

Coding Verbal Data

What to Optimize?

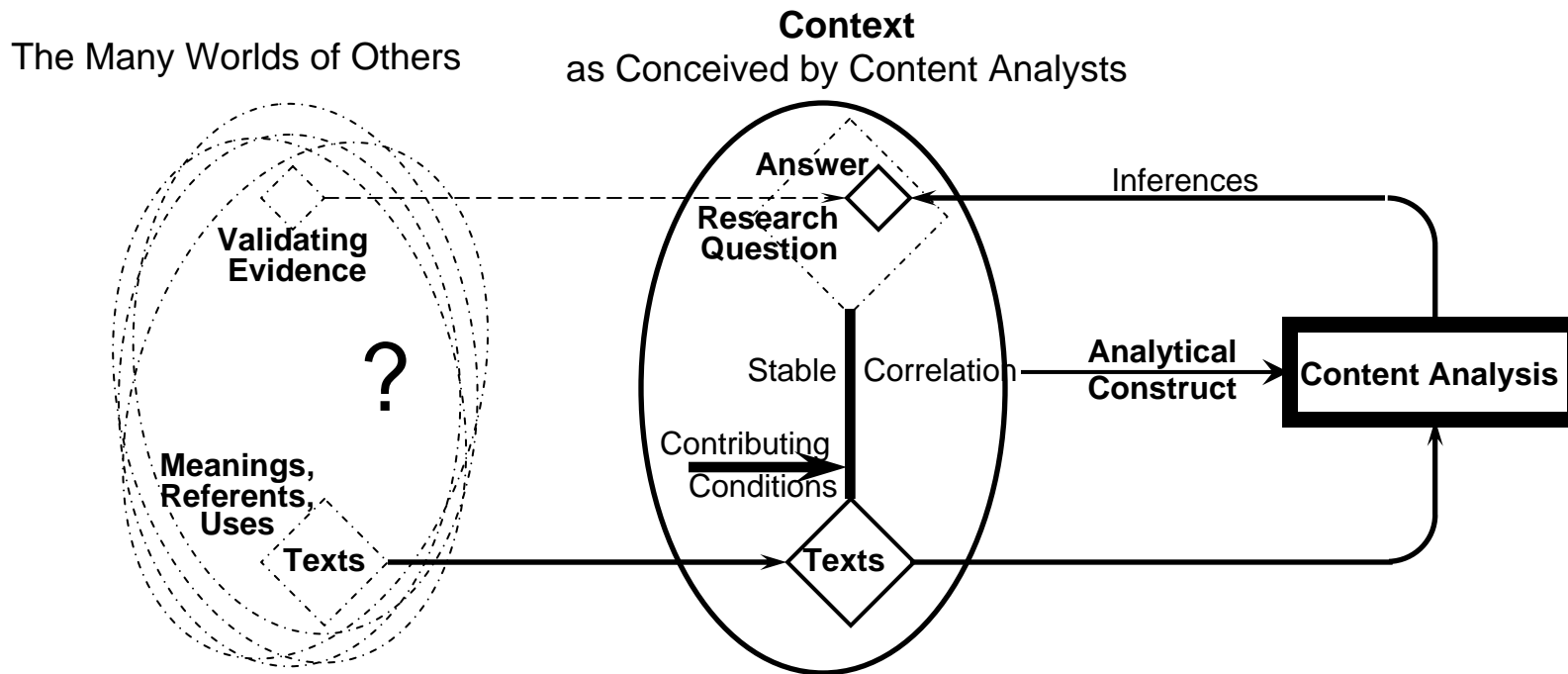
Klaus Krippendorff

Gregory Bateson Term Professor for Cybernetics, Language, and Culture
The Annenberg School for Communication
University of Pennsylvania, Philadelphia
kkrippendorff@asc.upenn.edu

Outline

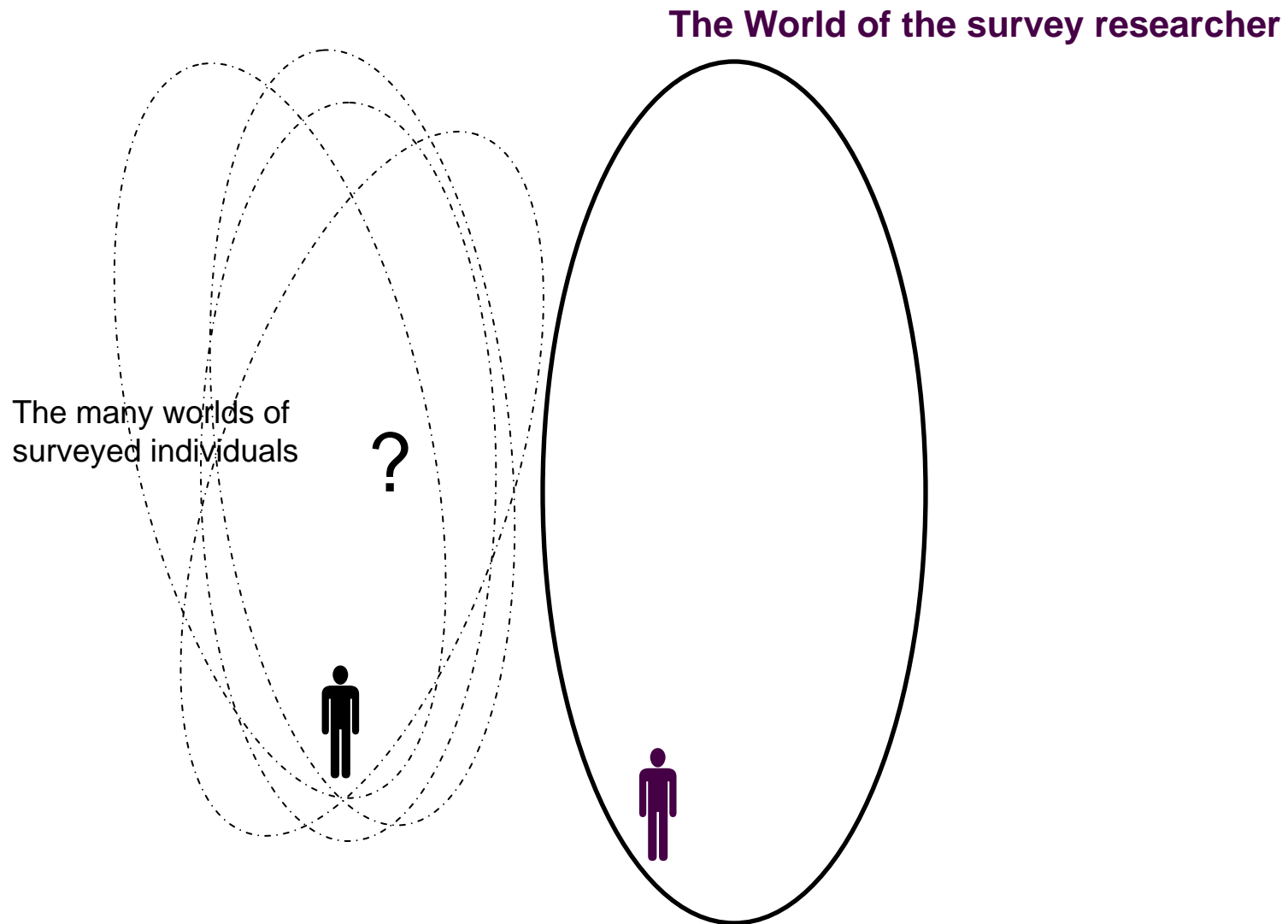
- **Familiar framework for content analysis**
- **Proposed framework for survey research**
- **Coding criteria a code needs to optimize**
- **How coding criteria interact**
- **Hopes for the content analysis of open-ended survey questions**
- **Real problems faced by coding open-ended survey questions**

Familiar framework for content analysis

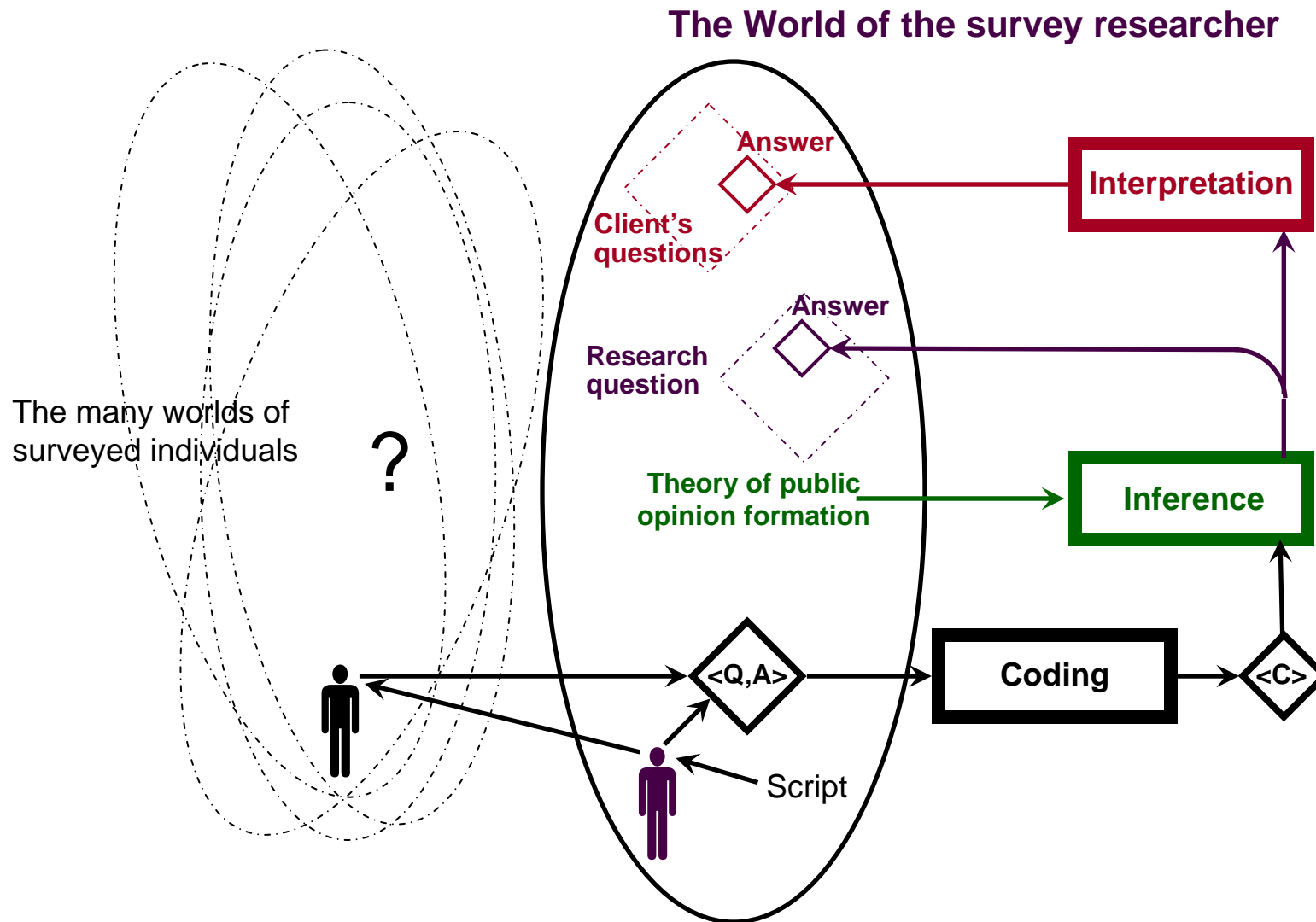


- Krippendorff, Klaus (2004). *Content Analysis; An Introduction to Its Methodology*. 2nd Ed. Thousand Oaks, CA: Sage
- Krippendorff, Klaus and Bock, Mary Angela (Eds.) (2009). *The Content Analysis Reader*. Thousand Oaks, CA: Sage

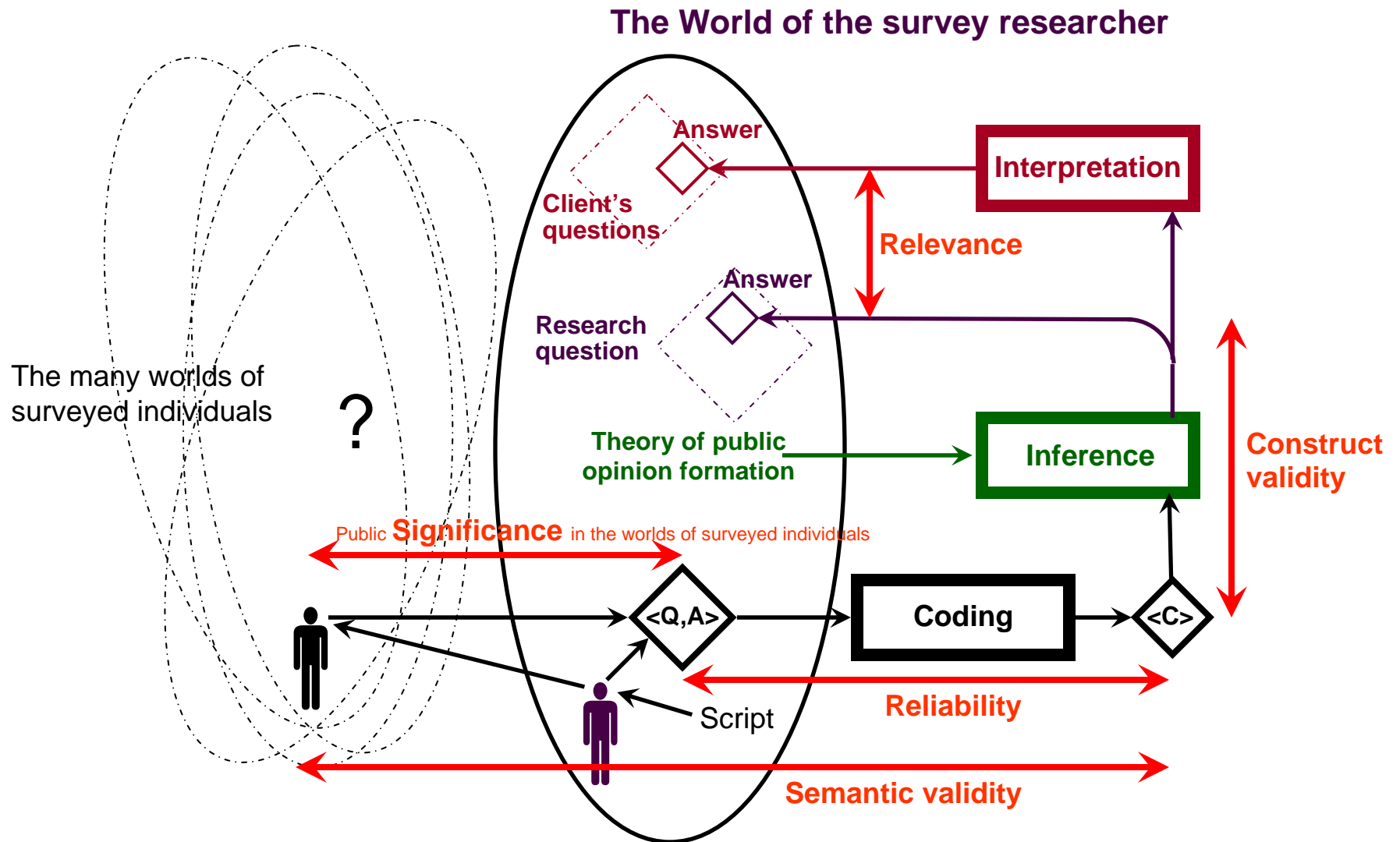
Proposed framework for survey research



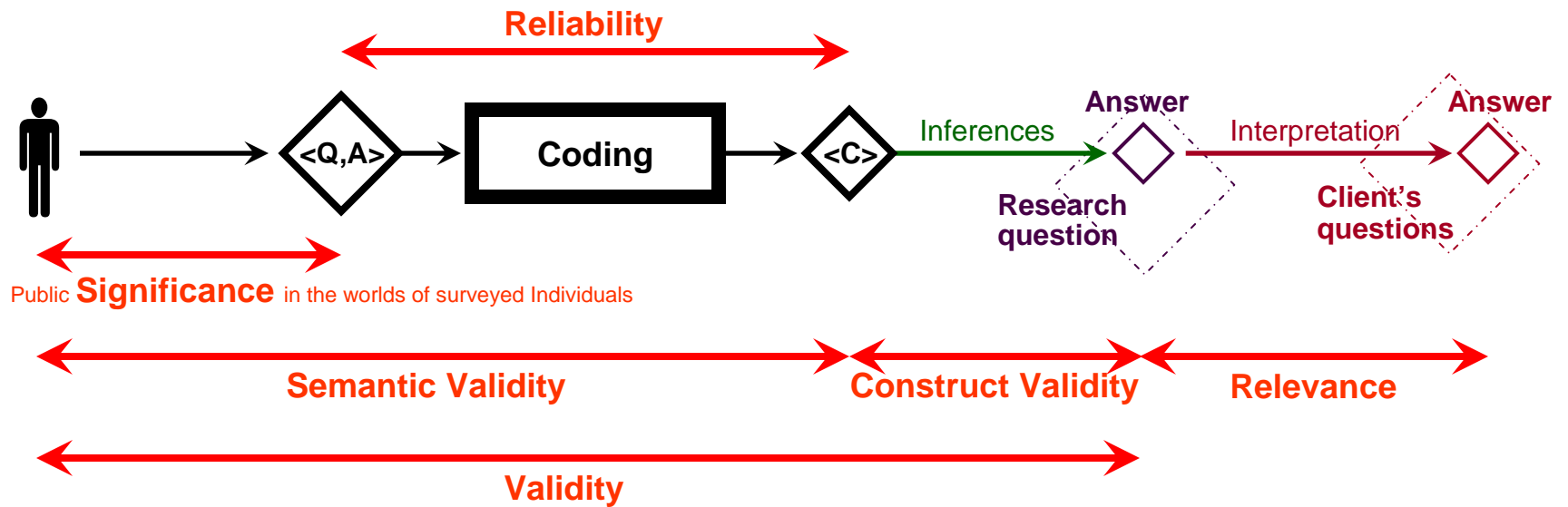
Proposed framework for survey research



Coding criteria a code needs to optimize



Coding criteria a code needs to optimize



How coding criteria interact

Human coders are reliable mainly when judgments are simple and conventional

Generally:

As reliability improves, semantic validity goes down – presenting a typical problem for optimization, tradeoffs

Computer content analysis obviates reliability but has major problems with semantic validity

Most software developers do not validate their claims

Sometimes:

Processing massive amounts of data compensates for the shallowness of computer reading of text – presenting another problem for optimization

Computer-**aided** coding probably is optimal – leaving reading to human judges and clerical work to computers

Hopes for the content analysis of open-ended survey questions

Traditional content analysis

Unobtrusive

Much irrelevant and unstructured text

Context and nature of **units** vary widely

Research questions are often **unique**

Analytical constructs are mostly ad hoc

Standardized content analyses are rare

Coding 1 : 1

Usually one category for each of many variables

Data = **points** in multi-variable spaces

Coding open-ended interview answers

Obtrusive

Data are more relevant, pre-structured by questions

Public significance of individual answers a problem

Units = convenient question-answer pairs

Questions = **context** for answers

Research questions are **recurrent**

Chance to develop valid **analytical constructs**

Standardizing analytical procedures makes sense

Coding 1 : many

Multiple categories for each of many variables

Data = **vectors of variable dimensionality**

Real problems faced by coding open-ended survey questions

Comments on **the most important problems facing the nation**

- There are multiple answers, mostly logically independent = vectors
- Some are single words = concepts, some are factual propositions
- Coding is 1 : 1, a seemingly arbitrary selection from several responses

Comments on **likes and dislikes about a candidate**

- There are multiple answers, not logically independent
E.g., “everything about him” – a mixture of concepts, personality attributes, positions, comparisons to Bush
- Most answers are factual propositions, predictions, quotations
- Some are reasoned, some are stated
- Many express respondent’s wishes, hopes, fears, uncertainties, agreements
- Often highly context dependent (on previous answer) e.g. “Abortion” “Family values”
- Some are individual opinions, some are public opinions (opinions of others)
There are multiple answers, mostly

Semantic validity and reliability is not achievable by a simple coding scheme

There is much more to be done than I initially thought

We need to embrace the challenge without ignoring the criteria coding needs to optimize

Thank you for listening

Klaus Krippendorff
The Annenberg School for Communication
University of Pennsylvania
kkrippendorff@asc.upenn.edu

Coding criteria a code needs to optimize

