CAN CORRELATIONS IMPLY CAUSATION? CAUSAL MODELING AND MUSIC PSYCHOLOGY RESEARCH

Daniel MÜLLENSIEFEN, Georgia FLORIDOU, Kelly JAKUBOWSKI

Music, Mind and Brain Group,

Department of Psychology, Goldsmiths, University of London, United Kingdom

The problem(s)

- Causal relationships often of main interest: “musical training (x) makes children more intelligent (y)”: x -> y
- Randomized Controlled Trials (RCTs) are main experimental paradigm for obtaining information about causal mechanisms
- But true RCTs often not possible
  - Not really R (e.g. no random allocation to experimental groups)
  - Not really C (e.g. no control group)
  - Not really T (e.g. observational instead of experimental data)

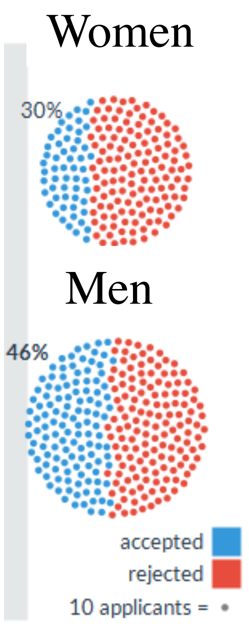
If not RCT, then potential problems:

– Confounding variables make cause -> effect relationship uncertain

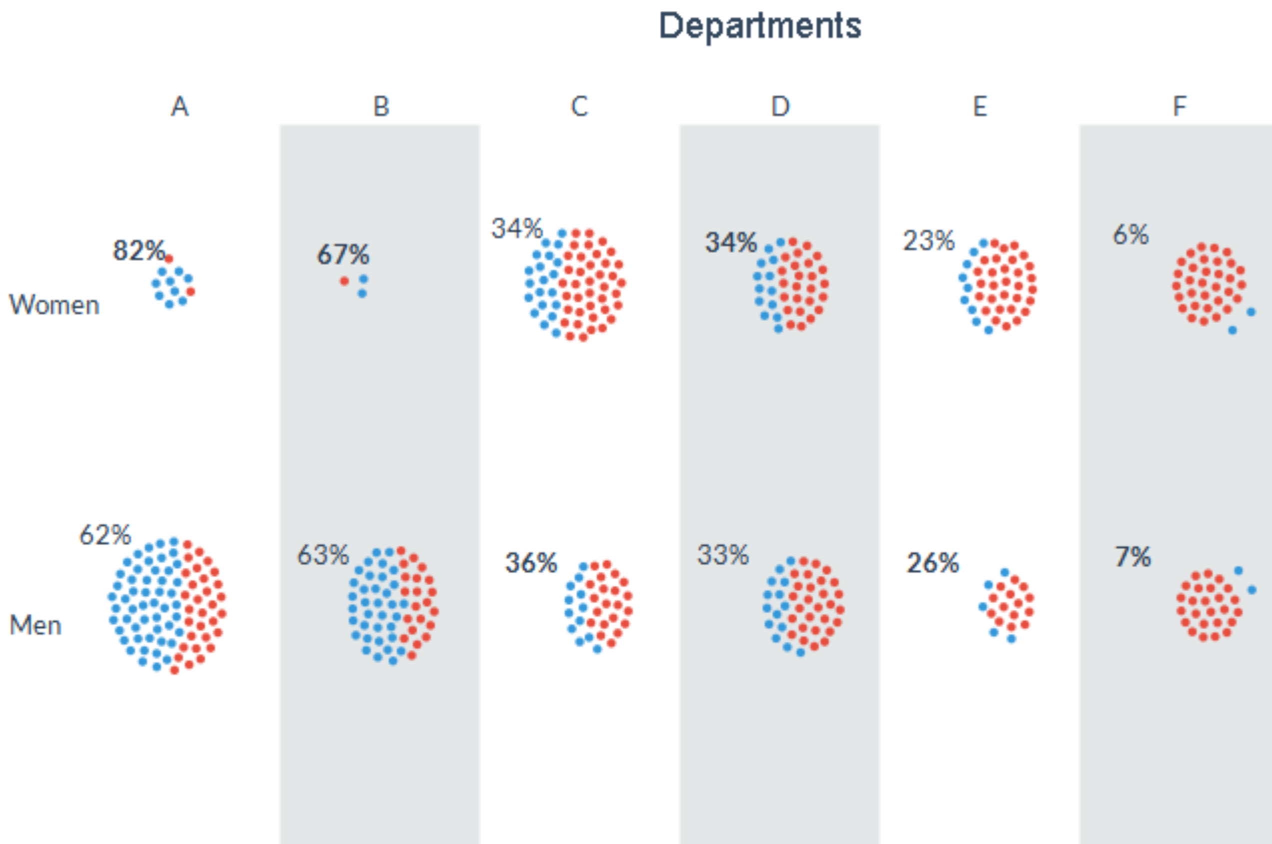
Example: Berkeley university admissions

(Bickel et al., 1975)

More women rejected than men



Women apply more to competitive departments



Omitted variable bias (Simpson’s Paradox; Simpson, 1951)

If not RCT, then potential problems:

- Confounding variables make cause -> effect relationship uncertain
- With several confounding variables, (causal) relationships among them unclear
- What variables do we need to control for?

But in music psychology a lot of non-RCT data is available:

- Surveys (Müllensiefen et al., 2014)
- Data from music behaviour in the real world (Pawley & Müllensiefen, 2012)
- Quasi-experimental data (Jakubowski & Müllensiefen, 2013)

- What to do with this data?
  - Not use it for causal inference because it is not RCT
  - Draw causal inference anyway
  - Control for effects of confounding variables with covariates in (hierarchical) multiple regression
  - Use causal models (Rubin, 1974; Pearl, 2000; Sprites et al., 2000)

Two scenarios for causal inference

1. Binary x -> Continuous y
   - BUT: need to control for potentially confounding variables
   - *Rubin's causal model*
2. Correlational data with many significant bivariate correlations
   - BUT:
     - Assumed network of underlying relationship is sparse
     - Causal processes can be assumed, but directions not always known
   - Graphical network from *PC algorithm*

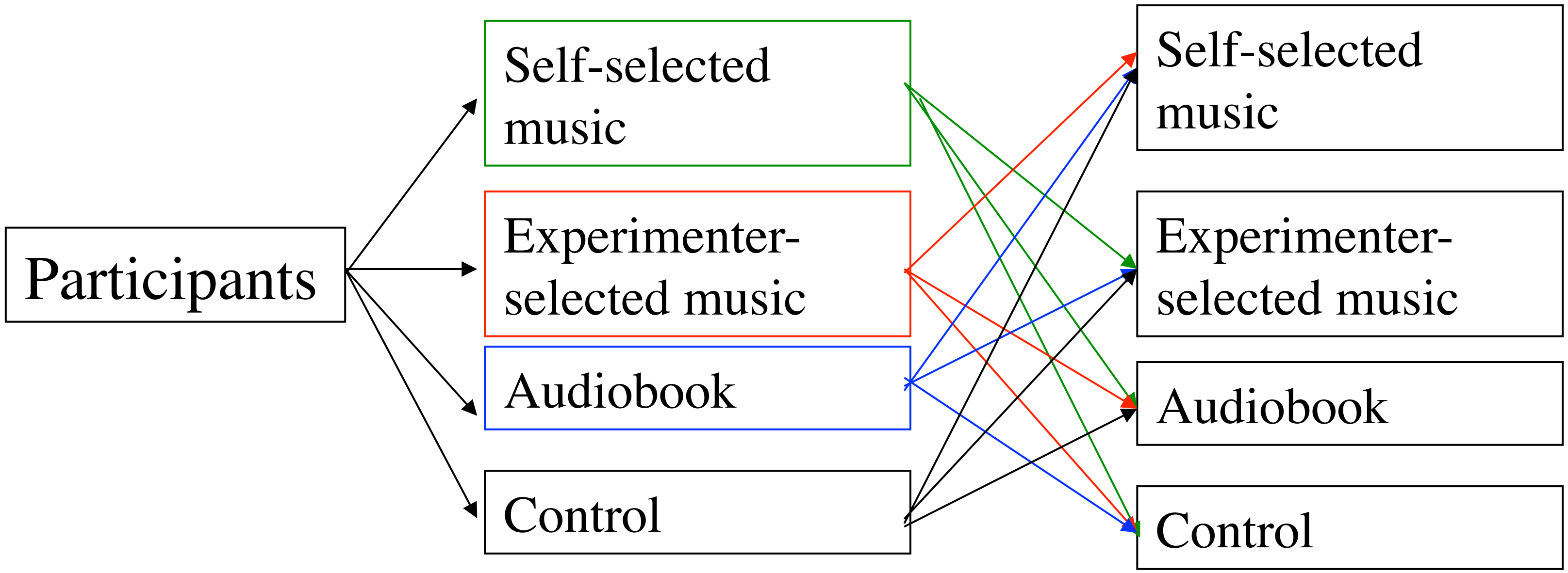
The ideal case

Goldsmiths
UNIVERSITY OF LONDON

The ideal case

Repeated measurements with cross-over design

Example: Music as sleep intervention (Trahan et al. 2015, ESCOM)



=> *How does each participant behave in treatment compared to control condition? ('subject-specific causal effect')*

Goldsmiths UNIVERSITY OF LONDON

The next best thing (1)

Between-groups design as RCT

Example: Foreign language learning with and without music (Kang & Williamson, 2014)

⇒ *How does treatment group behave compared to control group? ('population causal effect')*

⇒ Statistical association interpreted as Causal Effect, because participants are exchangeable between groups

⇒ Exchangeability derives from random allocation to groups

The next best thing (2)

Between-groups design with matched treatment +control subjects

Example: Musical and non-musical memory performance in amusics v healthy adults (Williamson & Stewart, 2010)

=> *How do participants from amusic group behave compared to matched control participants? ('population causal effect')*

What if RCT is not possible?

Because:

- Random group allocation is not possible
- Group sizes are small and groups differ on confounding variables by chance?
- Multiple causes contribute to effect

How does treatment group behave compared to control group? ('population causal effect')

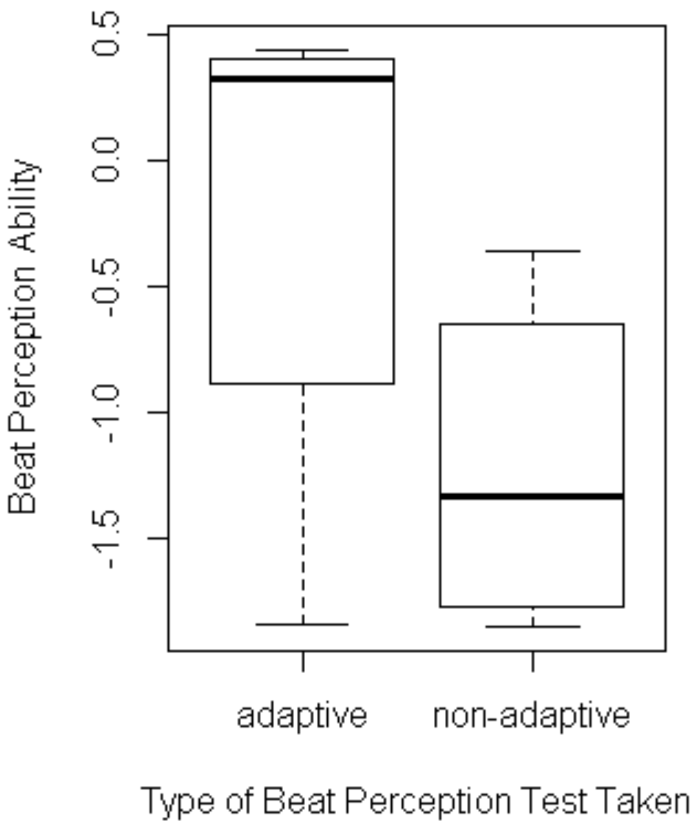
$\Rightarrow$Make participants exchangeable between groups by matching them on confounding variables

$\Rightarrow$Can't observe what happens to same participant in both conditions, but what happens to matched participant

1. Matching and Rubin's causal model

(Rubin 1974; Imbens & Rubin, 2015)

Example: Performance on adaptive v. non-adaptive version of Beat Perception Test (Harrison et al., 2015, ESCOM)



Causal effect of adaptive test?

- Match participants on IQ and 4 variables of musical background
- Average causal effect = 0.84, (SE = 0.31, t=2.7, p = .007)
- Validation in subsequent RM test: mean difference = 0.1 (t = 1.9, p = 0.03)

1. Matching and Rubin's causal model
(Rubin 1974; Imbens & Rubin, 2015)

– Matching works by (non-parametrically) finding closest participant in control group for any participant in treatment group

– Different from multiple regression:
  – Not assuming linearity
  – Not assuming constant variance
  – Not assuming additivity

– Drawback: Only works for small number of groups on independent variable

2. Correlational Data

2. Correlational Data

Example: Relations between self-beliefs, personality, cognitive and musical abilities

(Müllensiefen et al., under review)

1. Intelligence	1.00																			
2. Melodic Memory	0.30	1.00																		
3. Beat Perception	0.08	0.29	1.00																	
4. Musical Goals	-0.09	-0.28	-0.14	1.00																
5. Theory of Musical Ability	-0.07	-0.25	-0.14	0.35	1.00															
6. Age	0.22	0.14	0.18	-0.03	-0.05	1.00														
7. Musical Training	0.13	0.35	0.28	-0.34	-0.09	-0.04	1.00													
8. Goals Choice	0.11	0.08	0.05	-0.39	-0.18	-0.04	0.18	1.00												
9. Theory of Intelligence	-0.03	0.11	0.10	-0.28	-0.51	0.09	0.07	0.30	1.00											
10. Extraversion	-0.18	-0.22	-0.08	0.09	0.03	-0.29	0.02	-0.08	0.04	1.00										
11. Agreeableness	-0.01	-0.15	-0.04	-0.09	-0.06	-0.13	0.03	0.19	0.10	0.16	1.00									
12. Conscientiousness	0.02	-0.04	0.10	-0.13	-0.10	-0.06	0.12	0.10	0.25	0.14	0.35	1.00								
13. Emotional Stability	0.06	-0.02	0.02	-0.10	-0.11	-0.08	0.05	0.19	0.12	0.32	0.31	0.27	1.00							
14. Openness	-0.02	-0.13	0.05	-0.18	-0.12	-0.01	0.13	0.16	0.15	0.33	0.34	0.27	0.24	1.00						
15. Sound Similarity Perception	-0.07	-0.07	-0.01	0.08	0.06	-0.05	0.01	-0.05	-0.08	0.00	0.02	0.08	-0.04	-0.01	1.00					
16. Concurrent Musical Activities	0.14	0.29	0.18	-0.35	-0.14	-0.08	0.75	0.23	0.18	0.03	0.07	0.13	0.02	0.12	-0.05	1.00				
17. Academic Effort	0.22	0.06	0.11	0.02	-0.11	0.27	0.10	0.05	0.01	-0.05	0.14	0.43	0.02	0.18	0.00	-0.06	1.00			
18. Academic Achievement	0.35	0.23	0.26	-0.22	-0.25	0.00	0.21	0.16	0.24	-0.03	0.07	0.33	0.01	0.16	-0.02	0.22	0.78	1.00		
19. Academic Self-Concept	-0.26	-0.19	-0.18	0.06	0.07	-0.07	-0.18	-0.06	-0.05	0.15	-0.08	-0.31	-0.08	-0.06	-0.09	-0.18	-0.28	-0.46	1.00	
20. Social Self-Concept	0.02	0.01	0.04	-0.12	-0.15	-0.08	0.10	0.14	0.15	0.31	0.15	0.17	0.34	0.19	-0.07	0.02	0.20	0.17	0.19	1.00

Analysis options?

- Interpret pattern of bivariate correlations ?
- Factor analysis / principal component analysis ?
- Multiple regression ?
- Structural equation model ?

Which bivariate relationships still significant after accounting for influence from other variables?

- Graphical network analysis

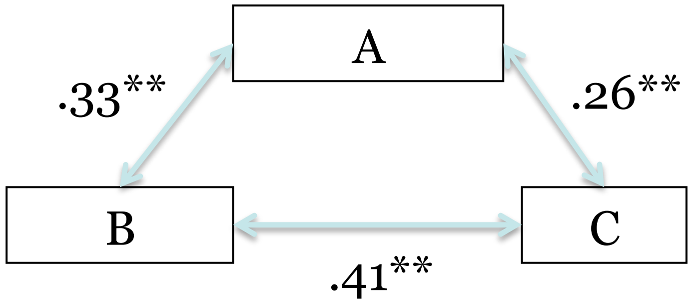
Can we work out the causal direction of the significant relationships?

- PC algorithm

Conditional independence tests and graphical models

Conditional independence tests (e.g. partial correlations):

- How does correlation of A and B change when controlling for C?

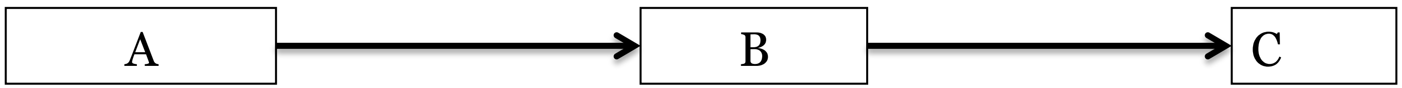


Partial correlation vanishes:

$r(A, B \mid C) = .20^{*}$

$r(C, B \mid A) = .36^{***}$

$r(A, C \mid B) = .16$ n.s.

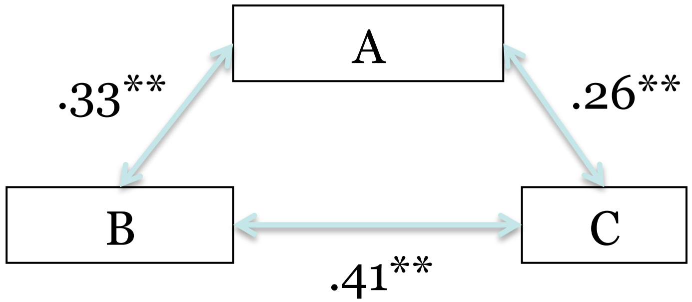


'A causes B and B causes C'

Canditional independence tests and graphical models

Conditional independence tests:

- How does correlation of A and B change when controlling for C?

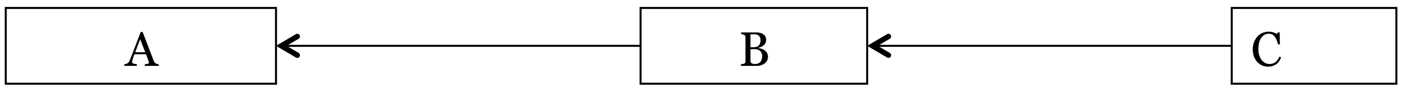


Partial correlation vanishes:

r(A, B | C) = .20*

r(C, B | A) = .36***

r(A, C | B) = .16 n.s.

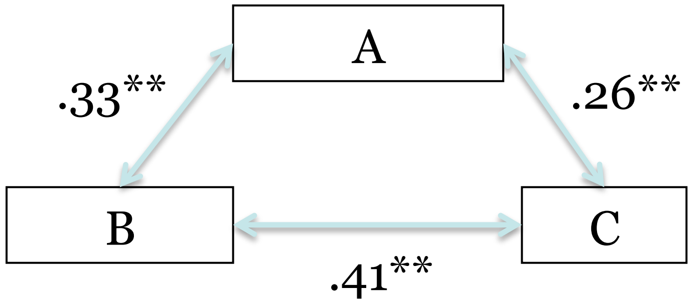


'C causes B and B causes A'

Canditional independence tests and graphical models

Conditional independence tests:

- How does correlation of A and B change when controlling for C?

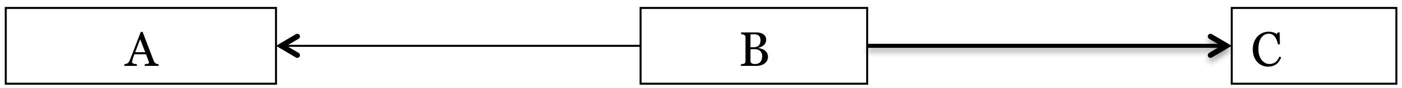


Partial correlation vanishes:

$r_{(A,\, B \mid C)} = .20^{*}$

$r_{(C,\, B \mid A)} = .36^{***}$

$r_{(A,\, C \mid B)} = .16$ n.s.



'B causes both, A and C'

Result of applying partial correlation tests repeatedly

- Network of conditional independence relationships

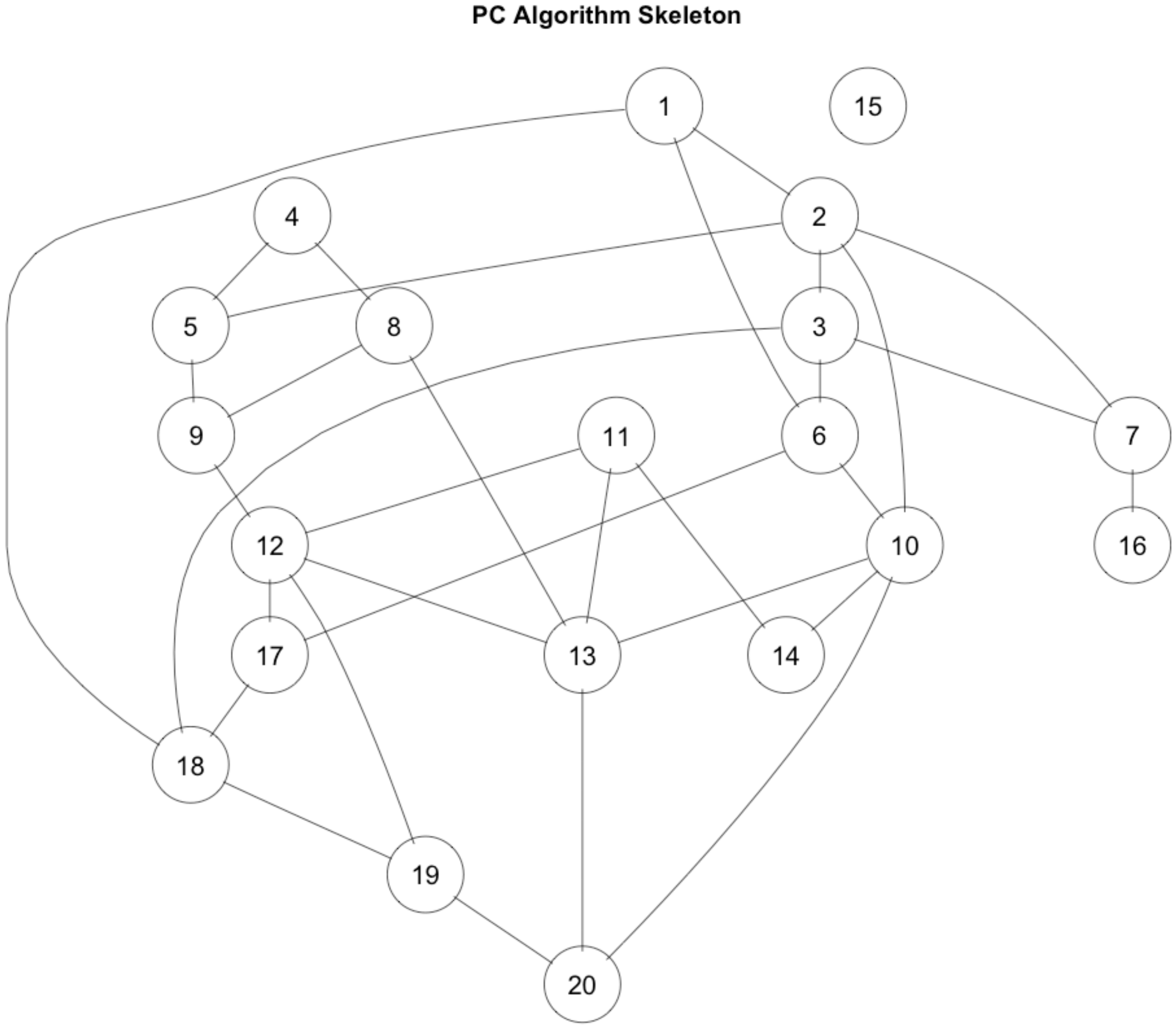
  => *Which two variables are still correlated after taking into account the influence of all other variables? (Remember the Berekely admission exmaple?)*

- Sparse network instead of many bivariate relationships

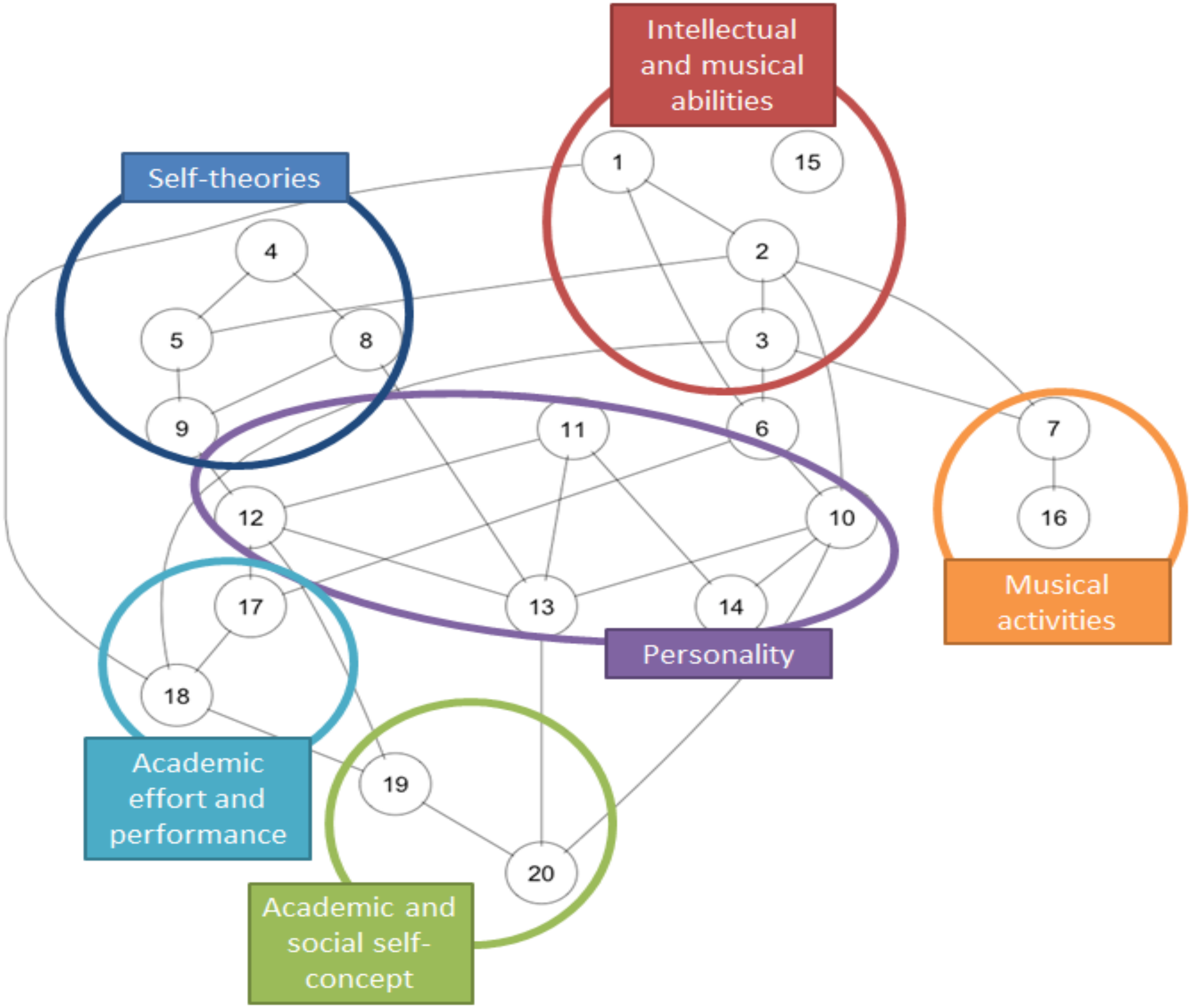
- But no information about direction of causal influence (yet)

Skeleton Network

After computing partial correlations



Skeleton Network (Müllensiefen et al, under review)



Can we work out causation from correlations?

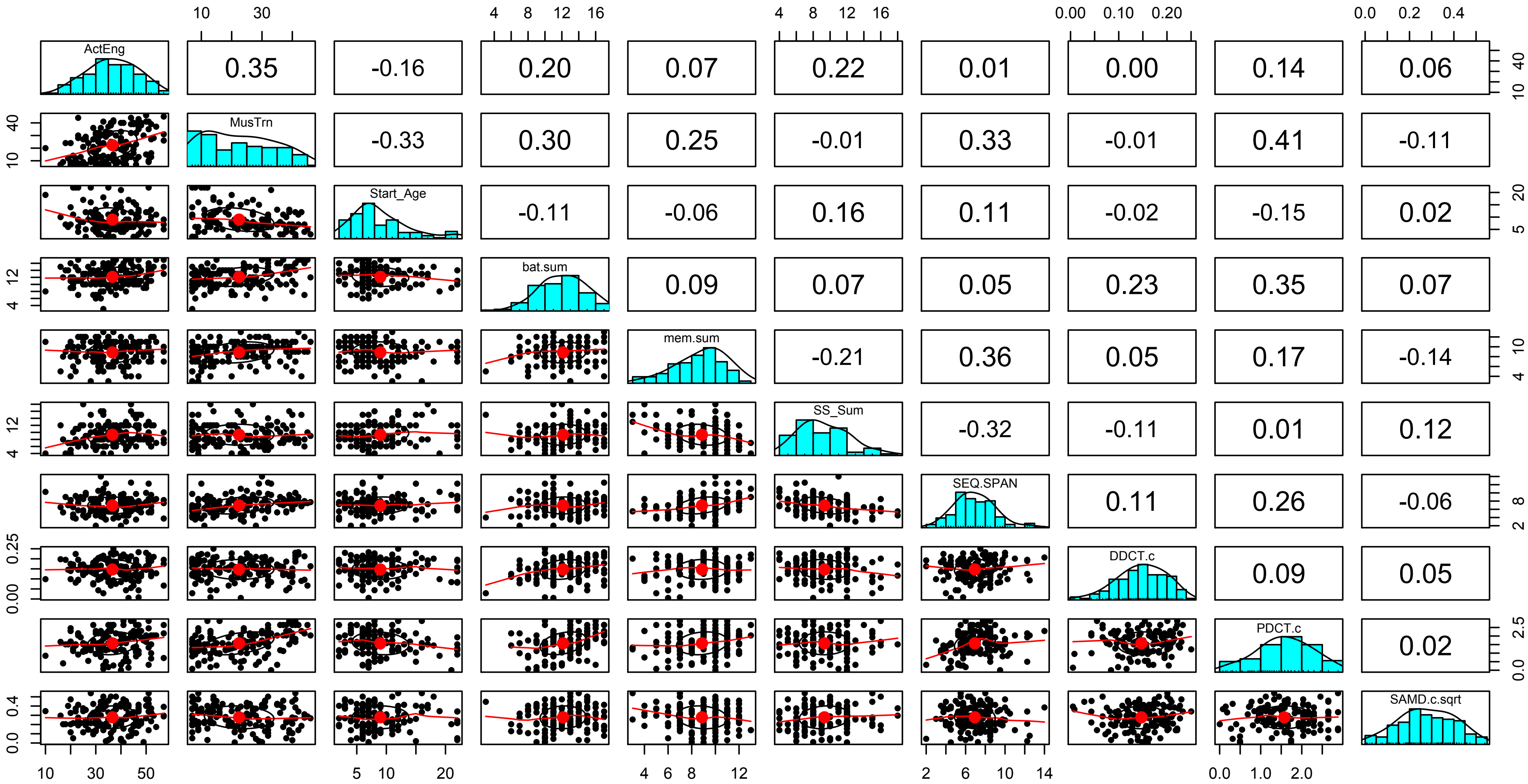
Example: Low-level, musical listening and formal training (Müllensiefen et al., in prep)

– 151 participants
– Controlled testing environment

– 3 Low-level hearing tests (Kidd et al., 2007)
– 1 Auditory sequence span test (Williamson & Stewart, 2010)
– 3 High-level musical listening tests (Müllensiefen et al., 2014)
– Background questionnaire on Musical Training, Active Music Engagement (Müllensiefen et al., 2014)
– (Audiometry, Figure Ground mid-level test, socio-economic status)

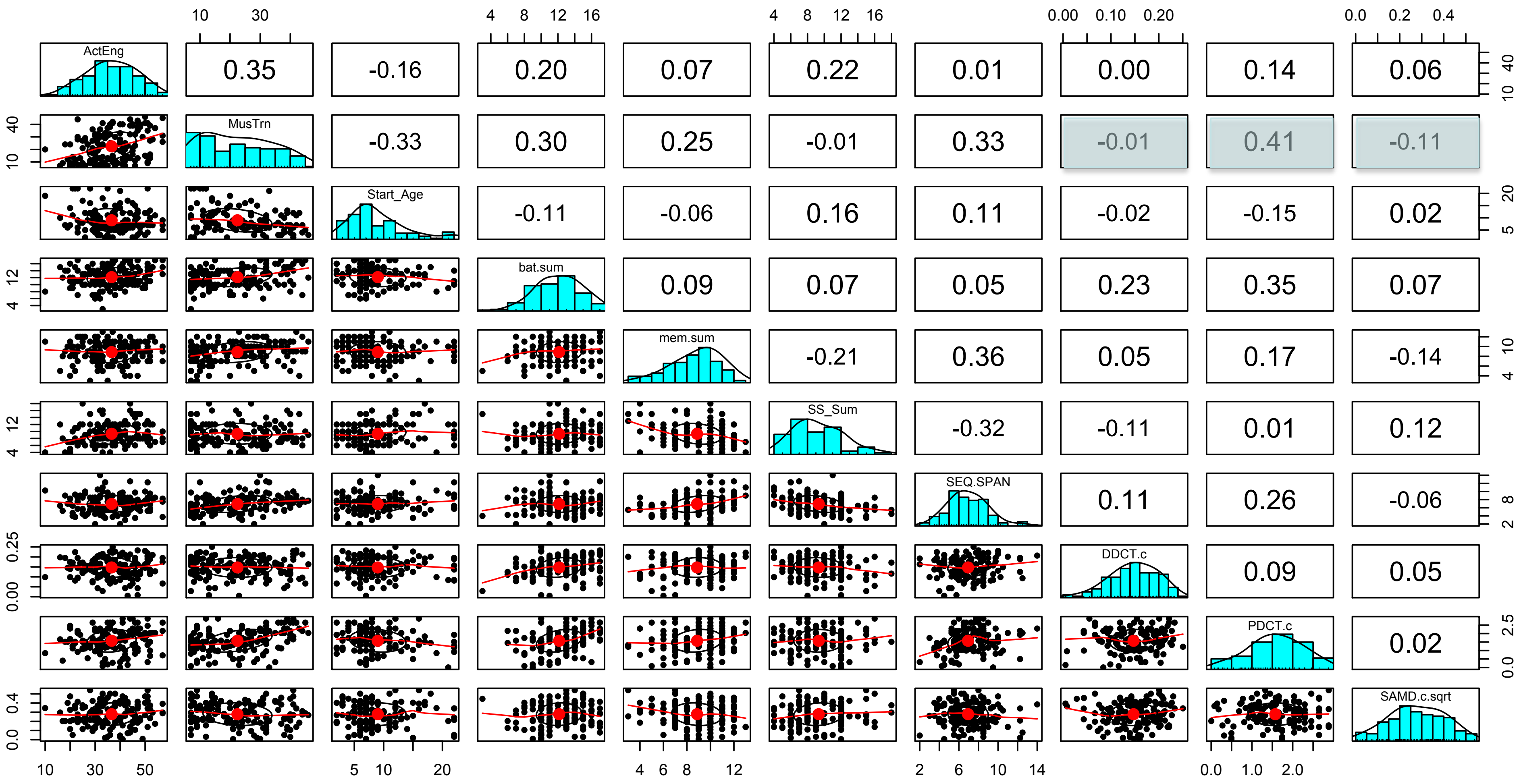
Goldsmiths
UNIVERSITY OF LONDON

Bivariate Correlations



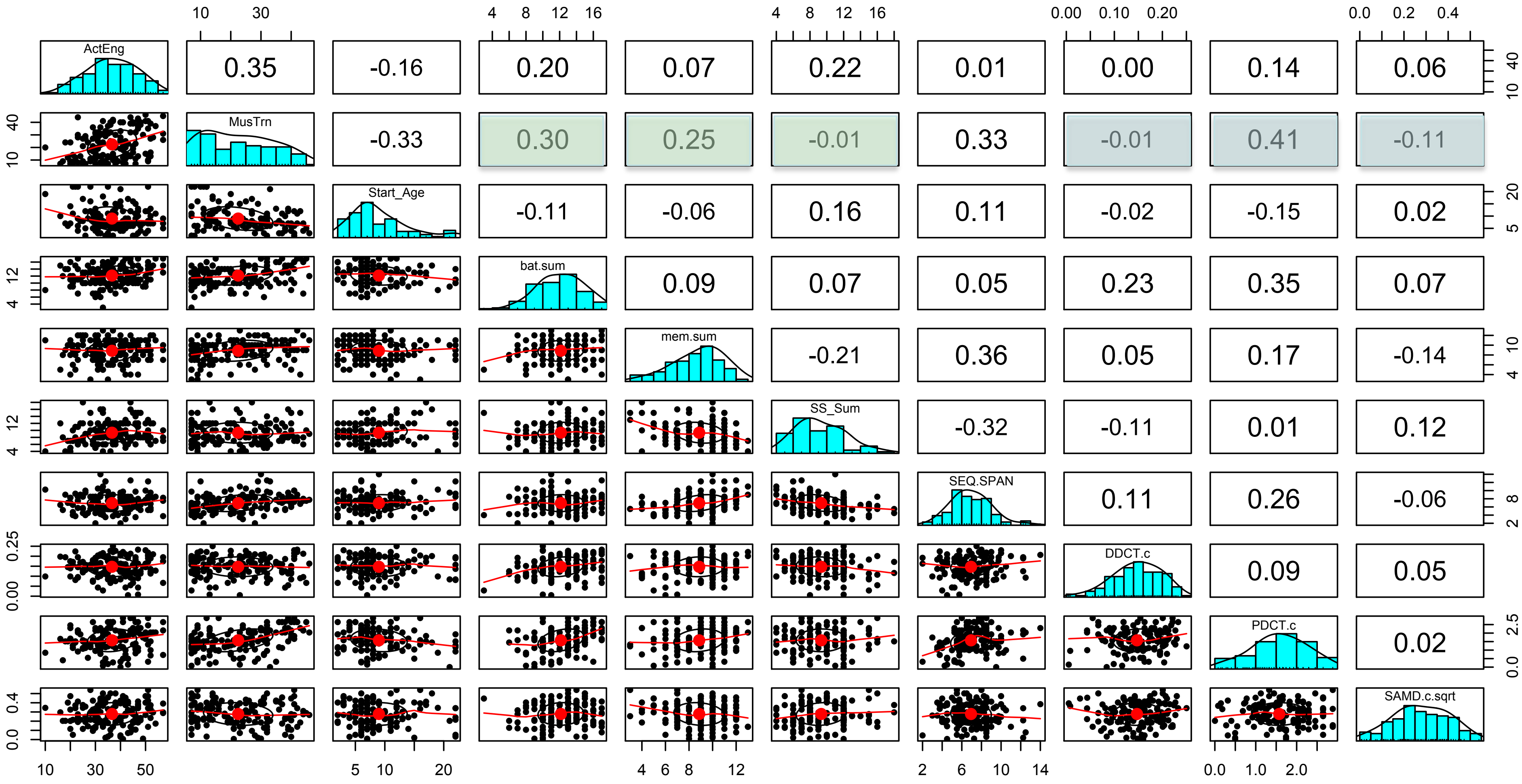
Bivariate Correlations

Musical Training correlates with one low-level hearing test (pitch discrimination)



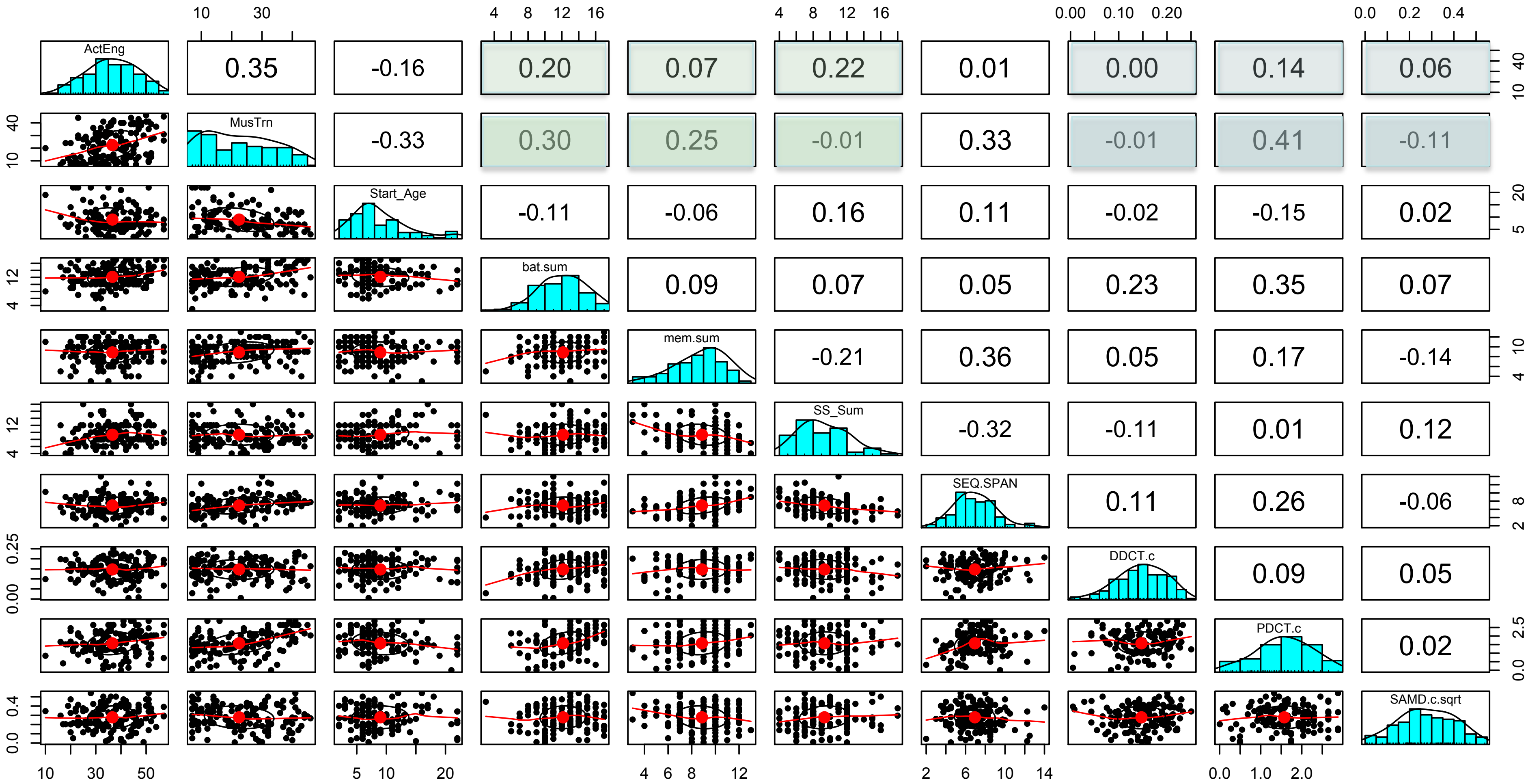
Bivariate Correlations

Musical Training correlates with melodic memory and beat perception ability



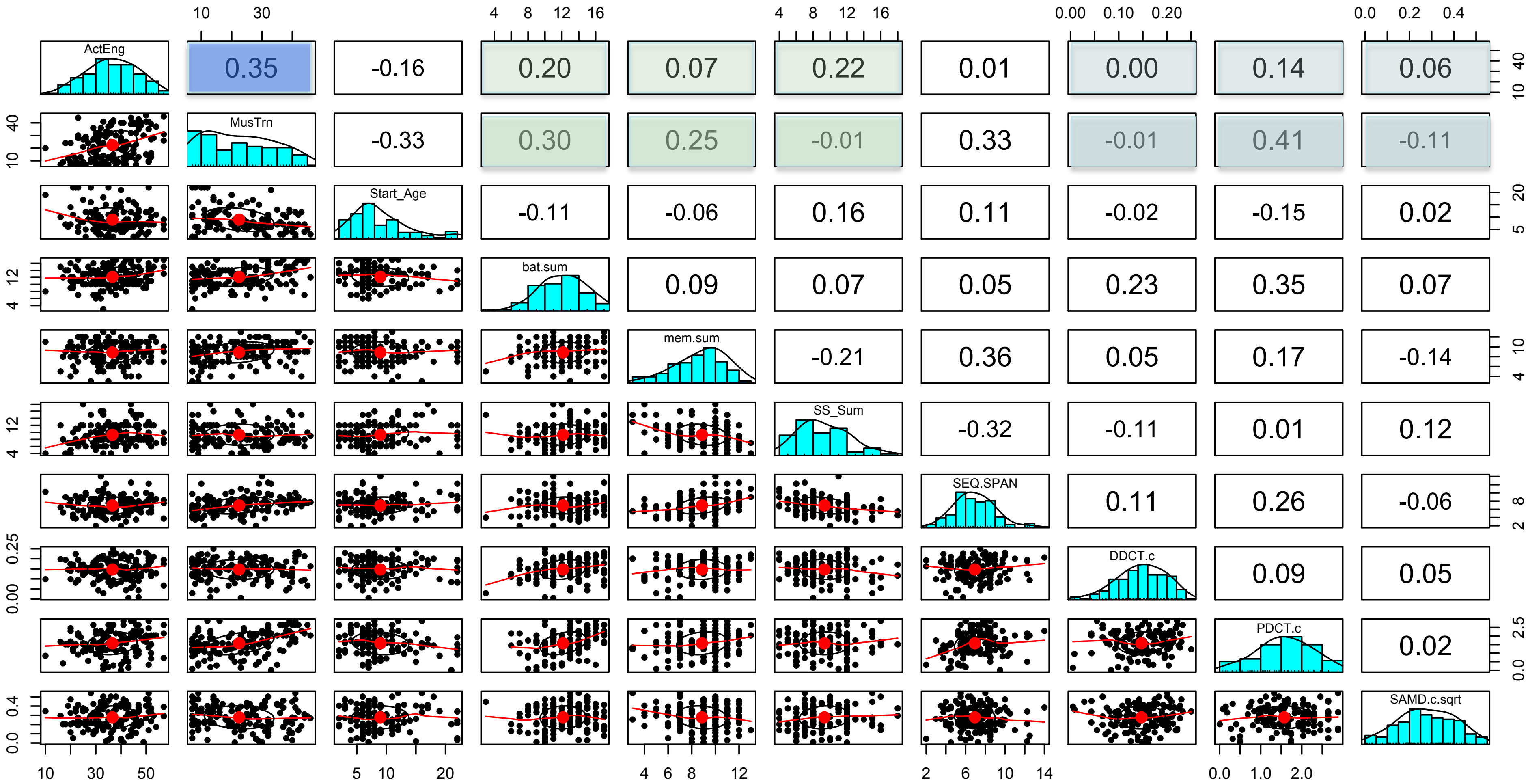
Bivariate Correlations

Active Musical Engagement also correlates with musical and hearing abilities



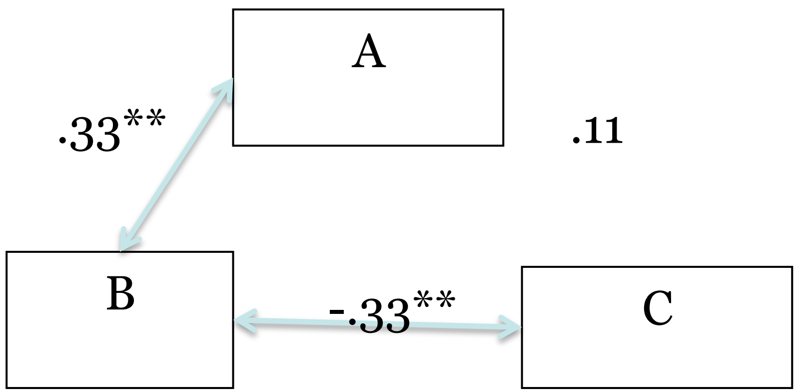
Bivariate Correlations

Active Musical Engagement and Musical Training are also correlated



Search for v-structures arising from partial correlations

Uncorrelated variables become conditionally dependent

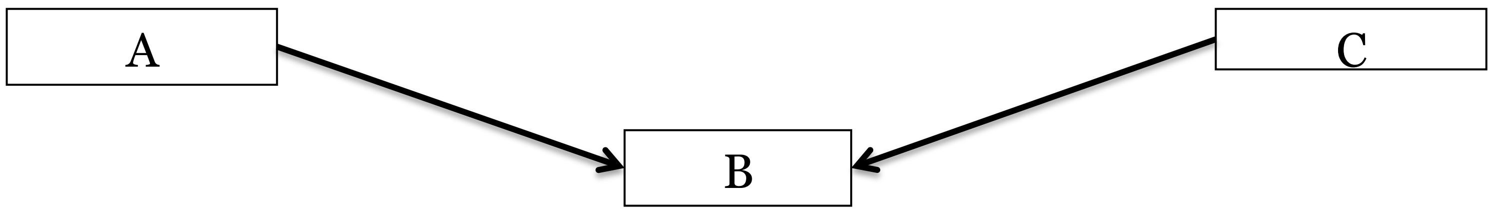


Partial correlation becomes significant where bivariate correlation is not:

$r_{(A,\, B \mid C)} = .31^{***}$

$r_{(B,\, C \mid A)} = -.37^{***}$

$r_{(A,\, C \mid B)} = .21^{*}$



„V-Structure“

A and C cause B

Goldsmiths UNIVERSITY OF LONDON

The intuition behind v-structures

The burglar alarm example:

$p(\text{Burglar}) = 0.001$

$p(\text{EQ}) = 0.001$

$p(\text{EQ} \mid \text{Burglar}) = p(\text{Burglar} \mid \text{EQ}) = 0.001 \times 0.001 = 0.00001$

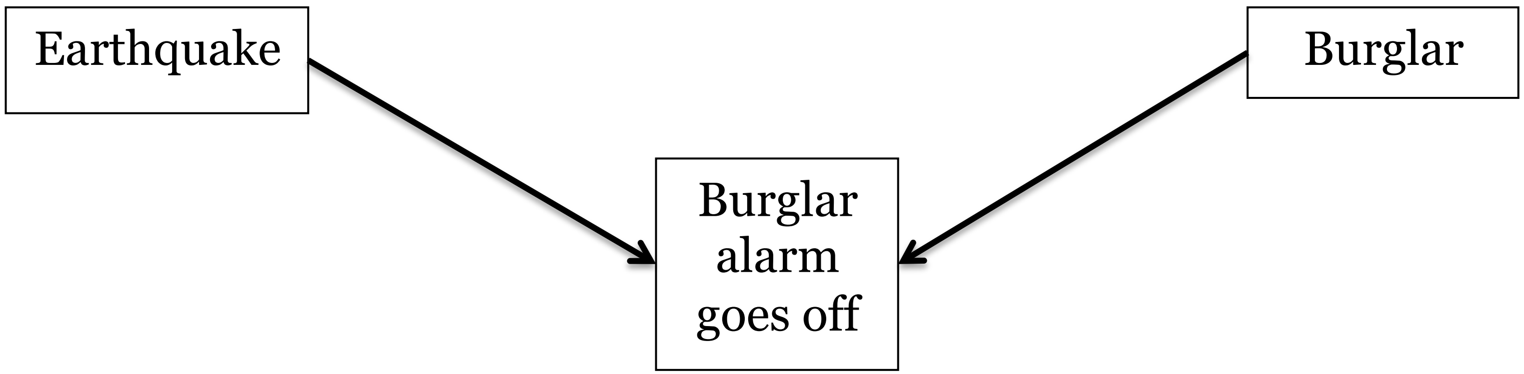
$p(\text{EQ} \mid \text{BA}) = 0.5$

$p(\text{Burglar} \mid \text{BA}) = 0.5$

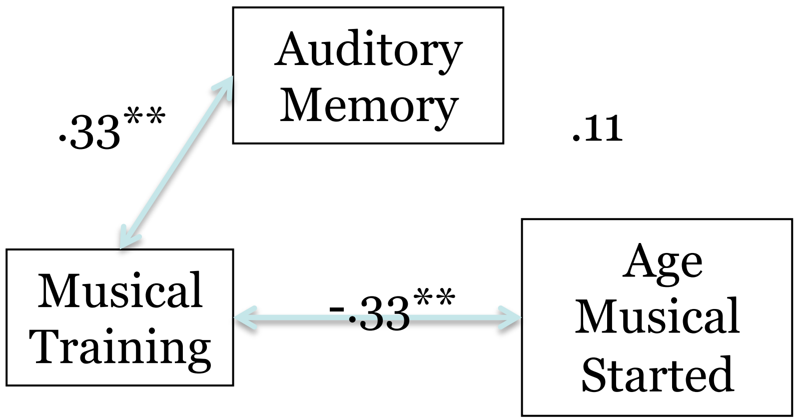
$p(\text{BA}, \text{EQ} \mid \text{Burglar}) = 0.001$

$p(\text{BA}, \text{Burglar} \mid \text{EQ}) = 0.001$

Having a burglar in the house 'explains away' the earthquake as a cause for the burglar alarm sounding



V-structure in real data



Partial correlation becomes significant where bivariate correlation is not:

r(AudMem, MusTrn | Start_Age) = .31***

r(MusTrn, Start_Age | AudMem) = -.37***

r(AudMem, Start_Age | MusTrn) = .21*

„V-Structure“

A and C cause B

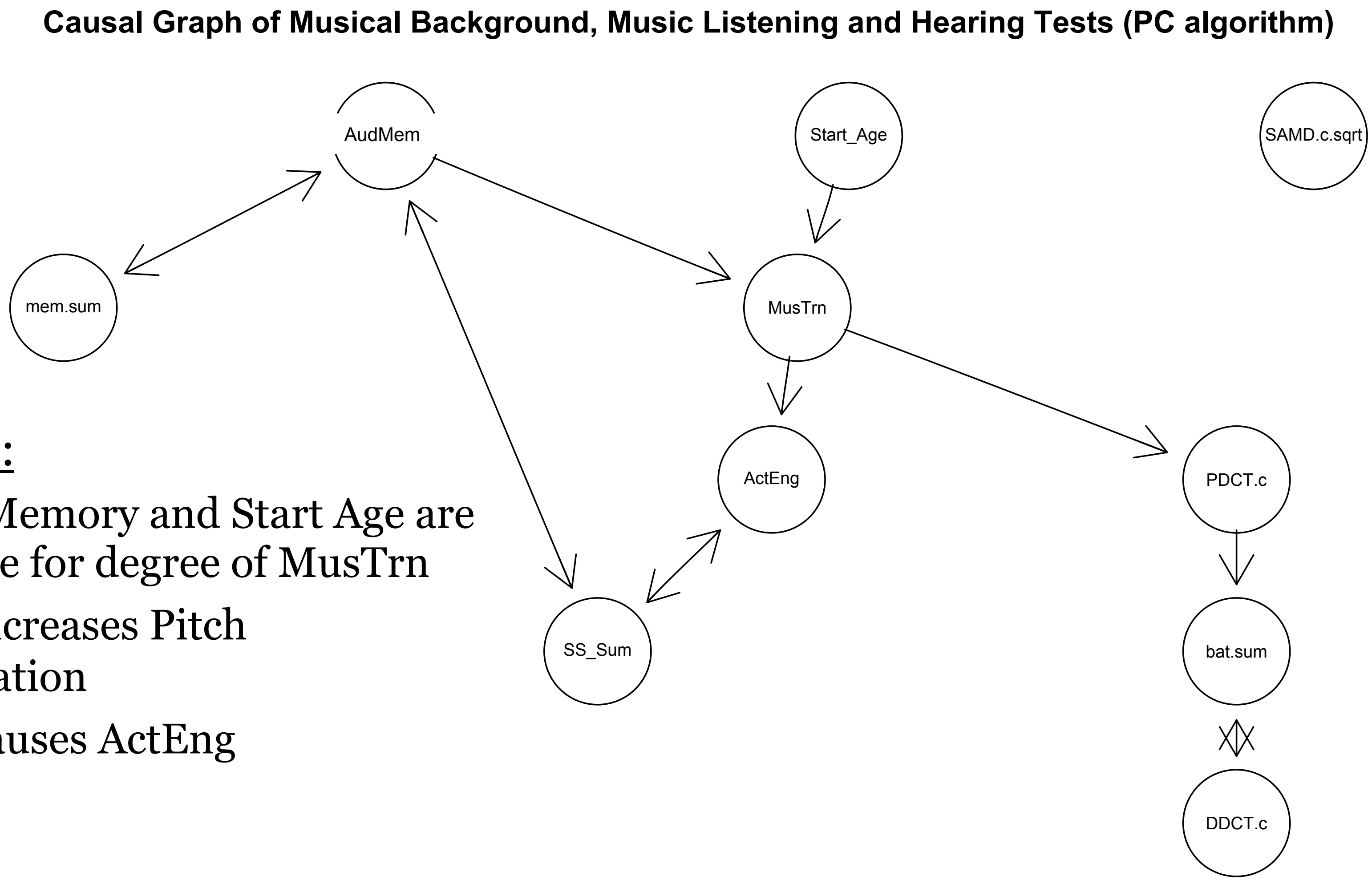
Goldsmiths
UNIVERSITY OF LONDON

The PC algorithm (Spirtes et al., 2000; Kalisch et al., 2012)

1. Start from full graph with undirected edges between all correlated variables
2. Run partial correlations tests and remove edges if two variables become independent conditional on other variables
3. Identify all V-structures and orient edges
4. Orient remaining edges (as far as possible) according to set of logical rules

Goldsmiths
UNIVERSITY OF LONDON

Output of pc algorithm: Causal network



Main results:

- Auditory Memory and Start Age are responsible for degree of MusTrn
- MusTrn increases Pitch Discrimination
- MusTrn causes ActEng

Conclusion: Invitation to causal research

– Use repeated measures or RCT where possible
– But: Correlational data is abundant in music psychology and can be used to inform causal hypotheses
– Non-parametric matching and graphical networks are two accessible techniques that enable reasoning in causal terms
– There is a lot more to discover …

Useful references

Literature:

Angrist, J. & Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: University Press.

Imbens, G. & Rubin, D. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge: University Press.

Kalisch, et al. (2012). Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, 47(11), 1-26.

Kalisch, M. & Bühlmann, P. (2014). Causal structure learning and inference: a selective review. *Quality Technology & Quantitative Management* 11, 3-21.

Morgan, S., L. (Ed), (2013). *Handbook of Causal Analysis for Social Research*. New York: Springer.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. Massachusetts: The MIT Press.

Software:

R Package *Matching*

R Package *MatchIt*

R Package *pcalg*

R Package *Simpsons*

TETRAD

CAN CORRELATIONS IMPLY CAUSATION? CAUSAL MODELING AND MUSIC PSYCHOLOGY RESEARCH

Daniel MÜLLENSIEFEN, Georgia FLORIDOU, Kelly JAKUBOWSKI

Music, Mind and Brain Group,

Department of Psychology, Goldsmiths, University of London, United Kingdom

Goldsmiths
UNIVERSITY OF LONDON