Method one for getting a subject average across two scans

1st-level feat (subject)		2nd-level feat (subject)	3rd-level feat (group model)	2nd-level feat (subject)		1st-level feat (subject)
fixed effects		*fixed effects*	*fixed effects*	*fixed effects*		*fixed effects*
Scan 1						**Scan 2**
Run 1	↘				↙	Run 1
Run 2	→	Scan1 avg →	Subject average (output: one per subject)	← Scan2 avg	←	Run 2
Run 3	→				←	Run 3
Run 4	↗		↓		↖	Run 4

4th-level feat (2 groups)
mixed effects

Inputs: 3rd-level COPEs of subject avgs

Use EVs to specify Group A or Group B

Output:
Group A average vs. Group B average

Method two for getting a subject average across two scans

1st-level feat (subject)		2nd-level feat (subject)		1st-level feat (subject)
fixed effects		*fixed effects*		*fixed effects*
Scan 1				**Scan 2**
Run 1	↘		↙	Run 1
Run 2	→	Subject average	←	Run 2
Run 3	→		←	Run 3
Run 4	↗	↓	↖	Run 4

3rd-level feat (2 groups)
mixed effects

Inputs: 2nd-level COPEs of subject avgs

Use EVs to specify Group A or Group B

Output:
Group A average vs. Group B average

Could someone explain to me the source of the difference in results for the subject average? Is there a good reason why one of these approaches is better or worse than the other?