WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Center for Demography and Ecology

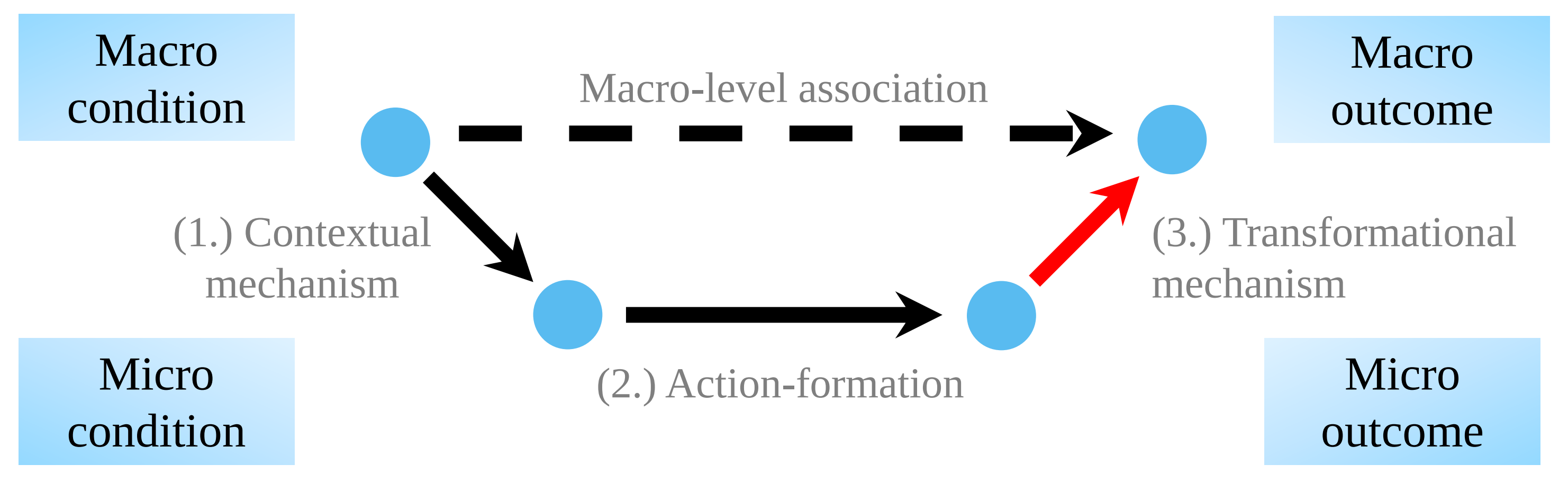
Utrecht University

The true heart of the social sciences:
the aggregation problem.
Without the aggregation problem, there is only
psychology. The problem of creating a whole from some
irreducible function of the parts, a social fact, is the
raison d'être of social science.
- Morris, Martina (2004)

Mechanism-based explanations and regression modeling

Benjamin Rosche (UW-Madison) and Jeroen Weesie (Utrecht)

The macro-micro-macro scheme



Mechanism-based explanations

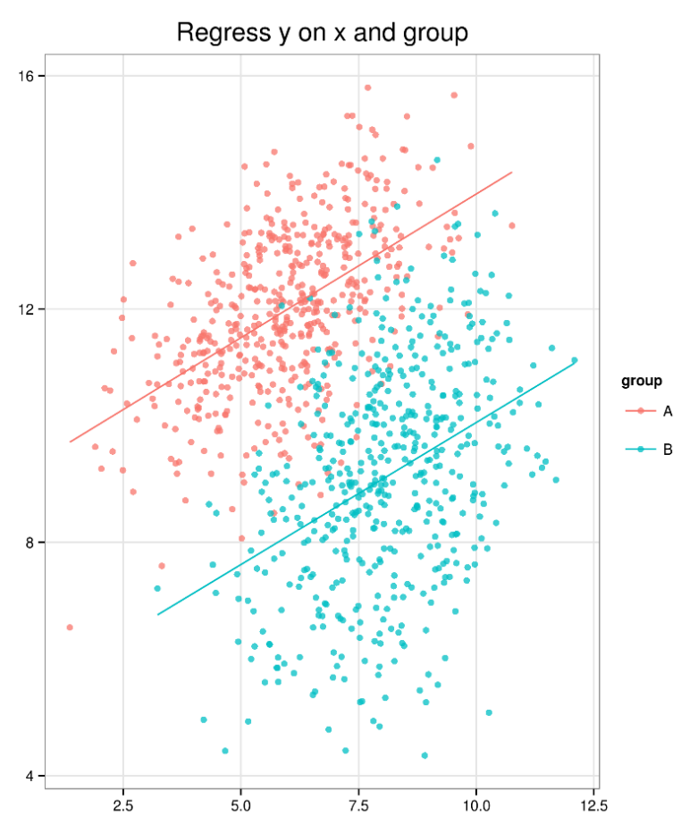
- We can conceptualize sociological phenomena that we seek to explain as outcomes at a higher (macro) level.
- To understand a phenomenon, we then delve into the processes that generate them at a lower (micro) level.
- That is, we try to understand their generative process, the ‘cogs and wheels’ inside the black box.
- To fully comprehend the generative process, we have to understand three different parts (cf. macro-micro-macro scheme; Coleman 1994):

1. How does the macro-level context influence the processes at the micro level?
2. Which processes are at work at the micro level?
3. How do the micro-level processes aggregate to a macro-level phenomenon?

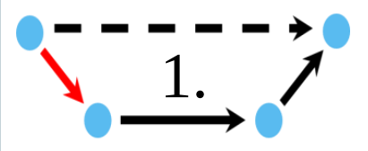
An example: Max Weber’s The Protestant Ethic and the Spirit of Capitalism
Contextual mechanism: Religious values of a society influence its members.
Action formation: Individuals adopt certain economic behaviors (anti-traditionalism, duty to one’s calling).
Transformational mechanism: How the change in economic behavior gave rise to modern capitalism is missing in Weber’s work. Maybe: Rise of capitalist economic organizations? Increasing number of entrepreneurs? Politicians creating a capitalistic regulative system?

- This approach to sociological explanations is consistent with the ‘methodological individualism’ and at the very heart of a new research agenda, analytical sociology.
- Analytical sociology places greater weight on analyzing transformational mechanisms, such as social dynamics, than on developing theories of action.

Mechanism-based explanations in regression modeling



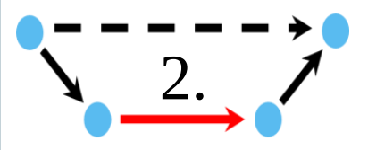
Let’s go through the three parts of a mechanism-based explanation step-by-step



The solution to explaining contextual effects is multilevel modeling:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{1ij} + e_{ij} \text{ and } \beta_{0j} = \gamma_0 + \gamma_1 Z_j + u_j,$$

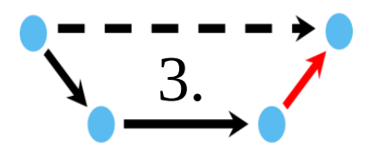
where i are the units at the micro level, j are the macro contexts



Micro-level processes can be modeled via simple regression modeling:

$$Y_i = \beta_0 + \beta_1 X_{1i} + e_i,$$

where i are the units on the micro level



Grow and Bavel (2017, p.8) in their book ‘Agent-based modeling in population studies’:

> As noted above, multilevel models make it to some extent possible to model the way in which macro-level variables affect individuals, but even with this approach it is not possible to model the processes by which individual interactions feed back into the macro level.

We will see that this is not correct

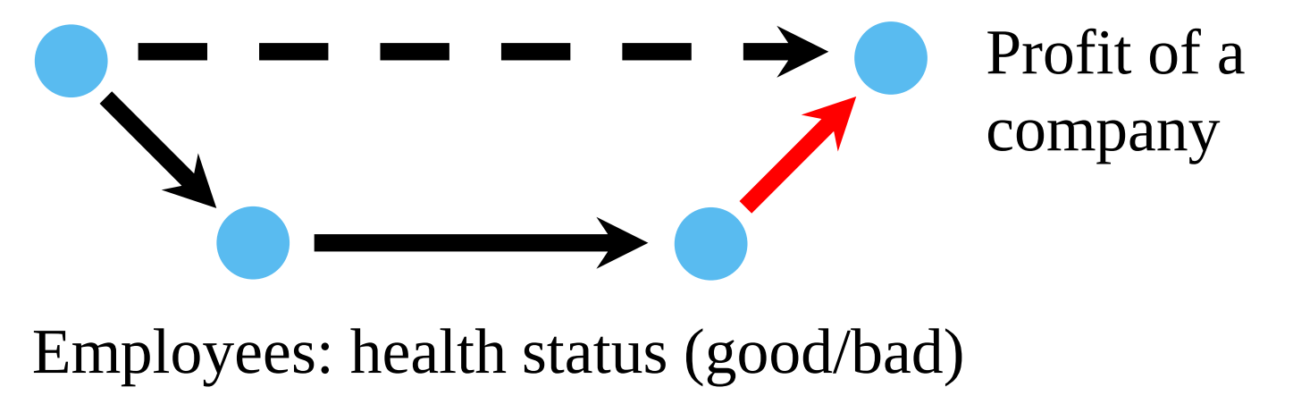
Transformational mechanisms

Our contribution: We discuss four methods to model transformational mechanisms in regression analysis. While simple dis/aggregation have been employed before, we show that social network analysis methods are better suited to model micro-macro transformations.

- There is computational (Macy 2009; Manzo 2014) and analytical work (Granovetter 1978; Buskens et al 2012) on transformational mechanisms.
- However, few (?) have considered to model the micro-macro link using regression analysis.

An example and two problematic solutions

Are healthy employees more profitable to a company?



1 Disaggregation

Employee i	Health status	Company j	Profit
1	Good = 1	1	100
2	Good = 1	1	100
3	Bad = 0	1	100
4	Good = 1	1	100
6	Bad = 0	2	80

- We can disaggregate the data (see table) and perform a regression at the micro level.
- Problem: We run the risk of drawing wrong conclusions regarding the association between leadership training and profit if we make cross-level inference on disaggregated data because we exaggerate the sample size and induce a problematic independence assumption.
  - The profit of company 1, was observed only once and by disaggregating the data, we miraculously multiply this number.
  - We treat employee 1 to 4 as independent observations. That is, we assume that the observed values for employee 3 are independent of employee 1, 2 and 4.
- Consequence: Excessive Type I error!

2 Aggregation

- We aggregate the information on employees to the company level, e.g. $\sum_{i \in c(1)} X_i = 3/4$ of the employees are in good health at company 1, and perform a regression at the macro level.

$$Y_j = \beta_0 + \beta_1 \sum_{i \in c(j)} X_i + e_j$$

where i are the units at the micro level (here: employees), j are the macro units (here: companies).

- Problem: While we account for unobserved differences at the company level (e_j), there is no residual term at the employee level!
- We thereby implicitly assume that differences among employees are completely described by their health status.
- Consequence: Excessive Type I error! Any unobserved variation at the employee level will affect our estimates:
  - Biased standard errors in case of linear regression
  - Biased standard errors & regression coefficients in case of logistic or event-history models
- A better solution: the aggregation problem can be conceptualized as social network problem.

Two better solutions: social network analysis

- By treating employees as *alters* and the companies as *egos*, which are affected by their employees, network autocorrelation models (originally developed as spatial autocorrelation models) can be used:

3 Network effect model and Network disturbance model

$$Y_j = \beta_0 + \beta_1 X_{1j} + \rho \sum\nolimits_{i \in C(j)} \boldsymbol{w_{ij} Y_i} + e_j \quad Y_j = \beta_0 + \beta_1 X_{1j} + \rho \sum\nolimits_{i \in C(j)} \boldsymbol{w_{ij} e_i} + u_j$$

ρ as estimated (auto)correlation in the outcome and the in residual, respectively.

Problems:

- In our example, we are interested in the impact of *features* of employees (X=leadership training) and not the impact of *outcomes* (Y) or its *residual* (e).
- Moreover, the regression equation at the micro and macro level needs to be the same for network autocorrelation models to make sense. However, in our case, we model two different (dependent) variables at each level, health status and profit, respectively.
- Finally, network autocorrelations models are restricted to quantitative (continuous) dependent variables. An equivalent solution from the realm of multilevel modeling comes with a whole suite of options.

4 Multiple membership multilevel model (Goldstein 2011a, chapter 13)

$$Y_j = \gamma_0 + \beta_1 X_{1j} + \sum\nolimits_{i \in C(j)} \boldsymbol{w_{ij} \theta_i} + e_j$$

$$\text{and } \boldsymbol{\theta_i = \gamma_1 Z_i + u_i}$$

where i are the units at the micro level (here: employees),
j are the macro contexts (here: companies)

- Multilevel approach: alters as context of ego. The multiple membership version of the multi-level model allows egos to be nested in multiple contexts (i.e. all employees of company i)
- Observed features of employees (fixed effect Z_i) as well as a residual effect of unobserved employee features (random effect u_i) can be added and specified separately from the company-level regression
- Estimated measures of network effects: Variance of the random parameter $U_i \sim N\left(0, \boldsymbol{\sigma_U^2}\right)$ and the intra-class correlation $ICC = \frac{\sigma_U^2}{\sigma_E^2 + \sigma_U^2}$ (as compared to the autocorrelation parameter)
- Many extensions possible because multilevel literature well developed: binary and multivariate outcomes, other forms of dependencies, …
- The weights represent the theory on how the employee effects aggregate to the company level (Leenders 2002)
  - Do we expect that all employees have the same effect? Or do we think that health status only affects a company’s profits if, say, more than 50% are in good/bad health?
  - In other words, the weights w_{ij} specifiy the transformation function, which can be
    - ✓ arithmetic average, (weighted) sum
    - ✓ more complex aggregation functions: threshold functions, other emergence mechanisms
    - ✓ estimated from the data: regression of unobserved w_{ij}^* on observed explanatory variables
    - ✓ E.g. 50% threshold function $w_{ij} = \begin{matrix} w_{ij}^* \text{ if } \frac{1}{n_i} \sum_{j \in C(i)} X_j > 0.5 \\ 0 \quad \text{otherwise} \end{matrix}$

References, further reading, and applications

References and further reading
- Buskens, V., W. Raub, and M. van Assen (eds.). 2012. Micro-Macro Links and Micro-Foundations, London: Routledge
- Coleman, James S. 1994. Foundations of Social Theory. Cambridge, Mass.: Belknap Press.
- Goldstein, Harvey. 2011a. Multilevel statistical models. 4th ed. Wiley Series in Probability and Statistics. Oxford: John Wiley & Sons. (esp. Chapter 13).
- Grow, A, Van Bavel, J. 2017. Agent-based modeling in population studies. Springer.
- Granovetter, M. 1978. Threshold models of collective behavior. American Journal of Sociology, 83(6), 1420-1443.
- Leenders, R. T. A. 2002. Modeling social influence through network autocorrelation. Social Networks, 24(1), 21-47.
- Macy et al. 2009. Life in the network: the coming age of computational social science. Science 323(5915), 721–723.
- Manzo, G. (Ed.). 2014. Analytical sociology: Actions and Networks. Oxford: John Wiley & Sons.

Applications
- Goldstein, Harvey. 2011b. Estimating research performance by using research grant award gradings. Journal of the Royal Statistical Society: Series A (Statistics in Society) 174 (1): 83–93.
- Rosche, Benjamin. 2018. Putting parties into the analysis of government survival using multilevel modeling. Unpublished Master’s Thesis. Utrecht University.

This research was supported by the core grants to the Center for Demography and Ecology at the University of Wisconsin-Madison (P2C HD047873) and to the Center for Demography of Health and Aging at the University of Wisconsin-Madison (P30 AG017266).

Contact: benrosche.net