

Classifying Open Occupation Descriptions in the Current Population Survey

Fred Conrad

Mick Couper

University of Michigan

Classification and measurement

- Open ended responses must be classified (coded) in order to be quantified
- Coding is subject to human error because
 - It is different from everyday classification
 - Instances belong to everyday categories to varying degrees
 - people can grow new categories to handle atypical cases
 - Coders can't do either of these
 - It is similar to everyday classification
 - coders, like anyone, may be blinded by superficial similarity between descriptions and categories and fail to consider category size (base rate)

CPS occupation questions

What kind of work do you do, that is, what is your occupation?

What are your usual activities or duties at this job?

- Both responses are considered together in assigning a (numeric) occupation code
 - US Census occupational classification system
 - 500 categories (3 digits); 46 detailed groups (2 digits); 14 major groups (1 digit)

Three part study of CPS occupation coding

1. What characteristics of occupation descriptions hurt coding reliability?
 - double-coded descriptions (n=32,362)
2. How do characteristics jointly affect reliability?
 - double-coded experimental descriptions (n=800)
3. What do coders think about while coding?
 - verbal reports while classifying experimental descriptions (n=100)

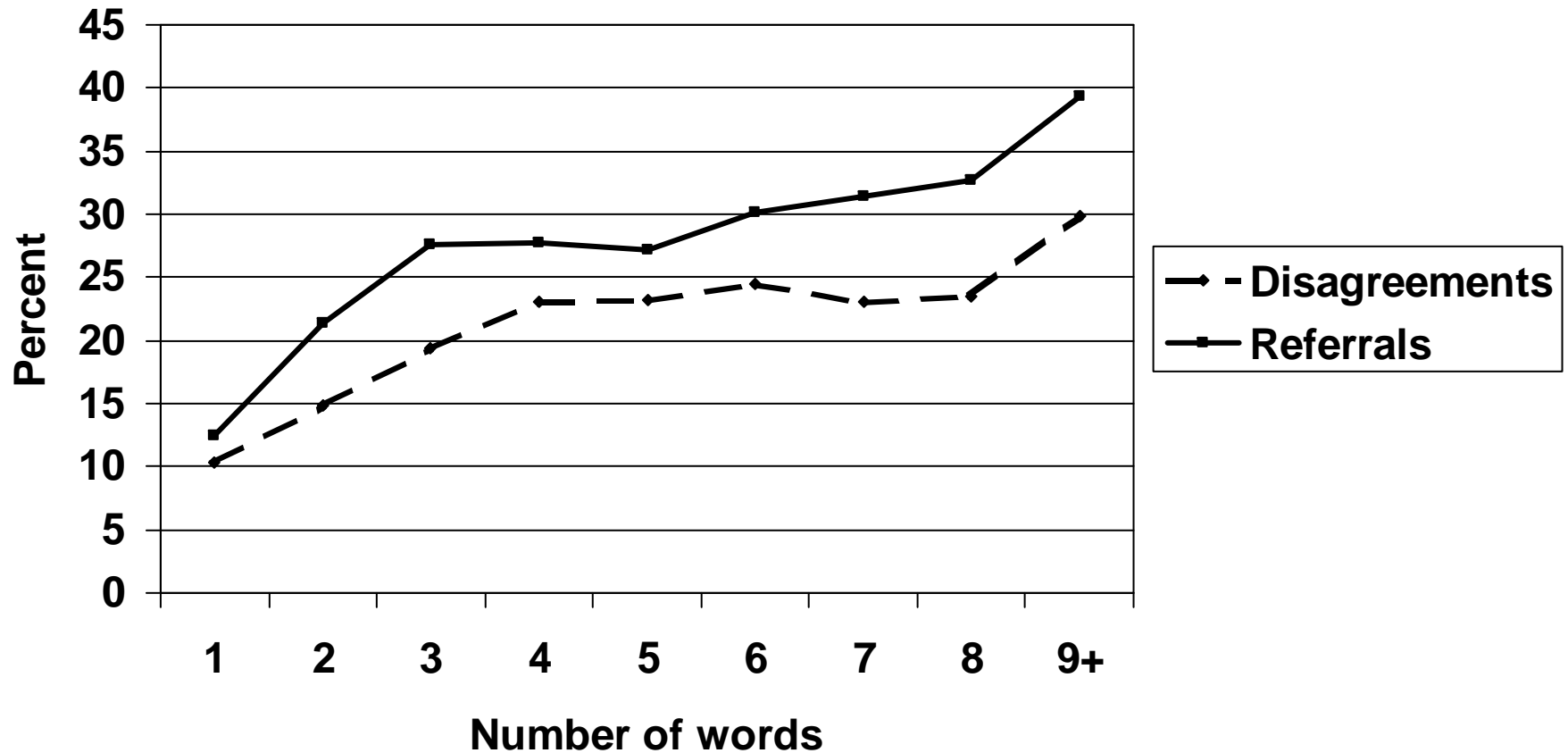
Characteristics of descriptions

- Quality Assurance (QA) data:
 - 10% descriptions in CPS routinely classified by 2 independent coders for QA purposes
 - current analyses for QA data from 3/97-2/98
 - n=32,362
- Coders disagreed on 15% (n≈4800) of the occupation descriptions
 - many coded as very different occupations
 - 78% disagreements at first two digits
 - agreement does not guarantee accuracy, but disagreement guarantees someone is wrong

Characteristics (cont'd)

- Not due to data entry errors
 - simple transposition errors (e.g. “234” vs. “243”) account for less than 1% of disagreements
 - one digit discrepancies (e.g. “123” vs. “223”) account for almost 7% of disagreements but many of these may be intended

Length of description and disagreement/referral rates



Length of description

- Disagreements involve longer descriptions
 - when disagree, description length: 2.10 words
 - when agree, descriptions length: 2.56 words
- Couper & Conrad ('96) found lower agreement if, in addition to basic occupation question Rs also asked for job title
- Cost of increased length surprised us because extra info might resolve ambiguity
- Suggests probing for more detail may be counter-productive

Length of description (cont'd)

- Esposito & Canter ('92) asked coders if interviewers should have probed for more information
 - Coders expressed a desire for more information only when a single code could not be chosen
 - No desire for additional information if a decision was possible on basis current information
 - It seems these coders recognized the potential costs of more detailed descriptions

Difficulty of particular words

- Certain words in description involved in more disagreement (high ratio disagree/agree):
 - “administrative,” “services,” “research,” “assist,”
“maintenance,” “administrator,” “general,” “service,” “helper,”
“equipment”
- Other words in description involved in less disagreement (high ratio agree/disagree):
 - “waitress,” “registered,” “guard,” “electrician,” “secretary,”
“accountant”

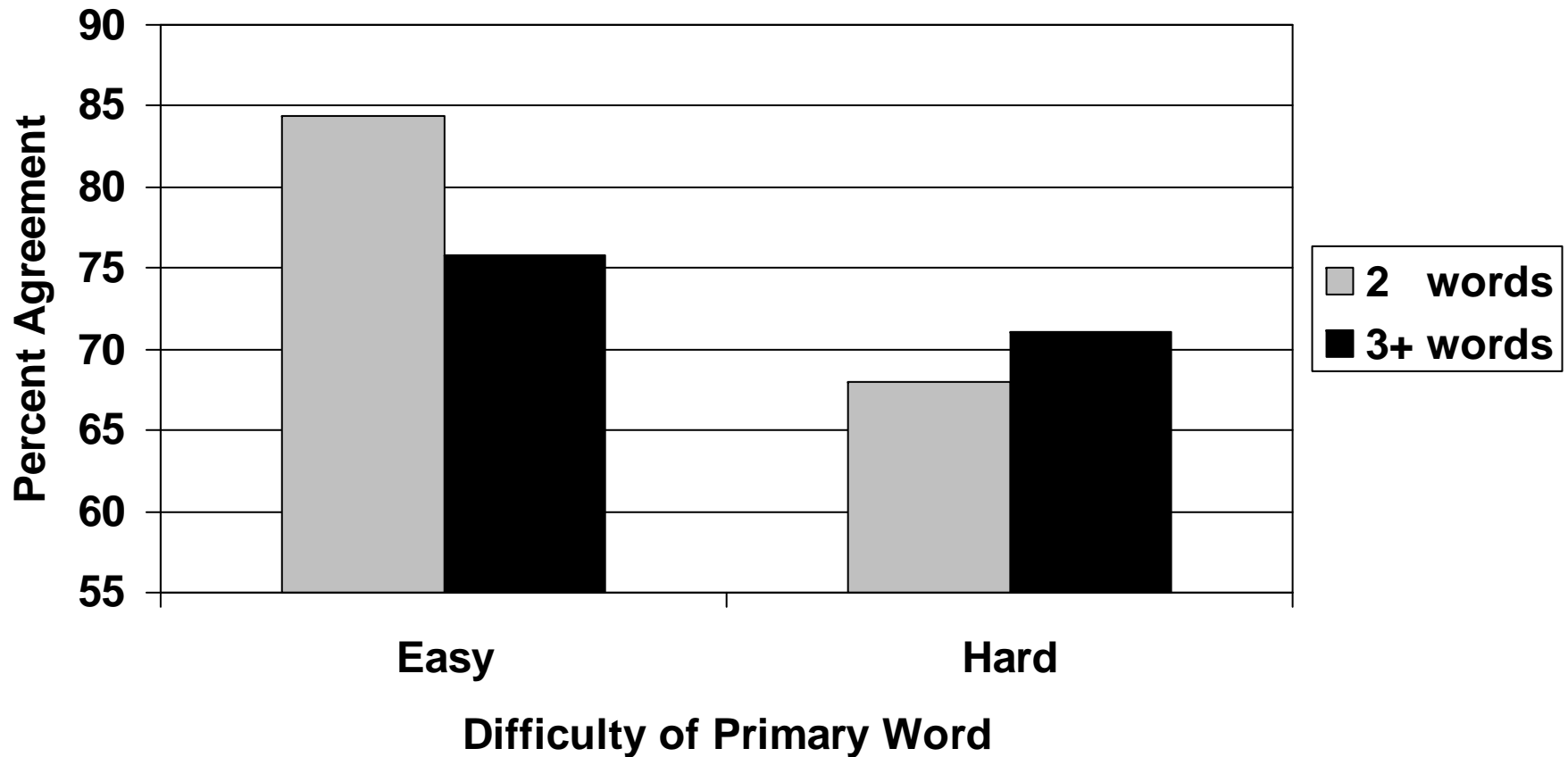
Difficulty of particular words (cont'd)

- How are difficulty and number of words related?
 - Cannot address systematically in QA data
 - corpus does not contain enough comparable cases
 - need long and short descriptions involving same easy and difficult words
 - best addressed through experiment

Joint effects difficulty and length

- Descriptions (n=800) systematically varied in:
 - length (1, 2 and 3+ words)
 - difficulty of “primary” words (easy, hard)
 - hard=highest QA ratio disagreement to agreement:
owner, operator, laborer, director, technician, clerk, supervisor,
administrator
 - easy=highest QA ratio agreement to disagreement:
secretary, cashier, driver, cook, teacher, nurse, waitress,
carpenter
- selected from QA corpus or crafted to match

Increased length only hurts when words are easy



Increased length only hurts when words are easy (cont'd)

- Extra words may be confusing when primary word is easy
- But may clarify description when primary word is hard

Coder observations

- 4 coders provided verbal reports while coding
 - 100 descriptions (50 per pair of coders)
 - multiple words
 - led to high disagreement in experiment

Special purpose rules

- All 4 coders reported using special purpose rules
 - Concerned superficial aspects of descriptions
 - Not definitions of coding categories
 - No coders could produce the rules in writing
- Should improve inter-coder reliability but no reason should improve accuracy
 - Martin, Bushnell, Campanelli and Thomas, '95 observed more “correlated coder variance” among “office” than field coders
 - Hak & Berndt ('96) observed coders formulating rules:
 - “I think that ... if the word ‘being able to’ occurs in the answer we can almost always choose VALUE if it is possible uh .. to read the answer in that way”

More than one occupation

EMP : **DOMINOS PIZZA**
IND : **PIZZA**
OCC : **COOK, DRIVER**
DUTIES : **DELIVERY, COOKING**

“When two different occupations are described, match to duties and go with the first duty listed.”

More than one occupation (cont'd)

EMP : NEWBERRYS
IND : VARIETY STORE
OCC : CASHIER STOCKING
DUTIES : STOCKING AND CASHIER

“When two different occupations are described, match to duties and go with the first duty listed.”

More than one occupation (cont'd)

- Rule could introduce coding error by giving priority to first duty if in fact job is combination or relies more on second duty
 - to make the *I*'s “double request” *cooperative* (in Grice's sense) *R* may reverse order of duties irrespective of which best describes her job

Directly code certain occupations

EMP : ROSS AUTO BODY SHOP
IND : AUTOMOBILE REPAIR BUSINESS
OCC : SECRETARY, CUSTOMER SERVICE ADVISOR
DUTIES : BILLING CUSTOMERS, SCHEDULING
SERVICE, ADVISING

“If secretary appears on the occupation line, don’t look at anything else; code it as secretary [313].”

Drop the modifier

EMP	:	MUSCOGEE COUNTY
IND	:	SCHOOL DISTRICT
IND_TYPE	:	SOMETHING ELSE
OCC	:	ASSISTANT TEACHER
DUTIES	:	ASSISTING TEACHER SPECIAL ED

“If ‘assistant’ is in front of the [occupation] word, ignore it. If ‘assistant’ comes after the word, assistant is the key word.”

- In this case, the rule leads to one code (158) for teacher, special education

Drop the modifier (cont'd)

EMP : MILL RIVER UNION HIGH SCHOOL

IND : HIGH SCHOOL

OCC : TEACHER ASSISTANT

DUTIES : WORKING WITH CHILDREN WITH
SPECIAL NEEDS AND

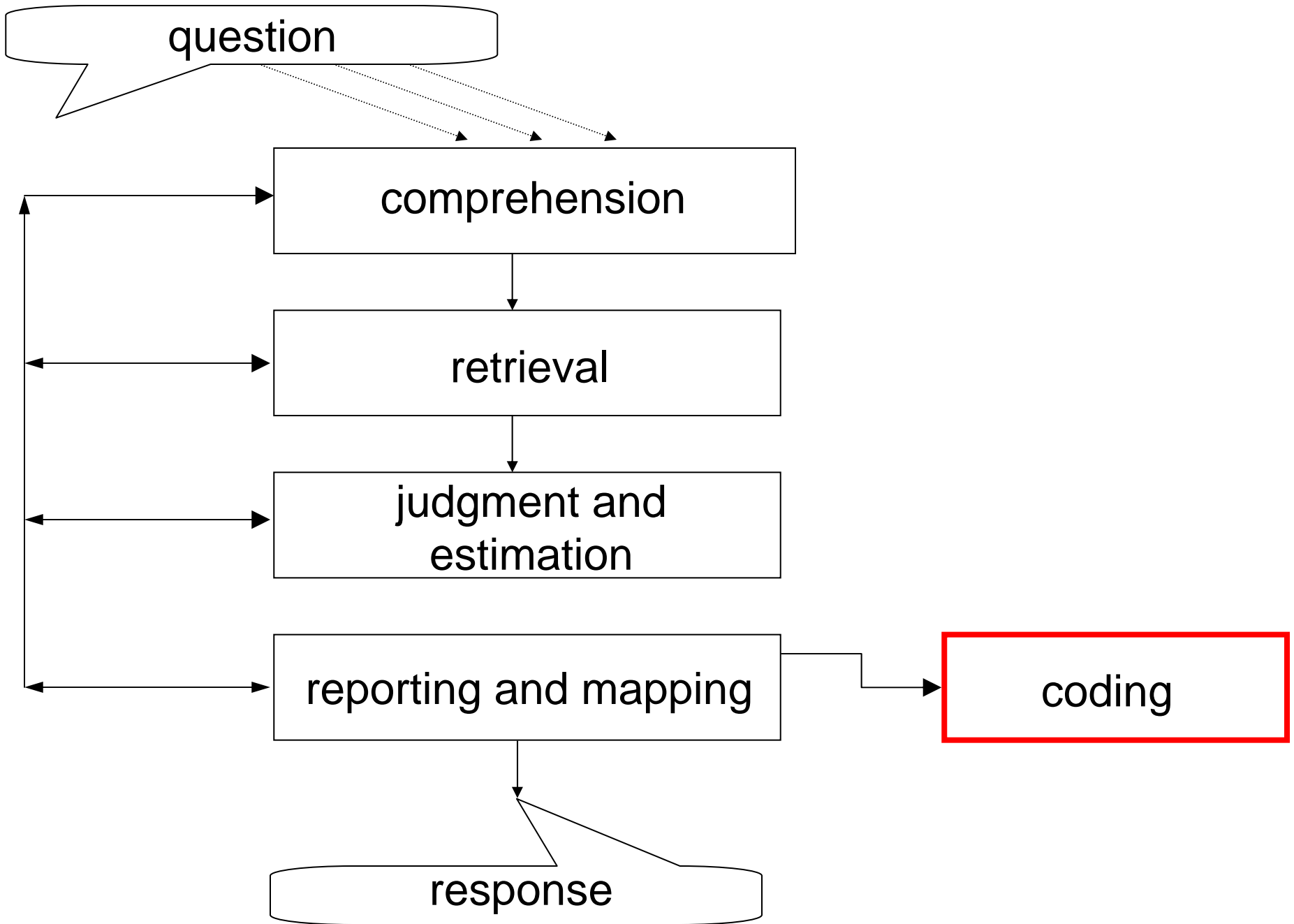
- but in this case, the rule leads to a different code for teacher's aid (387)

Summary

- Characteristics of open responses affect data quality, in particular, intercoder agreement
 - longer descriptions reduce agreement
 - although may depend on particular words
- Coders develop special purpose rules for classifying potentially ambiguous cases
 - may improve agreement (reliability) but could harm accuracy (validity)
 - if not applied consistently, will harm agreement

Conclusions and Implications

- More accurate, not just more reliable, coding should be priority
 - Coding logic should be conveyed to interviewers
 - through training and automated support in interview
 - Coders can be trained to code terms known to be difficult, in theoretically sound way
 - Special rules can be documented and re-written on basis category definitions
- If accurate coding is a priority, reliable coding will follow



Main Points

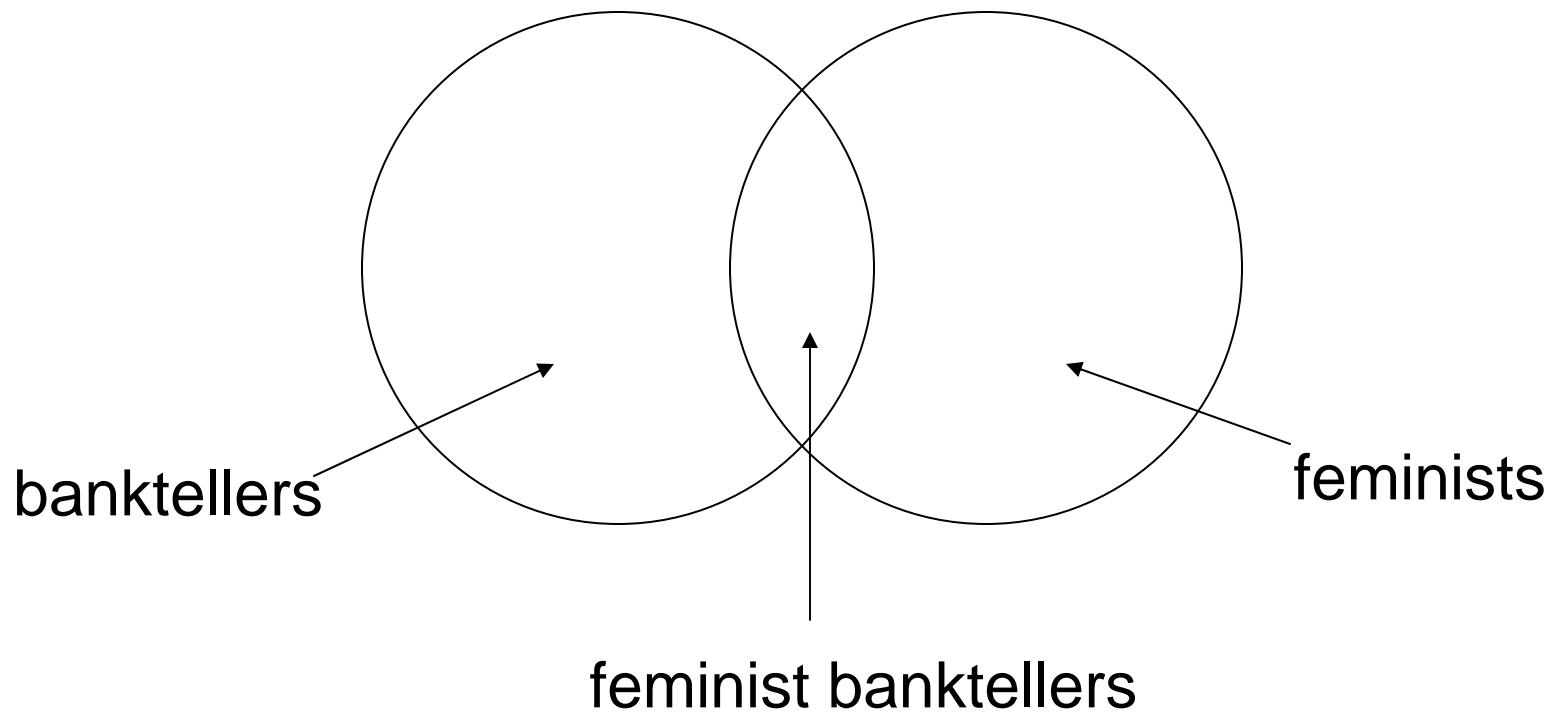
- Characteristics of respondents' descriptions affect reliability of coding
 - Number of words
 - Concreteness of words
- Coders develop special purpose (superficial) rules to maximize reliability but which may be unrelated to accuracy (validity)

Human error in classification

- Everyday classification is flexible
 - people can create new specialized categories to classify atypical cases
 - Kunda and Oleson, 1995
 - presented atypical case, e.g. “introverted lawyer”
 - people either created “introverted lawyer” category or modified “lawyer” category
- Coders using a formal classification system cannot add or modify categories

Human error in classification (cont'd)

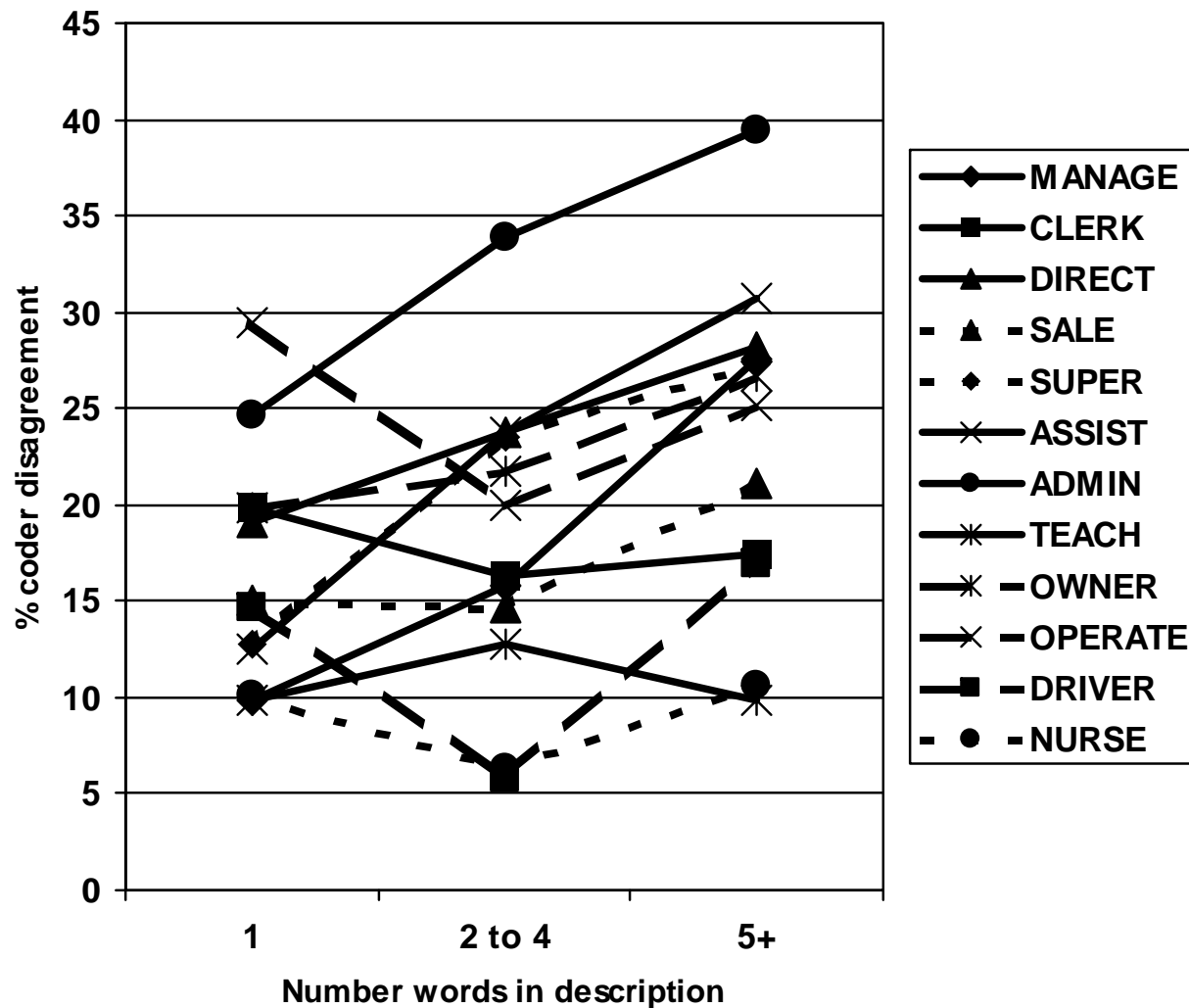
- Mistaking similarity or typicality for probability
 - Tversky and Kahneman's representativeness heuristic leads to insensitivity to base rate info:
 - someone who seems like an engineer is more likely to be judged an engineer than a lawyer *even if there are more lawyers in the sample*
 - Tversky and Kahneman's conjunction fallacy:
 - someone who seems more like a feminist bank teller than a bank teller is judged more *likely* to be a feminist bank teller *even though there are more bank tellers*
- When coding ambiguous cases, coders could, but don't, take probability into account



Linda may be more *typical* of feminist banktellers than of banktellers

but the *probability* is greater that she is a bankteller than feminist bankteller

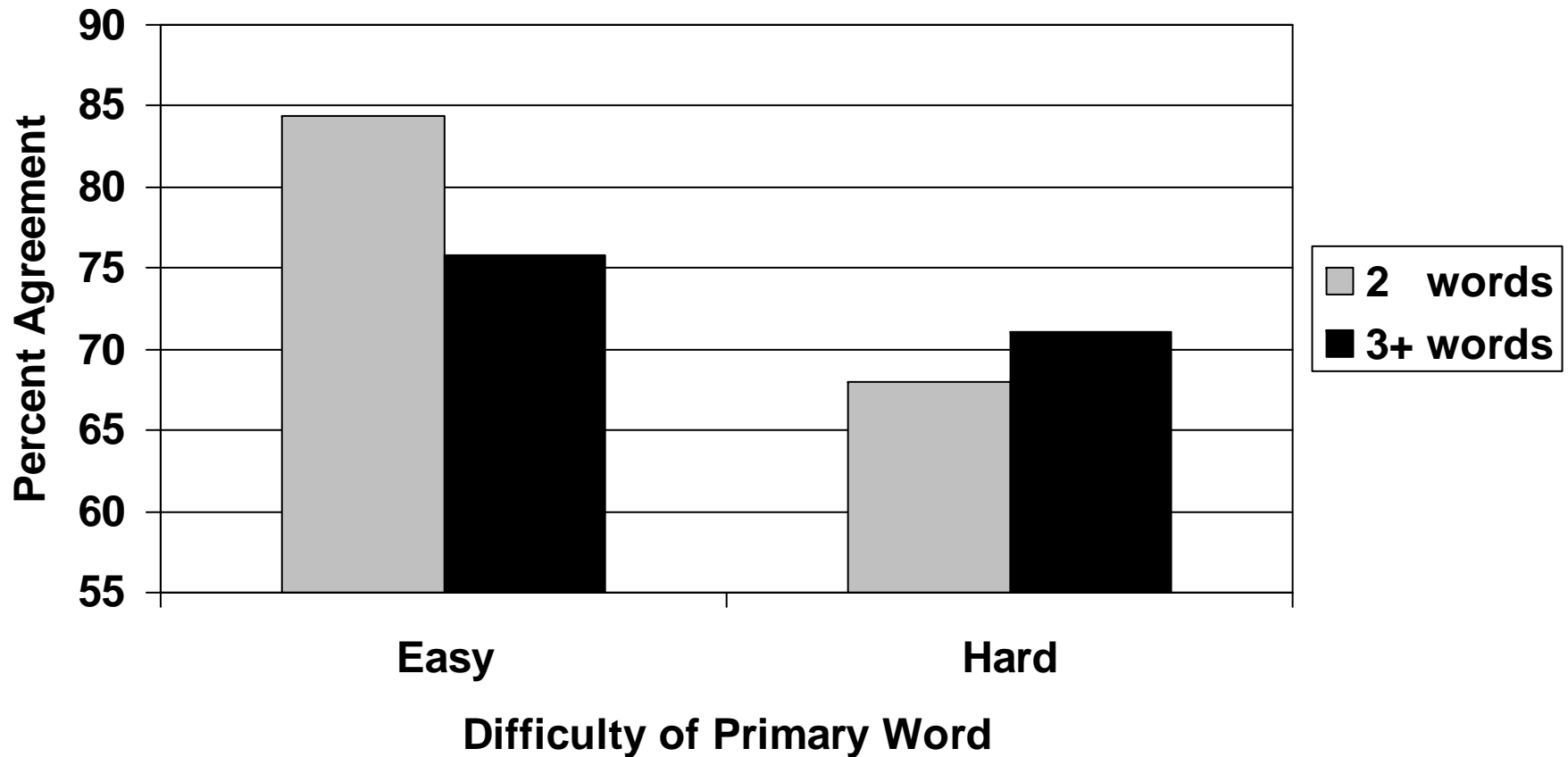
Length affects agreement differently for different words



More than one occupation (cont'd)

- Rule use could contribute to pattern of length and difficulty observed in experiment
 - only applied for relatively long descriptions
 - 2 or more occupation terms
 - rules are not used consistently and, so, agreement is imperfect
 - inconsistent rule use in easy-long descriptions may be worse than no rule use in easy, short descriptions
 - inconsistent rule use in hard-long descriptions may be about the same as no rule use in hard, short descriptions

Length only hurts agreement when words are easy



More than one occupation

EMP : DOMINOS PIZZA
IND : PIZZA
OCC : COOK, DRIVER
DUTIES : DELIVERY, COOKING

“When two different occupations are described, match to duties and go with the first duty listed.”