



**BERLIN  
BUZZWORDS  
2019** JUNE 16-18

# Dealing with pain points of recommenders in the real world

Sophie Watson

@sophwats

sophie@redhat.com



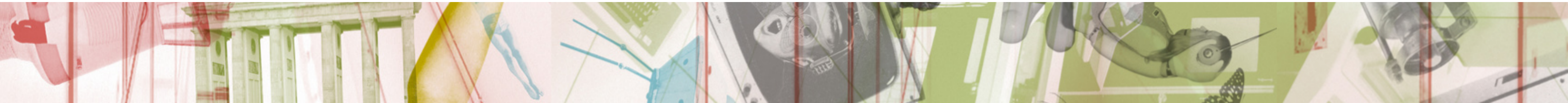
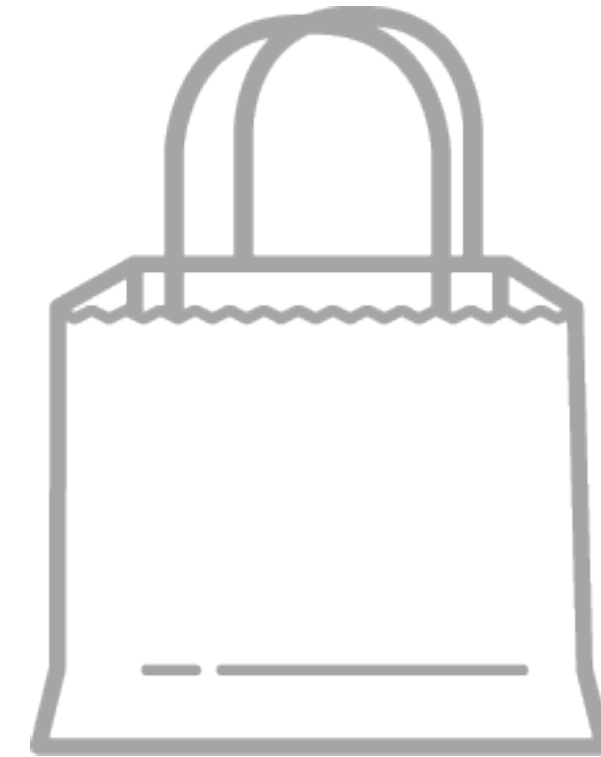
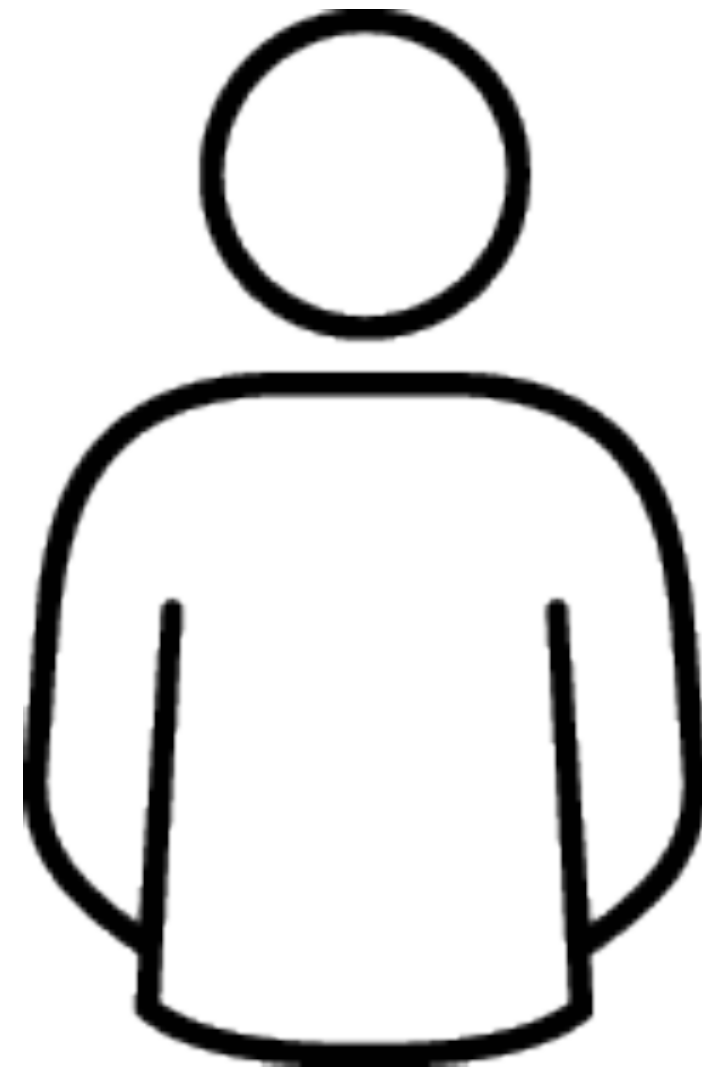
**BERLIN  
BUZZWORDS  
2019** JUNE 16-18

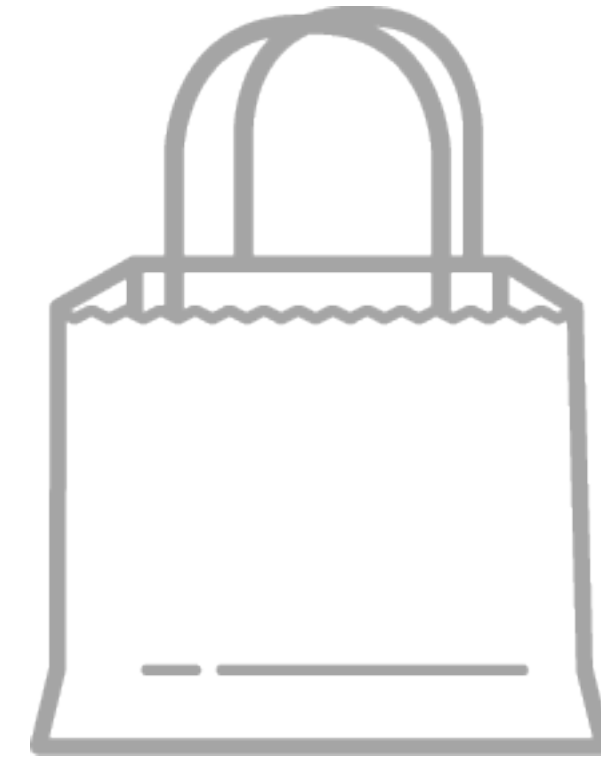
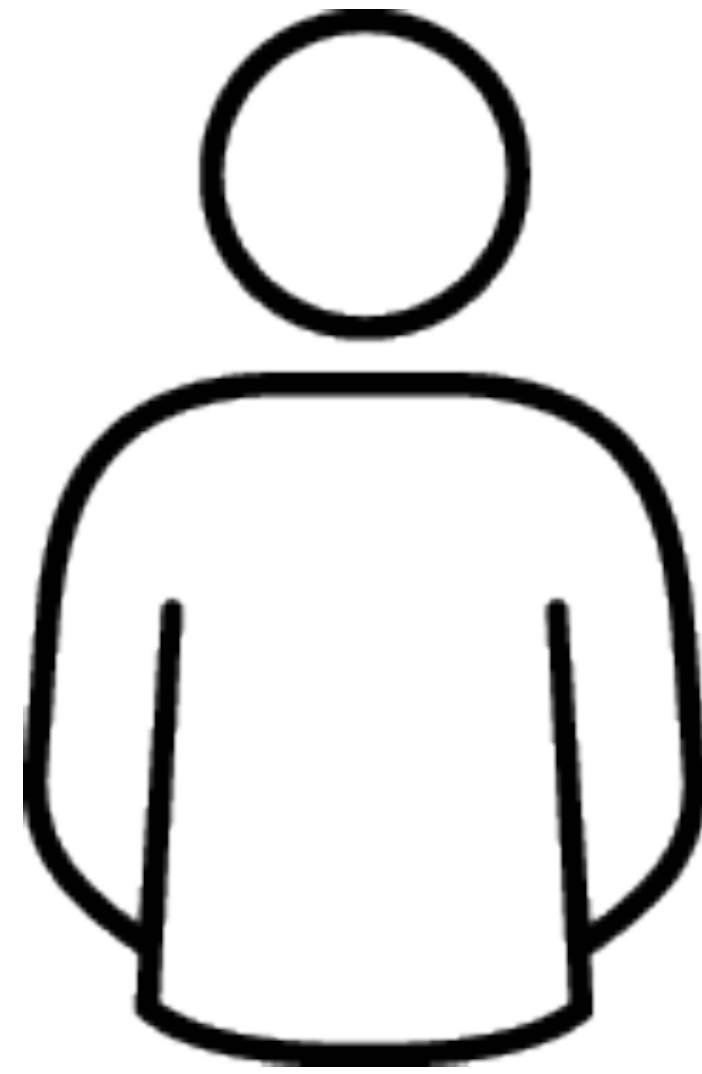
# Dealing with pain points of recommenders in the real world

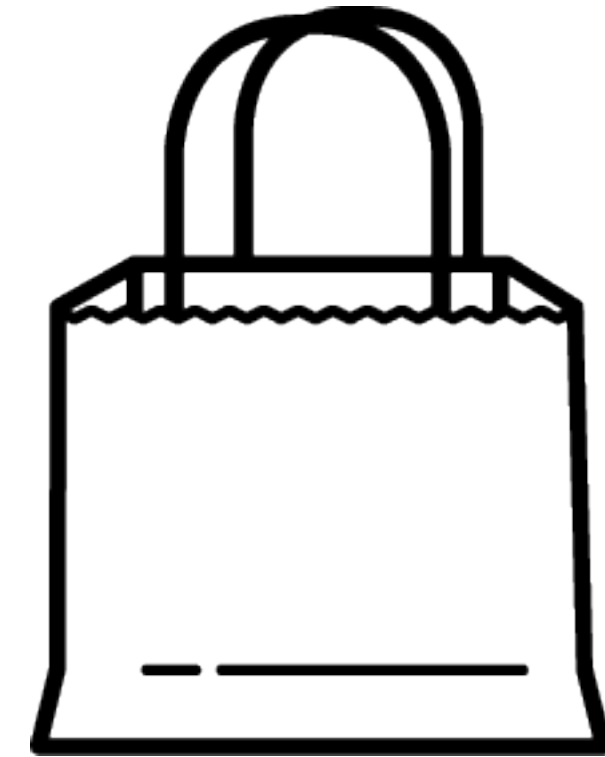
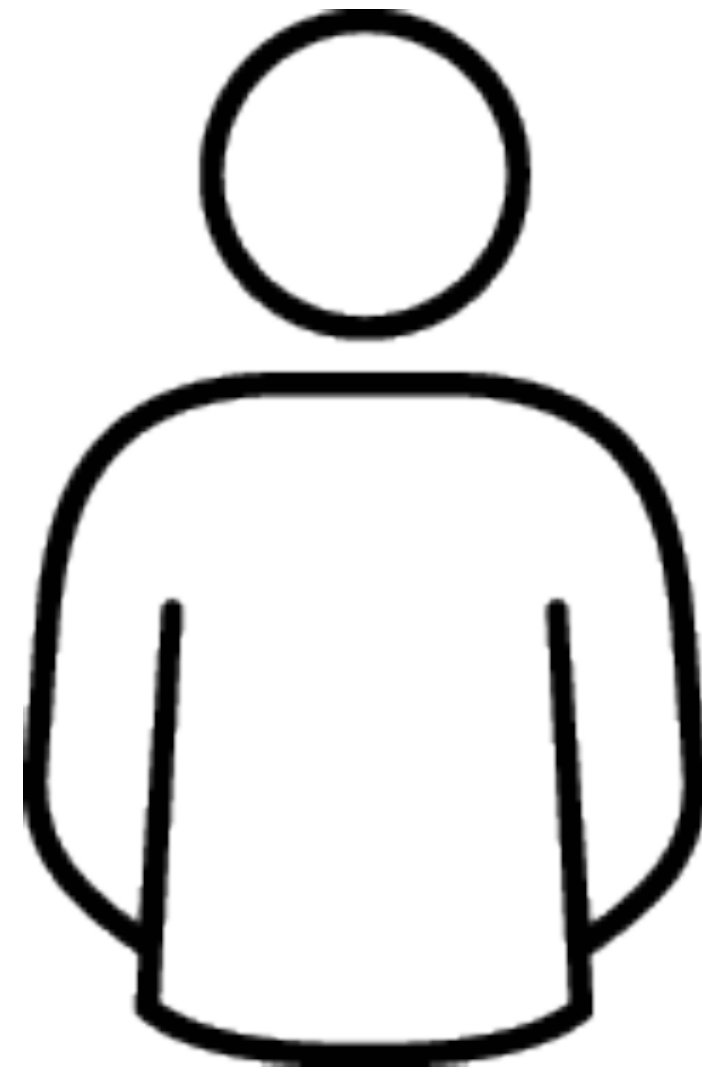
Sophie Watson

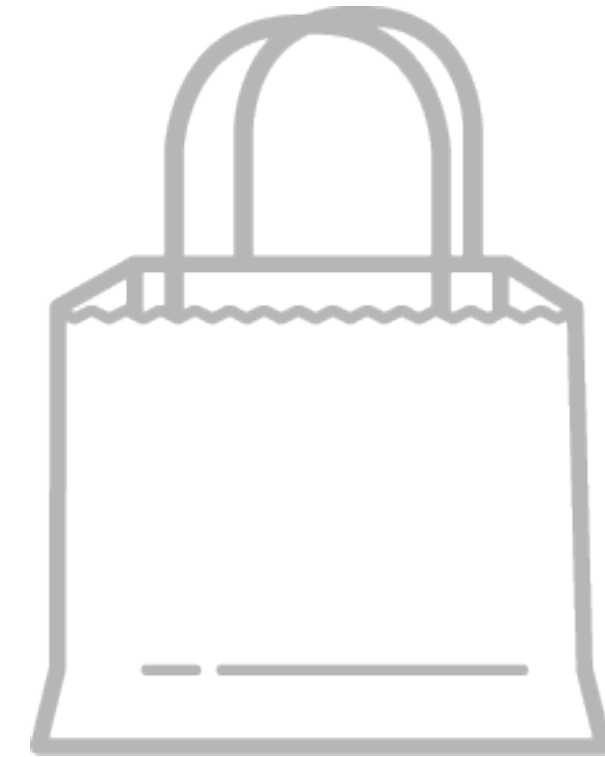
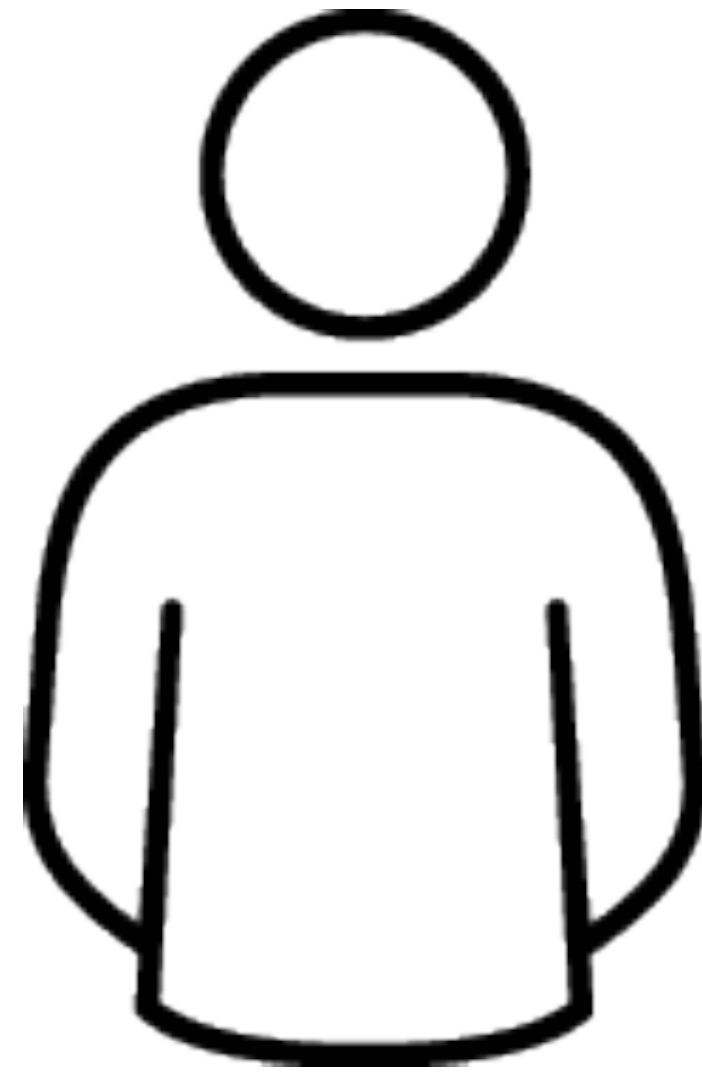
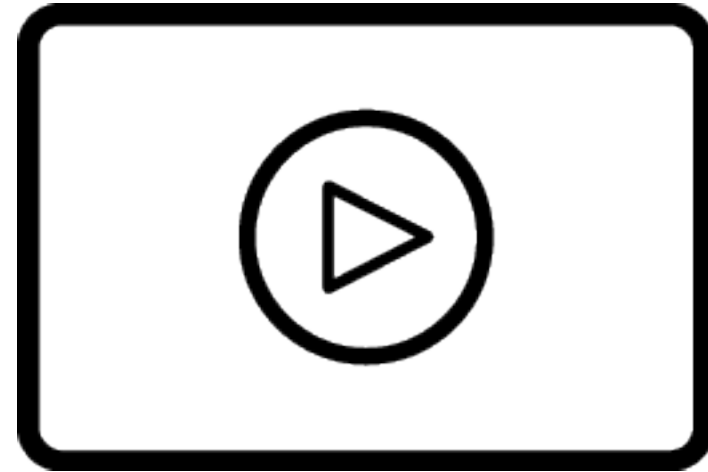
@sophwats

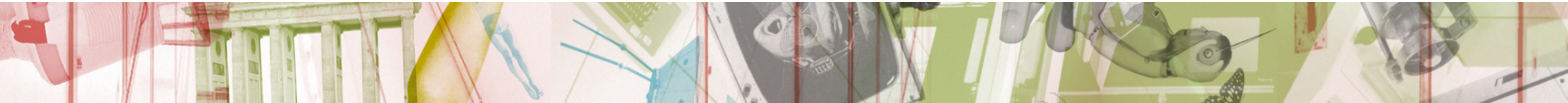
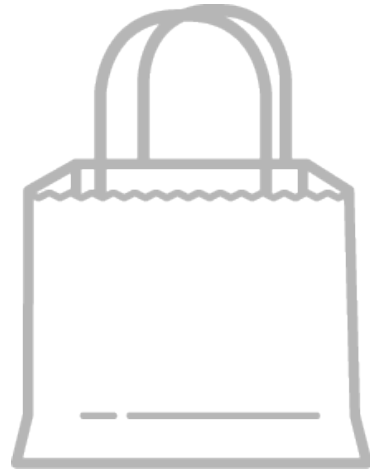
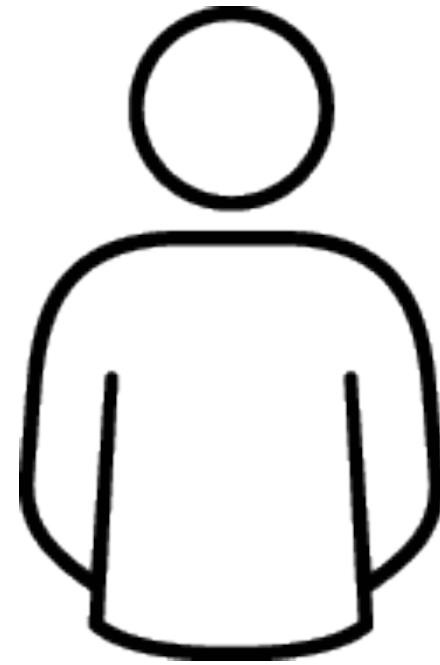
sophie@redhat.com

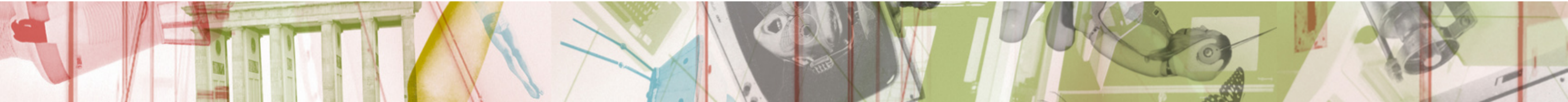
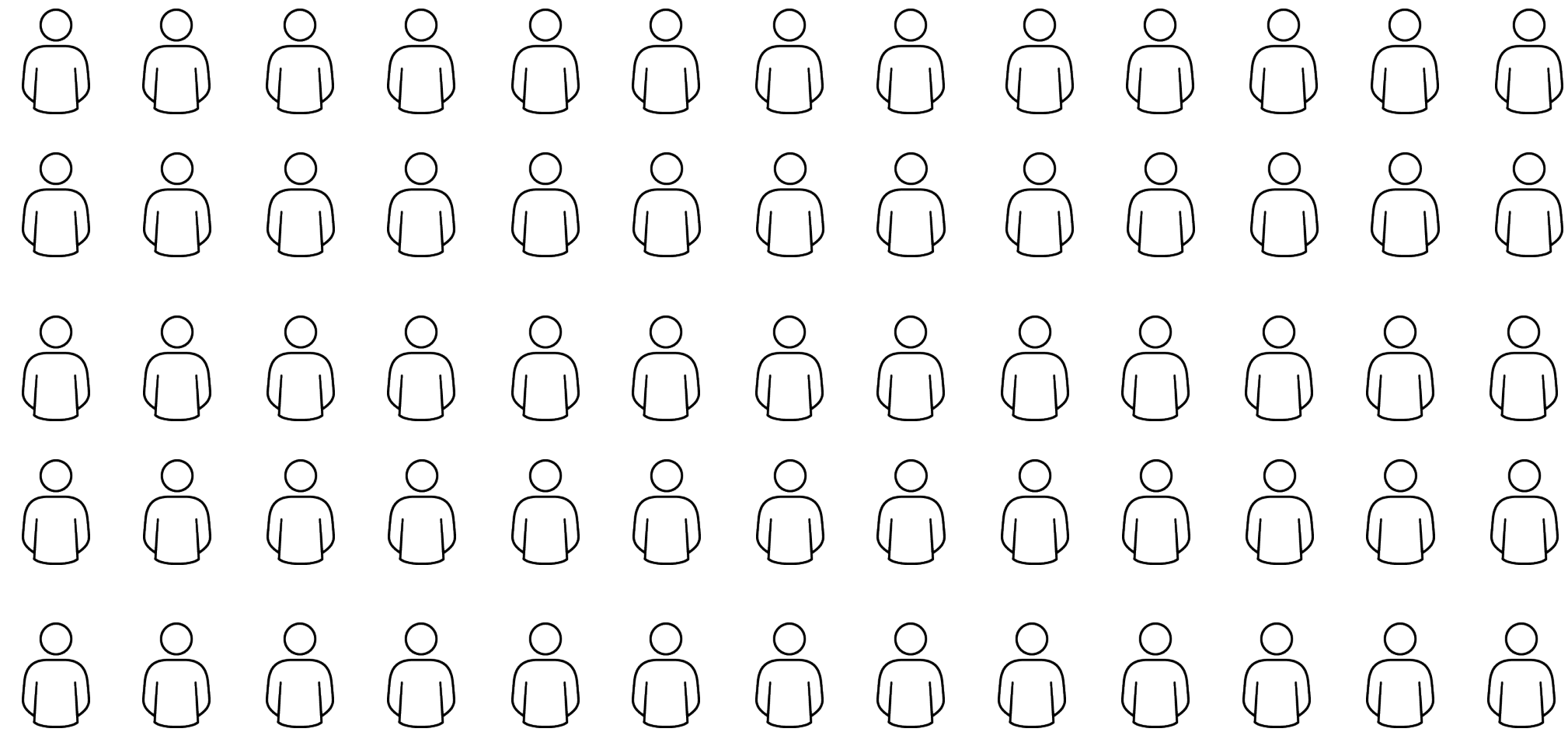
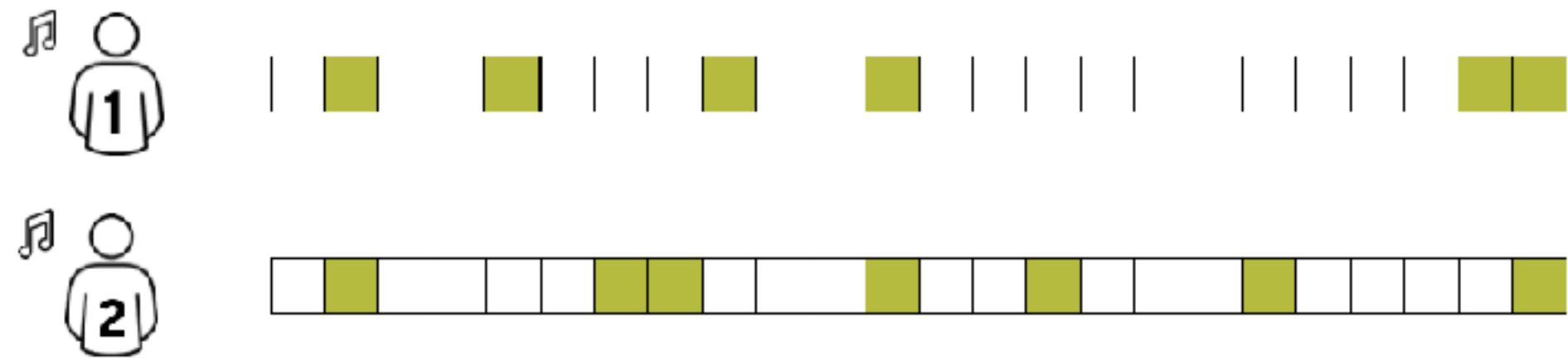
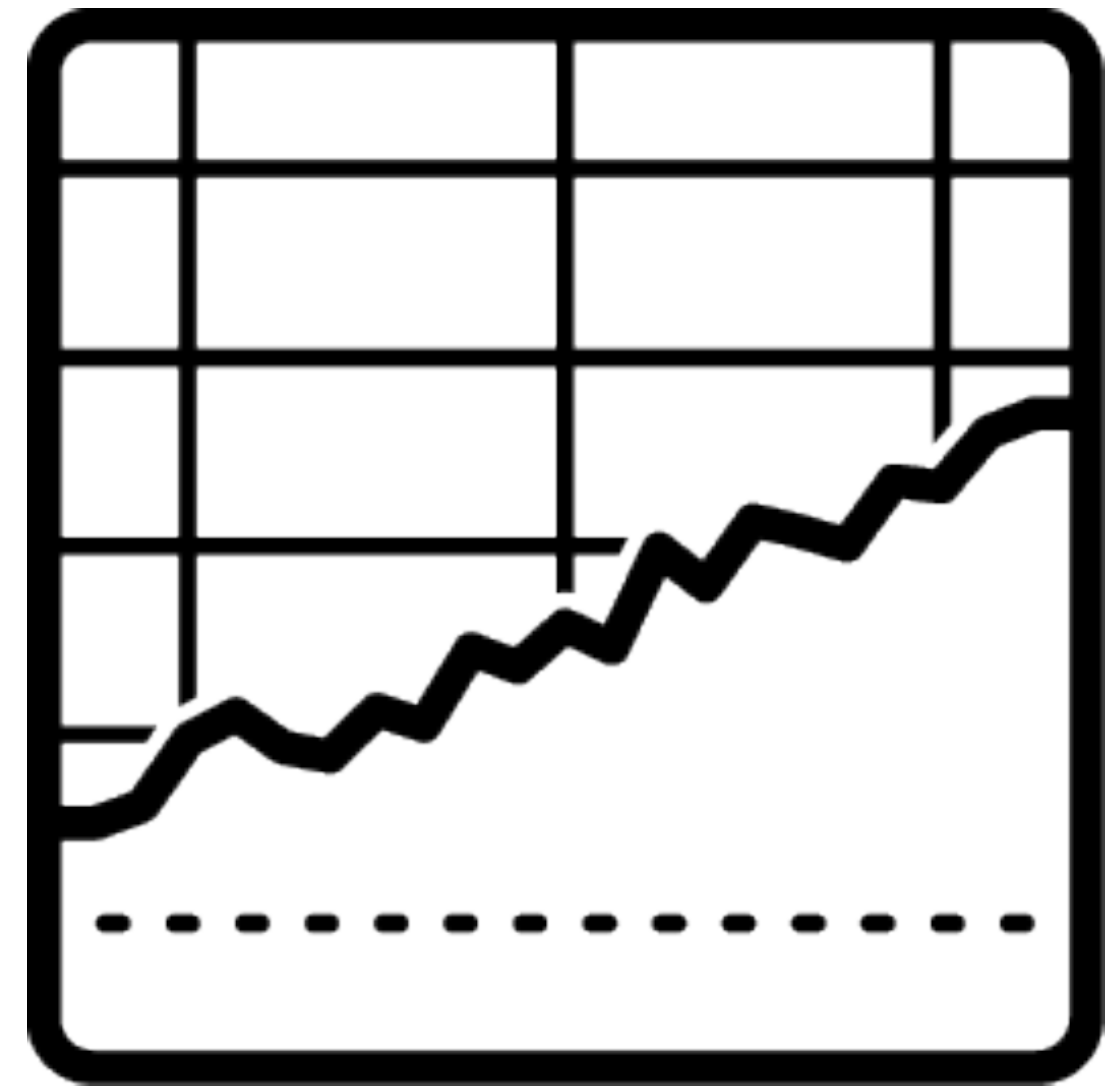
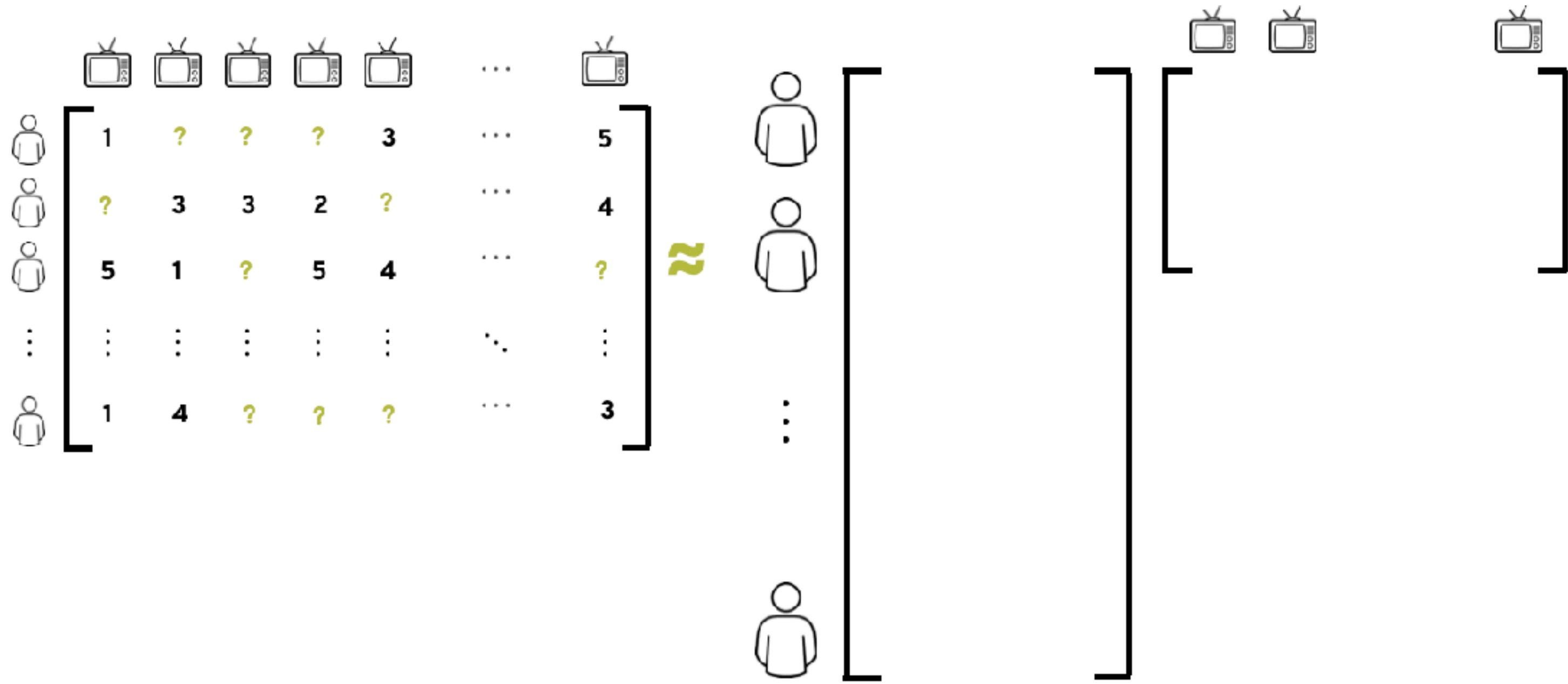






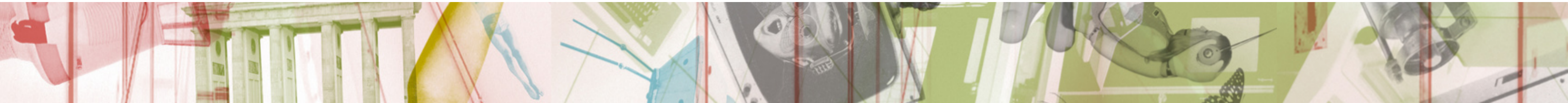
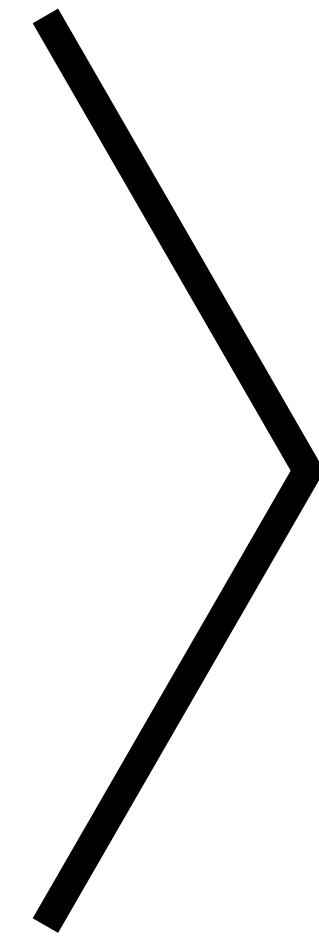
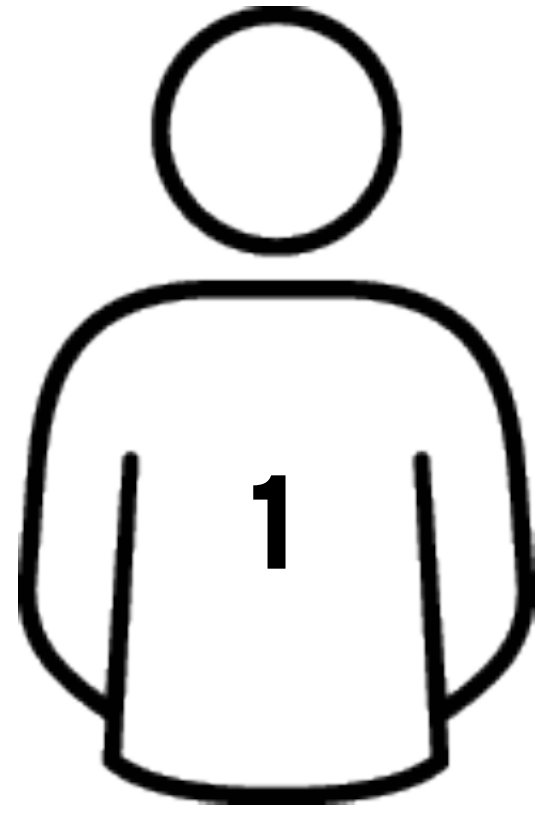
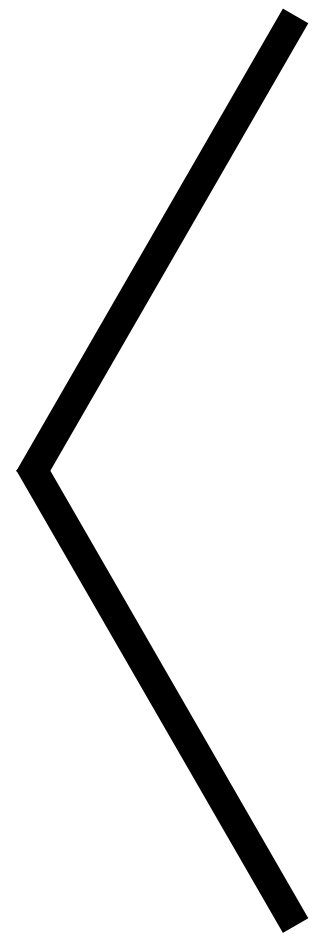


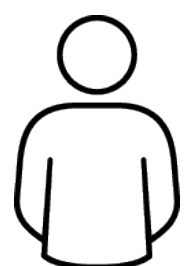
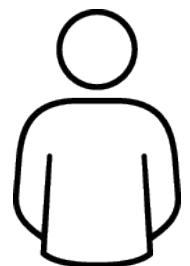
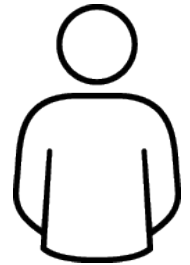
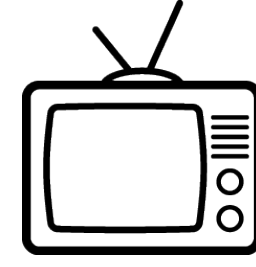
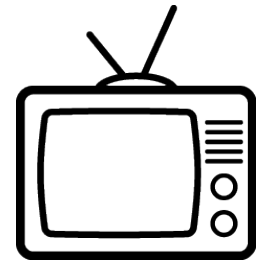




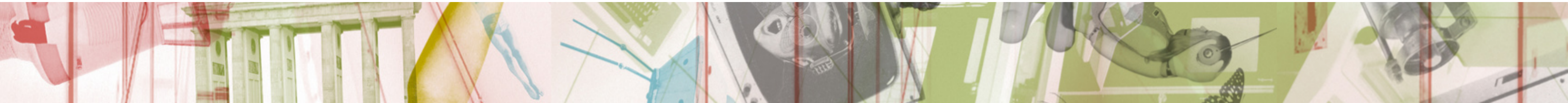
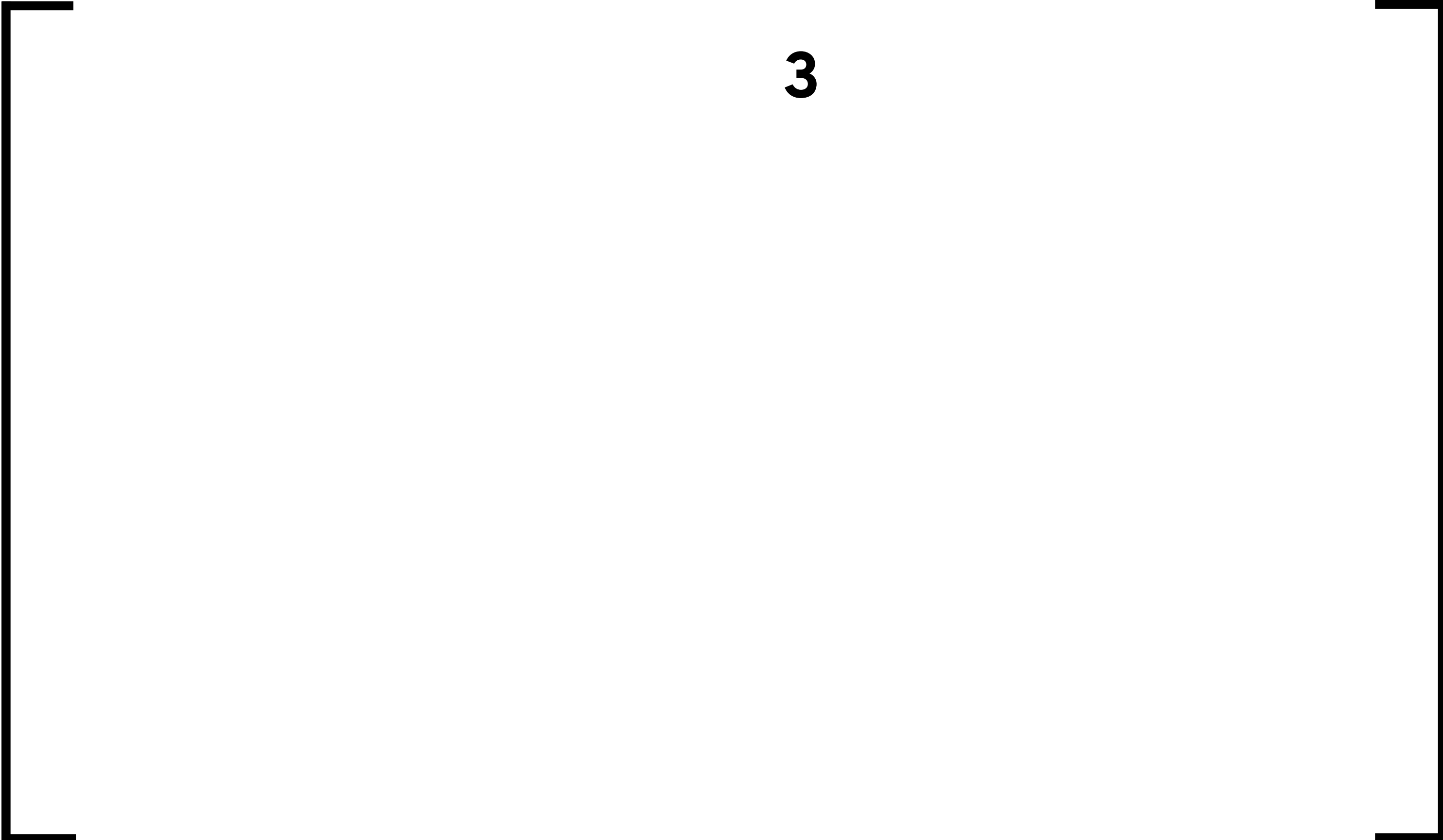


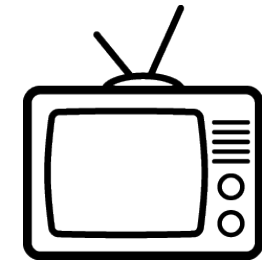
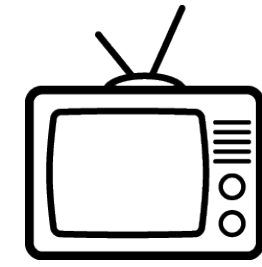
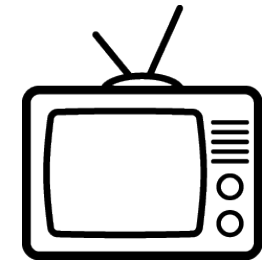
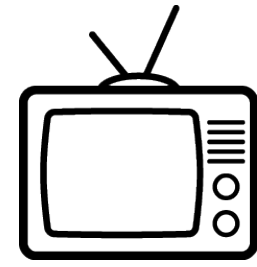
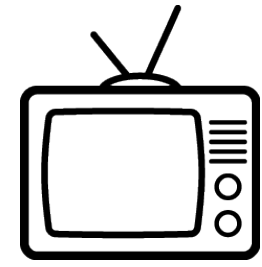
# Alternating Least Squares



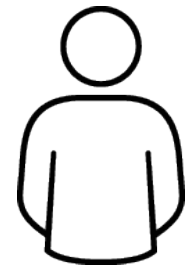
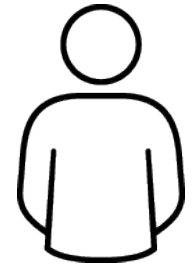
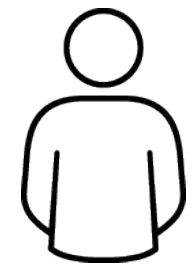
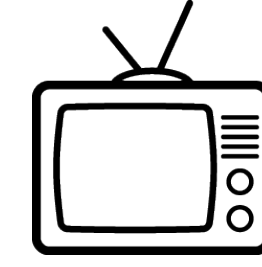


3

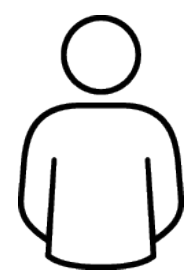




...



⋮



1

?

?

?

3

...

5

?

3

3

2

?

...

4

5

1

?

5

4

...

?

⋮

⋮

⋮

⋮

⋮

⋮

⋮

1

4

?

?

?

...

3



...



⋮



1

?

?

?

3

...

5

?

3

3

2

?

...

4

5

1

?

5

4

...

?

⋮

⋮

⋮

⋮

⋮

⋮

⋮

1

4

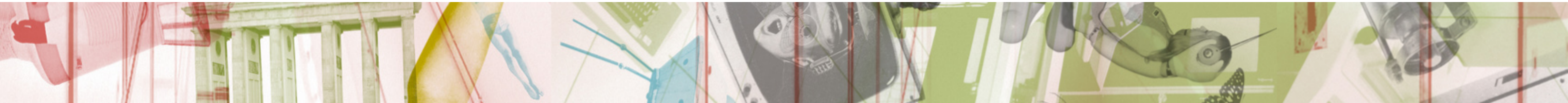
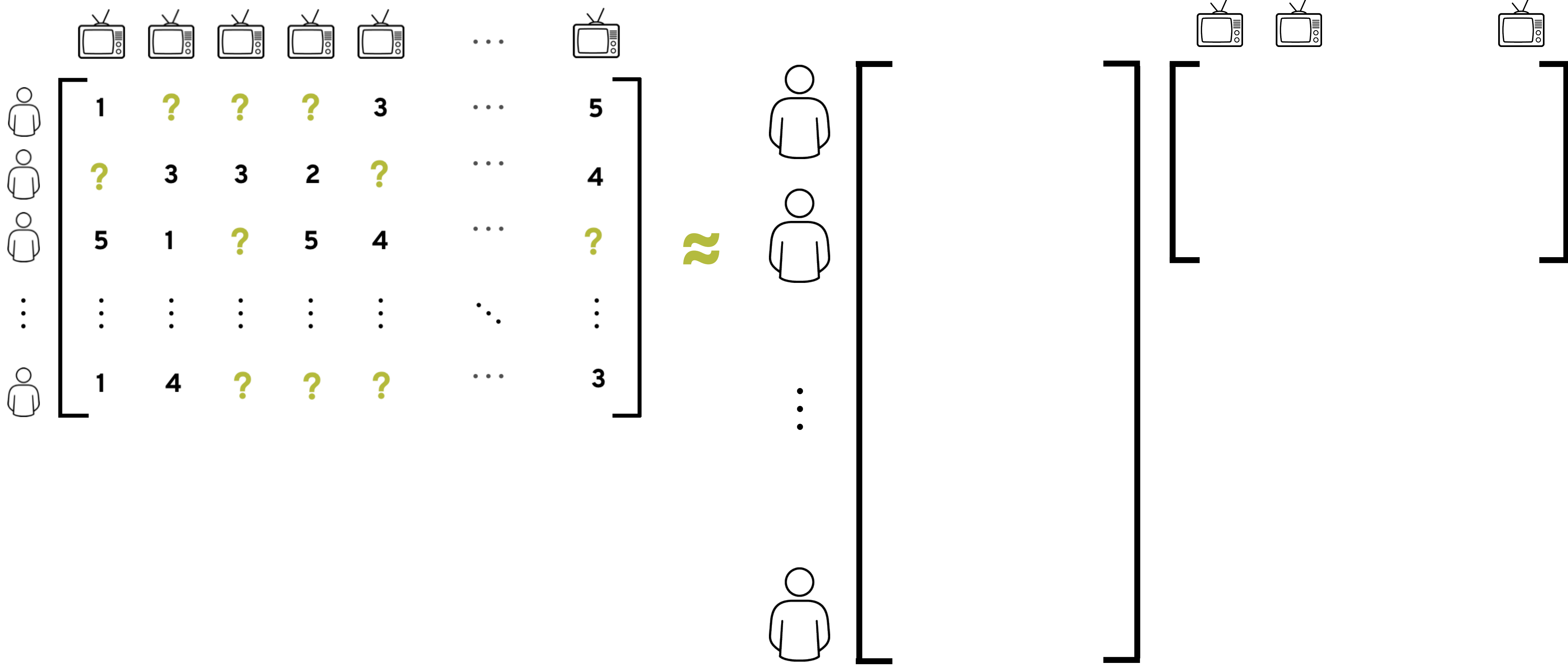
?

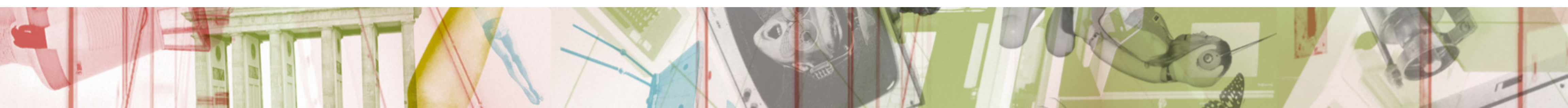
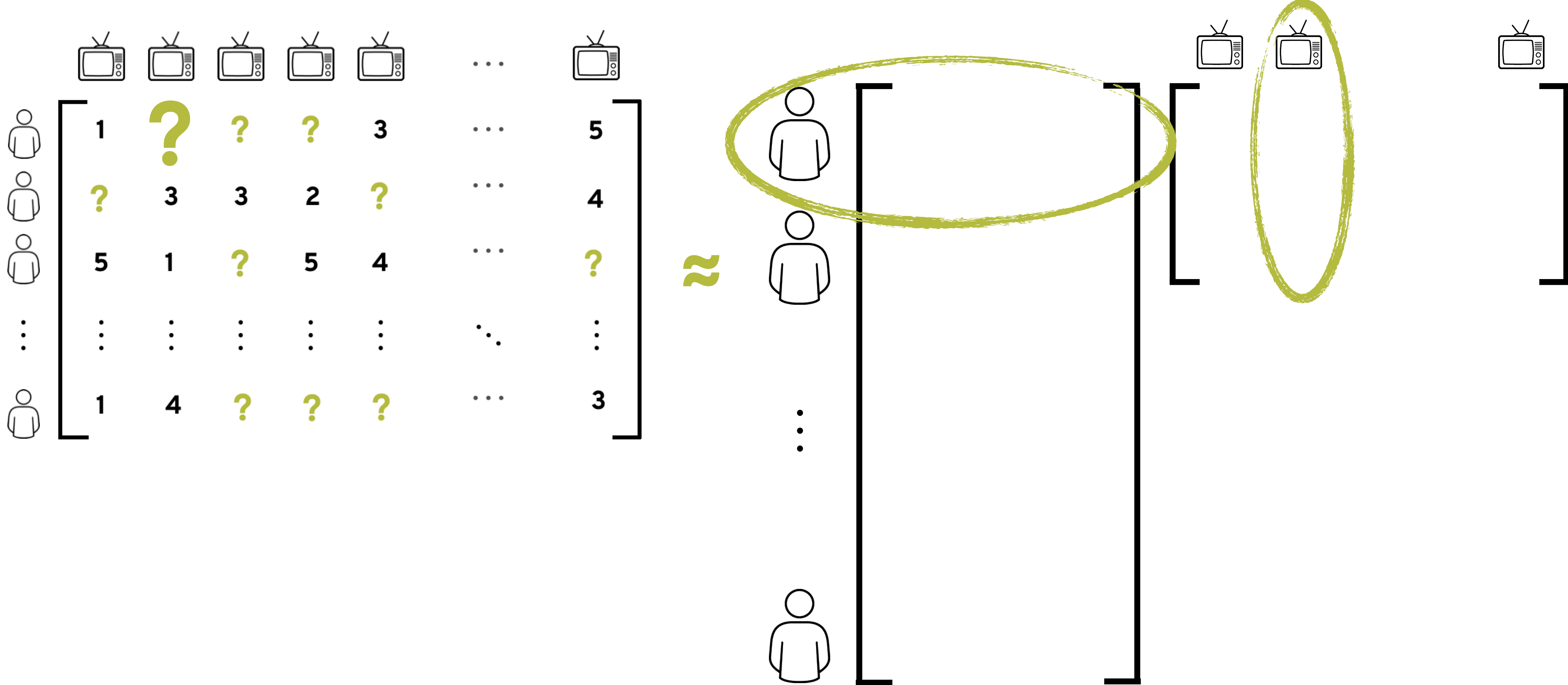
?

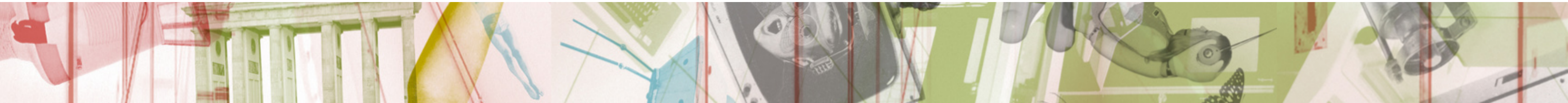
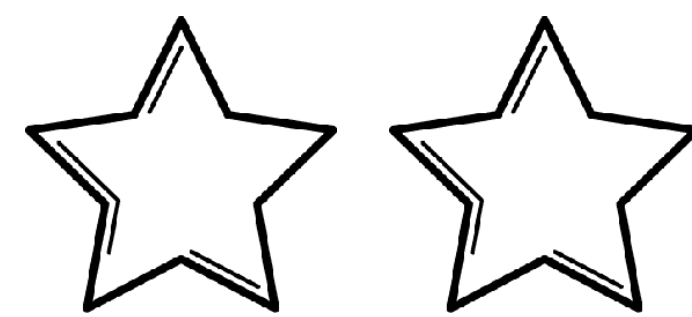
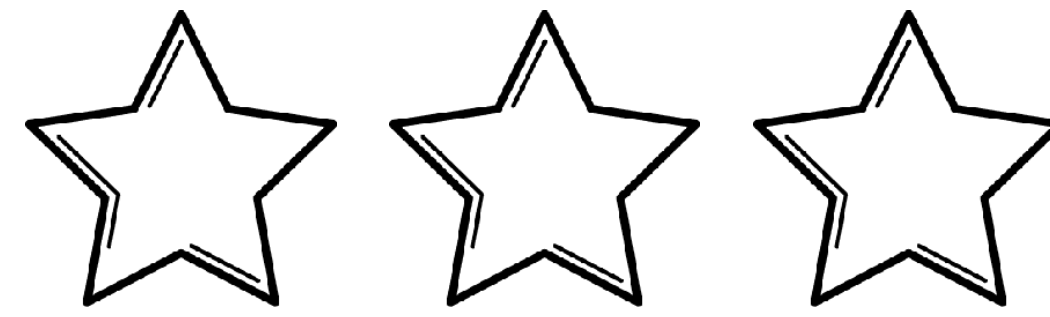
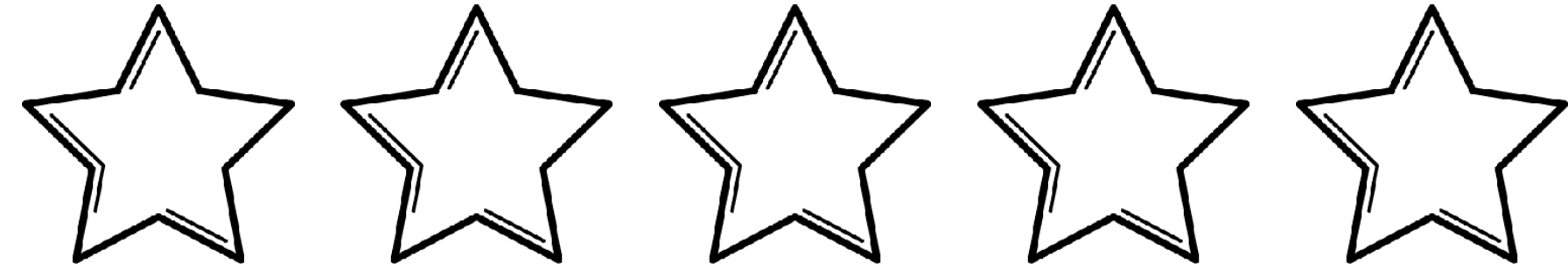
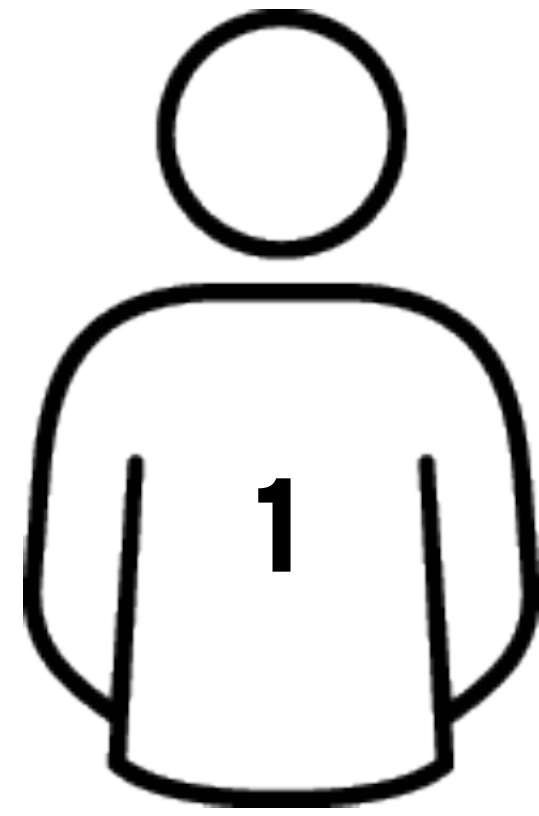
?

...

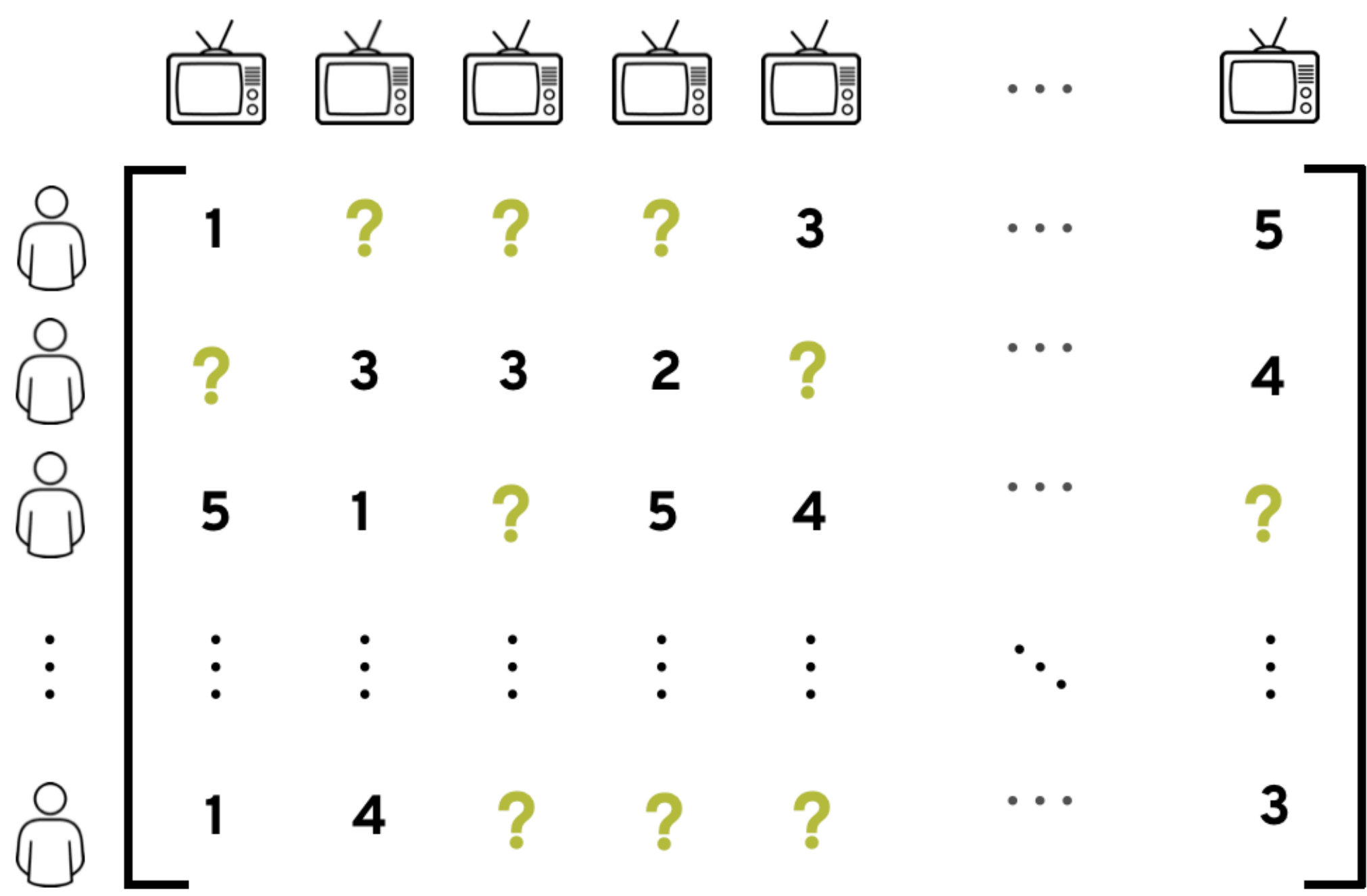
3



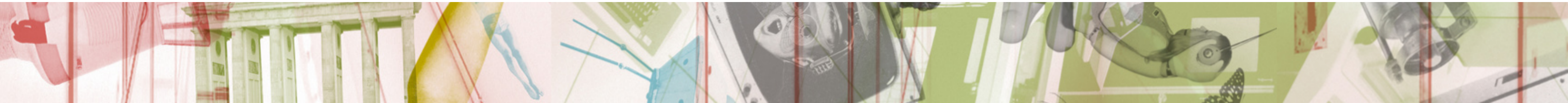
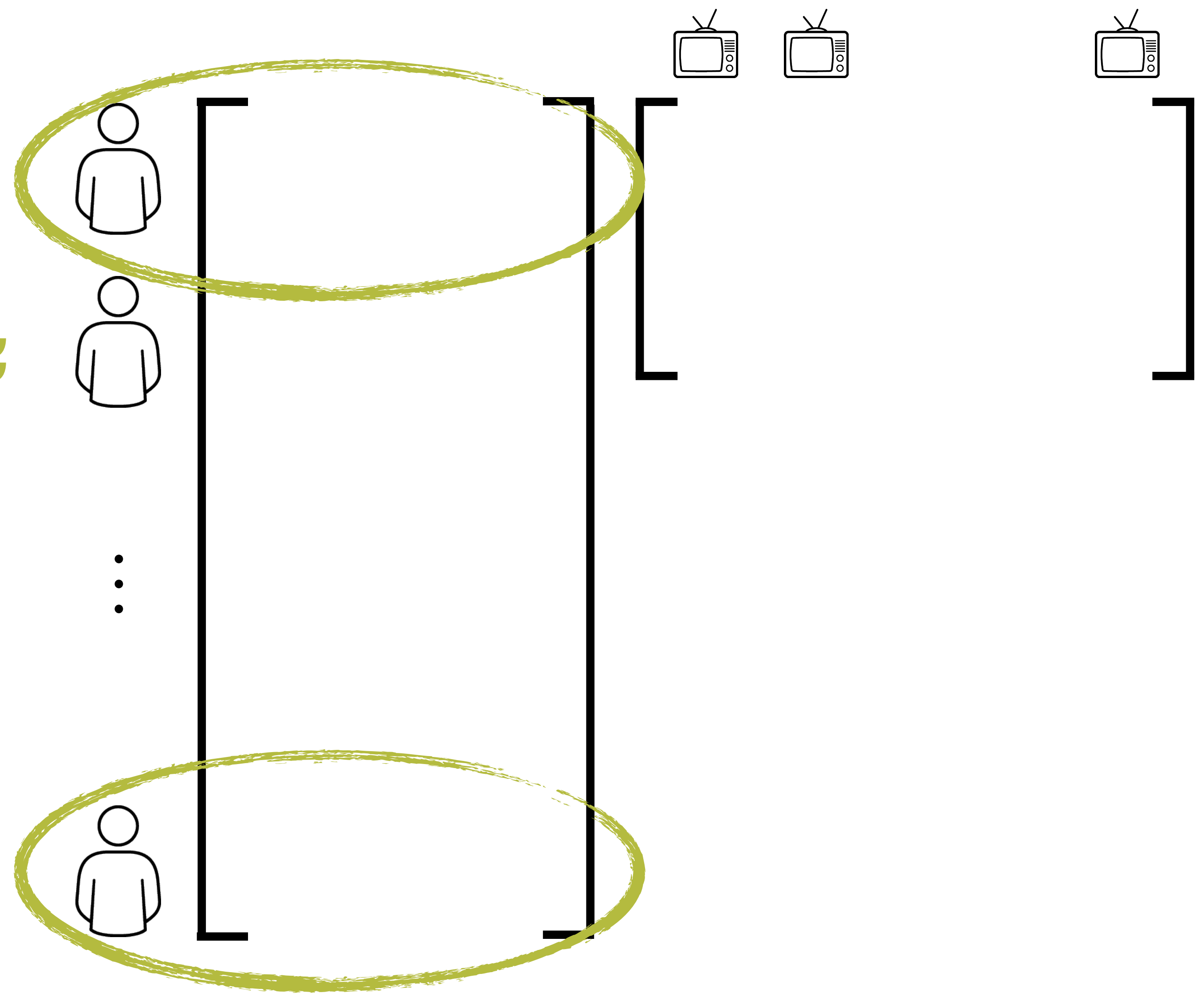


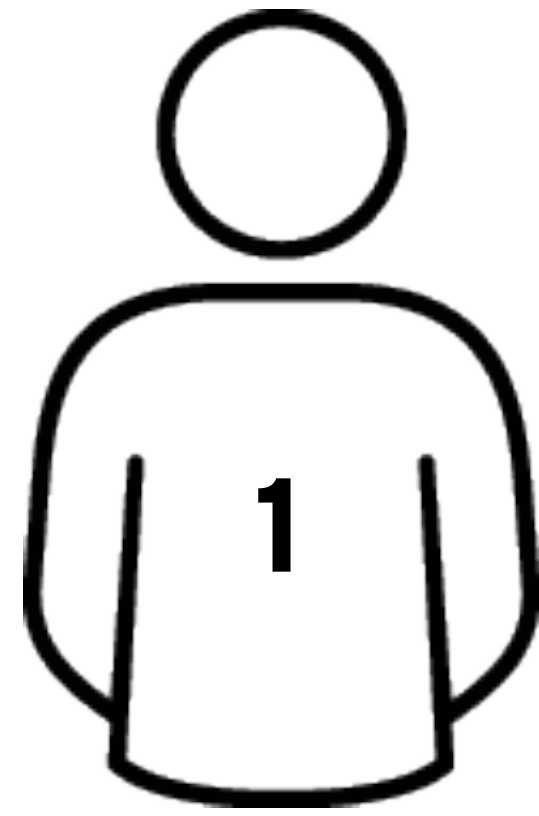






≈





**A**

**1 play**



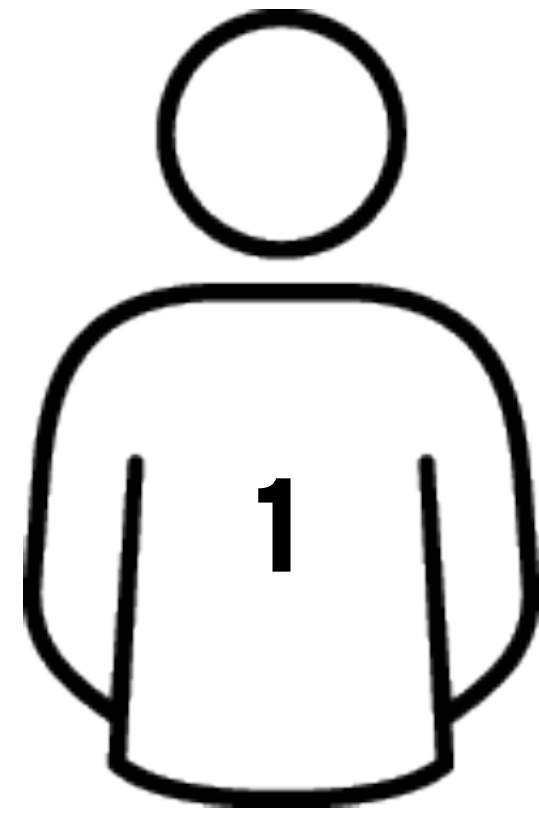
**B**

**10 plays**



**C**

**100 plays**



**A**

**1 play**

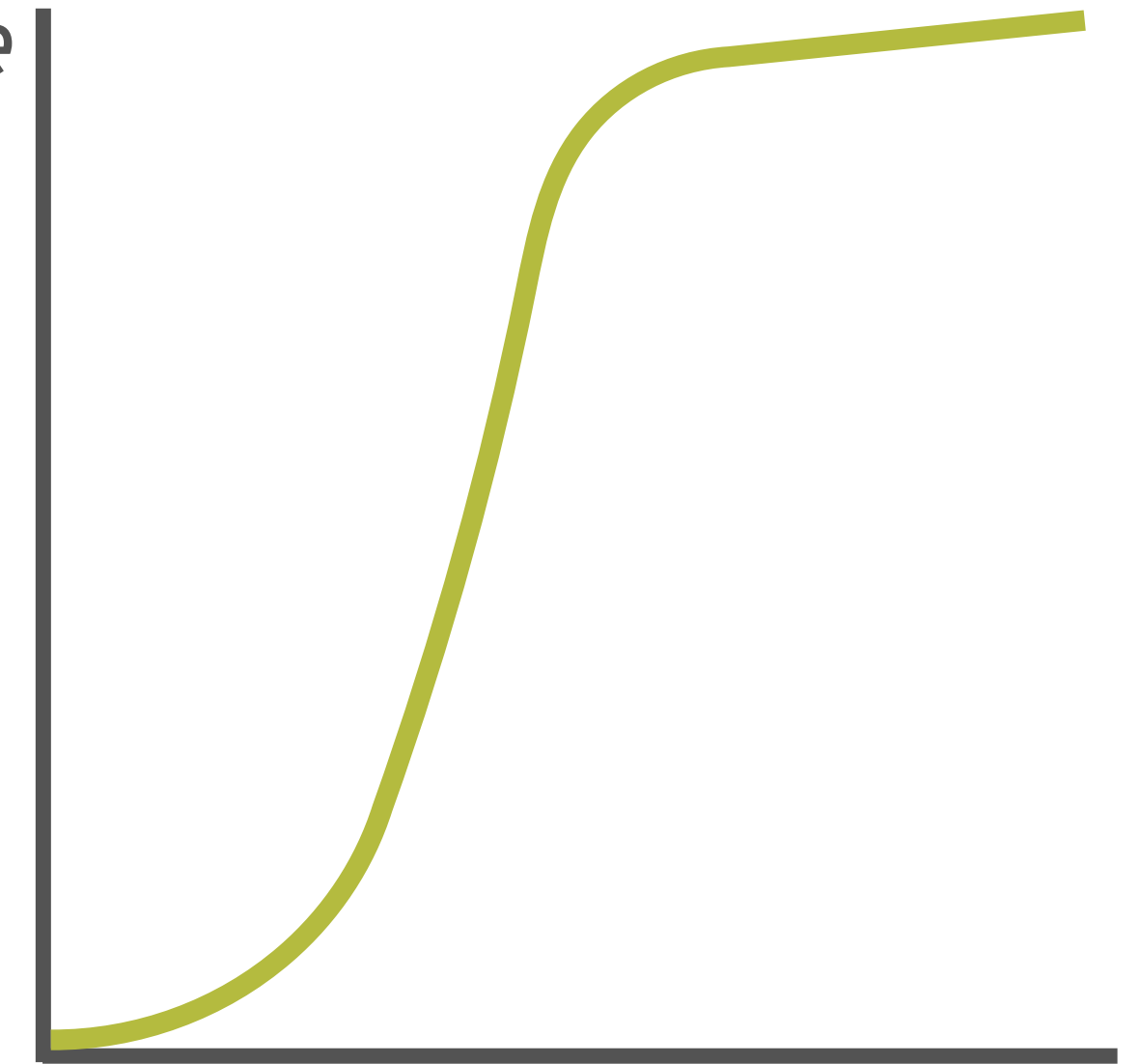
**B**

**10 plays**

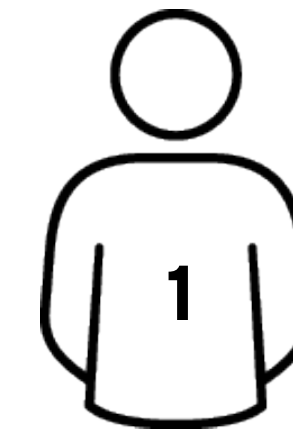
**C**




**100 plays**

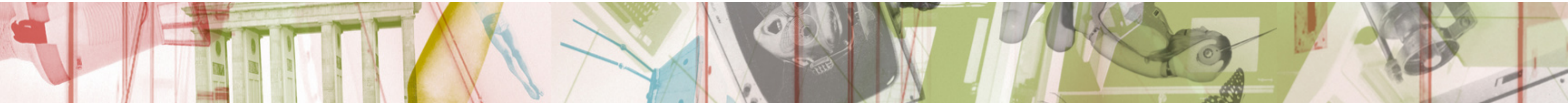
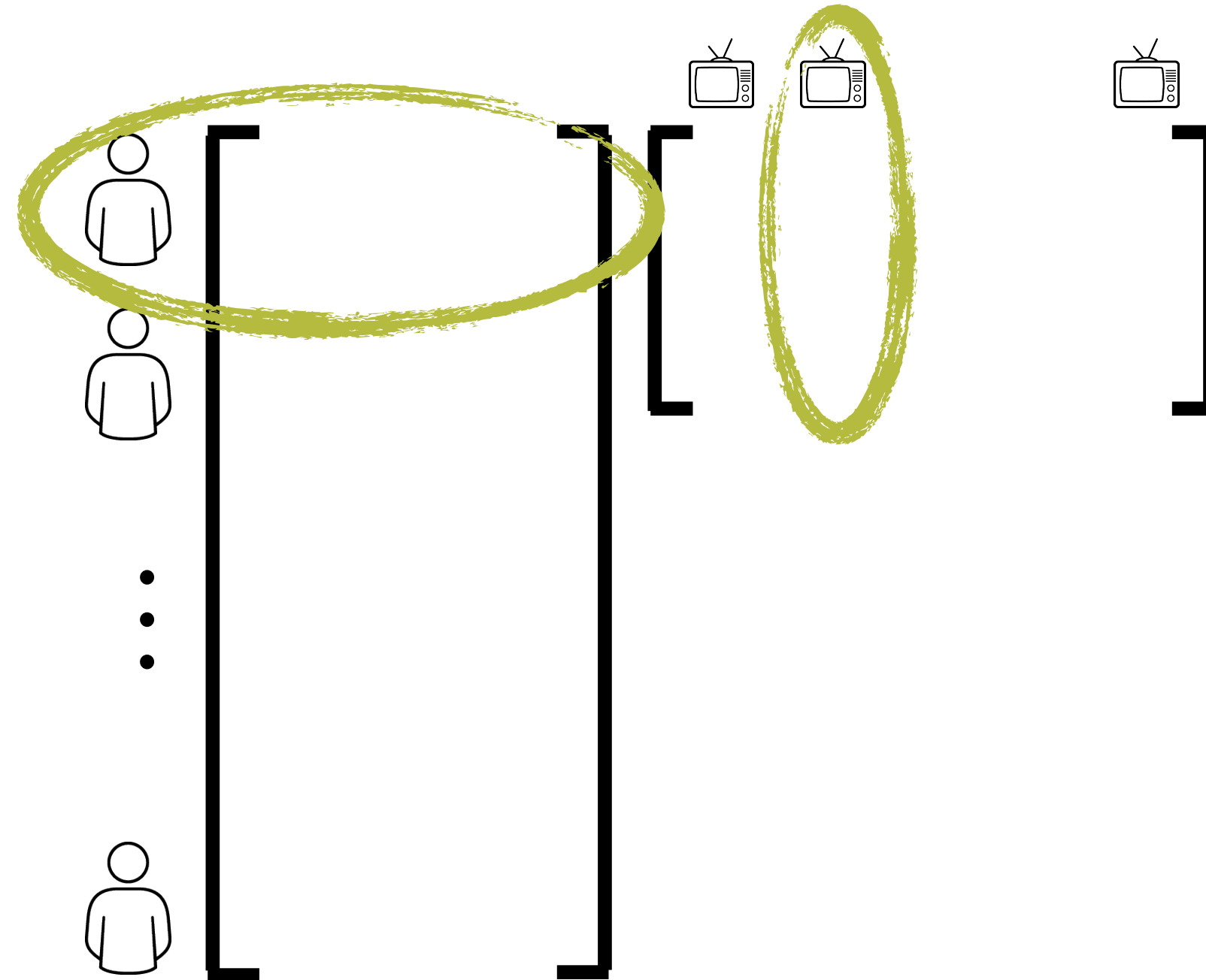
*Confidence*



*Plays*

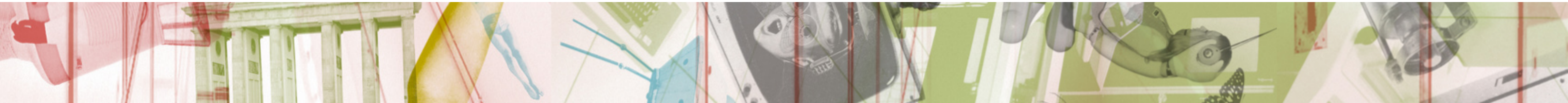
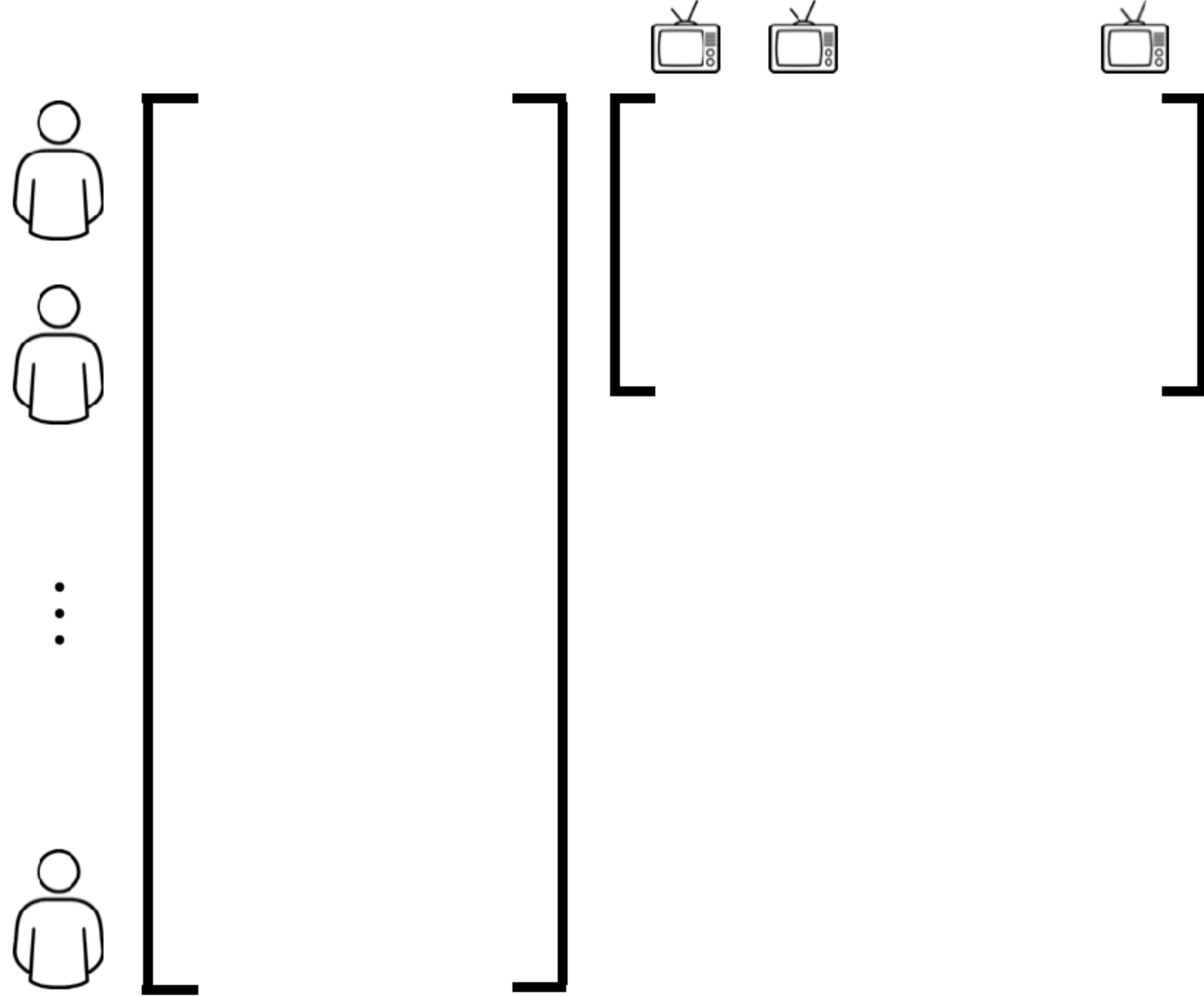


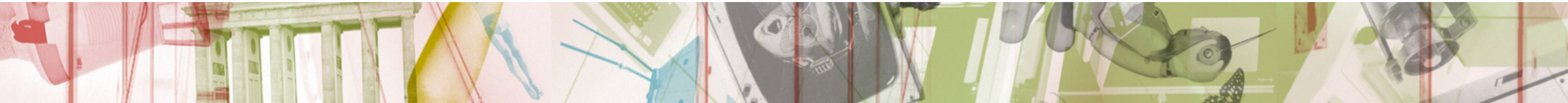
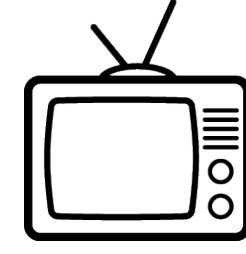
-  A      1 play
-  B      10 plays
-  C      100 plays



# Pain Points of ALS







?



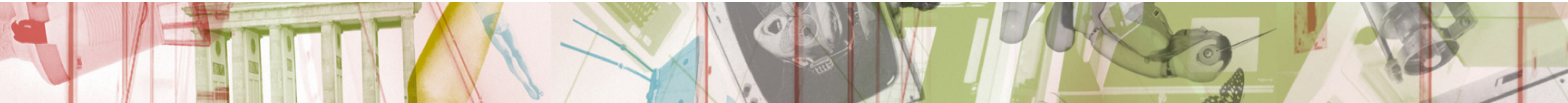
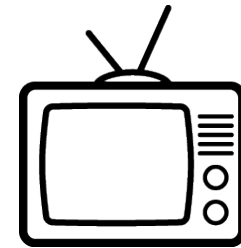
?



?

⋮

?





?



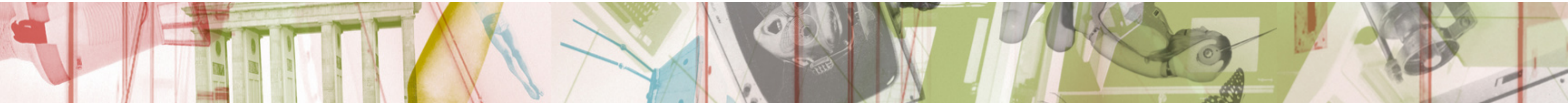
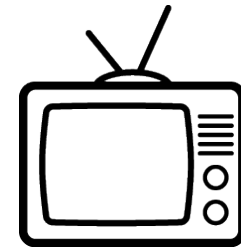
?

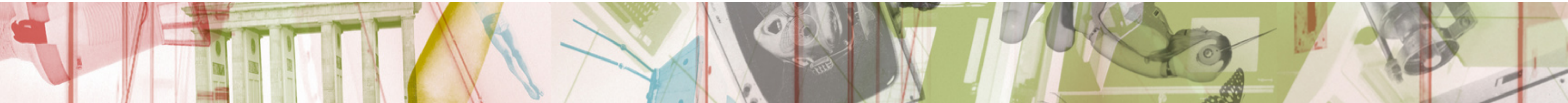
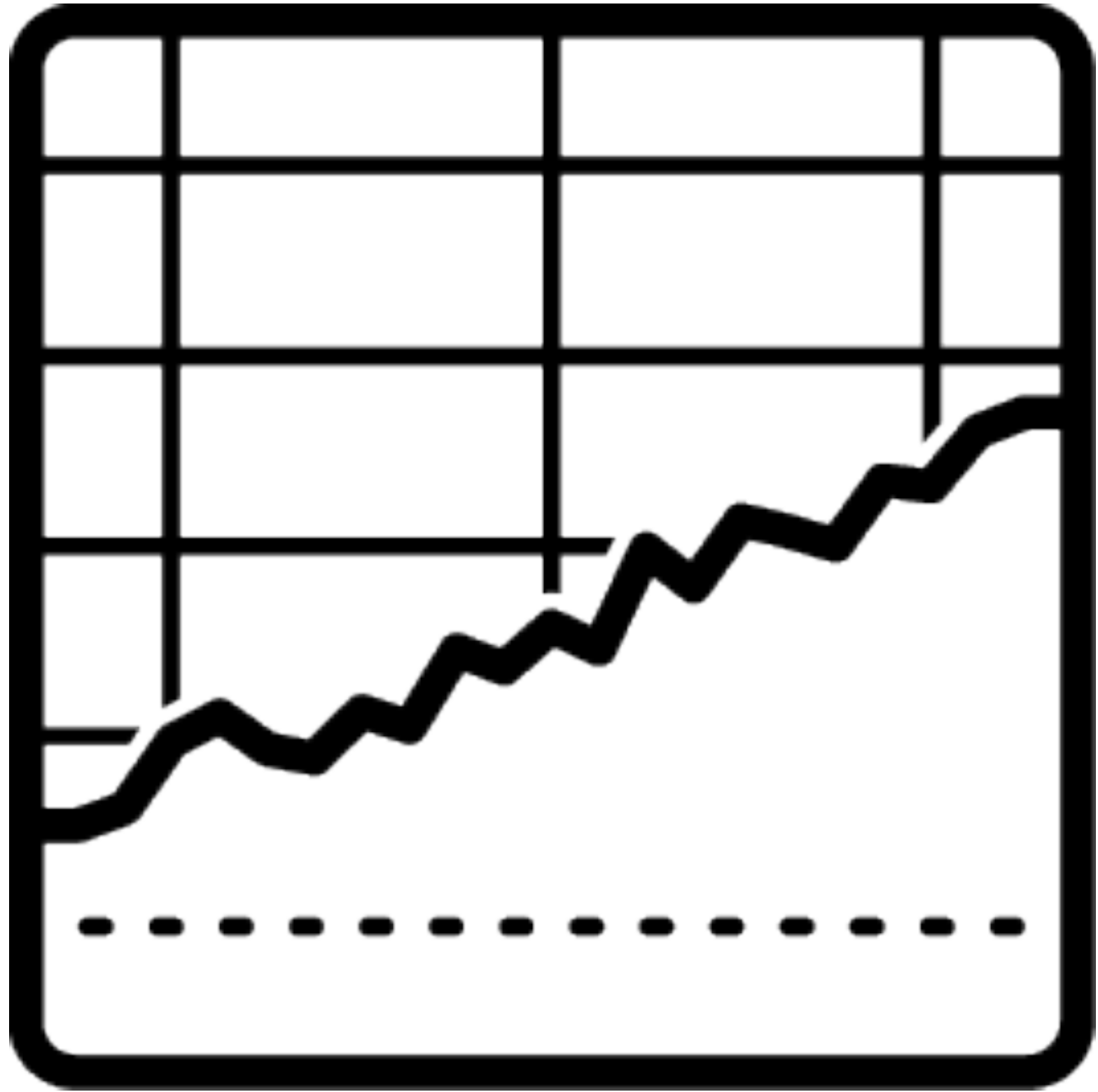


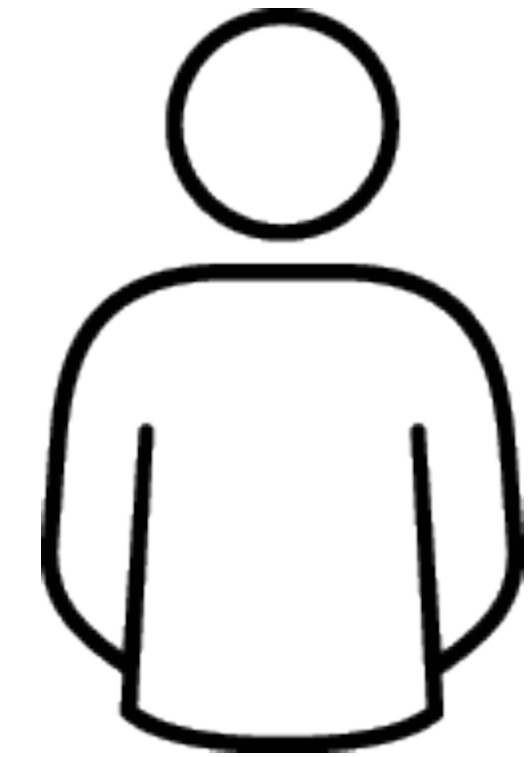
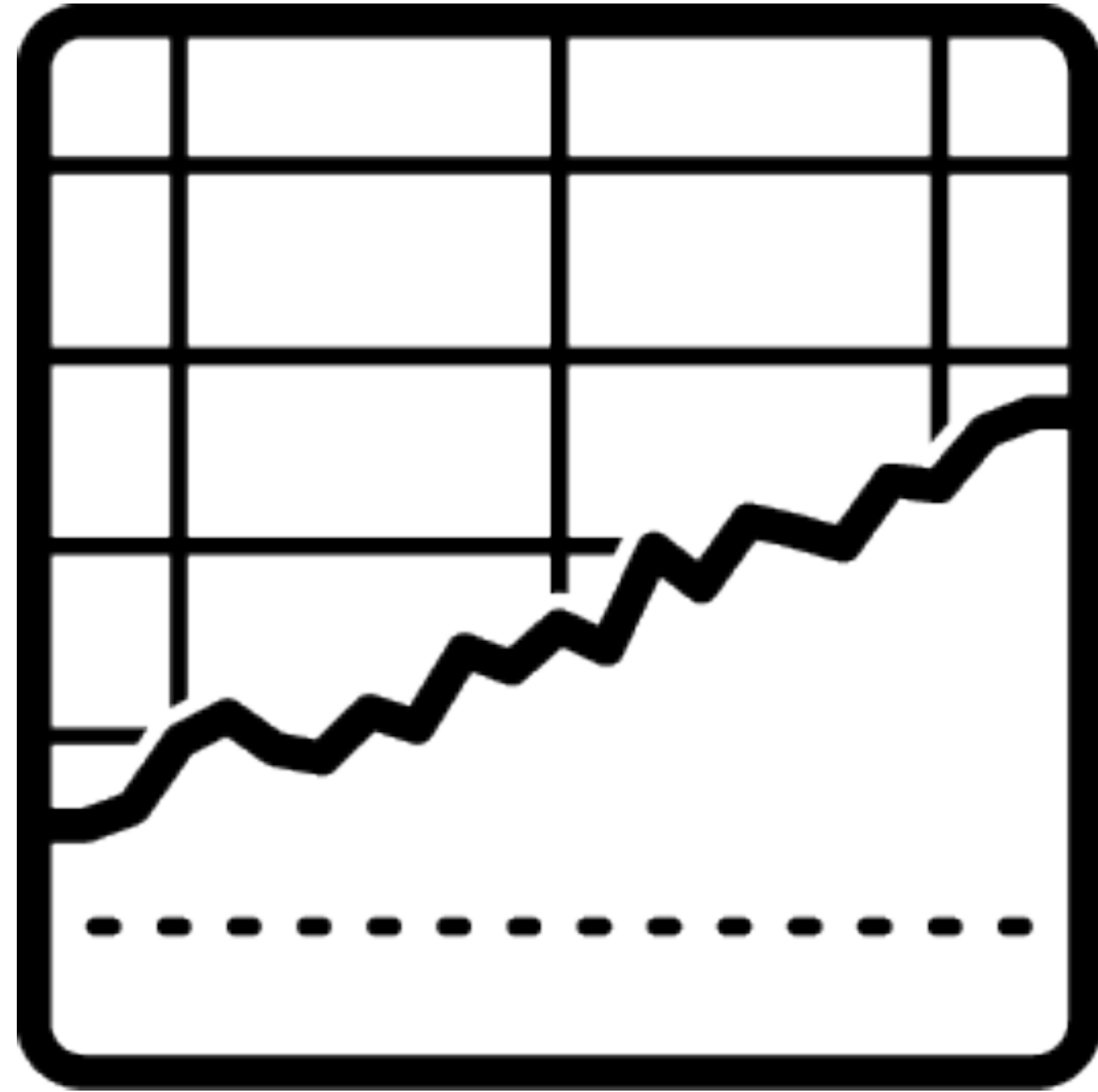
?

⋮

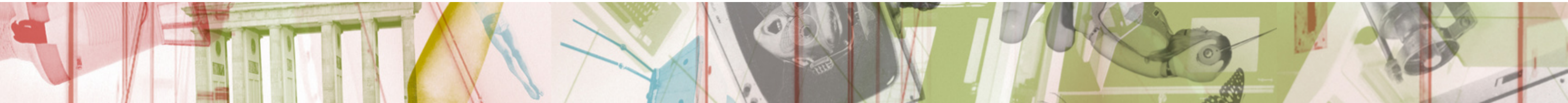
?

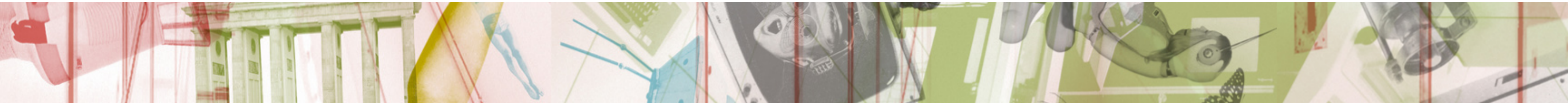
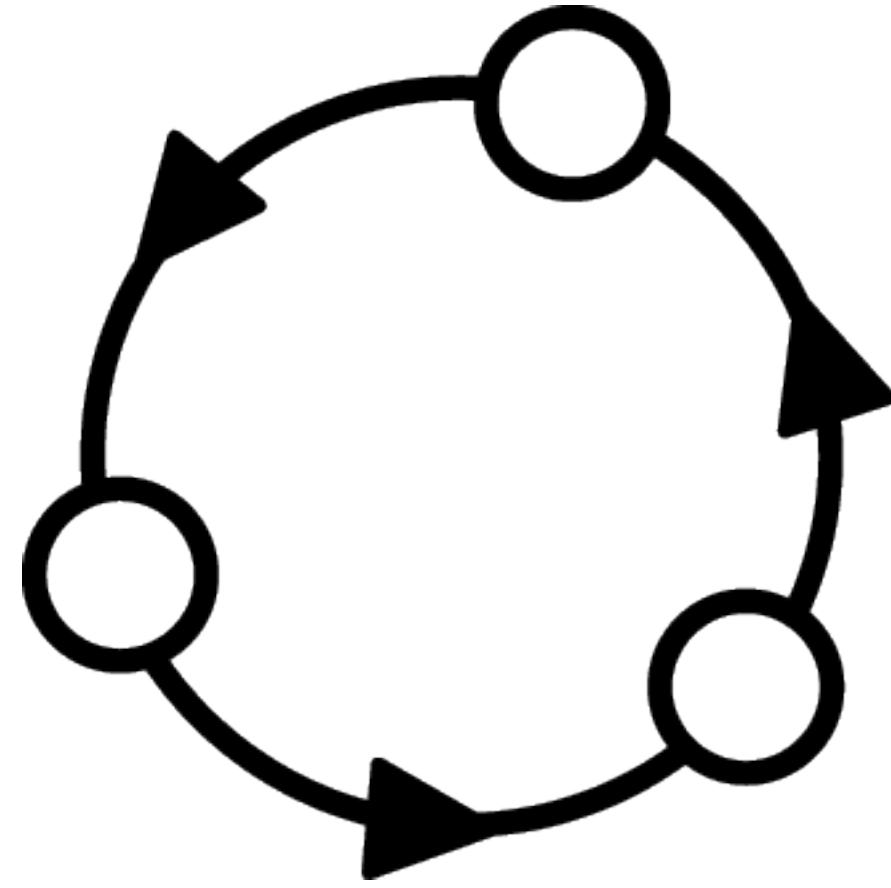
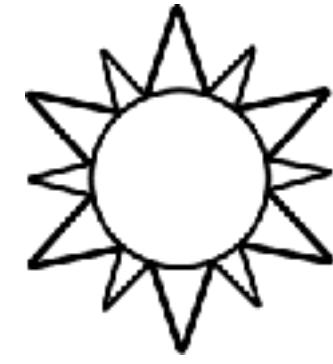
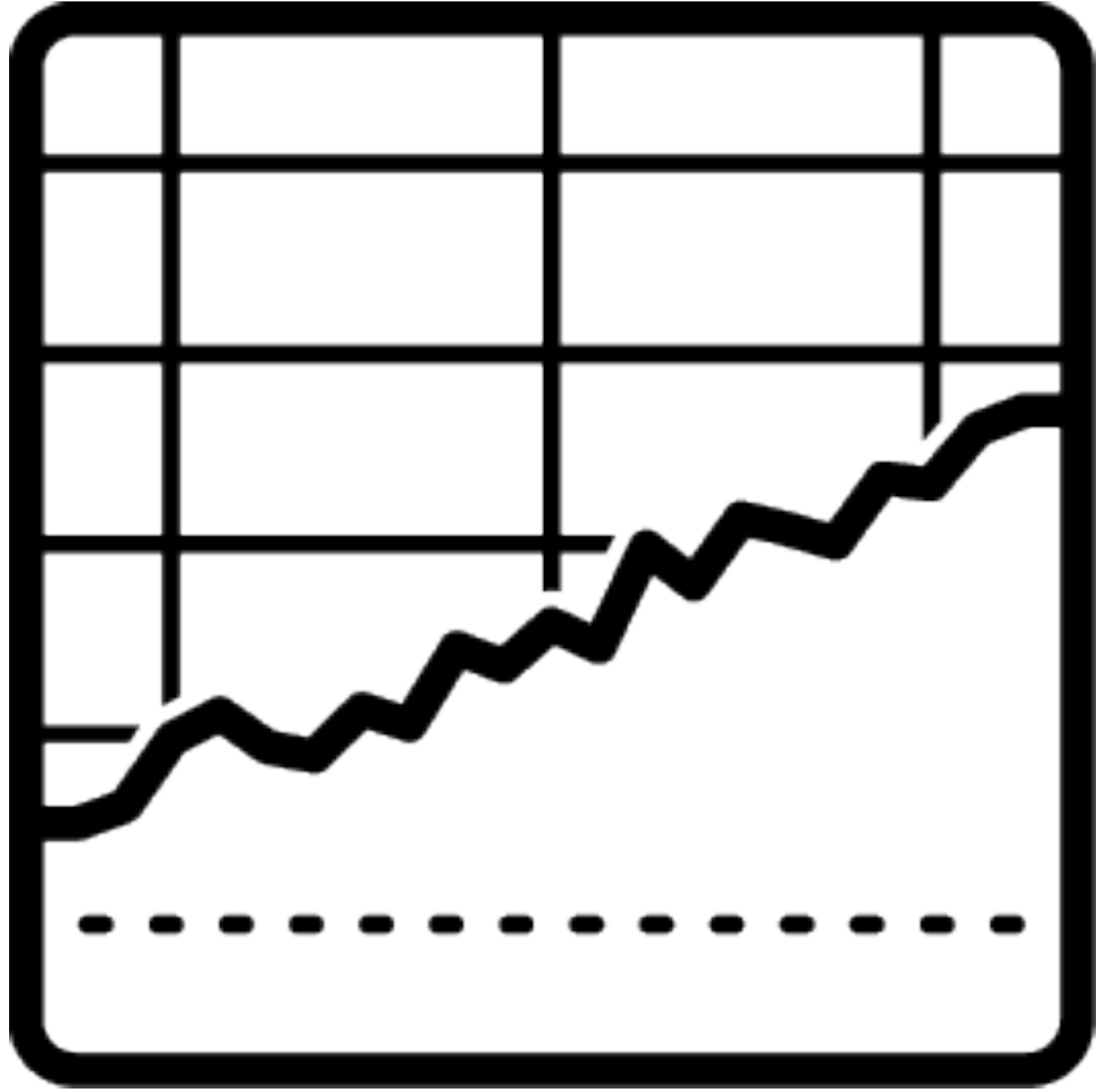


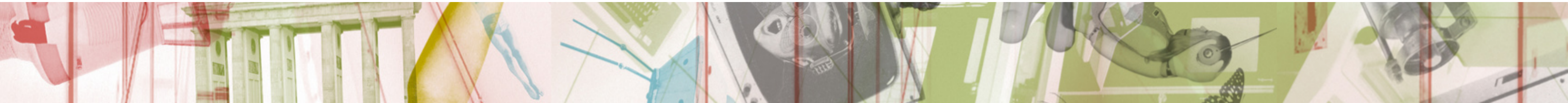
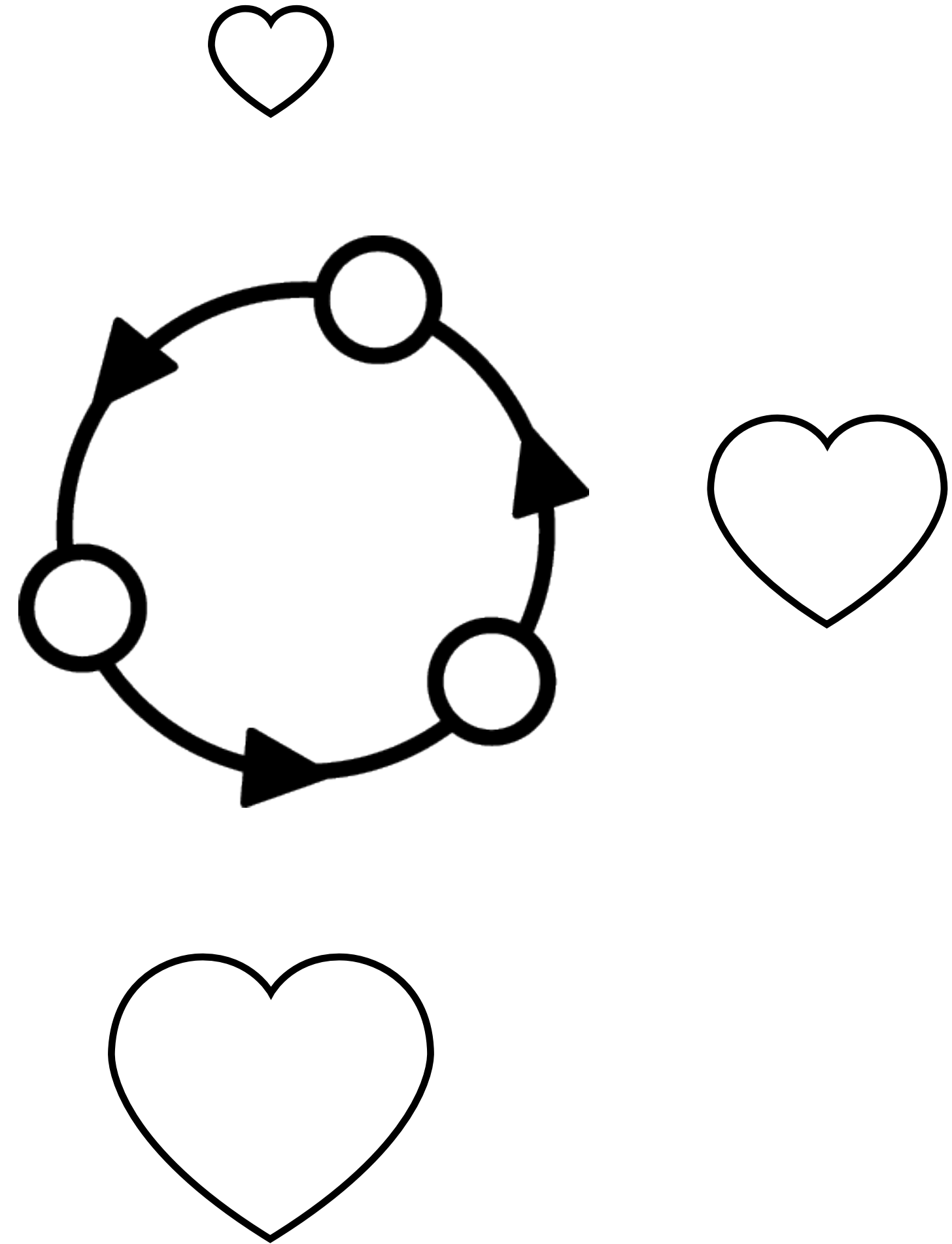
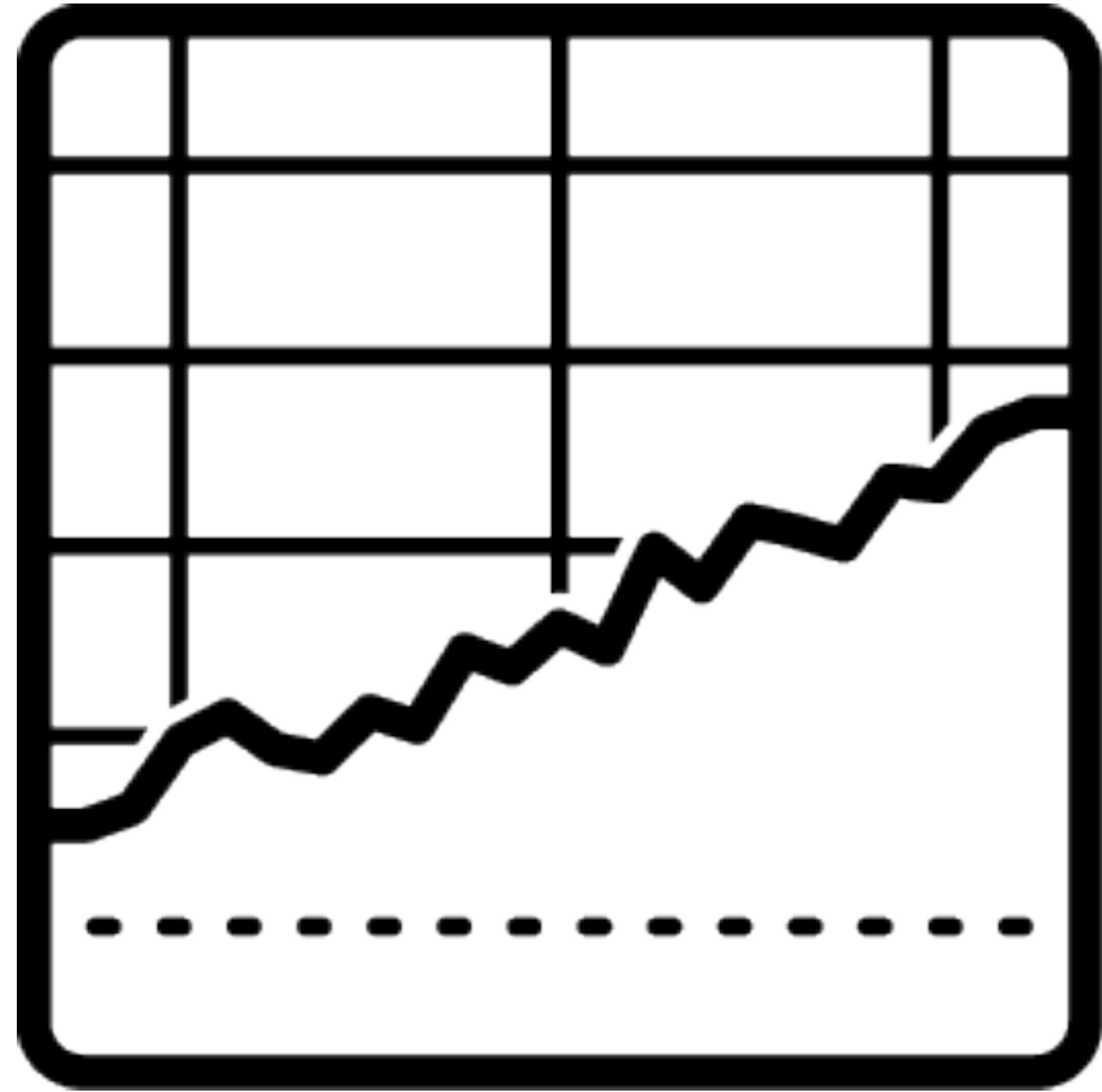


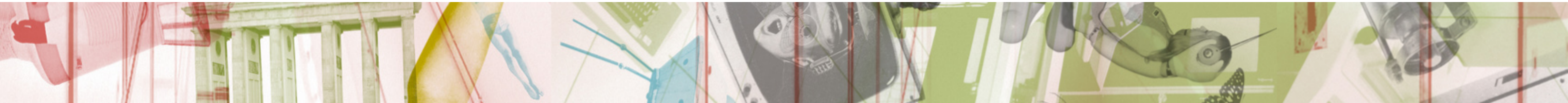
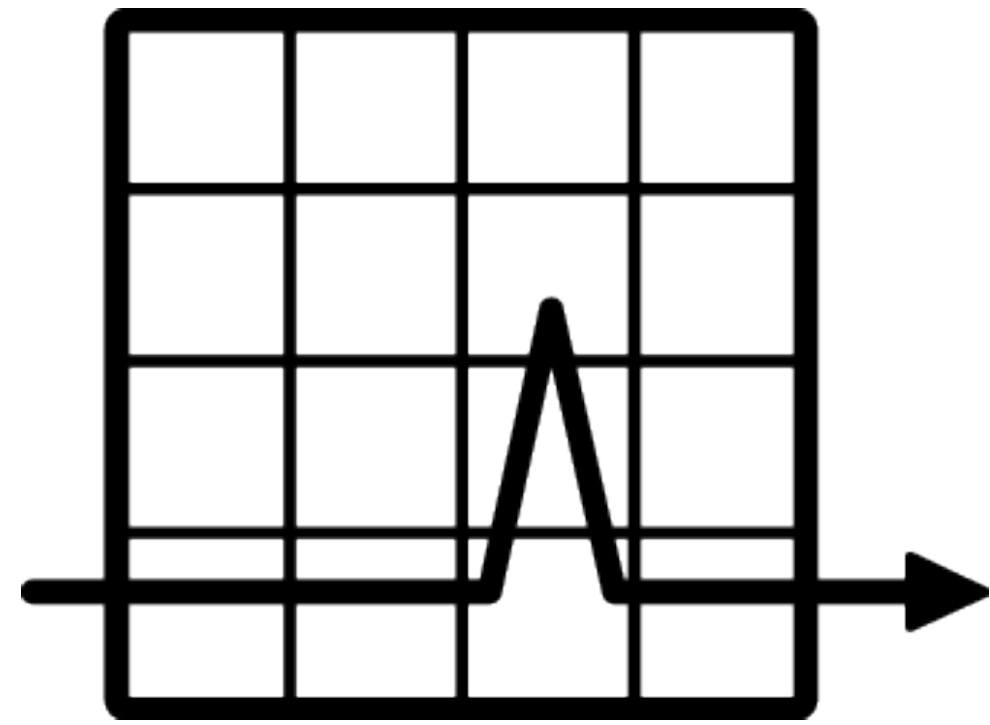


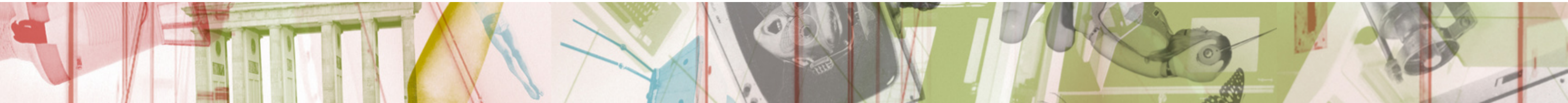
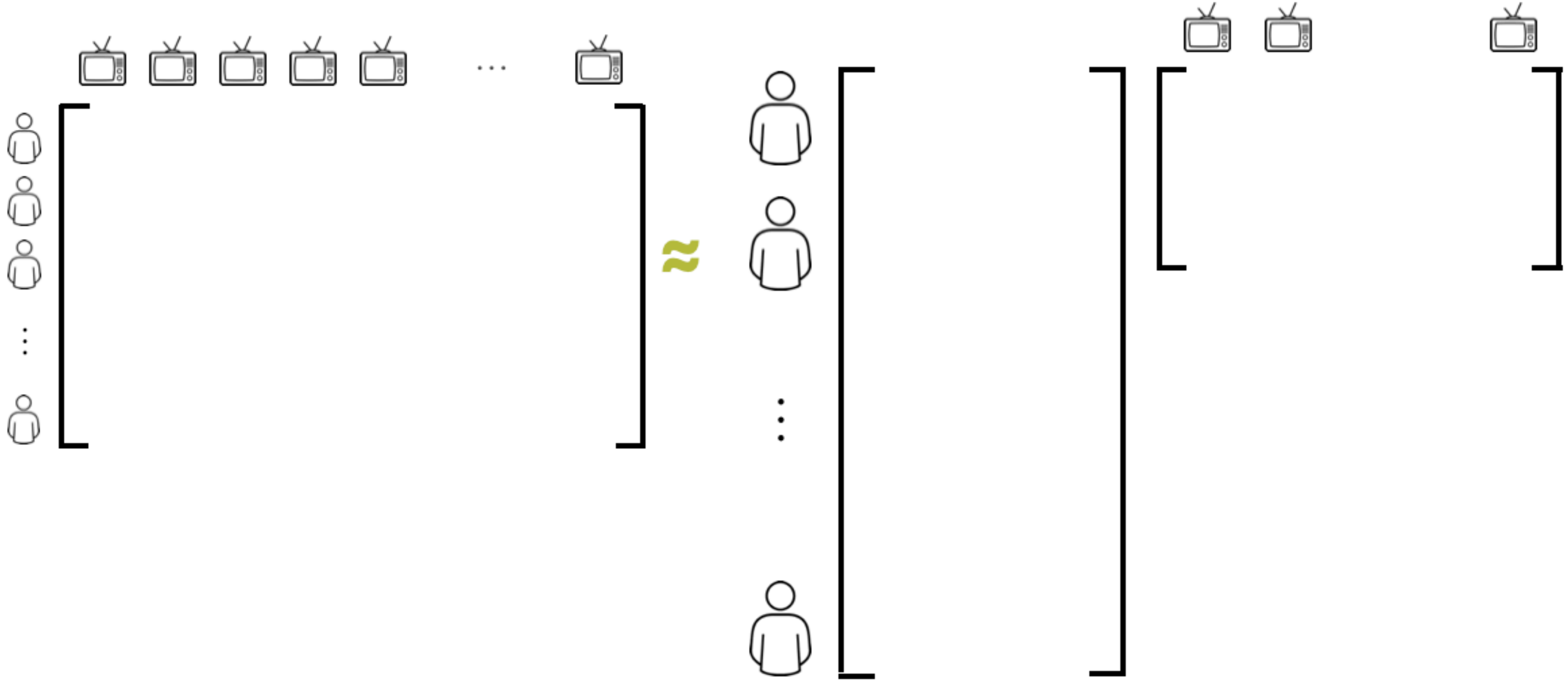
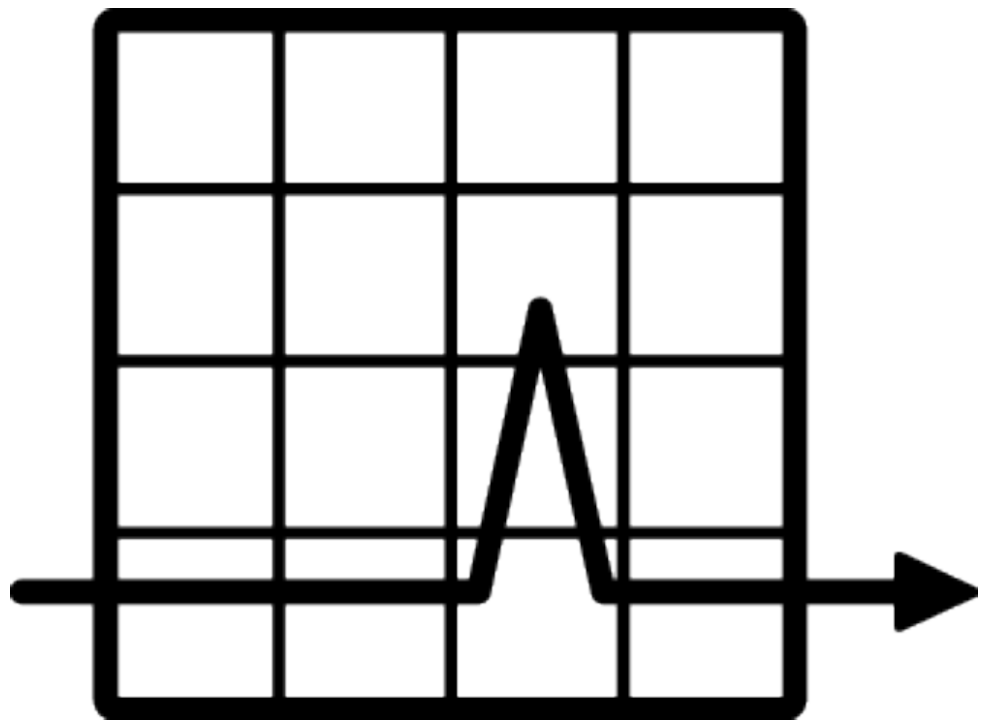
**time**







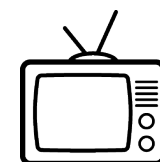
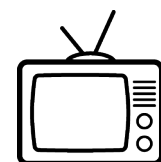
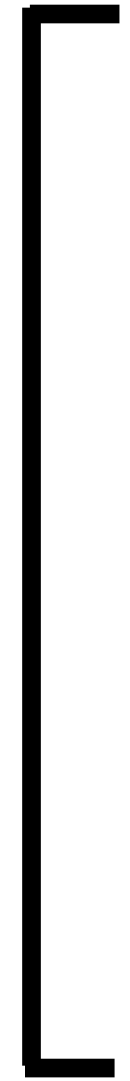




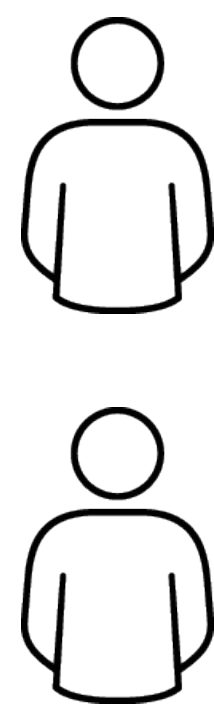
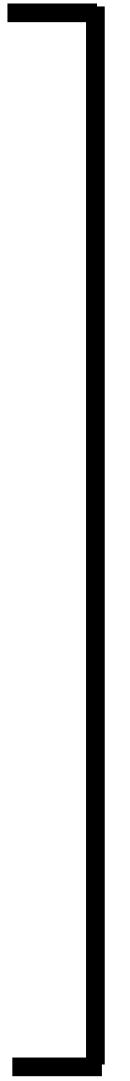
# Solutions



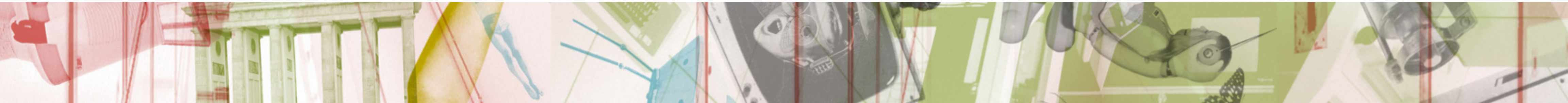
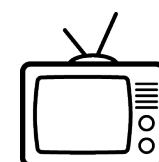
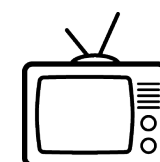
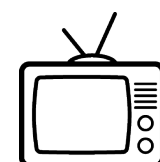
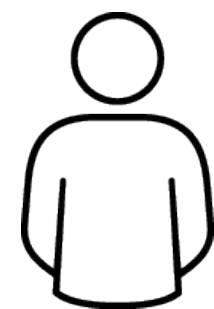




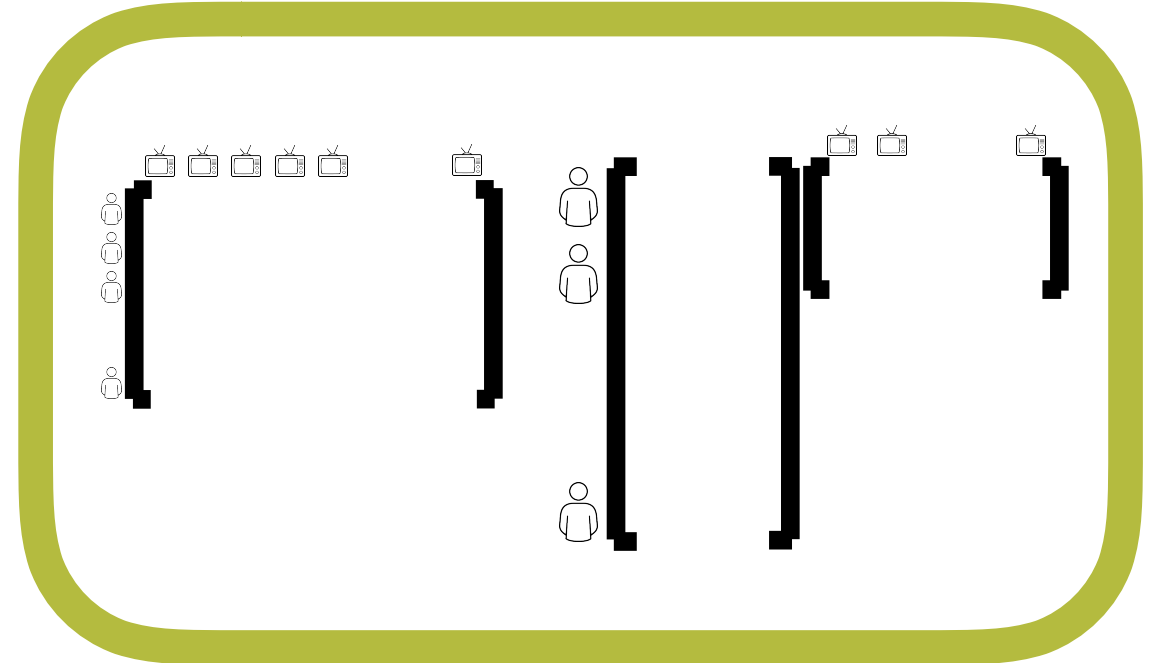
...



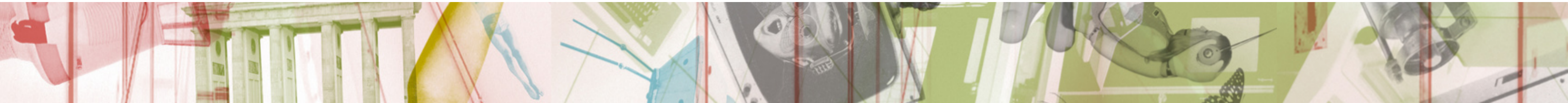
...



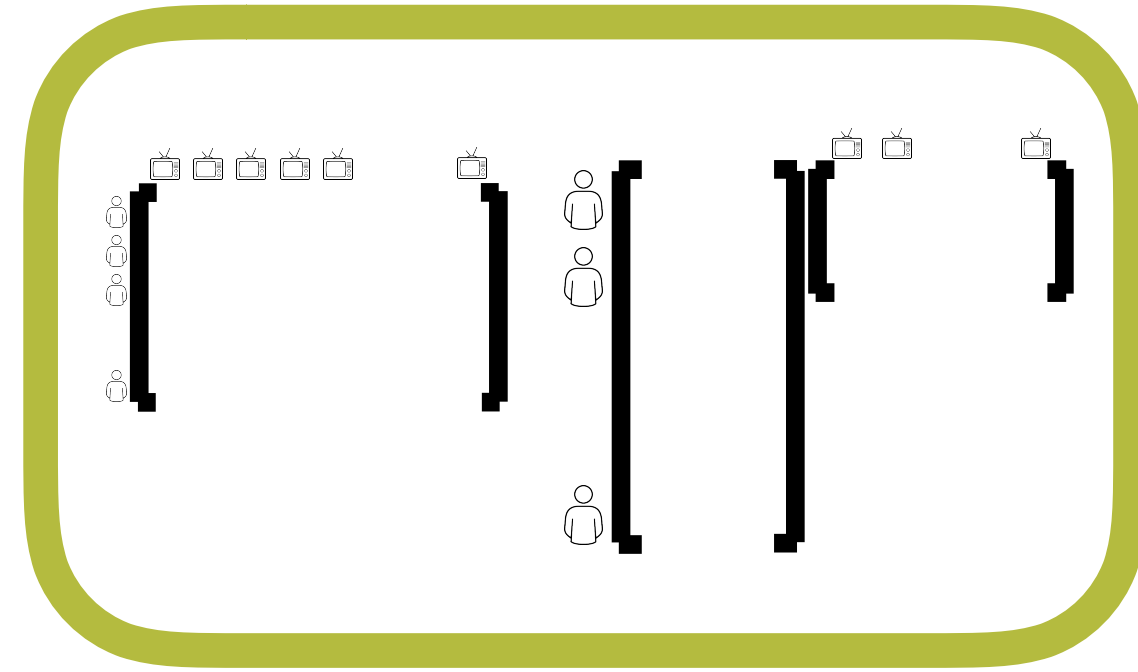
**Data**



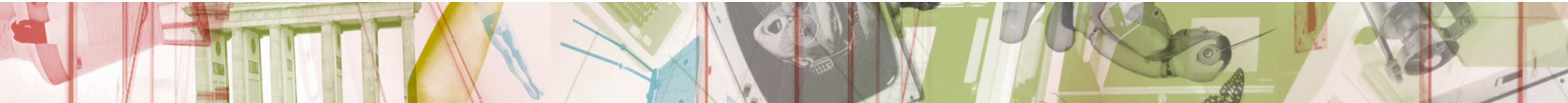
**Recommendations  
returned to users**

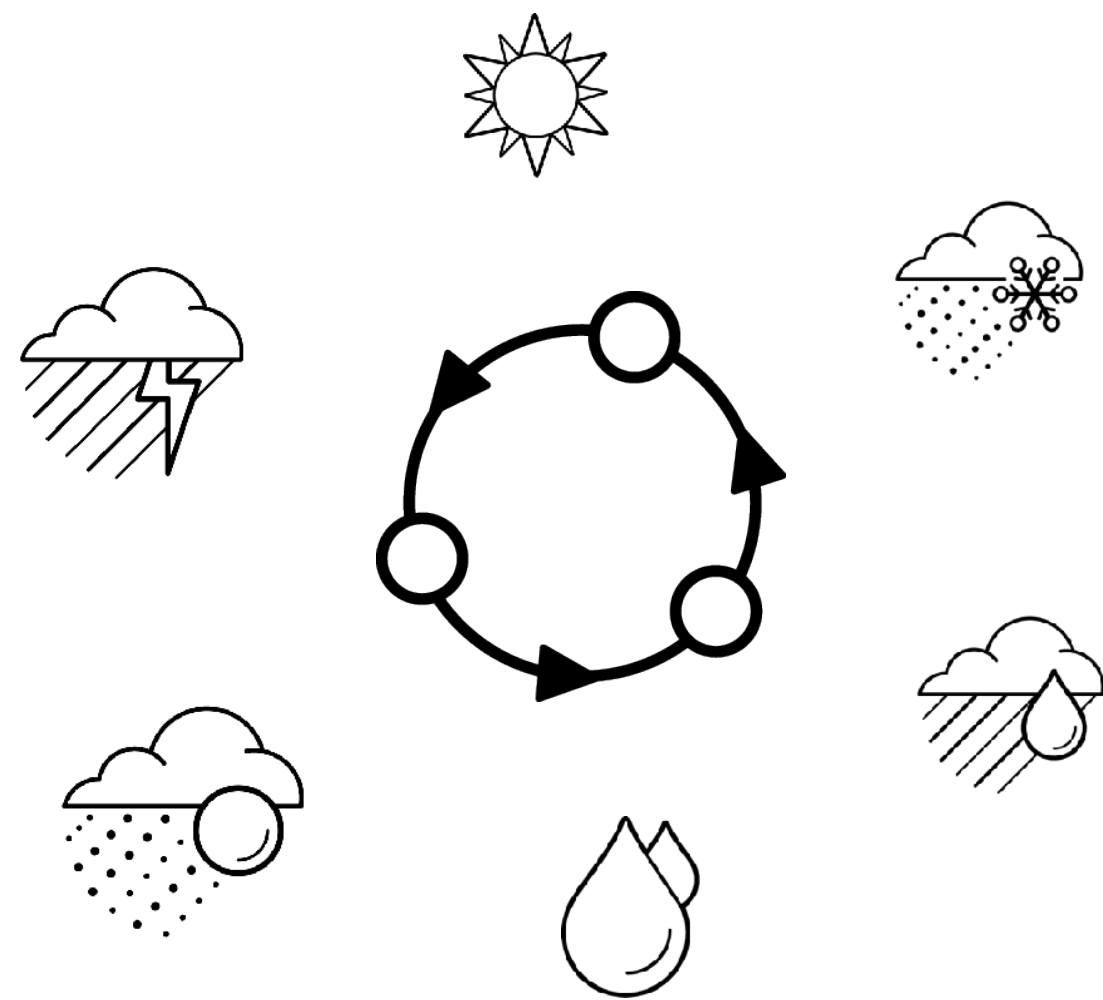
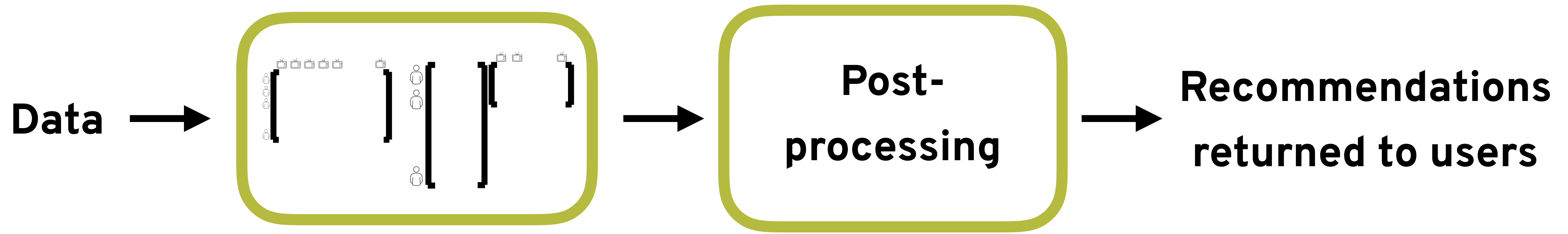


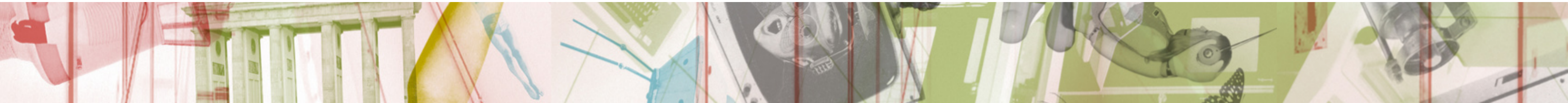
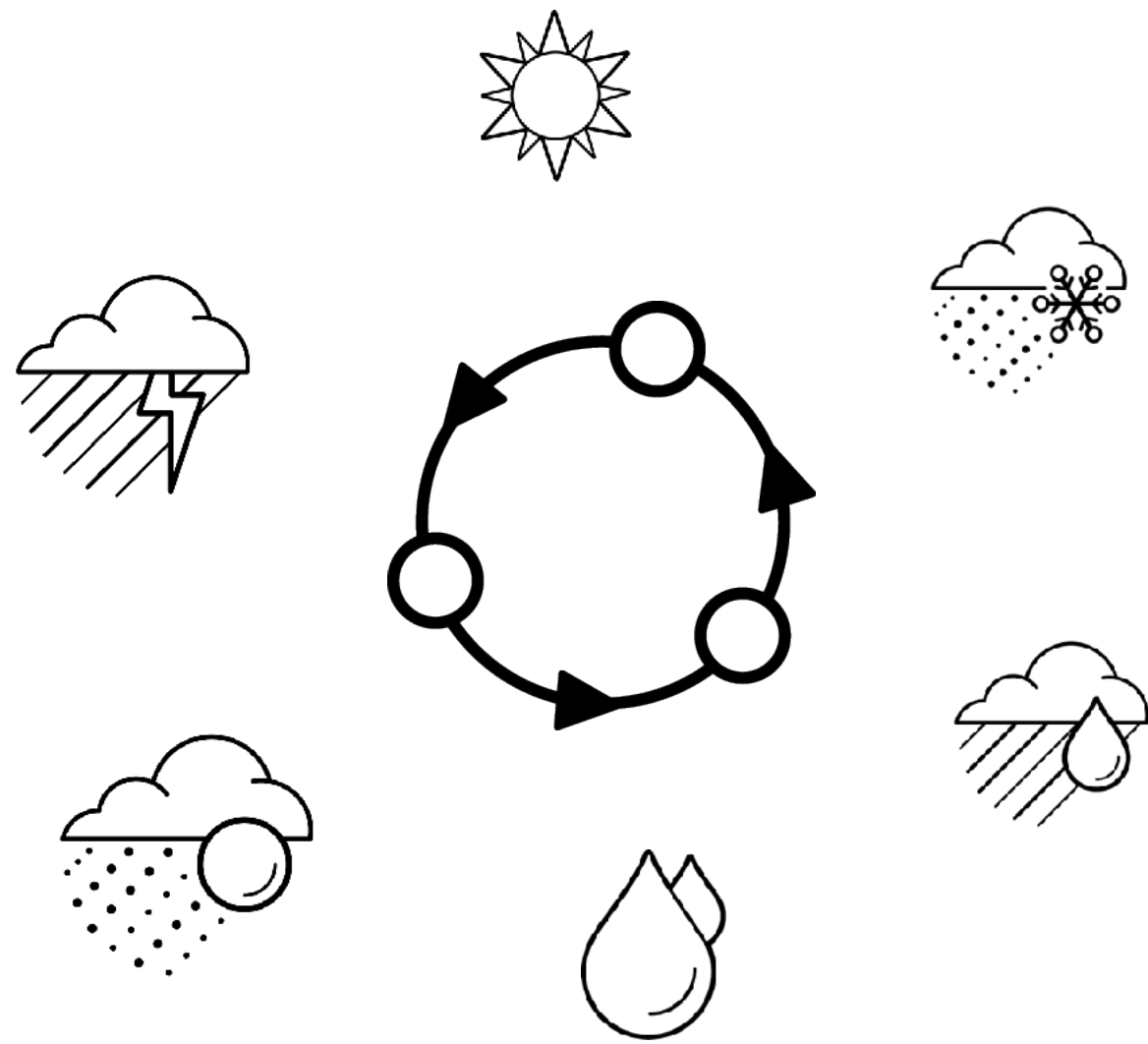
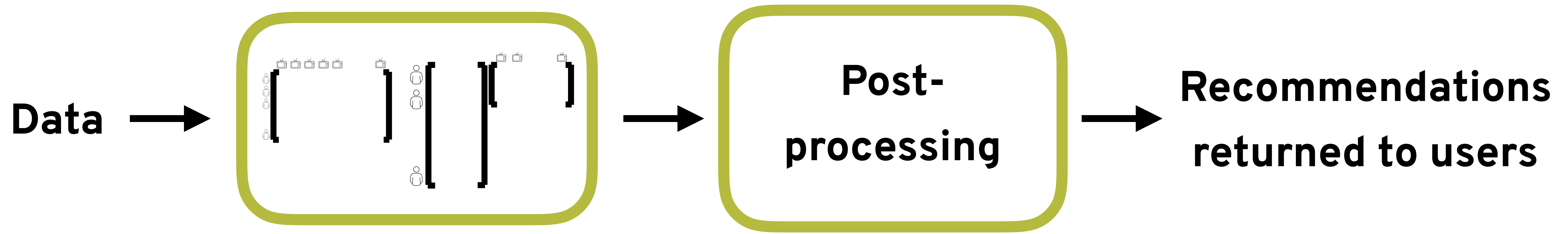
**Data**

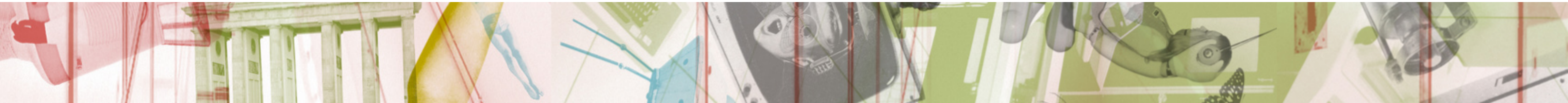
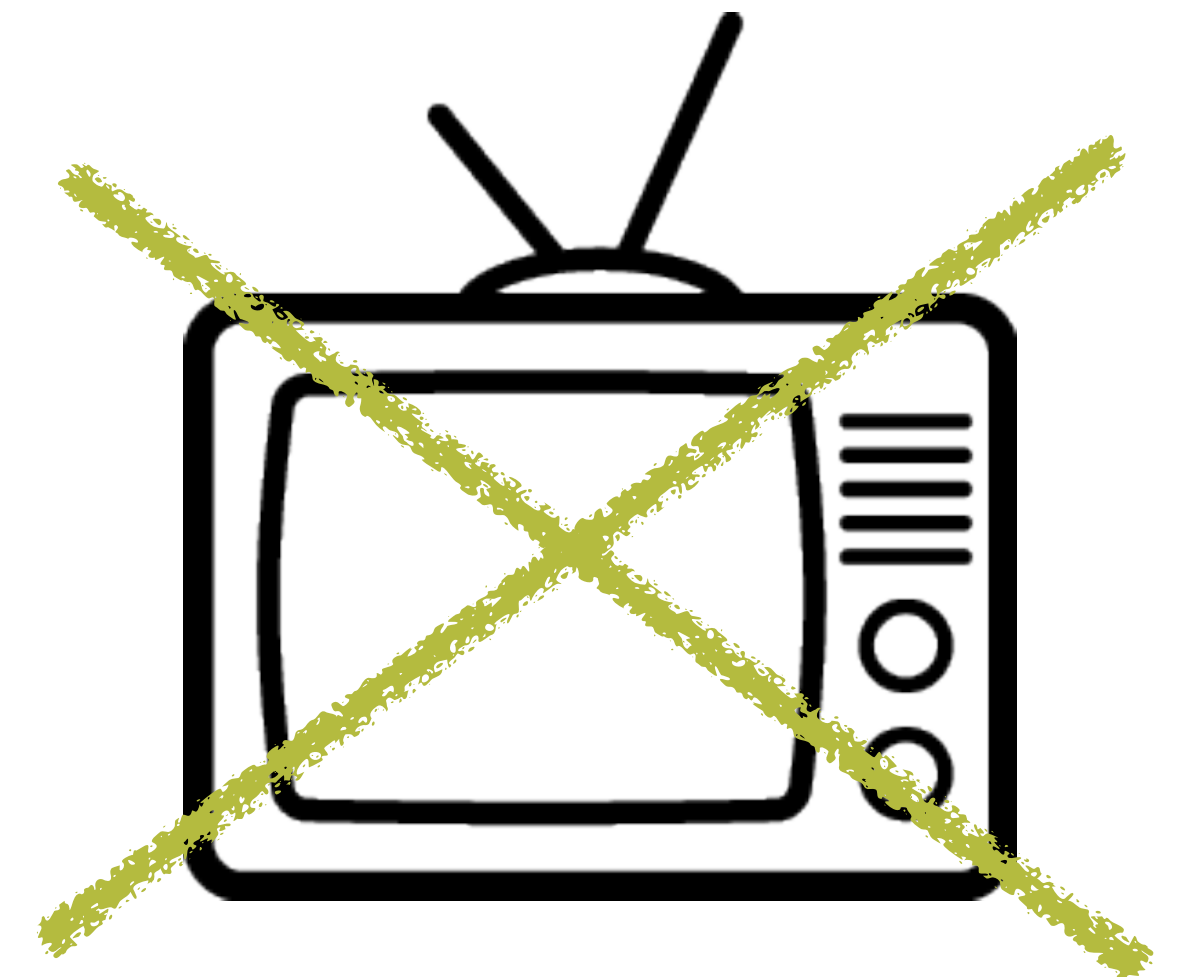
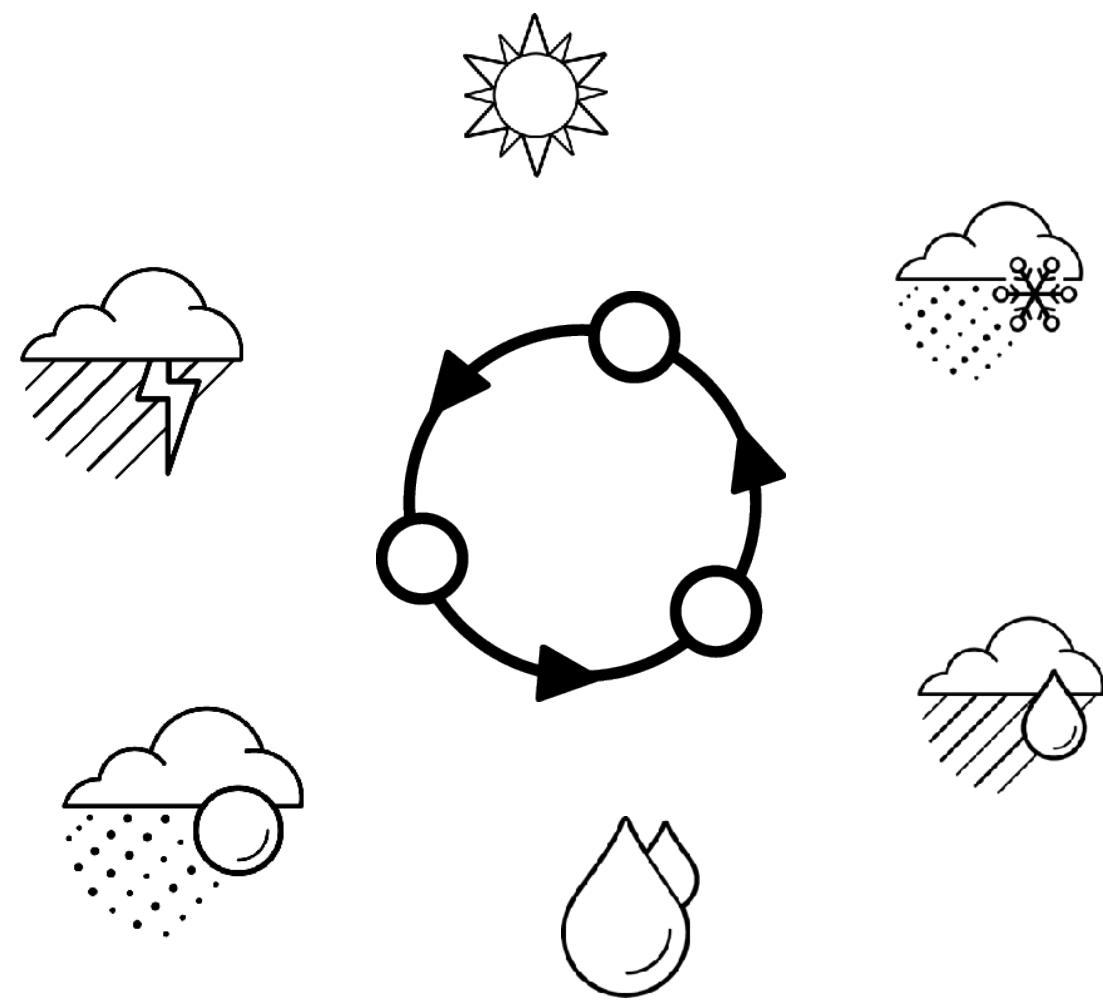
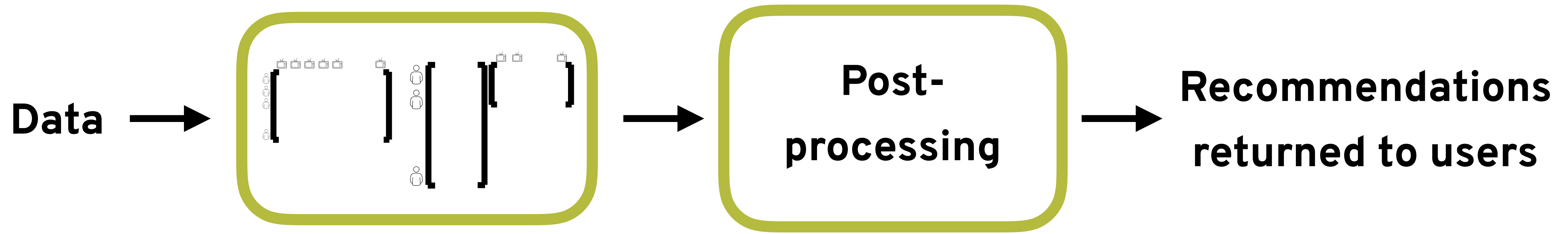


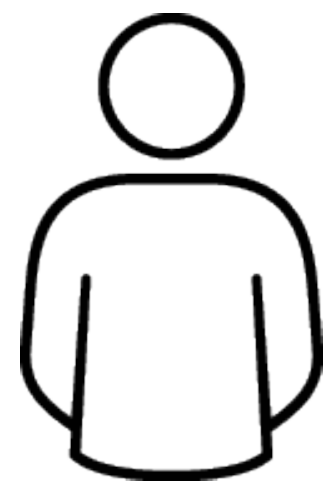
**Recommendations  
returned to users**



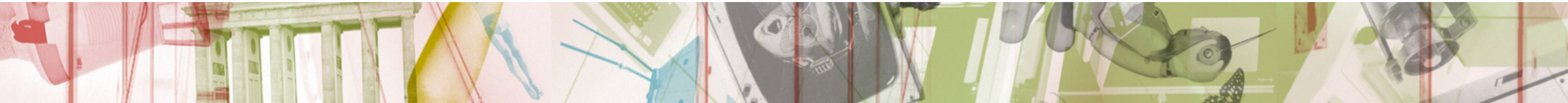








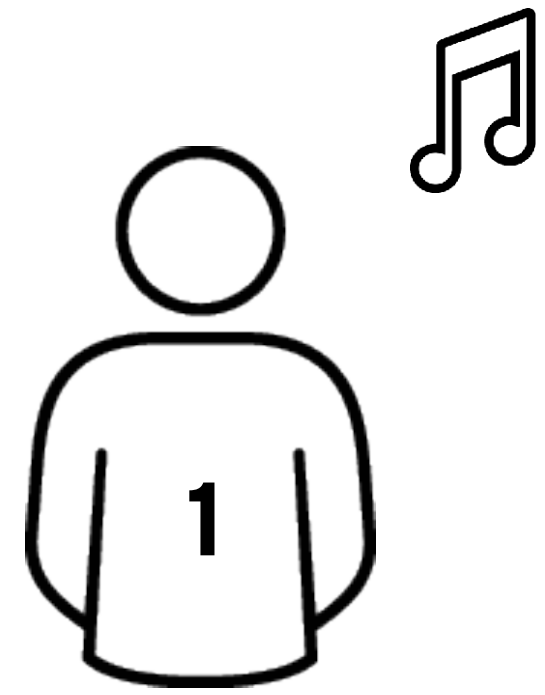
**time**



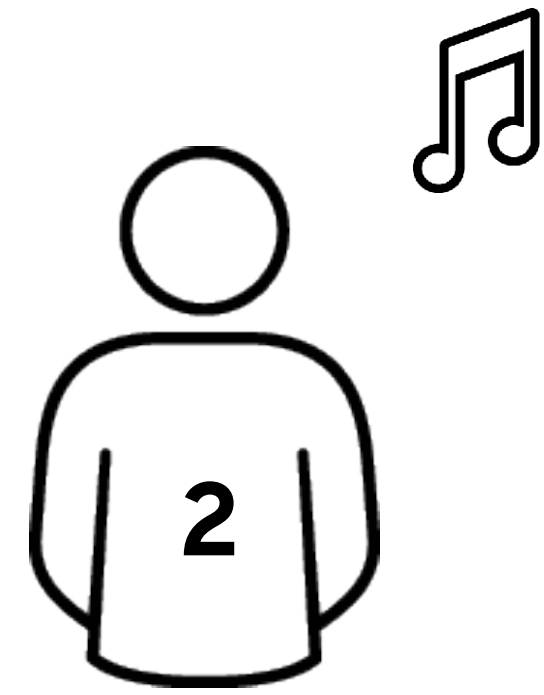
# Composable signatures for Recommendation



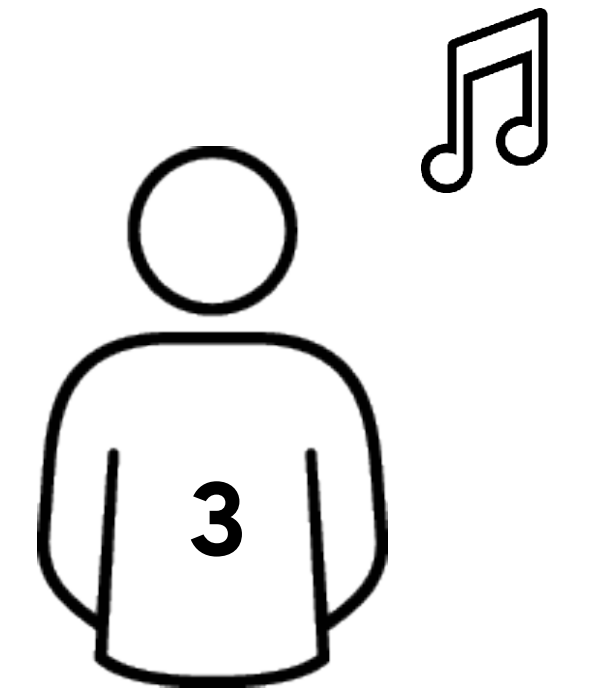




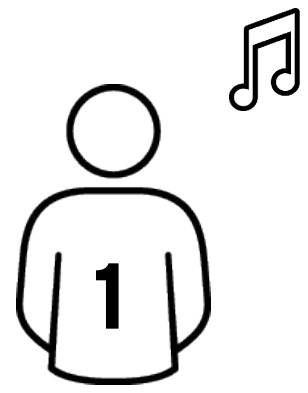
**Taylor Swift**  
**Pulp**  
**The Smiths**  
**Joy Division**  
**Billy Bragg**



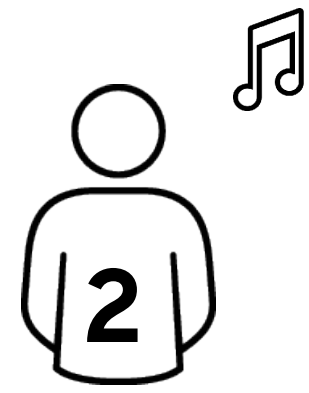
**Taylor Swift**  
**Pulp**  
**The Smiths**  
**Joy Division**  
**New Order**



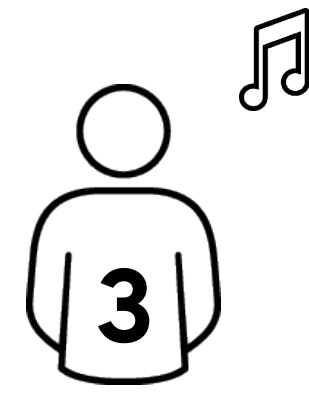
**Taylor Swift**  
**Miley Cyrus**  
**Little Mix**  
**Billie Eilish**  
**Spice Girls**



Taylor Swift  
Pulp  
The Smiths  
Joy Division  
Billy Bragg

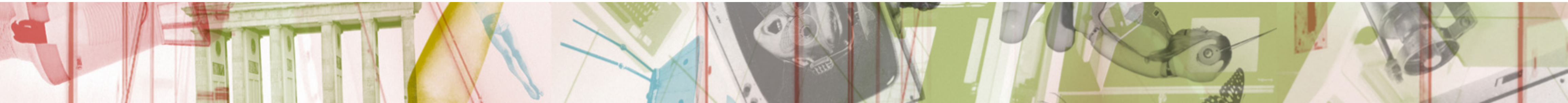


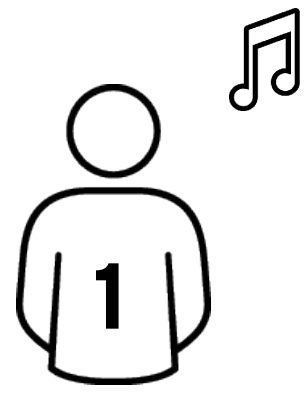
Taylor Swift  
Pulp  
The Smiths  
Joy Division  
New Order



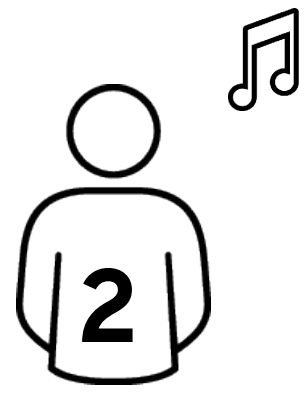
Taylor Swift  
Miley Cyrus  
Little Mix  
Billie Eilish  
Spice Girls

$$\text{Jaccard}(\text{1}, \text{2}) = \frac{|\text{Artist 1 listened to} \cap \text{Artist 2 listened to}|}{|\text{Artist 1 listened to} \cup \text{Artist 2 listened to}|}$$

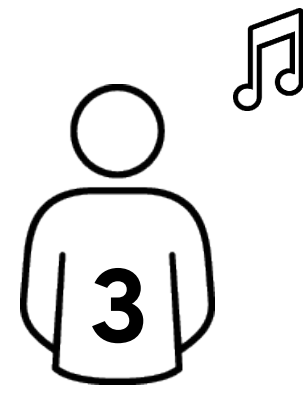




Taylor Swift  
Pulp  
The Smiths  
Joy Division  
Billy Bragg

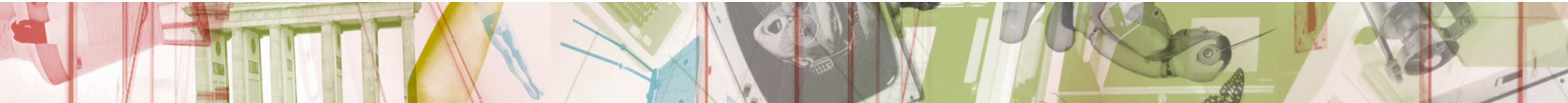


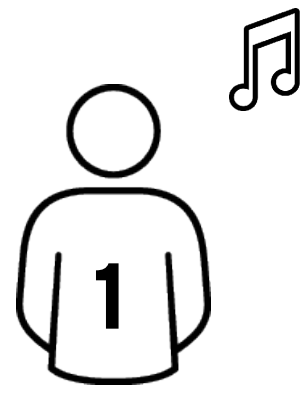
Taylor Swift  
Pulp  
The Smiths  
Joy Division  
New Order



Taylor Swift  
Miley Cyrus  
Little Mix  
Billie Eilish  
Spice Girls

$$\text{Jaccard}(\text{1}, \text{2}) = \frac{|\text{Artist 1 listened to} \cap \text{Artist 2 listened to}|}{|\text{Artist 1 listened to} \cup \text{Artist 2 listened to}|} = \frac{4}{6}$$

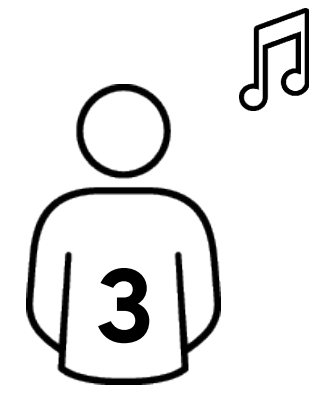




Taylor Swift  
Pulp  
The Smiths  
Joy Division  
Billy Bragg



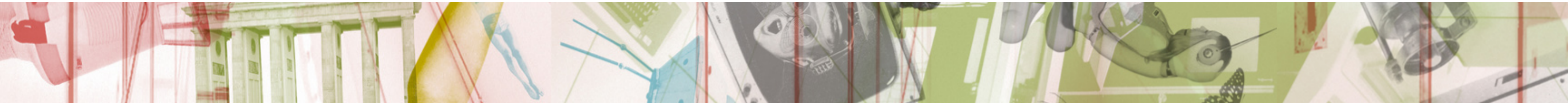
Taylor Swift  
Pulp  
The Smiths  
Joy Division  
New Order



Taylor Swift  
Miley Cyrus  
Little Mix  
Billie Eilish  
Spice Girls

$$\text{Jaccard}(\text{1}, \text{2}) = \frac{4}{6}$$

$$\text{Jaccard}(\text{1}, \text{3}) = \frac{|\text{Artist 1 listened to} \cap \text{Artist 3 listened to}|}{|\text{Artist 1 listened to} \cup \text{Artist 3 listened to}|} = \frac{1}{9}$$



$$\text{Jaccard}(\text{1}, \text{2}) = 0.3$$

$$\text{Jaccard}(\text{1}, \text{3}) = 0.5$$

$$\text{Jaccard}(\text{1}, \text{4}) = 0.1$$

$$\text{Jaccard}(\text{1}, \text{5}) = 0.1$$

$$\text{Jaccard}(\text{1}, \text{6}) = 0.6$$

$$\text{Jaccard}(\text{1}, \text{2}) = 0.3$$

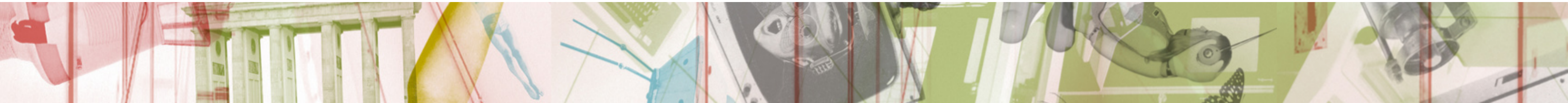
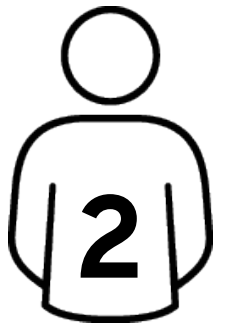
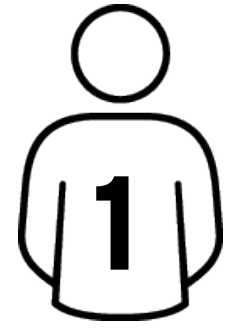
$$\text{Jaccard}(\text{1}, \text{3}) = 0.5$$

$$\text{Jaccard}(\text{1}, \text{4}) = 0.1$$

$$\text{Jaccard}(\text{1}, \text{5}) = 0.1$$

$$\text{Jaccard}(\text{1}, \text{6}) = 0.6$$

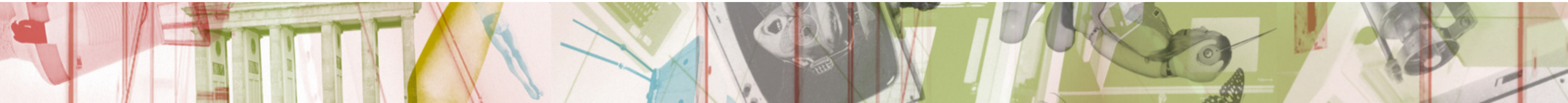


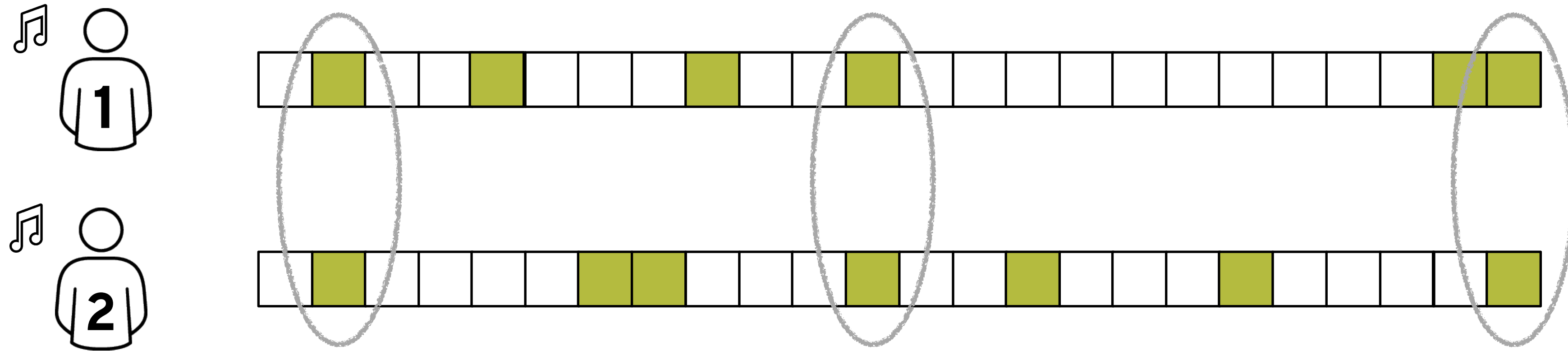




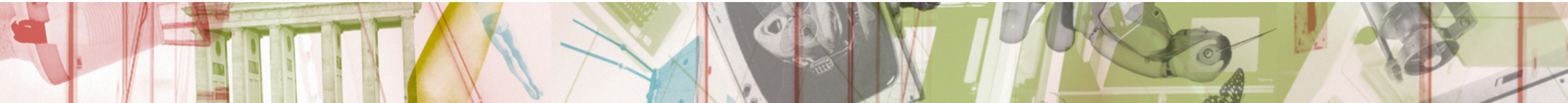


$$\text{Jaccard}(\text{Person 1}, \text{Person 2}) = \frac{|\text{Intersection}|}{|\text{Union}|}$$

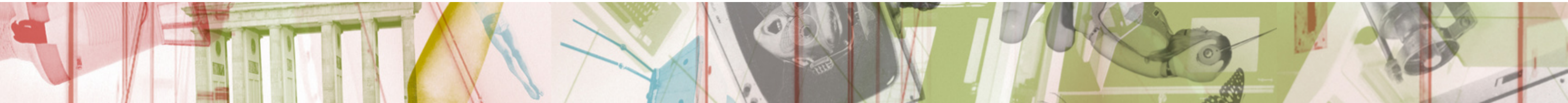
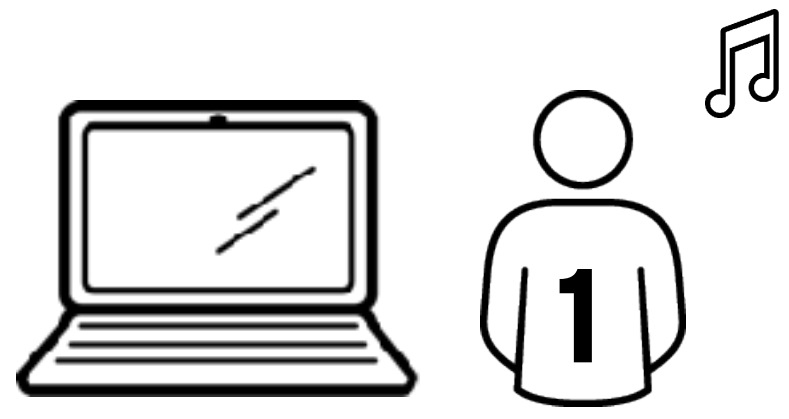


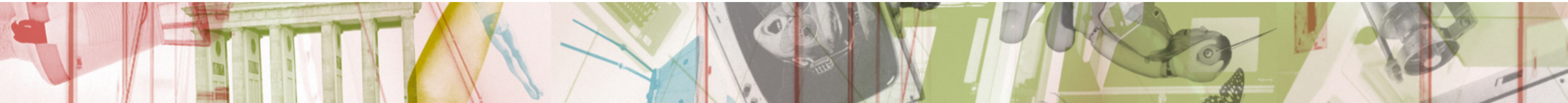
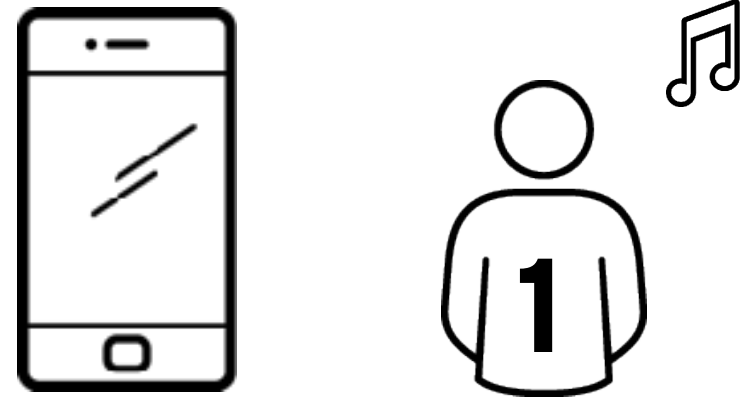
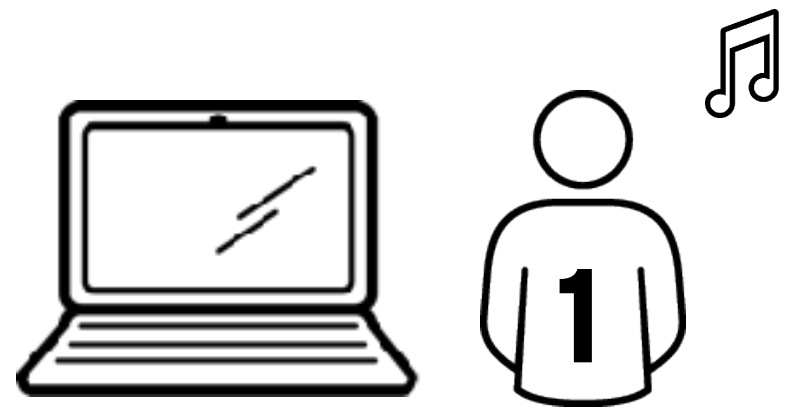


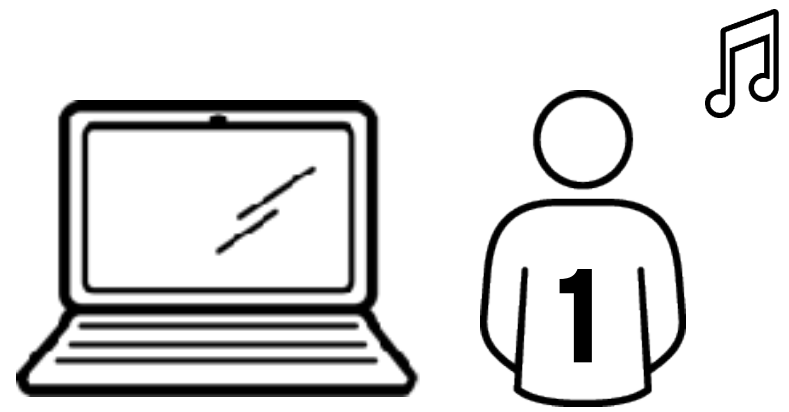
$$\text{Jaccard}(\text{1}, \text{2}) = \frac{\text{Intersection}}{\text{Union}}$$



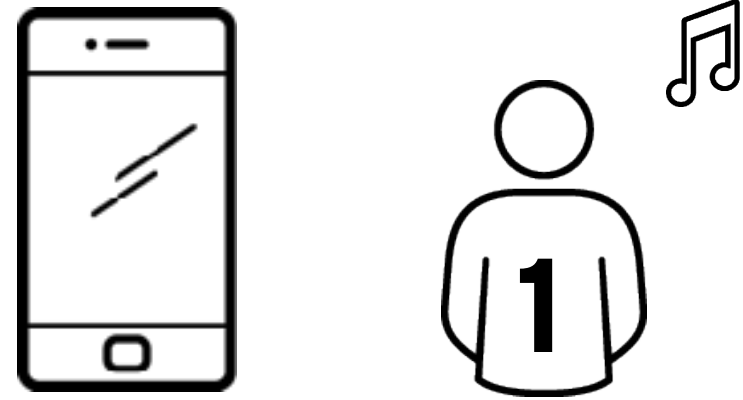




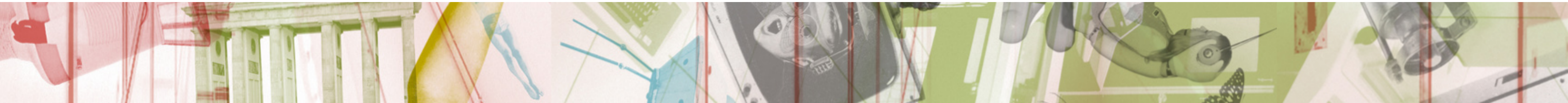


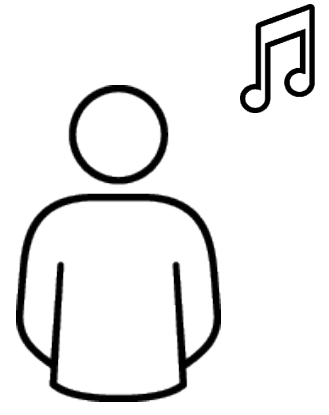


+



=





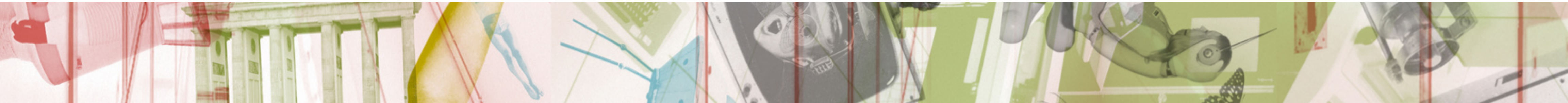
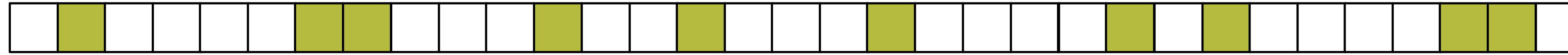
Artist id

Date-time

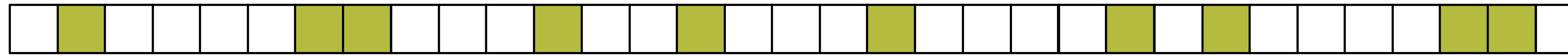
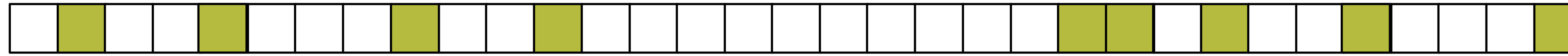
Summer

Winter

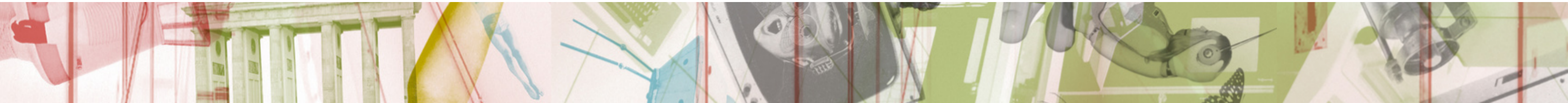
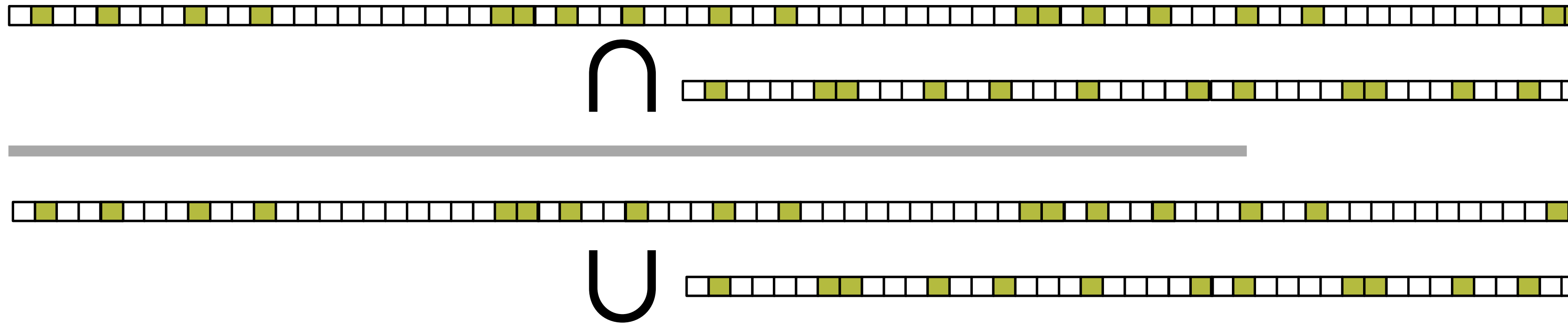


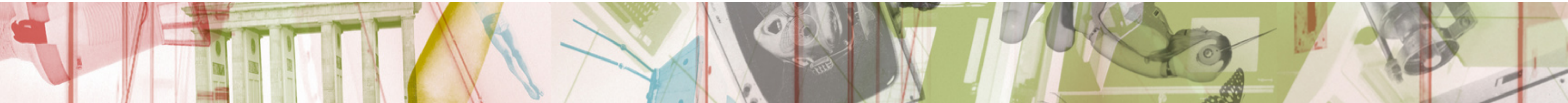
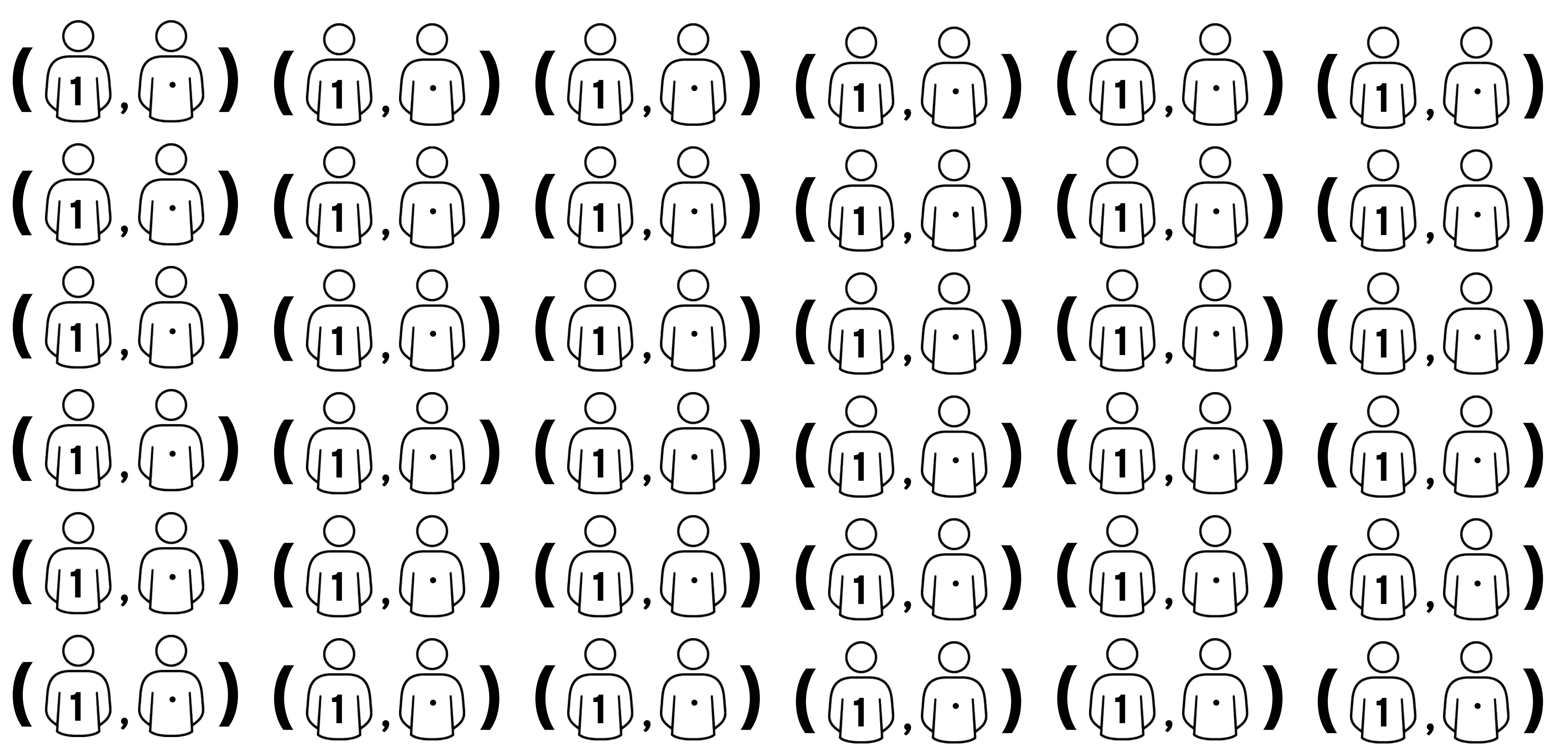






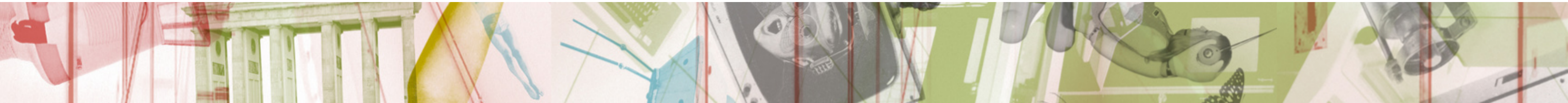
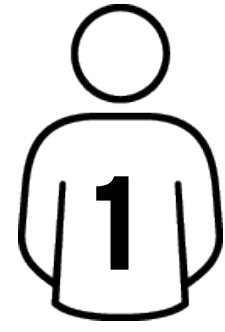
Jaccard (1, 2) =

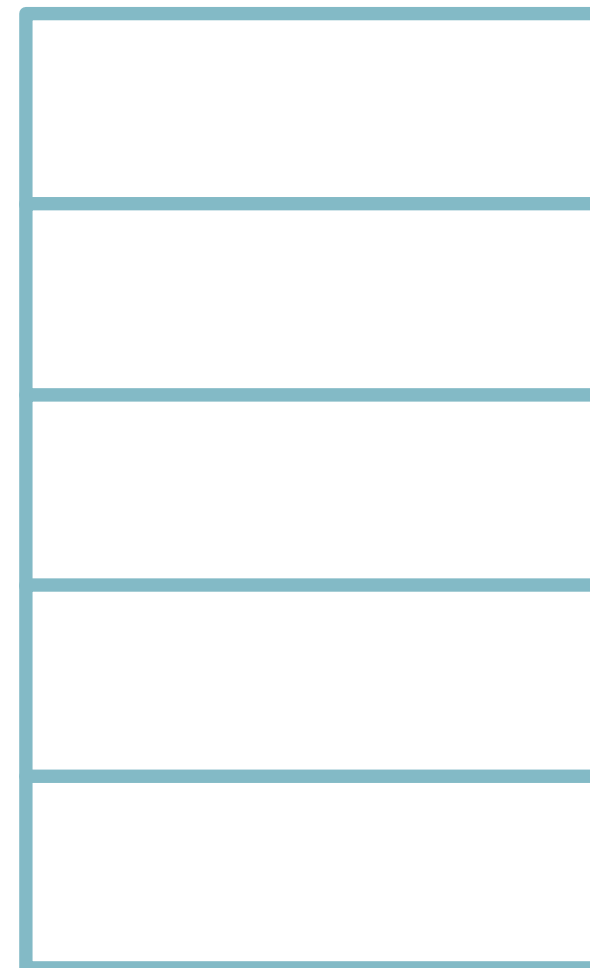




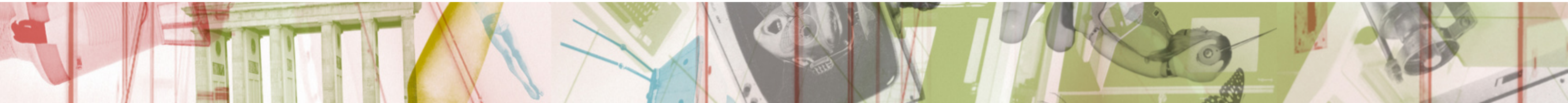
# MinHash for recommendation







**Min-hash signature**





Map from # artists to # artists.

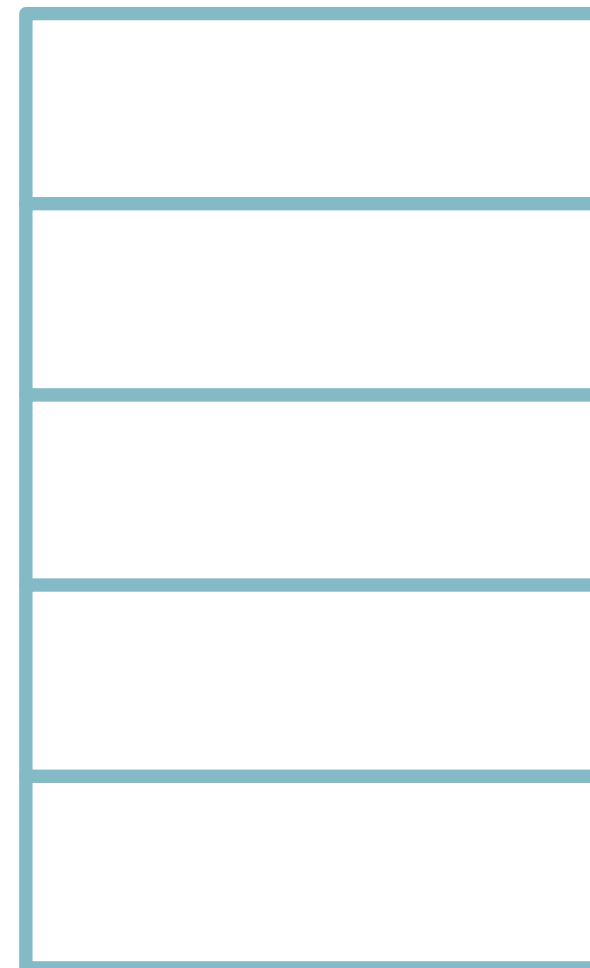
$h_1( )$

$h_2( )$

$h_3( )$

$h_4( )$

$h_5( )$



**Min-hash signature**



Map from # artists to # artists.

$h_1( )$

$\infty$

$h_2( )$

$\infty$

$h_3( )$

$\infty$

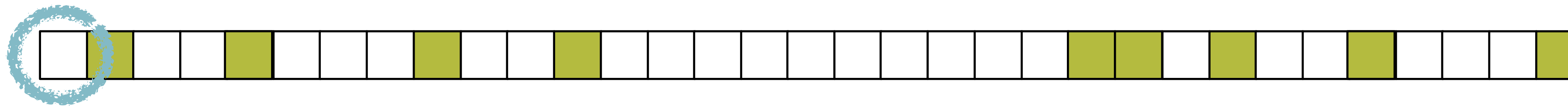
$h_4( )$

$\infty$

$h_5( )$

$\infty$

**Min-hash signature**



Map from # artists to # artists.

$h_1( )$

$\infty$

$h_2( )$

$\infty$

$h_3( )$

$\infty$

$h_4( )$

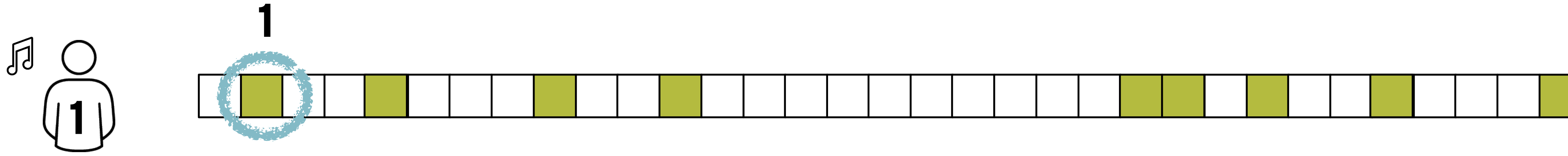
$\infty$

$h_5( )$

$\infty$

**Min-hash signature**





Map from # artists to # artists.

$h_1( )$

$\infty$

$h_2( )$

$\infty$

$h_3( )$

$\infty$

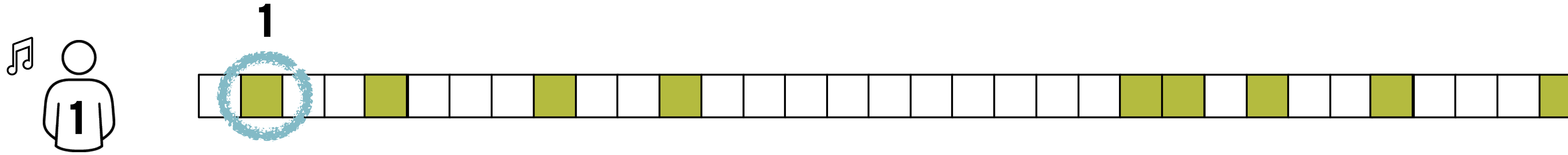
$h_4( )$

$\infty$

$h_5( )$

$\infty$

**Min-hash signature**



Map from # artists to # artists.

$h_1(\mathbf{1})$

$\infty$

$h_2(\mathbf{1})$

$\infty$

$h_3(\mathbf{1})$

$\infty$

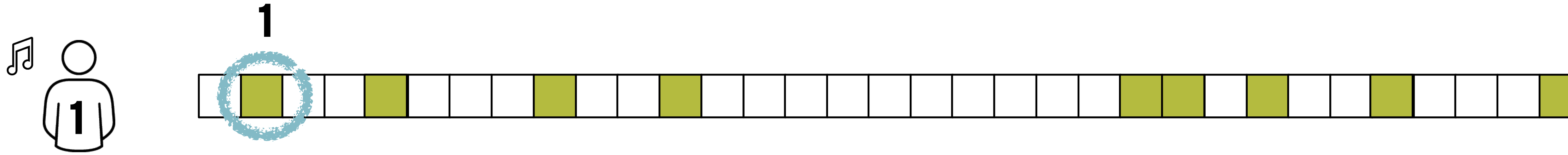
$h_4(\mathbf{1})$

$\infty$

**Min-hash signature**

$h_5(\mathbf{1})$

$\infty$



Map from # artists to # artists.

$h_1(\mathbf{1})$

$\infty$

6

$h_2(\mathbf{1})$

$\infty$

12,048

$h_3(\mathbf{1})$

$\infty$

98

$h_4(\mathbf{1})$

$\infty$

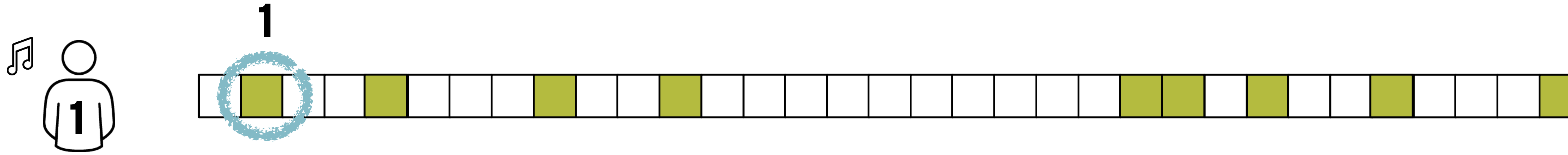
564

**Min-hash signature**

$h_5(\mathbf{1})$

$\infty$

71



Map from # artists to # artists.

$h_1(\mathbf{1})$

**6**

$h_2(\mathbf{1})$

$\infty$

**12,048**

$h_3(\mathbf{1})$

$\infty$

**98**

$h_4(\mathbf{1})$

$\infty$

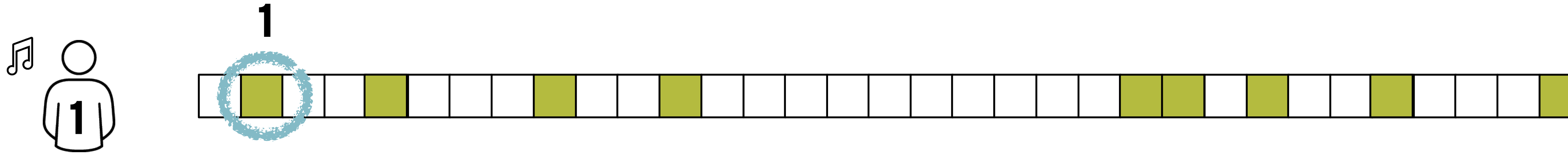
**564**

**Min-hash signature**

$h_5(\mathbf{1})$

$\infty$

**71**



*Map from # artists to # artists.*

$h_1(\mathbf{1})$

**6**

$h_2(\mathbf{1})$

**12,048**

$h_3(\mathbf{1})$

**98**

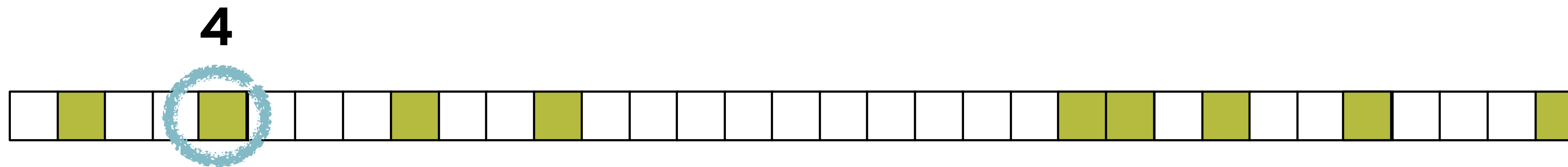
$h_4(\mathbf{1})$

**564**

$h_5(\mathbf{1})$

**71**

**Min-hash signature**



Map from # artists to # artists.

$h_1(4)$

6

17

$h_2(4)$

12,048

11,097

$h_3(4)$

98

43,861

$h_4(4)$

564

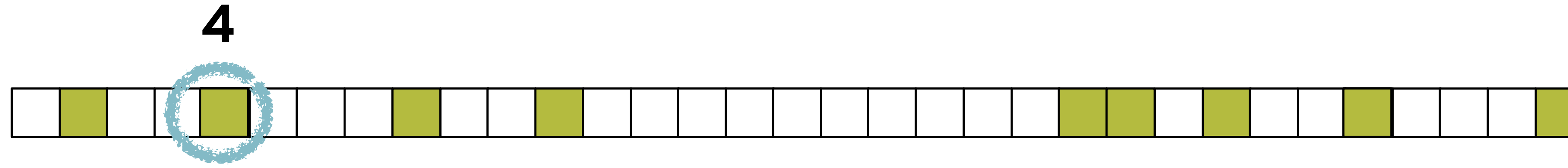
93

**Min-hash signature**

$h_5(4)$

71

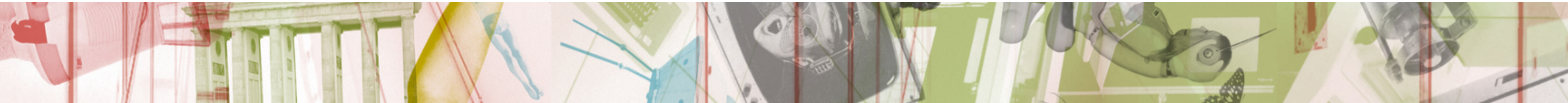
111,123



Map from # artists to # artists.

$h_1(4)$	6
$h_2(4)$	11,097
$h_3(4)$	98
$h_4(4)$	93
$h_5(4)$	71

**Min-hash signature**

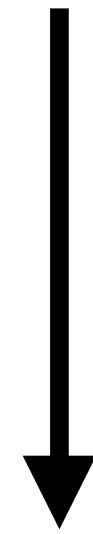




6
1,987
91
93
43

**Min-hash signature**



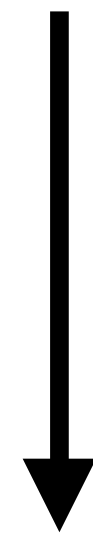


6
1,987
91
93
43

9,872
1,987
486
32,821
43

**Min-hash signature**



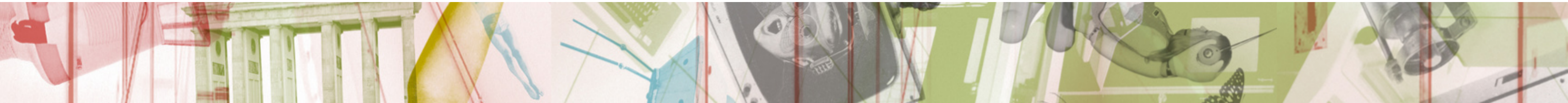
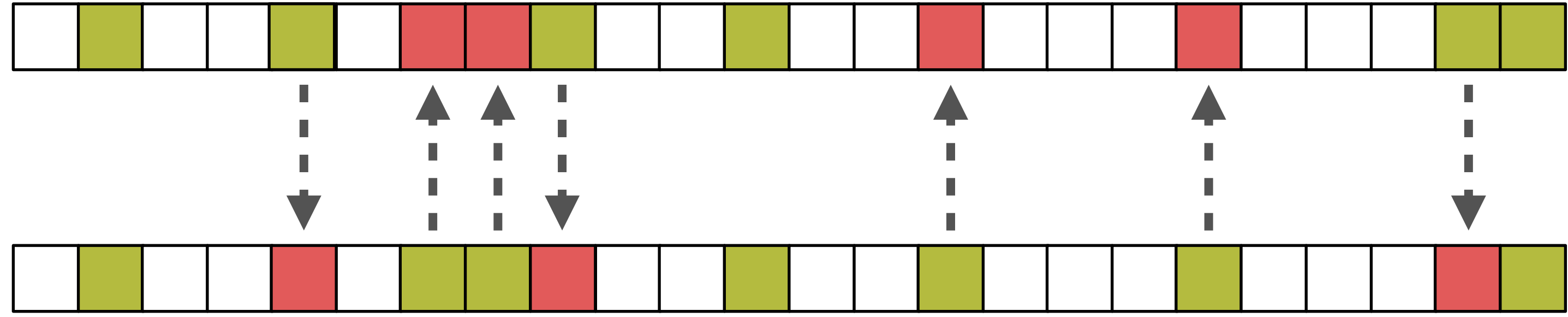


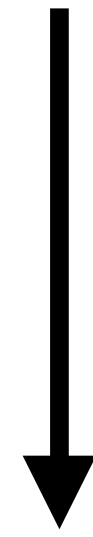
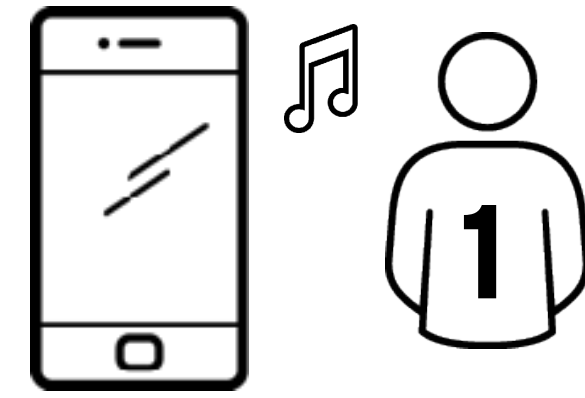
6
1,987
91
93
43

9,872
1,987
486
32,821
43

**Min-hash signature**





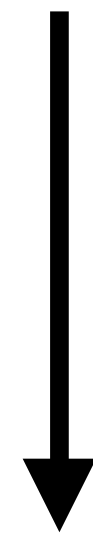
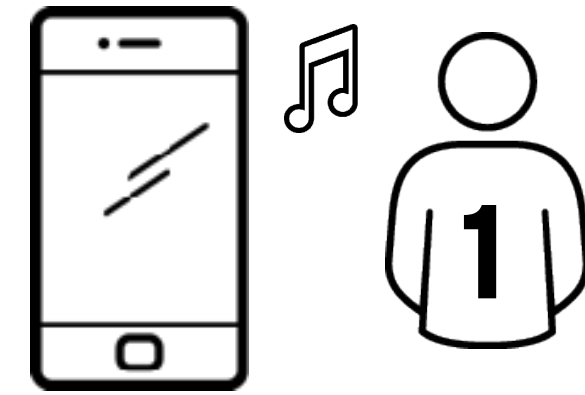
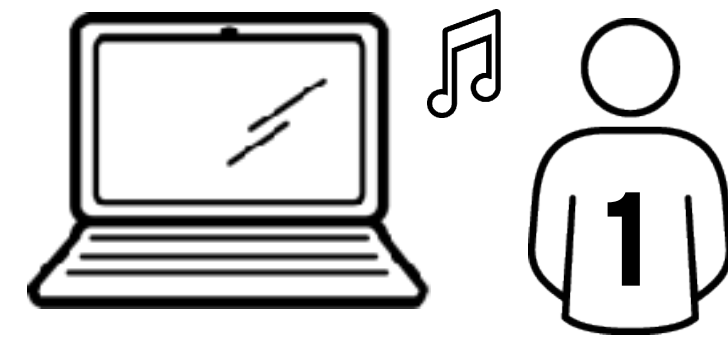


6
1,987
91
93
43

9,872
12
486
84
43

**Min-hash signature**





6
1,987
91
93
43

+

9,872
12
486
84
43

=

6
12
91
84
43

**Min-hash signature**



Artist	Date
--------	------

Map from *ARTIST NAMES* to # artists.

**Min-hash signature**

$h_1( )$

$h_2( )$

$h_3( )$

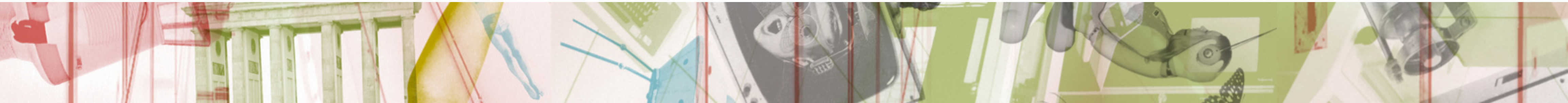
$h_4( )$

$h_5( )$






**BERLIN  
BUZZWORDS  
2019** JUNE 16-18

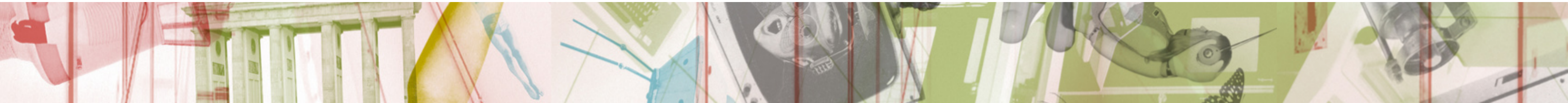


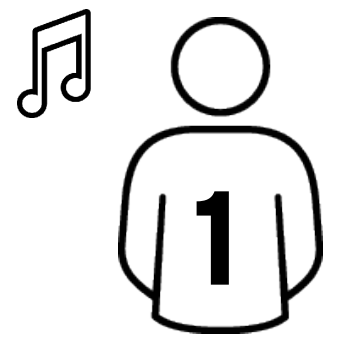


# Locality sensitive MinHash







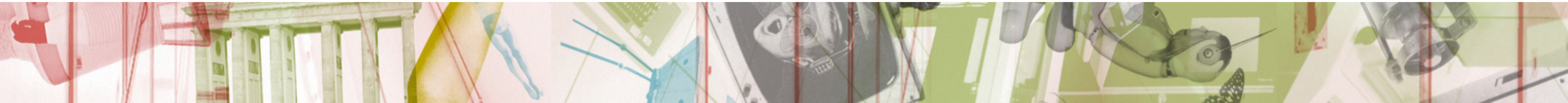

**Band 1**

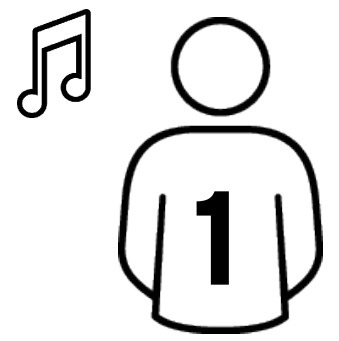
**Band 2**

**Band 3**

**Band 4**

**Band 5**





Map from vecs of band size to large numbers.  $h(\ )$

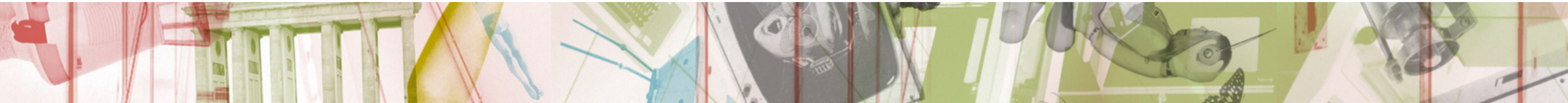

} **Band 1**

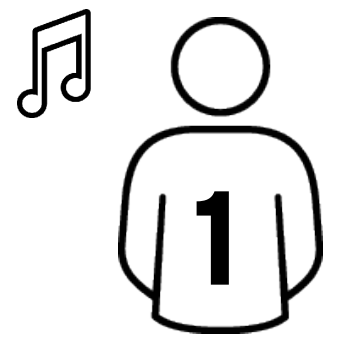
} **Band 2**

} **Band 3**

} **Band 4**

} **Band 5**





Map from vecs of band size to large numbers.  $h(\ )$

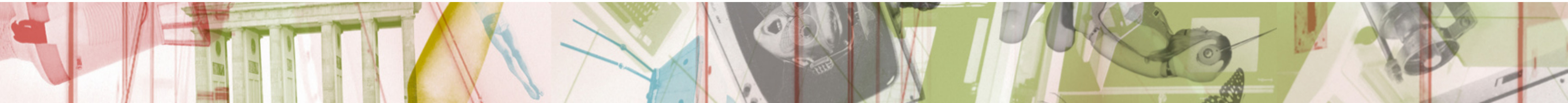

**Band 1**

**Band 2**

**Band 3**

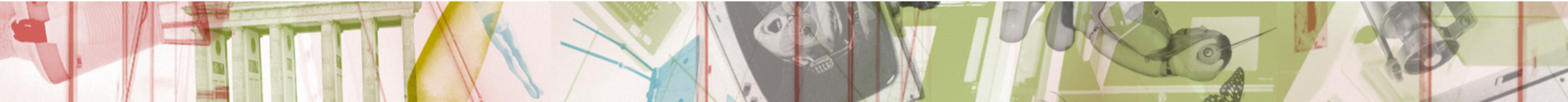
**Band 4**

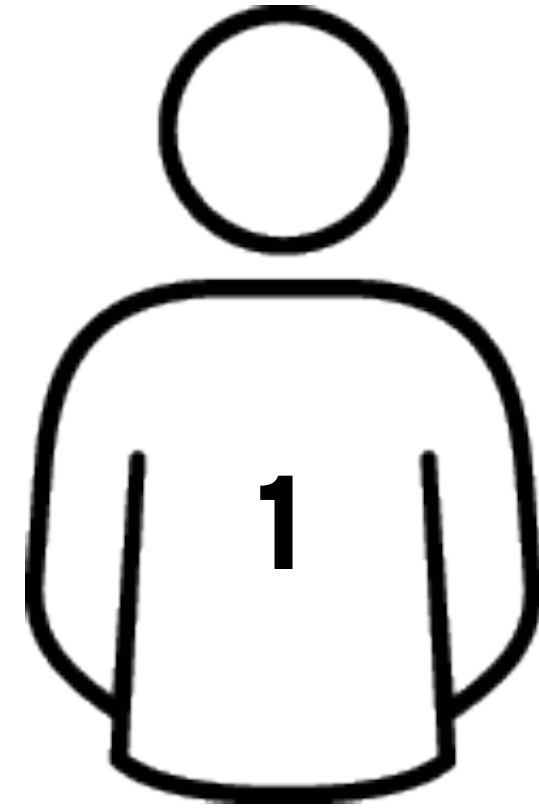
**Band 5**



# Things we didn't have time for



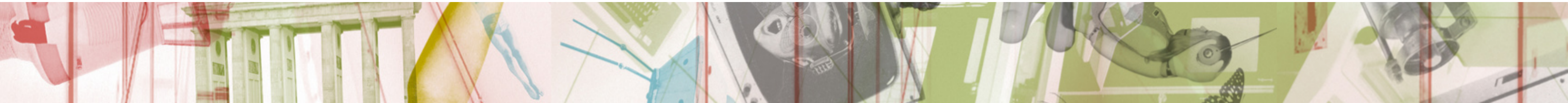




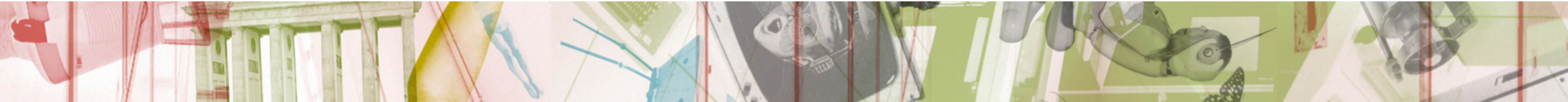
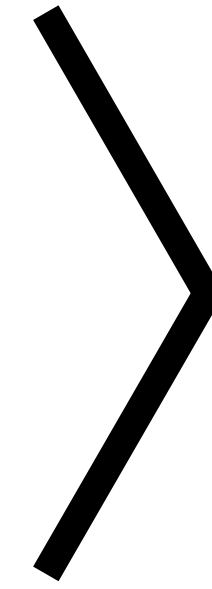
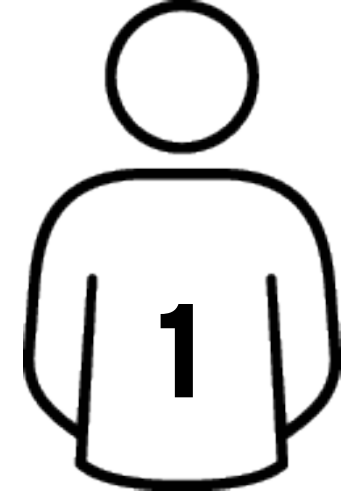
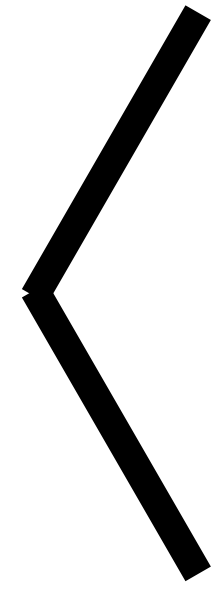
~~A 1 play~~

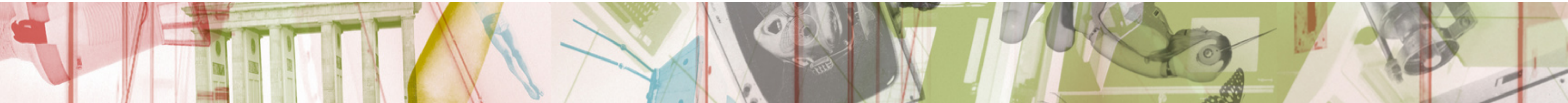
B 10 plays

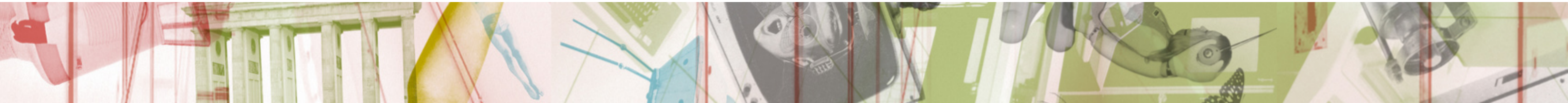
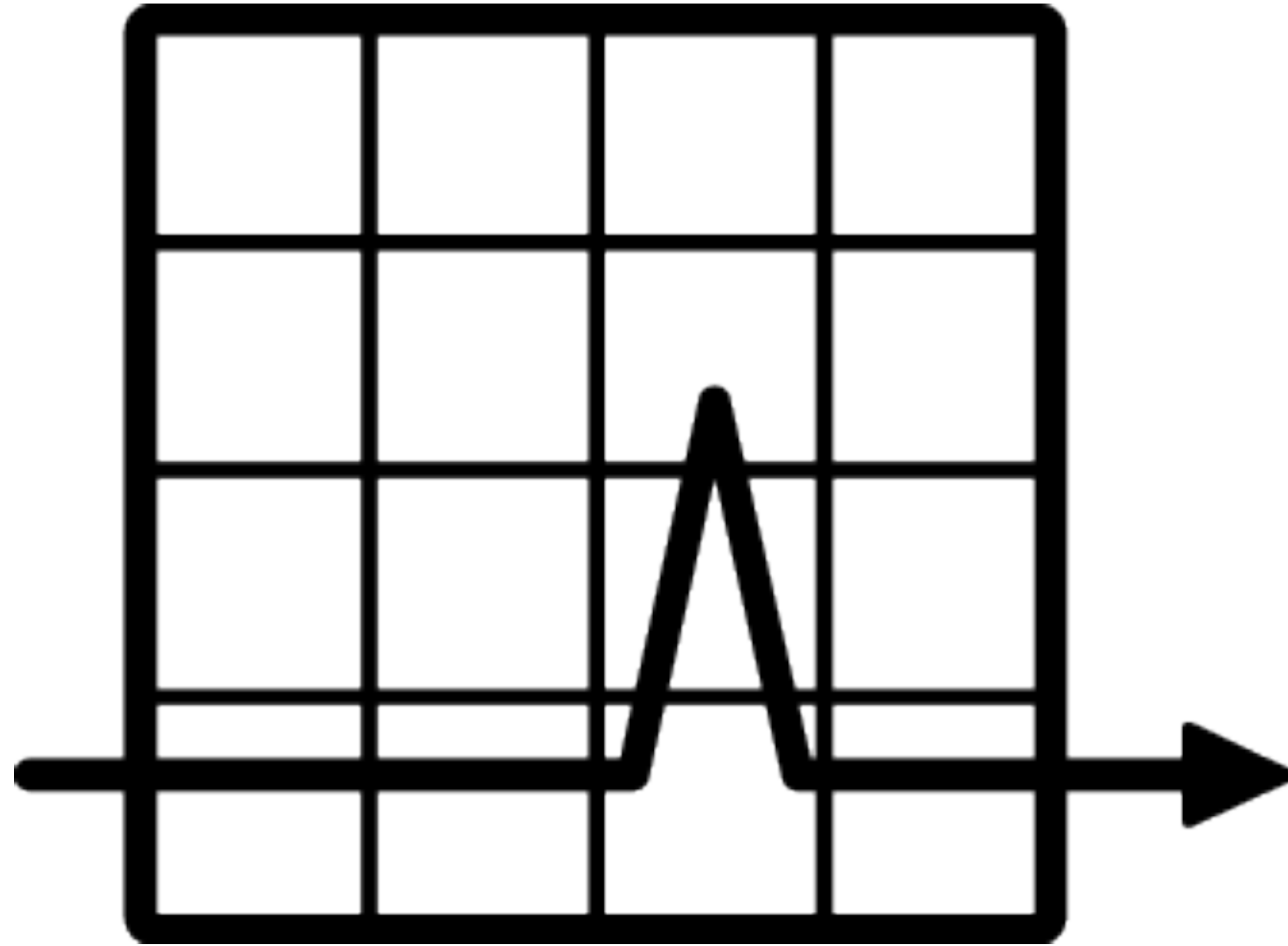
C 100 plays

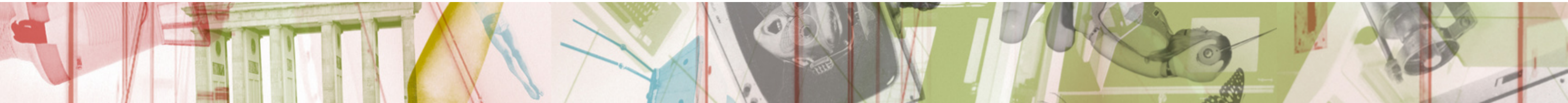












# Things we have talked about



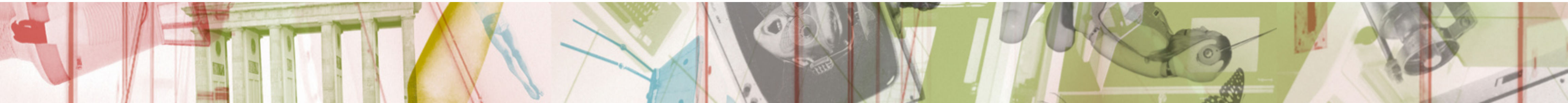


6
1,987
91
93
43

Min-hash signature

**sophie@redhat.com**

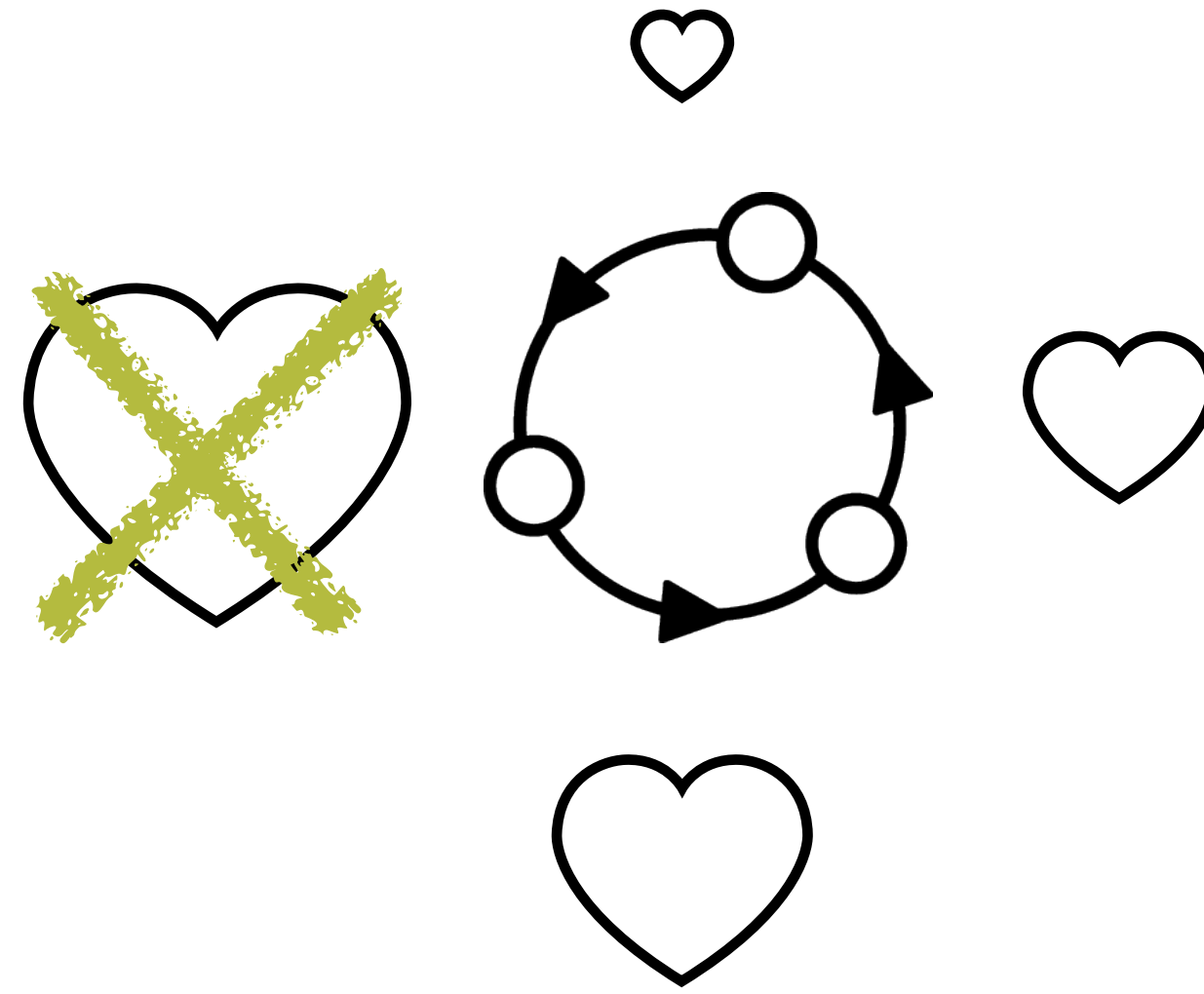
**@sophwats**



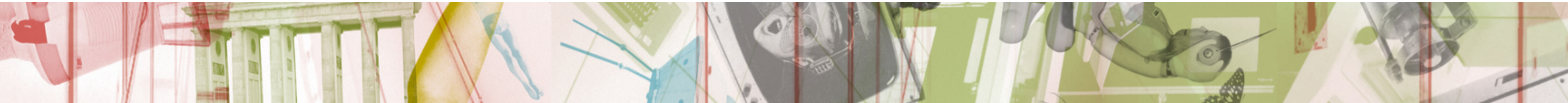


6
1,987
91
93
43

Min-hash signature



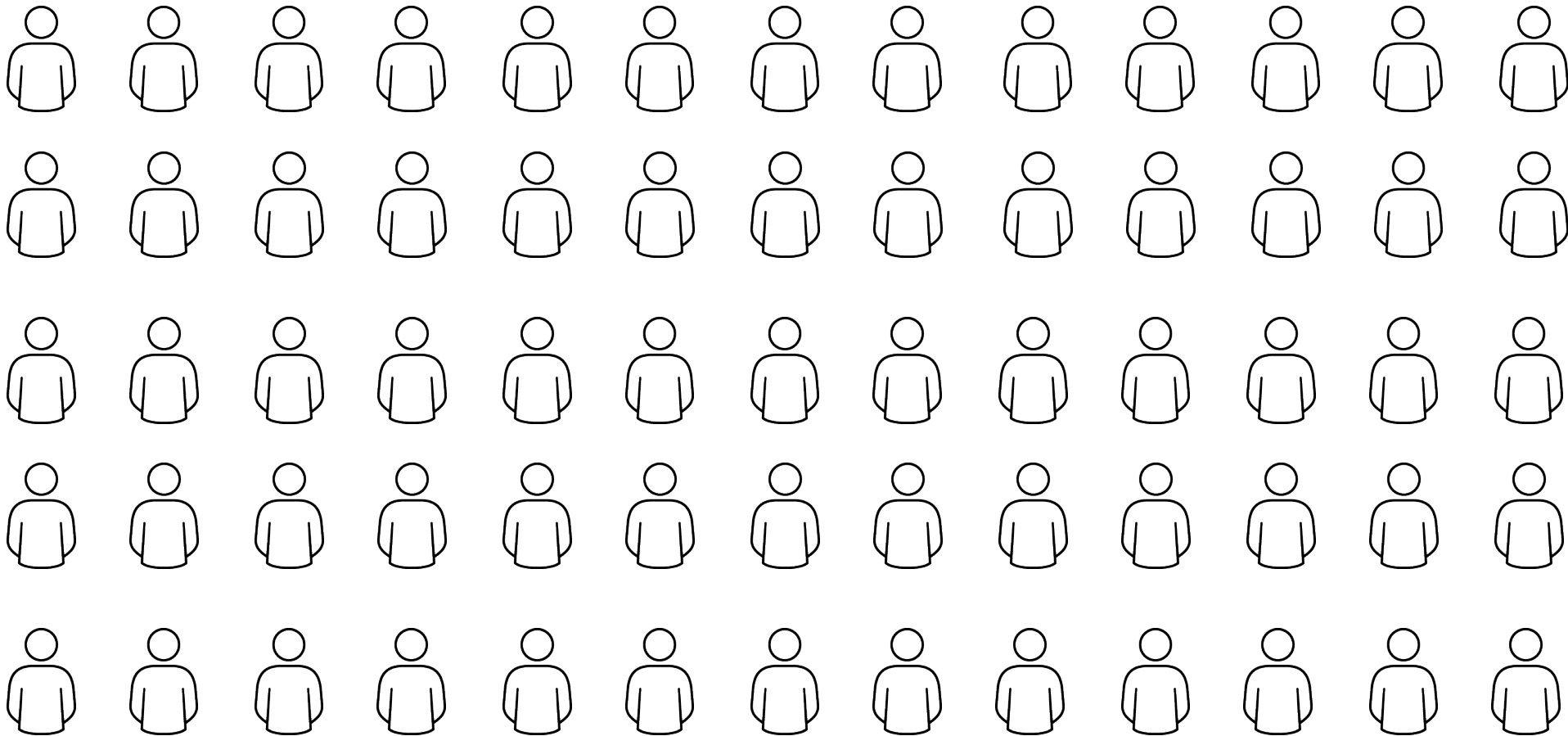
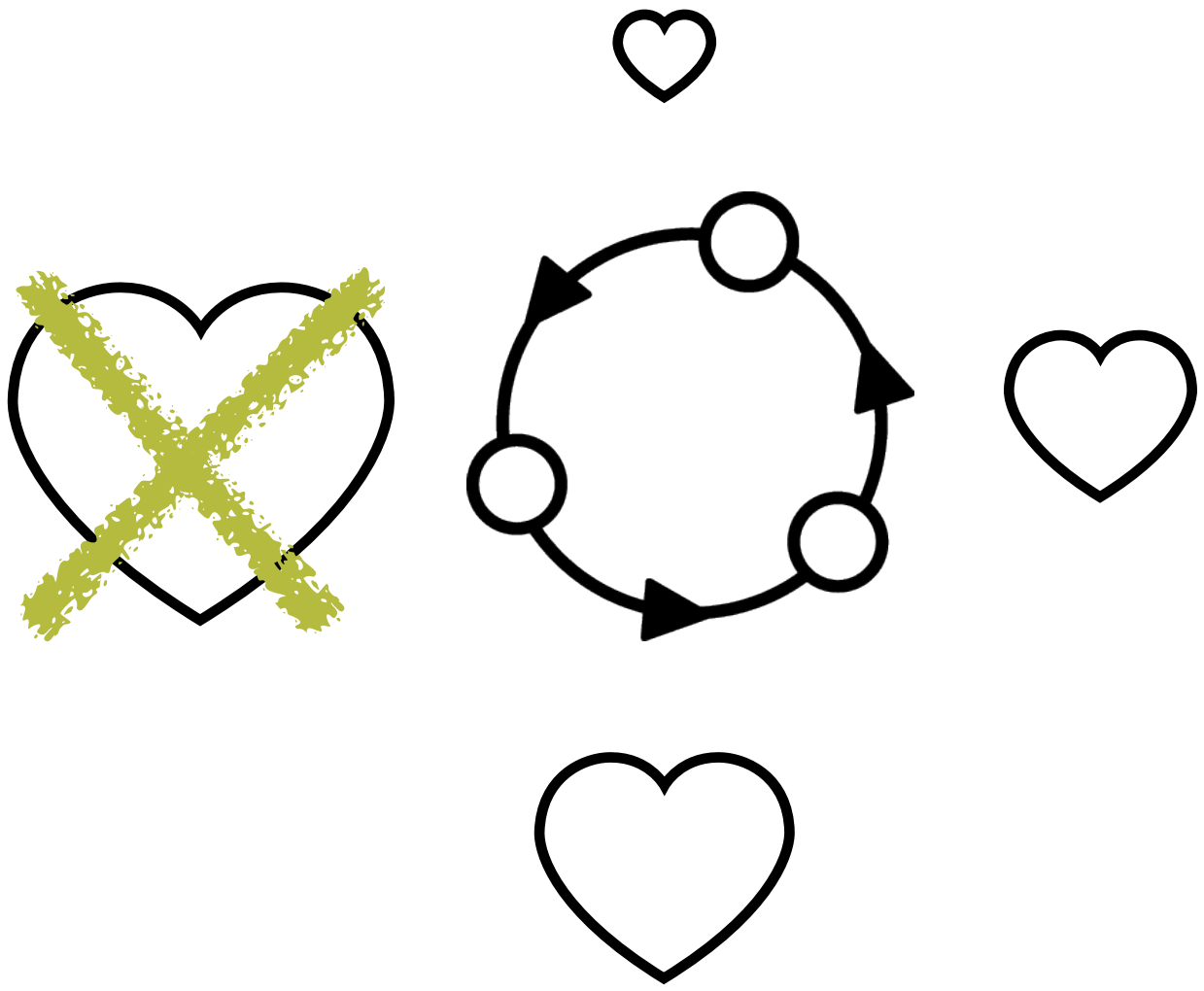
**sophie@redhat.com**  
**@sophwats**





6
1,987
91
93
43

Min-hash signature



**sophie@redhat.com**  
**@sophwats**

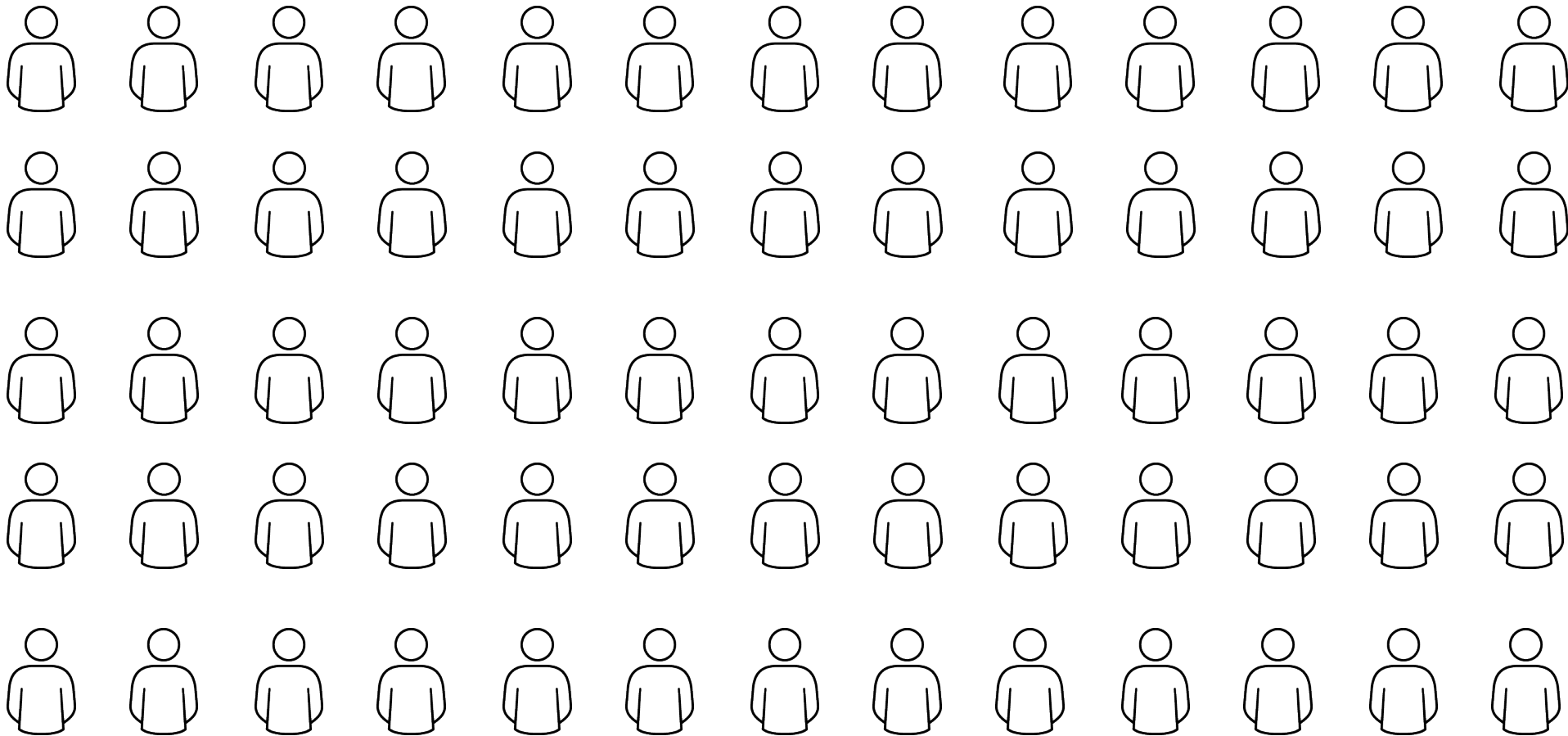
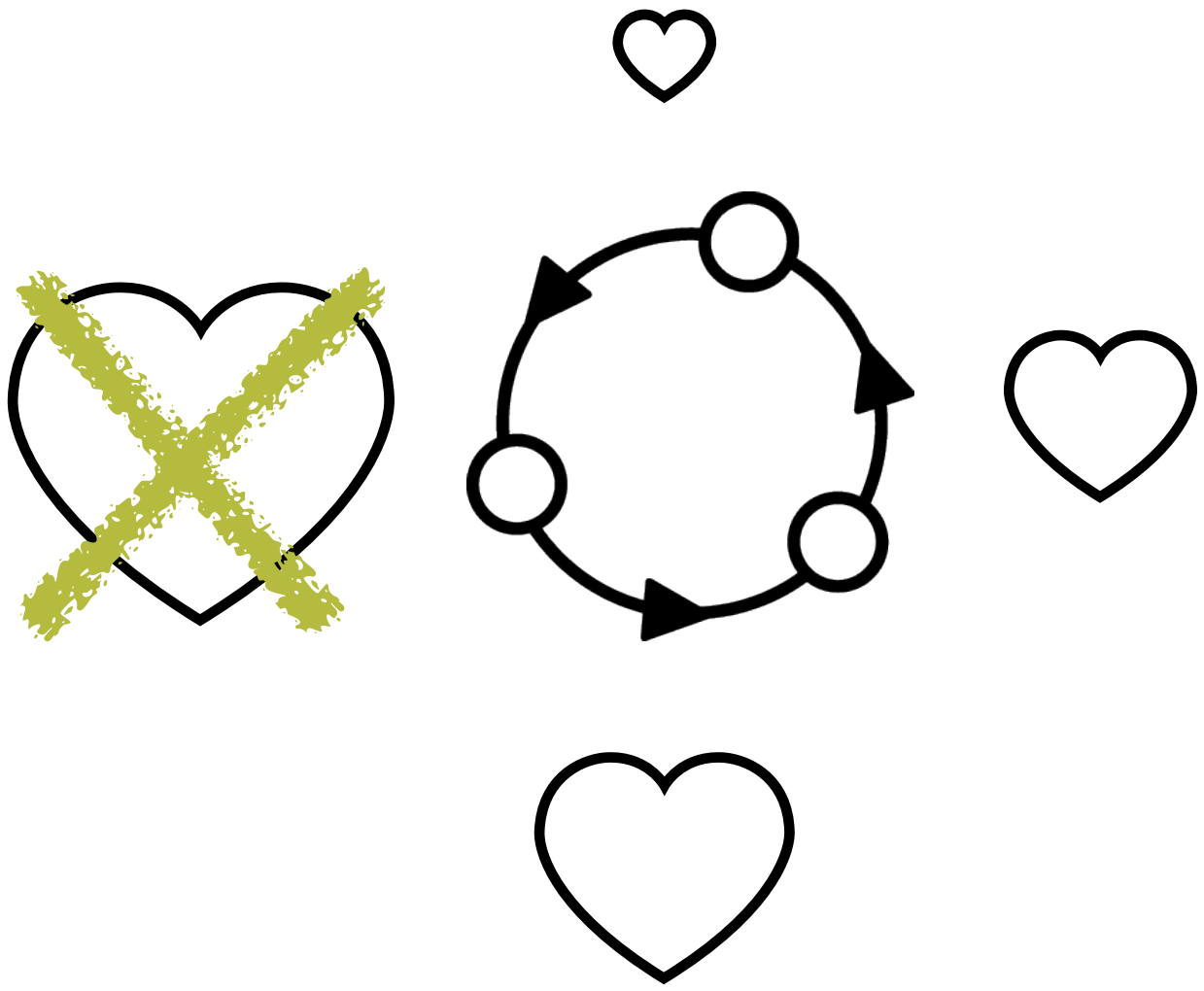






6
1,987
91
93
43

Min-hash signature



**sophie@redhat.com**  
**@sophwats**

