Design of two-phase experiments

R. A. Bailey

r.a.bailey@qmul.ac.uk

DEMA, Isaac Newton Institute, Cambridge
August 2008

Abstract

In a two-phase experiment, treatments are (typically) allocated to experimental units in the first phase, and the products from those experimental units are allocated to a second sort of experimental unit in the second phase.

Abstract

In a two-phase experiment, treatments are (typically) allocated to experimental units in the first phase, and the products from those experimental units are allocated to a second sort of experimental unit in the second phase. The appropriate data analysis (and therefore the quality of the overall design) depends on the designs used for the two phases and on how they fit together.

Abstract

In a two-phase experiment, treatments are (typically) allocated to experimental units in the first phase, and the products from those experimental units are allocated to a second sort of experimental unit in the second phase. The appropriate data analysis (and therefore the quality of the overall design) depends on the designs used for the two phases and on how they fit together.

Usually we want to estimate the most important contrasts with low variance and with a large number of degrees of freedom for the appropriate residual. In a two-phase experiment, these criteria may conflict.

Abstract

In a two-phase experiment, treatments are (typically) allocated to experimental units in the first phase, and the products from those experimental units are allocated to a second sort of experimental unit in the second phase. The appropriate data analysis (and therefore the quality of the overall design) depends on the designs used for the two phases and on how they fit together.

Usually we want to estimate the most important contrasts with low variance and with a large number of degrees of freedom for the appropriate residual. In a two-phase experiment, these criteria may conflict.

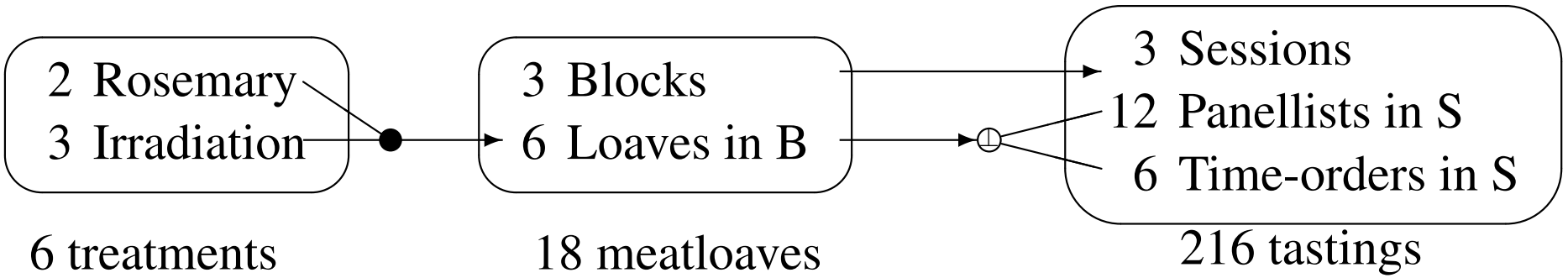
I will discuss some of the issues to think about when designing such experiments, and show how sometimes Patterson's design key can help.

Thanks

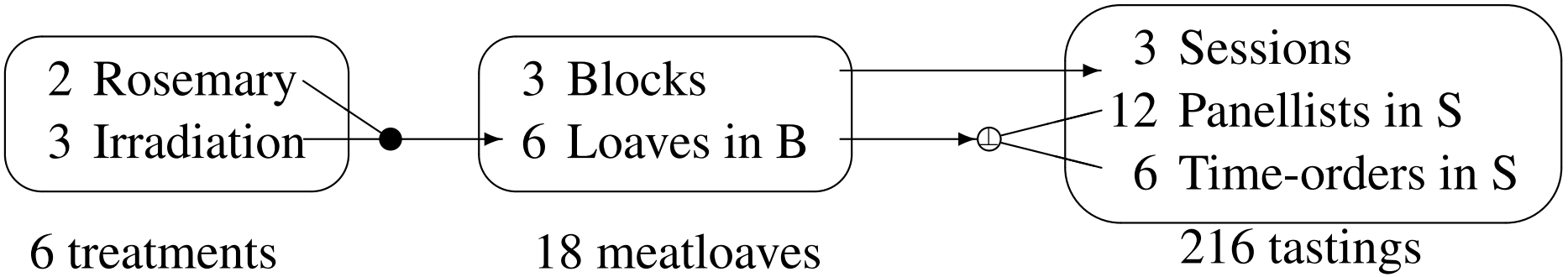
Based on

- joint work with Chris Brien
- discussions about particular examples with
  - Ruth Butler
  - Andrew Mead
  - Kathy Ruggiero
  - Andy Lynch
  - Tristan Mary-Huard
  - Terry Speed
  - Carla Vivacqua
- discussions with Ching-Shui Cheng

Meatloaves (T. B. Bailey): a tasting experiment

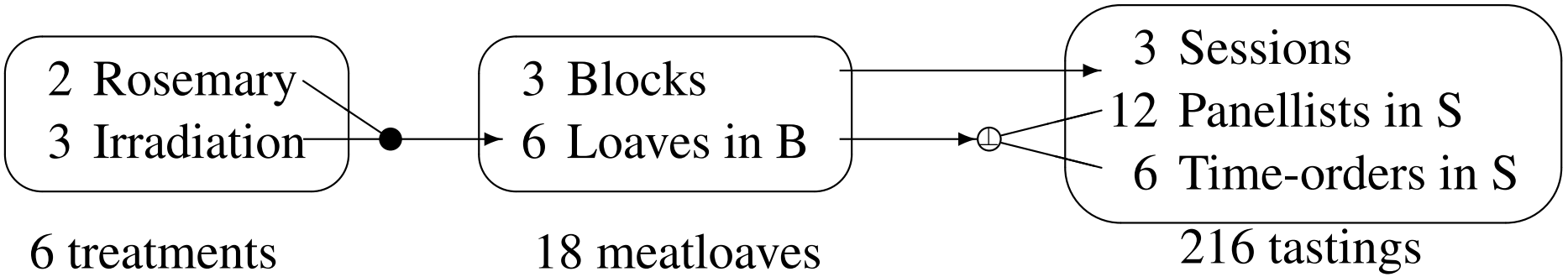


Meatloaves (T. B. Bailey): a tasting experiment



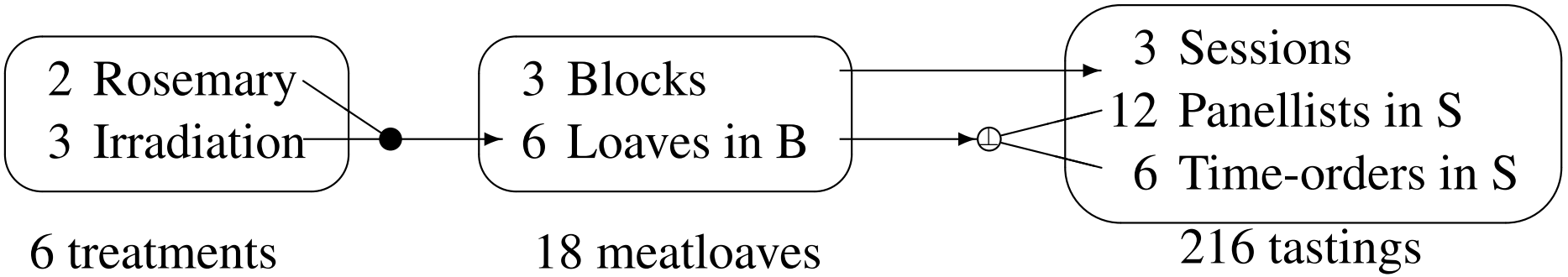
The systematic design for Phase I is randomized by randomly permuting blocks and randomly permuting loaves within each block.

Meatloaves (T. B. Bailey): a tasting experiment



The systematic design for Phase I is randomized by randomly permuting blocks and randomly permuting loaves within each block. This gives the covariance model $\gamma_{BL}\mathbf{I}_L + \gamma_B(\mathbf{J}_B - \mathbf{I}_L) + \gamma_0(\mathbf{J} - \mathbf{J}_B)$

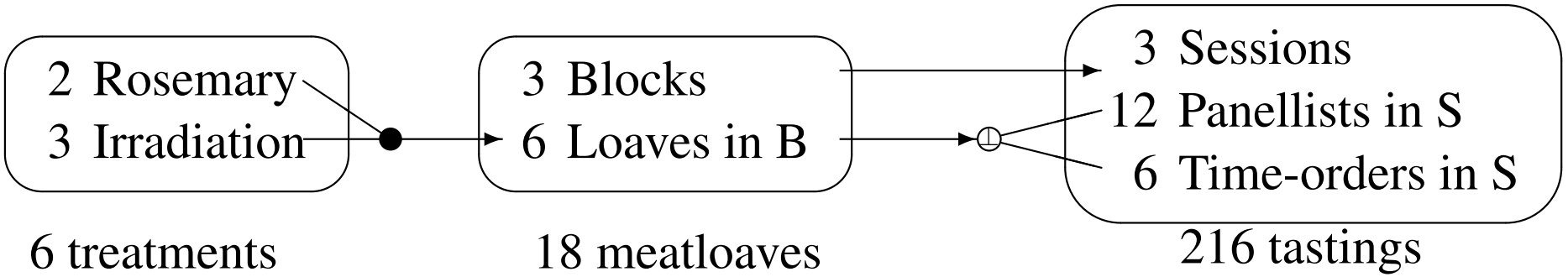
Meatloaves (T. B. Bailey): a tasting experiment



The systematic design for Phase I is randomized by randomly permuting blocks and randomly permuting loaves within each block. This gives the covariance model $\gamma_{BL}\mathbf{I}_L + \gamma_B(\mathbf{J}_B - \mathbf{I}_L) + \gamma_0(\mathbf{J} - \mathbf{J}_B)$

$$= \eta_{BL}\left(\mathbf{I}_L - \frac{\mathbf{J}_B}{6}\right) + \eta_B\left(\frac{\mathbf{J}_B}{6} - \frac{\mathbf{J}}{18}\right) + \eta_0\frac{\mathbf{J}}{18}$$

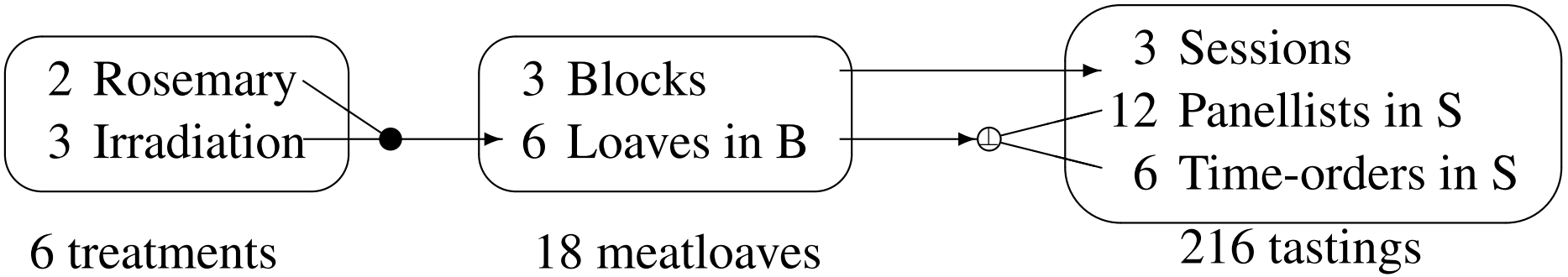
Meatloaves (T. B. Bailey): a tasting experiment



The systematic design for Phase I is randomized by randomly permuting blocks and randomly permuting loaves within each block. This gives the covariance model $\gamma_{BL}\mathbf{I}_L + \gamma_B(\mathbf{J}_B - \mathbf{I}_L) + \gamma_0(\mathbf{J} - \mathbf{J}_B)$

$$
\begin{aligned}
&= \eta_{BL}\left(\mathbf{I}_L - \frac{\mathbf{J}_B}{6}\right) + \eta_B\left(\frac{\mathbf{J}_B}{6} - \frac{\mathbf{J}}{18}\right) + \eta_0\frac{\mathbf{J}}{18} \\
&= \eta_{BL}\mathbf{Q}_{BL} + \eta_B\mathbf{Q}_B + \eta_0\mathbf{Q}_0
\end{aligned}
$$

Meatloaves (T. B. Bailey): a tasting experiment



The systematic design for Phase I is randomized by randomly permuting blocks and randomly permuting loaves within each block. This gives the covariance model $\gamma_{BL}\mathbf{I}_L + \gamma_B(\mathbf{J}_B - \mathbf{I}_L) + \gamma_0(\mathbf{J} - \mathbf{J}_B)$

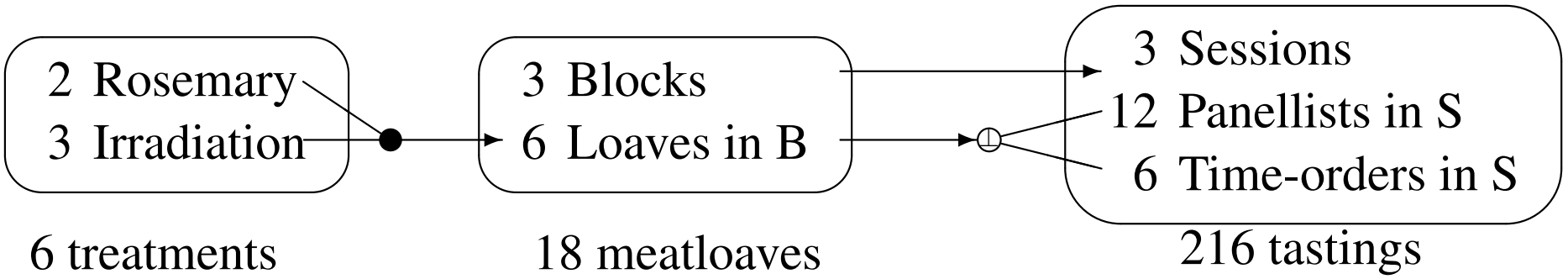
$$= \eta_{BL}\left(\mathbf{I}_L - \frac{\mathbf{J}_B}{6}\right) + \eta_B\left(\frac{\mathbf{J}_B}{6} - \frac{\mathbf{J}}{18}\right) + \eta_0\frac{\mathbf{J}}{18}$$

$$= \eta_{BL}\mathbf{Q}_{BL} + \eta_B\mathbf{Q}_B + \eta_0\mathbf{Q}_0 \qquad \leftarrow \text{orthogonal idempotents}$$

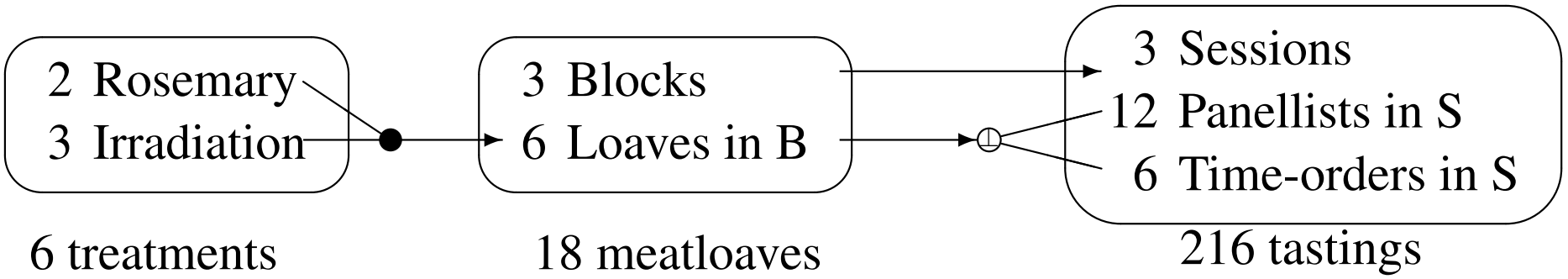
$\uparrow \qquad \uparrow \qquad \uparrow$

positive

Meatloaves (T. B. Bailey): a tasting experiment

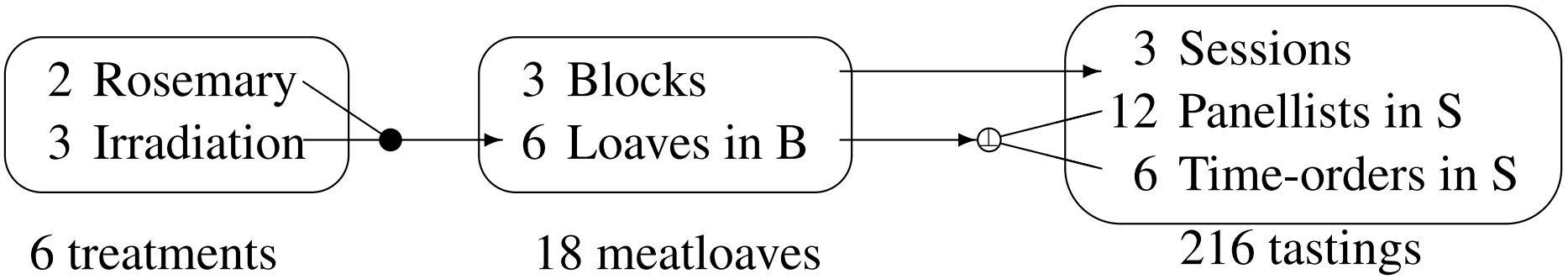


Meatloaves (T. B. Bailey): a tasting experiment



The systematic design for Phase II (a pair of 6×6 Latin squares) is randomized by randomly permuting sessions, randomly permuting panellists within each session, and randomly permuting time-orders within each session.

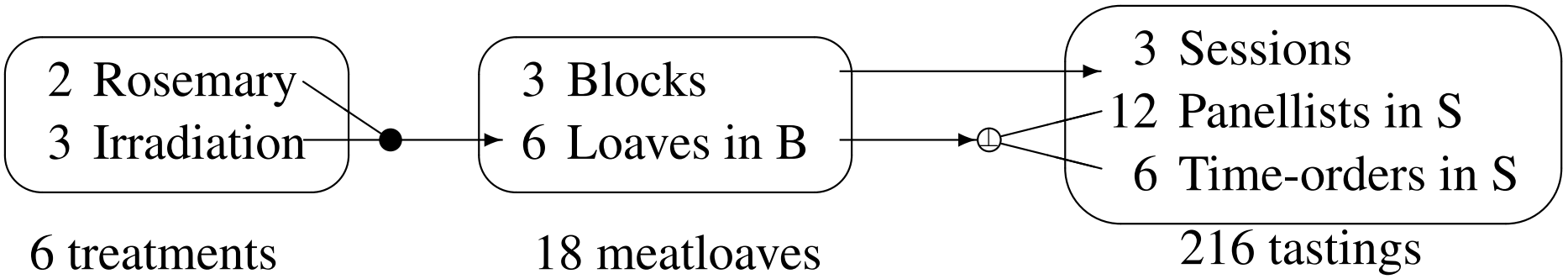
Meatloaves (T. B. Bailey): a tasting experiment



The systematic design for Phase II (a pair of 6×6 Latin squares) is randomized by randomly permuting sessions, randomly permuting panellists within each session, and randomly permuting time-orders within each session. This gives the covariance model

$$\xi_0 \mathbf{P}_0 + \xi_S \mathbf{P}_S + \xi_{SP} \mathbf{P}_{SP} + \xi_{ST} \mathbf{P}_{ST} + \xi_{SPT} \mathbf{P}_{SPT}.$$

Meatloaves (T. B. Bailey): a tasting experiment



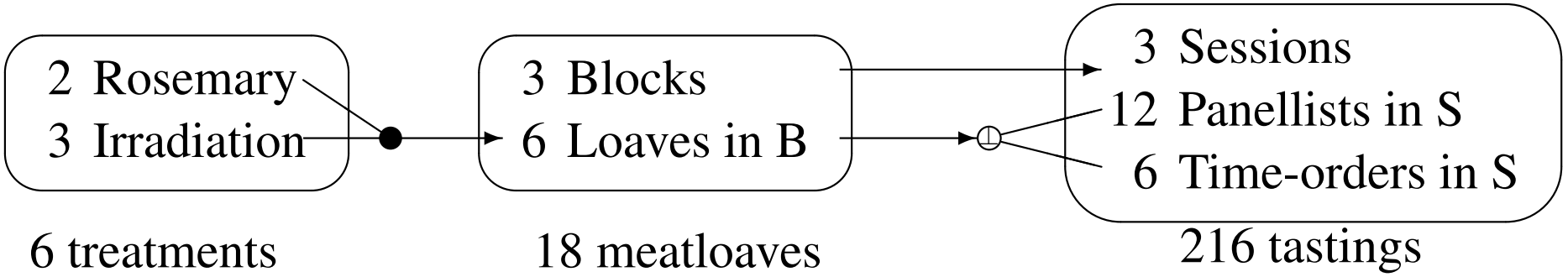
The systematic design for Phase II (a pair of 6×6 Latin squares) is randomized by randomly permuting sessions, randomly permuting panellists within each session, and randomly permuting time-orders within each session. This gives the covariance model

$$\xi_0 \mathbf{P}_0 + \xi_S \mathbf{P}_S + \xi_{SP} \mathbf{P}_{SP} + \xi_{ST} \mathbf{P}_{ST} + \xi_{SPT} \mathbf{P}_{SPT}.$$

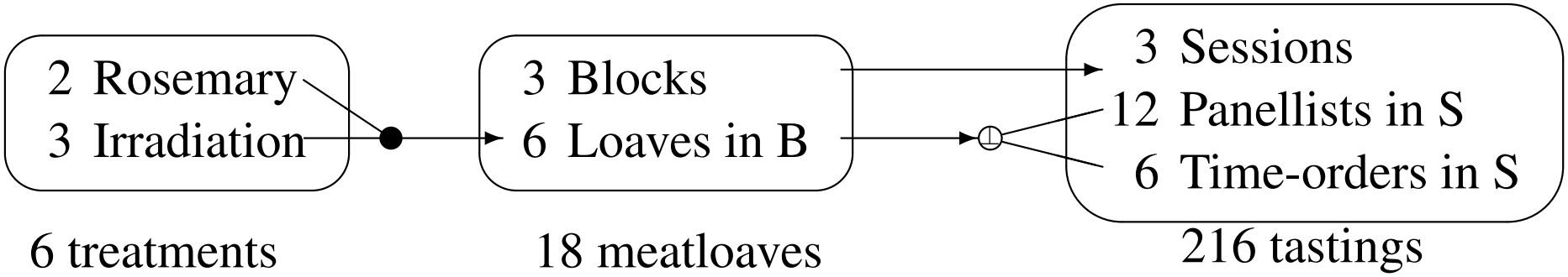
Because the design at Phase II has equal replication 12,
the overall covariance matrix for the 216 responses on tastings is

$$\xi_0 \mathbf{P}_0 + \xi_S \mathbf{P}_S + \xi_{SP} \mathbf{P}_{SP} + \xi_{ST} \mathbf{P}_{ST} + \xi_{SPT} \mathbf{P}_{SPT} + 12\eta_0 \tilde{\mathbf{Q}}_0 + 12\eta_B \tilde{\mathbf{Q}}_B + 12\eta_{BL} \tilde{\mathbf{Q}}_{BL}$$

Meatloaves: skeleton anova

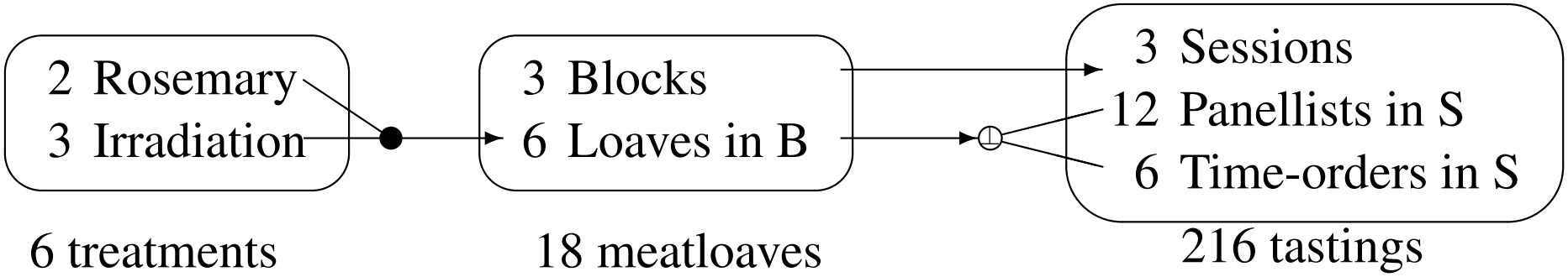


Meatloaves: skeleton anova



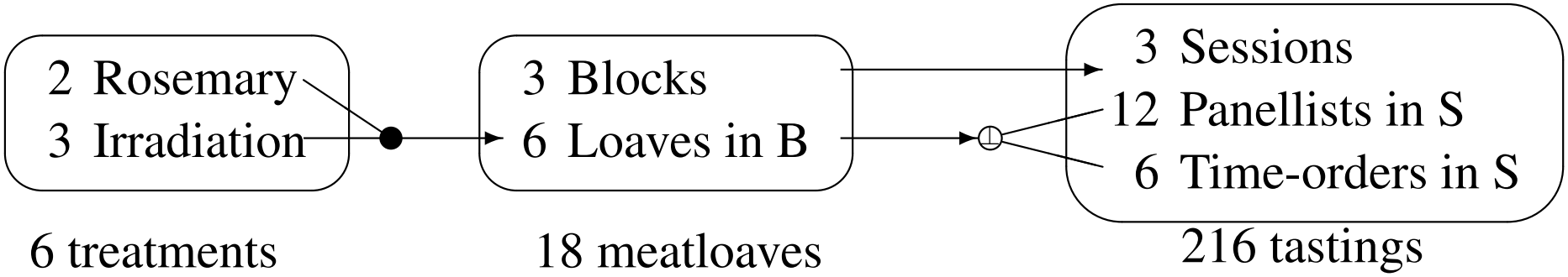
tastings				
source	df			
Mean	1			
Sessions	2			
Panellists[S]	33			
Time-orders[S]	15			
P#T[S]	165			

Meatloaves: skeleton anova



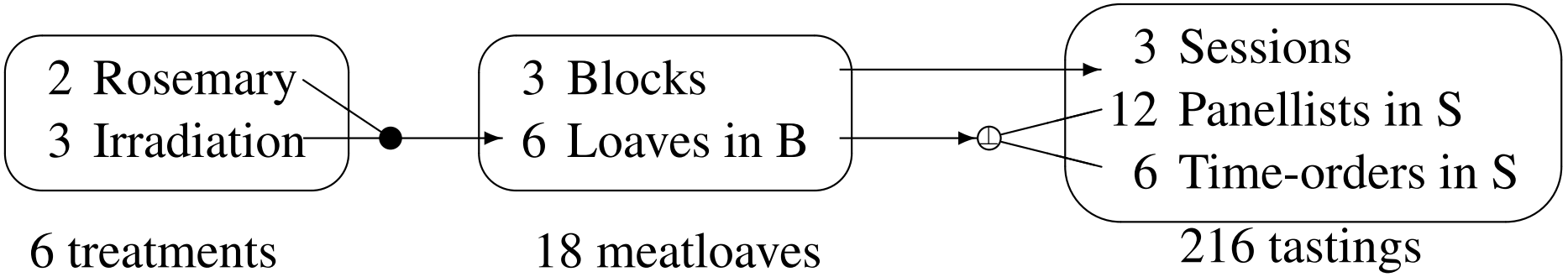
tastings				
source	df			EMS
Mean	1			ξ_0
Sessions	2			ξ_S
Panellists[S]	33			ξ_{SP}
Time-orders[S]	15			ξ_{ST}
P#T[S]	165			ξ_{SPT}
				ξ_{SPT}
				ξ_{SPT}
				ξ_{SPT}
				ξ_{SPT}

Meatloaves: skeleton anova



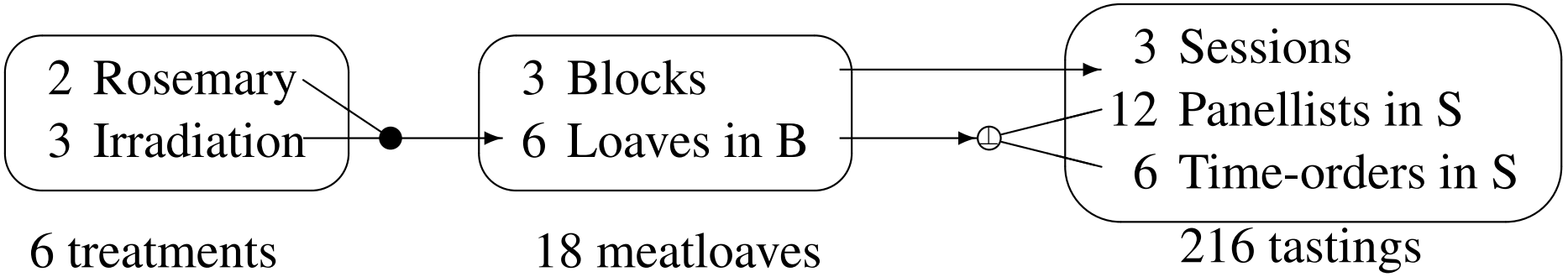
tastings		meatloaves			
source	df	source	df		EMS
Mean	1				ξ_0
Sessions	2				ξ_S
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165				ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}

Meatloaves: skeleton anova



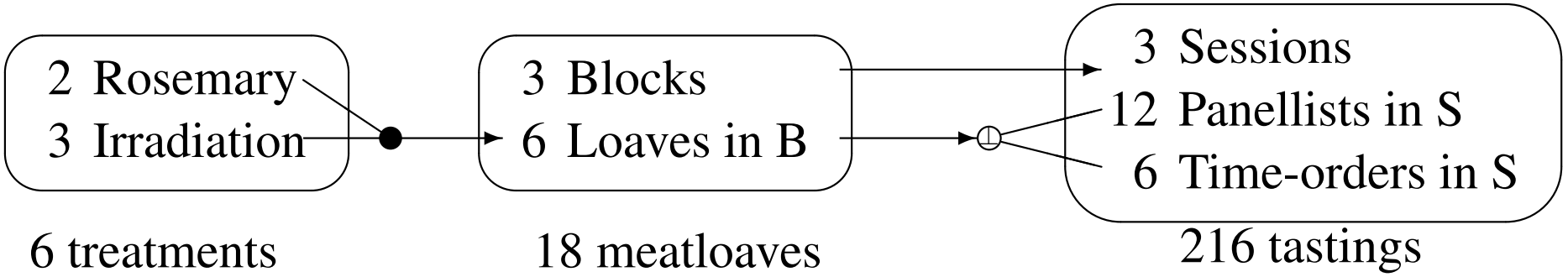
tastings		meatloaves			
source	df	source	df		EMS
Mean	1	Mean	1		ξ_0
Sessions	2				ξ_S
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165				ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}

Meatloaves: skeleton anova



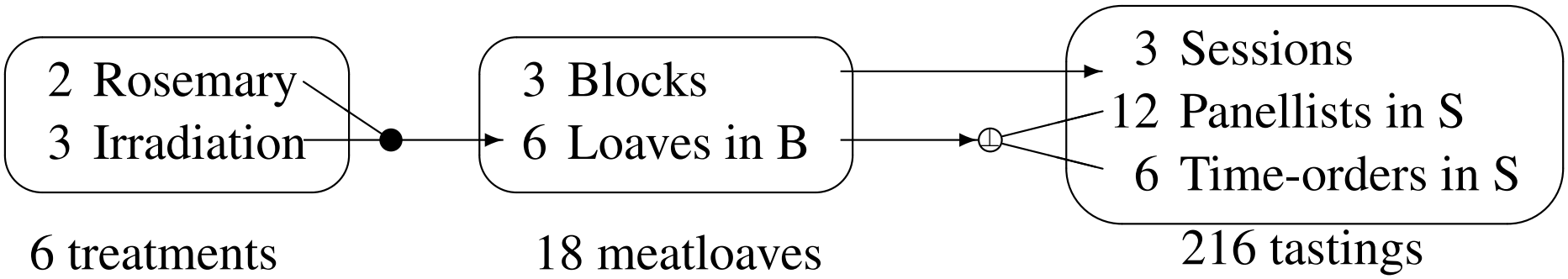
tastings		meatloaves			
source	df	source	df		EMS
Mean	1	Mean	1		ξ_0
Sessions	2	Blocks	2		ξ_S
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165				ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}

Meatloaves: skeleton anova



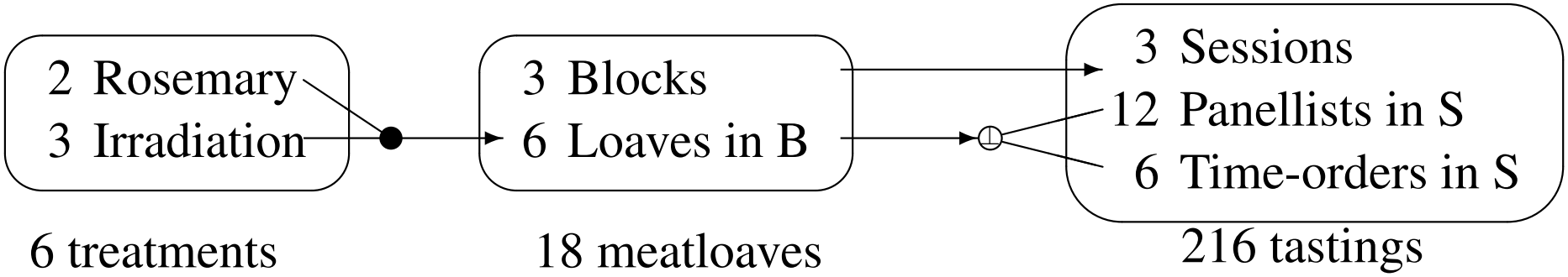
tastings		meatloaves			
source	df	source	df		EMS
Mean	1	Mean	1		ξ_0
Sessions	2	Blocks	2		ξ_S
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165	Loaves[B]	15		ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}

Meatloaves: skeleton anova



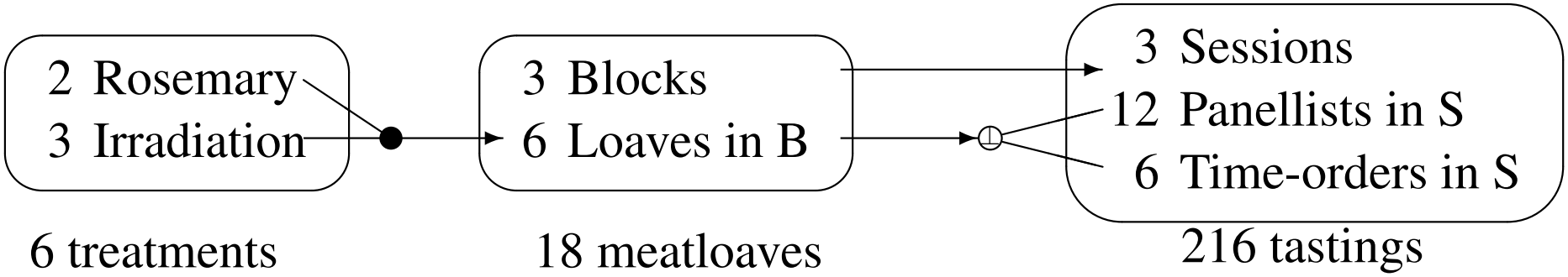
tastings		meatloaves			
source	df	source	df		EMS
Mean	1	Mean	1		ξ_0
Sessions	2	Blocks	2		ξ_S
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165	Loaves[B]	15		ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
					ξ_{SPT}
		Residual	150		ξ_{SPT}

Meatloaves: skeleton anova



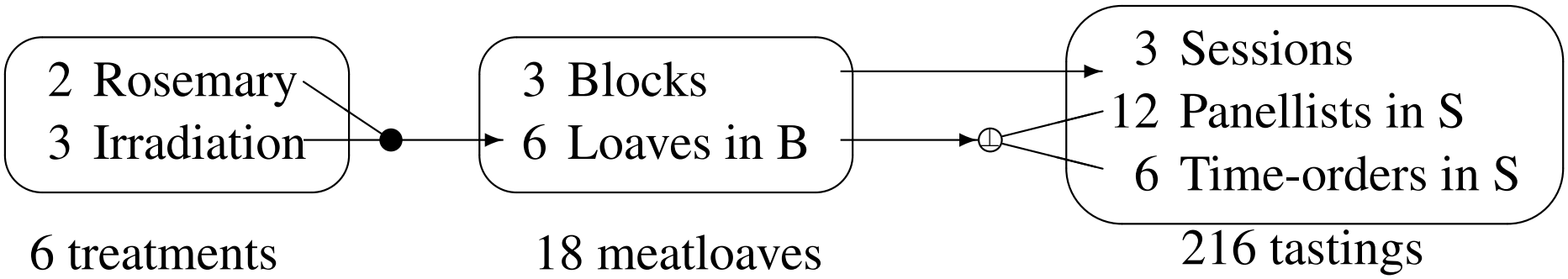
tastings		meatloaves			
source	df	source	df		EMS
Mean	1	Mean	1		$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2		$\xi_S + 12\eta_B$
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165	Loaves[B]	15		$\xi_{SPT} + 12\eta_{BL}$
					$\xi_{SPT} + 12\eta_{BL}$
					$\xi_{SPT} + 12\eta_{BL}$
					$\xi_{SPT} + 12\eta_{BL}$
		Residual	150		ξ_{SPT}

Meatloaves: skeleton anova



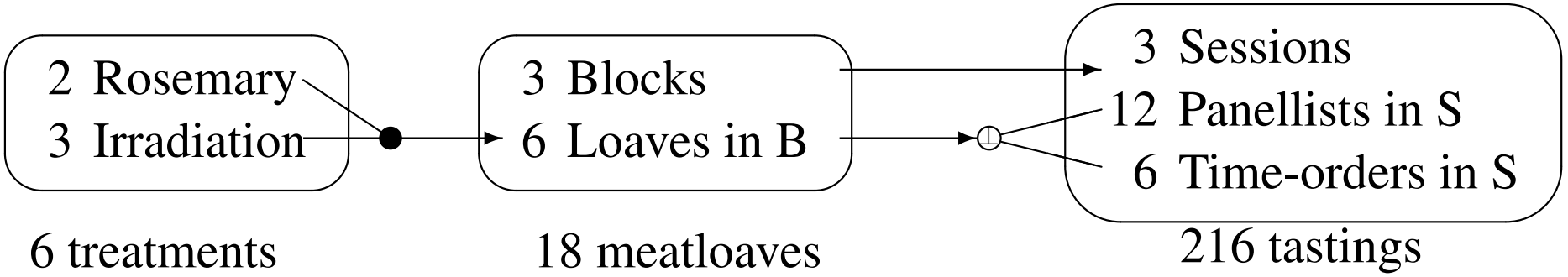
tastings		meatloaves		treatments	
source	df	source	df	source	df	EMS
Mean	1	Mean	1		$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2		$\xi_S + 12\eta_B$
Panellists[S]	33				ξ_{SP}
Time-orders[S]	15				ξ_{ST}
P#T[S]	165	Loaves[B]	15		$\xi_{SPT} + 12\eta_{BL}$
					$\xi_{SPT} + 12\eta_{BL}$
					$\xi_{SPT} + 12\eta_{BL}$
					$\xi_{SPT} + 12\eta_{BL}$
		Residual	150		ξ_{SPT}

Meatloaves: skeleton anova



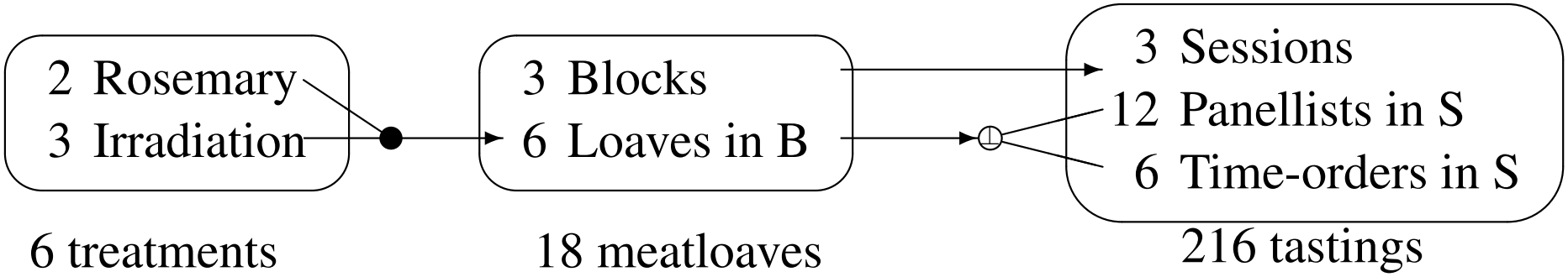
tastings		meatloaves		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2			$\xi_S + 12\eta_B$
Panellists[S]	33					ξ_{SP}
Time-orders[S]	15					ξ_{ST}
P#T[S]	165	Loaves[B]	15			$\xi_{SPT} + 12\eta_{BL}$
						$\xi_{SPT} + 12\eta_{BL}$
						$\xi_{SPT} + 12\eta_{BL}$
						$\xi_{SPT} + 12\eta_{BL}$
		Residual	150			ξ_{SPT}

Meatloaves: skeleton anova



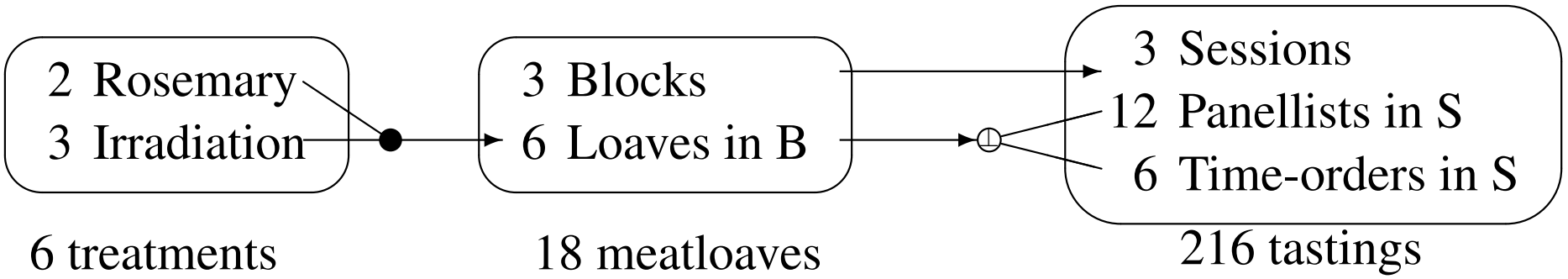
tastings		meatloaves		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2			$\xi_S + 12\eta_B$
Panellists[S]	33					ξ_{SP}
Time-orders[S]	15					ξ_{ST}
P#T[S]	165	Loaves[B]	15	Rosemary	1	$\xi_{SPT} + 12\eta_{BL} + q(R)$
						$\xi_{SPT} + 12\eta_{BL}$
						$\xi_{SPT} + 12\eta_{BL}$
						$\xi_{SPT} + 12\eta_{BL}$
		Residual	150			ξ_{SPT}

Meatloaves: skeleton anova



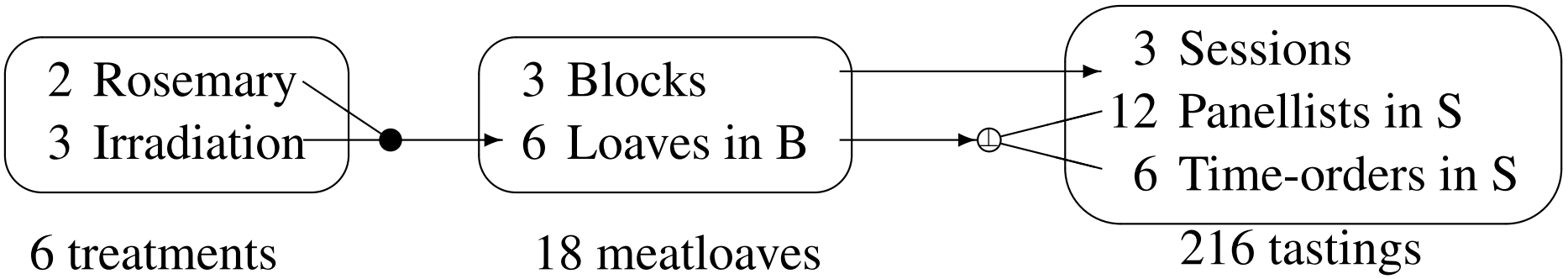
tastings		meatloaves		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2			$\xi_S + 12\eta_B$
Panellists[S]	33					ξ_{SP}
Time-orders[S]	15					ξ_{ST}
P#T[S]	165	Loaves[B]	15	Rosemary	1	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{R})$
				Irradiation	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{I})$
						$\xi_{SPT} + 12\eta_{BL}$
						$\xi_{SPT} + 12\eta_{BL}$
		Residual	150			ξ_{SPT}

Meatloaves: skeleton anova



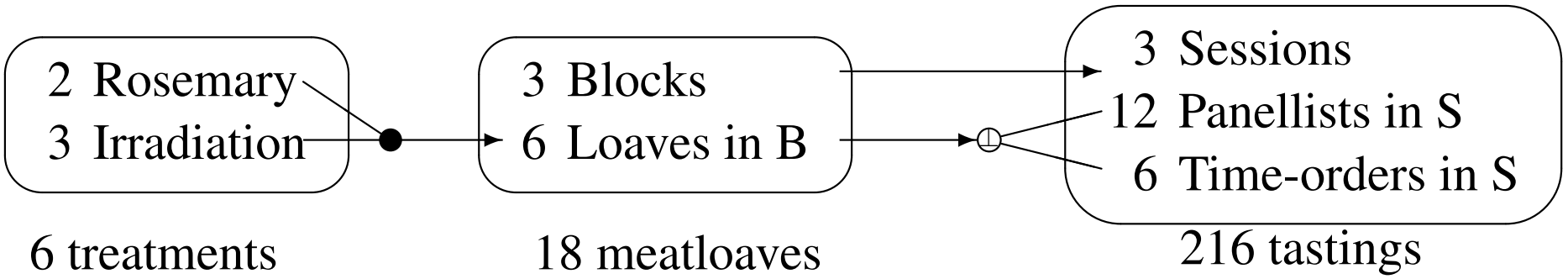
tastings		meatloaves		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2			$\xi_S + 12\eta_B$
Panellists[S]	33					ξ_{SP}
Time-orders[S]	15					ξ_{ST}
P#T[S]	165	Loaves[B]	15	Rosemary	1	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{R})$
				Irradiation	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{I})$
				R# I	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{RI})$
						$\xi_{SPT} + 12\eta_{BL}$
		Residual	150			ξ_{SPT}

Meatloaves: skeleton anova



tastings		meatloaves		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2			$\xi_S + 12\eta_B$
Panellists[S]	33					ξ_{SP}
Time-orders[S]	15					ξ_{ST}
P#T[S]	165	Loaves[B]	15	Rosemary	1	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{R})$
				Irradiation	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{I})$
				R# I	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{RI})$
				Residual	10	$\xi_{SPT} + 12\eta_{BL}$
		Residual	150			ξ_{SPT}

Meatloaves: skeleton anova



tastings		meatloaves		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 12\eta_0$
Sessions	2	Blocks	2			$\xi_S + 12\eta_B$
Panellists[S]	33					ξ_{SP}
Time-orders[S]	15					ξ_{ST}
P#T[S]	165	Loaves[B]	15	Rosemary	1	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{R})$
				Irradiation	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{I})$
				R# I	2	$\xi_{SPT} + 12\eta_{BL} + q(\mathrm{RI})$
				Residual	10	$\xi_{SPT} + 12\eta_{BL}$
		Residual	150			ξ_{SPT}

Residual degrees of freedom

Lesson

If treatments are applied in Phase I,
the number of degrees of freedom for the relevant residual
cannot increase in Phase II.

Residual degrees of freedom

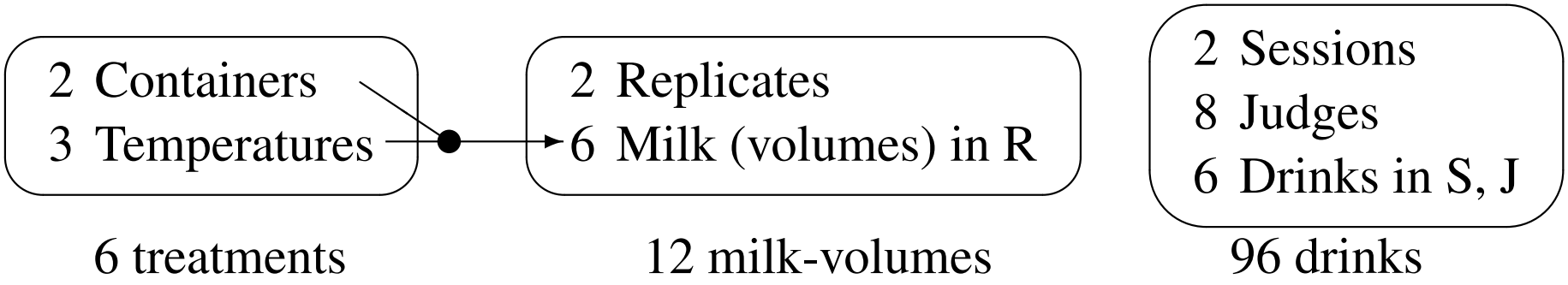
Lesson

If treatments are applied in Phase I,
the number of degrees of freedom for the relevant residual
cannot increase in Phase II.

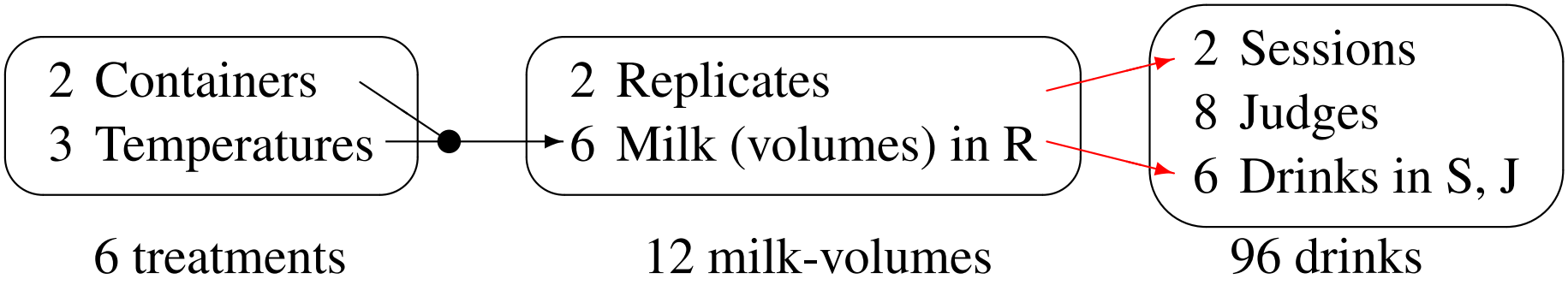
Principle

If treatments are orthogonal to 'large blocks' in Phase I,
then those large blocks should be confounded with
"large blocks" in Phase II.

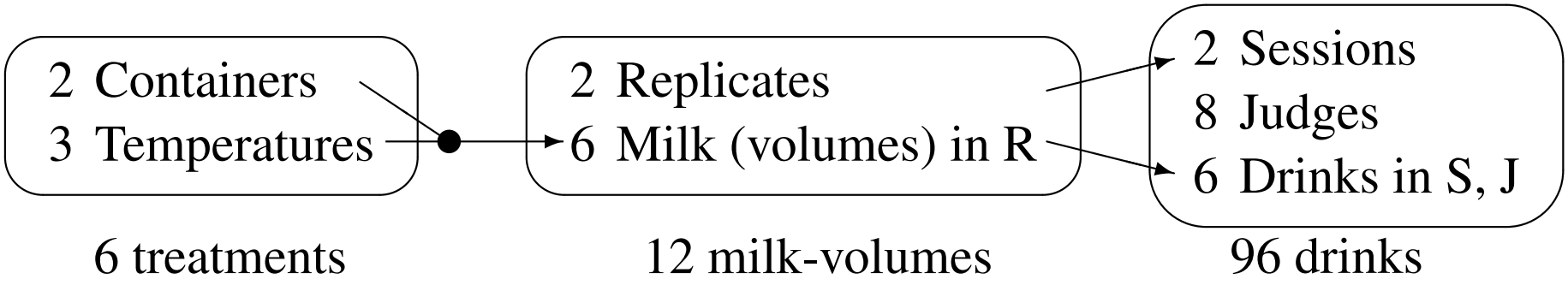
Milk storage (Wood, Willams and Speed)



Milk storage (Wood, Willams and Speed)

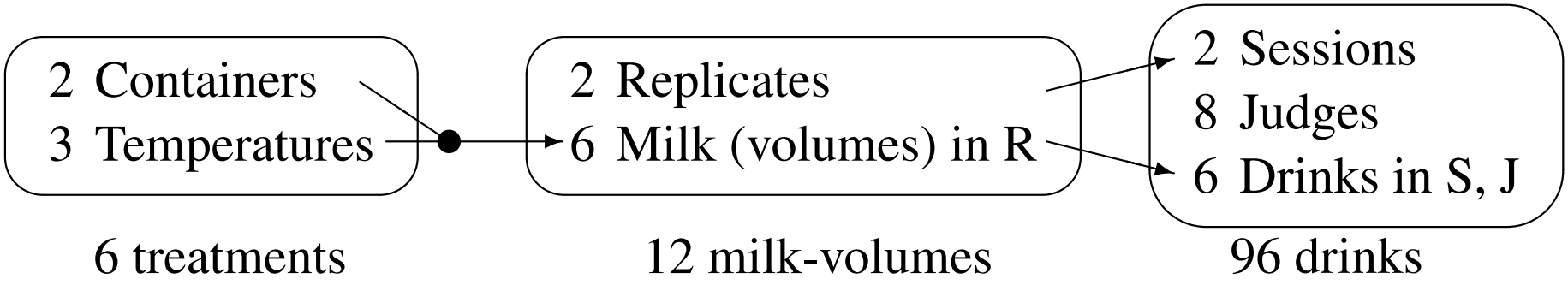


Milk storage (Wood, Willams and Speed)



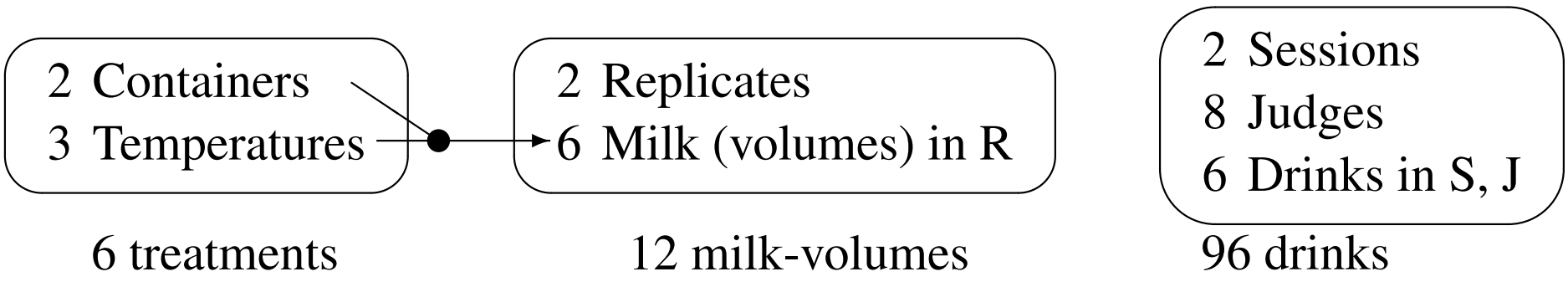
drinks		milk-volumes		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1	Replicates	1			$\xi_S + 8\eta_R$
Judges	7					ξ_J
S# J	7					ξ_{SJ}
Drinks[S∧J]	80	Milk[R]	10	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
				Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
				C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
				Residual	5	$\xi_{SJD} + 8\eta_{RM}$
		Residual	70			ξ_{SJD}

Milk storage (Wood, Willams and Speed)

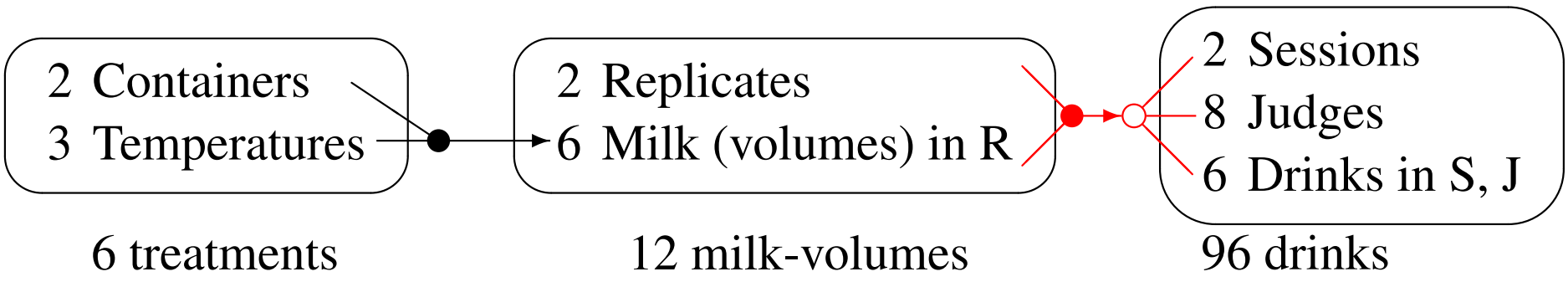


drinks		milk-volumes		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1	Replicates	1			$\xi_S + 8\eta_R$
Judges	7					ξ_J
S# J	7					ξ_{SJ}
Drinks[S∧J]	80	Milk[R]	10	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
				Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
				C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
				Residual	5	$\xi_{SJD} + 8\eta_{RM}$
		Residual	70			ξ_{SJD}

Milk storage (Wood, Willams and Speed): design 2



Milk storage (Wood, Willams and Speed): design 2



Milk storage (Wood, Willams and Speed): anova 2

drinks		milk-volumes			treatments		
source	df	eff	source	df	source	df	EMS
Mean	1	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1						ξ_S
Judges	7						ξ_J
S#J	7	$\frac{1}{3}$	Milk[R]$_1$	3			$\xi_{SJ} + \frac{8}{3}\eta_{RM}$
			Residual	4			ξ_{SJ}
Drinks[S$\wedge$J]	80	1	Replicates	1			$\xi_{SJD} + 8\eta_R$
		$\frac{2}{3}$	Milk[R]$_1$	3			$\xi_{SJD} + \frac{16}{3}\eta_{RM}$
		1	Milk[R]$_2$	7	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
					Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
					C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
					Residual	2	$\xi_{SJD} + 8\eta_{RM}$
			Residual	69			ξ_{SJD}

Milk storage (Wood, Willams and Speed): anova 2

drinks		milk-volumes			treatments		
source	df	eff	source	df	source	df	EMS
Mean	1	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1						ξ_S
Judges	7						ξ_J
S#J	7	$\frac{1}{3}$	Milk[R]$_1$	3			$\xi_{SJ} + \frac{8}{3}\eta_{RM}$
			Residual	4			ξ_{SJ}
Drinks[S$\wedge$J]	80	1	Replicates	1			$\xi_{SJD} + 8\eta_R$
		$\frac{2}{3}$	Milk[R]$_1$	3			$\xi_{SJD} + \frac{16}{3}\eta_{RM}$
		1	Milk[R]$_2$	7	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
					Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
					C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
					Residual	2	$\xi_{SJD} + 8\eta_{RM}$
			Residual	69			ξ_{SJD}

Milk storage (Wood, Willams and Speed): anova 2

drinks		milk-volumes			treatments		
source	df	eff	source	df	source	df	EMS
Mean	1	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1						ξ_S
Judges	7						ξ_J
S#J	7	$\frac{1}{3}$	Milk[R]$_1$	3			$\xi_{SJ} + \frac{8}{3}\eta_{RM}$
			Residual	4			ξ_{SJ}
Drinks[S$\wedge$J]	80	1	Replicates	1			$\xi_{SJD} + 8\eta_R$
		$\frac{2}{3}$	Milk[R]$_1$	3			$\xi_{SJD} + \frac{16}{3}\eta_{RM}$
		1	Milk[R]$_2$	7	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
					Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
					C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
					Residual	2	$\xi_{SJD} + 8\eta_{RM}$
			Residual	69			ξ_{SJD}

Milk storage (Wood, Willams and Speed): anova 2

drinks		milk-volumes			treatments		
source	df	eff	source	df	source	df	EMS
Mean	1	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1						ξ_S
Judges	7						ξ_J
S#J	7	$\frac{1}{3}$	Milk[R]$_1$	3			$\xi_{SJ} + \frac{8}{3}\eta_{RM}$
			Residual	4			ξ_{SJ}
Drinks[S$\wedge$J]	80	1	Replicates	1			$\xi_{SJD} + 8\eta_R$
		$\frac{2}{3}$	Milk[R]$_1$	3			$\xi_{SJD} + \frac{16}{3}\eta_{RM}$
		1	Milk[R]$_2$	7	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
					Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
					C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
					Residual	2	$\xi_{SJD} + 8\eta_{RM}$
			Residual	69			ξ_{SJD}

efficiency factor

Milk storage (Wood, Willams and Speed): anova 2

drinks		milk-volumes			treatments		
source	df	eff	source	df	source	df	EMS
Mean	1	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1						ξ_S
Judges	7						ξ_J
S#J	7	$\frac{1}{3}$	Milk[R]$_1$	3			$\xi_{SJ} + \frac{8}{3}\eta_{RM}$
			Residual	4			ξ_{SJ}
Drinks[S$\wedge$J]	80	1	Replicates	1			$\xi_{SJD} + 8\eta_R$
		$\frac{2}{3}$	Milk[R]$_1$	3			$\xi_{SJD} + \frac{16}{3}\eta_{RM}$
		1	Milk[R]$_2$	7	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
					Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
					C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
					Residual	2	$\xi_{SJD} + 8\eta_{RM}$
			Residual	69			ξ_{SJD}

efficiency factor

Milk storage (Wood, Willams and Speed): anova 2

drinks		milk-volumes			treatments		
source	df	eff	source	df	source	df	EMS
Mean	1	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Sessions	1						ξ_S
Judges	7						ξ_J
S#J	7	$\frac{1}{3}$	Milk[R]$_1$	3			$\xi_{SJ} + \frac{8}{3}\eta_{RM}$
			Residual	4			ξ_{SJ}
Drinks[S$\wedge$J]	80	1	Replicates	1			$\xi_{SJD} + 8\eta_R$
		$\frac{2}{3}$	Milk[R]$_1$	3			$\xi_{SJD} + \frac{16}{3}\eta_{RM}$
		1	Milk[R]$_2$	7	Containers	1	$\xi_{SJD} + 8\eta_{RM} + q(C)$
					Temperatures	2	$\xi_{SJD} + 8\eta_{RM} + q(T)$
					C#T	2	$\xi_{SJD} + 8\eta_{RM} + q(CT)$
					Residual	2	$\xi_{SJD} + 8\eta_{RM}$
			Residual	69			ξ_{SJD}

efficiency factor

Factors which are ‘hard to set’ or which must be applied to large areas

Lesson

If a treatment factor is ‘hard to set’in Phase I,
then it is probably in a Phase I stratum with large variance.
Stratum variances from the two phases are added.

Factors which are 'hard to set' or which must be applied to large areas

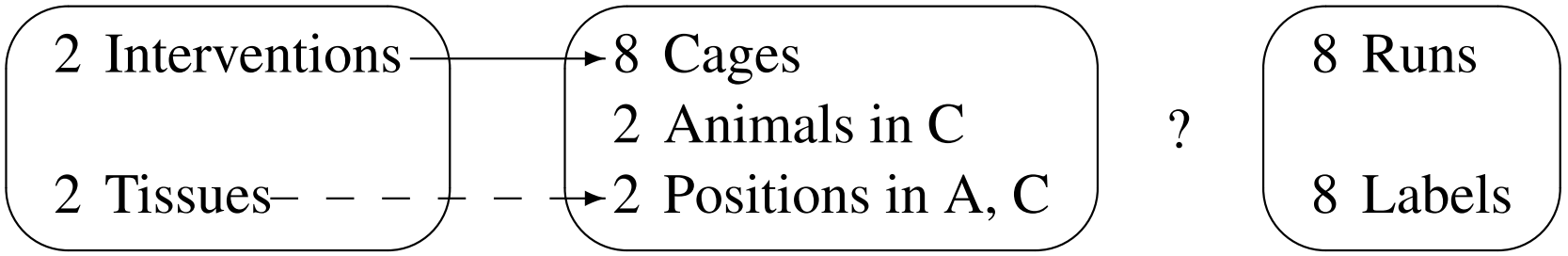
Lesson

If a treatment factor is 'hard to set'in Phase I,
then it is probably in a Phase I stratum with large variance.
Stratum variances from the two phases are added.

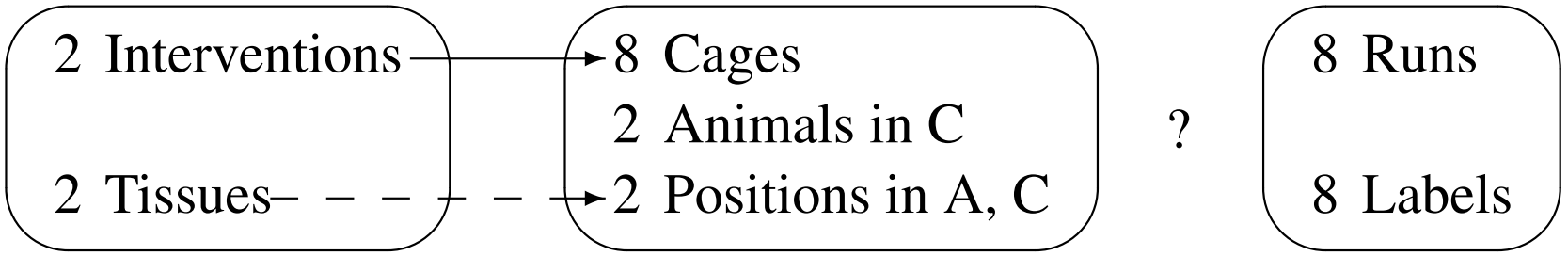
Principle

If a treatment factor is 'hard to set' in Phase I,
then it should be allocated to a Phase II stratum with small variance.

Proteomics (Ruggiero)

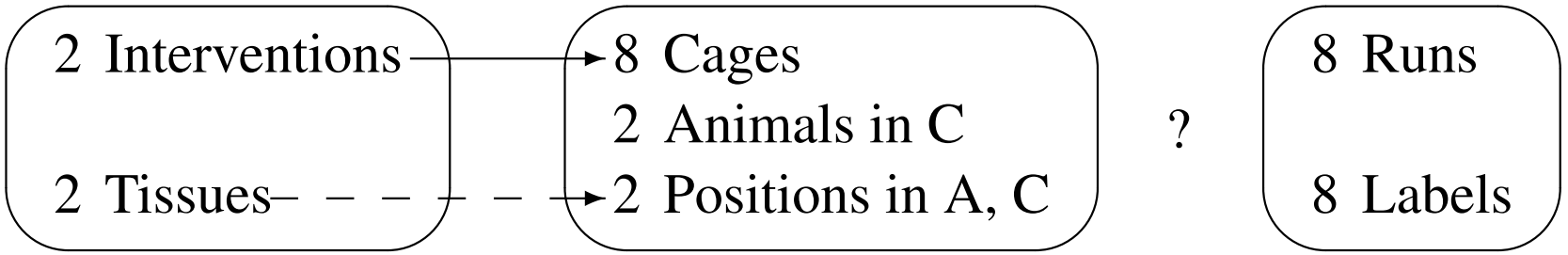


Proteomics (Ruggiero)



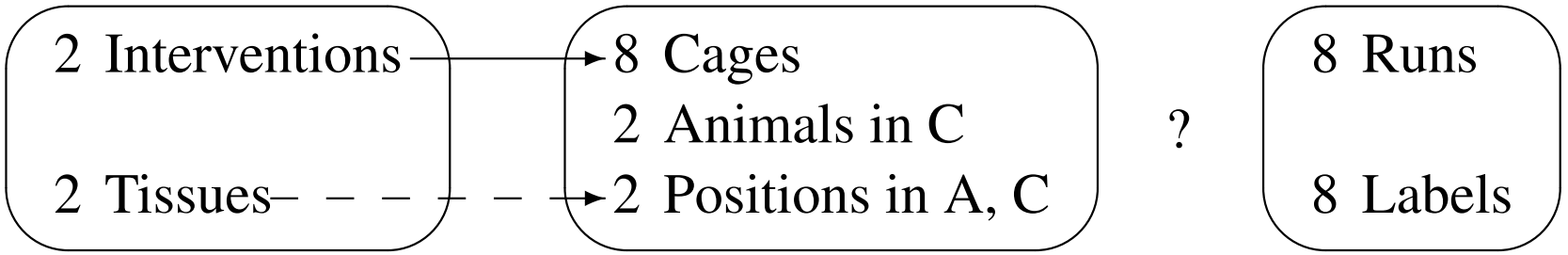
- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.

Proteomics (Ruggiero)



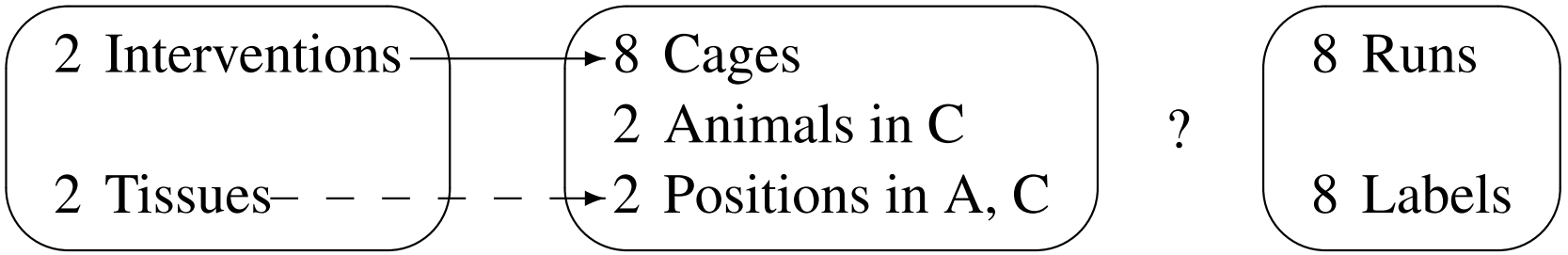
- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.

Proteomics (Ruggiero)



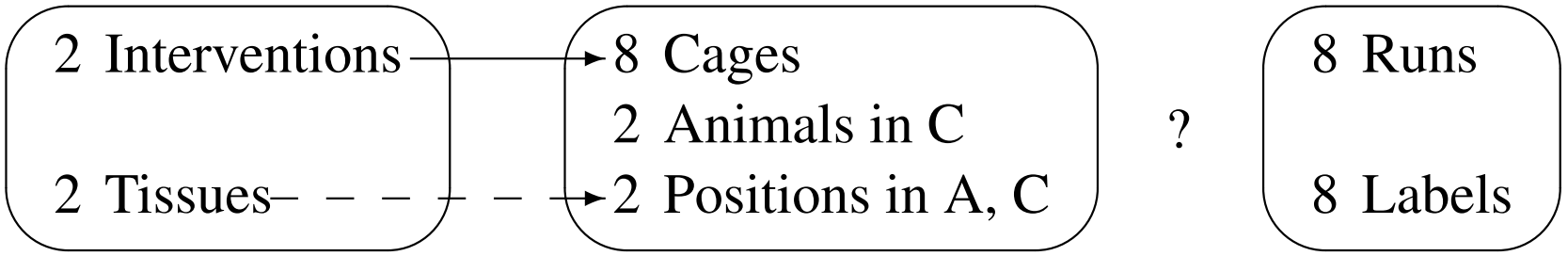
- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Proteomics (Ruggiero)



- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

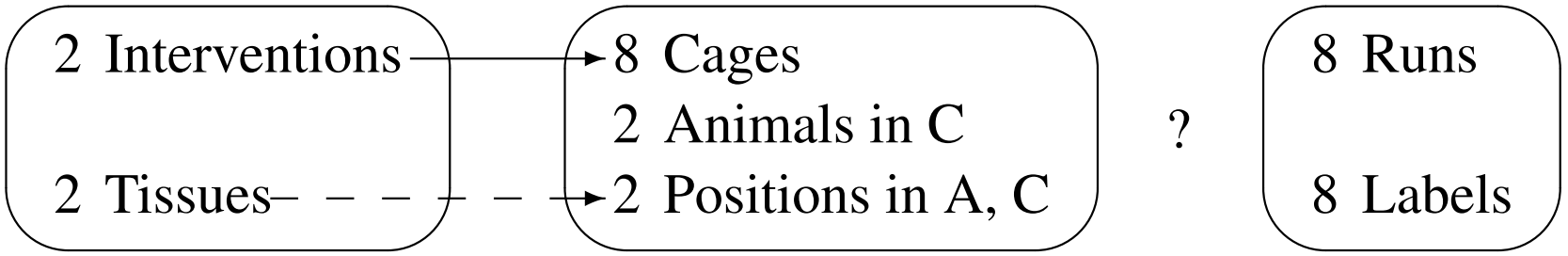
Proteomics (Ruggiero)



- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

Proteomics (Ruggiero)

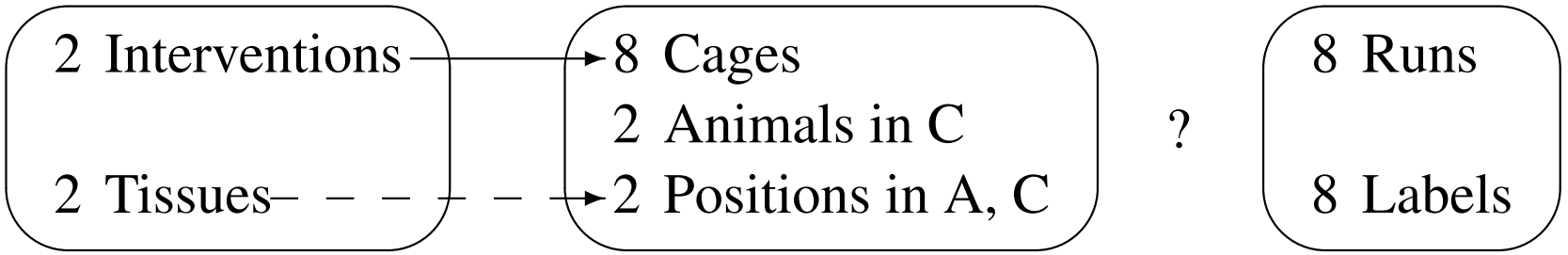


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

I	C_1 C_2 C_3	R_1 R_2 R_3
	A	
T	P	L_1 L_2 L_3

Proteomics (Ruggiero)

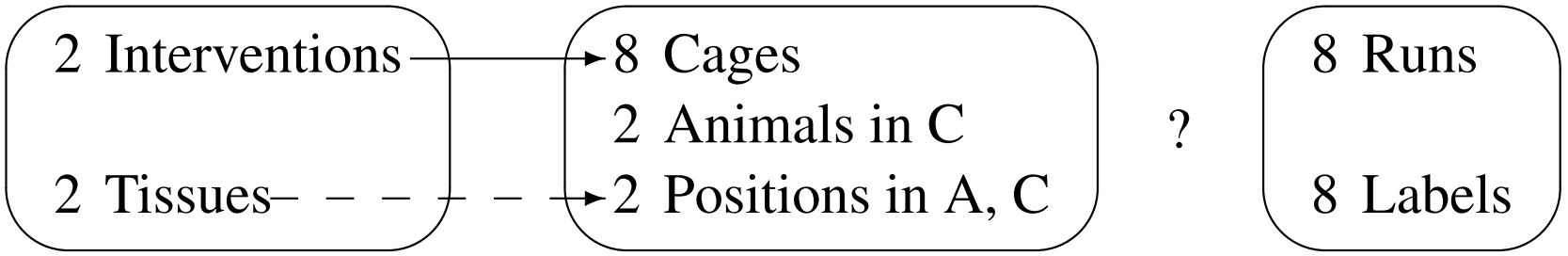


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

I	$I \equiv C_1$	C_1	C_2	C_3		R_1	R_2 R_3
			A				
T			P			L_1	L_2 L_3

Proteomics (Ruggiero)

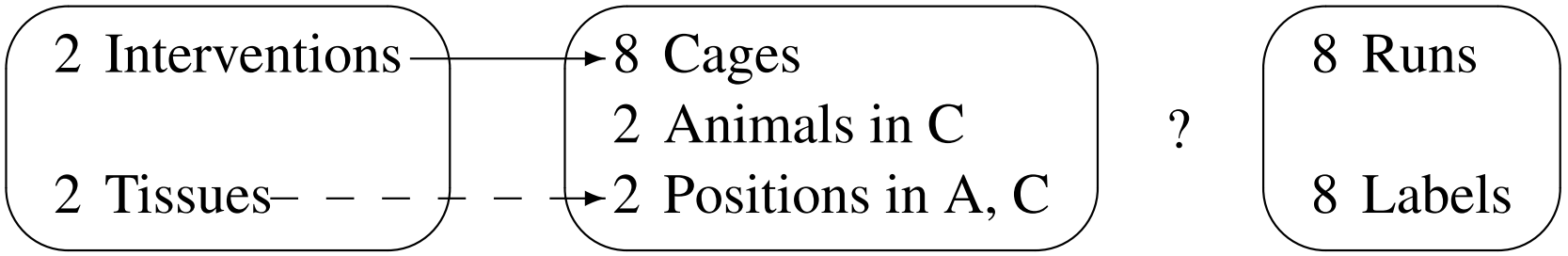


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A		
T		P		L_1 L_2 L_3

Proteomics (Ruggiero)

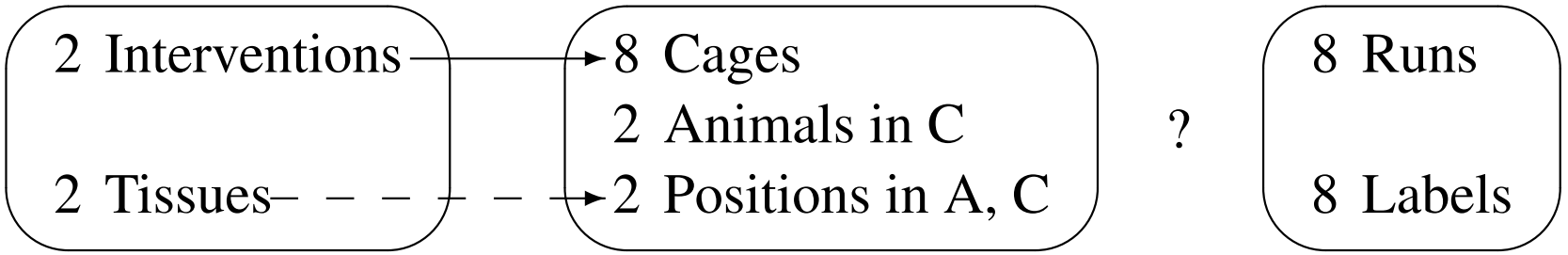


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

I	$I \equiv C_1$	$C_1 \;\; C_2 \;\; C_3$	$C_i \equiv R_i + L_i$	$R_1 \;\; R_2 \;\; R_3$
		A	$A \equiv R_1$	
T		P		$L_1 \;\; L_2 \;\; L_3$

Proteomics (Ruggiero)

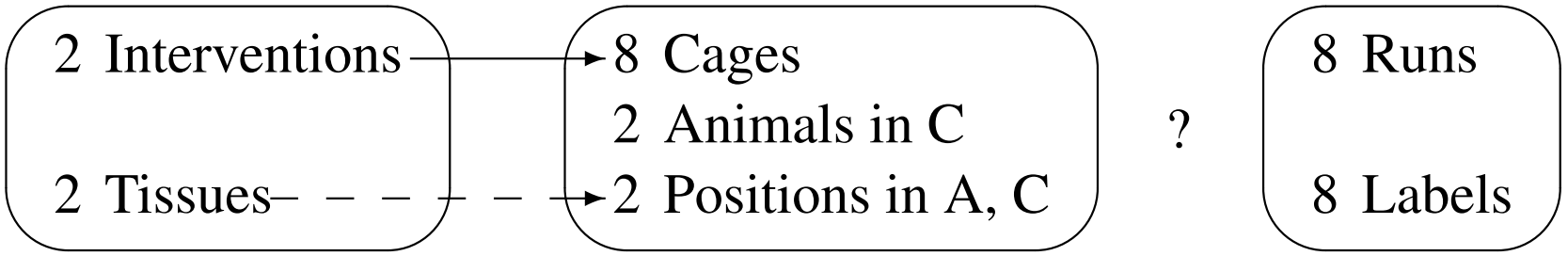


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T		P	$P \equiv L_2$	L_1 L_2 L_3

Proteomics (Ruggiero)

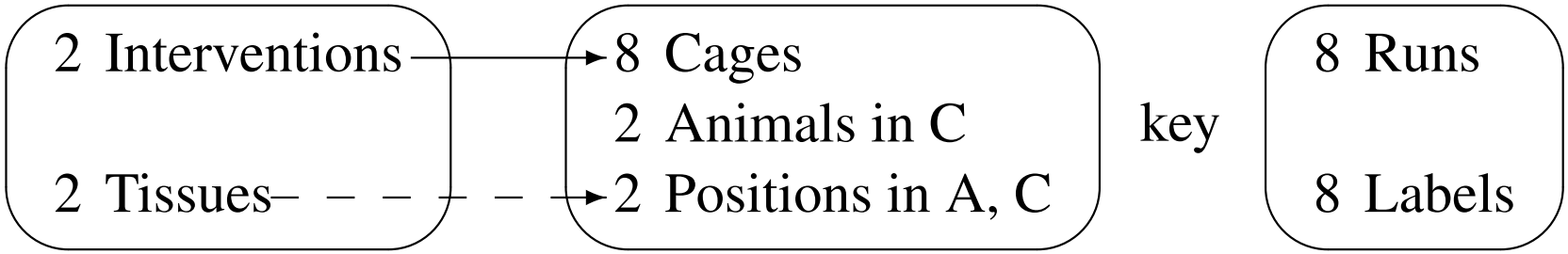


- Interventions probably has the biggest variance from Phase I, so try to confound this with a low-variance term in Phase II.
- If possible, confound the rest of Cages with the same term, to avoid losing degrees of freedom for the residual.
- If possible, make Tissues and I#T orthogonal to Runs and Labels.

Use 2-level pseudofactors and Patterson's design key.

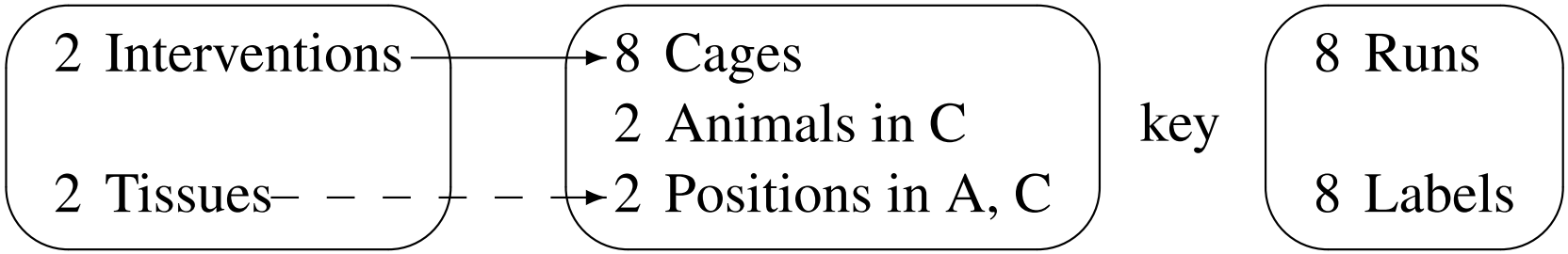
I	$I \equiv C_1$	$C_1 \;\; C_2 \;\; C_3$	$C_i \equiv R_i + L_i$	$R_1 \;\; R_2 \;\; R_3$
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	$L_1 \;\; L_2 \;\; L_3$

Calculations using the design key



I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	L_1 L_2 L_3

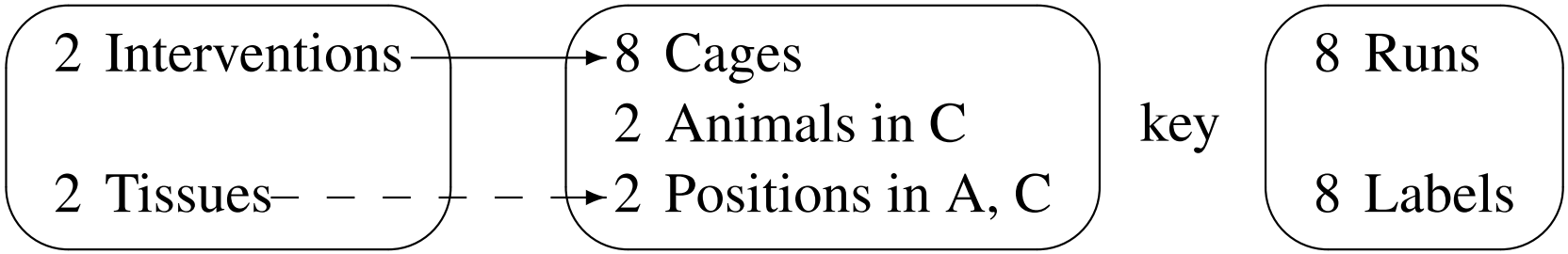
Calculations using the design key



I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	L_1 L_2 L_3

$$P + C_2 = R_2$$

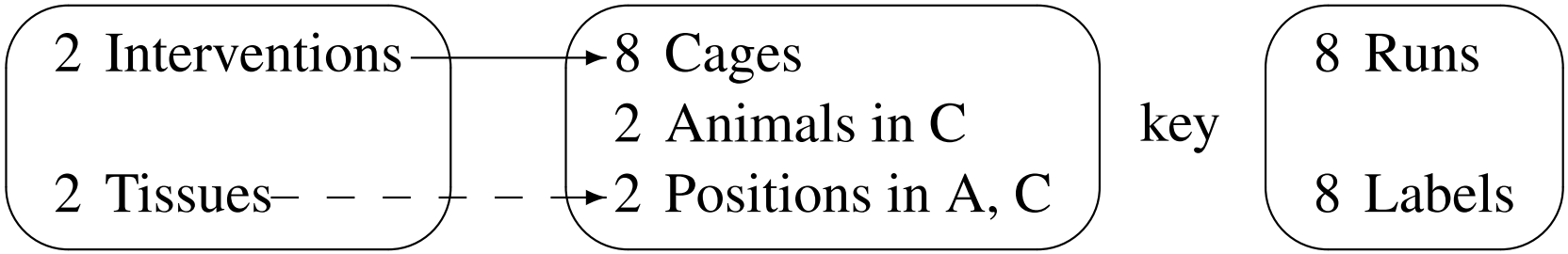
Calculations using the design key



I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	L_1 L_2 L_3

$$P + C_2 = R_2 \qquad \text{Positions[A,C], Runs}$$

Calculations using the design key

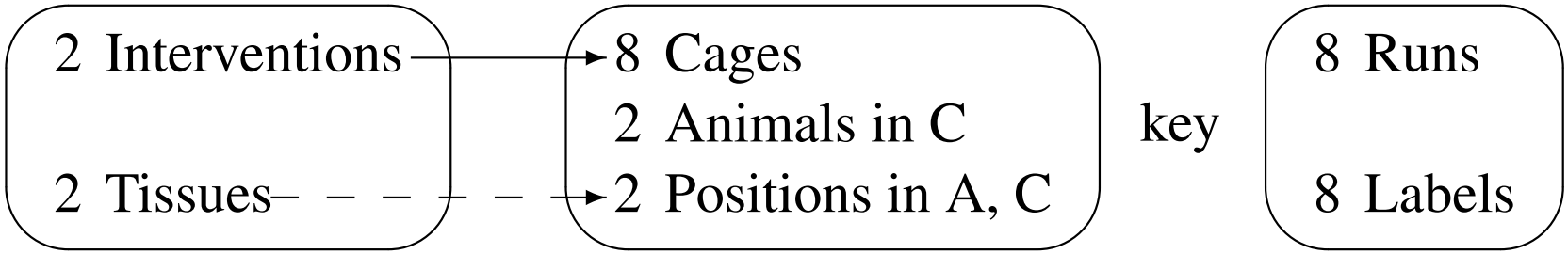


I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	L_1 L_2 L_3

$$P + C_2 = R_2 \qquad \text{Positions[A,C], Runs}$$

$$T \equiv P + C_3 \equiv L_2 + R_3 + L_3$$

Calculations using the design key

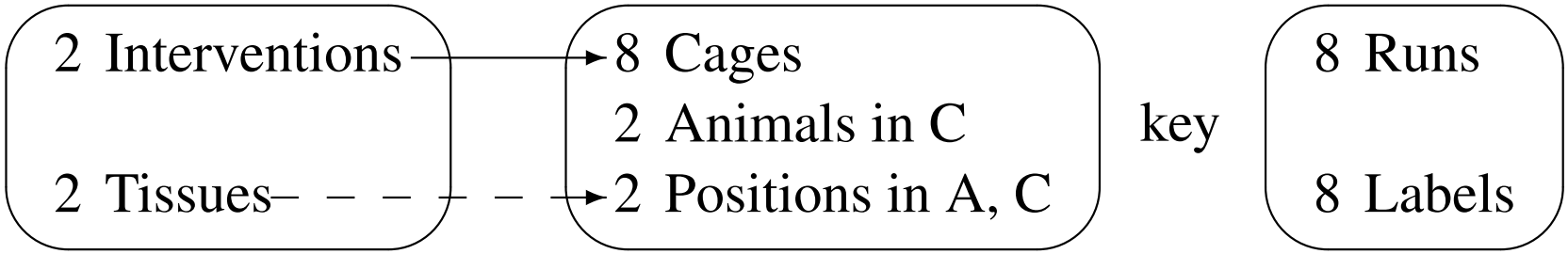


I	$I \equiv C_1$		C_1	C_2	C_3	$C_i \equiv R_i + L_i$	R_1	R_2 R_3
				A		$A \equiv R_1$		
T	$T \equiv P + C_3$			P		$P \equiv L_2$	L_1	L_2 L_3

$$P + C_2 = R_2 \qquad \text{Positions[A,C], Runs}$$

$$T \equiv P + C_3 \equiv L_2 + R_3 + L_3 \qquad \text{T, P[A,C], R\#L}$$

Calculations using the design key



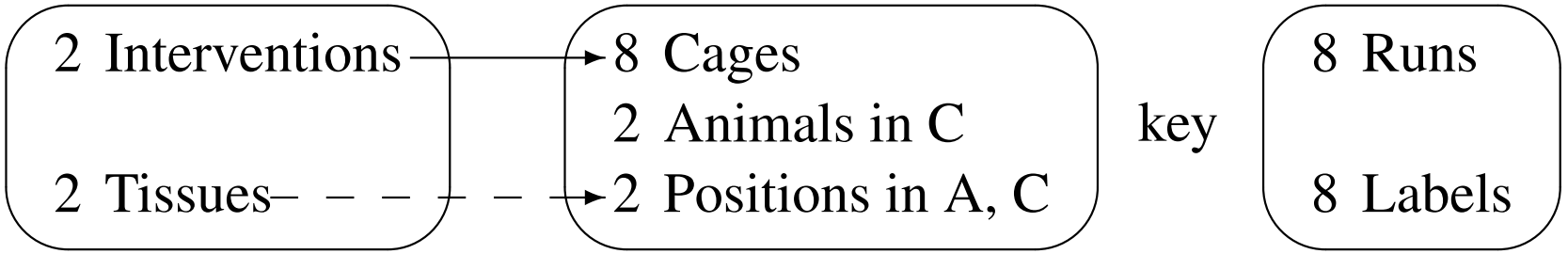
I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	L_1 L_2 L_3

$$P + C_2 = R_2 \qquad \text{Positions[A,C], Runs}$$

$$T \equiv P + C_3 \equiv L_2 + R_3 + L_3 \qquad \text{T, P[A,C], R\#L}$$

$$I + T = C_1 + P + C_3 = R_1 + L_1 + L_2 + R_3 + L_3$$

Calculations using the design key



I	$I \equiv C_1$	C_1 C_2 C_3	$C_i \equiv R_i + L_i$	R_1 R_2 R_3
		A	$A \equiv R_1$	
T	$T \equiv P + C_3$	P	$P \equiv L_2$	L_1 L_2 L_3

$$P + C_2 = R_2 \qquad \text{Positions[A,C], Runs}$$

$$T \equiv P + C_3 \equiv L_2 + R_3 + L_3 \qquad \text{T, P[A,C], R\#L}$$

$$I + T = C_1 + P + C_3 = R_1 + L_1 + L_2 + R_3 + L_3 \qquad \text{I\#T, P[A,C], R\#L}$$

Proteomics: skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(I)$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(T)$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(IT)$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Proteomics: skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(I)$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(T)$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(IT)$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Proteomics: skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(I)$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(T)$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(IT)$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Proteomics: skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(I)$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(T)$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(IT)$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Proteomics: skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(I)$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(T)$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(IT)$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Proteomics: skeleton anova

units		animal-bits		treatments		
source	df	source	df	source	df	EMS
Mean	1	Mean	1	Mean	1	$\xi_0 + 8\eta_0$
Runs	7	Animals[C]$_1$	1			$\xi_R + 2\eta_{CA}$
		Positions[A,C]$_1$	2			$\xi_R + 2\eta_{CAP}$
		Residual	4			ξ_R
Labels	7	Animals[C]$_2$	1			$\xi_L + 2\eta_{CA}$
		Positions[A,C]$_2$	2			$\xi_L + 2\eta_{CAP}$
		Residual	4			ξ_L
R#L	49	Cages	7	Interventions	1	$\xi_{RL} + 2\eta_C + q(I)$
				Residual	6	$\xi_{RL} + 2\eta_C$
		Animals[C]$_3$	6			$\xi_{RL} + 2\eta_{CA}$
		Positions[A,C]$_3$	12	Tissues	1	$\xi_{RL} + 2\eta_{CAP} + q(T)$
				I#T	1	$\xi_{RL} + 2\eta_{CAP} + q(IT)$
				Residual	10	$\xi_{RL} + 2\eta_{CAP}$
		Residual	24			ξ_{RL}

Some lessons

Lesson

If the design in both phases is orthogonal,
the using the design key gives a simple method of establishing the confounding.

Some lessons

Lesson

If the design in both phases is orthogonal,
the using the design key gives a simple method of establishing the confounding.

Lesson

The skeleton anova (decomposition table) shows
EMS for treatment terms and for residual terms,
as well as residual degrees of freedom,
so it is a useful tool for evaluating designs.

Some lessons

Lesson

If the design in both phases is orthogonal,
the using the design key gives a simple method of establishing the confounding.

Lesson

The skeleton anova (decomposition table) shows
EMS for treatment terms and for residual terms,
as well as residual degrees of freedom,
so it is a useful tool for evaluating designs.

Lesson

Equating mean squares to their expectations may give
several inconsistent estimators of the ξ_i and η_j,
each with potentially few degrees of freedom.

Data analysis after Phase II

Query

Is it better to analyse the data with

- *ANOVA (for three tiers)*
- *REML*
  (but beware!—small degrees of freedom can lead to silly results)
- *other mixed model software*
  (can it cope with the confounding?)
  (For example, the single degree of freedom for R_1 is part of both Runs and Animals[Cages]—
  does the software give the same result regardless of which is written down first in the list of random effects?)

?

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

Field then laboratory (Butler)

27 Varieties

V_1 V_2 V_3

3 Rows
3 Columns
9 Plots in R, C

R
C
P_1 P_2

9 Batches

9 Samples in B

B_1 B_2

S_1 S_2

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \quad P_2$

$B_1 \quad B_2$

$S_1 \quad S_2$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \quad P_2$

$R \equiv B_1$
$C \equiv B_2$
$P_i \equiv S_i$

$B_1 \quad B_2$

$S_1 \quad S_2$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \quad P_2$

$R \equiv B_1$
$C \equiv B_2$
$P_i \equiv S_i$

$B_1 \quad B_2$
$S_1 \quad S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Columns	2			$\xi_B + \eta_C$
		R#C	4	V_3	2	$\xi_B + \eta_{RC} + q(V_3)$
				Residual	2	$\xi_B + \eta_{RC}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \quad P_2$

$R \equiv B_1$
$C \equiv B_2$
$P_i \equiv S_i$

$B_1 \quad B_2$
$S_1 \quad S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Columns	2			$\xi_B + \eta_C$
		R#C	4	V_3	2	$\xi_B + \eta_{RC} + q(V_3)$
				Residual	2	$\xi_B + \eta_{RC}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \quad P_2$

$R \equiv B_1$
$C \equiv B_2$
$P_i \equiv S_i$

$B_1 \quad B_2$
$S_1 \quad S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Columns	2			$\xi_B + \eta_C$
		R#C	4	V_3	2	$\xi_B + \eta_{RC} + q(V_3)$
				Residual	2	$\xi_B + \eta_{RC}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \;\; V_2 \;\; V_3$

$V_3 \equiv R + C$

$V_1 \equiv P_1$

$V_2 \equiv P_2$

R

C

$P_1 \;\; P_2$

$B_1 \;\; B_2$

$S_1 \;\; S_2$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \;\; V_2 \;\; V_3$

$V_3 \equiv R + C$

$V_1 \equiv P_1$

$V_2 \equiv P_2$

R

C

$P_1 \;\; P_2$

$R \equiv B_1$

$P_1 \equiv B_2$

$C \equiv S_1,\; P_2 \equiv S_2$

$B_1 \;\; B_2$

$S_1 \;\; S_2$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

V_1 V_2 V_3

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
P_1 P_2

$R \equiv B_1$
$P_1 \equiv B_2$
$C \equiv S_1,\ P_2 \equiv S_2$

B_1 B_2

S_1 S_2

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Plots[R,C]$_1$	6	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	4	$\xi_B + \eta_{RCP}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

V_1 V_2 V_3

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
P_1 P_2

$R \equiv B_1$
$P_1 \equiv B_2$
$C \equiv S_1,\ P_2 \equiv S_2$

B_1 B_2

S_1 S_2

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Plots[R,C]$_1$	6	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	4	$\xi_B + \eta_{RCP}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

V_1 V_2 V_3

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
P_1 P_2

$R \equiv B_1$
$P_1 \equiv B_2$
$C \equiv S_1,\ P_2 \equiv S_2$

B_1 B_2

S_1 S_2

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Rows	2			$\xi_B + \eta_R$
		Plots[R,C]$_1$	6	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	4	$\xi_B + \eta_{RCP}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$

$V_1 \equiv P_1$

$V_2 \equiv P_2$

R

C

$P_1 \quad P_2$

$B_1 \quad B_2$

$S_1 \quad S_2$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \quad V_2 \quad V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \quad P_2$

$R \equiv B_1 + S_2$
$P_1 \equiv B_2$
$C \equiv S_1,\ P_2 \equiv S_2$

$B_1 \quad B_2$

$S_1 \quad S_2$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

V_1 V_2 V_3

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
P_1 P_2

$R \equiv B_1 + S_2$
$P_1 \equiv B_2$
$C \equiv S_1,\ P_2 \equiv S_2$

B_1 B_2

S_1 S_2

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Plots[R,C]$_1$	8	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	6	$\xi_B + \eta_{RCP}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

$V_1 \;\; V_2 \;\; V_3$

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
$P_1 \;\; P_2$

$R \equiv B_1 + S_2$
$P_1 \equiv B_2$
$C \equiv S_1,\; P_2 \equiv S_2$

$B_1 \;\; B_2$

$S_1 \;\; S_2$

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Plots[R,C]$_1$	8	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	6	$\xi_B + \eta_{RCP}$

Field then laboratory (Butler)

27 Varieties

3 Rows
3 Columns
9 Plots in R, C

9 Batches

9 Samples in B

V_1 V_2 V_3

$V_3 \equiv R + C$
$V_1 \equiv P_1$
$V_2 \equiv P_2$

R
C
P_1 P_2

$R \equiv B_1 + S_2$
$P_1 \equiv B_2$
$C \equiv S_1,\ P_2 \equiv S_2$

B_1 B_2

S_1 S_2

samples		plots		varieties		
source	df	source	df	source	df	EMS
Batches	8	Plots[R,C]$_1$	8	V_1	2	$\xi_B + \eta_{RCP} + q(V_1)$
				Residual	6	$\xi_B + \eta_{RCP}$

Consequences

Principle

If a treatment term is in a Phase I stratum with large variance, then it should be allocated to a Phase II stratum with small variance.

Lesson

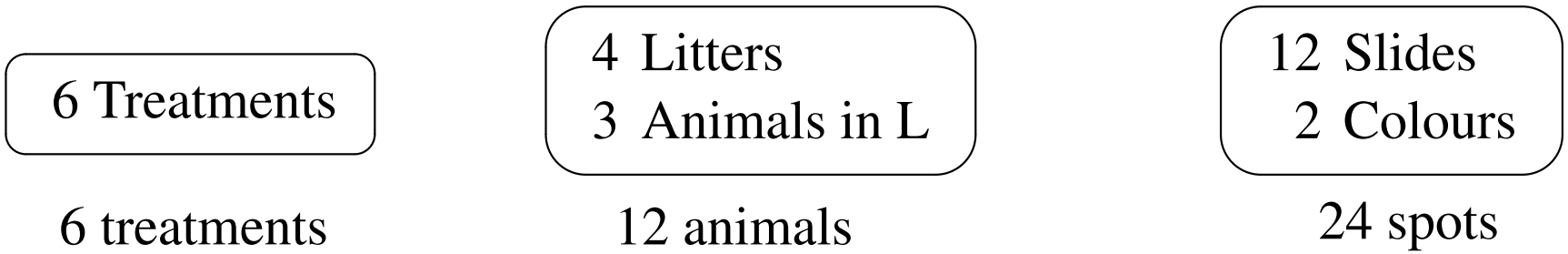
- *This may force sacrificing some information on another treatment term at Phase II.*
- *This may not be practicable.*

Nonorthogonality in Phase I

Query

What should we do if the design used in Phase I is not orthogonal (in the sense that there are efficiency factors other than 0 *and* 1*)?*

Two-colour microarrays (Mead; Ruggiero; Lynch)



Two-colour microarrays (Mead; Ruggiero; Lynch)

6 Treatments	4 Litters 3 Animals in L	12 Slides 2 Colours
6 treatments	12 animals	24 spots

Litter 1	Litter 2	Litter 3	Litter 4
1 2 3	1 4 5	2 4 6	2 5 6

Treatments form three groups of size two: $\{1,6\}, \quad \{2,5\}, \quad \{3,4\}$

Groups are orthogonal to litters;

treatments-within-groups have efficiency factor $\frac{1}{3}$ in litters.

Two-colour microarrays (Mead; Ruggiero; Lynch)

6 Treatments

6 treatments

4 Litters
3 Animals in L

12 animals

12 Slides
2 Colours

24 spots

	Litter 1			Litter 2			Litter 3			Litter 4		
Slide	1	2	3	4	5	6	7	8	9	10	11	12
	1	2	3	1	4	5	2	4	6	2	5	6

Treatments form three groups of size two: $\{1,6\}, \quad \{2,5\}, \quad \{3,4\}$

Groups are orthogonal to litters;

treatments-within-groups have efficiency factor $\frac{1}{3}$ in litters.

Two-colour microarrays (Mead; Ruggiero; Lynch)

6 Treatments	4 Litters 3 Animals in L	12 Slides 2 Colours
6 treatments	12 animals	24 spots

	Litter 1			Litter 2			Litter 3			Litter 4		
Slide	1	2	3	4	5	6	7	8	9	10	11	12
Red	1	2	3	1	4	5	2	4	6	2	5	6
Green												

Treatments form three groups of size two: $\{1,6\}, \quad \{2,5\}, \quad \{3,4\}$

Groups are orthogonal to litters;

treatments-within-groups have efficiency factor $\frac{1}{3}$ in litters.

Two-colour microarrays (Mead; Ruggiero; Lynch)

6 Treatments

6 treatments

4 Litters
3 Animals in L

12 animals

12 Slides
2 Colours

24 spots

	Litter 1			Litter 2			Litter 3			Litter 4		
Slide	1	2	3	4	5	6	7	8	9	10	11	12
Red	1	2	3	1	4	5	2	4	6	2	5	6
Green	2	3	1									

Treatments form three groups of size two: $\{1,6\}, \quad \{2,5\}, \quad \{3,4\}$

Groups are orthogonal to litters;

treatments-within-groups have efficiency factor $\frac{1}{3}$ in litters.

Two-colour microarrays (Mead; Ruggiero; Lynch)

6 Treatments	4 Litters 3 Animals in L	12 Slides 2 Colours
6 treatments	12 animals	24 spots

	Litter 1			Litter 2			Litter 3			Litter 4		
Slide	1	2	3	4	5	6	7	8	9	10	11	12
Red	1	2	3	1	4	5	2	4	6	2	5	6
Green	2	3	1	4	5	1	4	6	2	5	6	2

Treatments form three groups of size two: $\{1,6\}, \quad \{2,5\}, \quad \{3,4\}$

Groups are orthogonal to litters;

treatments-within-groups have efficiency factor $\frac{1}{3}$ in litters.

Two-colour microarrays (Mead; Ruggiero; Lynch)

6 Treatments	4 Litters 3 Animals in L	12 Slides 2 Colours
6 treatments	12 animals	24 spots

	Litter 1			Litter 2			Litter 3			Litter 4			
Slide	1	2	3	4	5	6	7	8	9	10	11	12	
Red	1	2	3	1	4	5	2	4	6	2	5	6	
Green	2	3	1	4	5	1	4	6	2	5	6	2	same animal

Treatments form three groups of size two: $\{1,6\}, \quad \{2,5\}, \quad \{3,4\}$

Groups are orthogonal to litters;

treatments-within-groups have efficiency factor $\frac{1}{3}$ in litters.

Two-colour microarrays: skeleton anova

	Litter 1			Litter 2			Litter 3			Litter 4		
Slide	1	2	3	4	5	6	7	8	9	10	11	12
Red	1	2	3	1	4	5	2	4	6	2	5	6
Green	2	3	1	4	5	1	4	6	2	5	6	2

spots		animals			treatments			
source	df	eff	source	df	eff	source	df	EMS
Mean	1	1	Mean	1	1	Mean	1	ξ_0
Slides	11	1	Litters	3	$1 \times \frac{1}{3}$	Tmts[G]	3	$\xi_S + 2\eta_L + \frac{1}{3}q(T[G])$
		$\frac{1}{4}$	Animals[L]	8	$\frac{1}{4} \times \frac{2}{3}$	Tmts[G]	3	$\xi_S + \frac{2}{4}2\eta_{LA} + \frac{1}{6}q(T[G])$
					$\frac{1}{4} \times 1$	Groups	2	$\xi_S + \frac{2}{4}2\eta_{LA} + \frac{1}{4}q(G)$
						Residual	3	$\xi_S + \frac{2}{4}2\eta_{LA}$
Colours	1							ξ_C
S#C	11	$\frac{3}{4}$	Animals[L]	8	$\frac{3}{4} \times \frac{2}{3}$	Tmts[G]	3	$\xi_{SC} + \frac{6}{4}2\eta_{LA} + \frac{1}{2}q(T[G])$
					$\frac{3}{4} \times 1$	Groups	2	$\xi_{SC} + \frac{6}{4}2\eta_{LA} + \frac{3}{4}q(G)$
						Residual	3	$\xi_{SC} + \frac{6}{4}2\eta_{LA}$
			Residual	3				ξ_{SC}

Two-colour microarrays: skeleton anova

	Litter 1			Litter 2			Litter 3			Litter 4		
Slide	1	2	3	4	5	6	7	8	9	10	11	12
Red	1	2	3	1	4	5	2	4	6	2	5	6
Green	2	3	1	4	5	1	4	6	2	5	6	2

spots		animals			treatments			
source	df	eff	source	df	eff	source	df	EMS
Mean	1	1	Mean	1	1	Mean	1	ξ_0
Slides	11	1	Litters	3	$1 \times \frac{1}{3}$	Tmts[G]	3	$\xi_S + 2\eta_L + \frac{1}{3}q(T[G])$
		$\frac{1}{4}$	Animals[L]	8	$\frac{1}{4} \times \frac{2}{3}$	Tmts[G]	3	$\xi_S + \frac{2}{4}2\eta_{LA} + \frac{1}{6}q(T[G])$
					$\frac{1}{4} \times 1$	Groups	2	$\xi_S + \frac{2}{4}2\eta_{LA} + \frac{1}{4}q(G)$
						Residual	3	$\xi_S + \frac{2}{4}2\eta_{LA}$
Colours	1							ξ_C
S#C	11	$\frac{3}{4}$	Animals[L]	8	$\frac{3}{4} \times \frac{2}{3}$	Tmts[G]	3	$\xi_{SC} + \frac{6}{4}2\eta_{LA} + \frac{1}{2}q(T[G])$
					$\frac{3}{4} \times 1$	Groups	2	$\xi_{SC} + \frac{6}{4}2\eta_{LA} + \frac{3}{4}q(G)$
						Residual	3	$\xi_{SC} + \frac{6}{4}2\eta_{LA}$
			Residual	3				ξ_{SC}

Nonorthogonality in Phase I: revisited

Query

Suppose that, in Phase I, treatment term i is partially confounded,
with a small efficiency factor λ_{ij},
with stratum j, which has a large variance η_j.

Nonorthogonality in Phase I: revisited

Query

Suppose that, in Phase I, treatment term i is partially confounded, with a small efficiency factor λ_{ij}, with stratum j, which has a large variance η_j.

At Phase II, should we

- *try to confound stratum j with a Phase II stratum with small variance, or*
- *cut our losses on this part of the information about treatment term i?*

Nonorthogonality in Phase I: a special case

Suppose that,

- in Phase I, there is a treatment term that has effiency factors

$$\begin{array}{lll} p & \text{in stratum 1 with variance} & \eta_1 \\ q & \text{in stratum 2 with variance} & \eta_2, \end{array}$$

  where $p+q=1$ and $\eta_1 > \eta_2$;
- the design for Phase II has replication r, there are two Phase II strata where these Phase I strata might be confounded, but they cannot both go in the one with the smaller variance.

Label the Phase I strata 1, 2 so that

$$\begin{array}{lll} p & \eta_1 & \xi_1 \\ q & \eta_2 & \xi_2. \end{array}$$

Query

Should we do this so that $\xi_1 > \xi_2$ or $\xi_2 > \xi_1$?

How to confound two pairs of strata from the two phases

tmt efficiency factor	Phase I variance	Phase II variance
p	$\eta_1 (> \eta_2)$	ξ_1
q	η_2	ξ_2

How to confound two pairs of strata from the two phases

tmt efficiency factor	Phase I variance	Phase II variance
p	$\eta_1 (> \eta_2)$	ξ_1
q	η_2	ξ_2

Theorem

Smaller variance is obtained if the Phase II strata are labelled so that

$$\xi_1 > \xi_2 \quad \text{if} \quad \frac{q}{p} > \frac{(r\eta_2 + \xi_1)(r\eta_2 + \xi_2)}{(r\eta_1 + \xi_1)(r\eta_1 + \xi_2)};$$

in particular, if $\frac{q}{p} > 1;$

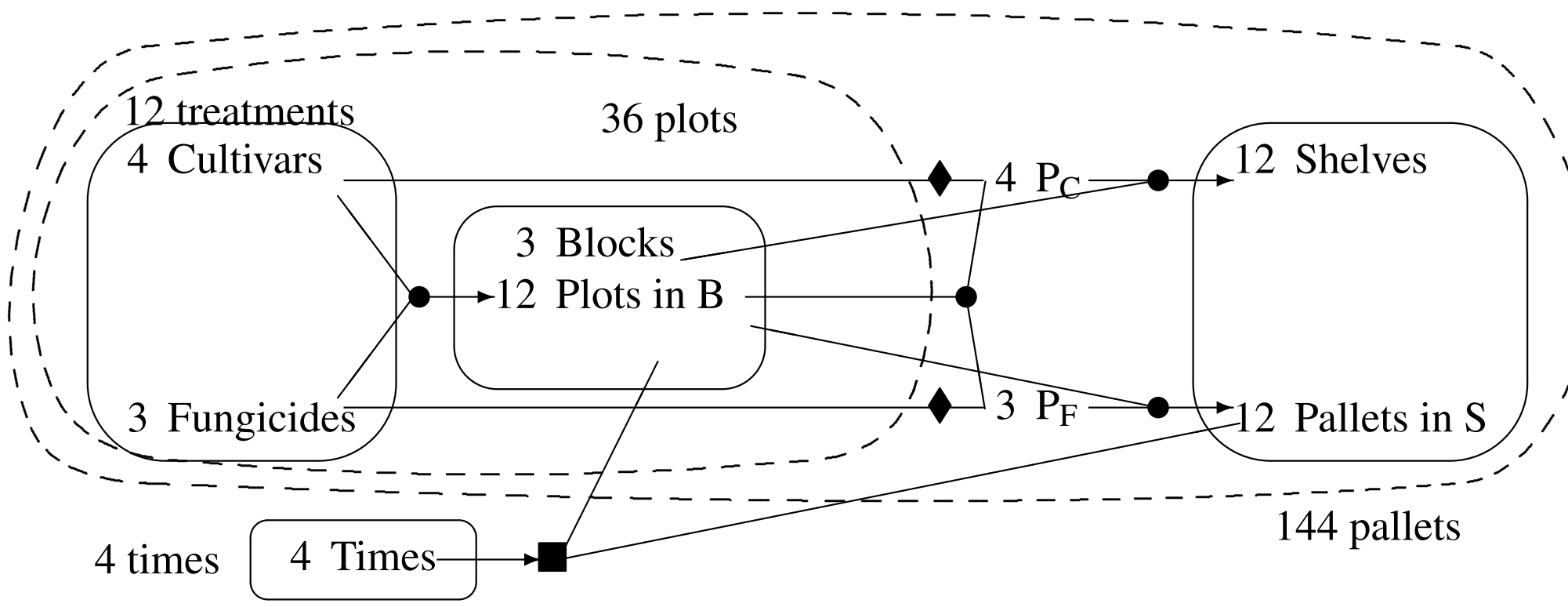
$$\xi_1 < \xi_2 \quad \text{if} \quad \frac{q}{p} < \frac{(r\eta_2 + \xi_1)(r\eta_2 + \xi_2)}{(r\eta_1 + \xi_1)(r\eta_1 + \xi_2)}.$$

Some treatment factors may be applied only in Phase II

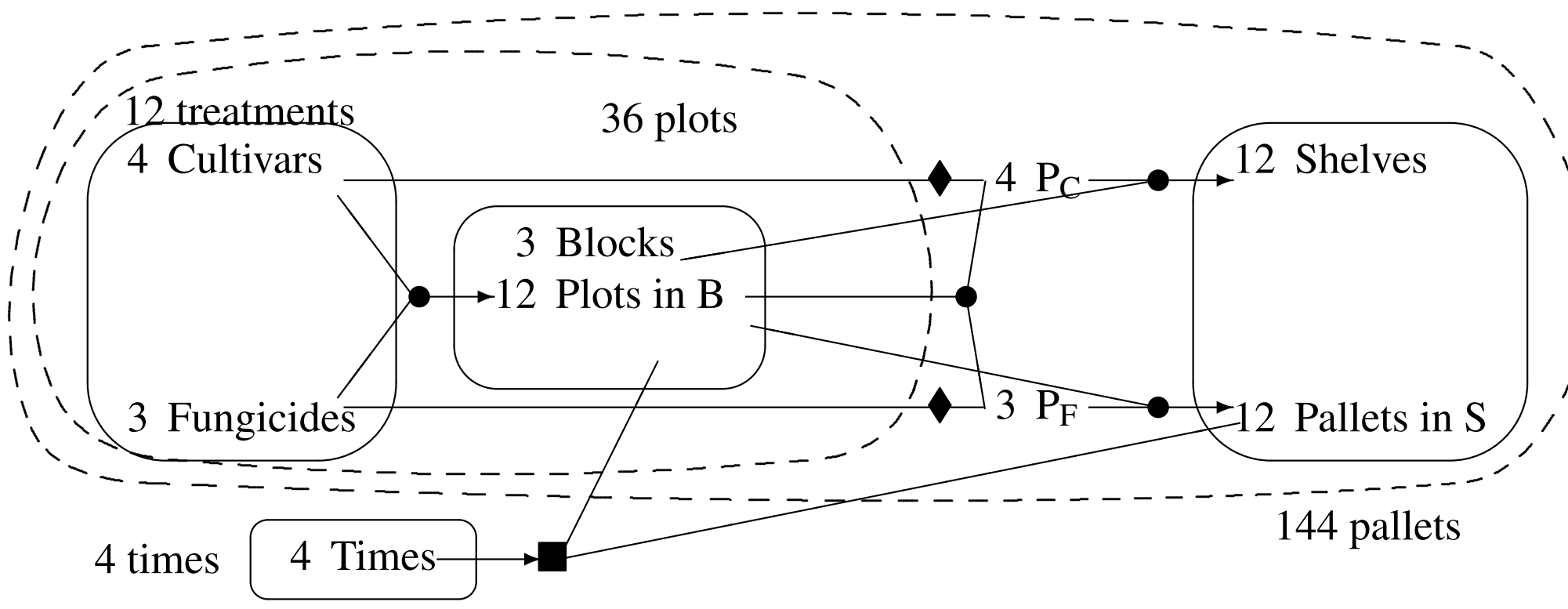
Principle

Plan the whole experiment in advance,
especially if you want to estimate interactions
between the Phase I treatments and the Phase II treatments.

Potato storage (Payne)

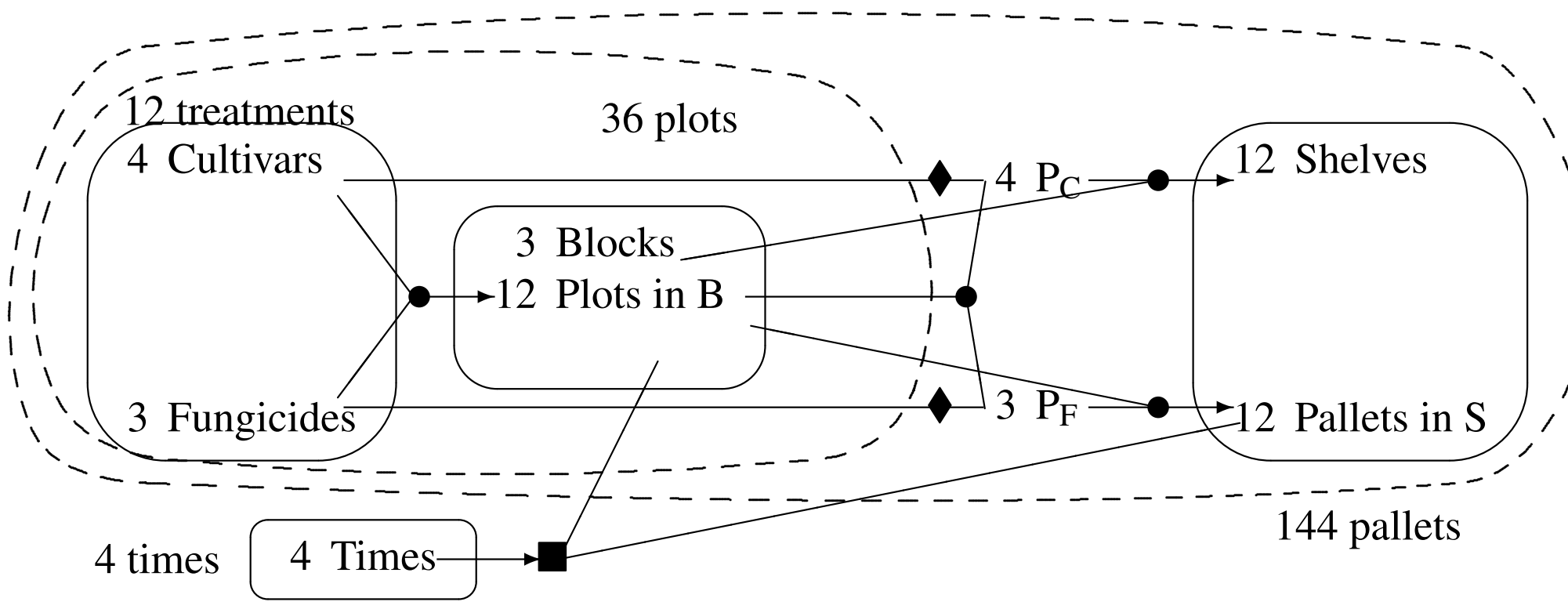


Potato storage (Payne)



levels	tmts	plots	pallets	times
4	C	P_2	S_2, Q_2	T
3	F	B, P_1	S_1, Q_1	

Potato storage (Payne)

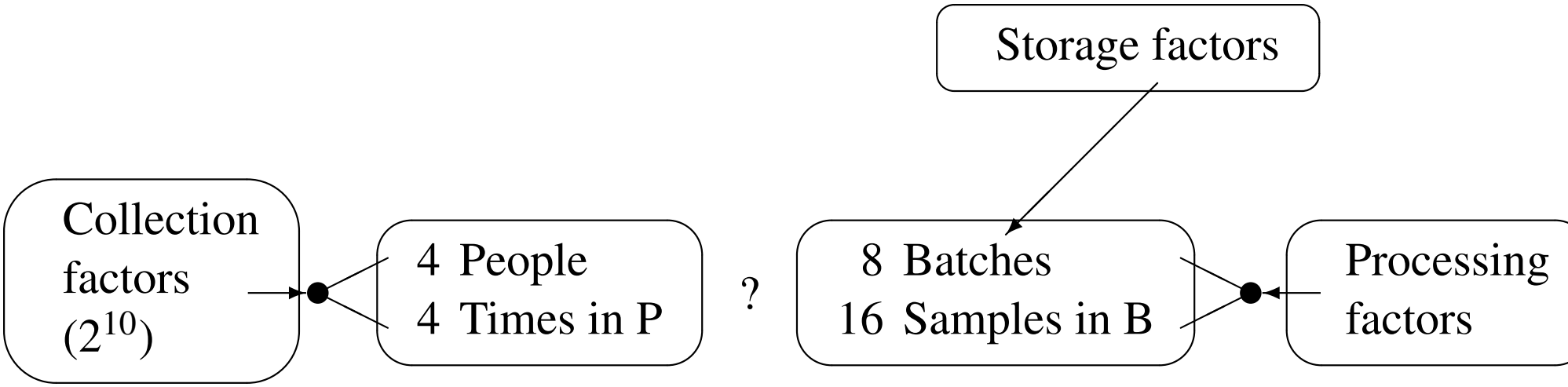


levels	tmts	plots	pallets	times
4	C	P_2	$S_2,\ Q_2$	T
3	F	$B,\ P_1$	$S_1,\ Q_1$	

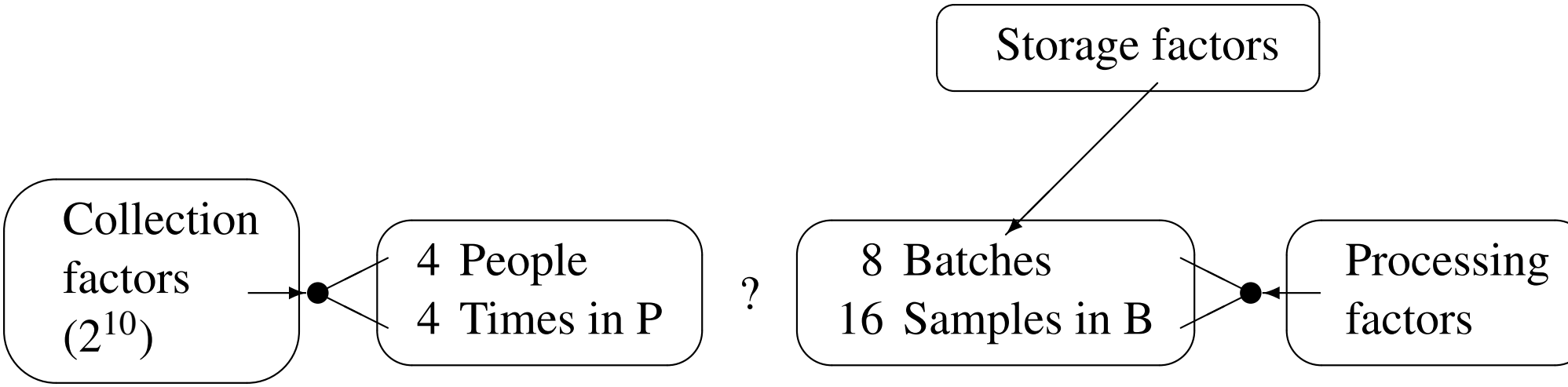
$$C \equiv P_2 \qquad P_2 \equiv S_2 \qquad T \equiv Q_2$$

$$F \equiv P_1 \qquad B \equiv S_1,\ P_1 \equiv Q_1$$

Blood (Speed)

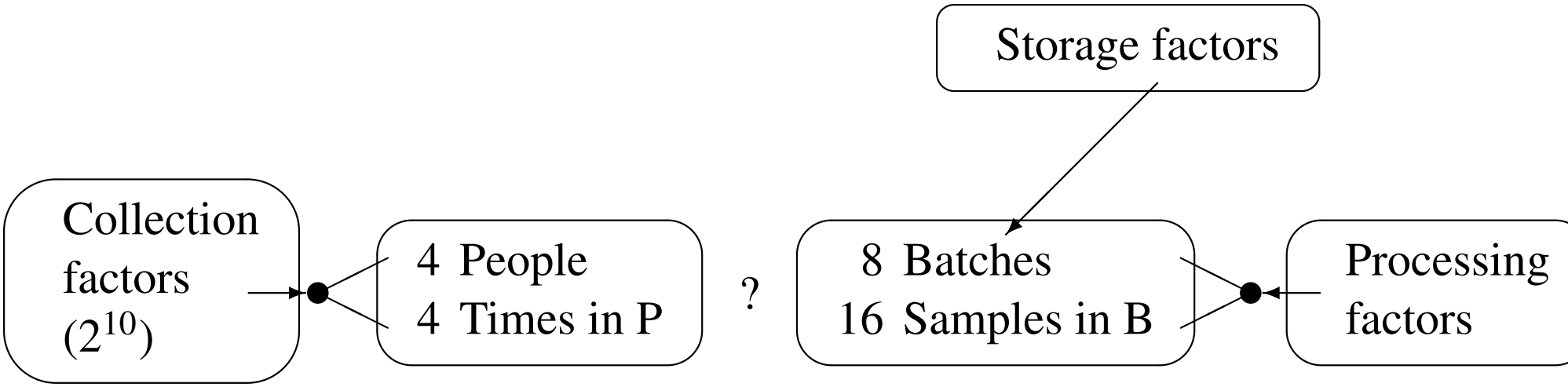

Storage factors
Collection factors (2^{10})
4 People
4 Times in P
?
8 Batches
16 Samples in B
Processing factors

Blood (Speed)



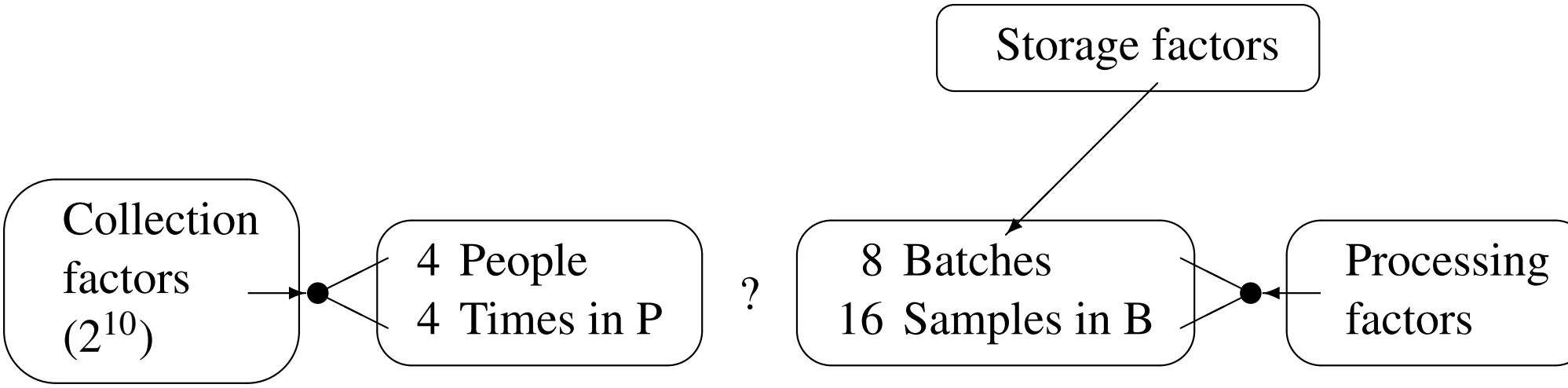
- Confounding People with Batches increases the variance for some Storage factors.

Blood (Speed)



- Confounding People with Batches increases the variance for some Storage factors.
- Making People orthogonal to Batches uses up more degrees of freedom.

Blood (Speed)



- Confounding People with Batches increases the variance for some Storage factors.
- Making People orthogonal to Batches uses up more degrees of freedom.
- The compromise has a single df for People $\vee$ Batches, with a variance which is large and different from all other variances (Cheng; Vivacqua).

	4	4
2	64	0
2	0	64

Conclusions: I

When designing a two-phase experiment
(or a multi-stage batch reprocessing experiment)

- design the whole thing in advance;
- pay attention to making variance small;
- pay attention to residual degrees of freedom;
- if treatments are orthogonal to 'large blocks' in Phase I, then those large blocks should be confounded with "large blocks" in Phase II;
- if a treatment term is in a Phase I stratum with large variance (in particular, if a treatment factor is 'hard to set' in Phase I), then it should be allocated to a Phase II stratum with small variance.
- if a treatment term is partially confounded with more than one Phase I stratum, and cannot be wholly allocated to a 'small variance' stratum in Phase II, then the best design depends on the ratios of the stratum variances and on the efficiency factors from Phase I.

Conclusions: II

When designing a two-phase experiment
(or a multi-stage batch reprocessing experiment),
the following are useful concepts and tools.

- diagrams with panels to show tiers;
- Hasse diagrams to elucidate strata and degrees of freedom;
- the design key method of construction,
  which can also be used to elucidate the confounding;
- decomposition tables (skeleton anova)
  to assess the qualities of the design.