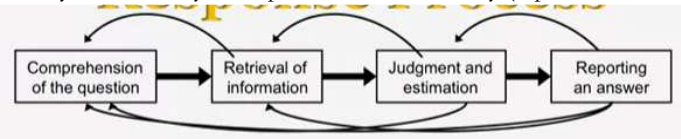
Questionnaire Design for Social Surveys

Week 2 Response Process

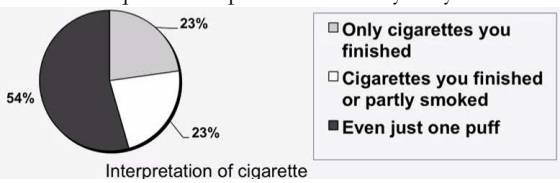
Response Process: Comprehension

Simple Model of the Response Process
• Processes are sequential although R can backtrack
• Models ideal performance, but Rs:
- may misunderstand question
- may not record ("encode") relevant events in memory
- may forget relevant details
- may take shortcuts: satisficing, acquiescence (give acceptable answers but not most fully thought answers for convenience and saving time)
- may intentionally misreport: social desirability (report non-honestly and instead report in social desirable ways)



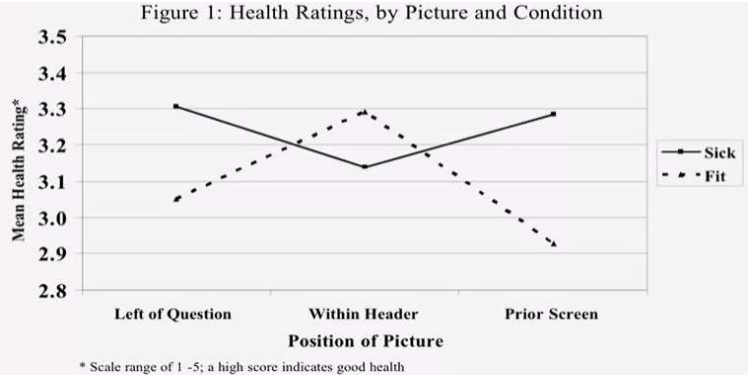
Comprehension Processes
• Most problems understanding survey questions are
- lexical: what do the individual words mean?
- semantic: what does the overall question mean, literally, when those words are assembled?
- pragmatic: what did the author of the question intend for me to understand?

Lexical Processes
• Words have many interpretations, even for sampe person
• Suessbrick, Schober & Conrad (2000) administered CPS Tobacco ; Use supplement followed by interpretation post-test:
- most frequent interpretation held by only 54%



- after interview and interpretation post-test, Rs answer original items again (self administered)
- Half Rs also given a standard definition
- More change when given definitions than not, indicating initial misconceptions
- 10% changed on filter questions possibly leading down wrong path
- Differences in interpretation of individual words affect responses

Lexical Processes: Context
• When interpreting words, Rs are affected by context
- Question context: Info used to answer prior questions:
- if asked about life satisfaction after answering specific question of marital satisfaction, answers correlated 0.67
- but if order reversed (general before specific) r = 0.32 (Schwarz, Strack & Mai, 1991)
• Visual Context: Content of images may be incorporated into interpretation
- Rs rate health lower when image emphasizes fitness rather than illness – except when image in header



Semantic Processes

• Key issue for R is determining how survey concepts correspond – or map – to his/her circumstances

"Have you purchased or had expenses for household furniture?"

"I purchases a floor lamp. Does that count as a household furniture?"

• R knows that words mean but not how they map to his/her situation

• In standardized interviewing, it would not be possible to directly answer the respondent's question

Pragmatic Inferences

• Grice's Cooperative Principle (CP) allows listeners (including Rs) to infer speakers' intended meaning:

- Based on four sub-principles (called maxims):
  - quantity: say as much, but no more, than necessary
  - quality: do not say what you believe to be false
  - relation: be relevant
  - manner: avoid obscurity and ambiguity

• Assuming speaker is following CP, what could she have meant by her utterance

e.g. can you pass the salt – is not about my ability to pass the salt – she already knows that I can; must be request for salt

• speaker would not say something if she did not intend to inform me, so I'll include that information in my interpretation

• Schwarz, Knauper, Hippler, Noelle-Neuman and Clark (1991) asked Rs:

"How successful would you say you have been in life?"

|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

not at all successful — extremely successful

0 — 10

- 34% 0 to 10 group responded in lower half (0 to 5)
- 13% of -5 to +5 group responded in lower half (-5 and 0)
- zero scale anchor combined with label: absence of failure
- negative scale values combined with label: presence of failure

➔ Rs going beyond literal meaning to make sense of labels presented in the question; using numbers to interpret number labels but probably not what questionnaire designer intended

Implications for Questionnaire Design

• Lexical Issues
- try to use terms that most people interpret the same way; will require pre-testing

• Semantic issues
- provide (clickable) definition or train Interviewers to offer definitions as needed

• Pragmatic issues
- try to block Rs' unintended inferences, e.g. avoid gratuitous design features

Response Process: Retrieval

• After R understands the question, and task being asked to perform, will need to retrieve relevant information
- this is true whether Q is factual or subjective
- If factual, generally needs to recall >= 1 event
- If subjective (opinion),
  - needs to recall already-formed opinion
  - OR if R does not currently hold opinion, recall opinions on similar topics or relevant considerations

Different Kinds of Memory
• Working Memory (WM) v.s. Long Term Memory (LTM)
- WM: temporary (short term) storage of
  - words of a sentence (question) as they arrive
  - partial results e.g. when counting up recalled episodes
  - spoken response options
- LTM is more permanent (through forgetting occurs)
• Types of LTM
- Episodic: events in one's life e.g. I ate Ethiopian food on Saturday
- Semantic: knowledge about the world and oneself e.g. Trees have leaves, I meet with my reading club every Tuesday
- Procedural: doing things e.g. riding a bike

Events v.s. Self-knowledge
• Menon (1993) asked Rs to judge the frequency of mundane activities e.g. snacking, washing one's hair etc.
• Activities occurred on regular or irregular schedule and were similar or dissimilar
• Predicted most use episodic recall and least use of rates when events distinctive (dissimilar) and irregular; least use of episodic recall and most use of rates when events similar and regular

	Similar	Dissimilar
Regular	89.5	62.5
Irregular	44.7	24.1

% Responses based on Rates
(knowledge of one's recurring patterns of behavior)

For similar & regular events, more likely to be using "rate" information but for irregular & dissimilar events, more likely to be using "episodic recall".

Forgetting from Episodic Memory
• Interference
- the longer the time period in question (e.g. 1 year v.s. 1 month) the more likely other similar events will have occurred
- hard to distinguish details of one evens from others
- tend to blend into single generic memory
• Decay
- the more time that has passed since events occurred, the weaker the memory
- forgetting most rapid in period immediately after event experienced
- forgetting continues after as many as 50 years(!)

Retrieval Cues
• Cues can remind people of events:
- NCVS item on shopping lists: "drug, clothing, grocery, hardware and convenience stores."
- open question whether Rs search only for matches to cues or use cues to define broader conceptual space
• Cues can affect what events come to mind:
- Couper (2004) asked Qs like "How many sporting events have you attended in the past year?" and presented images with Qs
- Images were either low frequency (major league ball park) or high frequency (little league game)
- low frequency images led to lower reports than high frequency images

Encoding
• Failure to encode
- Lee ('99) intercepted parents leaving "injection clinic"
- Had poor recall for immunizations children had just received
- recall no worse over time, and not improved when Rs provided calendar or were asked if recognized immunizations
• Encoding specificity (Tulving & Thompson 1973)
- Memory better when context of encoding and retrieval match than when differ
- Godden & Baddley (1975) asked divers to study words on lando r 20 feet under
- tested them in either same or different context
- divers showed vastly superior memory when recalled in same context as studied
- Context reinstatement (e.g. Fisher & Quigley, 1992)

Response Process: Judgement

Estimation and Judgement
• Compensating for imprecise or incomplete memory: two examples
- availability heuristic
People can infer frequency or probability on basis of retrieval difficulty (Tverksy and Kahneman)
I can't recall many instances of the event so it must be rare
- frequency estimation when recall is not possible or involves more effort that Rs willing to invest

Availability Heuristic
• Tversky & Kahneman 1974
- people sometimes use the ease with which examples come to mind as an indication of their frequency or probability of occurrence
- easier to retrieve words starting with r than words whose 3rd letter is r; former judged more frequent even though latter actually is
- easier to recall famous than less famous names; when list had more famous females than males, former judged more frequent even though equal numbers of males and females

Availability in Survey Context
• Schwarz 1991
- asked Rs to list either 6 or 12 situations in which were assertive
- then asked to rate own assertiveness
- Rs who listed 6, rated themselves more assertive than Rs who listed 12 instances
- Presumably, easier to bring few instances to mind so Rs conclude more likely

Frequency Estimation
• In the past 2 years, how many times did you donate blood?
• For someone who donates every 8 weeks in the same workplace location, there is little to distinguish one donation from the next so it's hard to recall and count up each one
• To supplement memory, Rs may estimate on basis of
- rate
"I donate every time there is a blood drive which seems to be about every 8 weeks which is two so I will say 12 times in the last two years"
- qualitative impression and convert to a number
- "I do this a lot so I'll say 10 times in the last year"

Estimation Problems: Behavioral Frequency Questions
• At least three broad strategies, each lending to different type of error
- recall and count: underestimation
- more likely to forget than invent an event
- rate based estimation: overestimation
- for regularly occurring events, rates can be quite accurate except when Rs fail to take exceptions into account (e.g. not doing the behavior)
- impression based estimation: overestimation

- translation of impression to number cannot be any lower than 0 but is unbounded on the high end

Strategies for Estimating Event Frequency
• Conrad, Brown & Cashman (1998) asked telephone Rs to answer frequency questions
• For each, Rs also
  - reported how they arrived at their answer; later coded for evidence of strategy
  - rated the regularity of occurrence
  - rated the similarity of occurrence

Response Strategy	%	Regularity	Similarity	Mean Reported Frequency	RT Slope
1.Enumeration	27	2.39	2.86	2.3	.84
2.Rate Retrieval	15	3.66	3.54	7.8	.02
3.Rate Estimation	12	3.43	3.31	11.1	.15
4.Rate and Adjust	9	3.65	3.43	11.9	.07
5.General	18	2.91	3.11	12.3	.01
6.Uncodable	18				

Conrad, Brown & Cashman (1998)

➔ Enumeration (episodic recall): reported frequency quite low… so you can actually recall specific episodes (also regularity and similarity are lower than other strategies which matches with Menon(1993)'s study earlier)
➔ Rate Retrieval: high reported frequency… so you can't really recall every single episode…so you instead rely on "rates" with high regularity and similarity (matches with Menon(1993)'s study results earlier)
➔ General: highest mean reported frequency… so frequent that solely relying on either episode or rates seems infeasible… so regularity and similarity scores are kind of in the middle level

Estimation and Aging
➔Ziniel (2008) conducted two studies to see how reduced memory ability in aging affects selection of estimation strategy
**Study 1
• reanalyzed Conrad, brown, and Cashman (1998) based on respondent age and observed less enumeration among older Rs
• conducted lab experiment with young (20-30 yrs) and older (>= 65 yrs)

	Young	Old
Episodic Enumeration	36.05%	26.32%
Rate retrieval	17.53%	21.80%
Rate estimation	14.32%	13.53%
Rate and adjustment	12.10%	8.27%
General impression	20.00%	30.08%
	100%	100%
N	405	133

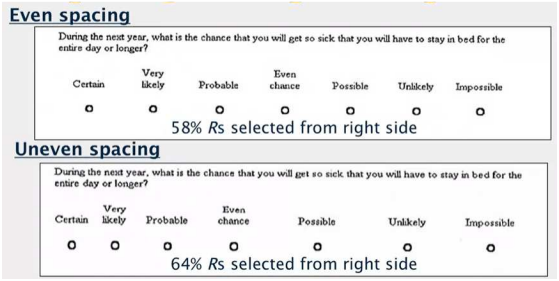
Study 1: Reanalysis of Conrad, Brown & Cashman (1998) on basis of *R* age

**Study 2
• compared likelihood of selecting "cognitively less demanding" strategies in young v.s. older Rs
  - Rs indicate strategy by picking from menu (card)
• For older Rs, odds of choosing Rate Estimation or Rate-and-Adjust greater than Enumeration; for young Rs opposite pattern
• For older Rs, odds of choose general impression greater than enumeration; for younger Rs opposite pattern
• Same pattern if Rs divided on basis cognitive ability score
  - further evidence that Rs consider difficult of task and their own mental abilities

Response Process: Response

Mapping and reporting
• Assigning a judgement into an answer category
- result of judgement and estimation stage
May not be expressed precisely
May not match a response option (if close question)
- requires judgement to be transformed into
- a number (if a vague quantity)
- one (ore more) of the options provided
• Transformation can introduce error

Spacing of Response Options
• Conceptual v.s. visual midpoint
- Rs' mental representation of scale when presented visually involves underlying concept and visual appearance
• Tourangeau, Couper & Conrad (2004) compared endorsement of middle option when scale spaced evenly v.s. unevenly
- uneven spacing made scale points on right side (conceptually) look more central



Primacy Effects with Visual Presentation
• Krosnick & Alwin (1987) embedded experiment in General Social Survey
- Rs select qualities of children most desirable for their children
- presented visually, on show cards, in FTF interviews
- For 1/3 of Rs, order reversed

1.... has good manners (MANNERS)
2.... tries hard to succeed (SUCCESS)
3.... is honest (HONEST)
4.... is neat and clean (CLEAN)
5.... has good sense and sound judgment (JUDGMENT)
6.... has self-control (CONTROL)
7.... he acts like a boy or she acts like a girl (ROLE)
8.... gets along well with other children (AMICABLE)
9.... obeys his parents well (OBEY)
10... is responsible (RESPONSIBLE)
11... is considerate of others (CONSIDERATE)
12... is interested in how and why things happen (INTERESTED)
13... is a good student (STUDIOUS)

➔ When options appeared early, selected substantially more often than when late – primacy effect

Quality	Standard order	Reversed order	Difference
	%	%	%
Manners	26.4	10.1	16.3
Success	19.1	14.6	4.5
Honest	65.7	48.4	17.3
.	.	.	.
.	.	.	.
.	.	.	.
Considerate	24.9	39.5	-14.6
Interested	17.9	24.9	-7.0
Studious	6.5	16.4	-9.9

➔ Krosnick & Alwin ('87) call this "survey satisficing" i.e. tendency to select first item that is good enough

Recency Effects (forgetting about earlier option) with Auditory Presentation
• Schwarz (1992) report preference for final item in short lists presented orally

	Response Order	
	First	Second
Form of Government		
Authoritarian	15%	26%
Democratic	58%	67%

• Age might impact recency effects due to deteriorating memory capacity as people age
➔ Knauper (1998, 1999) controlled for working memory capacity using respondent age as proxy
- reanalyzed shuman & presser ('81) study on housing
- older respondents showed larger recency effect
- younger respondents rarely produced order effect

Rounding
• Questions requiring (open) numerical responses may exhibit distortion in mapping stage
• Visible example is prototypical or rounded reports
- may indicate imprecision in underlying representation
- may simplify mapping task by creating "categories"
- may signal uncertainty
- may signal embarrassment
• Example
➔ Evaluation of Candidates on 100 point feeling thermometer: Huge numbers evaluating with numbers that end with 0 (multiples of 10) so this indicates rounding happened a lot

	Number Responses	% Total
Clinton		
Multiples of 10	1907	78.9
15, 85	310	12.8
All other values	199	8.2
Bush		
Multiples of 10	1948	79.3
15, 85	412	16.8
All other values	98	4.0

American National Election Survey, 1992
source: Tourangeau, Rips & Rasinski, 2000

Sensitive Questions and Self Administration
• Relative to I-administration, self-administration increases Rs willingness to report socially undesirable (including illegal) behavior
• Tourangeau & Smith (1996) presented sensitive questions, e.g. drug use, under I- and self-administration

	Time Frame		
	Past Month	Past Year	Lifetime
Cocaine	1.74	2.84	1.81
Marijuana	1.66	1.61	1.48

Ratios of Estimated Prevalence under Self-(ACASI) and Interviewer-(CAPI) administration
➔ Editing Response
• Motivating misreporting (Tourangeau & Yan 2008)
- evidence suggests that socially desirable reporting is the result of deliberate misreporting
- misreporting in one direction
- self administration affects answers to sensitive but not non-sensitive questions
- most likely result of editing a formulated response rather than selectively retrieving positive attributes about self
- response time longer for sensitive than equally demanding non-sensitive questions
- holtgraves (2004) observed longer RTs when introduction emphasized social desirability concerns
- suggests edit step even if Rs don't change their answer