Feature allocations, probability functions, and paintboxes

Tamara Broderick
UC Berkeley
(MIT starting 2015)

Clustering/Partition

$$\pi_7 = \{\{1, 2, 7\}, \{3, 5\}, \{4\}, \{6\}\}$$

Clustering/Partition

$$\pi_7 = \{\{1, 2, 7\}, \{3, 5\}, \{4\}, \{6\}\}$$

Document 1

Document 2

Document 3

Document 4

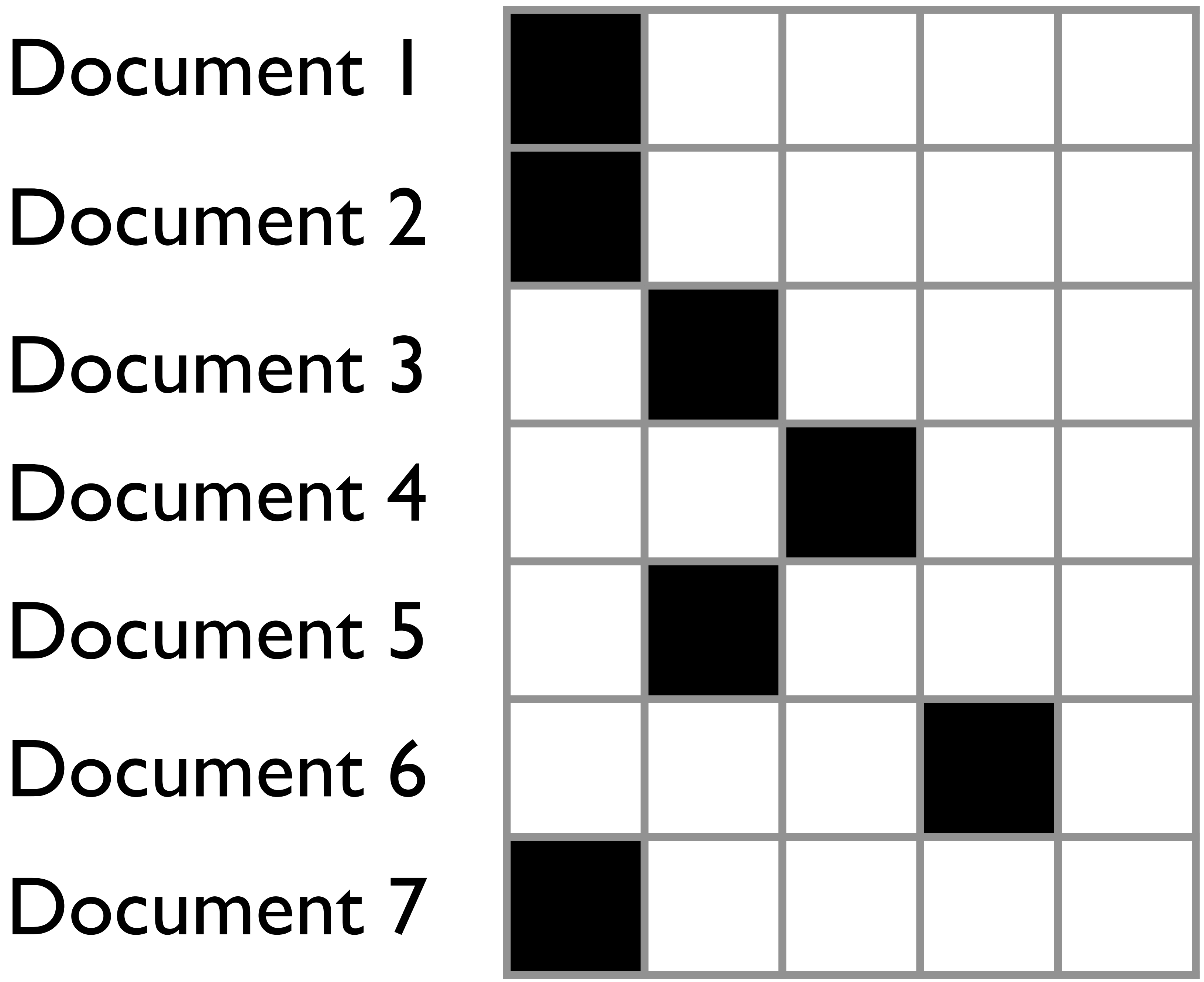
Document 5

Document 6

Document 7

Clustering/Partition

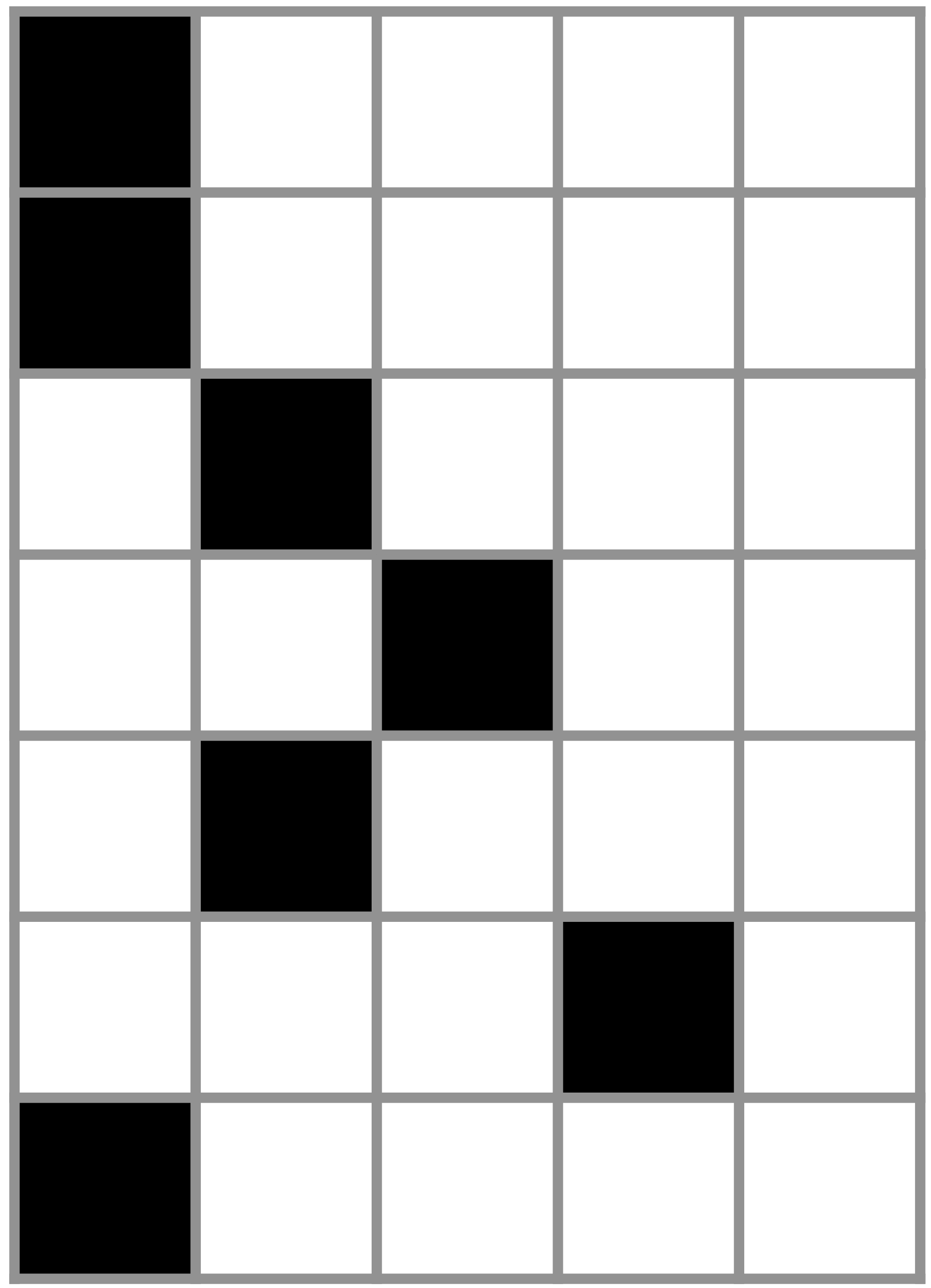
$$\pi_7 = \{\{1,2,7\},\{3,5\},\{4\},\{6\}\}$$



Clustering/Partition

$$\pi_7 = \{\{1, 2, 7\}, \{3, 5\}, \{4\}, \{6\}\}$$

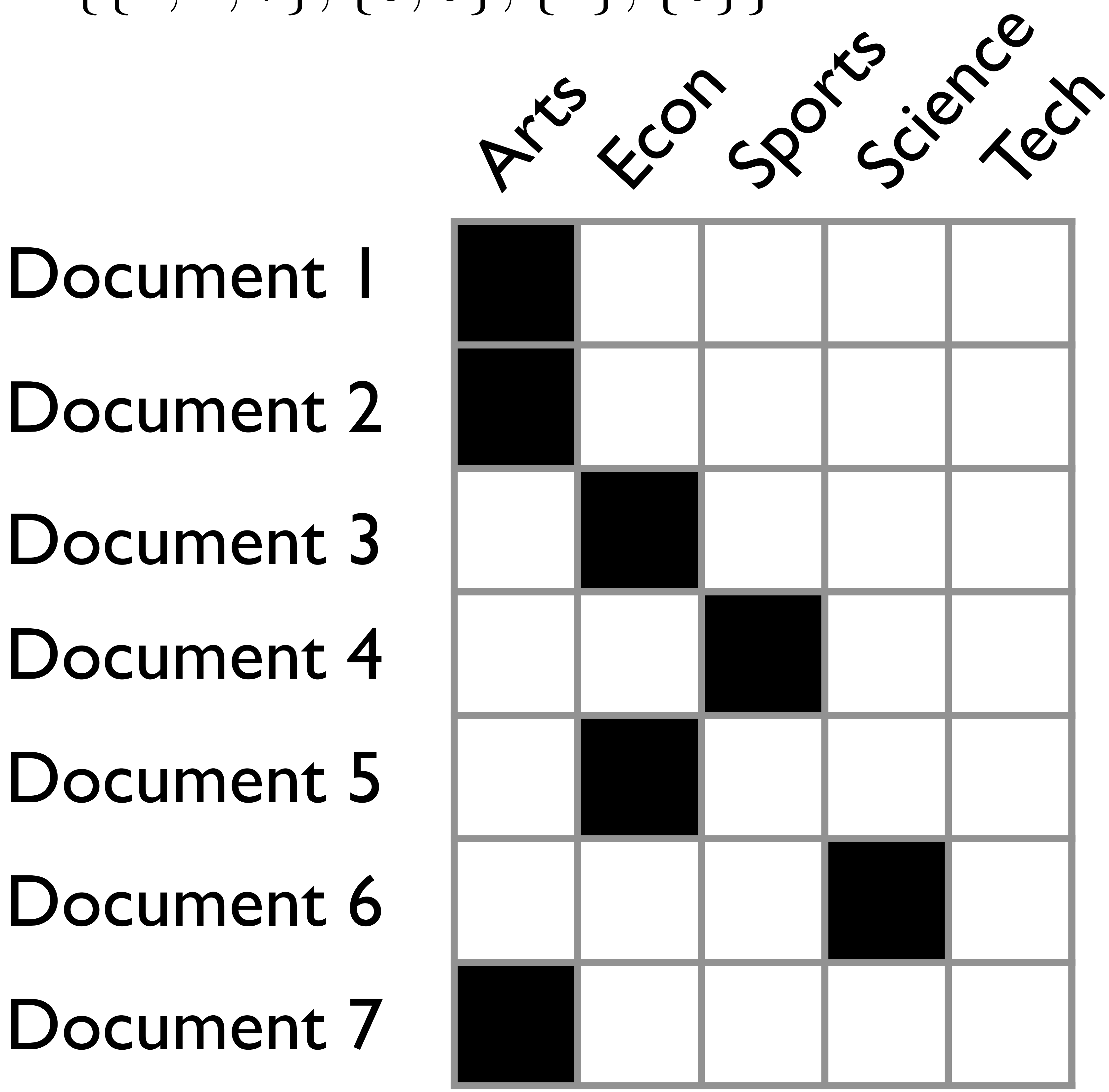
Document 1
Document 2
Document 3
Document 4
Document 5
Document 6
Document 7



"clusters", "blocks (of a partition)"

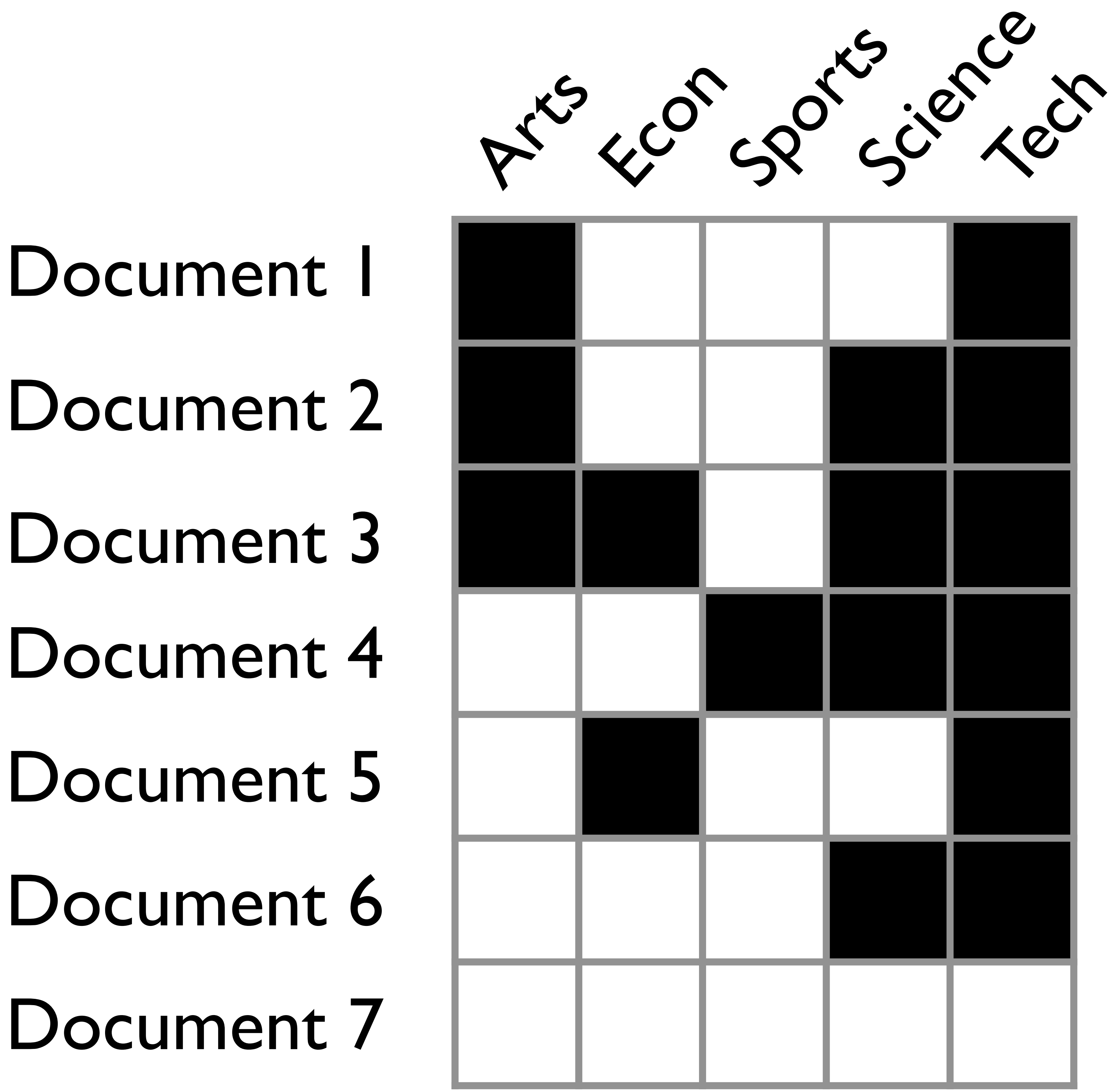
Clustering/Partition

$\pi_7 = \{\{1,2,7\},\{3,5\},\{4\},\{6\}\}$

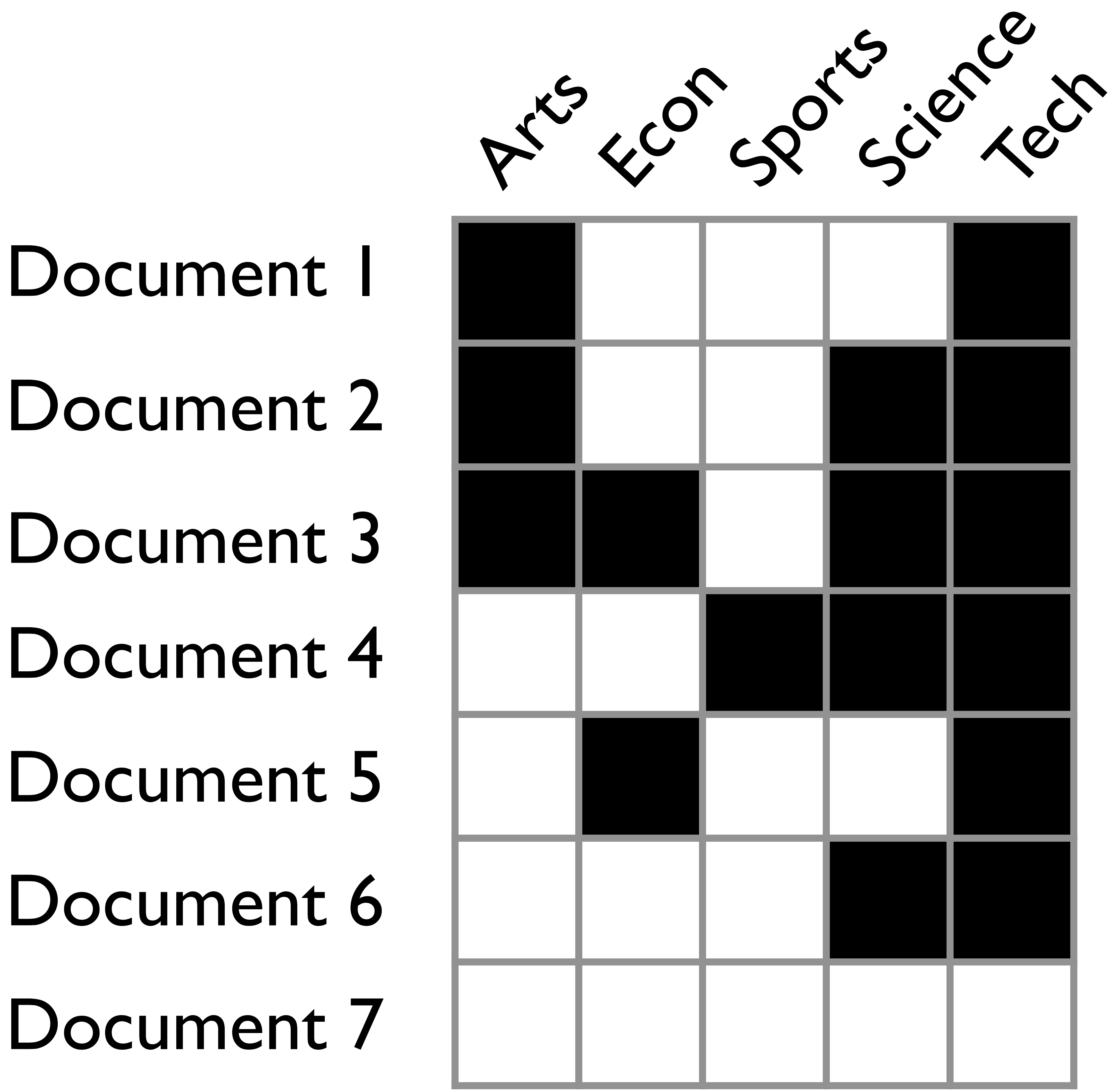


“clusters”,
“blocks (of a partition)”

Latent feature allocation



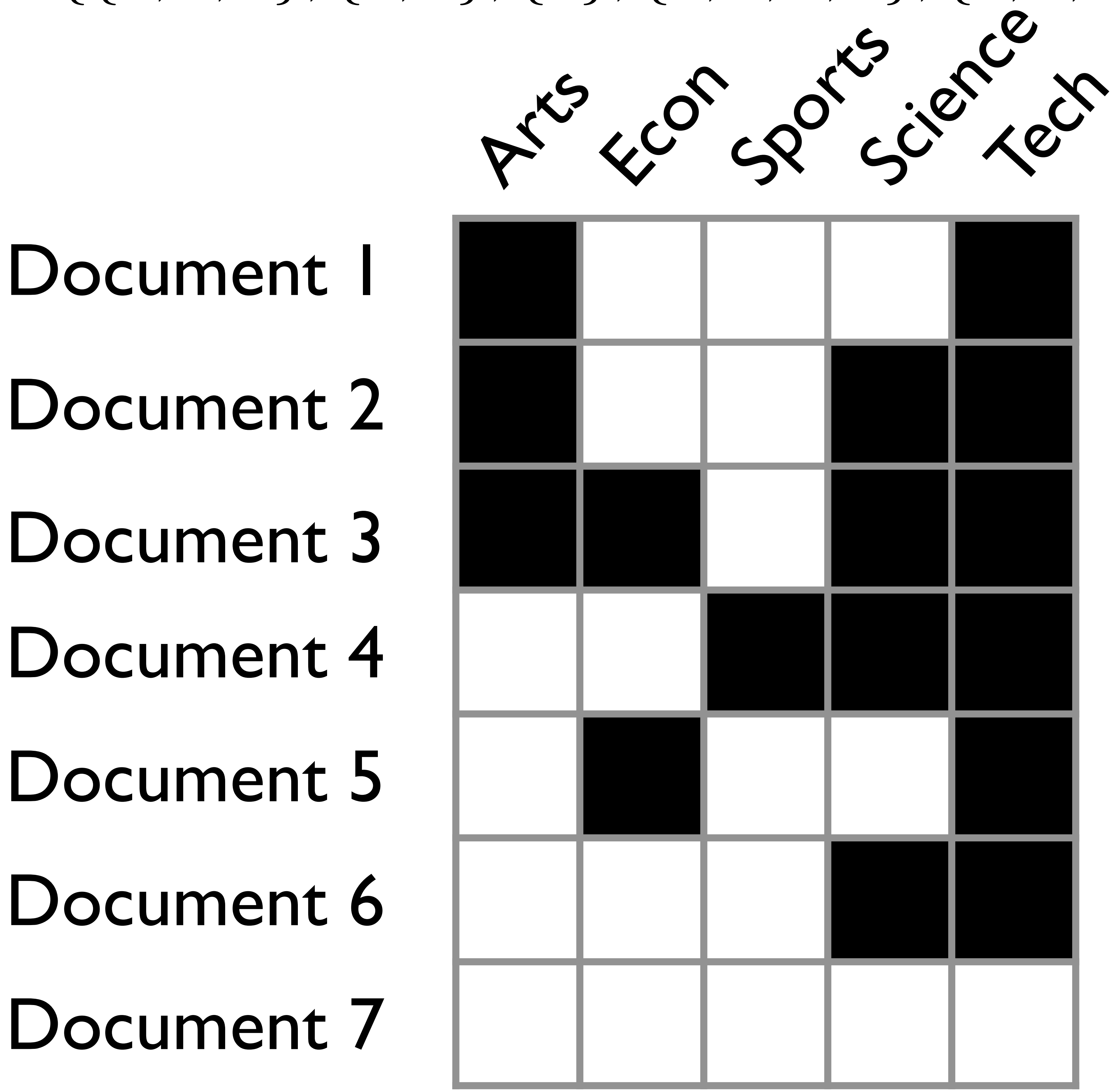
Latent feature allocation



“features”,
“topics”

Latent feature allocation

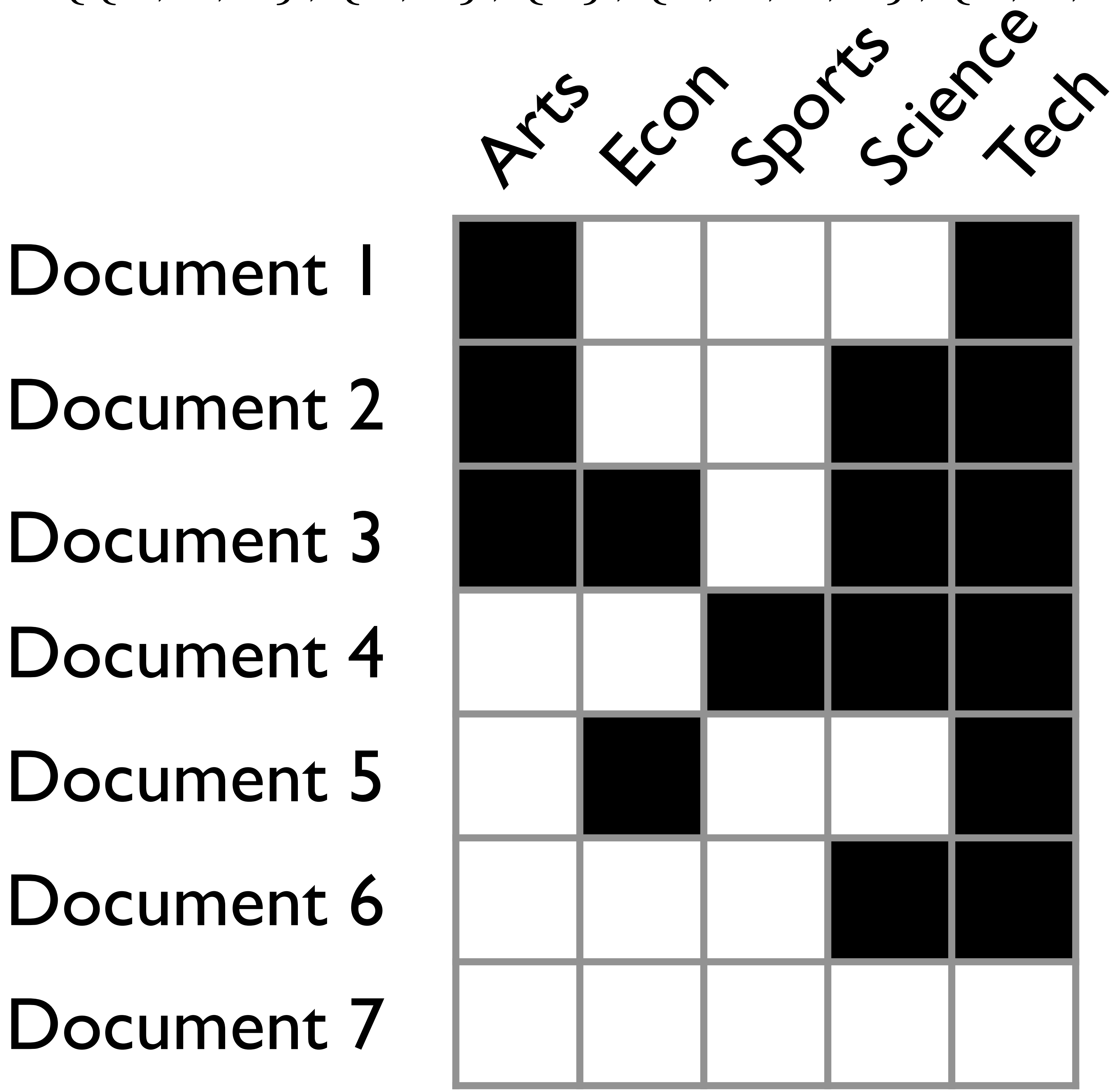
$$f_7 = \{\{1, 2, 3\}, \{3, 5\}, \{4\}, \{2, 3, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$$



"features",
"topics"

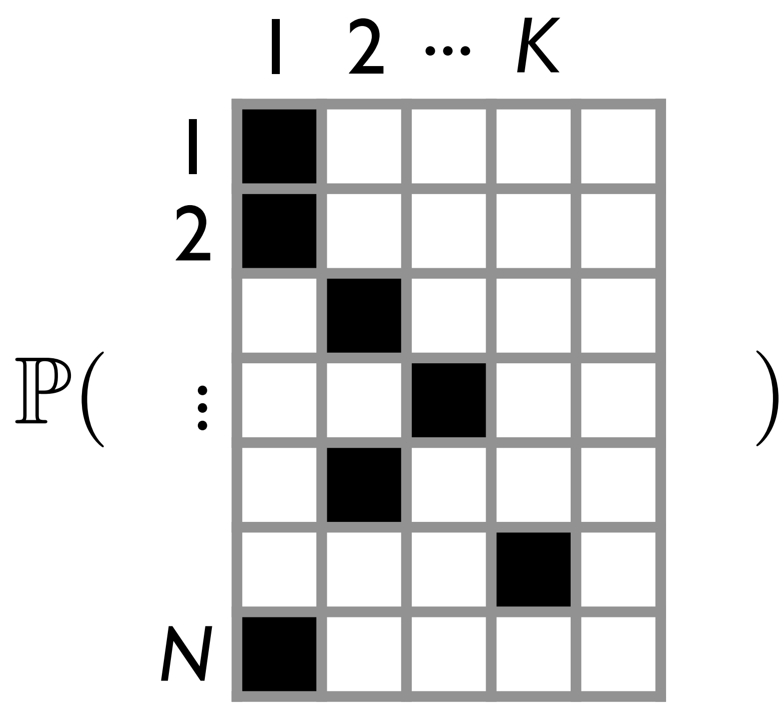
Latent feature allocation

$f_7 = \{\{1,2,3\},\{3,5\},\{4\},\{2,3,4,6\},\{1,2,3,4,5,6\}\}$



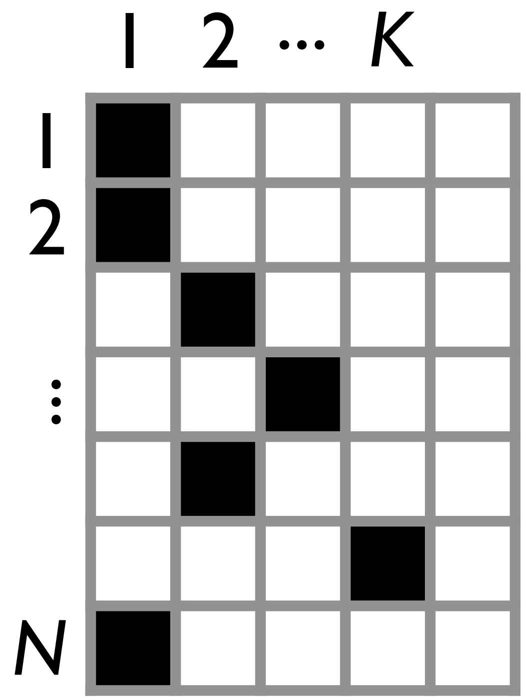
<u>Question</u>:
Can we characterize exchangeable feature distributions?

Exchangeable probability functions



Exchangeable probability functions

$\mathbb{P}($

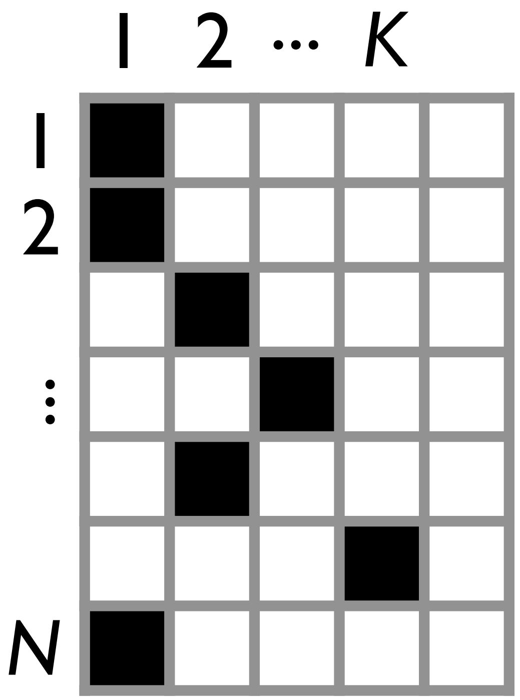


$) = p\left(S_{N,1}, \ldots, S_{N,K}\right)$

[Pitman 1995]

Exchangeable probability functions

Size of *K*th cluster

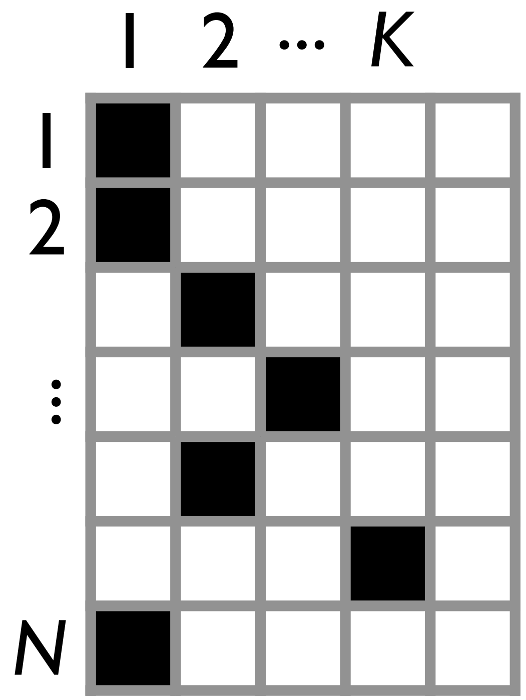
$\mathbb{P}($ 

$) = p\left(S_{N,1}, \ldots, S_{N,K}\right)$

[Pitman 1995]

Exchangeable probability functions

Exchangeable partition probability function (EPPF)



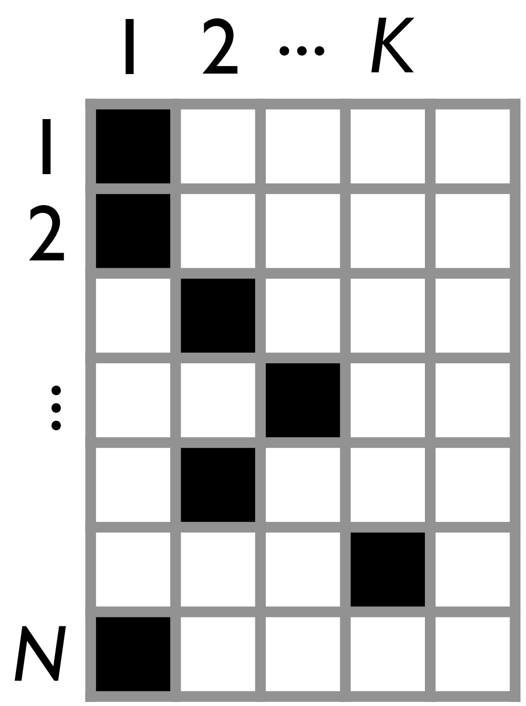
$$\mathbb{P}(\quad [\text{figure}] \quad) = p\left(S_{N,1}, \ldots, S_{N,K}\right)$$

[Pitman 1995]

Exchangeable probability functions

Exchangeable partition probability function (EPPF)

$\mathbb{P}($



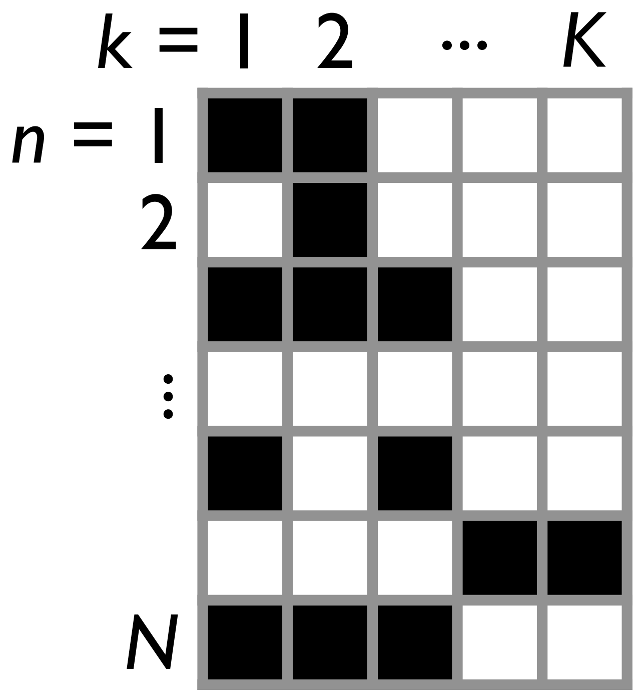
$) = p\left(S_{N,1}, \ldots, S_{N,K}\right)$

"Exchangeable **feature** probability function" (EFPF)?

Example: Indian buffet process

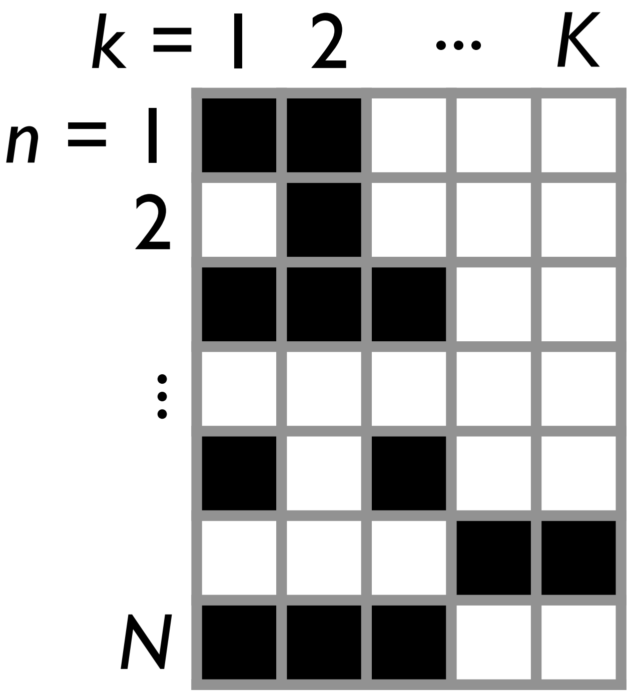
[Griffiths, Ghahramani 2006]

Example: Indian buffet process



[Griffiths, Ghahramani 2006]

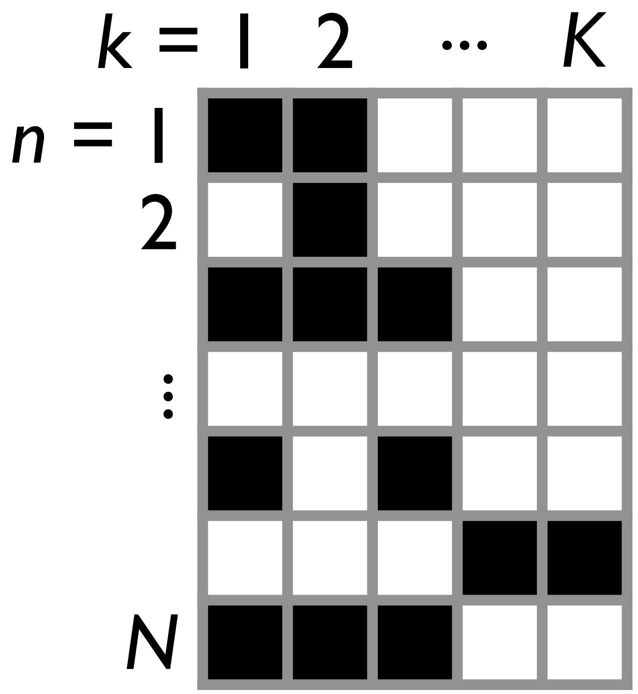
Example: Indian buffet process



For n = 1, 2, ..., N

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

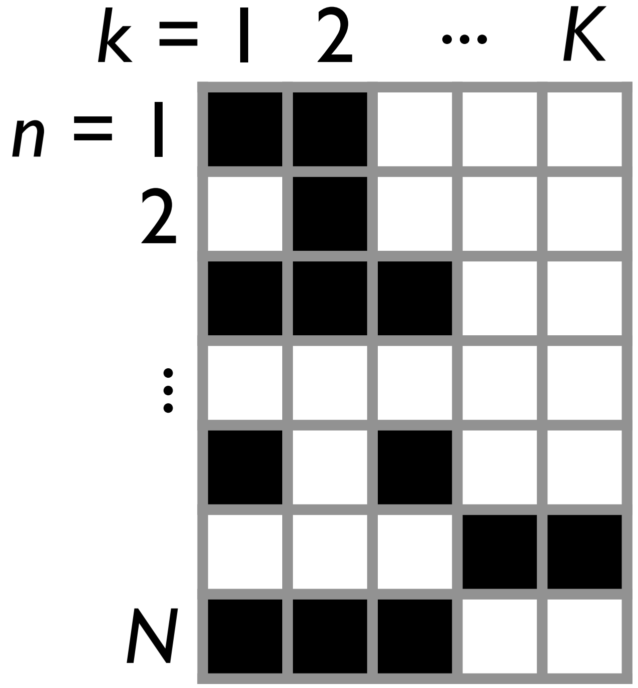


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



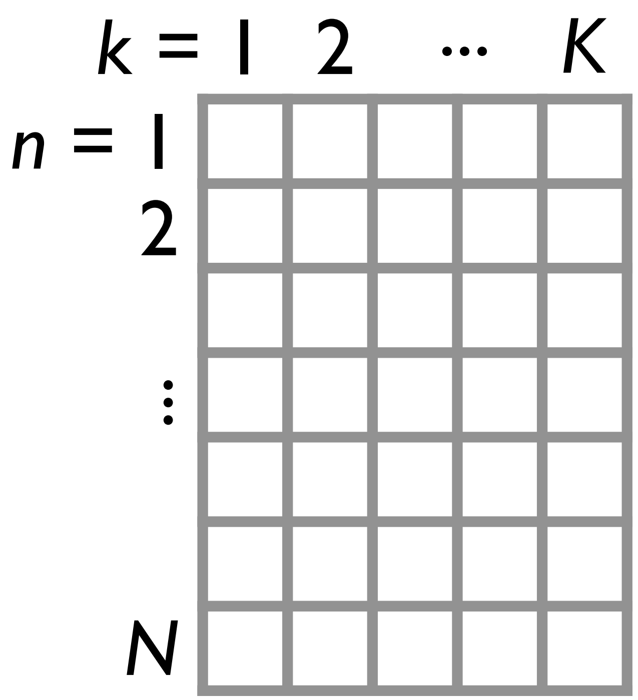
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



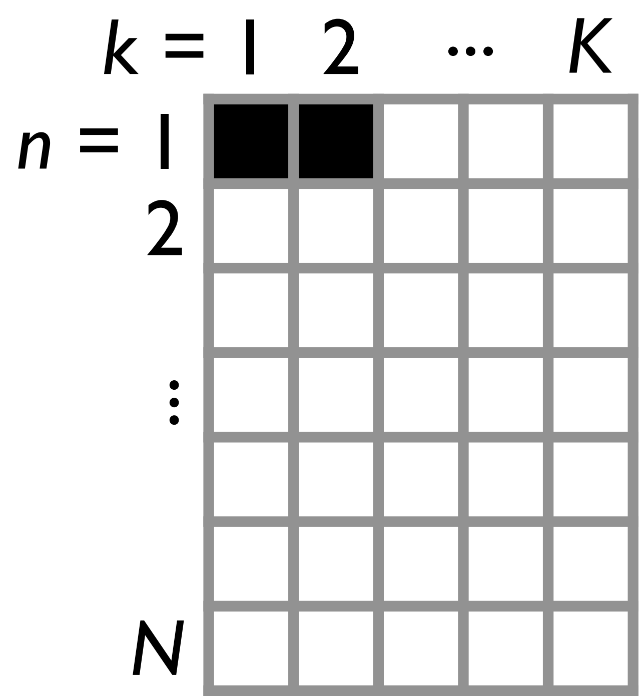
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

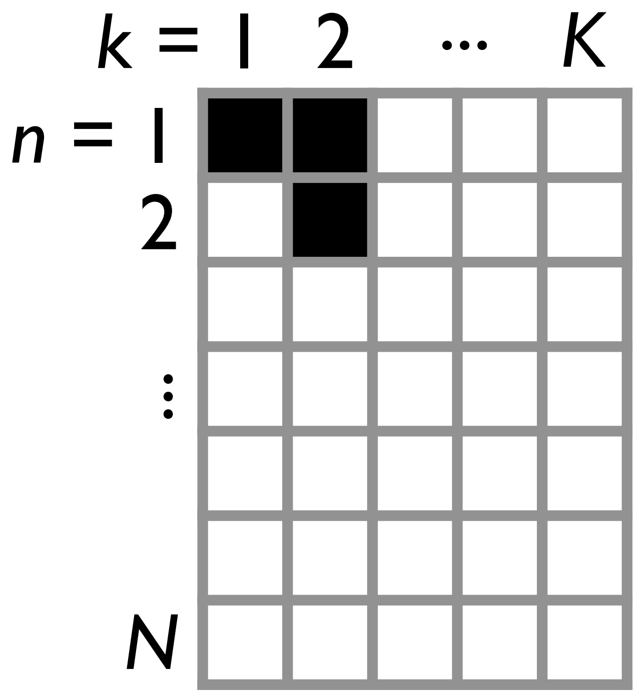


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



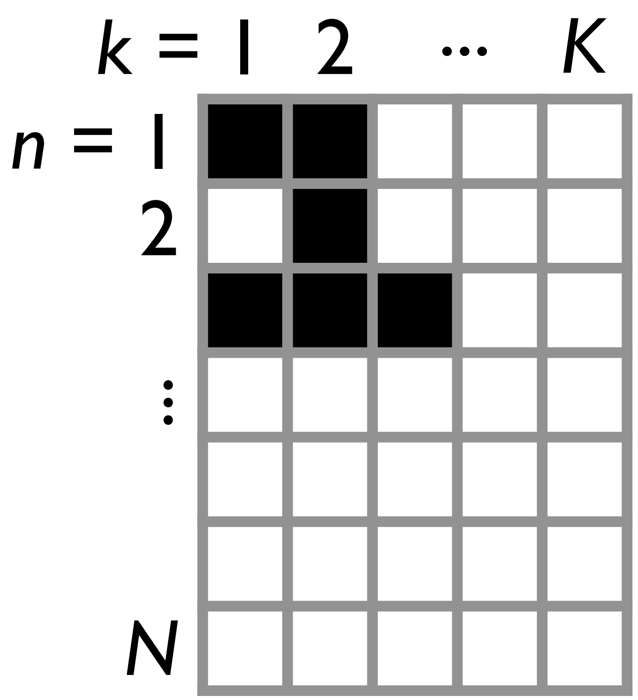
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



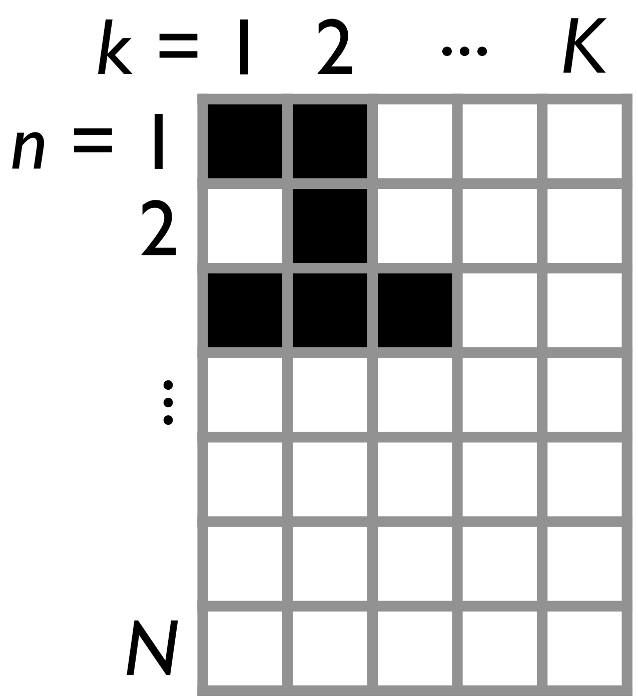
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



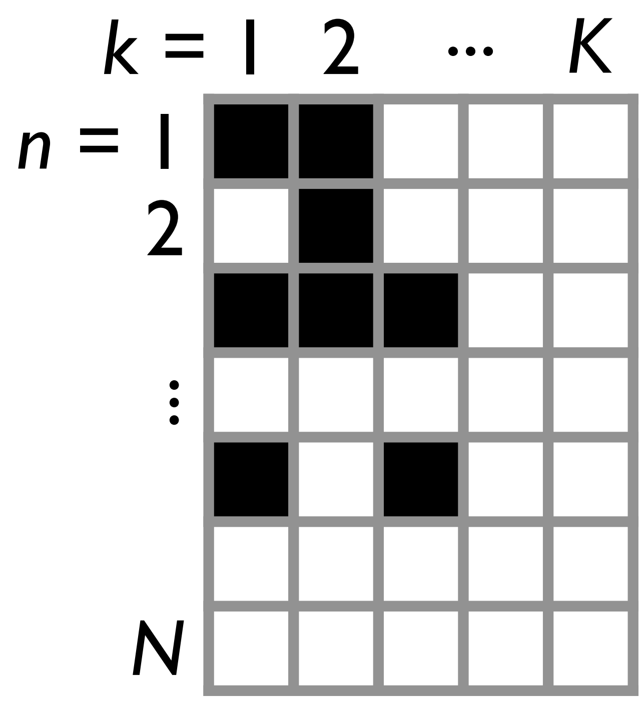
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

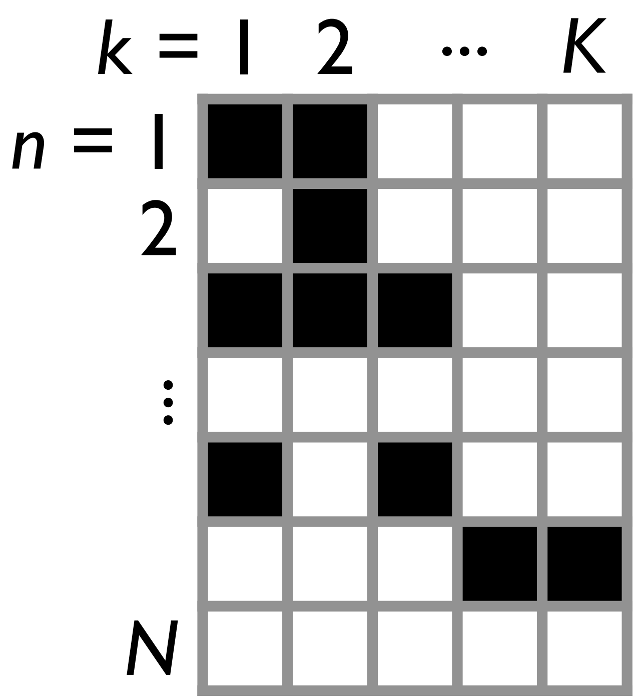


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



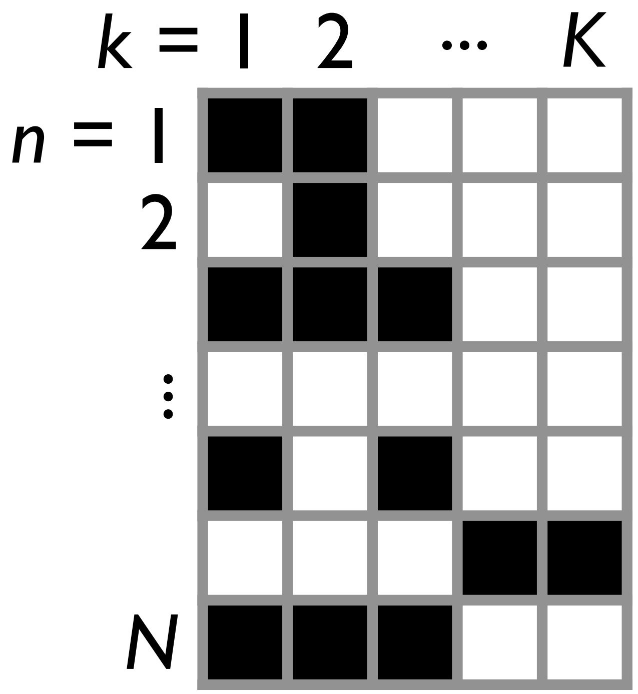
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

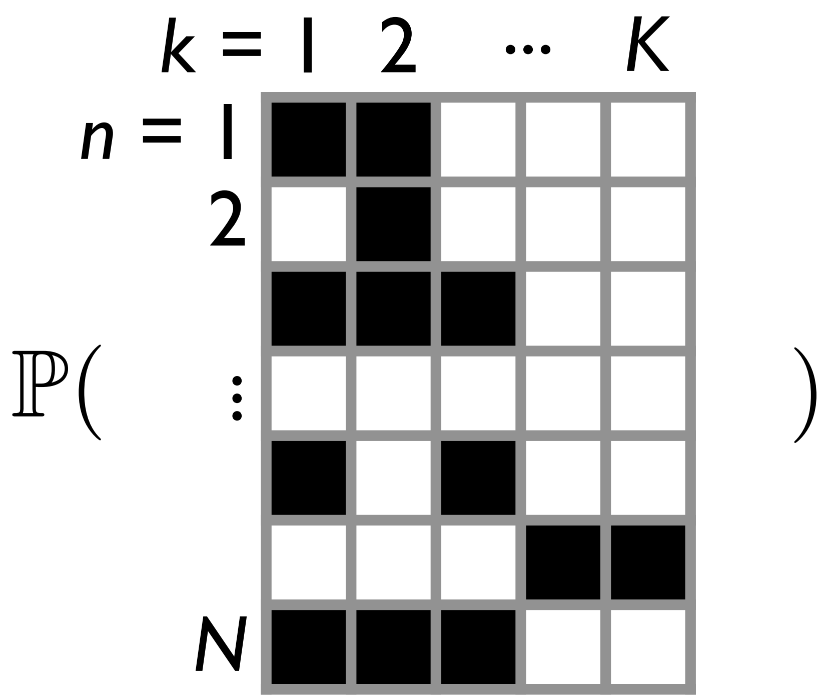
Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

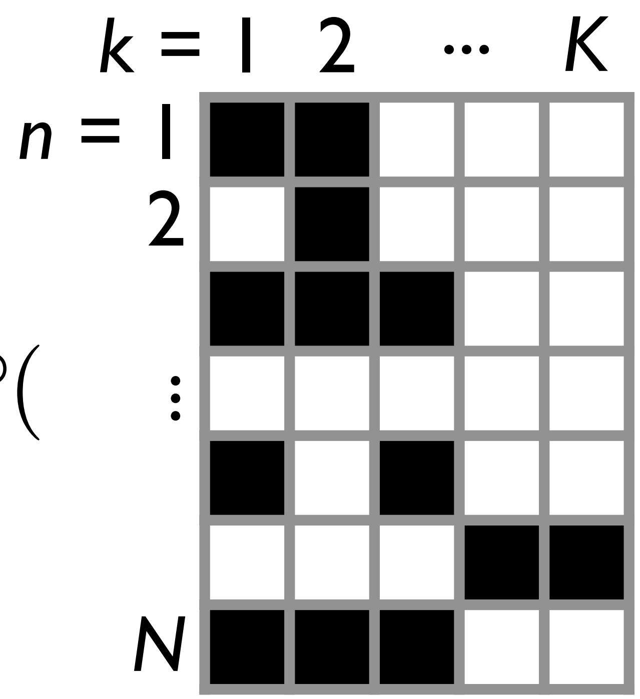


Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($



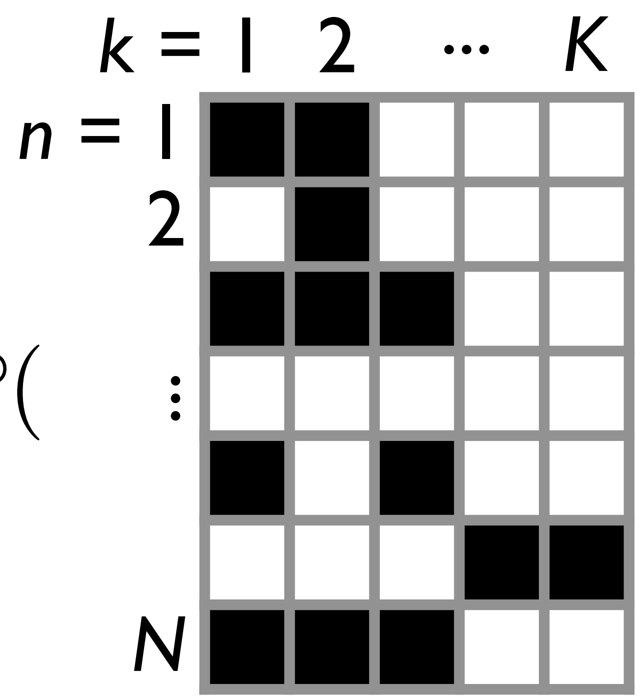
$)$

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($  $)$

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

Size of *k*th feature

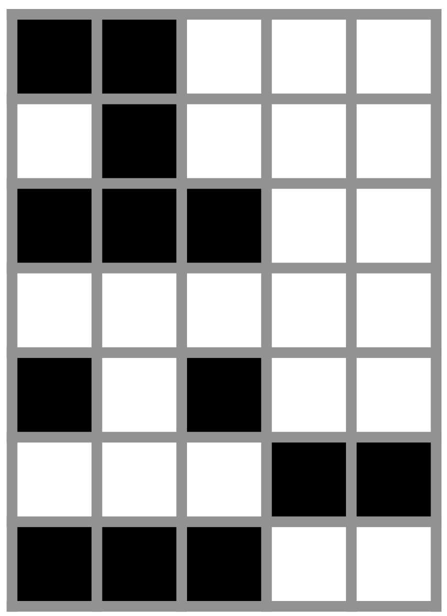
Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

k = 1 2 ... K

n = 1, 2, ⋮, N

$\mathbb{P}($  $)$

Number of features

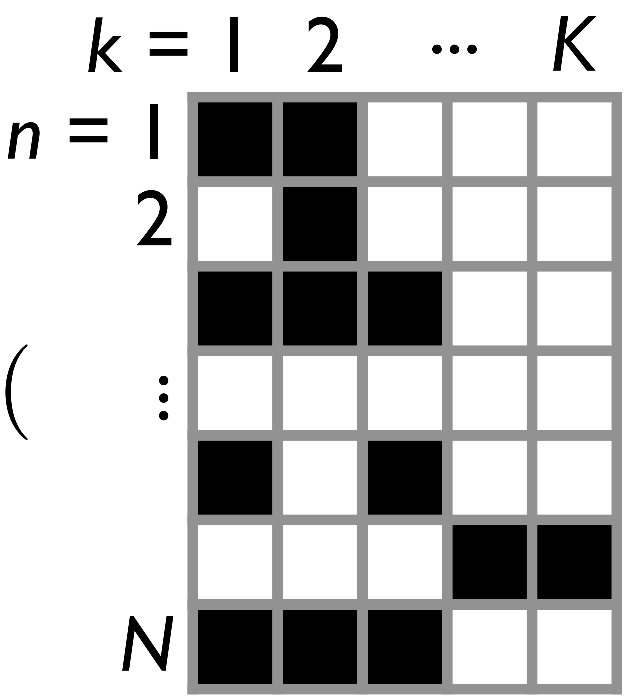
Size of kth feature

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N+\theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($  $)$

Number of data points

Size of kth feature

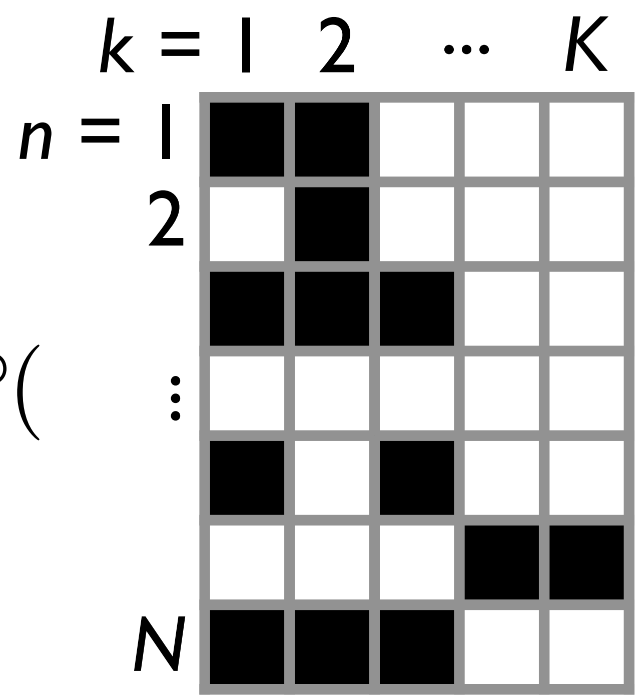
Number of features

$$= \frac{1}{K_N!} (\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($ 

$)$

Number of data points

Size of *k*th feature

Number of features

$$= \frac{1}{K_N!} (\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

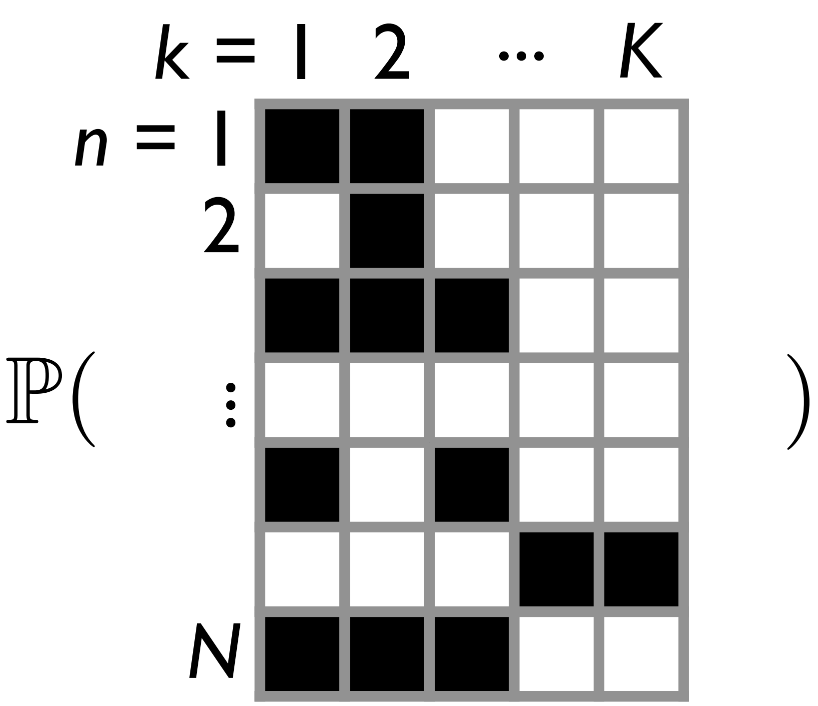
$$= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)



Number of data points

Size of *k*th feature

Number of features

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma\sum_{n=1}^{N}(\theta+n-1)^{-1}\right)\prod_{k=1}^{K_N}\frac{\Gamma(S_{N,k})\Gamma(N-S_{N,k}+\theta)}{\Gamma(N+\theta)}$$

$$= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

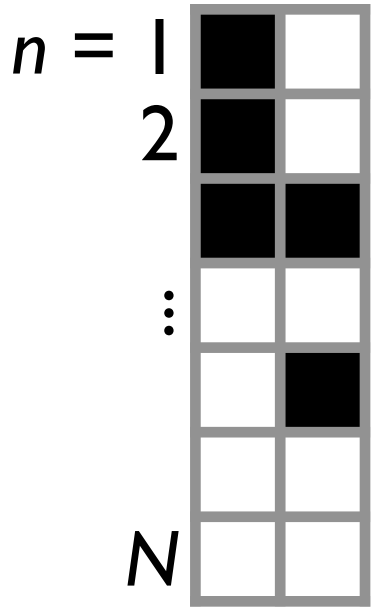
"EFPF"

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Counterexample

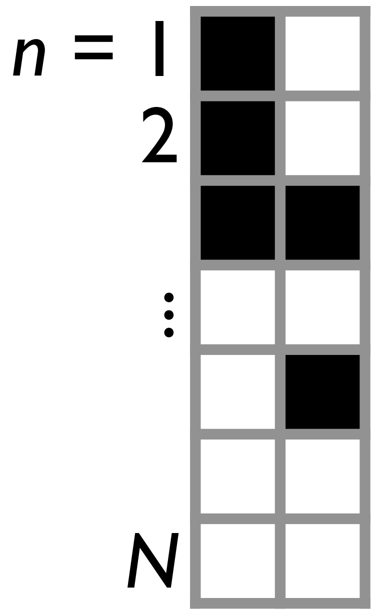


[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

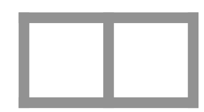
Counterexample



$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

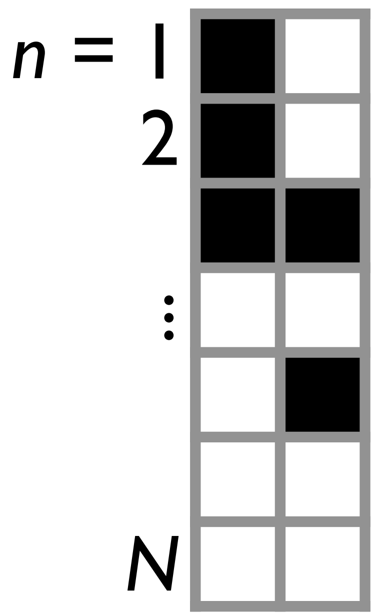
$\mathbb{P}(\text{row} =$  $) = p_4$

[Broderick, Jordan, Pitman 2013]

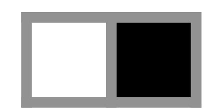
Exchangeable probability functions

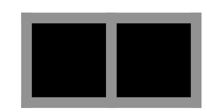
"Exchangeable feature probability function" (EFPF)?

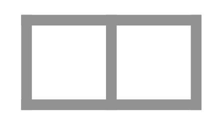
Counterexample

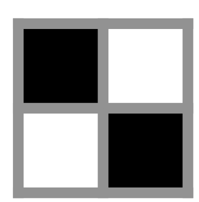


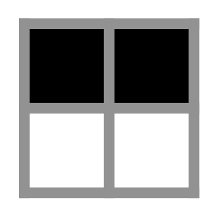
$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$



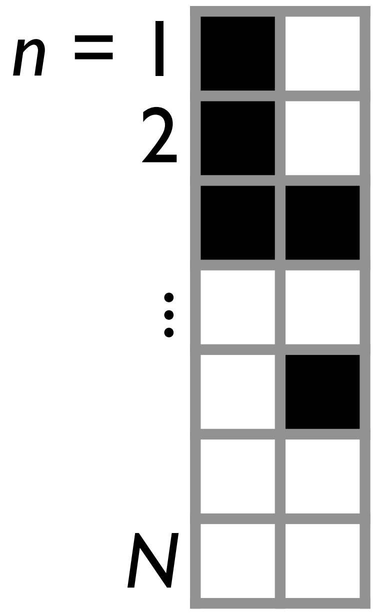


[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

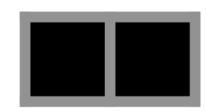
"Exchangeable feature probability function" (EFPF)?

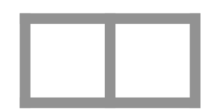
Counterexample

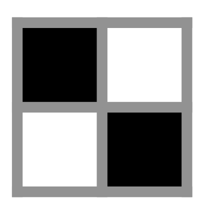
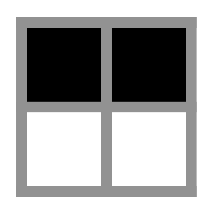


$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$

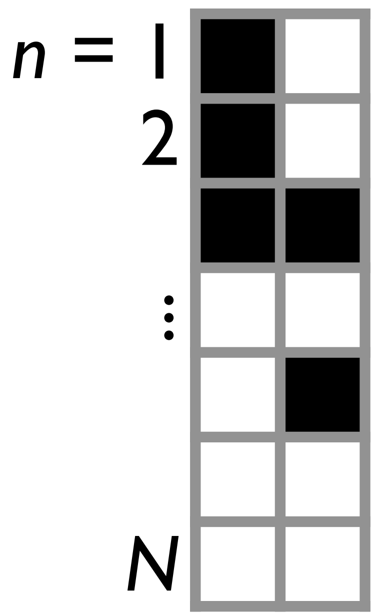
$\mathbb{P}($  $)$ $\mathbb{P}($  $)$

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

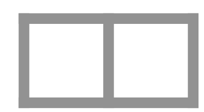
Counterexample

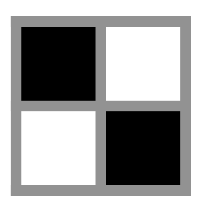
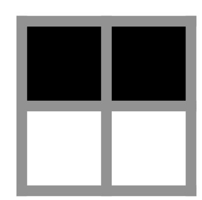


$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$

$\mathbb{P}($  $)$ $\quad$ $\mathbb{P}($  $)$

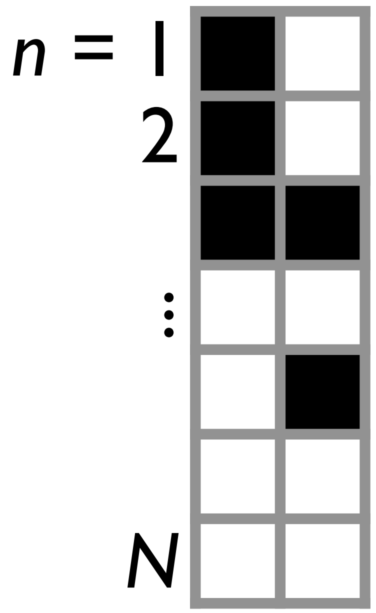
$p_1 p_2$ $\quad$ $p_3 p_4$

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

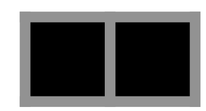
"Exchangeable feature probability function" (EFPF)?

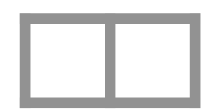
Counterexample

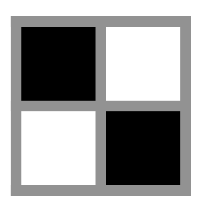
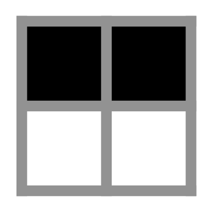


$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

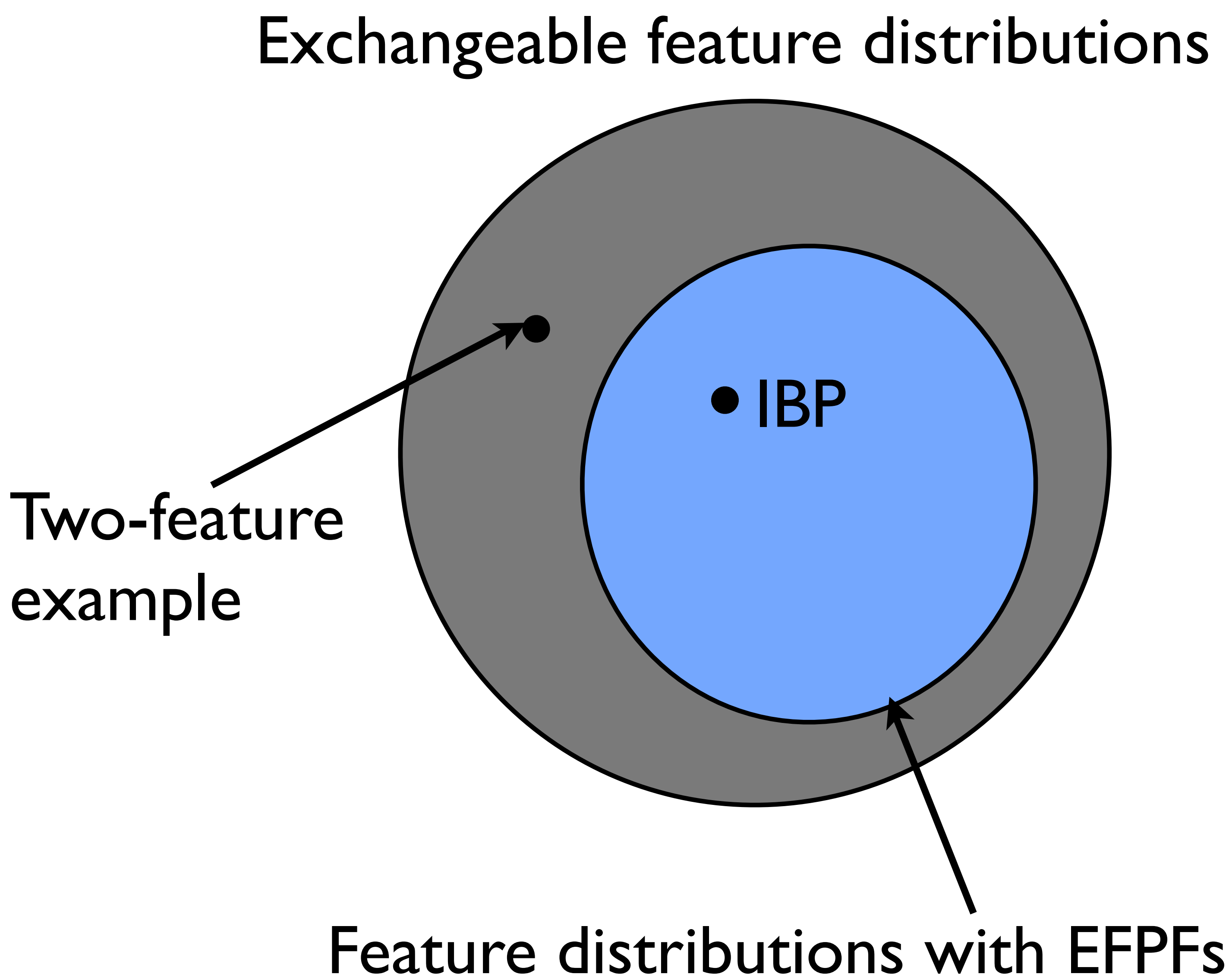
$\mathbb{P}(\text{row} =$  $) = p_4$

$\mathbb{P}($  $) \neq \mathbb{P}($  $)$

$$p_1 p_2 \neq p_3 p_4$$

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions



[Broderick, Jordan, Pitman 2013]

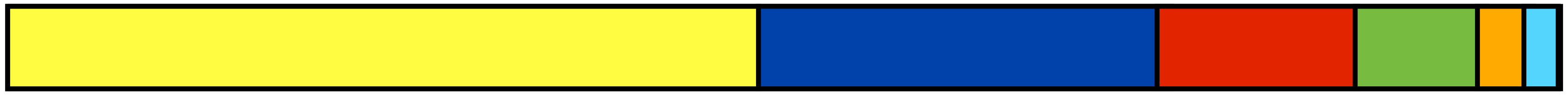
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]

Paintboxes

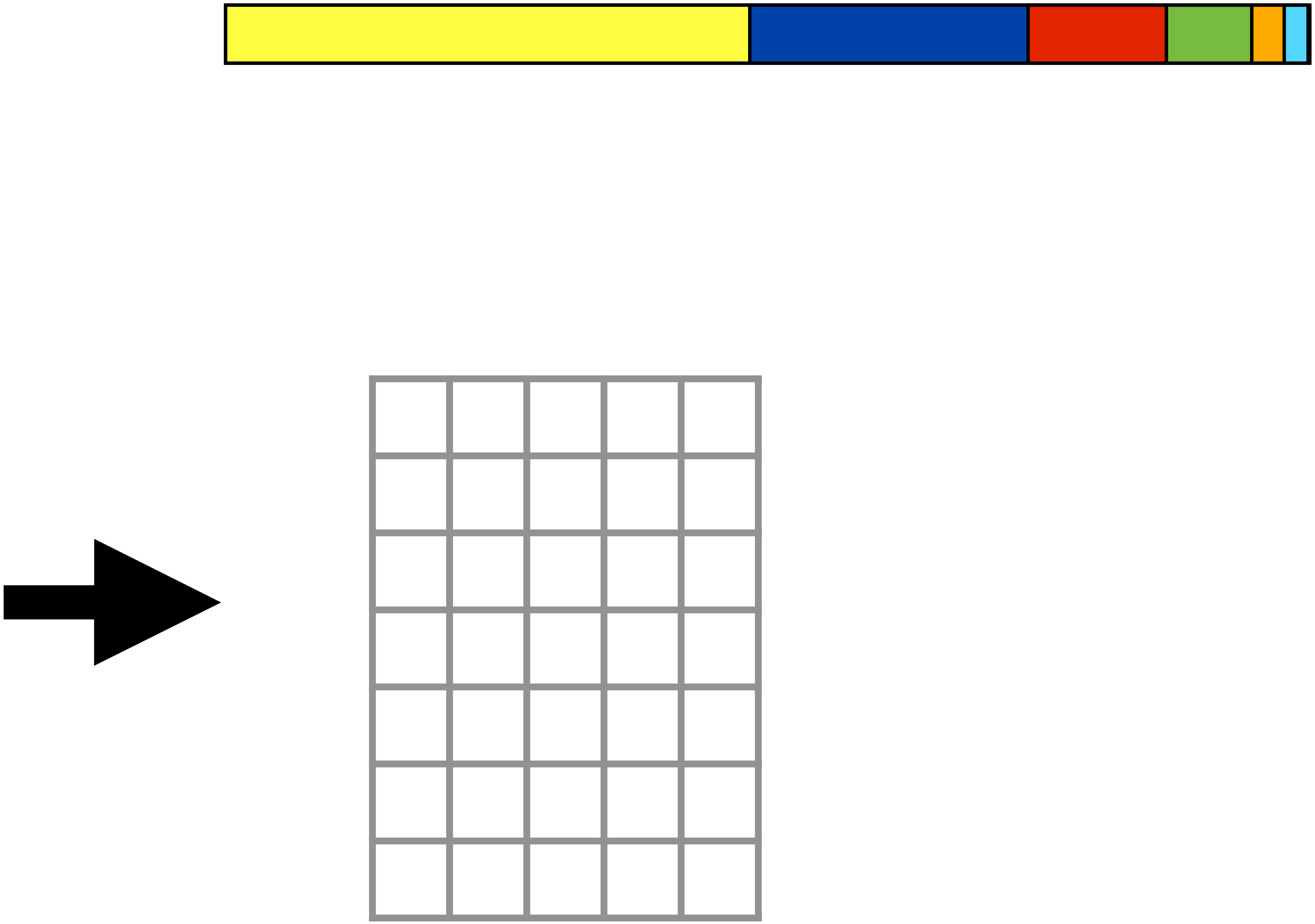
Exchangeable partition: Kingman paintbox



[Kingman 1978]

Paintboxes

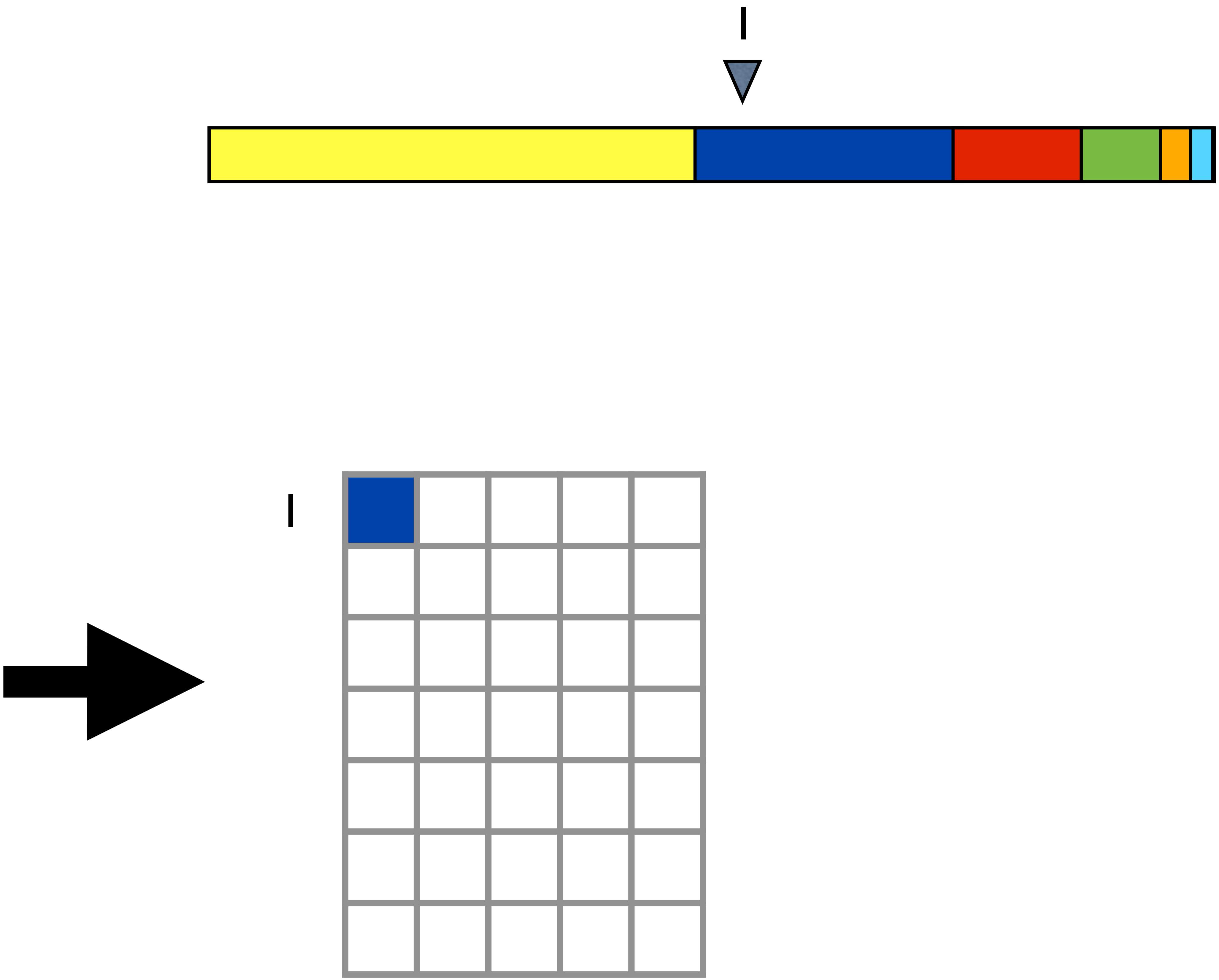
Exchangeable partition: Kingman paintbox



[Kingman 1978]

Paintboxes

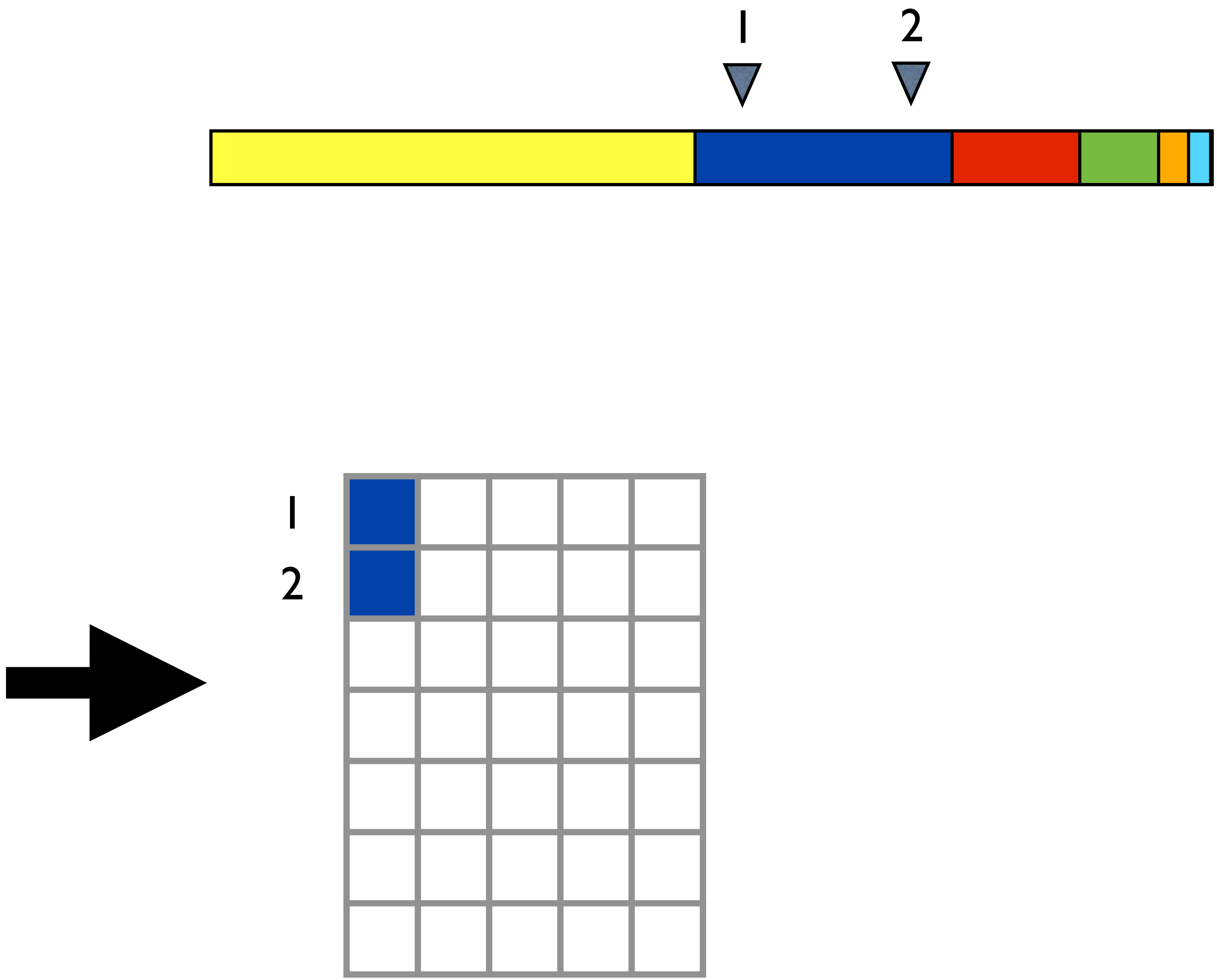
Exchangeable partition: Kingman paintbox



[Kingman 1978]

Paintboxes

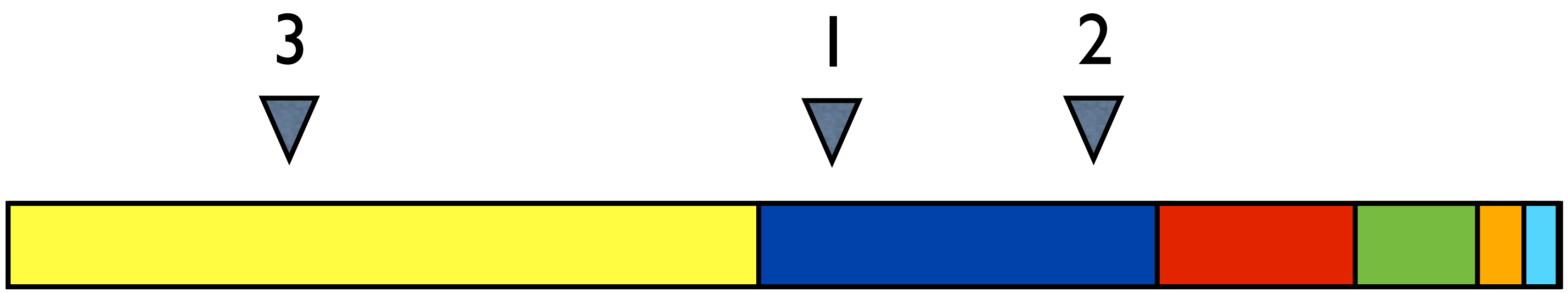
Exchangeable partition: Kingman paintbox

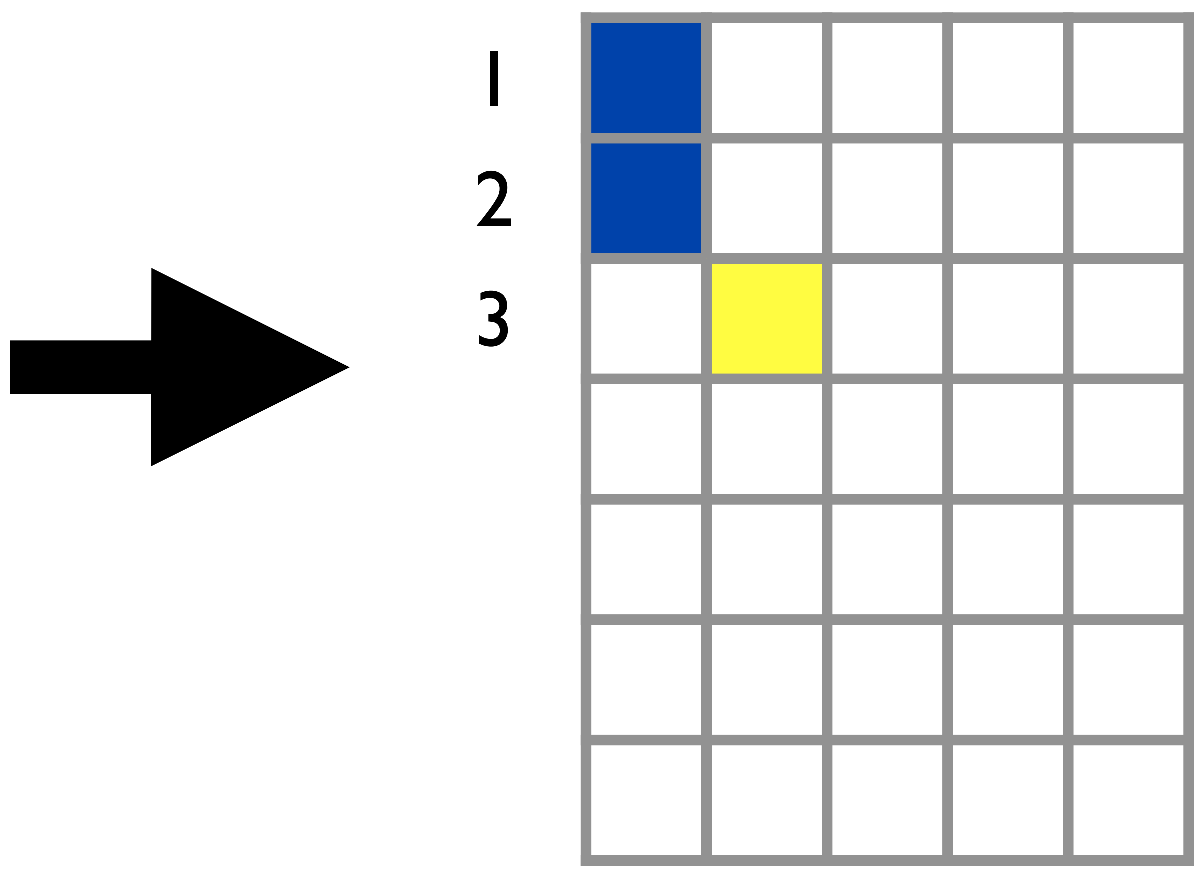


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

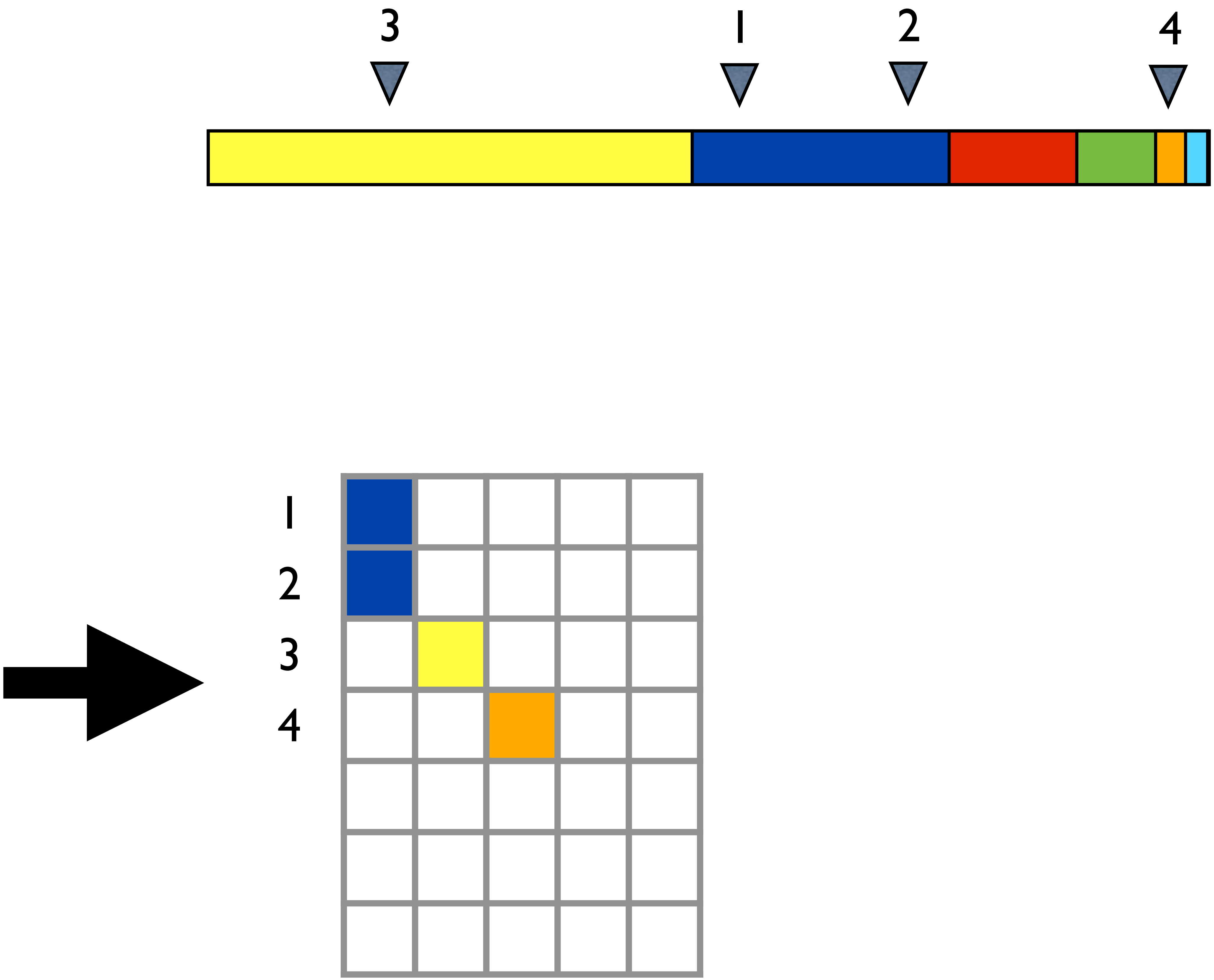




[Kingman 1978]

Paintboxes

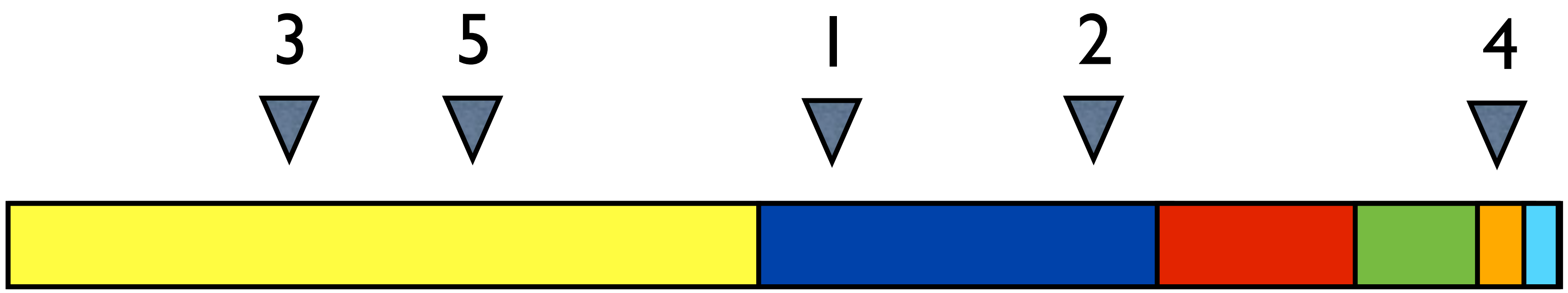
Exchangeable partition: Kingman paintbox

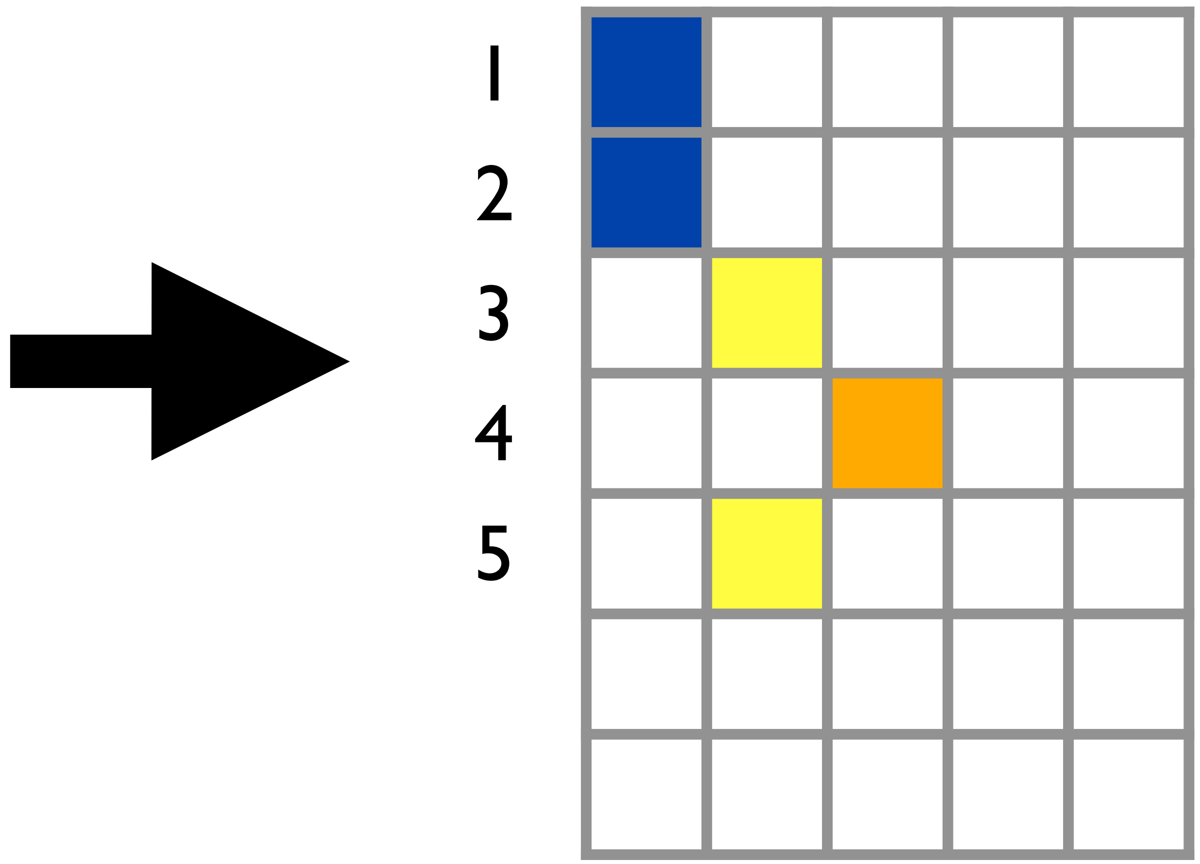


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

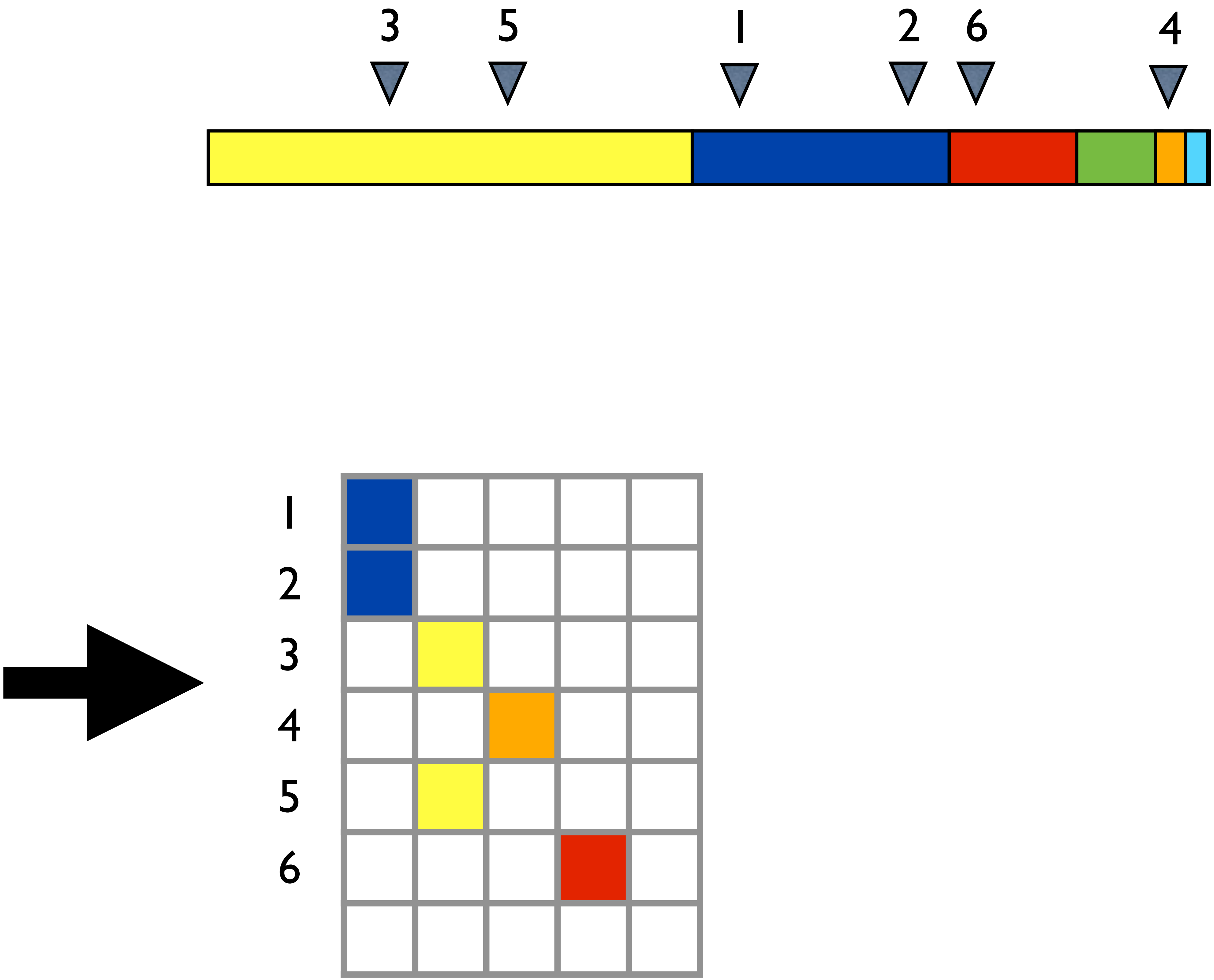




[Kingman 1978]

Paintboxes

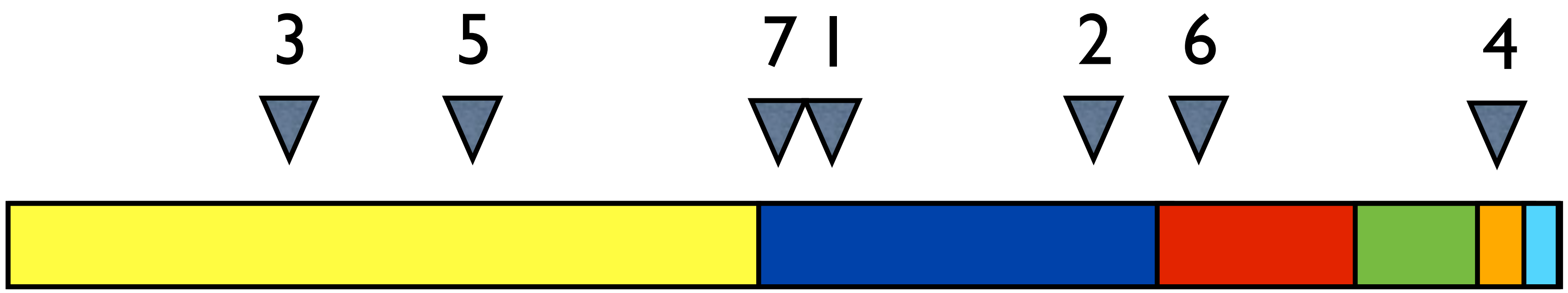
Exchangeable partition: Kingman paintbox

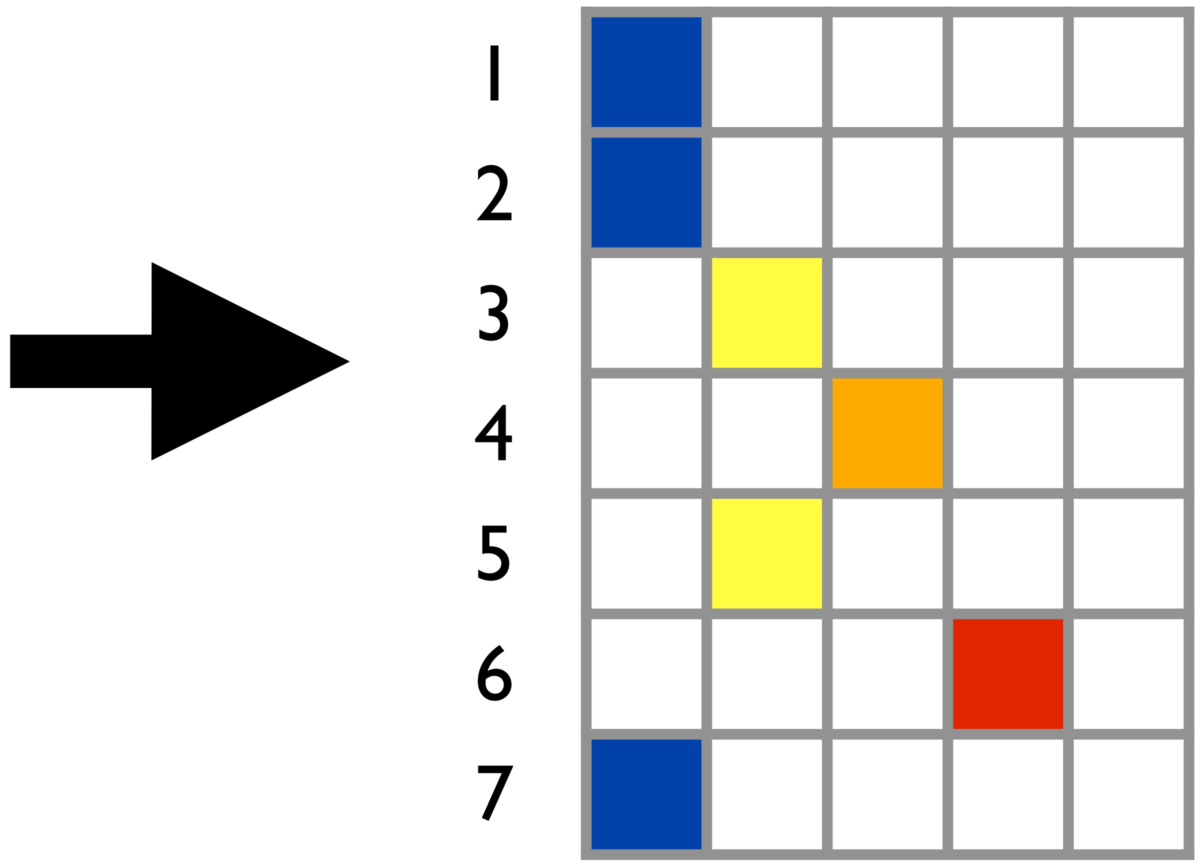


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

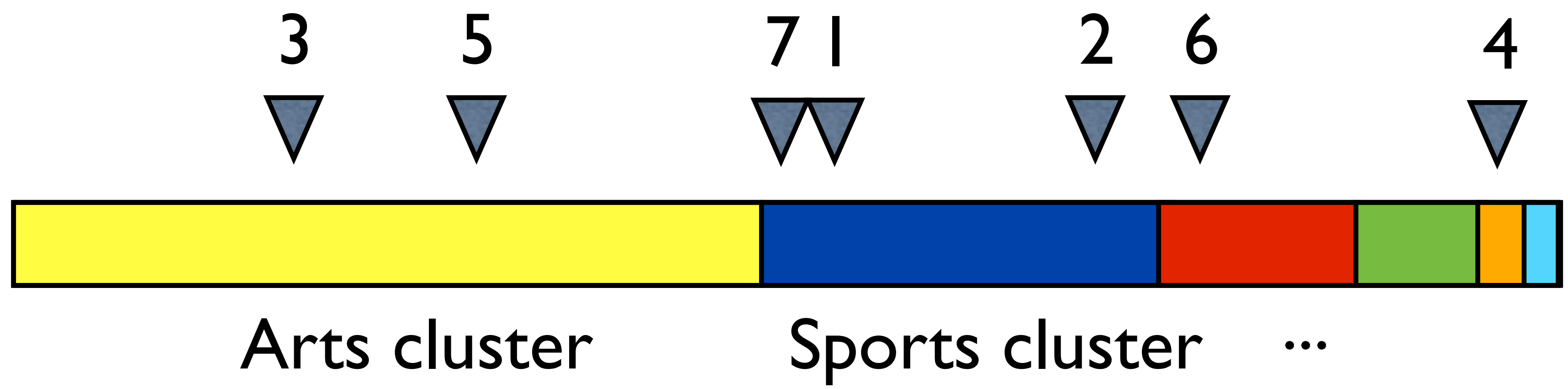


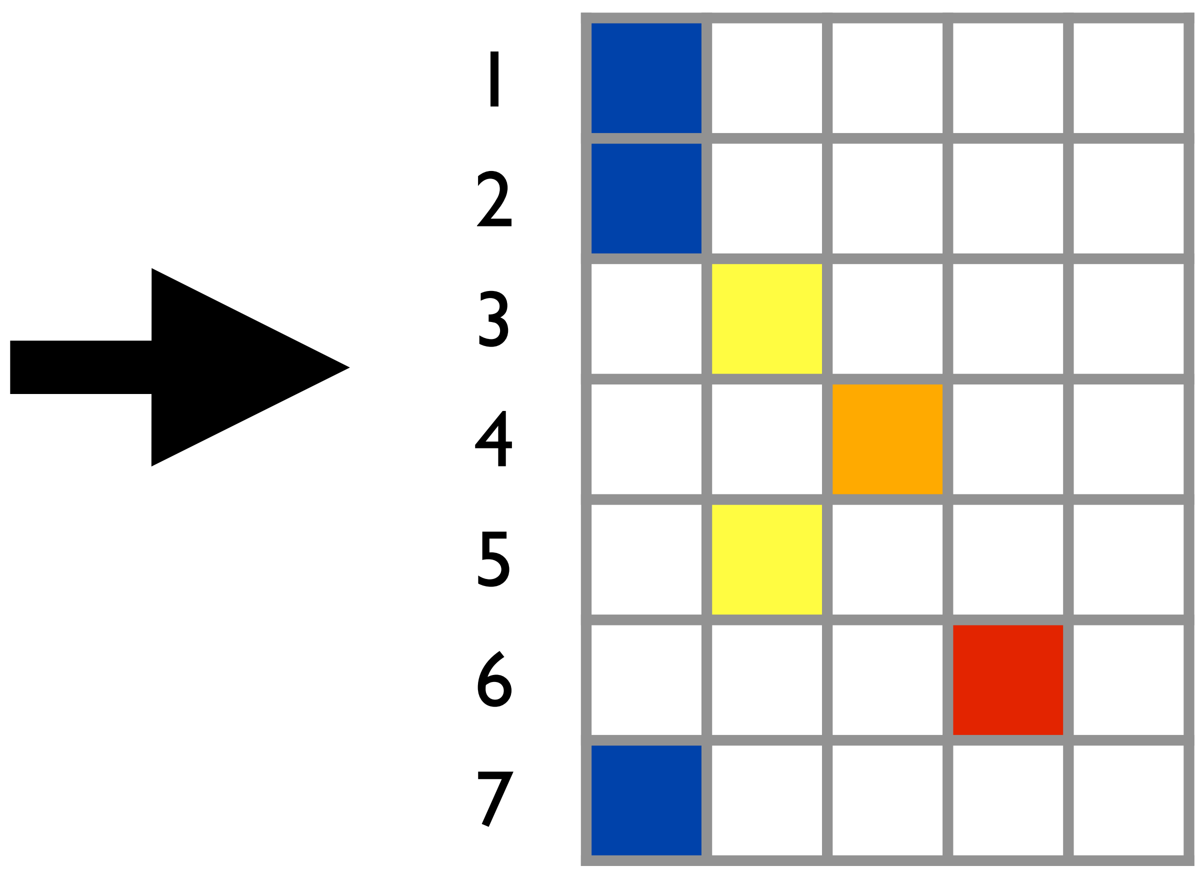


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

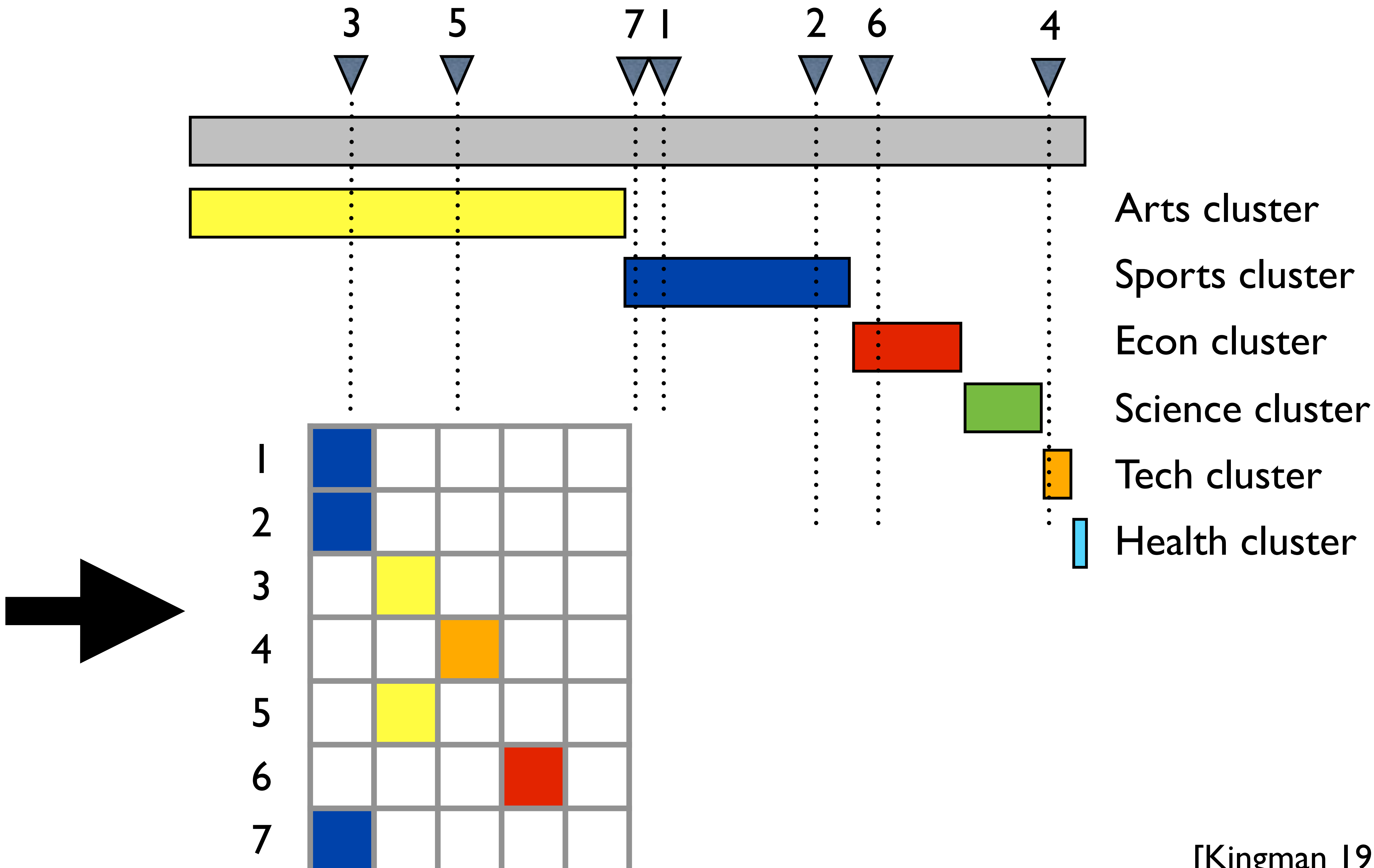




[Kingman 1978]

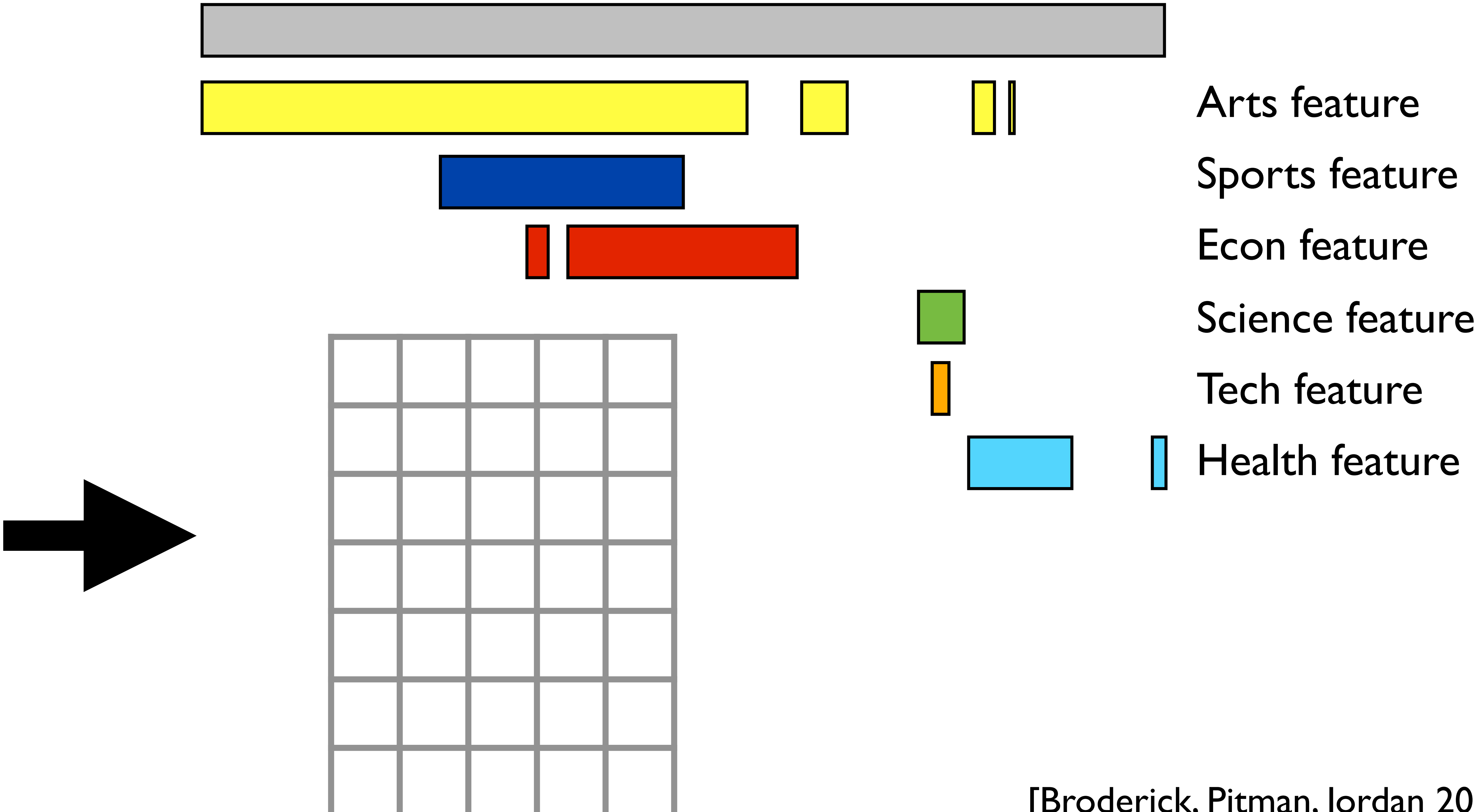
Paintboxes

Exchangeable partition: Kingman paintbox



[Kingman 1978]

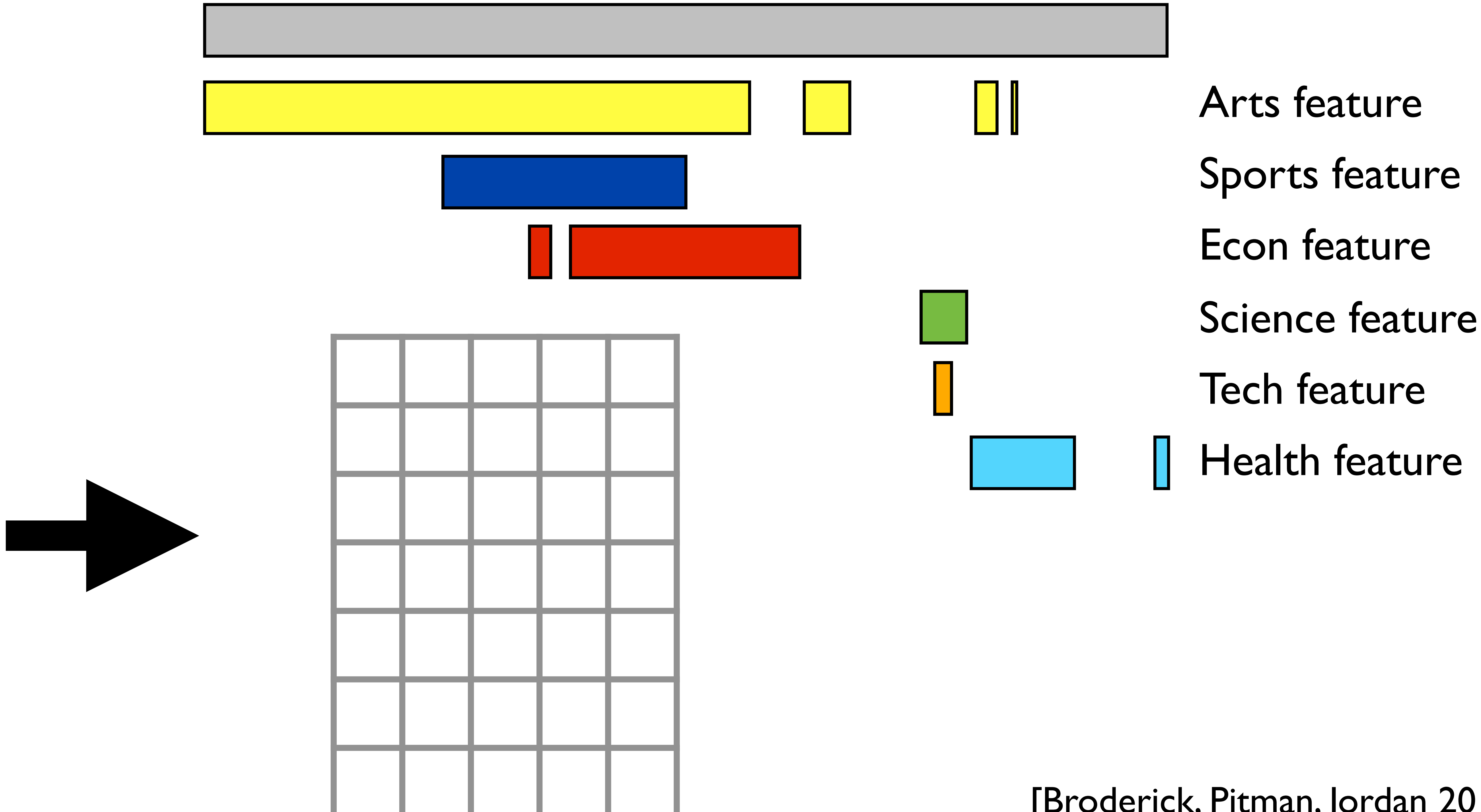
Paintboxes



[Broderick, Pitman, Jordan 2013]

Paintboxes

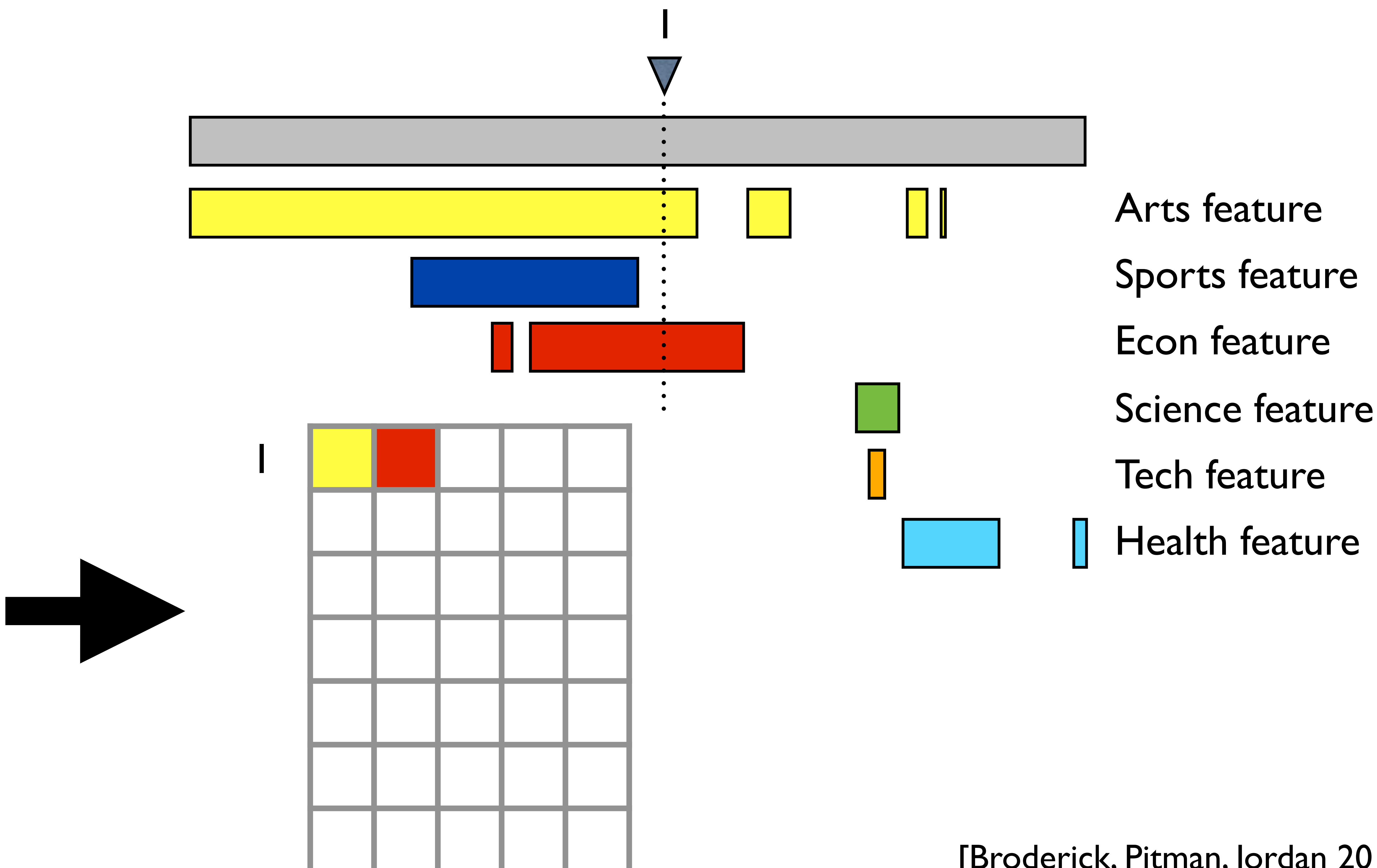
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

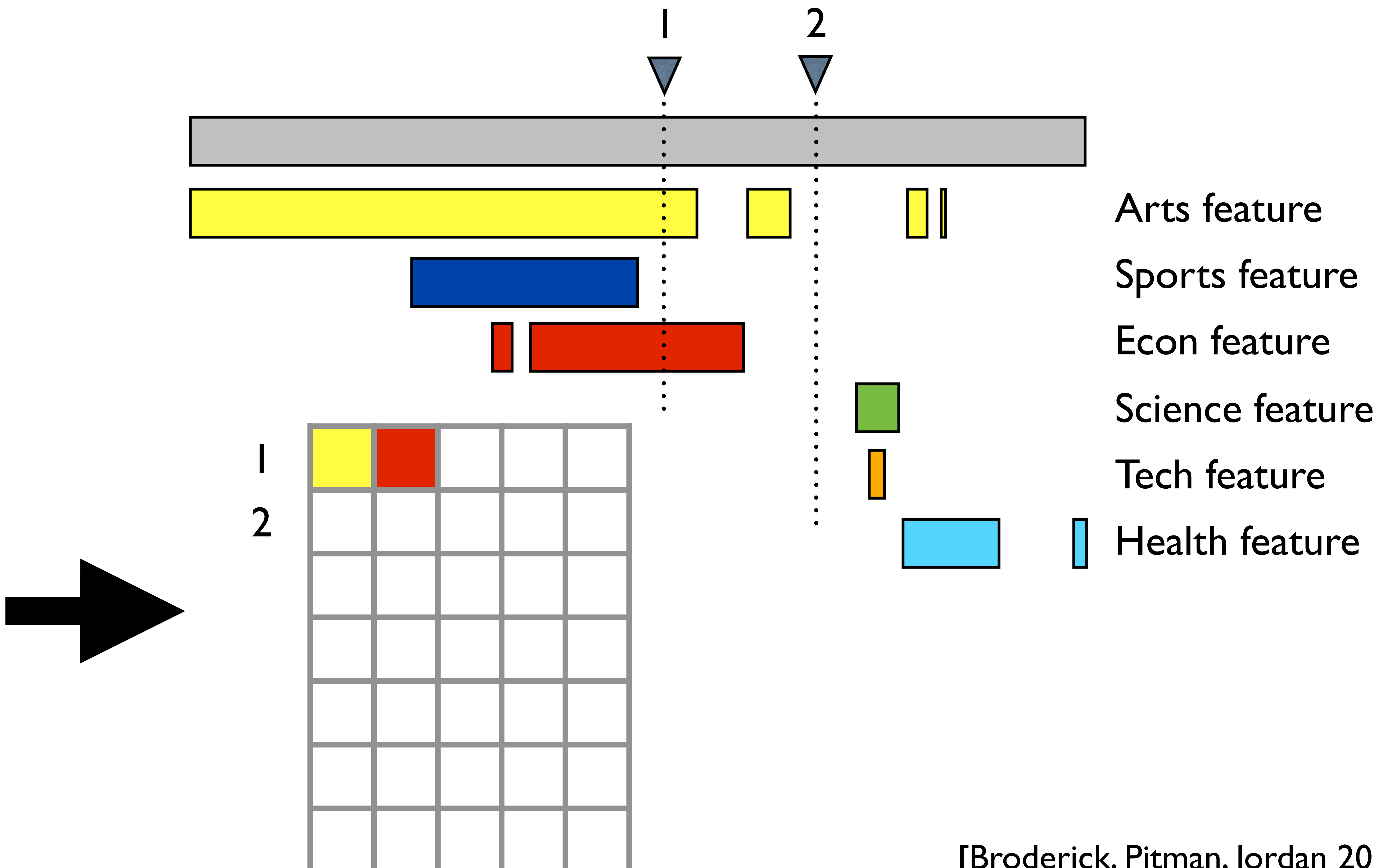
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

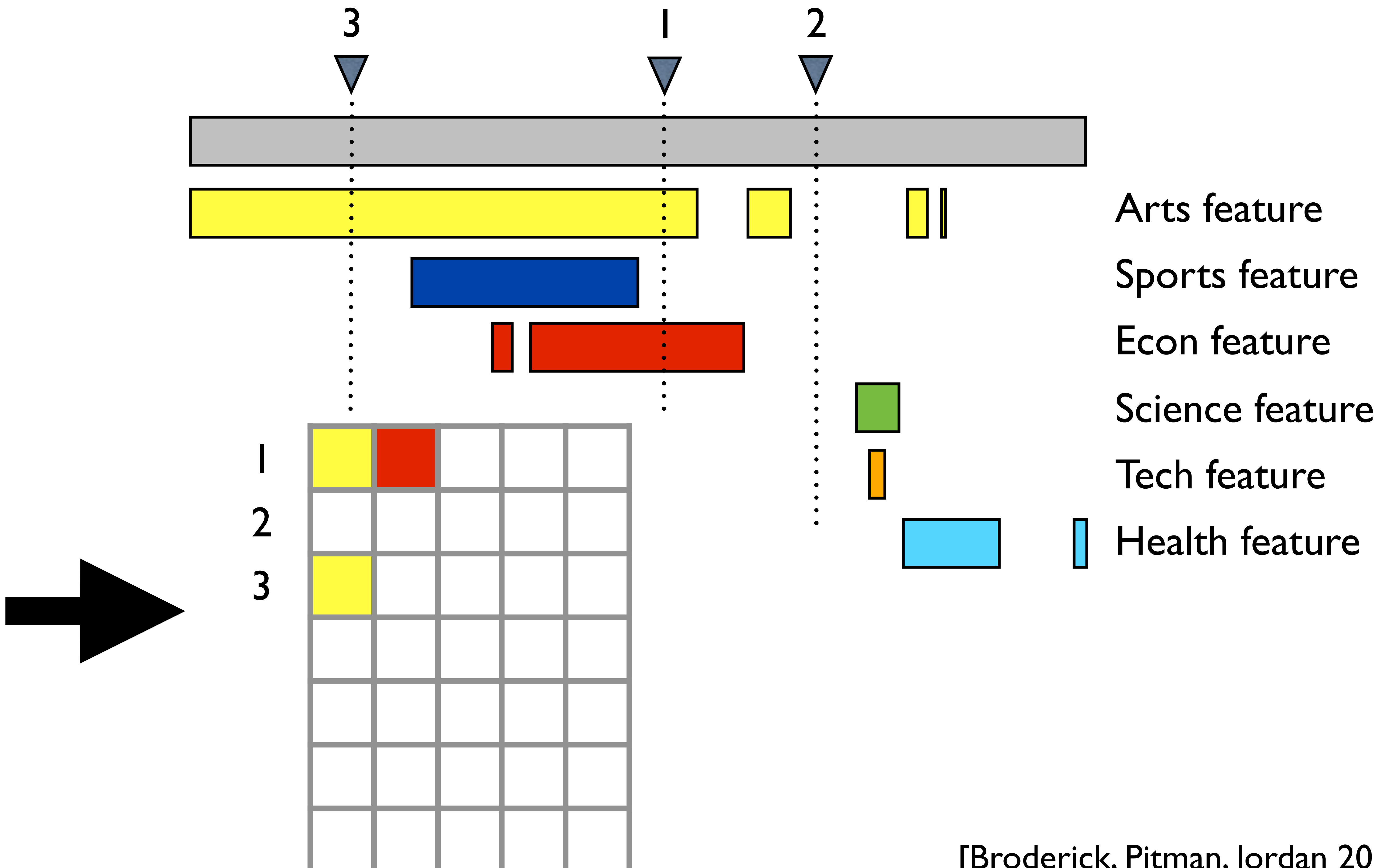
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

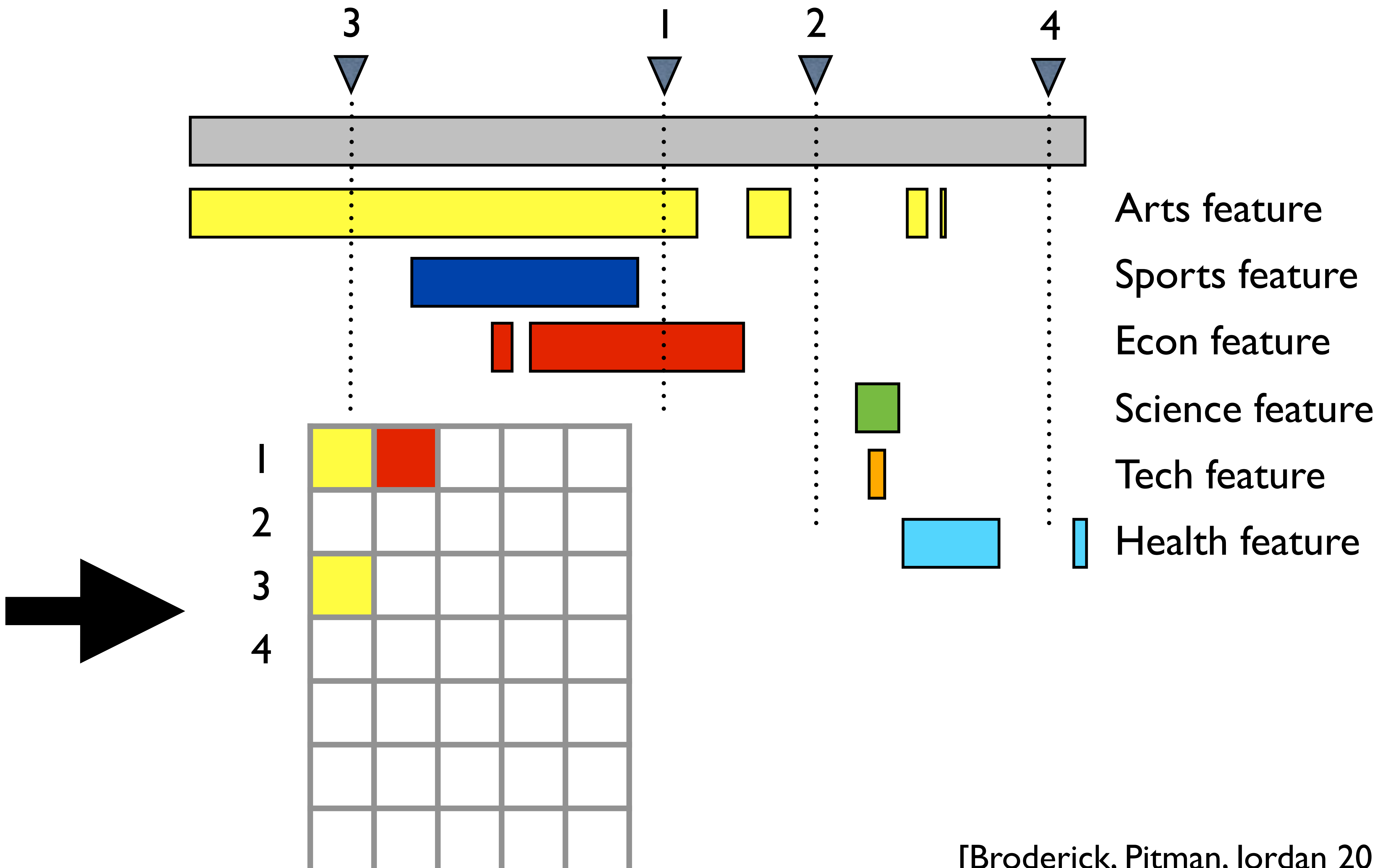
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

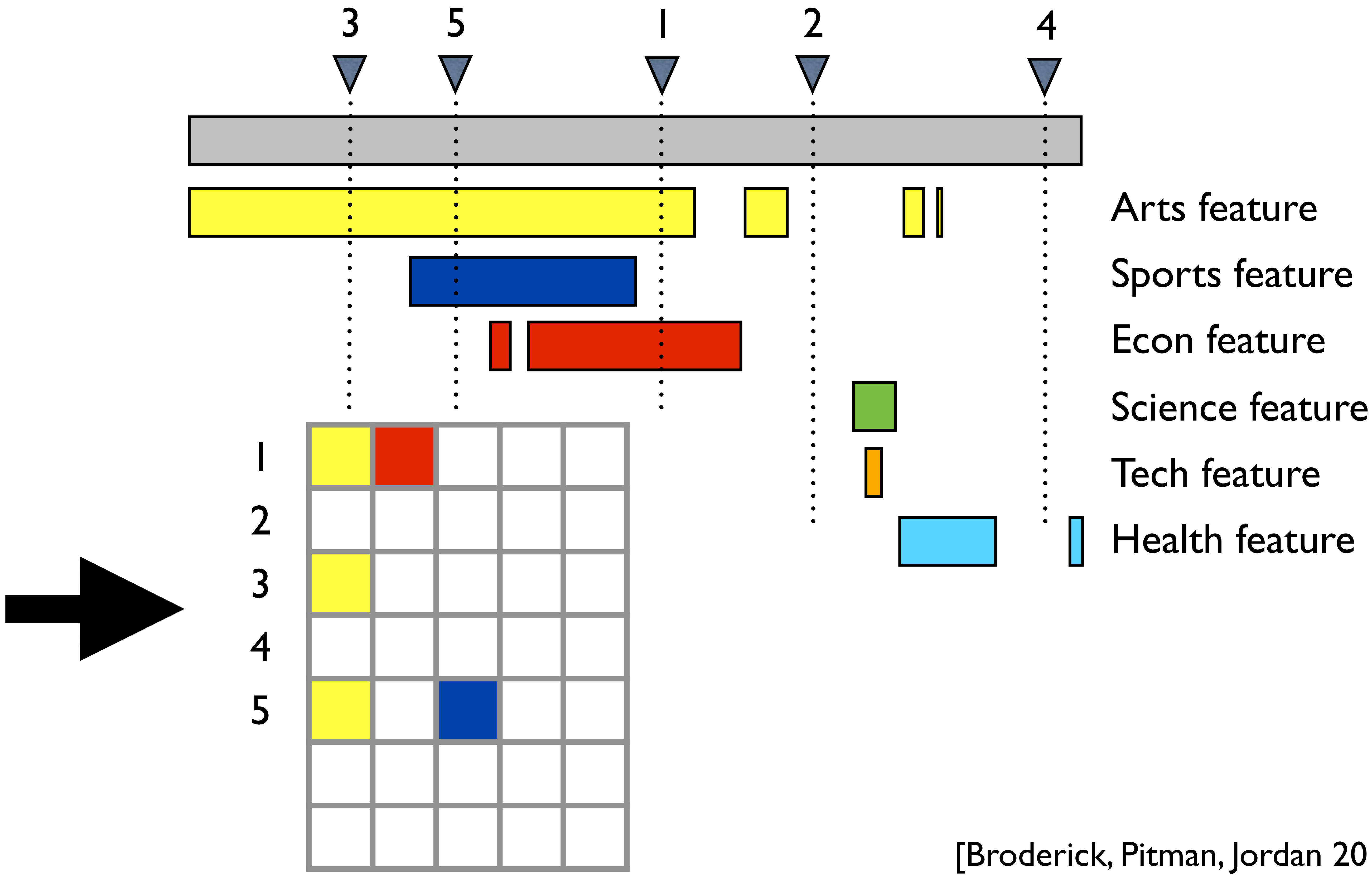
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

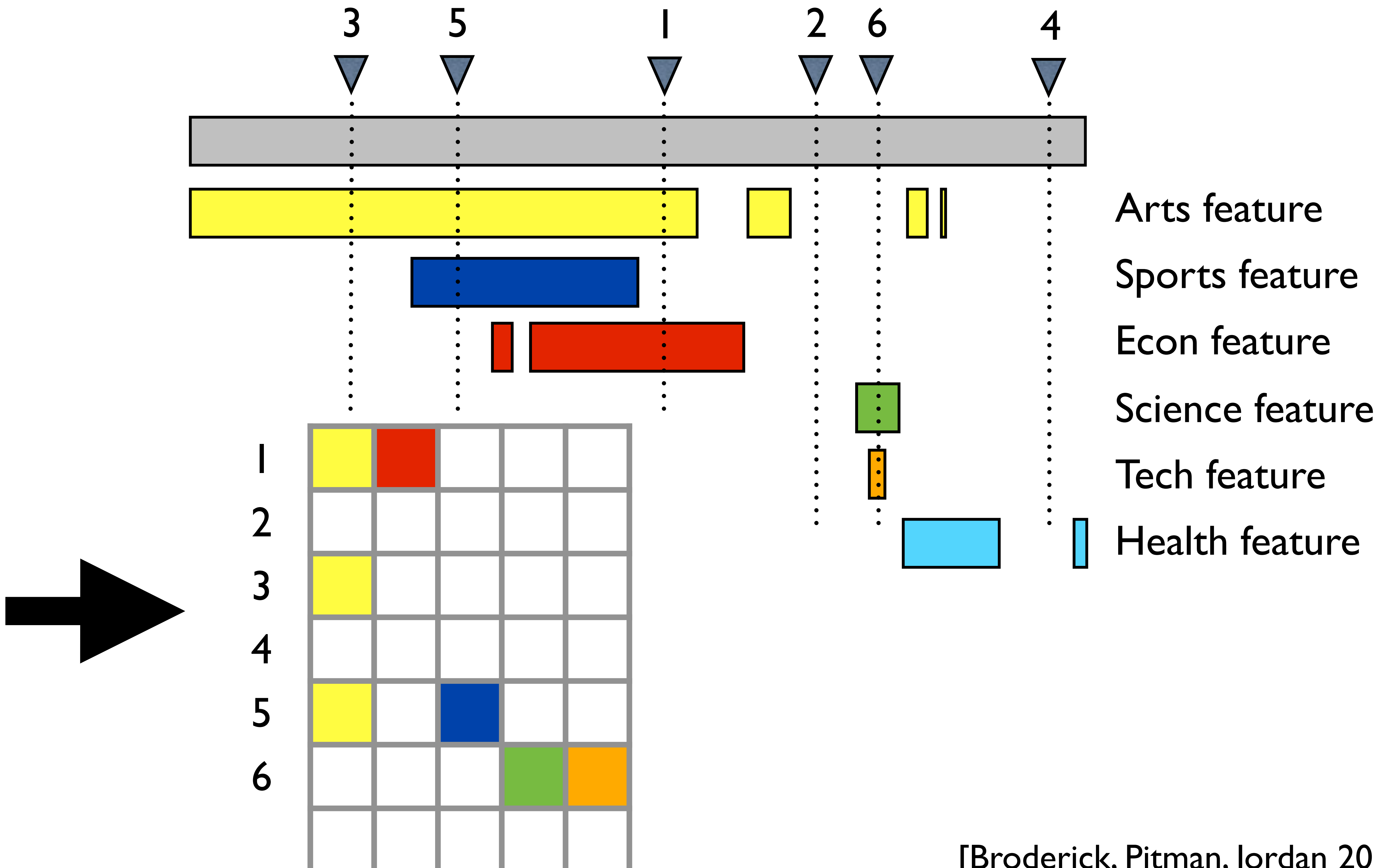
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

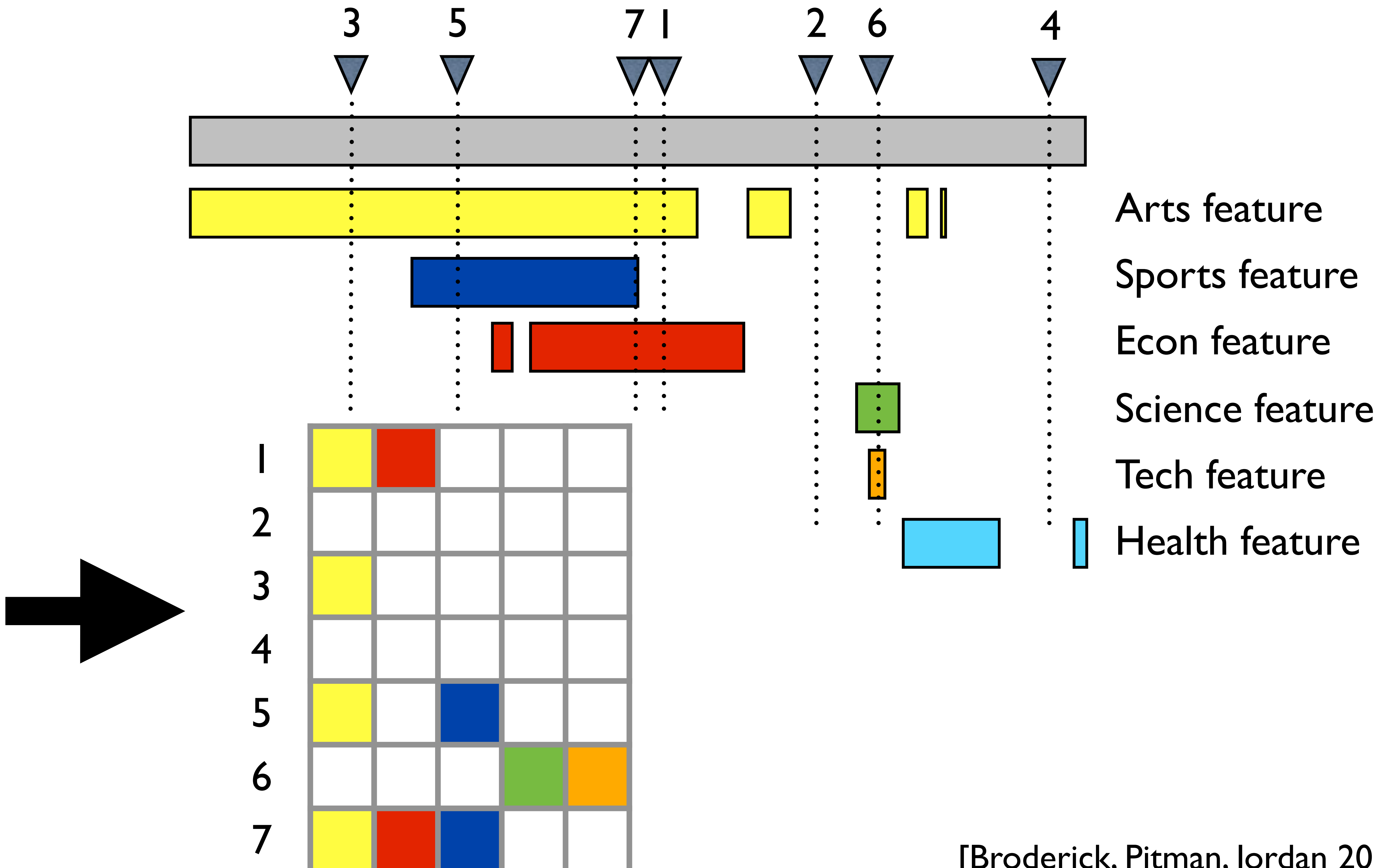
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

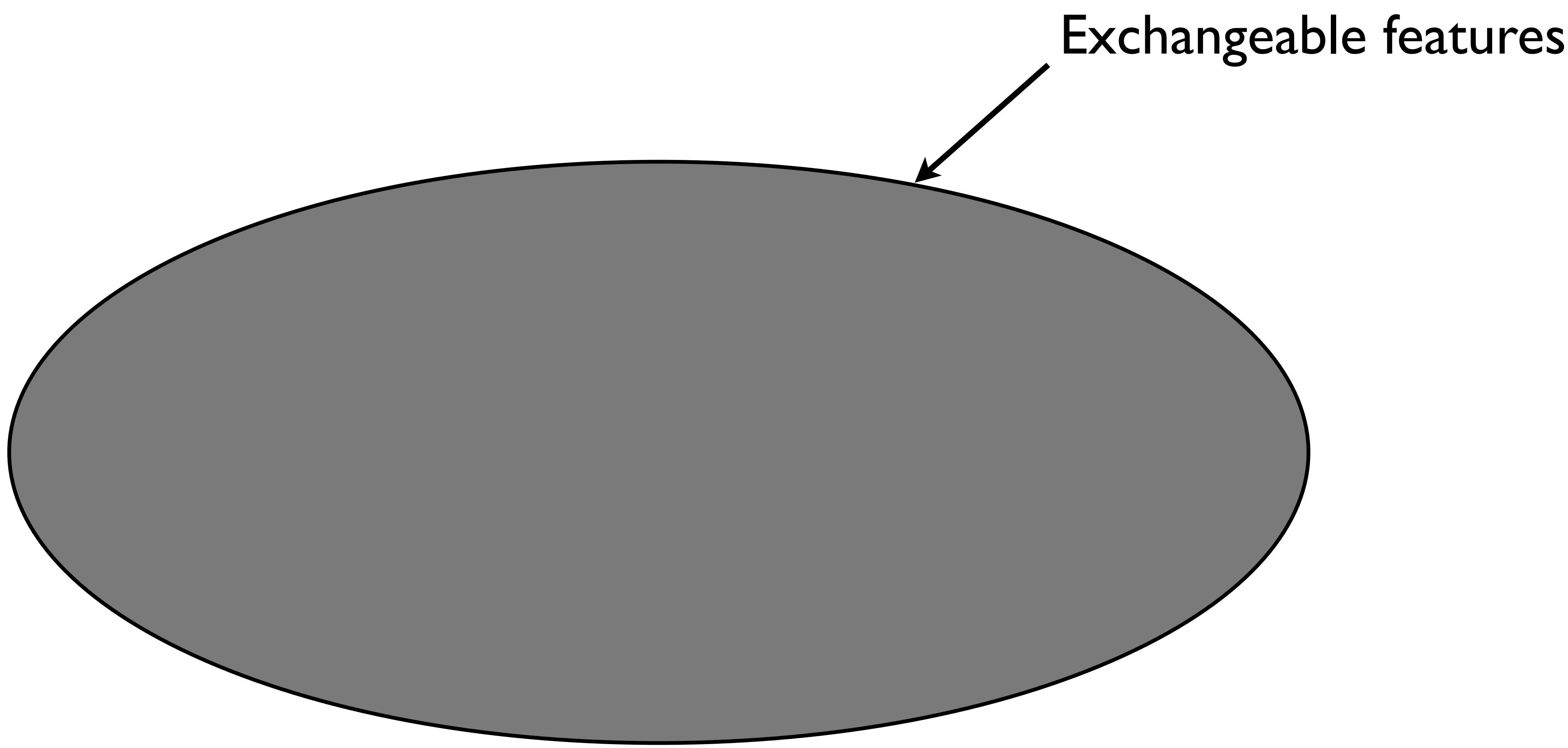
Paintboxes

Exchangeable feature allocation: feature paintbox



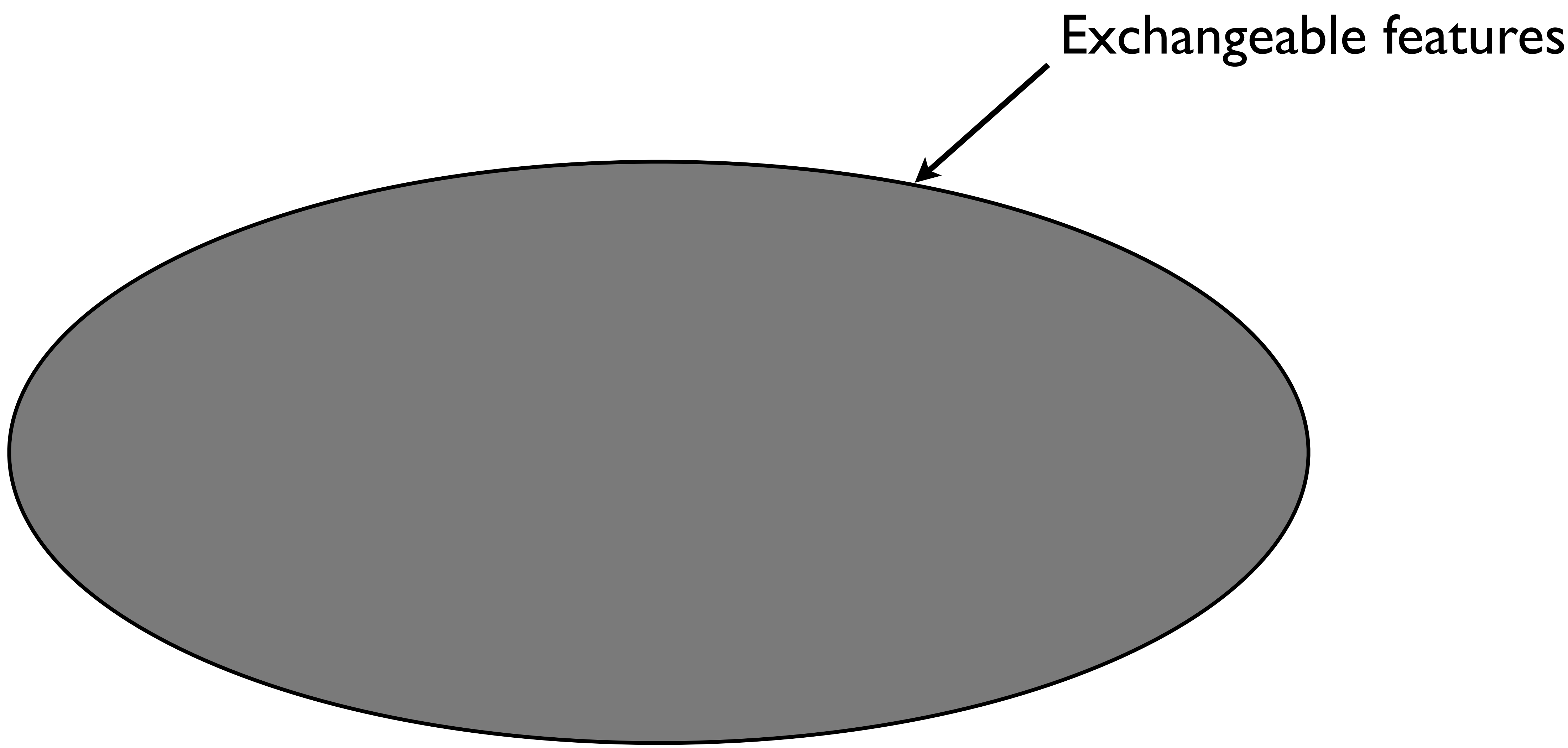
[Broderick, Pitman, Jordan 2013]

Conclusions



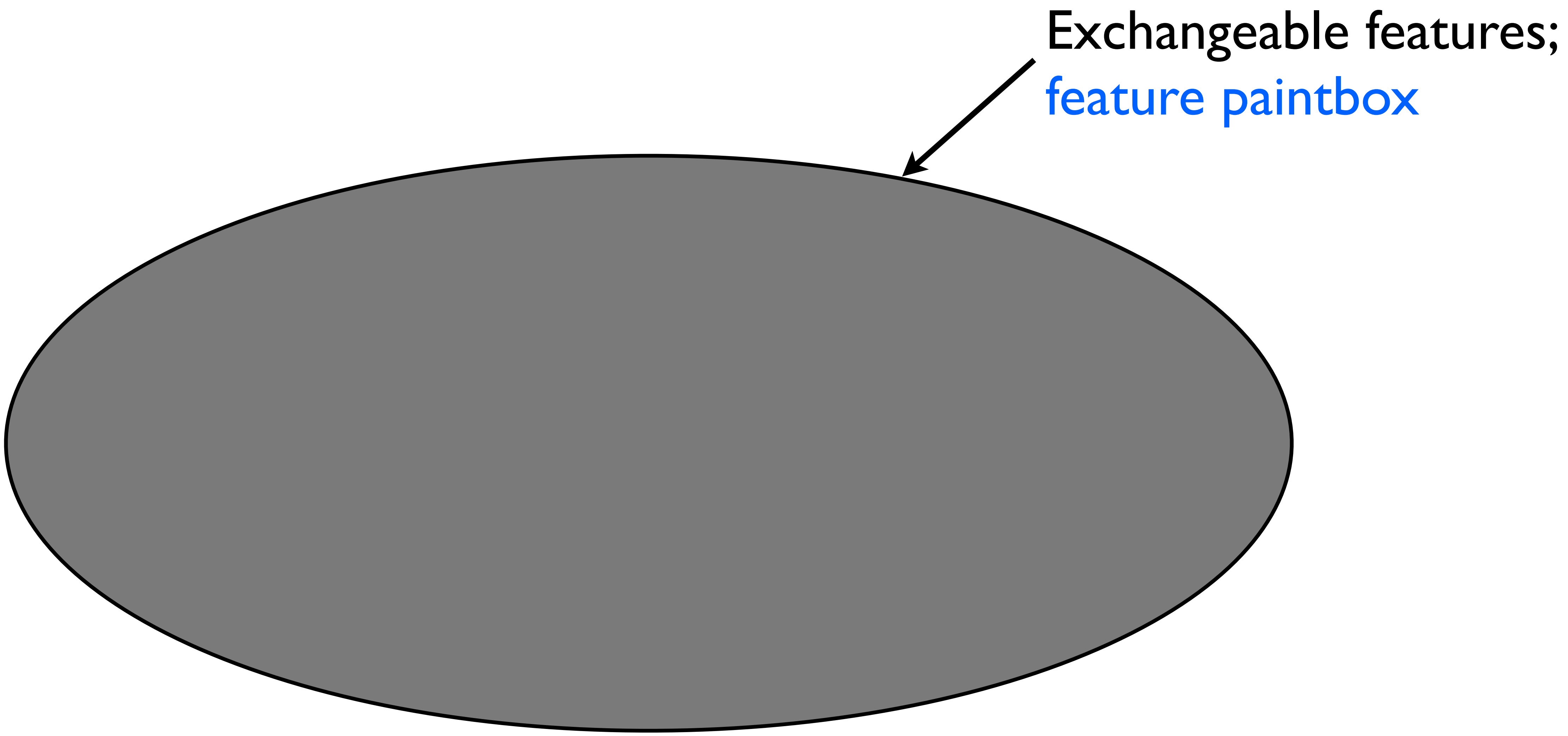
Conclusions

- Feature paintbox: characterization of exchangeable feature models



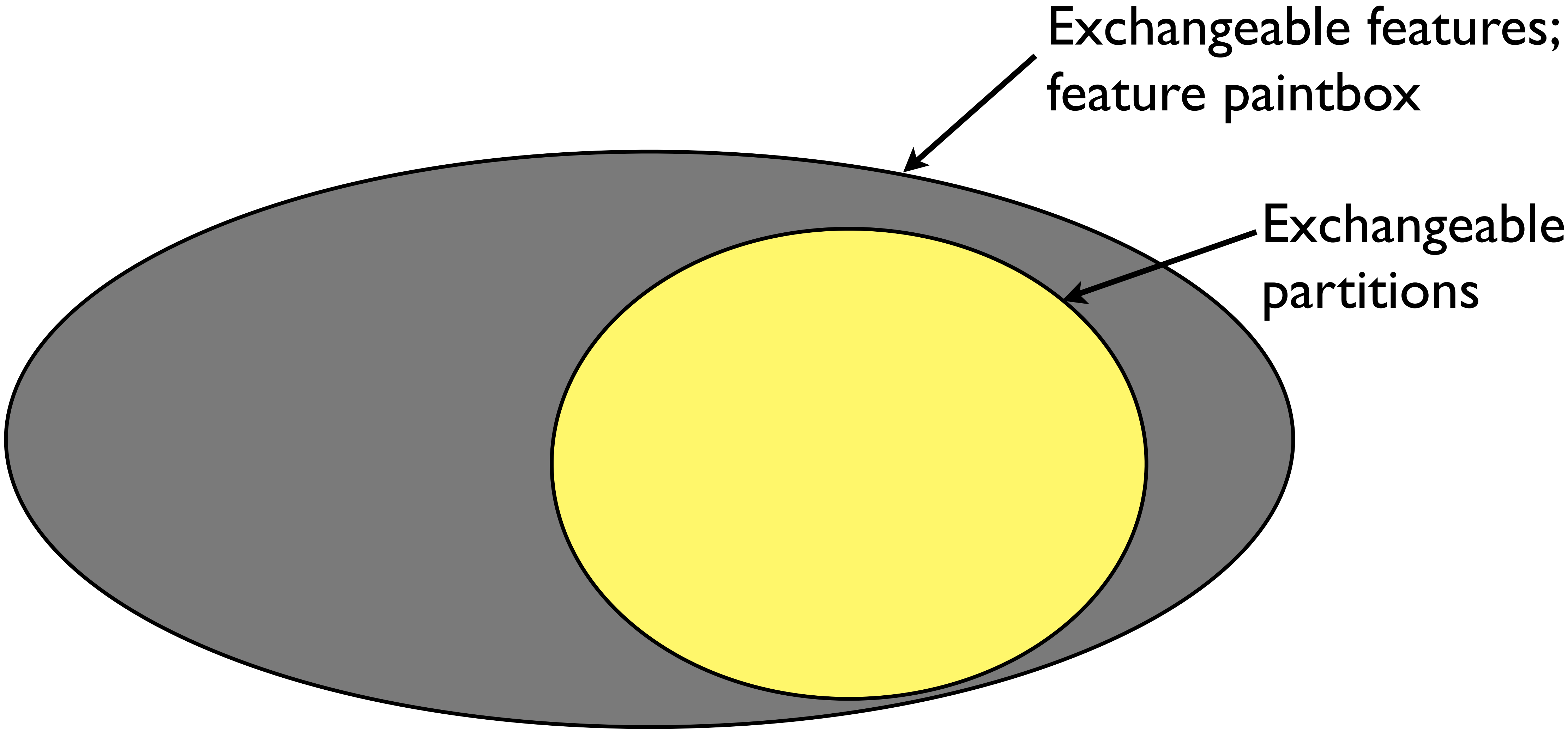
Conclusions

- Feature paintbox: characterization of exchangeable feature models



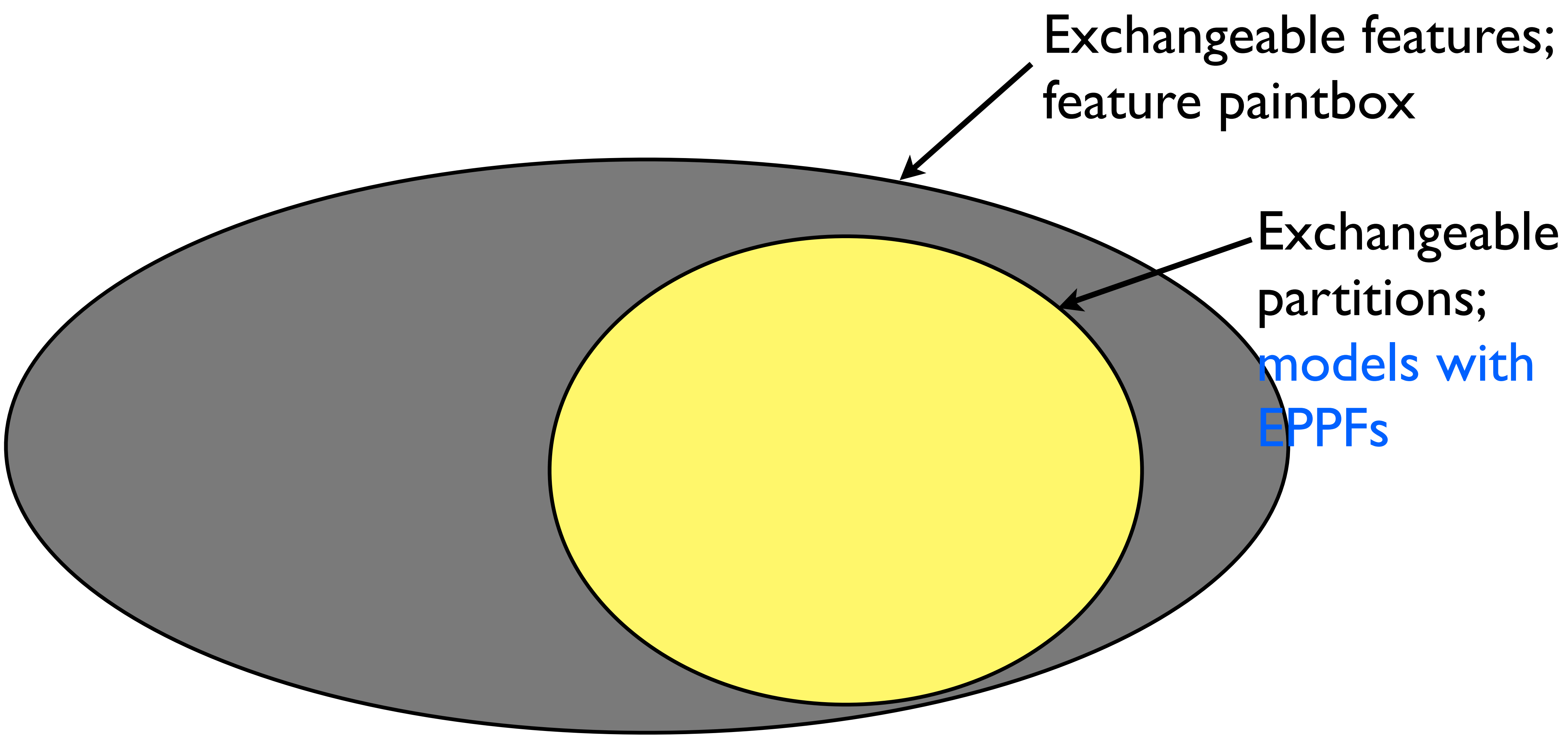
Conclusions

- Feature paintbox: characterization of exchangeable feature models



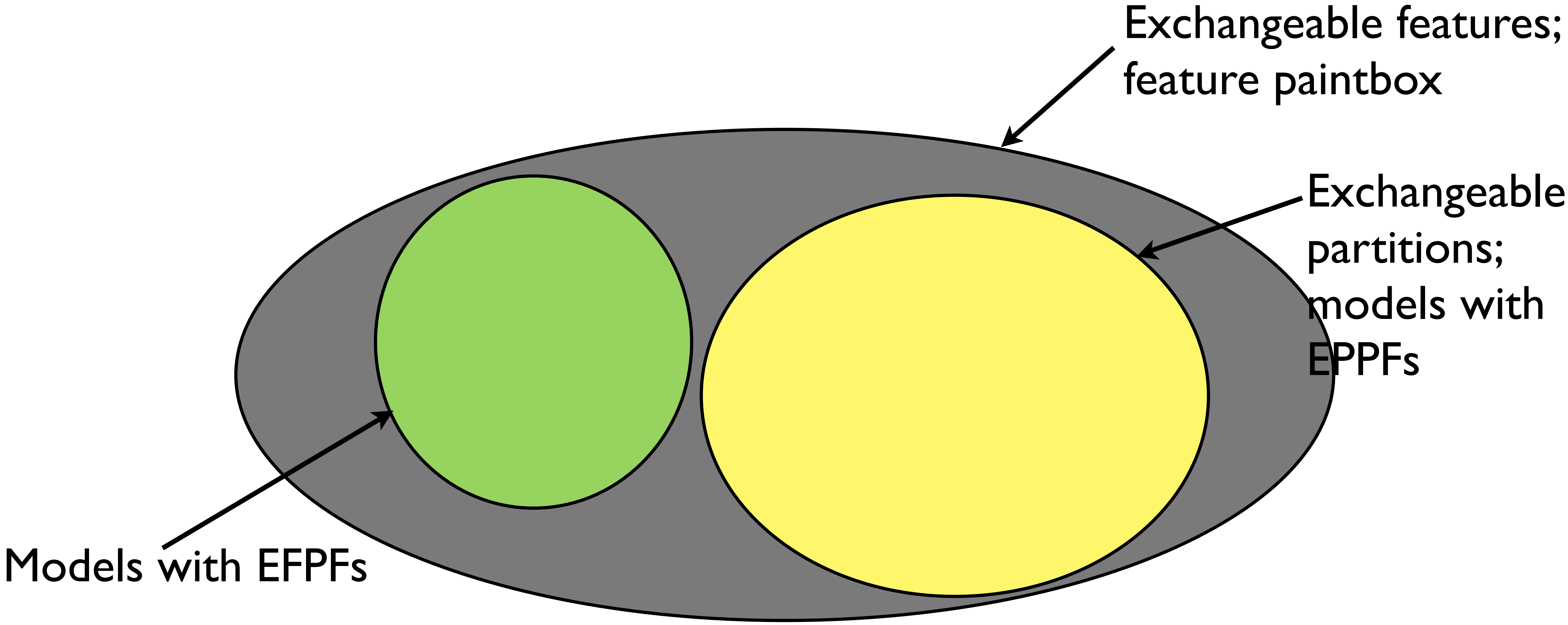
Conclusions

- Feature paintbox: characterization of exchangeable feature models



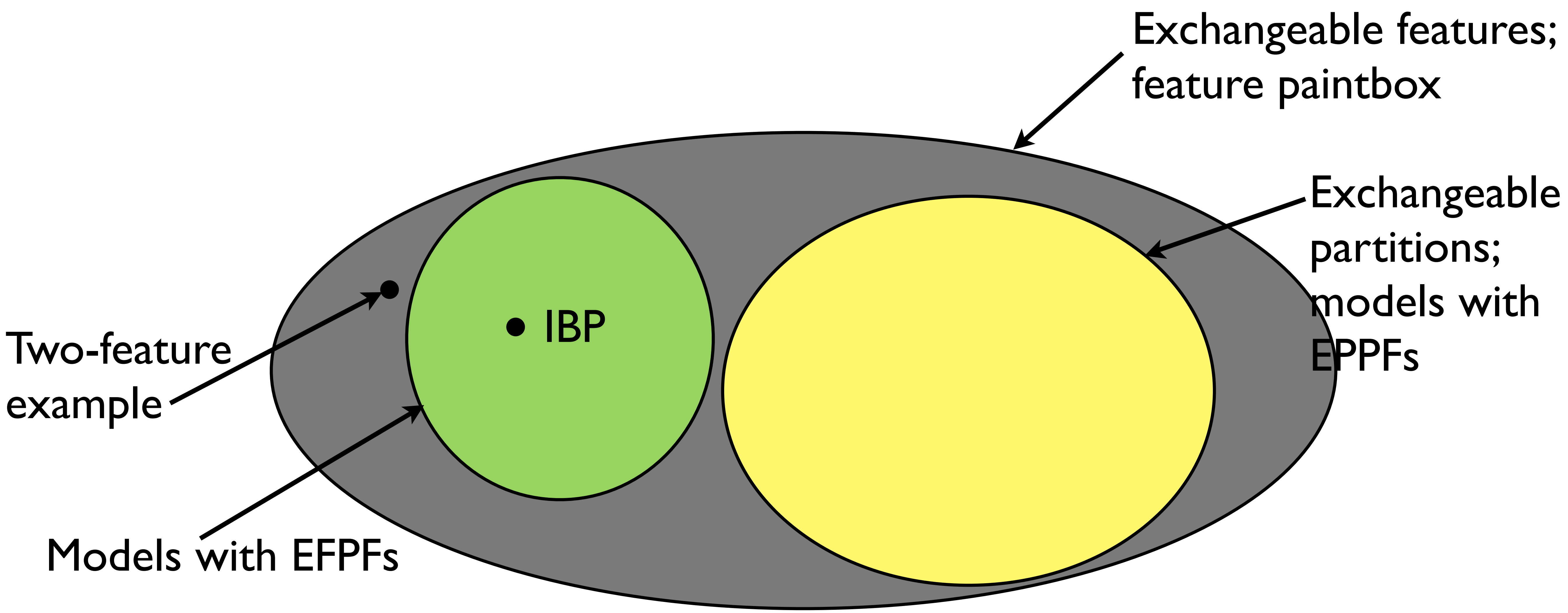
Conclusions

- Feature paintbox: characterization of exchangeable feature models



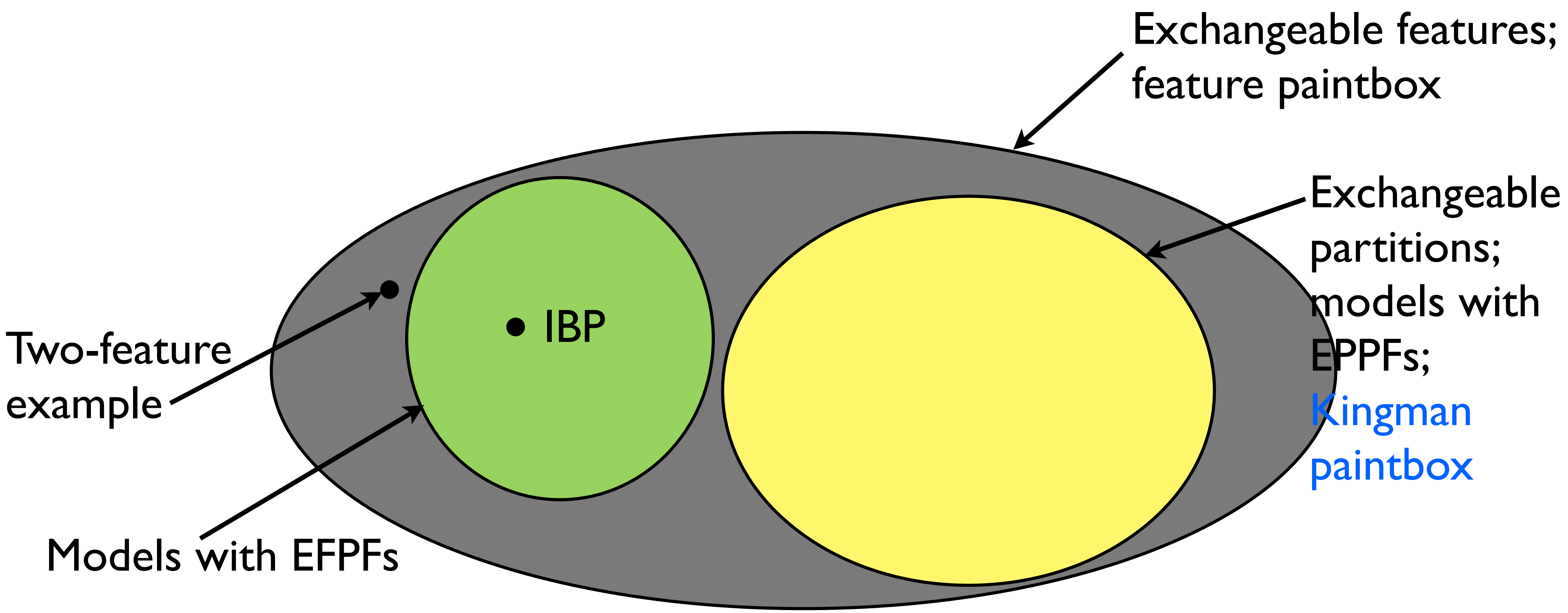
Conclusions

- Feature paintbox: characterization of exchangeable feature models



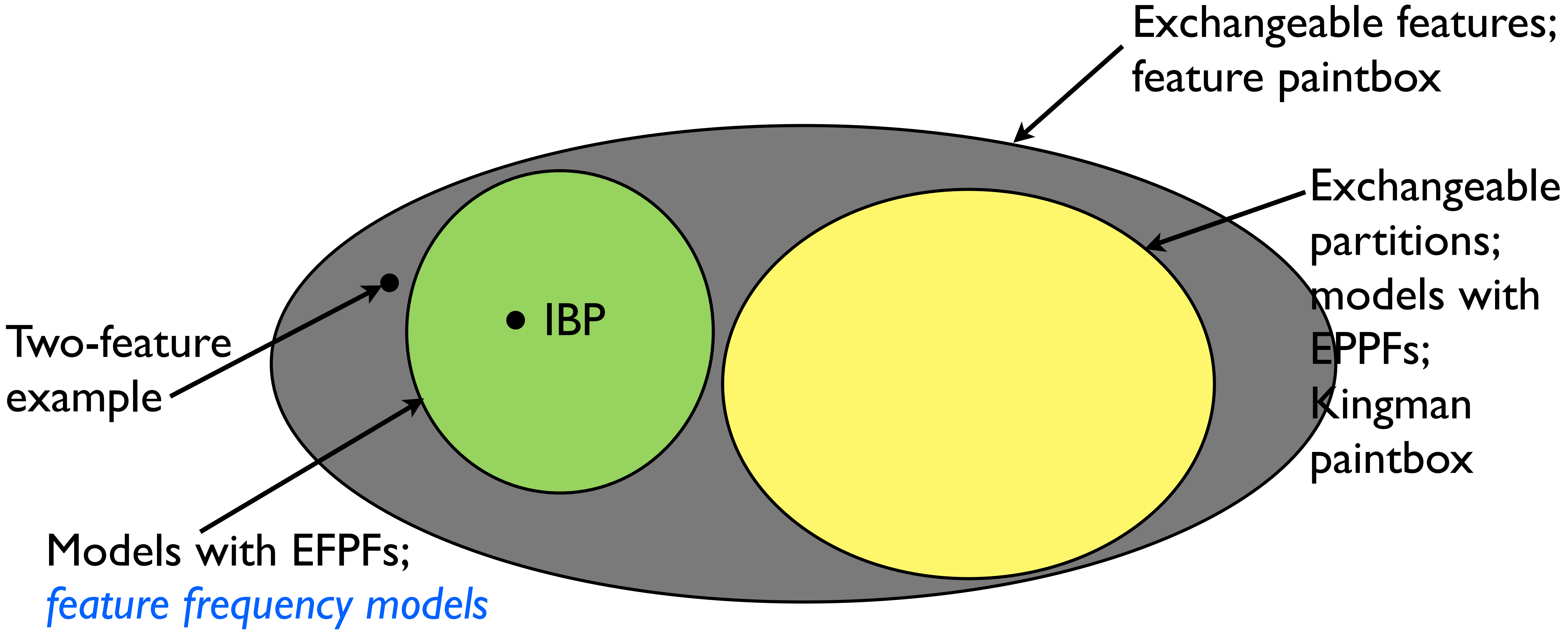
Conclusions

- Feature paintbox: characterization of exchangeable feature models



Conclusions

- Feature paintbox: characterization of exchangeable feature models



References

T. Broderick, J. Pitman, and M. I. Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 8(4):801-836, 2013.

T. Broderick, M. I. Jordan, and J. Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 28(3):289-312, 2013.

Further References

T. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *Neural Information Processing Systems*, 2006.

N. L. Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, 18(3):1259–1294, 1990.

J. F. C. Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, 2(2):374, 1978.

J. Pitman. Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, 102(2):145–158, 1995.

R. Thibaux and M. I. Jordan. Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, 2007.