Open Data, Reproducibility & the Reliability of Scientific Knowledge

Theresa Velden

Technische Universität Berlin

Zentrum für Technik und Gesellschaft (ZTG)

4S/EASST Barcelona

Sept 1, 2016

Project Overview

- Topic: Inquiry into Link Between Reliability of Scientific Knowledge, Scientific Replication, and Open Data
- Aim: Explore variation across scientific fields
- Approach: Secondary Literature Analysis & Field Research (Ethnographic & Expert Interviews)
- Funding: IPODI Fellowship of Technische Universität Berlin

`Reproducibility Crisis´

Overselling & questionable methods:

E.g. Large-scale replication effort in psychology could replicate only one third of significant results. Effect sizes were reduced on average by about 50%. (Open science Initiative 2015).

Scientific error:

E.g. in 2009, three clinical trials at Duke University suspended because selection of cancer treatments based on irreproducible genomic signatures (Ioannidis & Khoury 2011, Baggerly 2010, Baggerly & Coombes 2009)

Data Falsification:

Major cases of scientific fraud discovered e.g. in 2002 in physics, 2005 in stem cell research, 2010 in crystallography, 2011 in social psychology …

Data Sharing as a Remedy

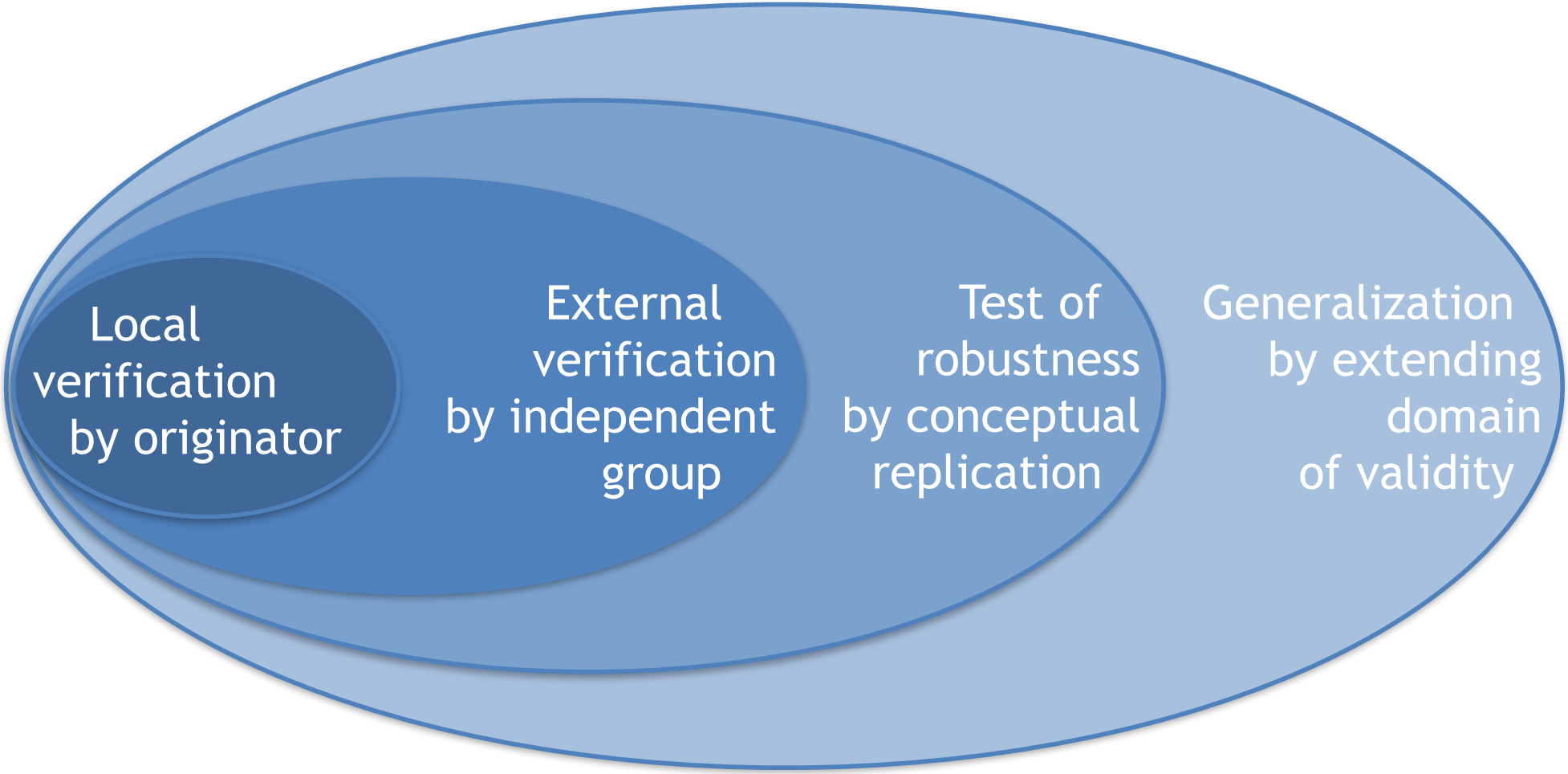
- Calls for “Reproducible research”
  - sharing of data, code, protocols, etc.
- Push For Data Sharing by Funding Agencies (NSF, NIH, DFG)
- Little data sharing happening so far
- Push-back from scientists who fear waste of resources

Open Questions

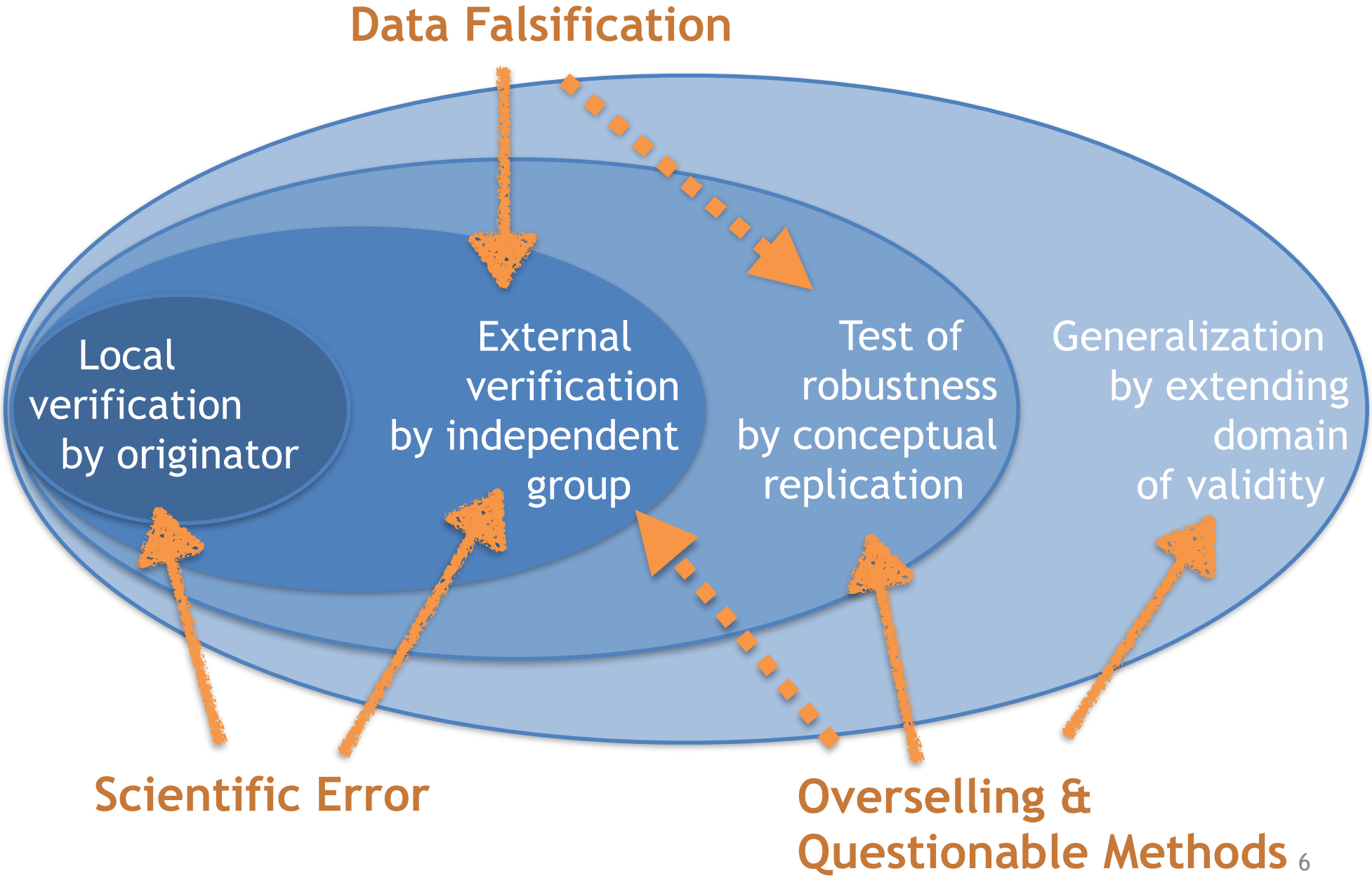
Will open data help improve the reliability of scientific knowledge?

How exactly are the reliability of scientific knowledge, scientific replication and data sharing are linked?

Range of Replication Purposes



Range of Replication Purposes



State of Research
Data Sharing

- Studies focus on data sharing for re-use
  - Link between data sharing and replication not well articulated
- **Obstacles to data sharing**
  - Lack of resources (infrastructure, time, reward) [Tenopir 2011]
  - Valuation of data in production contexts [Vertesi & Dourish 2011, Akmon 2013]
  - Field specific tensions between competition (secrecy) and cooperation (sharing) [Velden 2013]
  - `Data Streams', not unambiguous, bounded entities [Hilgartner 1997]
  - Data sharing as relational practice [Mauthner & Parry 2013]

State of Research

Scientific Replication Practices

- Role of Replication?
  - `Gold standard´ of science (Jasny et al 2011, Popper 1959)
  - Experimenter's regress (Collins 1992)
  - Feasibility elusive (Borgman 2012)
- Scarcity of empirical evidence on field variation in type and prevalence of replication (Zuckerman 1977)

State of Research

Scientific Replication Practices

1. ***Explicit replication*** versus ***implicit replication***
   - Explicit replication rare (Gläser 2006, Collins 1981, Zuckerman 1977)
   - Implicit replication routine in some fields (Velden 2013 - synthetic chemistry)
2. ***Exact replication*** versus ***conceptual replication***
   - Stroebe & Strack (2014) applied sciences —> exact replication; basic sciences —> conceptual replication

Image credits (clockwise): microbiologybites (farn), NIH-NEI (chemist), solarnu (CERN), Steve Jurvetson (DNA analyzers) , @matylda (hackatron)

Next steps in comparative study

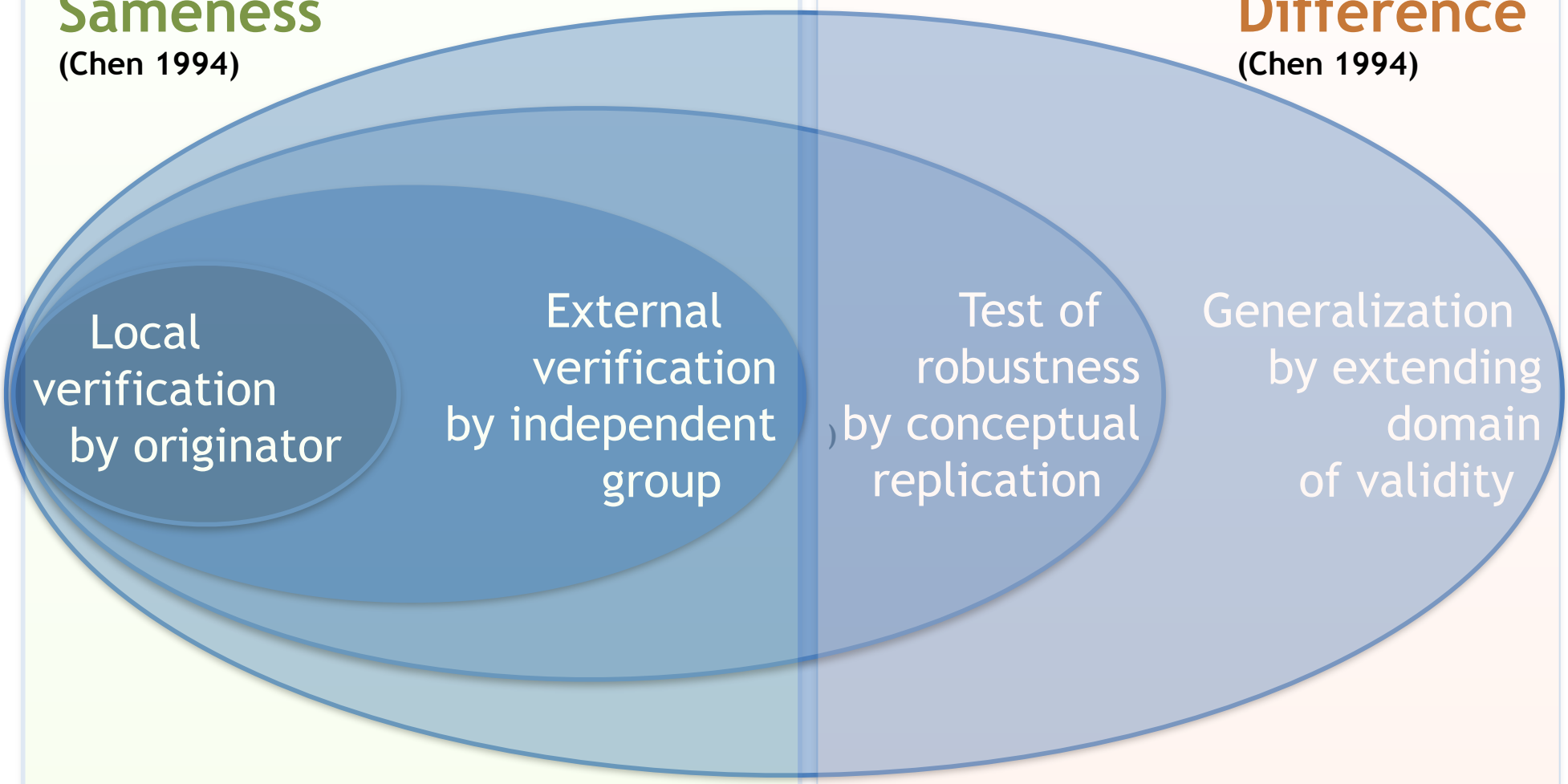
Cont'd Secondary Analysis of Literature

1. How do replication practices vary across fields & and when do they require data sharing?
2. How is data sharing for enabling scientific replication supported across fields?

—> Inform Design of Field Study (Case selection)

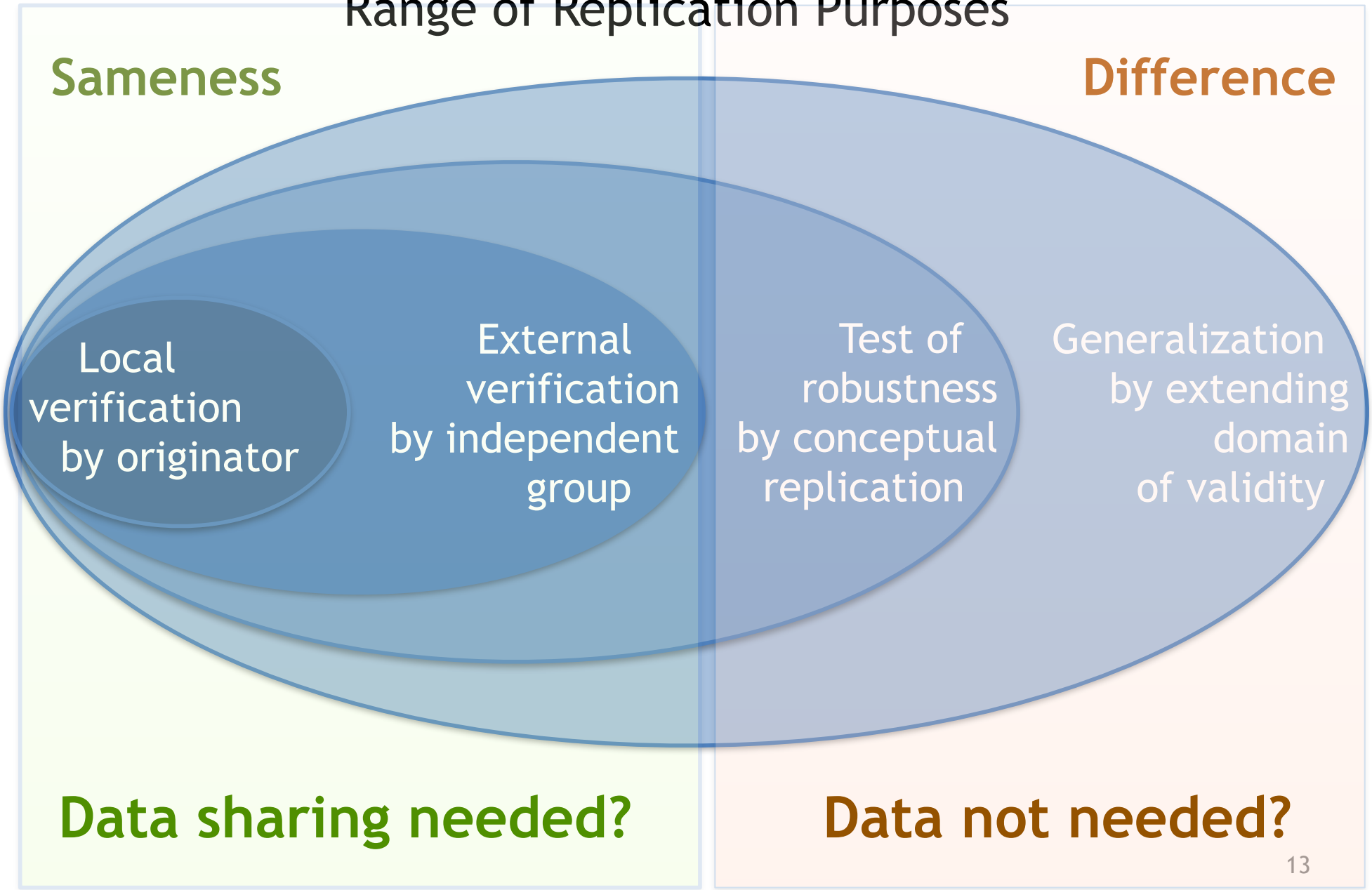
Discussion Point

Range of Replication Purposes



Discussion Point

Range of Replication Purposes



Thank You

Theresa Velden

velden@tu-berlin.de

(Some) References

Akmon, D. (2013). “The Role of Conceptions of Value in Data Practices: A Multi-Case Study of Three Small Teams of Ecological Scientists”, PhD Thesis, University of Michigan. Available online at: http://deepblue.lib.umich.edu/handle/2027.42/107162

Borgman, C. L. (2012). The conundrum of sharing research data. Journal of the American Society for Information Science and Technology, 63(6), 1059–1078.

Collins, H. (1992). Changing order: Replication and induction in scientific practice. University of Chicago Press.

Gläser, J. (2006). Wissenschaftliche Produktionsgemeinschaften. Die soziale Ordnung der Forschung. Campus Forschung Band 906, Frankfurt/New York.

Ioannidis, J. P. A. & Khoury, M. J. (2011). Improving Validation Practices in “Omics” Research. Science, 334(6060): 1230-1232.

Stroebe, W. & Strack, F. (2014). The alleged crisis and the illusion of exact replication. Perspec- tives on Psychological Science, 9(1), 59-71

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. PloS one, 6(6), e21101.

Velden, T. (2013). Explaining Field Di↵erences in Openness and Sharing in Scientific Communities. CSCW 2013: Computer Supported Cooperative Work, February 23 – 27 2013, San Antonio

Vertesi, J., & Dourish, P. (2011). The value of data: considering the context of production in data economies. In Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (pp. 533-542). ACM.

Zuckerman, Harriet (1977). ‘Deviant Behaviour and Social Control in Science?, in E. Sagarin (ed.), Deviance and Social Change, Sage Publications, Beverley Hills, pp. 87–138.