National Cancer Institute

Questionnaire Design, Development, Evaluation, and Testing: Where in the World Are We?

Gordon Willis, PhD
National Cancer Institute, NIH

QDET2 Keynote Address
November 10, 2016
Miami, Florida

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health
NIH...Turning Discovery Into Health®

Cognitive Interviews of Questionnaire Designers/Evaluators

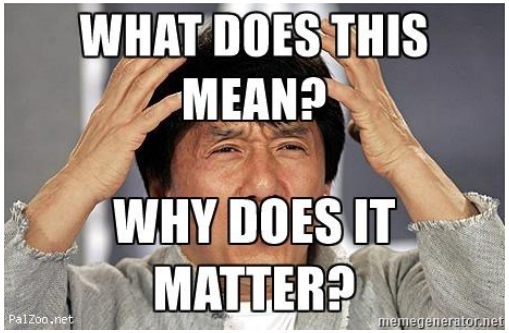
Special thanks to:

- Paul Beatty, US Census Bureau
- Jen Dykema, University of Wisconsin
- Jack Fowler, University of Massachusetts
- Emily Geisen, RTI
- Kerry Levin, Westat
- Kristen Miller, NCHS/CDC

Questions I address in my presentation

(I)

(II)



(III)

(I) Where are we… as a discipline ?

- "The world is full of well-meaning people who believe that anyone who can write plain English and has a modicum of common sense can produce a good questionnaire"
– (A. Oppenheim, 1966)

- Evidence?
  - Many examples of misguided applications
  - Some predating any attempt to develop QDET ->

1929 Little Review: American Literary Magazine

https://magmods.wordpress.com/2010/10/04/what-is-the-questionnaire/

QUESTIONNAIRE

1. What should you most like to do, to know, to be? (In case you are not satisfied).
2. Why wouldn't you change places with any other human being?
3. What do you look forward to?
4. What do you fear most from the future?
5. What has been the happiest moment of your life? The unhappiest? (if you care to tell).
6. What do you consider your weakest characteristics? Your strongest? What do you like most about yourself? Dislike most?
7. What things do you really like? Dislike? (Nature, people, ideas, objects, etc. Answer in a phrase or a page, as you will).
8. What is your attitude toward art today?
9. What is your world view? (Are you a reasonable being in a reasonable scheme?)
10. Why do you go on living?

A short history of questionnaire design models/theory

- Cantril and Fried (1944): Prehistoric cognitive probing
- Payne (1951): Rules – ***cookbook*** rather than ***theory***
- Lansing, Ginsburg, and Braaten (1961), U IL:
  - ***An Investigation of Response Error***
- Cannell, Miller, Oksenberg (1981), U Michigan
- ***Turning point*****:** Tourangeau (1984): First CASM conference--Four-stage cognitive model –>>>

The Four Horsemen of the Questionnaire Design Apocalypse (Tourangeau, 1984)

Cognition and Survey Methodology (CASM) as a catalyst for change

- CASM: Interdiscipline between cognitive psychology, survey methods

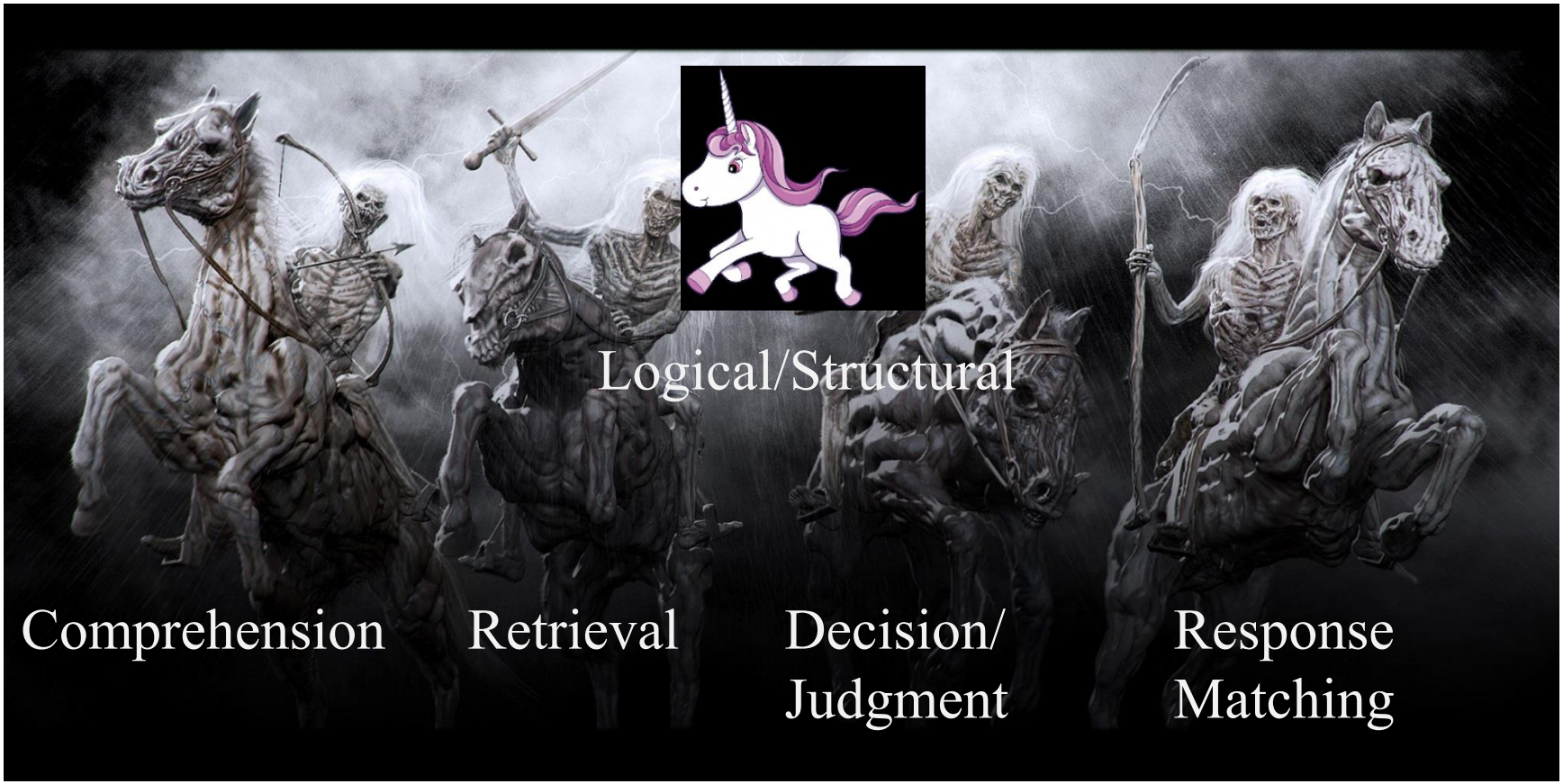
- Cognitive Laboratories (1980's):
  - Monroe Sirken, NCHS Cognitive Lab
    Census Bureau, BLS…
    Applied Lab as a tangible enterprise:
  - Took advantage of adaptation of ***Think-Aloud*** (Ericsson-Simon) as advocated by Elizabeth Loftus, and ***Verbal Probing***

Beyond Cognition

- Beyond HOW to ask: WHAT to ask- ***are we asking the right Q*** in the first place?
- This notion has persisted, under different guises:
  - Classically, as ***Specification Error*** – Lazarsfeld (1986) -- we want one thing, but measure another
  - Recognition that 4-stage cognitive model didn't capture all problem types:
    - Willis, Royston, Bercini (1991): Added ***Logical/Structural errors***
    - Problems are more 'in the question' than 'in the respondent'

      -- So, we need a fifth horseman ->>>

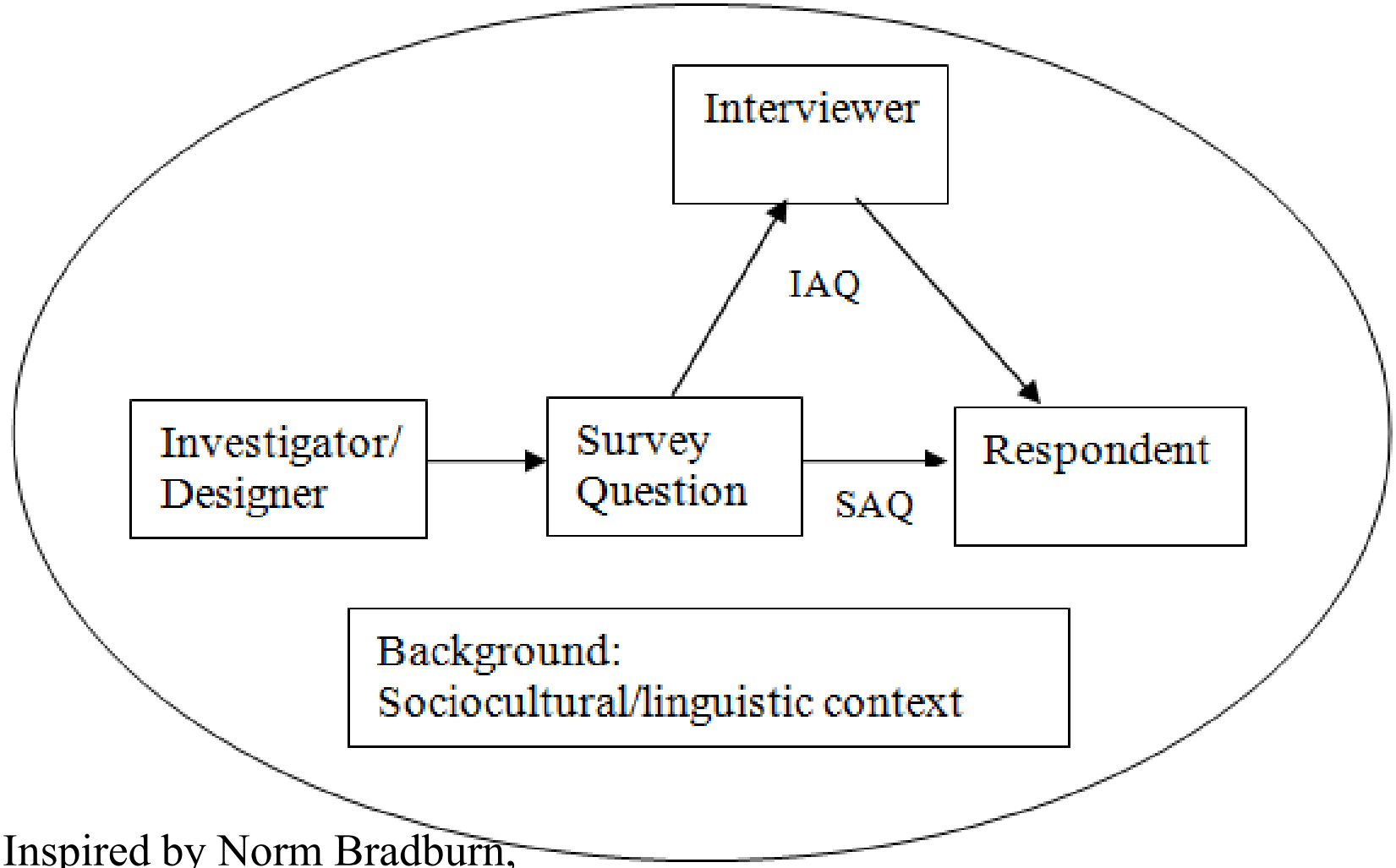
The Four+1 Horsemen of the Questionnaire Design Apocalypse



Socio/Anthro extensions to theory

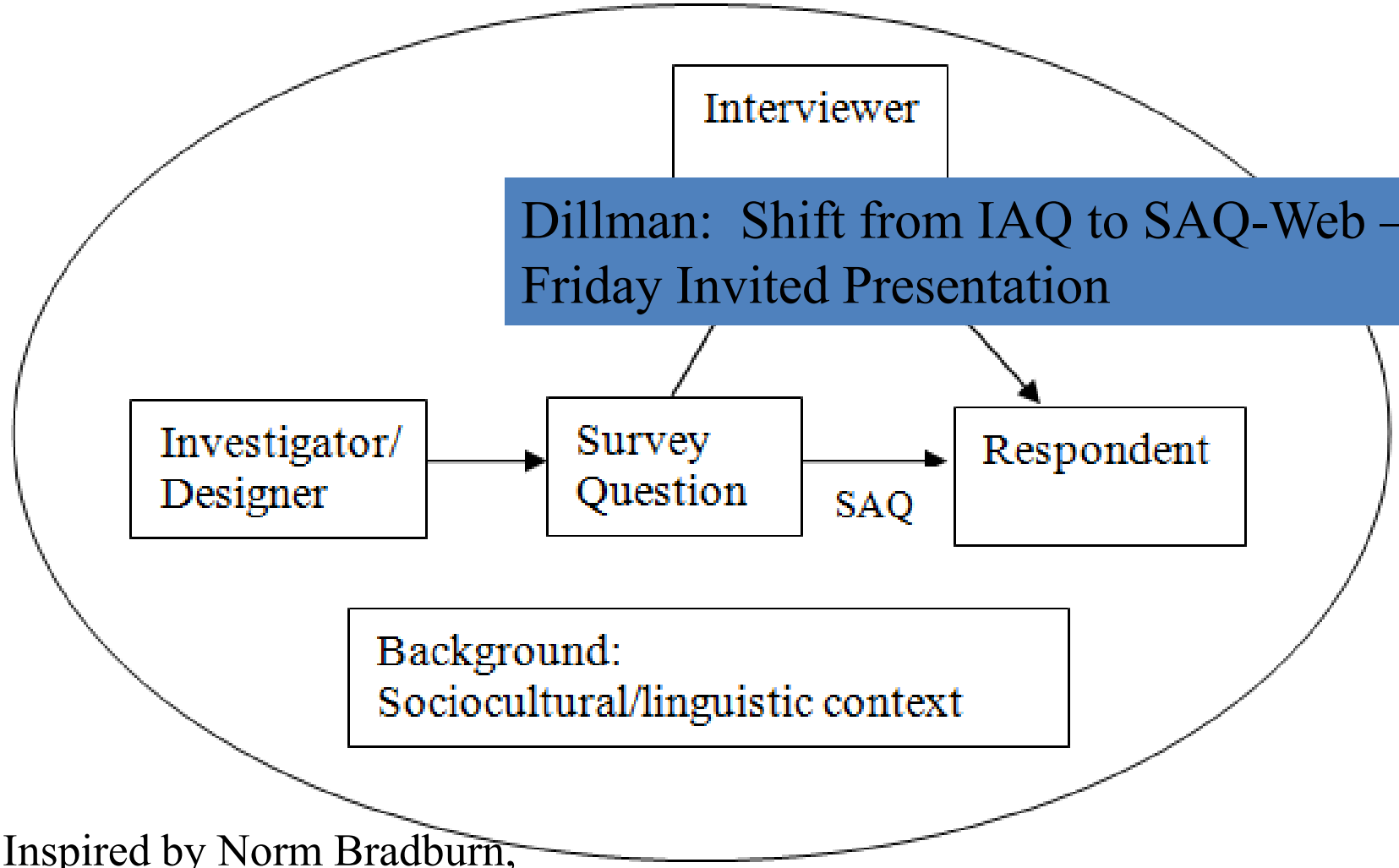
- The logical extension of CASM arguing that "there's something else out there" – linguistics, sociology, anthropology…
  - Eleanor Gerber, Census; Kristen Miller, NCHS
  - Royston: Little of what we do is really 'cognitive'; Gerber and Wellens (1997) – "Cognition" in the Cognitive Interview?
  - Schaeffer and Dykema (2011): Cognitive versus Interactional models
  - Multi-interdisciplinary view was pronounced by time of CASM II in 1997
  - Landscape shifted: CASM III → MIST->QUEST, QDET, QEM
  - Focus is on ***methods*** – how much is 'theory' involved?
    - Dillman: Our methods are dependent variables…

From Lite to Saturated model*



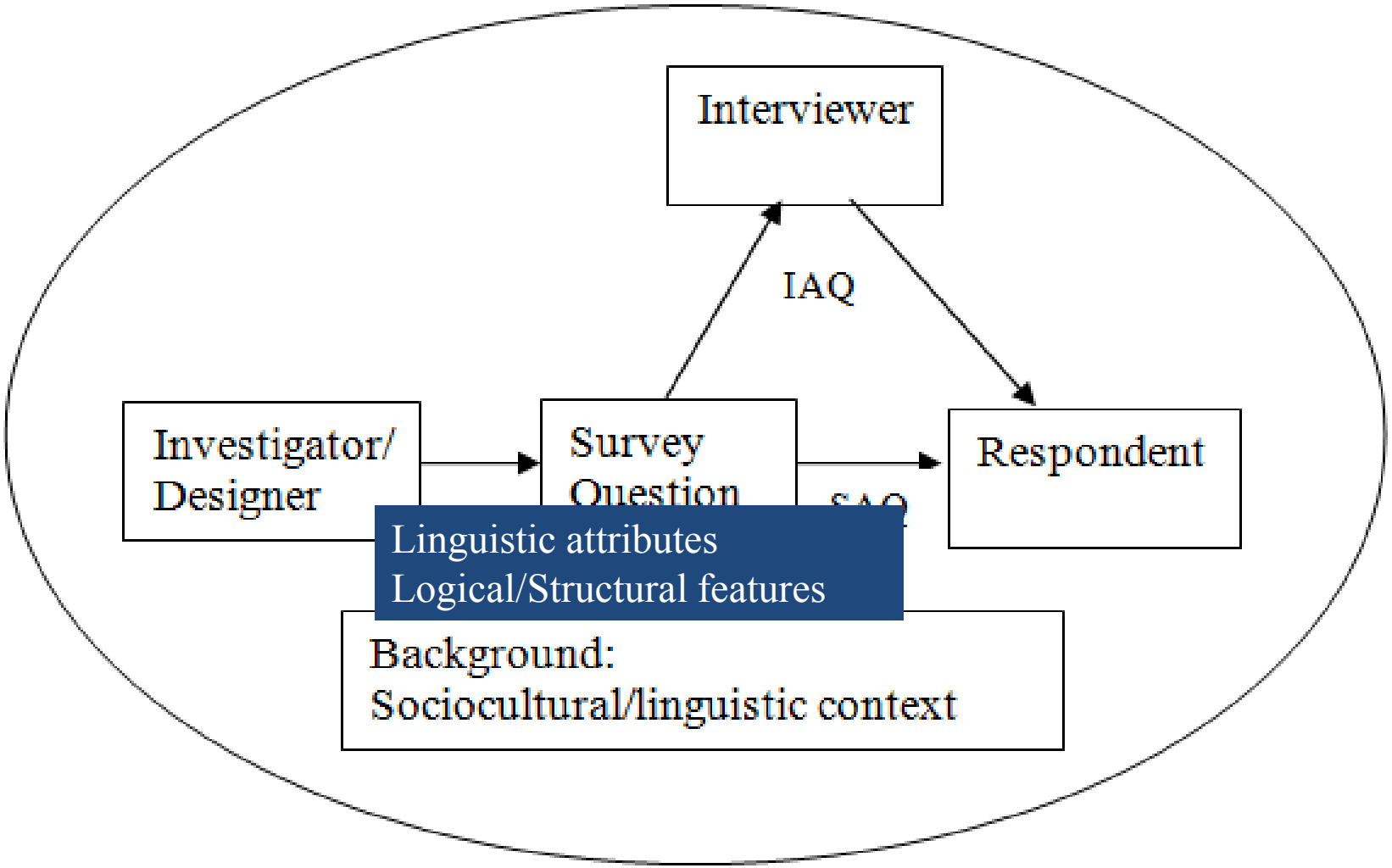
*Inspired by Norm Bradburn, QDET1

From Lite to Saturated model*

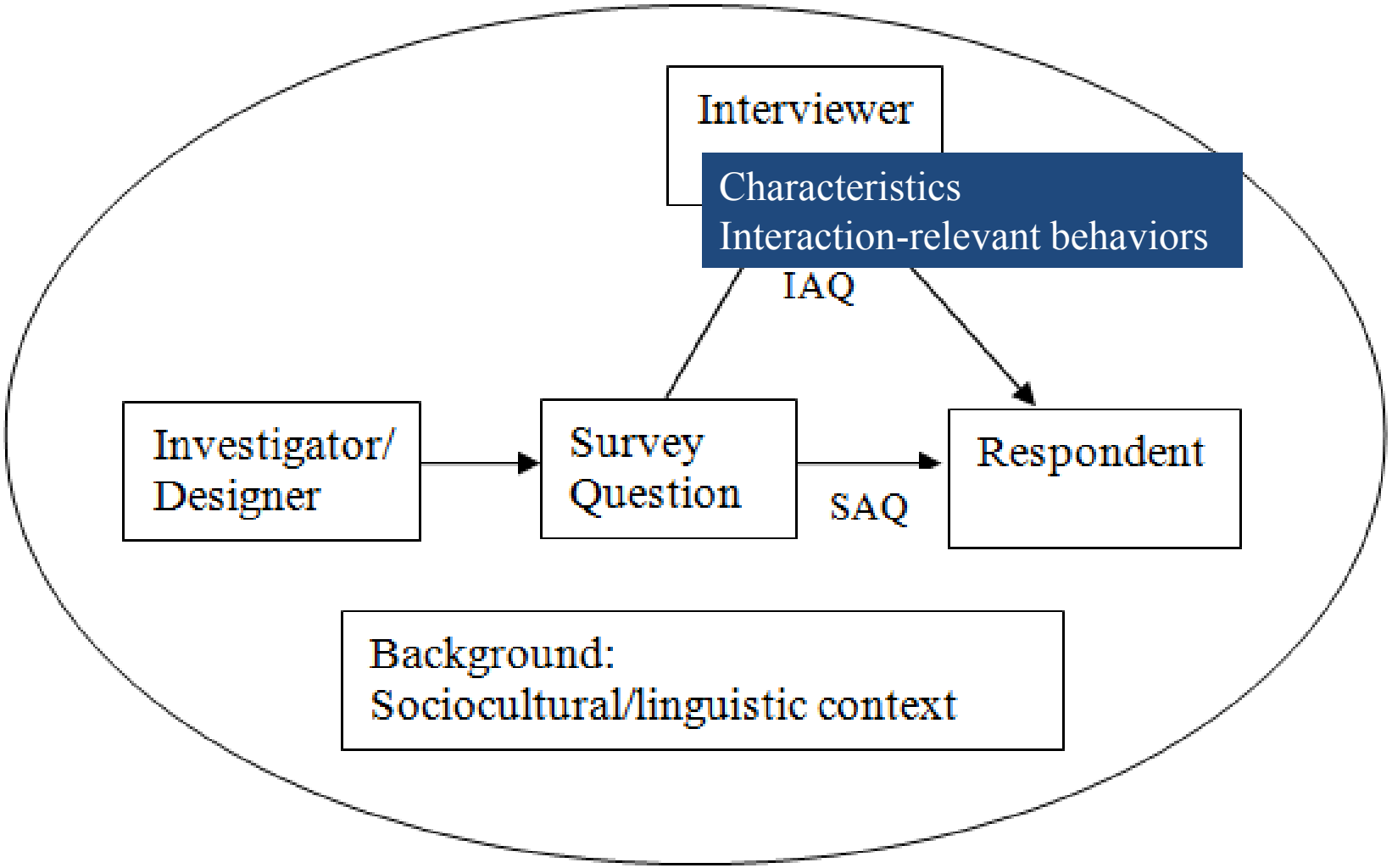


*Inspired by Norm Bradburn, QDET1

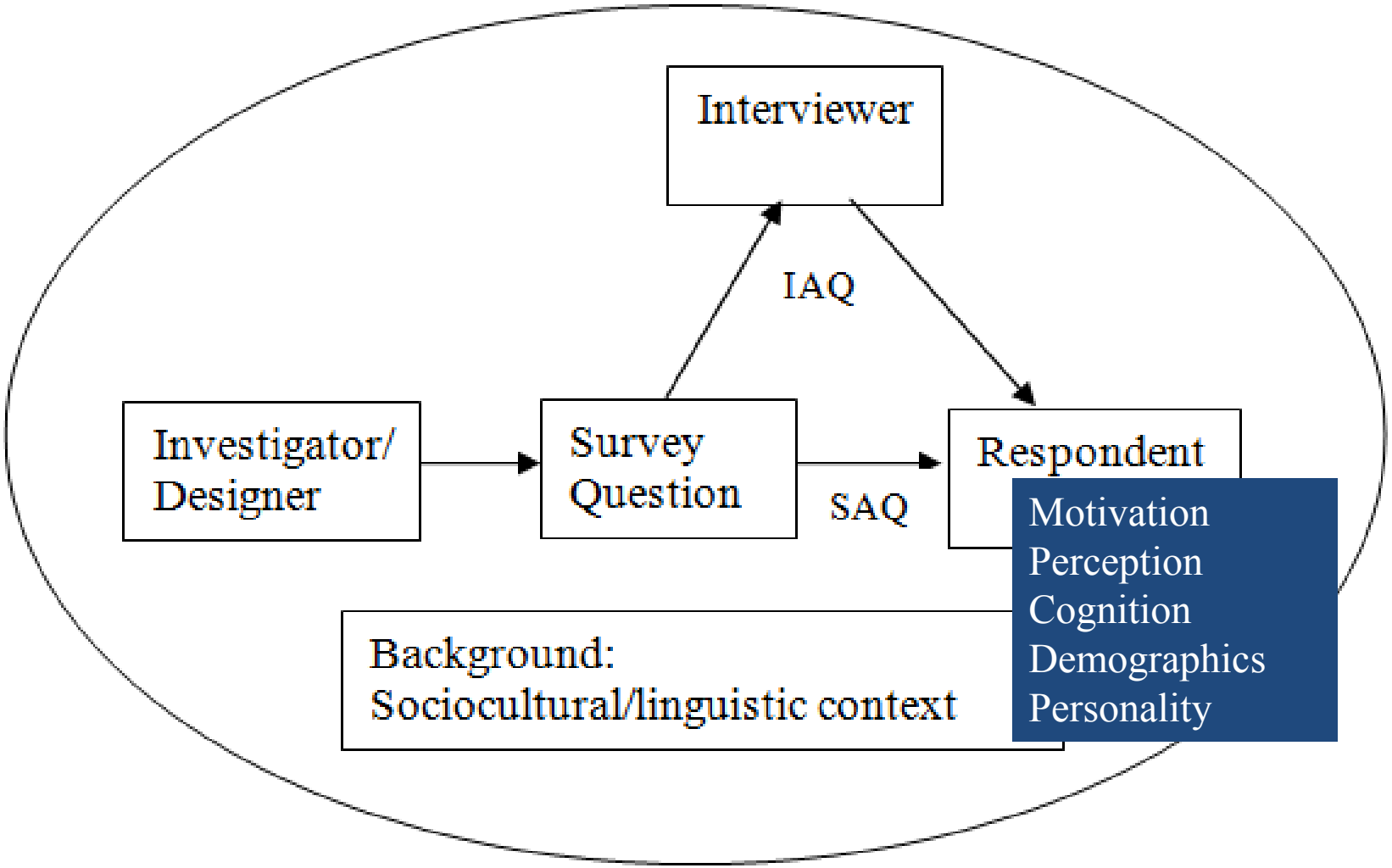
From Lite to Saturated model



From Lite to Saturated model

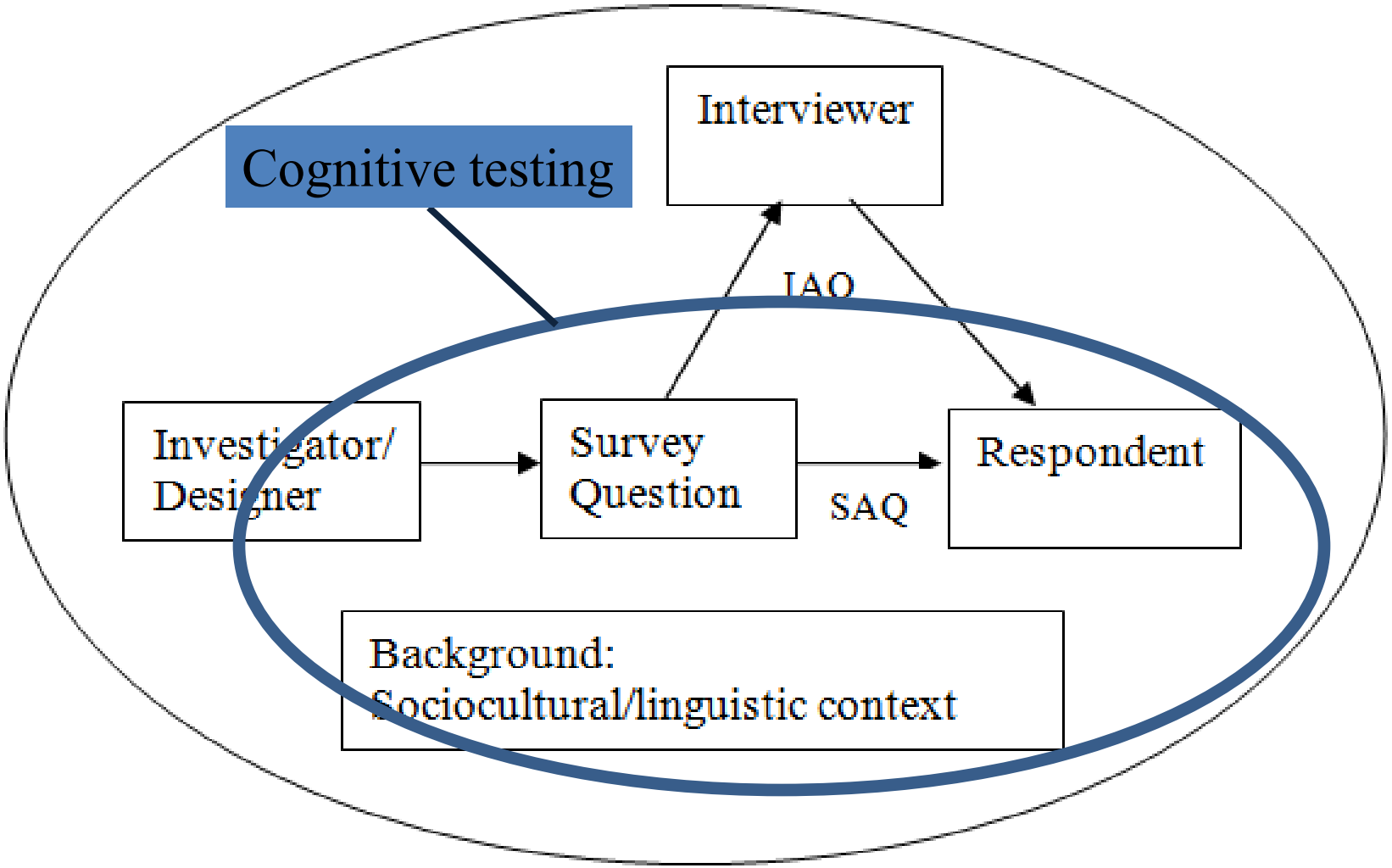


From Lite to Saturated model

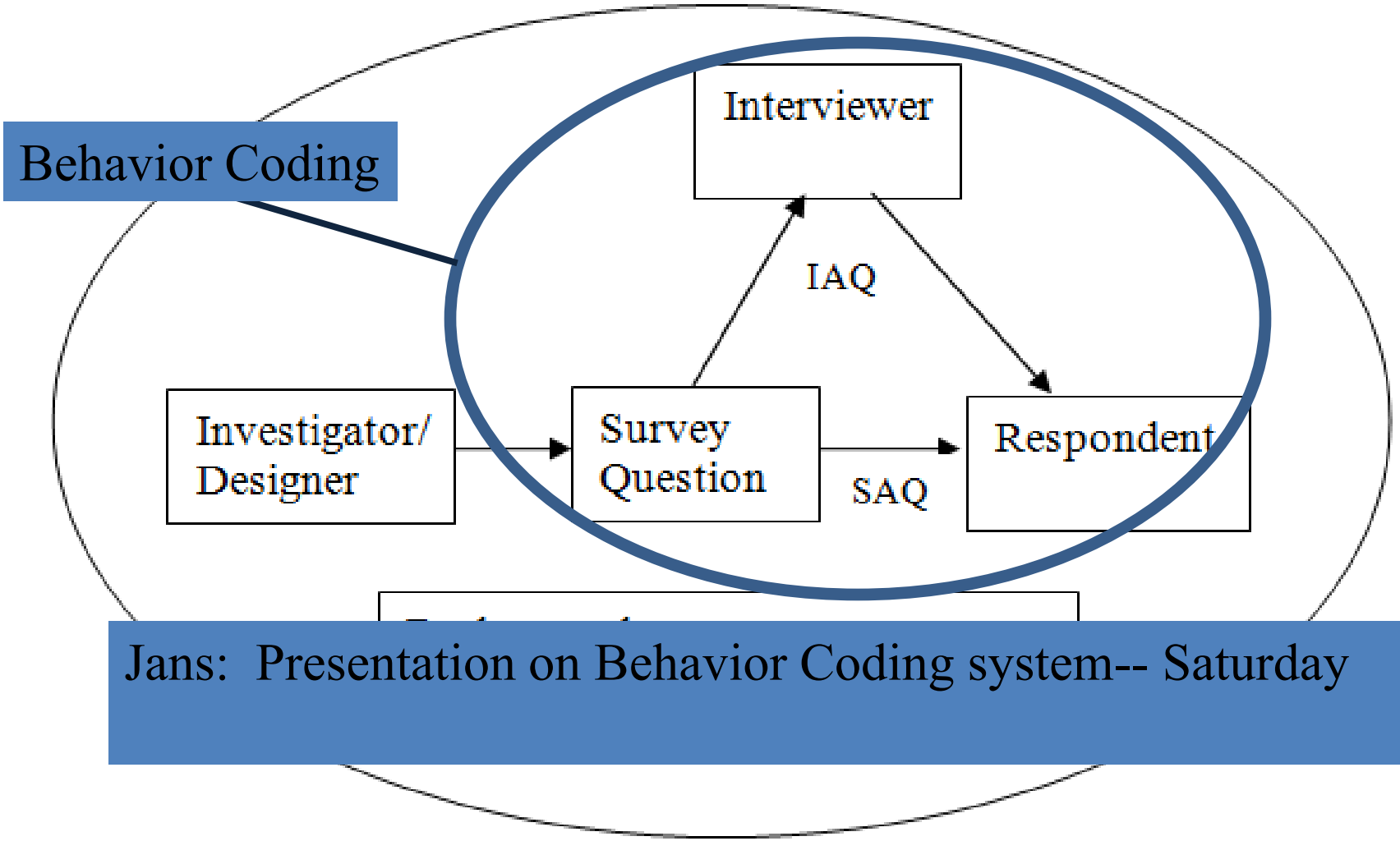


How do we deal with all this complexity?

Different methods target different subsets

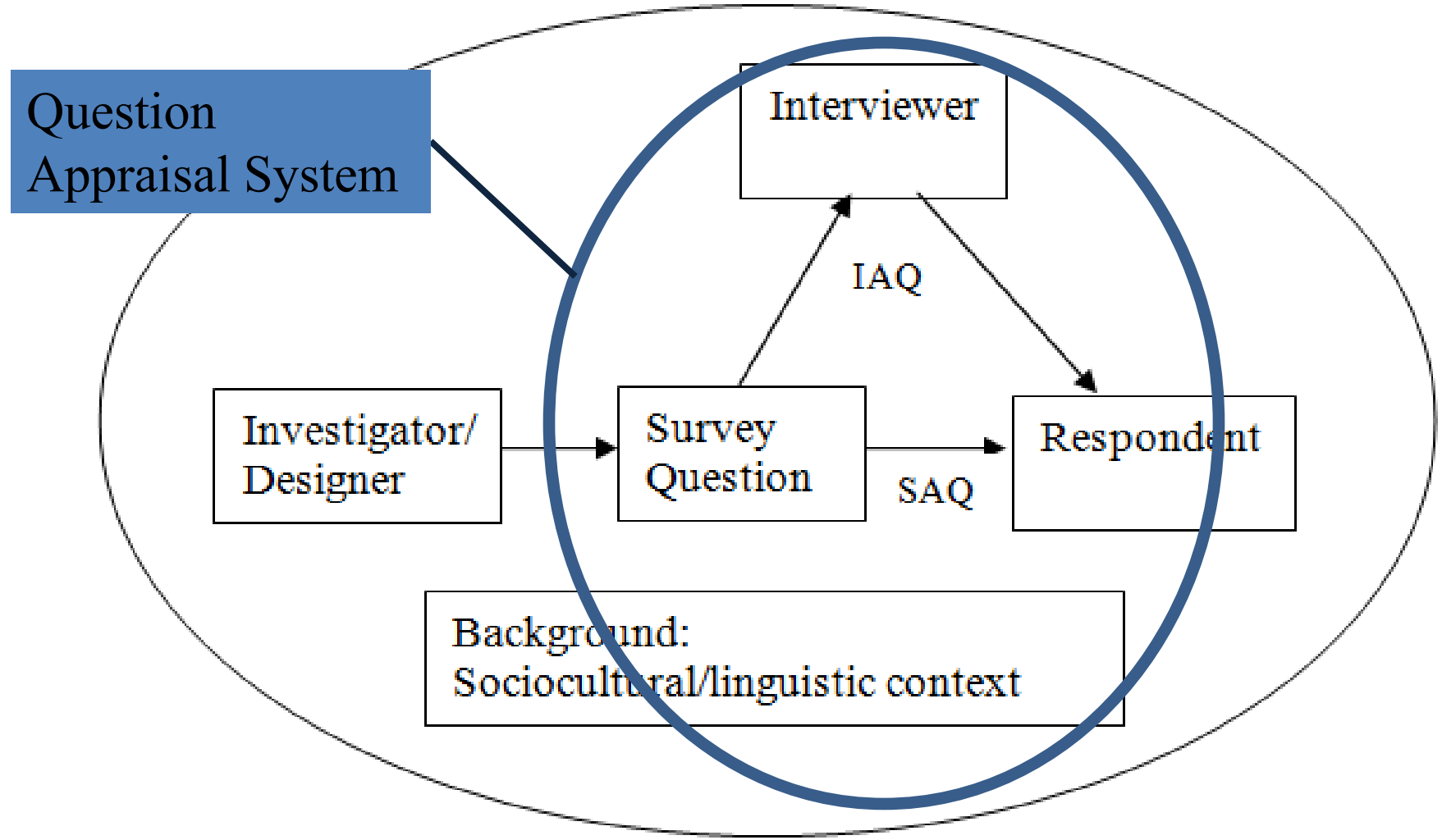


Different methods target different subsets

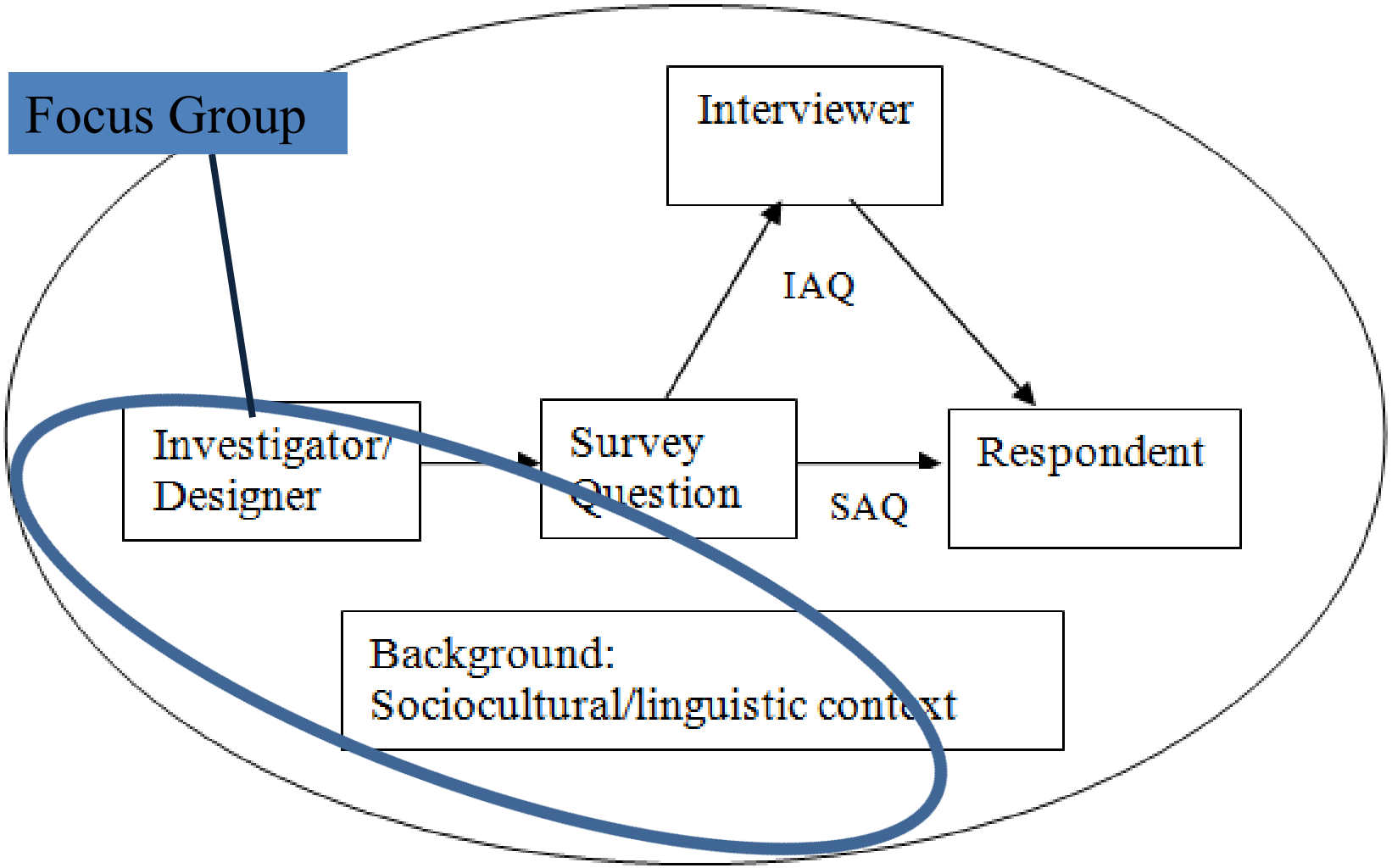


Jans: Presentation on Behavior Coding system-- Saturday

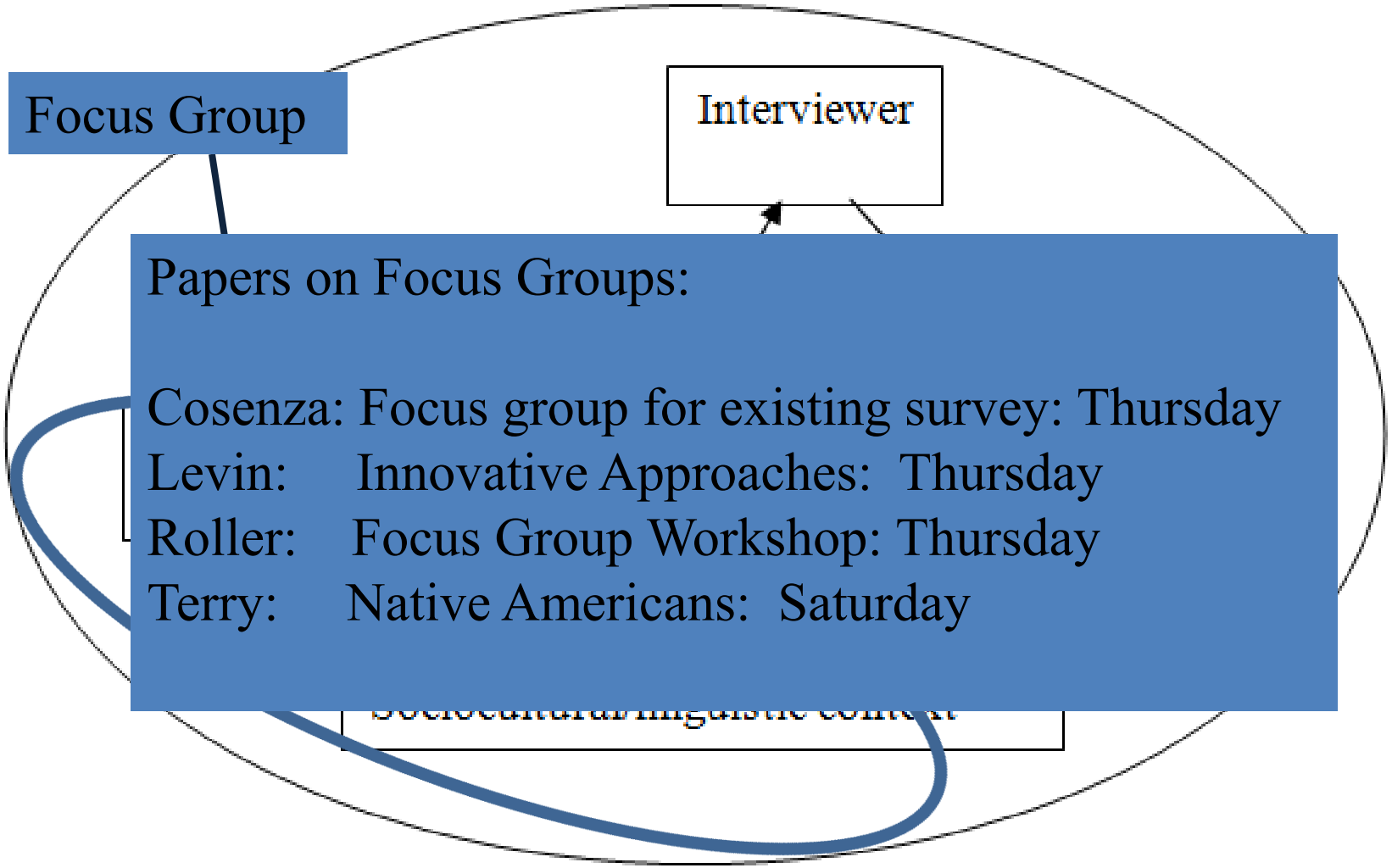
Different methods target different subsets



Different methods target different subsets



Different methods target different subsets



Papers on Focus Groups:

Cosenza: Focus group for existing survey: Thursday
Levin: Innovative Approaches: Thursday
Roller: Focus Group Workshop: Thursday
Terry: Native Americans: Saturday

Different methods target different subsets

Use as many methods as possible (E. Geisen)

Interviewer

Investi
Design

So, (I) Where are we?

- We have a wide variety of tools at our disposal, but not clear how to use them
- I'm not sure how much 'scientific theory' underlies this

Different methods target different subsets

Use as many methods as possible (E. Geisen)

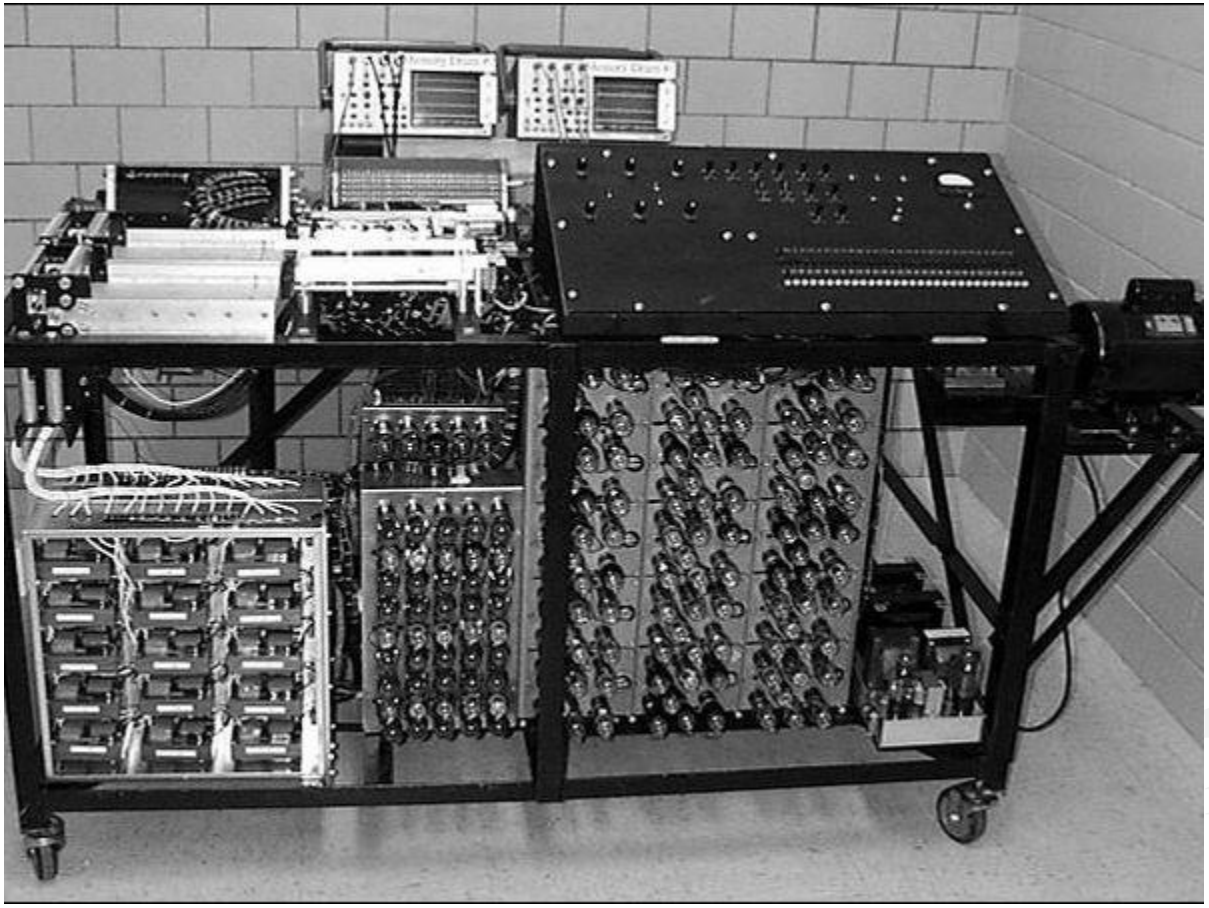
Interviewer

Investi
Design

Sociocultural/linguistic context

- Maitland, Steiger, and Tourangeau: *Framework for Making Decisions about Question Evaluation Methods* – Invited Presentation, Thursday

II. Relevance: Do we still matter?



Does the world care about QDET?

- Sometimes seems that all the action is in non-prob samples, response rates, big data, etc…
- Do we care about the questionnaire?
- 6/6 CI respondents: Yes!
  - Clients, too – K. Levin: "Now more than ever" – especially once they see CI, are converts
  - J. Fowler: Butterfly Effect: Small changes in wording have huge downstream measurement effects

Does the world care about QDET?

- BUT: Are we at the cusp of a change?
- **The world is full of well-meaning people who believe that anyone who can write plain English and has a modicum of common sense can produce a good…**

Key Question: Should we worry about how to ask questions?

If that train has run as far as it can… What else can we do?

1) Return to motivation as an explanation/leverage point – original Cannell, Krosnick (satisficing)
   - Increase attractiveness to Digital-Native-Multi-Taskers? →

Key Question: Should we worry about how to ask questions?

What else can we do?

1) Make things more attractive, motivating…
2) Switch to devices that better reach respondents, and work for their brains… -> Callegaro Keynote, QDET2
3) Go to another data source: ***Q -> Non-Q***

Are we moving away from Q data?

- Argument: We are moving from subjective to objective measures
  - No need for unreliable, unmotivated, hung-over survey respondents
  - Instead, get the real, objective data
  - Example: DNA changes (methylation alterations) could be used to reveal a person's smoking history, to better inform studies that explore risk factors for diseases like heart disease and lung cancer

Are we moving away from Q data?

- But, recent trend/revolution in biomedical/epi field:
  - No longer rely on doctor notes, records, superficial sources
  - Go right to the horse's mouth, and get Patient Reported Outcomes (PROs), which are… self-report

Please respond to each item by marking one box per row.

(QUESTIONNAIRE)

In the past 7 days…

		Not at all	A little bit	Somewhat	Quite a bit	Very much
PAININ1	How difficult was it for you to take in new information because of pain?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PAININ3	How much did pain interfere with your enjoyment of life?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
PAININ5	How much did pain interfere with your ability to participate in leisure activities?.....	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Are we moving away from Q data?

Q!

NON-Q!

Records > Report? NOT a new debate

- Charlie Cannell -> Don Camburn (U MI)
  - Set up survey record-check study with docs
  - "This is great – we'll finally know how good our records are."

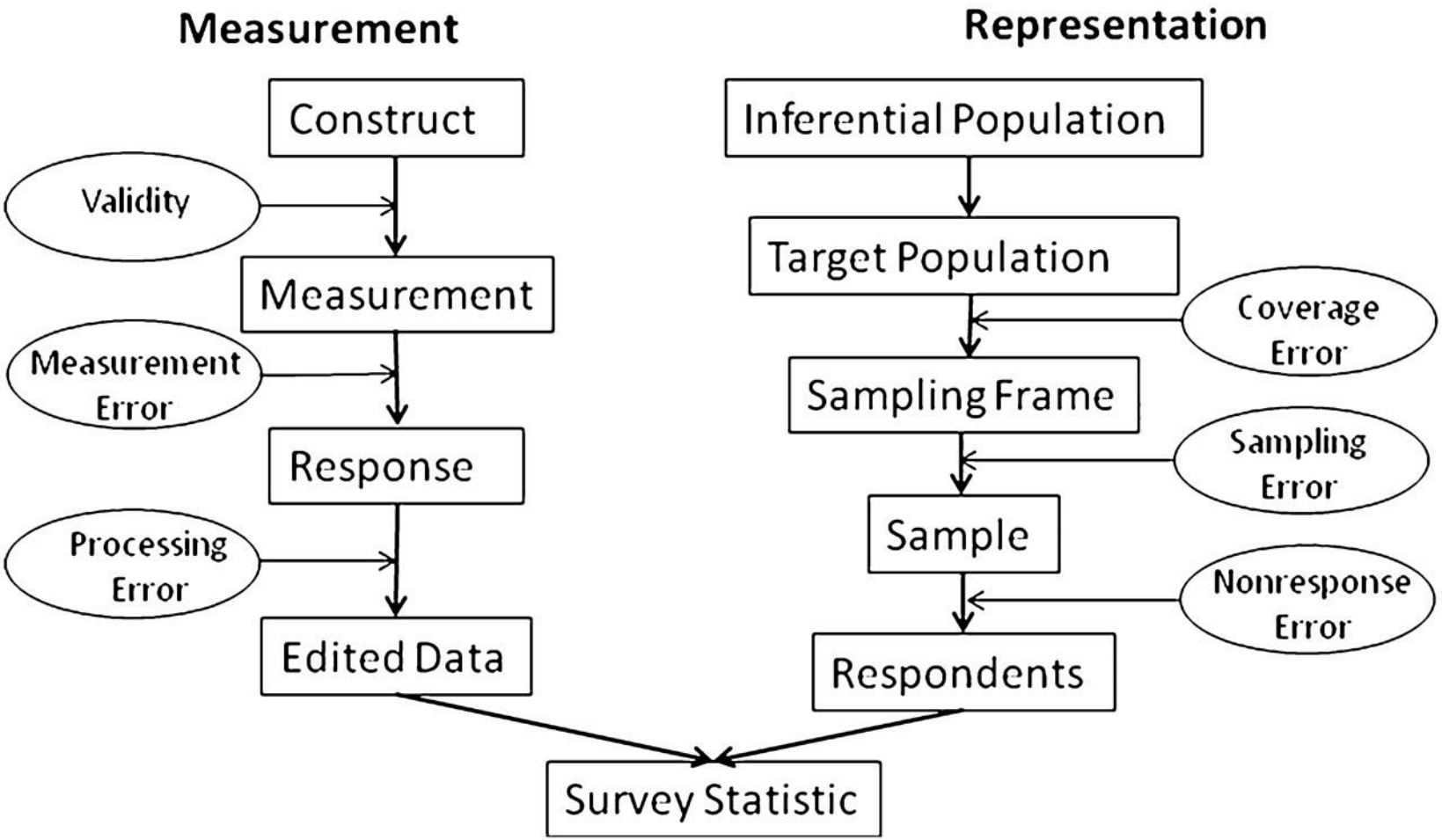
- Willson (NCHS): Cog Interviewing of birth certificates; Jobe, et al. (NCHS); CI of death certificates
  - From birth to death, there is error in the records
- Couper (2013): Self-report Q data are not going anywhere…
  - Challenge is when to use Q versus NON-Q

But: What about emerging non-Q data?

- Social media? Use of Facebook Posts/Tweets?
  - Not an either-or situation – social media may be useful in other ways than 'data collection'
- *e.g., ethnographic study in order to obtain initial input information*
- ***Social Media = Big Data Focus Group?*** *To learn about 'contours and boundaries of the concept under study'*

- ***How should QDET-ers look at Non-Q sources?***

Total Survey Error Components Linked to Steps in the Measurement and Representational Inference Process (Groves et al. 2004).



Robert M. Groves, and Lars Lyberg Public Opin Q 2010;74:849-879

© The Author 2011. Published by Oxford University Press on behalf of the American Association for Public Opinion Research. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Q versus Non-Q sources

Measurement category

Representation category	Q: survey self-report: ask questions	Non-Q: social media, device, records…
Prob Probability sampling	- AAPOR standard model: garden variety probability survey - Probability-based Web panels	Sampling based on demographics, etc.; processing of obtained (social media/record/device) data just as survey data
Non-Prob Not sampled via probability mechanisms	Non-probability surveys -Web panels	Big Data sources - Mega Focus Group? - Pseudo-survey?

Q versus Non-Q sources

Measurement category

Representation category

	Q: survey self-report: ask questions	**Non-Q:** social media, device, records…
Prob Probability sampling	- AAPOR standard model: garden variety probability survey - Probability based Web panels	Sampling based on demographics, etc.; processing of obtained (social media/record/device) data just as survey data
Non-Prob Not sampled via probability mechanisms	Non-probability surveys -Web panels	Big Data sources - Mega Focus Group? - Pseudo-survey?

Q versus Non-Q sources

Measurement category

Representation category

	Q: survey self-report: ask questions	**Non-Q:** social media, device, records…
Prob Probability sampling	- AAPOR standard model: garden variety probability survey - Probability-based Web panels	Sampling based on demographics, etc.; processing of obtained (social media/record/device) data just as survey data
Non-Prob Not sampled via probability mechanisms	Non-probability surveys -Web panels	Big Data sources - Mega Focus Group? - Pseudo-survey?

Too narrow →

Q versus Non-Q sources

(II) Does QDET matter?

Of course…

But… but we may need to start thinking about 'NonQ-DET' as well

mechanisms

III. Where are we going?

How do we reach the Emerald City of Minimal Standards/Best Practices?

Our constant refrain: Was ist eine gute frage? (Faulbaum, Prufer, Rexroth, 2009)

- How do we tell a good frage from a bad frage?

- *Fundamental question for QDET: D versus ET*
  - Huge latent debate: (a) Rules/guidance versus (b) Empirical testing
  - Debate subsumes several unresolved issues:
    - Value of ***Expert Review*** versus ***Cognitive Testing***
    - ***Absolute*** *(static) versus* ***Contextual*** *(dynamic) notion of question quality*

Short of agreeing to disagree… Do we even know that we disagree?

- Presser, Couper, Lessler, E. Martin, J. Martin, Rothgeb, and Singer (2004): *"…pretesting is the only way to evaluate in advance whether a questionnaire causes problems for interviewers or respondents."*
- This is the premise of Cognitive Interviewing
  - Collins (2015); Miller, Willson, and Padilla (2015); Willis (2005; 2015)
  - Kristen Miller (QDET2 cognitive interview):
    - Expert Review can't tell you whether a question is double-barreled; we need to test!
    - Lots of researchers lack appropriate training in ***questionnaire evaluation***

Short of agreeing to disagree… Do we even know that we disagree?

- But, how much is this view universally held?
  - Jen Dykema (QDET2 cognitive interview):

(a) Academic center relies on expert review, design rules, literature, rather than cog interviewing

(b) Practice is likely to be ***sector specific***

(c) Practices depend largely on ***resources/mandates*** rather than on science or theory

(d) "Lots of people have no training *in* ***questionnaire design***" (…as opposed to evaluation)

Dykema: Invited presentation on design based on question characteristics: Friday

Zavala-Rojas: Poster on Survey Quality Predictor (SQP): Thursday

Fundamental point of departure: Absolute versus Contextual notions of validity

- Not a new debate!
- **Design-Rule-Based = *Absolutism*:**
  - Meaning, and function, are inherent within the question
    - If it looks like a duck, it's a duck! – So design the duck right
  - Leads to notions of "Database of good/bad questions"
- **Empirical = *Contextualism:***
  - Oppenheim: Questions may have multiple validities, depending on objectives
  - Interpretive view: Meaning isn't static, but is a function of the social interaction/expectation/context
  - Understanding function in context (usually) requires testing

Toilet Paper Model: Approach Depends on Objectives/Context

How do we resolve this debate?

- Beatty (QDET2 Cognitive Interview): *"Both extremes are really really bad"*
  - Rules alone will never predict outcomes
  - But, there is danger in mindless, unguided empiricism
- The appropriate balance is not clear
  - Bottom line: We should be arguing more about this!
  - We should NOT just go along on parallel tracks as governed by operational constraints and institutional mandates
- My hope is that QDET2 talks won't just branch off into separate
  - Schools
  - Cliques
  - Echo chambers
  - Factions

So, (III) Where would I like to be going?

- Let's better describe what we do (Boeije and Willis: CIRF)
- Let's disseminate what we find (QBANK)
- Let's communicate what we think works (Miller, Dykema, Fowler, Geisen…)
- Let's be serious about criteria for declaring victory and going home!
  - Fowler: Cognitive Testing has been a clear v
  - But: Don't be TOO comfortable… ***Common***
  - *Cog testing is effective… because it improve*
  - *The questions are improved… because Cog t* *e*!

- To evaluate: We need an external measure of error/quality (*so say all my cognitive interview subjects*)

In closing…

1) *Remember that: 'The uncreative mind can spot wrong answers, but it takes a very creative mind to spot wrong questions." - Anthony Jay*
2) *Enjoy QDET2*
3) *See you at QDET3 (2002… 2016… 2030?)*