0.04455	4.611E-4	0.637	0.7316	0.81	1001	39000

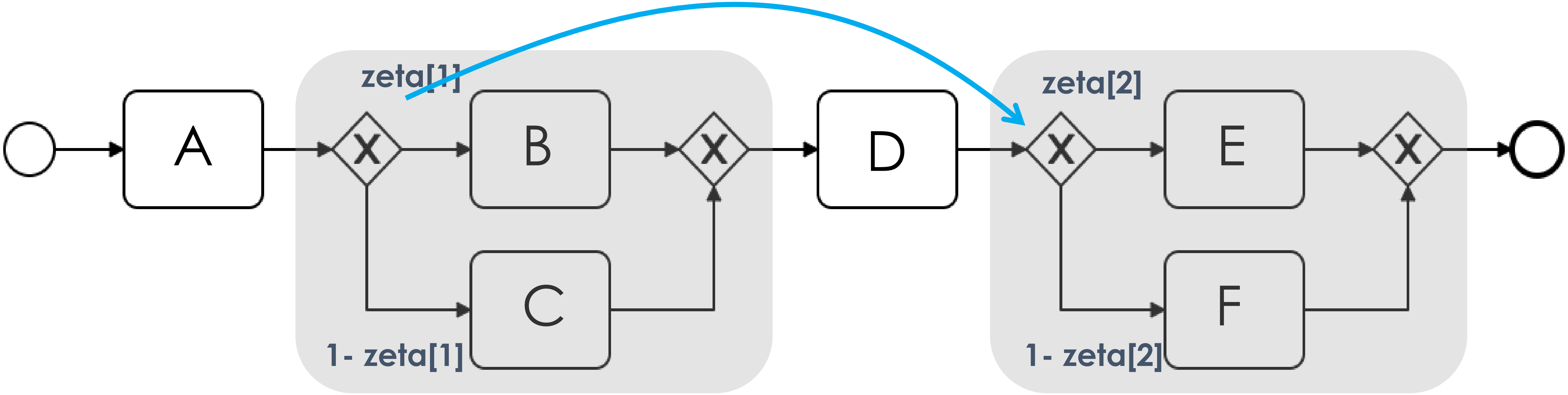
Table 4. Results Example 2 - Approach 1 - DIC

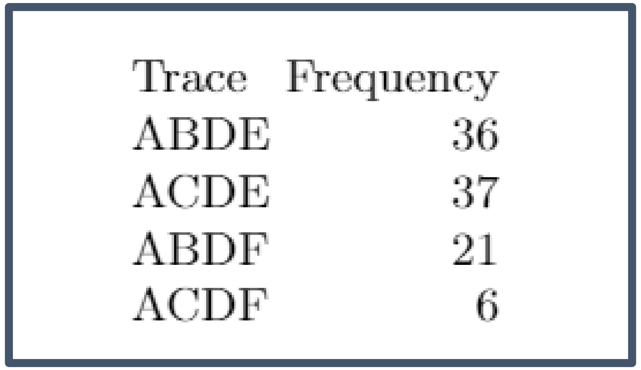
	Dbar	Dhat	DIC	pD
y	21.88	19.86	23.9	2.022
total	21.88	19.86	23.9	2.022



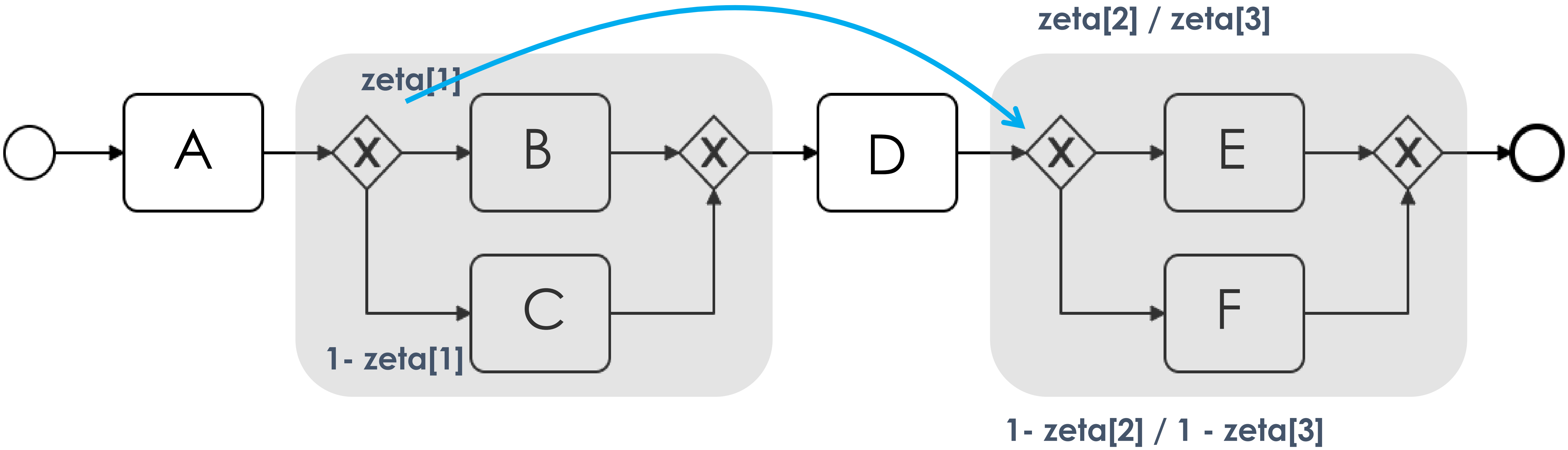


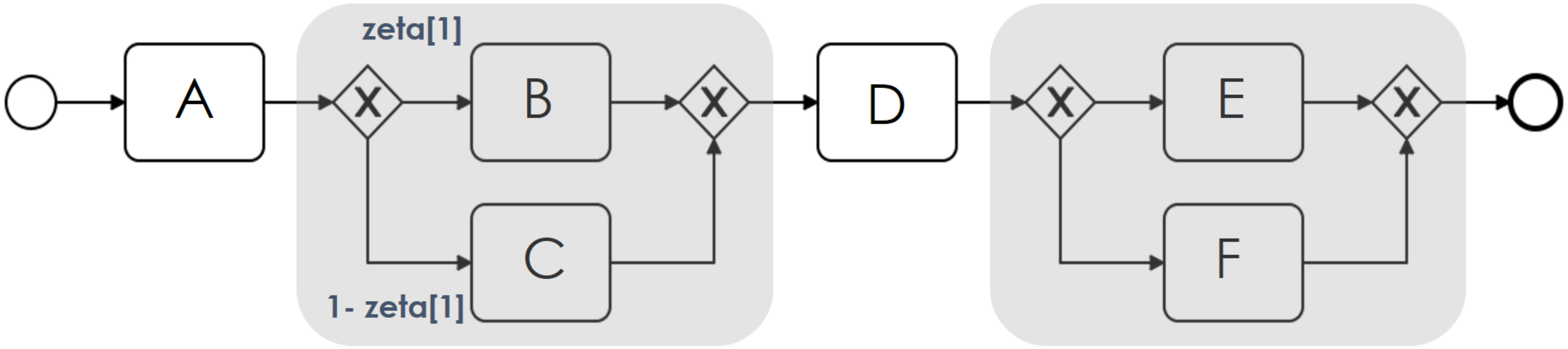
Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6





Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6



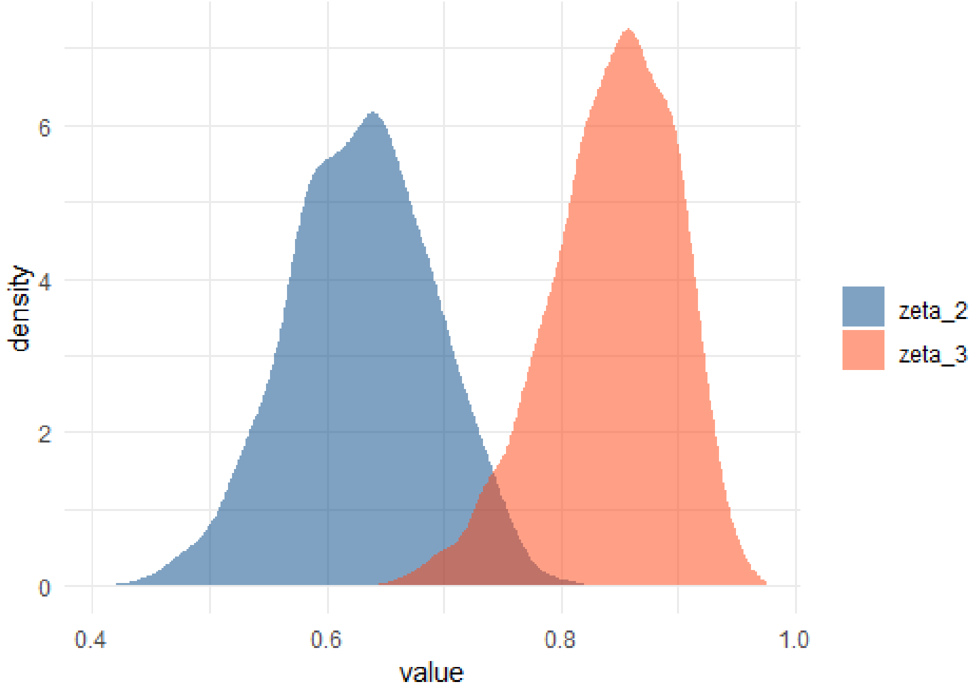


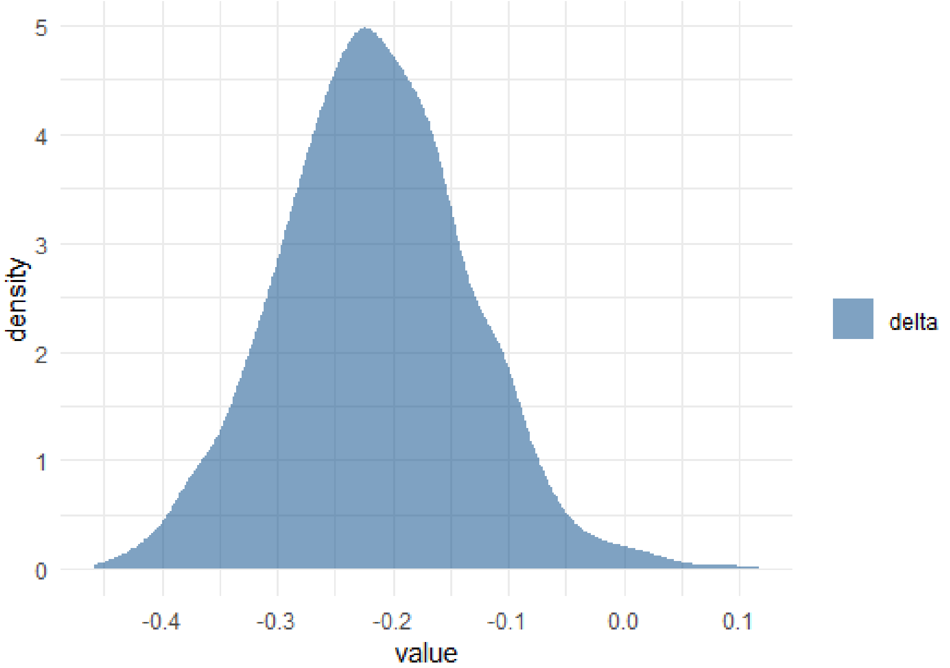
Trace	Frequency
ABDE	36
ACDE	37
ABDF	21
ACDF	6

```
model{

    y[1:4] ~ dmulti(theta[1:4], N)

    theta[1] <- zeta[1]*zeta[2]
    theta[2] <- (1-zeta[1])*zeta[3]
    theta[3] <- zeta[1]*(1 - zeta[2])
    theta[4] <- (1-zeta[1])*(1 - zeta[3])


    zeta[1] ~ dbeta(2,2)
    zeta[2] ~ dbeta(2,2)
    zeta[3] ~ dbeta(2,2)

    delta <- zeta[2] - zeta[3]
}

list(y = c(36,37,21,6), N = 100)
```

Table 5. Results Example 2 - Approach 2 - Parameters

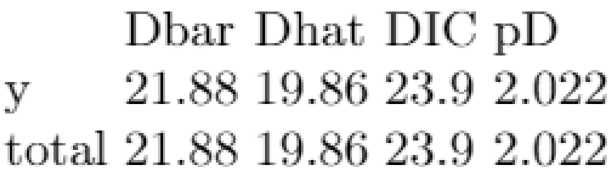
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
delta	-0.2285	0.08219	0.001177	-0.3869	-0.229	-0.06632	1001	19000
theta[1]	0.3589	0.04782	6.437E-4	0.2698	0.3579	0.4562	1001	19000
theta[2]	0.3706	0.04782	6.758E-4	0.2786	0.3698	0.4633	1001	19000
theta[3]	0.2098	0.03986	6.041E-4	0.1362	0.2084	0.2925	1001	19000
theta[4]	0.0606	0.02386	3.588E-4	0.02313	0.05739	0.1145	1001	19000
zeta[1]	0.5688	0.04889	7.399E-4	0.4734	0.5684	0.6619	1001	19000
zeta[2]	0.631	0.06294	8.945E-4	0.504	0.6318	0.7492	1001	19000
zeta[3]	0.8595	0.05282	7.747E-4	0.7437	0.866	0.9443	1001	19000





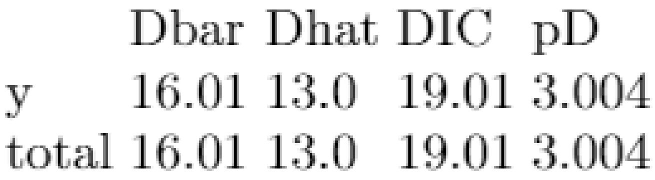
Which model fits best?

Fixed probability



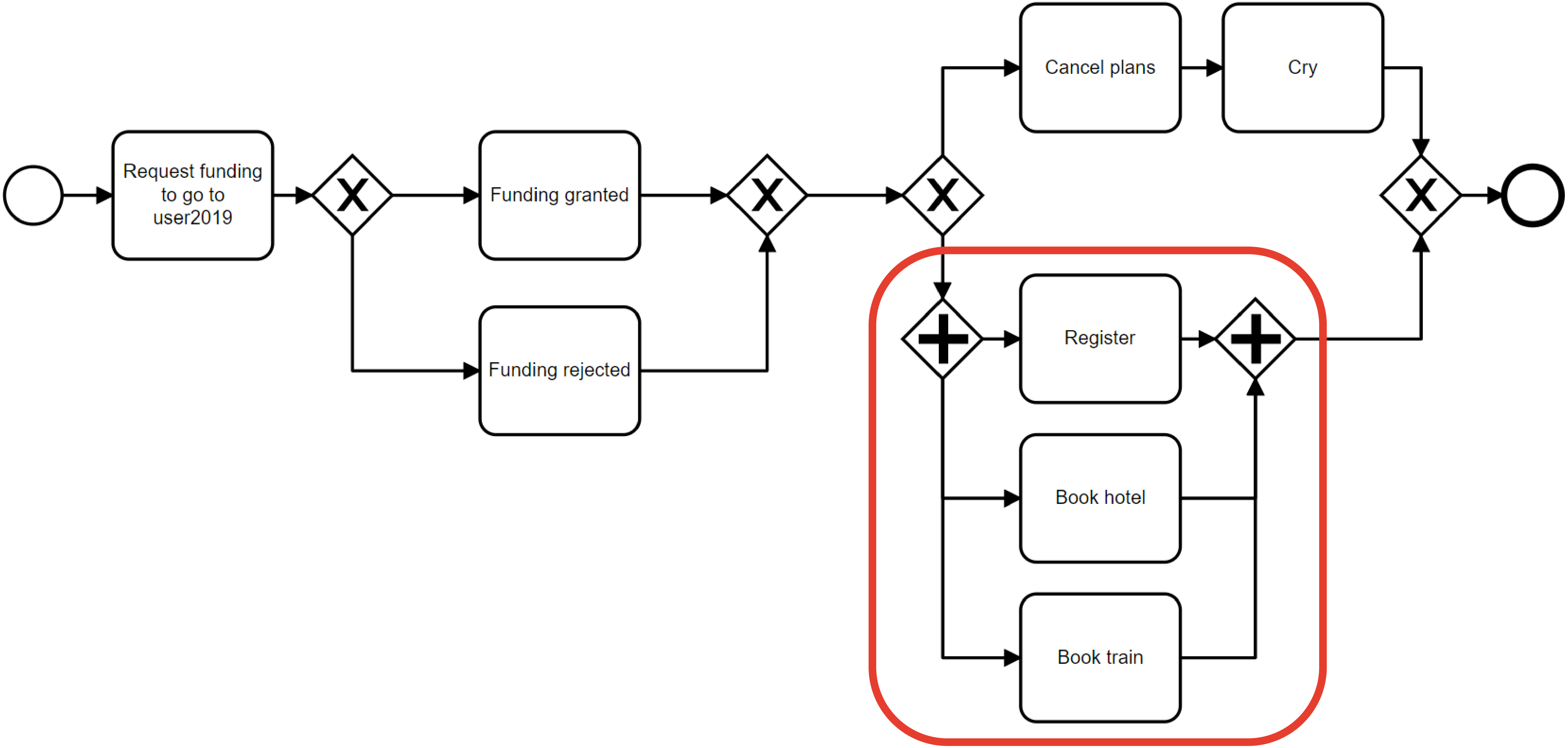
	Dbar	Dhat	DIC	pD
y	21.88	19.86	23.9	2.022
total	21.88	19.86	23.9	2.022

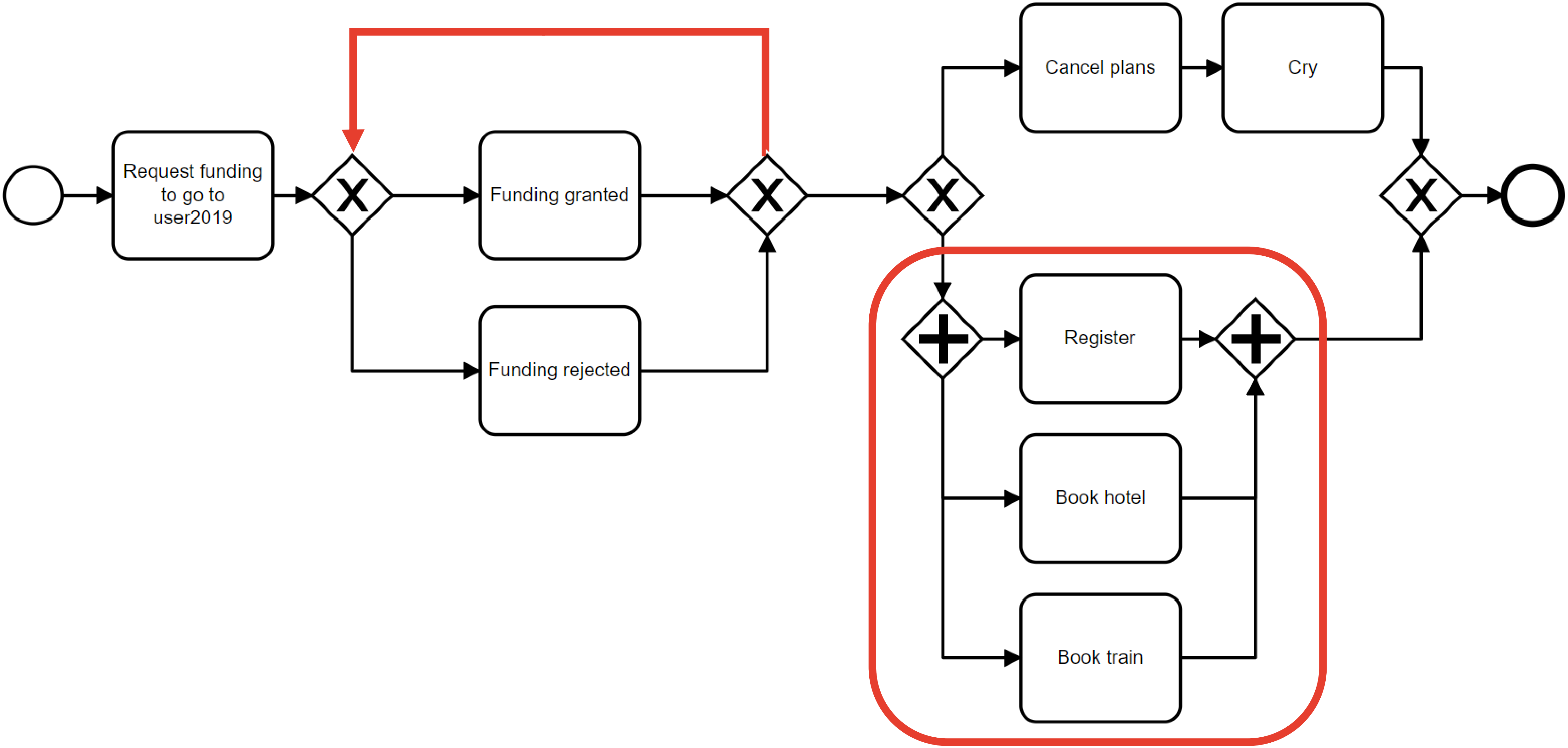
Dependent probability

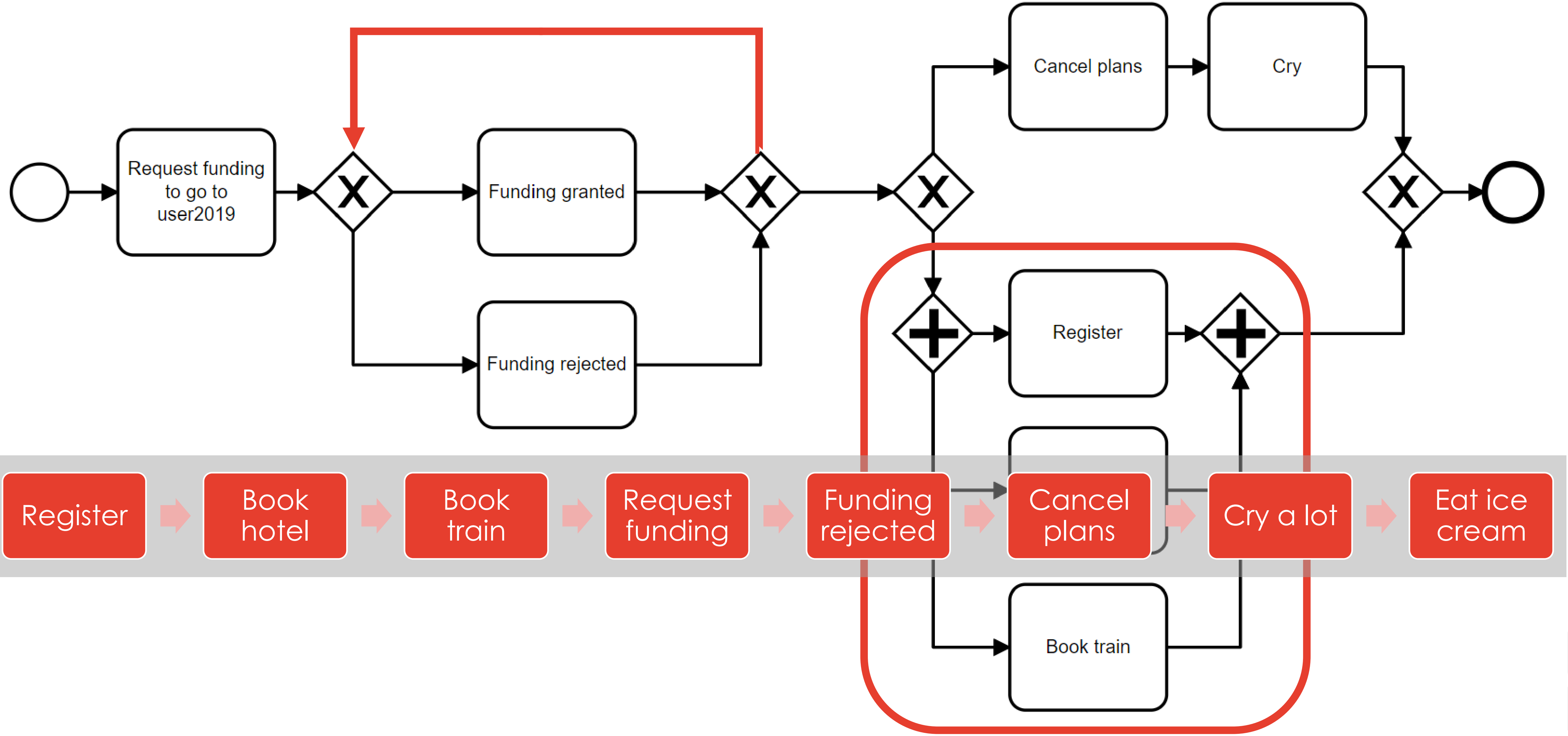


	Dbar	Dhat	DIC	pD
y	16.01	13.0	19.01	3.004
total	16.01	13.0	19.01	3.004

But?

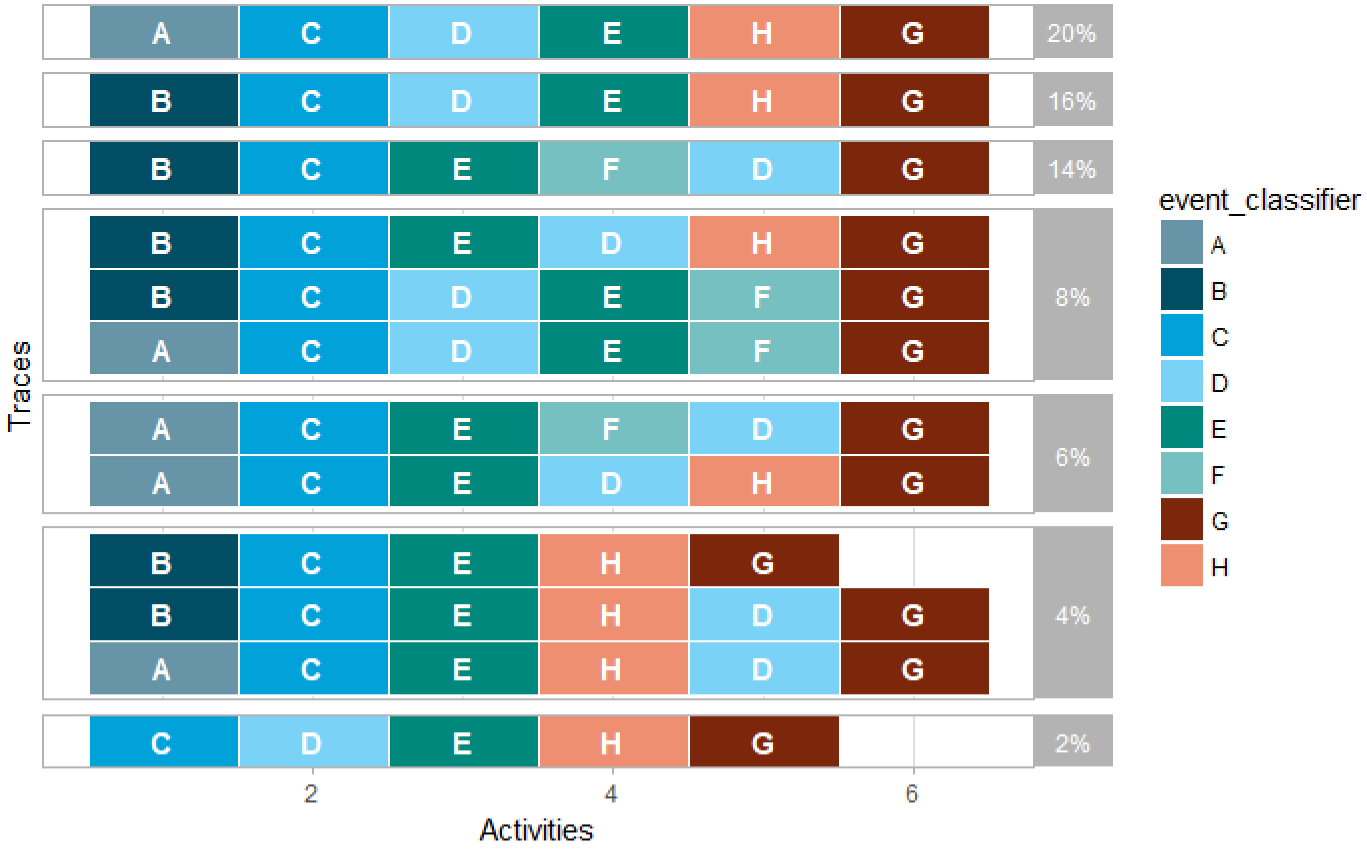

Request funding to go to user2019
Funding granted
Funding rejected
Cancel plans
Cry
Register
Book hotel
Book train

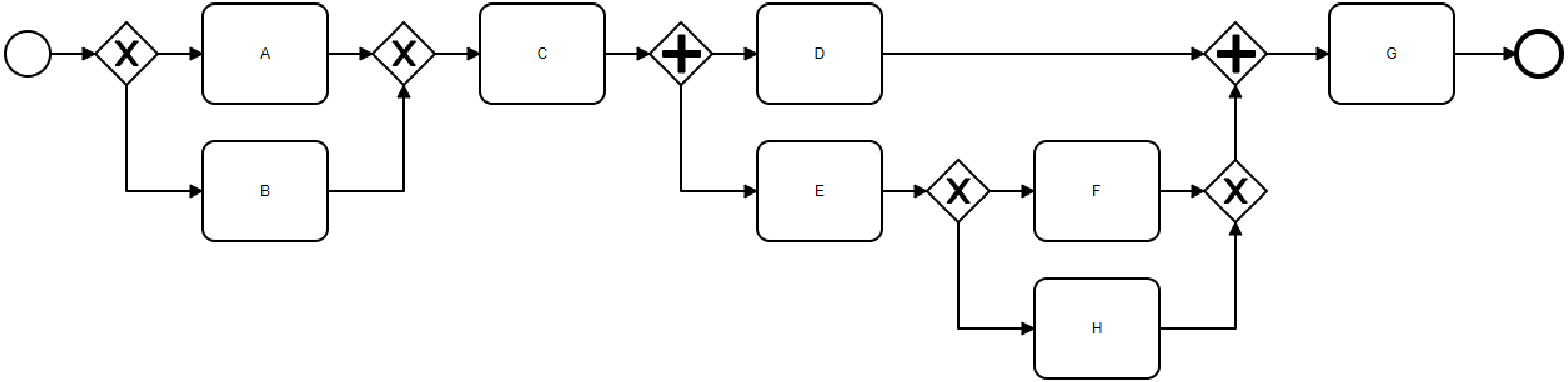

Request funding to go to user2019
Funding granted
Funding rejected
Cancel plans
Cry
Register
Book hotel
Book train



Register → Book hotel → Book train → Request funding → Funding rejected → Cancel plans → Cry a lot → Eat ice cream

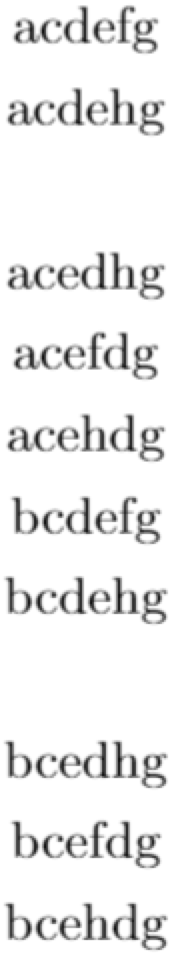
How to tackle these complexities?


A C D E H G 20%
B C D E H G 16%
B C E F D G 14%
B C E D H G
B C D E F G
A C D E F G 8%
A C E F D G
A C E D H G 6%
B C E H G
B C E H D G
A C E H D G 4%
C D E H G 2%
Traces
2
4
6
Activities
event_classifier
A
B
C
D
E
F
G
H


A
B
C
D
E
F
H
G

Create a prefix tree

based on the data



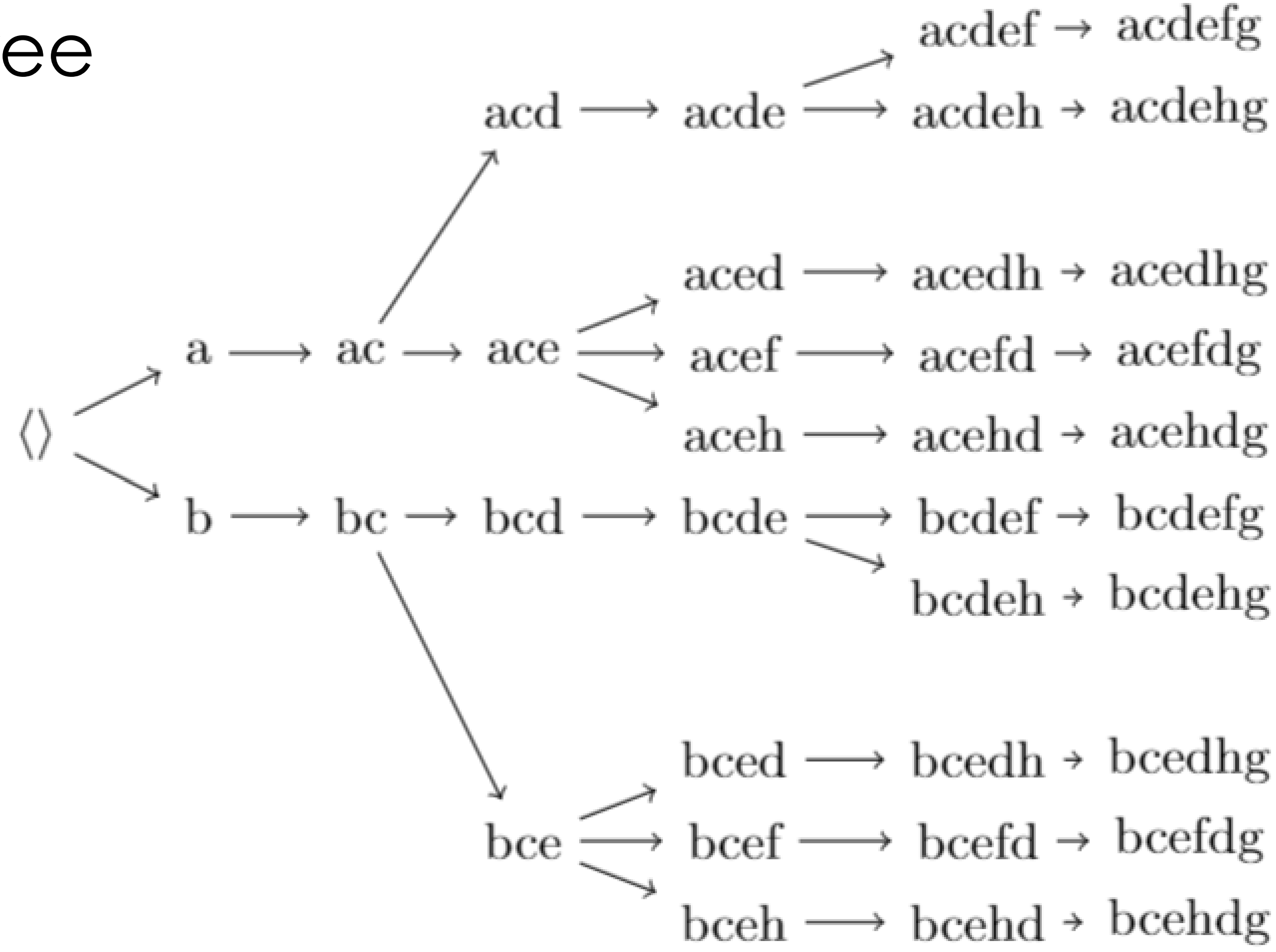
acdefg
acdehg

acedhg
acefdg
acehdg
bcdefg
bcdehg

bcedhg
bcefdg
bcehdg

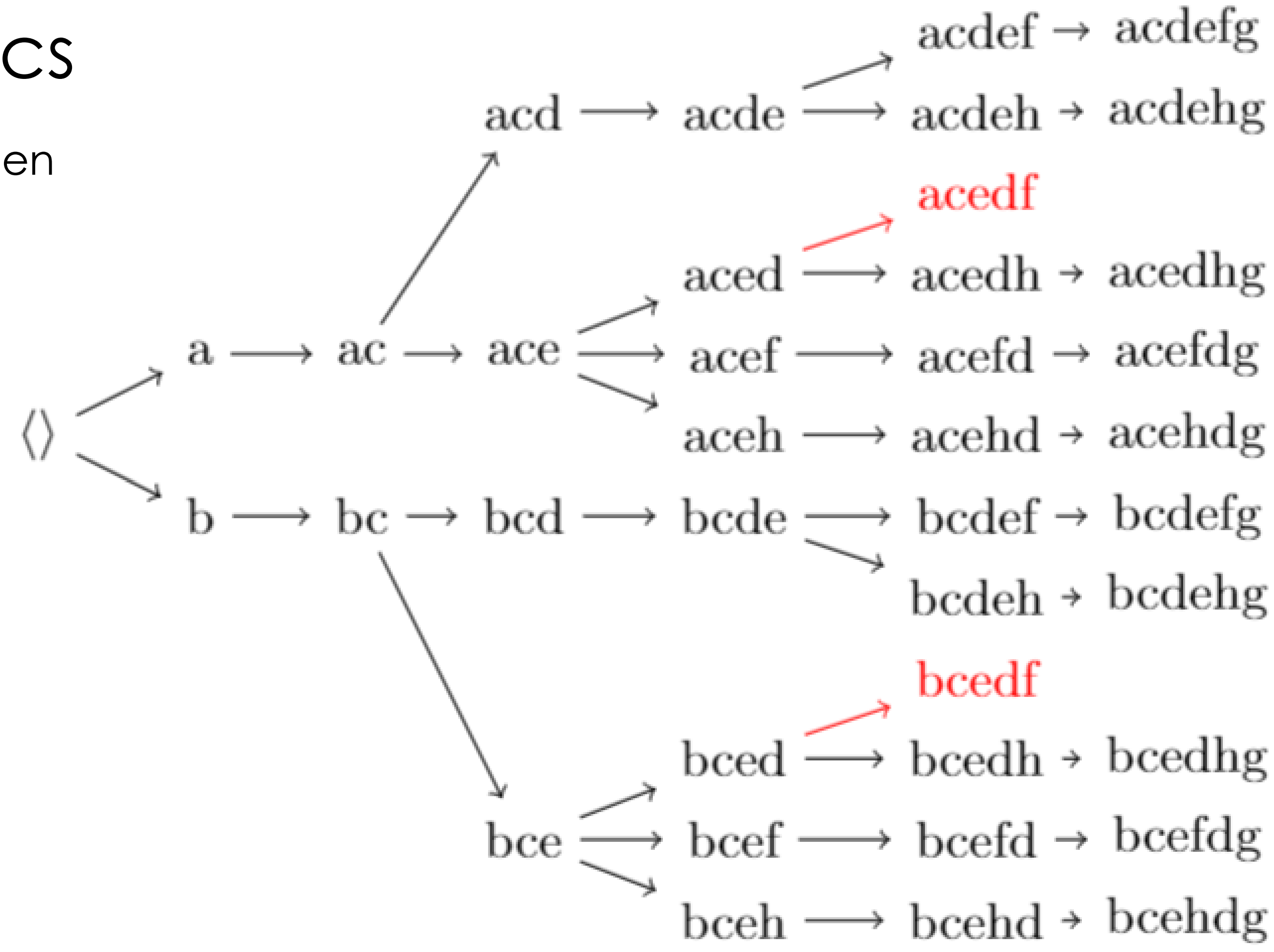
Create a prefix tree

based on the data

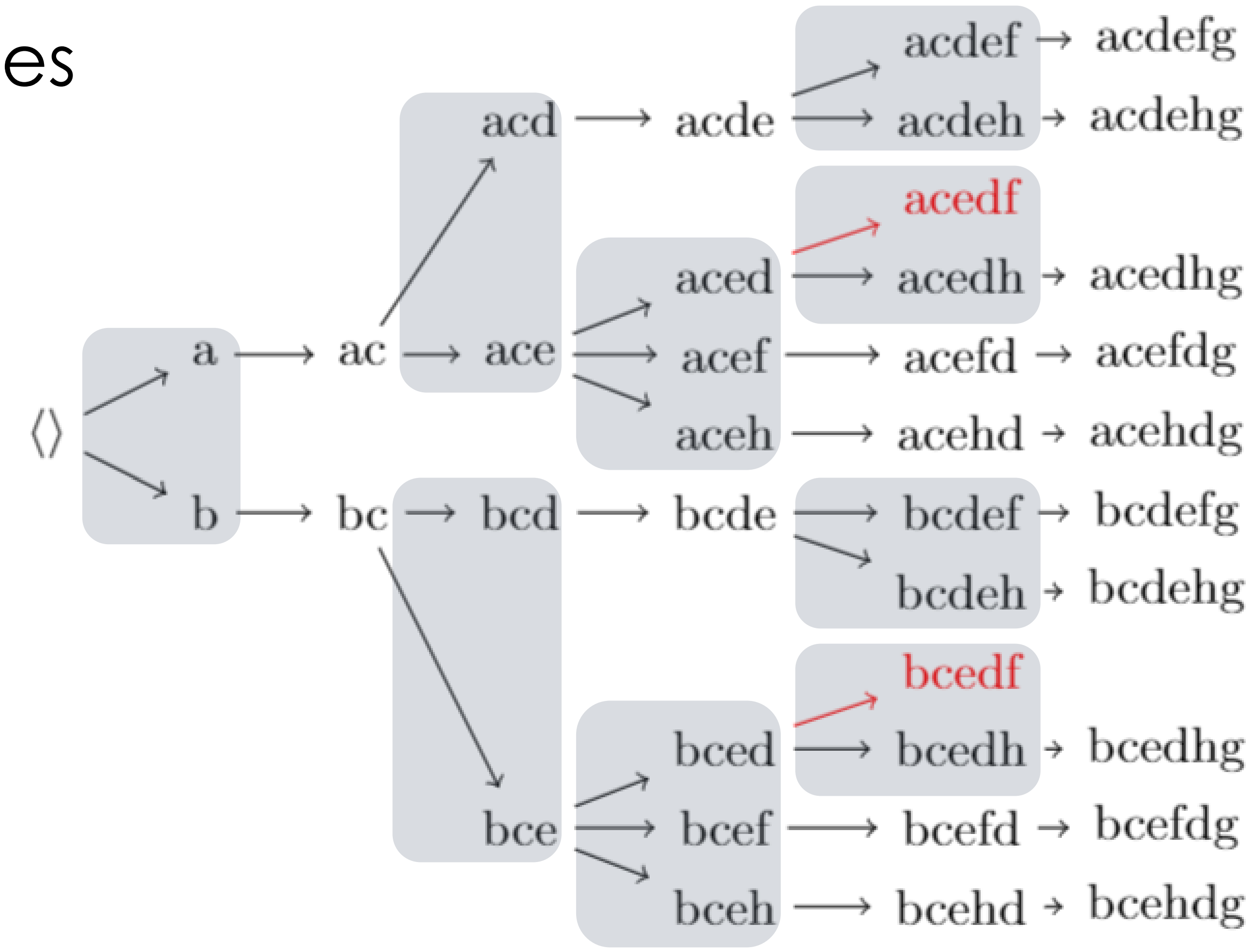


Add escaping arcs

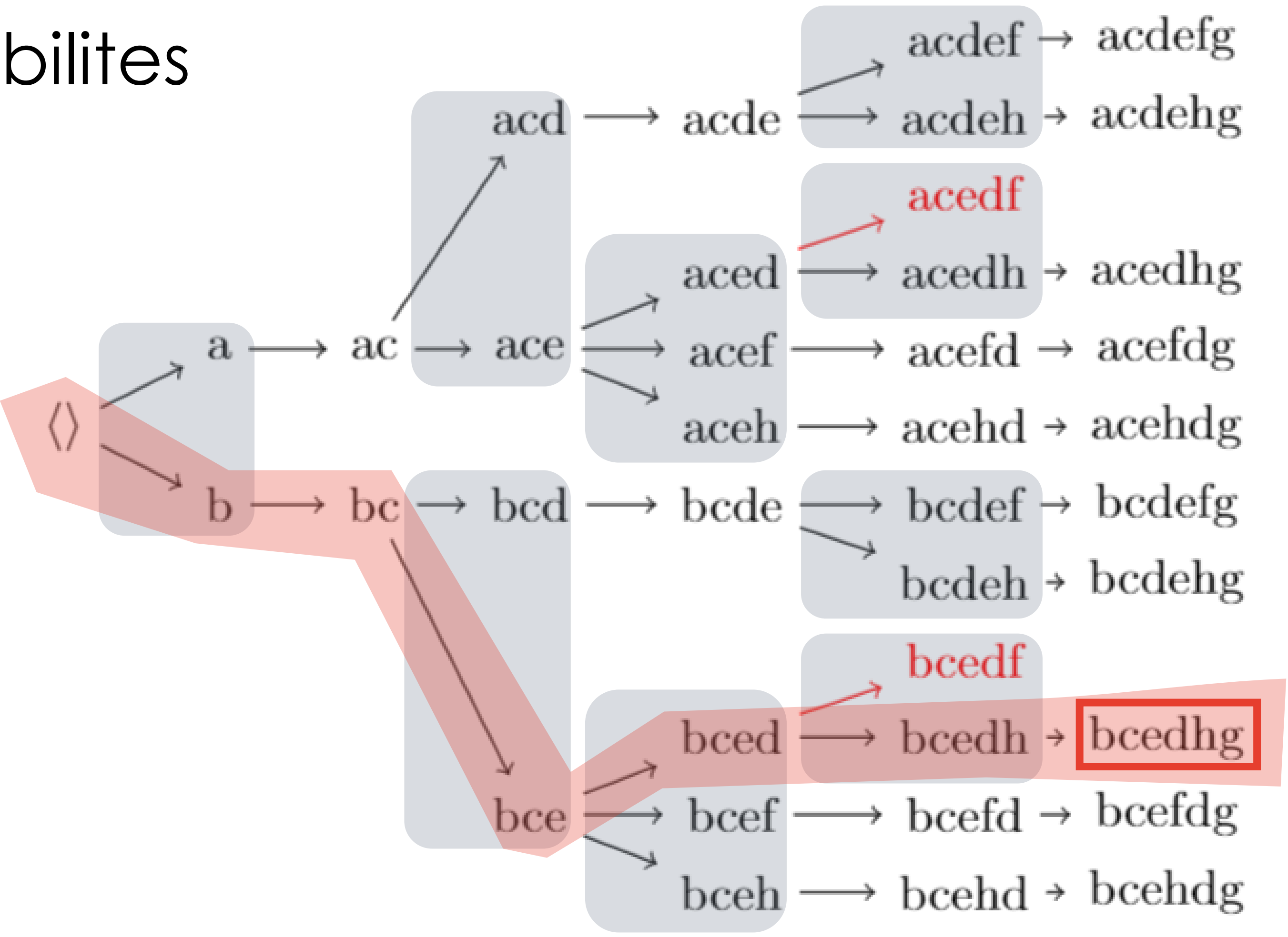
= choices possible but not seen



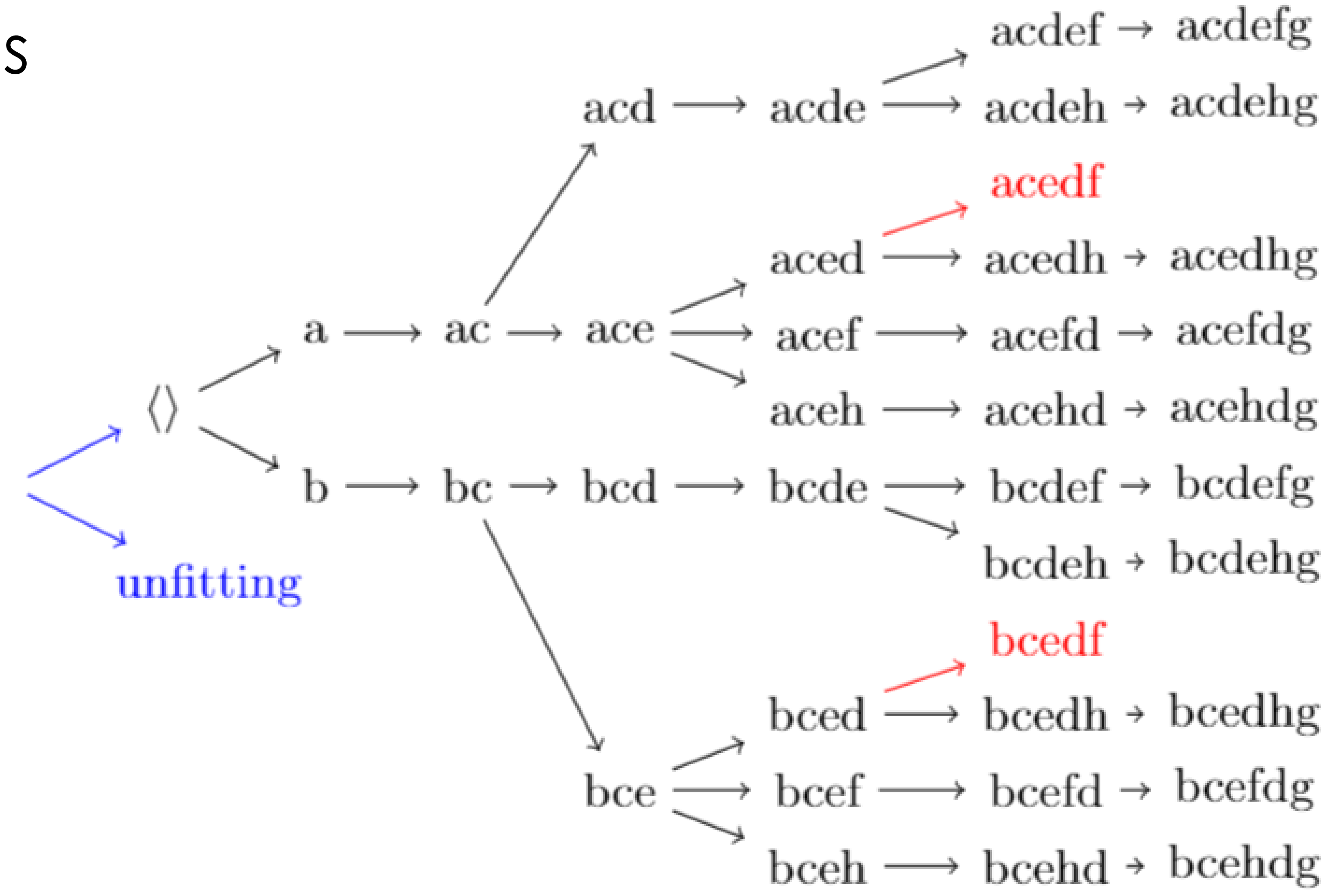
Define probabilites for each split



Combine probabilites for each trace



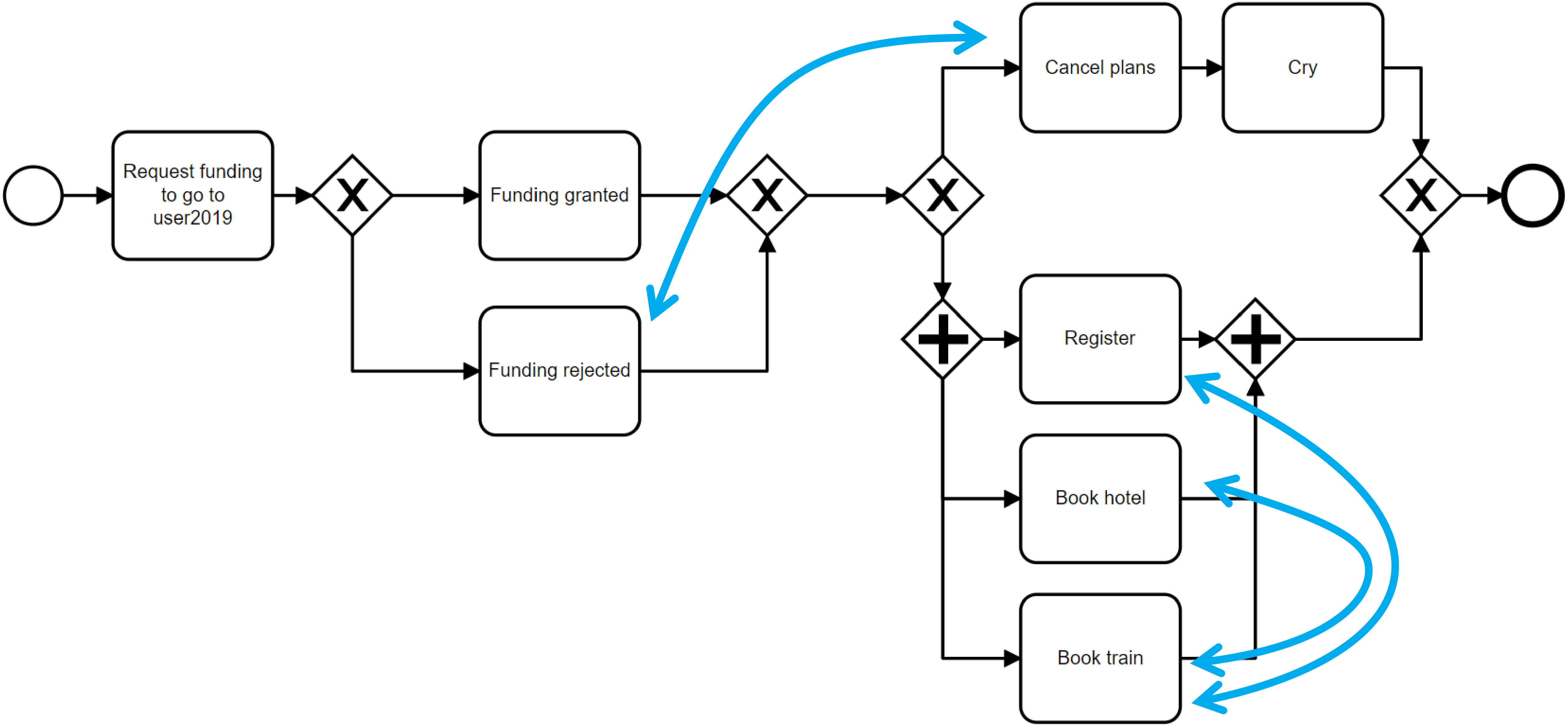
Define bypass



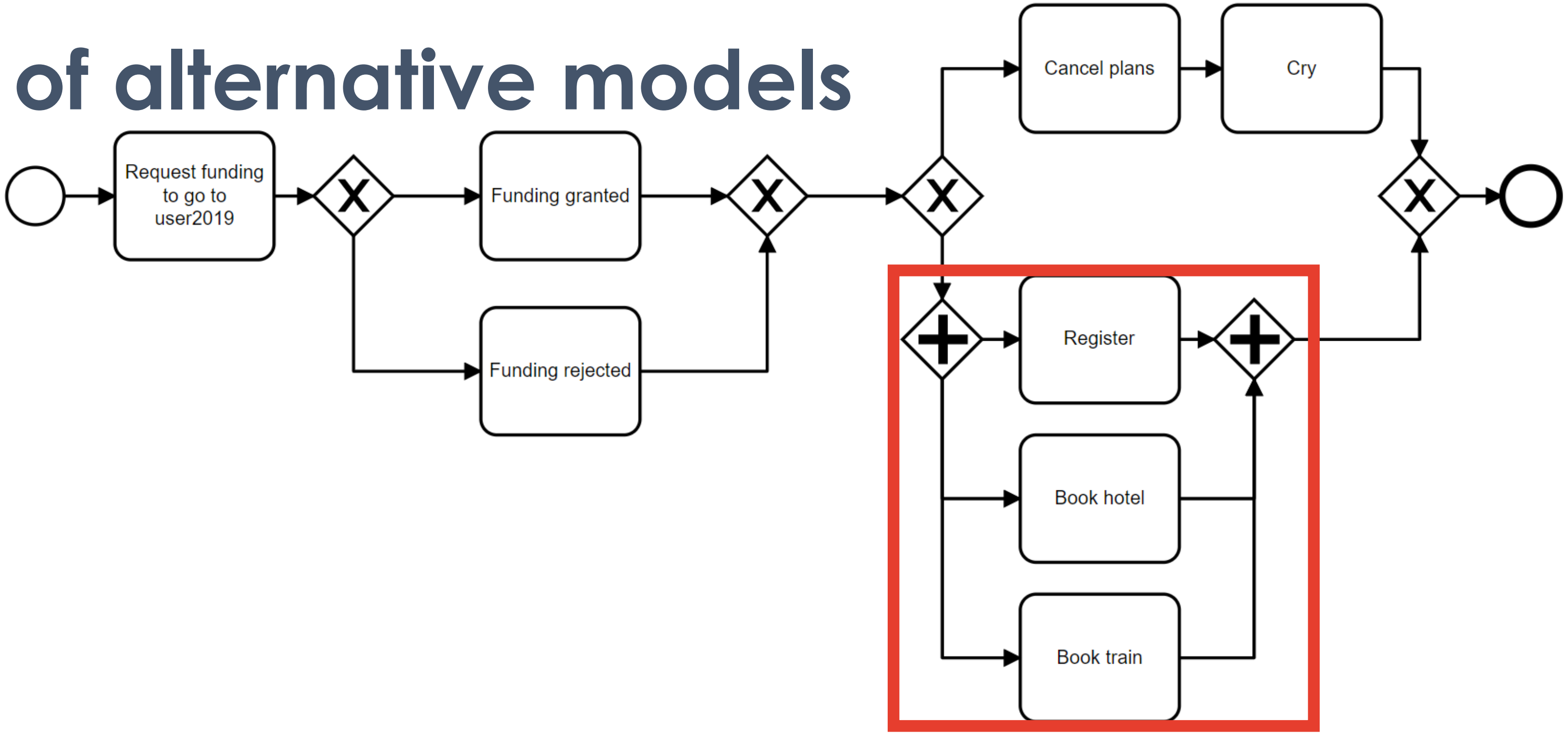
```
model{

    y[1:12] ~ dmulti(theta[1:12], N)

    theta[1] <- beta_f*beta[1]*beta[2]*(1-beta[6])
    theta[2] <- beta_f*(1-beta[1])*beta[2]*(1-beta[6])
    theta[3] <- beta_f*(1-beta[1])*(1-beta[2])*beta[4]
    theta[4] <- beta_f*(1-beta[1])*(1-beta[2])*beta[3]*(1-beta[6])
    theta[5] <- beta_f*(1-beta[1])*beta[2]*beta[6]
    theta[6] <- beta_f*beta[1]*beta[2]*beta[6]
    theta[7] <- beta_f*beta[1]*(1-beta[2])*beta[4]
    theta[8] <- beta_f*beta[1]*(1-beta[2])*beta[3]*(1-beta[6])
    theta[9] <- beta_f*(1-beta[1])*(1-beta[2])*beta[5]
    theta[10] <- beta_f*beta[1]*(1-beta[2])*beta[5]


    #nonfitting traces
    theta[11] <- (1-beta_f)

    #escaping arcs
    theta[12] <- beta_f*(1-beta[2])*beta[3]*beta[6]


    delta[1] <- beta[2] - beta[3]


    beta_f ~ dbeta(1,1)
    beta[1] ~ dbeta(1,1)
    beta[2] ~ dbeta(1,1)
    beta[3:5] ~ ddirich(alpha[])
    beta[6] ~ dbeta(1,1)



}
```

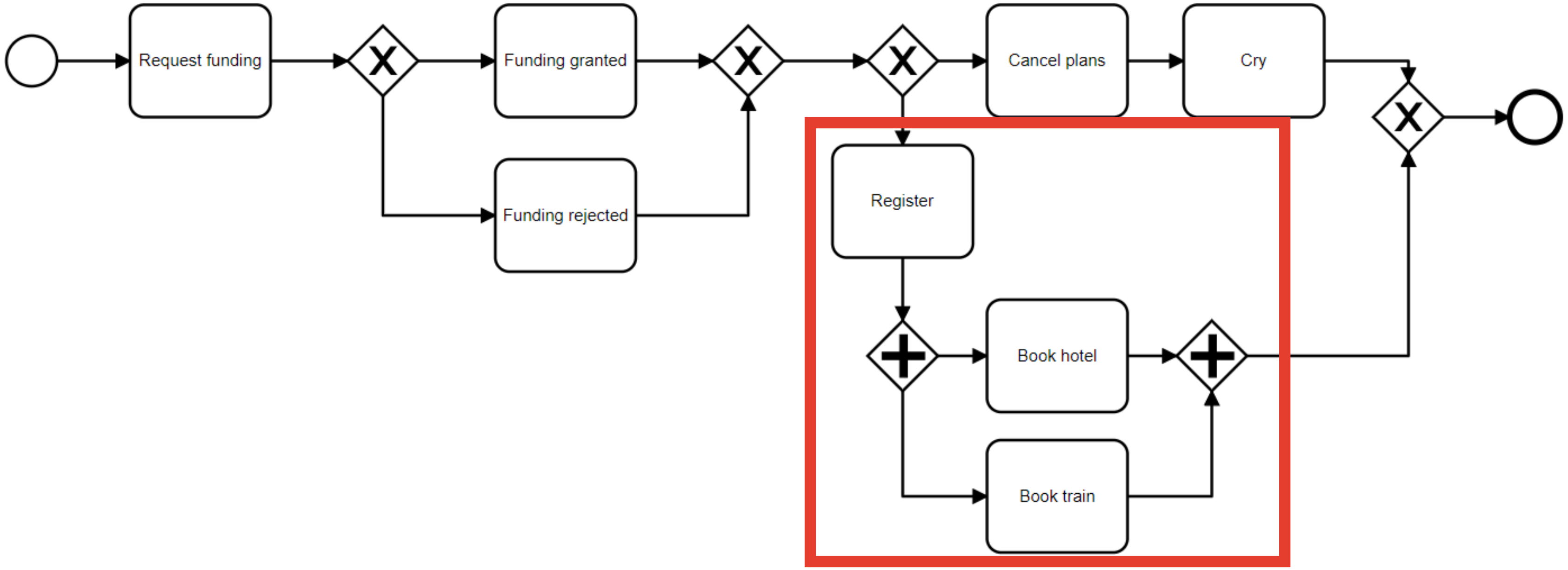
What can we do with it?

Test probabilistic dependencies

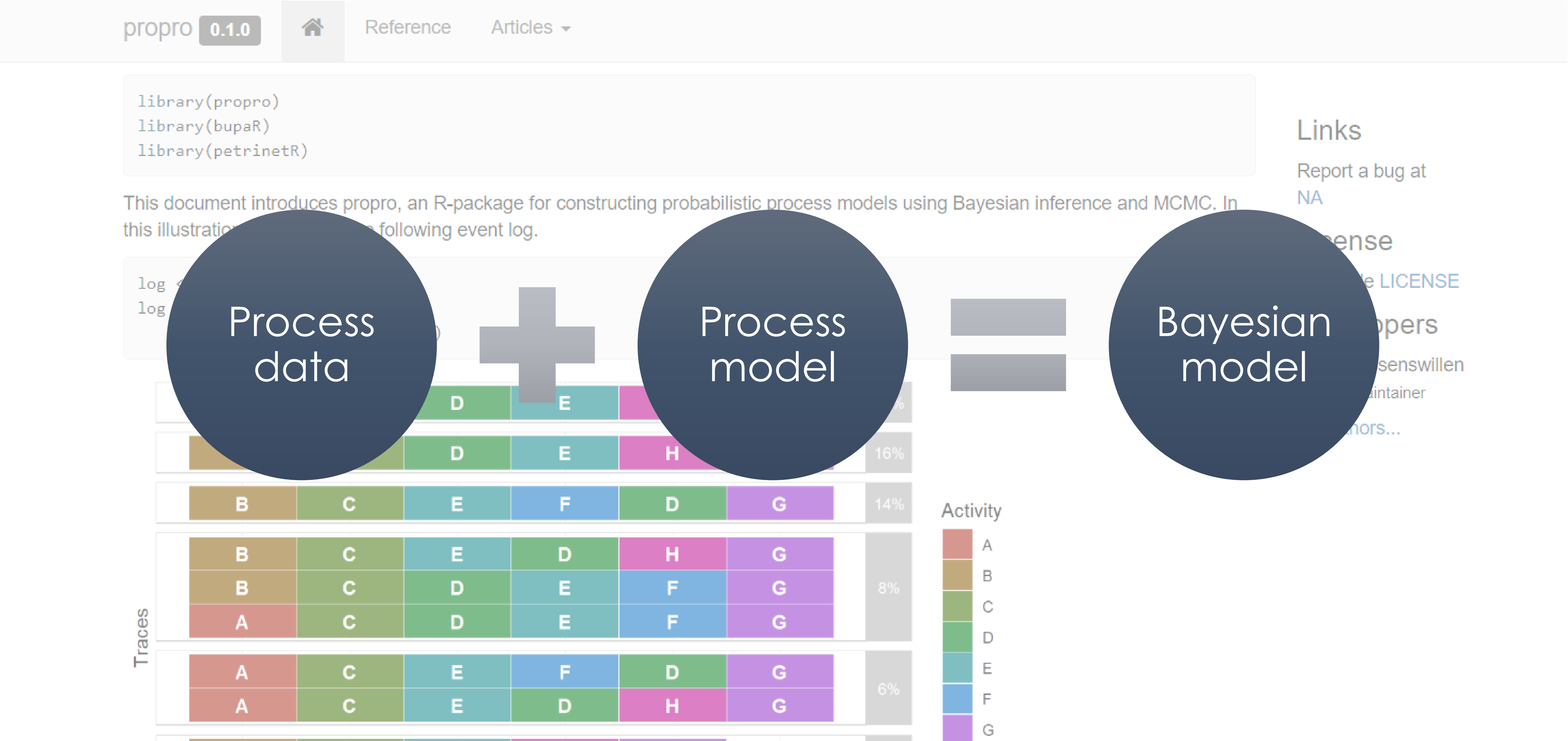


Test fit of alternative models





propro – Creating probabilistic process models



```
library(propro)
library(bupaR)
library(petrinetR)
```

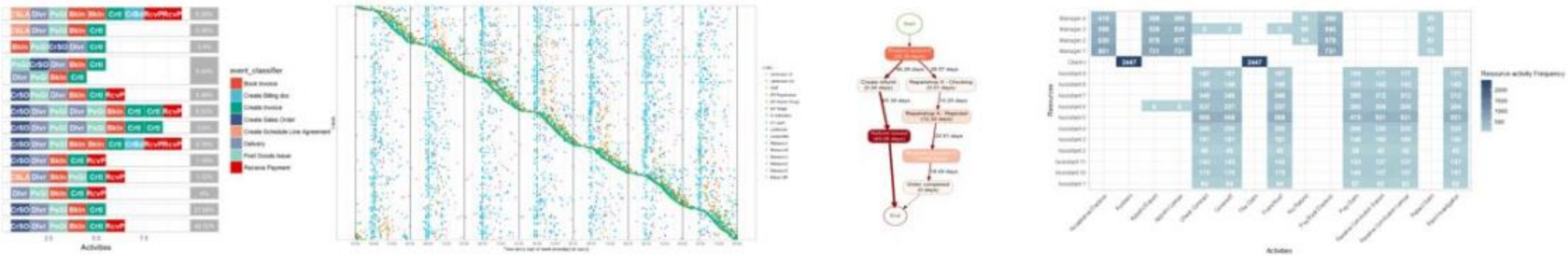
This document introduces propro, an R-package for constructing probabilistic process models using Bayesian inference and MCMC. In this illustratio following event log.

```
log
log
```

Process data + Process model = Bayesian model

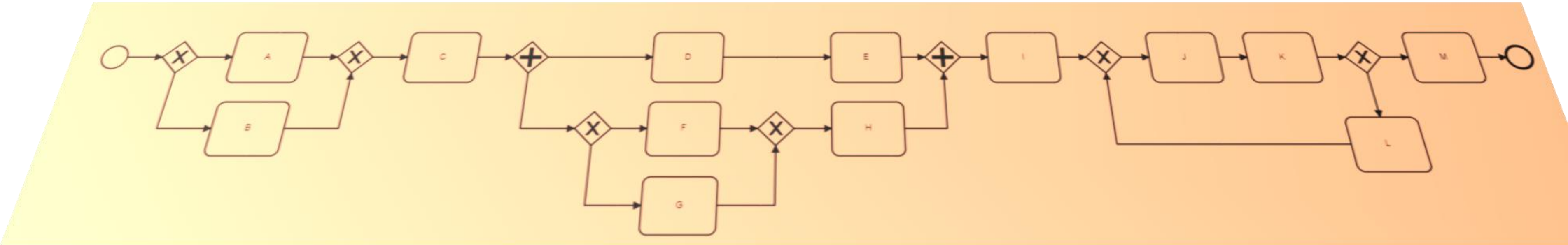
part of bupaR – Business process analysis in R

bupaR is an open-source, integrated suite of R-packages for the handling and analysis of business process data. It currently consists of 8 packages, including the central package, supporting different stages of a process mining workflow.

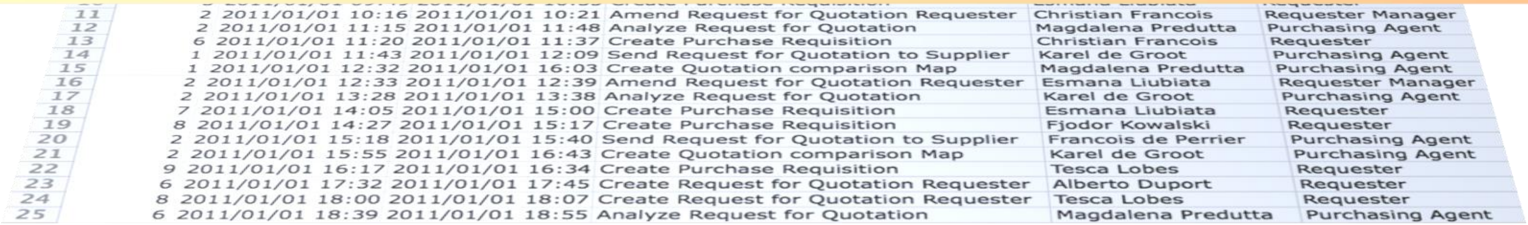


bupaR provides support for different stages in process analysis, such as importing event data, calculating descriptives, process monitoring and process visualization. The central package, **bupaR** includes basic functionality for creating event log objects in R. It contains several functions to get information about an event log and also provides specific event log versions of generic R functions.

Process Model



Event Log

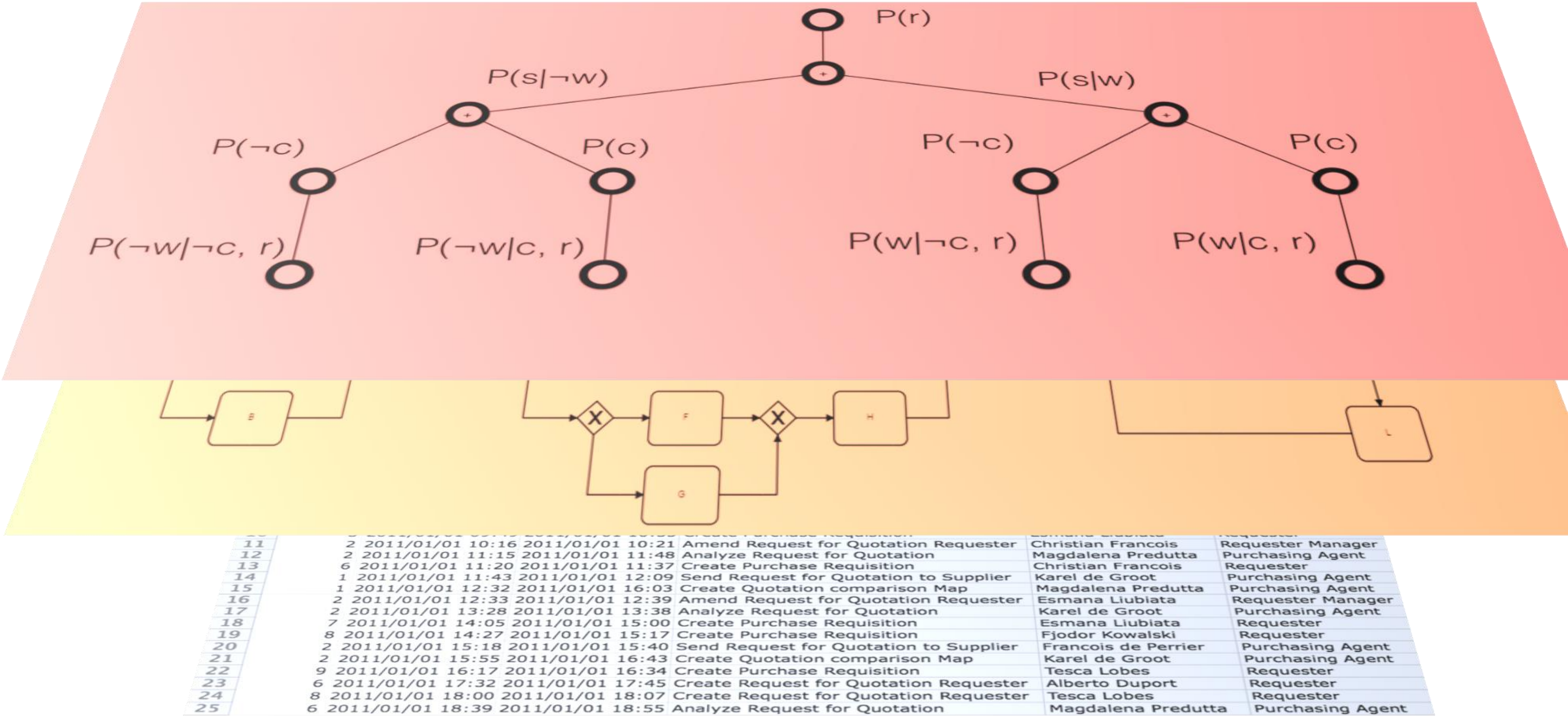


11	2	2011/01/01 10:16	2011/01/01 10:21	Amend Request for Quotation Requester	Christian Francois	Requester Manager
12	2	2011/01/01 11:15	2011/01/01 11:48	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent
13	6	2011/01/01 11:20	2011/01/01 11:37	Create Purchase Requisition	Christian Francois	Requester
14	1	2011/01/01 11:43	2011/01/01 12:09	Send Request for Quotation to Supplier	Karel de Groot	Purchasing Agent
15	1	2011/01/01 12:32	2011/01/01 16:03	Create Quotation comparison Map	Magdalena Predutta	Purchasing Agent
16	2	2011/01/01 12:33	2011/01/01 12:39	Amend Request for Quotation Requester	Esmana Liubiata	Requester Manager
17	2	2011/01/01 13:28	2011/01/01 13:38	Analyze Request for Quotation	Karel de Groot	Purchasing Agent
18	7	2011/01/01 14:05	2011/01/01 15:00	Create Purchase Requisition	Esmana Liubiata	Requester
19	8	2011/01/01 14:27	2011/01/01 15:17	Create Purchase Requisition	Fjodor Kowalski	Requester
20	2	2011/01/01 15:18	2011/01/01 15:40	Send Request for Quotation to Supplier	Francois de Perrier	Purchasing Agent
21	2	2011/01/01 15:55	2011/01/01 16:43	Create Quotation comparison Map	Karel de Groot	Purchasing Agent
22	9	2011/01/01 16:17	2011/01/01 16:34	Create Purchase Requisition	Tesca Lobes	Requester
23	6	2011/01/01 17:32	2011/01/01 17:45	Create Request for Quotation Requester	Alberto Duport	Requester
24	8	2011/01/01 18:00	2011/01/01 18:07	Create Request for Quotation Requester	Tesca Lobes	Requester
25	6	2011/01/01 18:39	2011/01/01 18:55	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent

Statistical Model

Process Model

Event Log



11	2	2011/01/01 10:16	2011/01/01 10:21	Amend Request for Quotation Requester	Christian Francois	Requester Manager
12	2	2011/01/01 11:15	2011/01/01 11:48	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent
13	6	2011/01/01 11:20	2011/01/01 11:37	Create Purchase Requisition	Christian Francois	Requester
14	1	2011/01/01 11:43	2011/01/01 12:09	Send Request for Quotation to Supplier	Karel de Groot	Purchasing Agent
15	1	2011/01/01 12:32	2011/01/01 16:03	Create Quotation comparison Map	Magdalena Predutta	Purchasing Agent
16	2	2011/01/01 12:33	2011/01/01 12:39	Amend Request for Quotation Requester	Esmana Liubiata	Requester Manager
17	2	2011/01/01 13:28	2011/01/01 13:38	Analyze Request for Quotation	Karel de Groot	Purchasing Agent
18	7	2011/01/01 14:05	2011/01/01 15:00	Create Purchase Requisition	Esmana Liubiata	Requester
19	8	2011/01/01 14:27	2011/01/01 15:17	Create Purchase Requisition	Fjodor Kowalski	Requester
20	2	2011/01/01 15:18	2011/01/01 15:40	Send Request for Quotation to Supplier	Francois de Perrier	Purchasing Agent
21	2	2011/01/01 15:55	2011/01/01 16:43	Create Quotation comparison Map	Karel de Groot	Purchasing Agent
22	9	2011/01/01 16:17	2011/01/01 16:34	Create Purchase Requisition	Tesca Lobes	Requester
23	6	2011/01/01 17:32	2011/01/01 17:45	Create Request for Quotation Requester	Alberto Duport	Requester
24	8	2011/01/01 18:00	2011/01/01 18:07	Create Request for Quotation Requester	Tesca Lobes	Requester
25	6	2011/01/01 18:39	2011/01/01 18:55	Analyze Request for Quotation	Magdalena Predutta	Purchasing Agent

Taking process understanding to a next level

dr. Gert Janssenswillen
@gjanssenswillen
gert.janssenswillen@uhasselt.be
FWO PhD Fellow
Research Group Business Informatics

T +32(0)11 26 86 39

www.uhasselt.be - www.bupar.net
Universiteit Hasselt - Campus Diepenbeek
Agoralaan Gebouw D- B-3590 Diepenbeek