# User study example

1. Participant needs to agree with informed consent
2. Participant fills in demographic information
3. Participant gets the following explanation:

Explanation of the model and visualization with a simple example

Basic event explanation.
Different activities occur after each other, for example, an occurrence of activity A followed by an occurrence of activity B and then an occurrence of activity C. These activity occurrences belong to the same entity or case and are ordered chronologically. The activity occurrences of a case form a trace. Each unique trace in a log is called a variant.
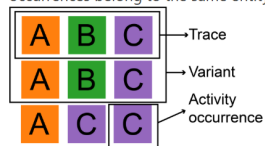

[model explanation of simple example based on the visualization the participant is about to see, see the options below. It is the same example for all visualizations.]

Interactions and filtering.
Possible interactions with the visualization include zooming with the mouse wheel or panning/dragging by using the holding the left mouse button. Also, there is a slider to filter out less frequent pairs of consecutive activity occurrences:

Frequency filtering:

 100%
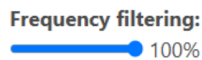
For example, 100% means that all the activity occurrences are included. The lower the percentage the more pairs of consecutive activity occurrences are filtered out. 80% does not necessarily mean only 80% off all the consecutive activity occurrences are kept. For example, assume we have the trace ABBCAC, see Figure below. Assume B followed by C and A followed by C are frequent pairs of consecutive activity occurrences in the current filter setting. However, C followed by A is not frequent enough. Then all four activity occurrences are included to have all the pairs of frequent consecutive activity occurrences. Assume this trace occurs 5 time then all the edges (B-C, C-A, A-C) will have weight 5.
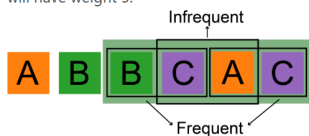
If B followed by B would be frequent and A followed by C, but B followed by C and C followed by A are infrequent, we filter out the occurrence of activity C.
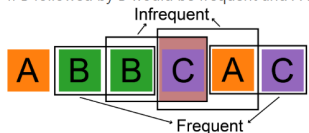
As a final example, if A followed by B is frequent and B followed by C, but B followed by B is not frequent. We filter out one B.
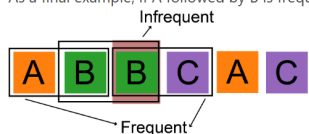
Next

Options for visualization explanations:

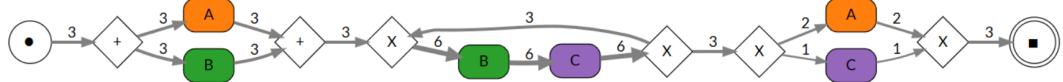## a. Variant timeline

Model explanation.

The visualization you will see next will display the activity occurrences per variant as horizontally aligned colored blocks based on the activity. The frequency of each variant is displayed in front of the variant. And a legend gives information on which color represents which activity. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. The variants are ordered based on frequency. Several relations can be present in the visualization. For example, parallelism, loops or choice. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order. For example, the first two activity occurrences of both traces. Either A is followed by B or B is followed by A. Loops are consecutive activities that are repeated again. In the Figure you can see that in both variants B followed by C is repeated twice. Finally, a choice is, for example, picking a certain activity occurrence in one variant and another activity occurrence in another variant. In the example below you can see one example of this at the end, C is either followed by A or by C.
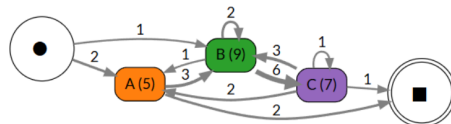

b. BPMN

Model explanation.

The visualization you will see next, a bpmn, will display the activity occurrences per variant as nodes and edges. Activity occurrences are represented by nodes. One unique activity can have multiple nodes. The edges indicate how many activity occurrences progress from one activity to the other. The start is indicated by a circle with a smaller black circle inside. The end is indicated by a double circle with a smaller square inside. The frequency of two consecutive activities is displayed on the edges. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. Several relations can be present in the model. For example, parallelism, loops or choice. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order indicated by diamonds with a plus sign inside. A choice is to pick a certain activity occurrence in one variant and another activity occurrence in another variant, indicated by diamonds with a X in them. In the example below you can see one variant either ends with A or C. Loops are consecutive activities that are repeated again. In the Figure you can see that B followed by C is repeated three times in total. You can see this by an edges looping back after a choice diamond to a choice diamond before this loop happened.

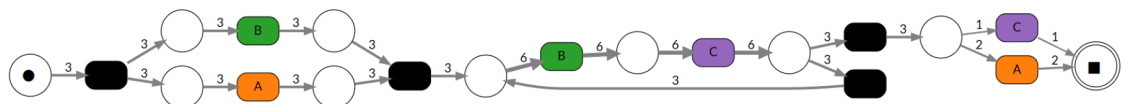## c. DFG

Model explanation.

The visualization you will see next, a dfg (directly follows graph), will display the activity occurrences per variant as nodes and edges. Each unique activity occurrence is aggregated in one node, e.g., there is only one node with activity A etc. The edges indicate how many activity occurrences progress from one activity to the other. The start is indicated by a circle with a smaller black circle inside. The end is indicated by a double circle with a smaller square inside. The frequency of each unique activity is displayed in the nodes and the frequency of two consecutive activities on the edges. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. Several relations can be present in the log. For example, parallelism, loops or choice. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order. For example, the first two activity occurrences of both variants. Either A is followed by B or B is followed by A. Loops are consecutive activities that are repeated again. In the Figure you can see that activity B has a self-loop of 2. Also, in the log, B followed by C is repeated twice within one variant. Finally, sometimes a choice is made, this means we pick a certain activity occurrence in one variant and another activity occurrence in another variant. In the example below you can see one variant either ends with A or C.

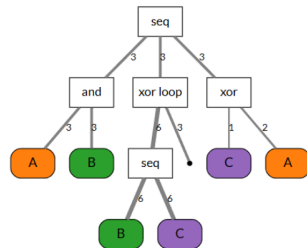## d. Petri net

Model explanation.

The visualization you will see next, a petri net, will display the activity occurrences per variant as nodes and edges. The activity occurrences are represented by nodes. One unique activity can have multiple nodes. The start is indicated by a circle with a smaller black circle inside. The end is indicated by a double circle with a smaller square inside. The frequency of two consecutive activities is displayed on the edges. A petri net has places, white circles, and transitions, the black or colored rectangles with an activity label. The edges are always from places to transitions or the other way around. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. The transitions happens if there are enough tokens in the places that have an edge to the transition and creates tokens in each place connected to it. For example, in the figure below, in the beginning the black rectangle would consume a token at the start and then create a token in both the white circles after it. Then both activity occurrence A and B consume a token and create a token. Both these created tokens in both the white circles are consumed by the second black rectangle to create a token in the white circle after it, etc. Several relations can be present in the log. For example, parallelism, loops or choice. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order. For example, the first two activity occurrences of both traces. Either A is followed by B or B is followed by A. You can recognize this by a transition, black rectangle, having multiple places to it afterward, and later on the in the models these places merge again in one transition. Loops are consecutive activities that are repeated again. In the Figure you can see that B followed by C is repeated three times in total. You can see this by a transition and edges looping back. Finally, sometimes a choice is to pick a certain activity occurrence in one variant and another activity occurrence in another variant. In the example below you can see one variant either ends with A or C. You can recognize this by one place followed by multiple transitions.
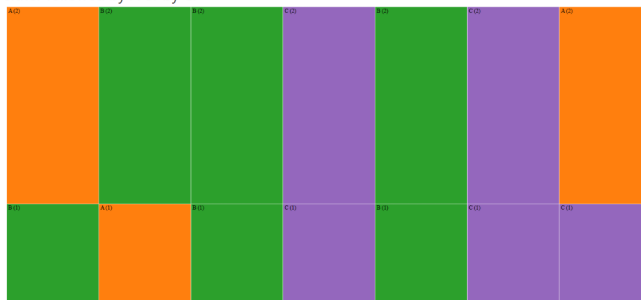
e. Process tree

Model explanation.

The visualization you will see next, a process tree, will display the activity occurrences per variant as nodes and edges in the form of a tree. Activity occurrences are represented by nodes. One unique activity can have multiple nodes. The start is indicated by the root. The end is indicated by the most right activity occurrence. The frequency of an activity is displayed on the edges. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. Several relations can be present in the log. A rectangle with the text seq inside means that you should sequentially walk the paths to its children from left to right in order to see a variant. Also, there can be parallelism, loops or choices. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order indicated by a rectangle with the text and inside. For example, the first two activity occurrences of both variants. Either A is followed by B or B is followed by A. A choice is to pick a certain activity occurrence in one variant and another activity occurrence in another variant, indicated by a rectangle with the text xor. In the example below you can see one variant either ends with A or C. Loops are consecutive activities that are repeated again. In the Figure you can see that B followed by C is repeated six times in total. You can see this by a rectangle with the text xor loop inside. So, in total, we have three traces and first A and B happen in any order, the and rectangle, this is the walk through the most left branch in this tree. Since the root node is a seq node, the branches are sequential. We move to the next branch, middle one, where we see that all three traces enter into a loop, this loop happens in total in the entire log 6 times and contains the two sequential activities B and C. The left branch indicates how often we loop back to this sequence of B followed by C again.

f. Icicle plot
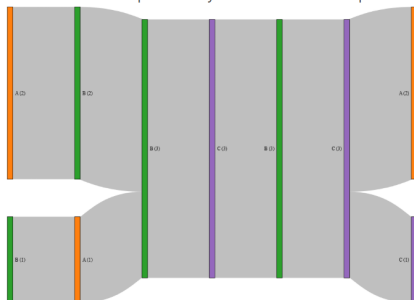
Model explanation.

The visualization you will see next, hierarchical icicle plot, will display the activity occurrences per variant as horizontally aligned colored rectangles based on the activity. If the previous activity occurrences of several variants at a certain index (e.g. the 5th activity occurrence) are exactly the same they are merged together. The frequency of each activity occurrence is displayed in the rectangle. If variants are different in the beginning but for example have the same activity occurrences later on they are treated and displayed on separate lines. And a legend gives information on which color represents which activity. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. Several relations can be present in the visualization. For example, parallelism, loops or choice. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order. For example, the first two activity occurrences of both variants. Either A is followed by B or B is followed by A. Loops are consecutive activities that are repeated again. In the Figure you can see that within one variant B followed by C is repeated twice. Finally, a choice is, for example, picking a certain activity occurrence in one variant and another activity occurrence in another variant. In the example below you can see one example of this at the end, C is either followed by A or by C.

g. Sankey

Model explanation.

The visualization you will see next, a sankey flow diagram, will display the activity occurrences per variant as connected colored rectangles based on the activity. At each index (e.g., all the first, second or third activity occurrences), the activity occurrences of all variants are merged together based on the activity. The frequency of each activity occurrence is displayed next to the rectangle and represented in the height of the rectangle. The different variants are not necessarily horizontally aligned. For example, the purple C at the end with value 1 does not necessarily belong to the same variant that starts with a B followed by a A. A legend gives information on which color represents which activity. In the Figure below, we have three traces: 2 times the trace ABBCBCA and 1 time the trace BABCBCC. Several relations can be present in the visualization. For example, parallelism, loops or choice. Parallelism means that a certain number of activity occurrences are parallel if they can occur in any order. For example, the first two activity occurrences of both traces. Either A is followed by B or B is followed by A. Loops are consecutive activities that are repeated again. In the Figure you can see that in one variant B followed by C is repeated twice. Finally, a choice is, for example, picking a certain activity occurrence in one variant and another activity occurrence in another variant. In the example below you can see one example of this at the end, C is either followed by A or by C.
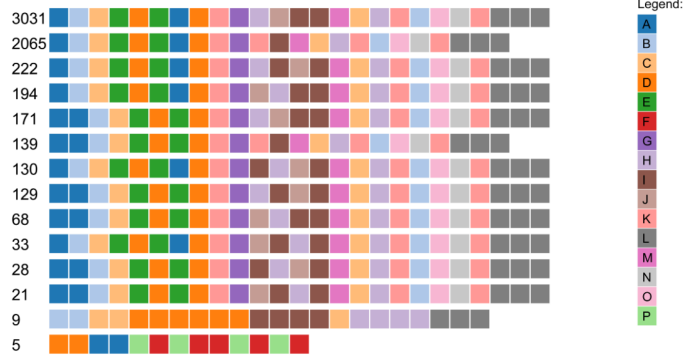
4. After the explanation the participant will see the visualization and can press next to start to fill in the subtasks (example of the timeline):

Subtask 1 (mouse is located in the dropdown menu to pick and activity occurrence):

Subtask 2:

Model Visualization

When you are ready press next to fill in the questions based on the visualization and start the interactions. Some questions can have multiple correct answers, we are only asking for one correct answer.
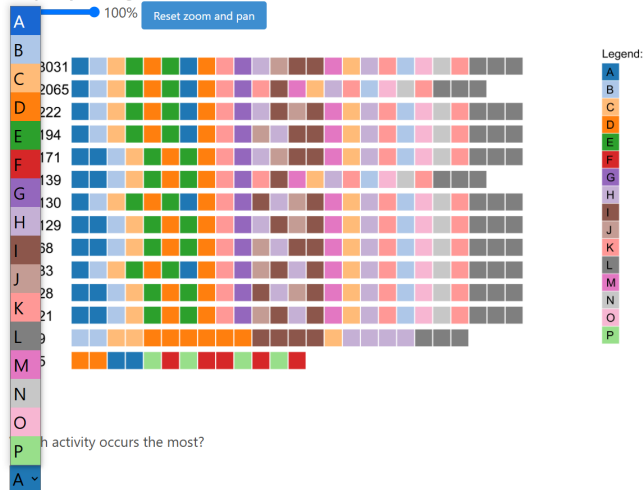
Frequency filtering:

100% [Reset zoom and pan]

Legend:
A B C D E F G H I J K L M N O P

Which two consecutive activities occur the most besides ■ followed by ■?

[A ⌄] [A ⌄]

[Next]

Subtask 3:

Model Visualization

When you are ready press next to fill in the questions based on the visualization and start the interactions. Some questions can have multiple correct answers, we are only asking for one correct answer.
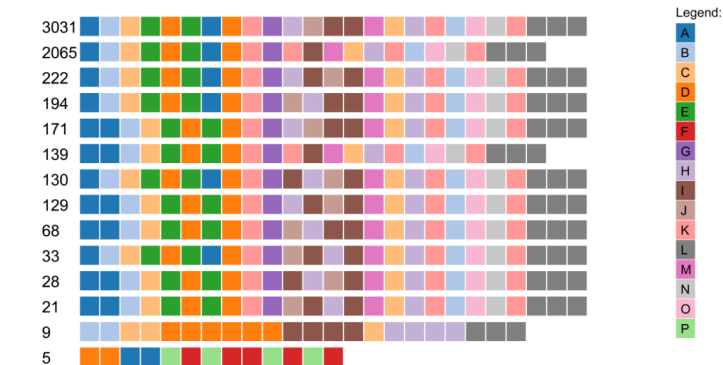
Frequency filtering:

100% [Reset zoom and pan]

Legend:
A B C D E F G H I J K L M N O P

Is parallelism (a set of activities at the same place in the trace can occur in any order) present?

○ Yes   ○ No

[Next]

Subtask 4:

Model Visualization

When you are ready press next to fill in the questions based on the visualization and start the interactions. Some questions can have multiple correct answers, we are only asking for one correct answer.

Frequency filtering:

100% Reset zoom and pan

Is there a loop of 3 consecutive activities, e.g. are three consecutive events repeated over time in one variant?
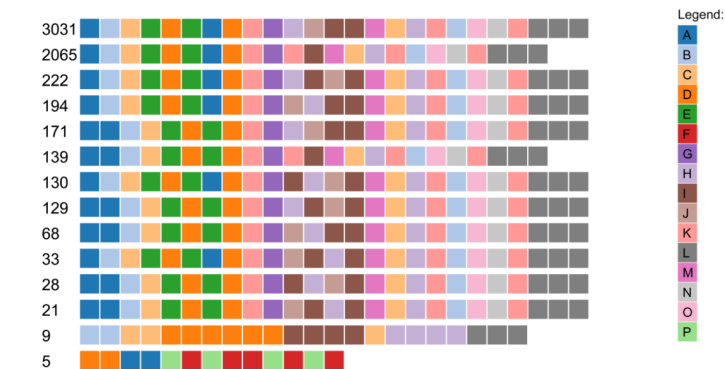
○ Yes ○ No

Next

Subtask 5:

Model Visualization

When you are ready press next to fill in the questions based on the visualization and start the interactions. Some questions can have multiple correct answers, we are only asking for one correct answer.

Frequency filtering:

100% Reset zoom and pan

If ▪ occurs does ▪ always occur before another ▪ occurs?
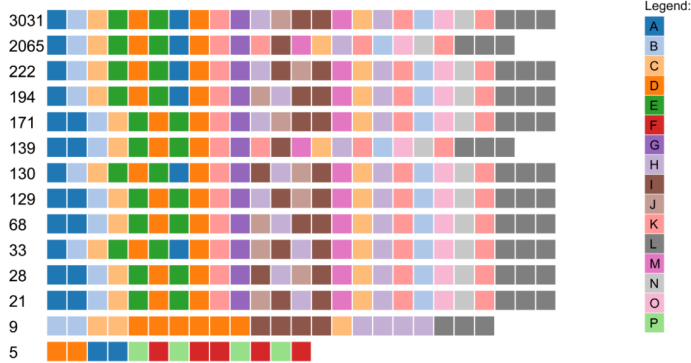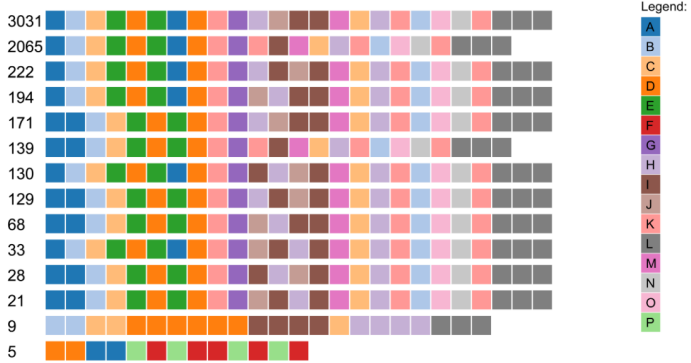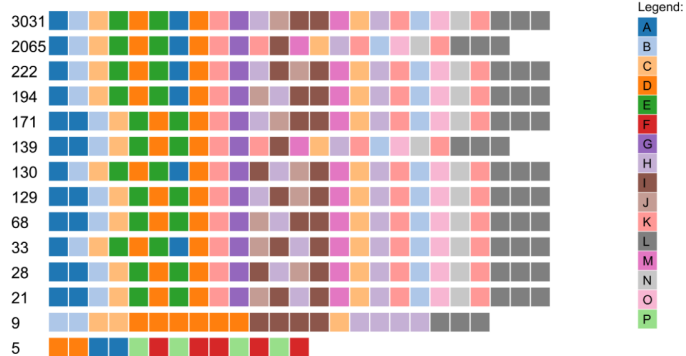
○ Yes ○ No

Next

Subtask 6:

Model Visualization

When you are ready press next to fill in the questions based on the visualization and start the interactions. Some questions can have multiple correct answers, we are only asking for one correct answer.

Frequency filtering:

100% [Reset zoom and pan]

What is the most frequent occurrence of an activity not in the happy flow (desired/most frequent path) is ▮-B-C- ▮-D-▮- ▮-D-K-▮ -(H,▮)- ▮-M-C-H -K-B-O-N- K-▮-▮-▮, where (H,▮) can occur in any order?

[A ⌄]

[Next]

5. Afterwards, the subjective questions are asked:

Please fill in the following questions.

It is difficult to find the parts of the visualization needed to answer the questions.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

The meaning of the nodes/rectangles and edges/arrows in the visualization are clear to me.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

I can easily read the activity flow (how the activities transition/progress from start to end) in the visualization.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

This visualization gives me a good overview of the activity flow (how the activities transition/progress from start to end).

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

[Next]

Please fill in the following questions.

I understand how the activity data is visualized.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

How the visualization displays the activity flow (how the activities transition/progress from start to end) is confusing.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

I can easily comprehend how the visualization displays the activity flow (how the activities transition/progress from start to end).

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

The visualization helped to increase my understanding of the activity flow (how the activities transition/progress from start to end).

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

[Next]

Please fill in the following questions.

The visualization of the activity flow (how the activities transition/progress from start to end) is informative.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

The visualization of the activity flow (how the activities transition/progress from start to end) is relevant for the questions I was given.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

The information provided by the visualization of the activity flow (how the activities transition/progress from start to end) is sufficient for me to answer the questions.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

The visualization of the activity flow (how the activities transition/progress from start to end) has insufficient detail to increase my knowledge about the activity flow.

○	○	○	○	○	○	○
Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly Agree

[Next]

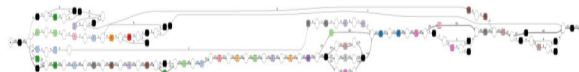6. Step 3-5 are repeated 4 times in total.
7. At the end the participant is asked to rank the visualizations

Please rank the previous visualizations
Where 1 is the best and 4 is the worst.

 -- select an option -- ⌄

 -- select an option -- ⌄

 -- select an option -- ⌄

 -- select an option -- ⌄

[Next]

8. A thank you screen appears

Thank you for participating! You can close the window.