

Effects on Employment Classifications of Conceptual Variability of Response Category Options—Implications for Data Quality

October 2012

John Bosley, Scott S. Fricker, and Dan Gillman
U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Washington, DC 20212

Abstract

Close-ended survey response options need to be exhaustive, mutually exclusive, and well understood by respondents, but sometimes differential conceptual complexity in the response categories can make response choice difficult and reduce data quality. The present study demonstrated this effect by examining respondent classification decisions in a class-of-worker (COW) question that asks respondents to select one of four employment categories: *government*, *private company*, *non-profit organization*, or *self employed*. Study participants (n=90) read a series of narrative vignettes describing fictional employment situations, and then classified each worker using two different groupings of the COW classification. Vignettes contained cues indicative of the job's membership in just one of the first three categories, but no clear indications about self-employment. One-half of the sample made choices from among the entire set of four response options, while in the other half-sample respondents were presented with only the "self-employed" option as a "yes-or-no" choice. The vignettes were presented a second time to both groups, and everyone classified the same jobs into just three employment categories —government, private or non-profit.

The data show that the stability and accuracy of respondents' answers were highly dependent on the set of response options provided. We focus particular attention on the *self employment* class of work, which appears conceptually distinct from the other three classes, and for which we observed the highest number of classification errors. We also examine the impact of including conceptually variable response categories in the set of response options of a single close-ended question. The results of this study are discussed in the context of cognitive theories of concept formation and categorization, along with broader implications for questionnaire designers.

KEY WORDS: categorization errors, response options

1. Introduction

Question design is an important sector of survey methods research; poorly designed questions produce poor quality data. For example, when creating a response set for a closed-ended question that asks respondents to choose exactly one of the options, survey designers attempt to provide respondents with category labels that identify concepts that are mutually exclusive and unambiguous. The present study focuses on one specific aspect of question design, i.e., how cognitive and linguistic relationships within a set of response options for a closed-ended question may affect respondents' interpretation and selection of the preferred response. If different options do not fit cleanly within a shared

interpretive framework, the choices offered may confuse respondents and their response choices will not meet quality standards.

In some instances, however, the categorical choices in the set of response options reflect the survey designers' technical viewpoint rather than the kinds of conceptual distinctions that respondents are likely (or even able) to make. This disconnect undermines the effectiveness of category labels and can reduce the accuracy of survey responses (Schober and Conrad, 1997; Tourangeau et al., 2006). In other instances, ambiguity results not from the use of some constrained, technical definition of a concept, but from the nature of the concept itself. Some concepts are unidimensional (i.e., membership is determined by a single factor, like concepts of *_height* or *_weight*) and others are multidimensional (i.e., membership is determined by a set of factors, as with concepts like *_healthy* or *_clothing*). Some have relatively dense representation (i.e., they have multiple intercorrelated features relevant for category membership, with only a few features being irrelevant), while others have sparse feature representation. Empirical work in the field of cognition has shown that people often have difficulty making categorical decisions about cognitively complex concepts (e.g., sparse concepts, or multidimensional concepts when *some* features fit the concept but others do not) (e.g., Rosch and Mervis, 1975; Kurtz and Gentner, 2001). These findings are mirrored in numerous survey methods studies that have demonstrated increases in respondent comprehension problems for questions asking about vague or broad concepts (see, e.g., Tourangeau, Rips, and Rasinski, 2000, for a review).

The present study explores errors attributable to confusion among the response options due to variability in how the concepts underlying those options are interpreted and cognitively represented by respondents. As the above discussion makes clear, a lack of shared meaning about a concept can result when there is misalignment between respondents' naive sense of the concept and the survey organization's official definition, or when a category label is insufficient to evoke concepts that are by nature very broad or complex. To explore this issue, the authors to choose the "Classification of Work" (COW) question in the monthly Current Population Survey (CPS) for the study. This is an actual closed-ended survey question that has been treated as if its response options are unambiguous, mutually exclusive, and exhaustive, but where there is some evidence that this assumption is of questionable validity.

Our decision to select this item for study was motivated in part by analyses showing discrepancies between COW-based estimates of employment and similar estimates from other sources. The estimates associated with the COW category of "self-employed" workers appear especially problematic from this comparative perspective. Bowler and Morisi (2006) found that total employment estimates derived from the CPS-COW question differed substantially from estimates taken from the other Bureau of Labor Statistics (BLS) employment survey – the Current Employment Statistics survey (CES). There are substantial discrepancies between CPS-COW estimates of self-employed and estimates derived from employer-provided administrative data (e.g., unemployment insurance records; Social Security Administration's Detailed Earnings Records), as well (e.g., Abraham et al., 2007; Abraham et al, 2009).

That discrepancies exist between CPS self-employment estimates and those from other sources is, in one sense, not surprising. CPS numbers come from self-classifications (or often classifications made by household proxies) using an unspecified definition that is partly derived from common use of the concept and partly from idiosyncratic

interpretations of it, whereas comparative estimates are derived from administrative sources based on specific extrinsic features (e.g., tax paid on earnings, unemployment insurance claims). The point is that economists and policy makers often treat measures like these as data that represent the “same” concept, even if non-specialists find it obvious that this assumption is unwarranted.

In addition to results from comparative, analytic studies of COW data, a cognitive research-based perspective leads us to hypothesize that for self-employed respondents the COW classification categories will not be understood to be mutually exclusive. They may be exhaustive, but the difference in clarity of defining characteristics between the three “dependent employment”¹ classes, as a group, and the self-employed category may cause the latter to be avoided when in fact it should have been chosen from a labor economist’s or policy researcher’s perspective.²

The COW question reads as follows:

“Now I have a few questions about the job at which you worked LAST WEEK. Were you employed by government, by a private company, a non-profit organization, or were you self employed (or working in the family business)?”³

The four response options for the COW question are presented verbally by the CPS interviewer, and the respondent is required to select just one of them. The team reasoned that respondents whose work fits in any of the three dependent employment options are aided in making the appropriate classification decision because the decision is based in part on selection of employer type from among three widely-used conceptual categories of employers, i.e., private (for-profit), government, or non-profit. These concepts are grounded in distinctive, observable differences and should be easy to distinguish, although there is anecdotal evidence from interviewers that some respondents are unsure of the private vs. non-profit distinction.

Respondents who view themselves as “self employed” often are performing work that has few common characteristics with that performed by other self-employed individuals. Also, the idea of “working for oneself” is not grounded in any social-structural concepts such as a “firm” or an “organization.” These respondents are “on their own” in their work, and are also on their own in deciding among the COW classes, opening the door to the possibility that highly subjective and idiosyncratic work features may play a significant role in self-classification.

¹ “Dependent employment” is a convenient way to refer to work classified in one of the private, non-profit, or government COW options. The term is widely used outside the US to describe any employment relationship that involves an employment contract between a worker and employing organization, to somewhat oversimplify the term’s use. (ILO, 1993)

² The decision to use the CPS-COW question in this study is not intended as a criticism of the CPS methodology, a methodology that is surely one of the soundest for any national household survey in the world. The study results also should not be viewed as any kind of quality assessment of COW data.

³ This question is asked *only* if the answer to a prior question about working for pay last week was answered affirmatively, which means that the respondent is “employed” for CPS purposes. No examples of such work were included in this study’s material.

2. Objectives

The purpose of this study was to explore respondents' interpretations of cognitively complex concepts, such as the closed-ended self-classification response options for the CPS Classification of Work question. The experimental design allowed us to investigate the impact on response distributions of varying the wording of the options, varying the number of options provided, and varying the content of the verbal stimuli presented for classification.

3. Methods

3.1 Study Respondent Characteristics

Ninety participants were recruited from a pool of local residents who had volunteered to participate in BLS research studies. These participants were chosen non-randomly from this self-selected pool. To ensure an adequate sample of respondents who were in the workforce, all 90 were screened to establish that they were currently in the labor force, recently retired, or unemployed but looking for work.

3.2 Stimuli

The present study used 20 *vignettes* – brief stories about a hypothetical situation – that described various jobs. Ten of the 20 vignettes described work that was designed to appear to be “private and for profit,” 5 of the stories placed the work in a governmental setting, and the other 5 were set in a non-profit environment⁴.

The vignettes described different workers and occupations, included some contextual information about the job setting, and deliberately varied the actors' gender. All 20 vignettes were carefully crafted to unambiguously describe the general environment or *functional framework* of the work (Private, Government, Non-Profit). At the same time each vignette deliberately avoided providing explicit indications of whether or not the worker was self-employed, in order to make the self-employment classification decision as subjective as possible for the study participants.

An example of each vignette type illustrates this strategy. These examples also give the reader a sense of how well the goals of clearly defining a type of functional work environment and at the same time omitting any clues about dependent vs. independent (self) employed job characteristics were achieved in practice.

1. Private For-Profit Work Example: ” Marcella has done graphic design work since she graduated from college. She is currently working on a long-term contract with an advertising agency in a city near her home. She works from her home as well as on-site at the client's offices. Marcella gets new business by word-of-mouth and by putting ads in the local paper.”
2. Non-Profit Work Example: —Nikolas does the maintenance and repair work at three inner-city schools for ‘at-risk’ young people who have dropped out of the local public school system but want to get more education or career training. A community

⁴ This doubling of the frequency of vignettes designed to indicate private employment, relative to the other two dependent types, was motivated by the desire to mimic the much higher frequency of private employment in the US economy than employment by government or non-profit organizations. The 2:1 ratio is a very rough approximation of this inequality, but is not intended to represent the actual proportions precisely

organization that is funded by corporate contributions operates these schools. Nicholas is skilled in plumbing, wiring, repairing HVAC equipment and other maintenance tasks.”

3. Government Work Example: —Lawrence is an experienced automotive technician who maintains or repairs the vehicles (police, fire, ambulance, etc.) owned by his town. Lawrence works on the vehicles in a town garage, using his own tools. The city transportation department buys needed parts and supplies from its approved vendors, from requisitions that Lawrence prepares and submits to the Town Clerk’s office.”

All 20 vignettes were worded like these examples, and contained clear indications of where the vignette should be placed in a 3-option classification system while leaving the self-employed vs. not self-employed nature of the work indeterminate. This set of vignettes is termed the Baseline (B) set.

Variations to Strengthen or Weaken Features of Self-Employment

The study team decided to add a second source of variation to see whether manipulating vignette content to make the jobs appear more or less “self-employed” in nature would shift choices toward or away from the self-employed option. Since no concise, formal definition of “self-employed” has been developed to apply to the CPS use of the term, the study team performed a quick survey of economists who either collect or analyze and interpret self-employment data to see if these experts would agree on any attributes of work that help define it clearly as “self-employment.” This expert panel showed almost complete agreement that two attributes are key defining characteristics of self-employment. One was “Extent of personal financial stake in employment,” while the other was “Degree to which work is supervised and/or directed by another.” These key attributes were labeled “Financial” and “Supervisory” factors, respectively.

The team added a sentence to the baseline vignette to indicate the presence of, or lack of, these defining characteristics. The resulting vignettes created four additional stimulus classes, labeled “Financial Positive” and “Supervisory Positive” (F+, S+) where the added description was intended to make the job appear more like self-employment. On the other hand, when descriptions made the job appear less like self-employment, the variations were termed “Financial Negative” (F-) and “Supervisory Negative” (S-). To give readers a flavor of the types of sentences used to create the four variants, we provide examples of the S+, S-, F+, and F- statements from one of the Non-Profit vignettes (Nicholas; see above for its baseline version), below.

Supervisory Positive (S+) Nicholas has to schedule his own work to cover all three schools so they depend on him to do all the needed work on time.

Supervisory Negative (S-) There is a staff member at each school who assigns the maintenance work to Nicholas.

Financial Positive (F+) Nicholas sometimes needs to work out a deal with the schools to get enough hours to keep a steady income.

Financial Negative (F-) Nicholas’s work is so important that he is guaranteed a certain number of hours of work and some benefits by the community organization.

The combination of the Baseline and the S+/S- and F+/F- variants yielded a total of five separate versions for each of 20 vignettes—a total of 50 Private For-Profit vignettes of all types, 25 Non-Profit vignettes and 25 for Government type workers. The study team chose “Variety” to refer to these five different versions of the vignettes as an experimental variable. As previously noted these variations were included in the stimulus blocking scheme to insure that all were presented equally frequently.

4. Procedure

4.1 Baseline Vignette Pretest

The study team conducted an informal pretest in order to assure that the Baseline vignette set would be consistently classified into the *intended* 3-OPTION categories. This paper adopts the term “*intended classification*” to denote the assignment of a vignette to the 3-OPTION category that is consistent with how its content describes the work environment and functional setting. Fifteen employed pretest participants read and classified all 20 *Baseline* vignettes into *only* the 3-option categories. Self-employed was not mentioned as a possibility. These proxy respondents also were encouraged to “think aloud” as they considered the options, to identify what elements of vignette content had the greatest influence on their final categorizations. Analysis of data from a first round of pretesting indicated that 15 of the 20 vignettes were appropriately classified by everyone in a small pretest sample. Content in the remaining 5 Baseline vignettes was modified to remove likely sources of misclassification uncovered by the “think-aloud” procedure. In a second round of pretesting with 6 newly-recruited participants, all 5 of these modified Baseline vignettes were classified to match their intended categories by this smaller sample. The resulting versions were the Baseline vignettes used in the study.

4.2 Main Study Procedures

The main study was conducted in a laboratory setting at the Bureau of Labor Statistics. Each participant was given a booklet containing two sets of the same 12 vignettes. Each vignette described a fictional worker, and the participant was instructed to classify the work in the vignette by choosing an option from a set that appeared below the vignette on the page. The two sets of 12 offered different sets of response options on the two passes through the set.

Each participant’s booklet had instructions printed on the cover page, and the classification tasks were performed at the participant’s own pace. Each vignette had its own page in the booklet. Participants were asked to read and consider each vignette carefully before classifying it. They were also discouraged from comparing their choices on the first pass through the vignettes with their second choices for any of workers described. Vignette order was randomized between halves of the booklet so respondents did not see vignettes in the same order twice, to minimize carry-over response effects.

Participants were also asked to rate the confidence they had in the correctness of the choice they made for every classification decision, using a 5-point rating scale that appeared on the page along with the classification categories.

After completing the classification task, respondents also completed a two-page debriefing form that asked them to provide definitions of each conceptual category. The following question was designed to elicit such definitions:

—Forgetting about the stories you just read in the study booklet, please give us a *brief* description in your own words about what it means to you for a worker to be:

- Employed by a government office or agency?
- Employed by a private for-profit company?
- Employed by a not-for-profit organization?
- Self-employed?”

Open-ended responses were invited in a large blank field on the page.

One of the team entered the debriefing responses verbatim in a spreadsheet as they were collected, and early in the process noted that the original question about definitions was not producing many informative responses; few respondents were able to *define* the categories in terms of features or attributes. This qualitative but strong impression led to a form revision halfway through the study to add a second part to the question that was designed to elicit examples of employment in each of the COW categories in addition to the feature-based definitions. This question had the following form: —Please give 1 to 3 examples of employers that are government offices or agencies” (similarly worded items were asked about the remaining employment categories). This additional inquiry enabled comparison of the content and quality of definitions and of examples for 47 of the 90 participants, and this comparison proved to be quite informative, as the Results section will show.

The 90 participants were randomly assigned to one of two groups: (1) Standard, or (2) Split condition. In their two passes through the 12 vignettes, the Standard group first saw a 4-option response set, followed by a 3-option set. The data derived from the pass that included Government, Private, Non-Profit, or Self-Employed choices, as in the actual COW question, will be referred to as the 4-option response set. The 3-option response set excluded the response of self-employed.

In the Split-condition group, one pass through the 12 vignettes asked participants to classify the vignettes as either self-employed or not (SE or not-SE.) These data will be referred to as the 2-option response set. In the second pass through the 12 vignettes, the Split-condition group used the 3-option response set.

Using the 3-option response set in both groups showed how well the study team had succeeded in indicating the government (G), private (P), or non-profit (N) character of the work described. Presenting the 3-option set in both groups allowed an estimate of the within-participant consistency of vignette assignment to these three categories independently of (positive) self-employment classification rates in both the Standard and Split-condition samples.

In both treatment groups, the orders of presenting the 12 vignettes with each different response set were alternated so the 3-option response set was offered randomly to half the participants on the first pass through the vignettes, and to the other half on the second pass. A randomized block design also was implemented to manipulate of the presence or absence of the workers’ supervisory and financial independence characteristics across the different vignettes.

5. Results

Pretest results demonstrated that when only the 3-option response set was offered to a small pretest sample, everyone in that sample assigned all Baseline cases as intended. This report will use the conventions of “intended” or “appropriate” classification as a substitute for “correct” 3-option classification, since these category assignments were based on respondent inferences from the cues provided and not on any “gold standard” of correct classifications. Additionally, since the vignettes provided no explicit cues regarding self-employment, any self-employed classification had to be based entirely on a subjective interpretation of what it means to be “self-employed.” This last consideration is what makes any shifts in the proportion of self-employed classifications between experiment conditions especially noteworthy.

Table 1. Percentages of 3-option Cases Placed in Intended G, P, or N Classifications

Intended Classification	Participants' Classification		
	Government (G)	Private (P)	Non-Profit (N)
Government (G)	73.3%	22.2%	4.4%
Private (P)	5.7	86.7	7.6
Non-Profit (N)	12.2	19.6	68.1

Table 1 displays the overall pattern of vignette classifications for the three-option classifications (ignoring S and F Variety manipulation), combined across both Standard and Split-condition treatments (n=1080 classifications). The distributions of 3-option responses did not differ significantly between the two presentation orders in any of the treatment groups, so we collapsed across passes in Table 1 and all subsequent analyses.

The bolded main diagonal of Table 1 shows that the proportion of cases where respondents appropriately mapped the vignettes onto their indicated categories (G, P, or N) varied considerably across employment types in the combined sample. The highest proportion (86.7%) of appropriate classification is seen for the Private category, with both Government (73.3%) and Non-Profit (68.2%) rates of appropriate classification as much as 13-18% below that level.

Table 1 not only shows nearly 87% appropriate classifications for the Private category, but it also clearly shows some bias toward use of the Private category for classifying *all* of the vignettes, even when the verbal cues indicate Government or Non-Profit type of work. Thus we see that vignettes that were intended as Government were misclassified as Private 22.2% of the time, but as Non-Profit only 4.4% of the time. Similarly, vignettes intended as Non-Profit were preferentially placed in the Private category 19.6% of the time, versus only 12.2% being misclassified as Governmental.

Stability of Individual Participant's Dependent Employment (G, P, N) Classifications

Table 2 shows how consistent (reliable) participants in the Standard condition (N=540) were in classifying the same vignette on different passes, using the 3-option COW question on one pass and the 4-option item on the other. Respondents' GPN classifications were similar, with vignettes classified into the same GPN category in 80-83% of cases. Interestingly, in the majority of the instances in which participants changed

their classifications, they placed G or N vignettes into the P category. This agrees with other results that the participants tended to resolve uncertainties about a vignette’s 3-option classification by assigning it to the Private category (as long as it was not assigned to the self-employed category in the 4-option condition.)

Table 2. Percent of Matching Vignette Classifications with 4- and 3-Option Item (Within Participants, Standard Condition Treatment Group)

Classifications based on 3-Option Item	Classifications based on 4-Option Item		
	Government (G)	Private (P)	Non-Profit (N)
Government (G)	83.5%	13.8%	2.8%
Private (P)	10.8	80.6	8.6
Non-Profit (N)	4.4	13.3	82.2

Table 3 highlights the results that bear most directly on the key study hypothesis; that is, presentation of COW response options in the 4-option set will yield a lower rate of self-employed vignette classification than will the 2-option response set that forces a dichotomous “yes-no” choice with respect to self-employment.

Table 3. Rates of “Self-Employment” Classifications in 4-Option vs. 2-Option Conditions

# of Response Options	Rate of Participants’ Classifications (and frequency)				
	Government (G) (n=135)	Private (P) (n=270)	Non-Profit (N) (n=135)	Self-Employed (SE)	2-Option Self-Employed (SE/Not SE)
Four	20.4% (110)	25.4% (137)	16.5% (89)	37.8% (204)	
Two				58.7 (316)	41.3 (222)

The bolded percentages in the two Self-Employed cells show the overall effect on self-employed classification rates of the 4- and 2-option response sets (ignoring supervision or financial variations). Fully 20% more of participants classified the work as ‘self-employed’ with the 2-option item than with the 4-option item. Clearly, forcing respondents to make a choice between “self-employed” or “not self-employed” dramatically increased the frequency of self-employed choices over the rate for those respondents who were offered four choices, including three distinct options other than self-employment.

As discussed earlier, half of the vignettes were intended as P and a quarter each were intended as G and N. Thus, the 4-option classifications in Table C data included 135 instances of vignettes intended for assignment to each one of the G or N dependent employment categories, and 270 instances of intended P vignettes. The first row of the table shows the pattern of actual 4-option classifications of these 540 vignettes. These row data thus trace the portions of vignettes intended for each of the three dependent categories that were instead classified as self-employed. They do not reflect any shifts away from the intended classification to one of the other dependent employment

classifications, like those displayed in Table 2. About 80% of the intended G vignettes were classified appropriately, as were about 50% of the vignettes intended for the P category, and about 66% of the vignettes intended for the N category. Thus, nearly half of the 38% of the 4-option total that were classified as self-employed were intended as P cases, but only about 18% of intended government vignettes were classified as self-employed, and participants only assigned about 34% of the intended non-profit vignettes to the self-employed category.

Effects of Supervisory and Financial Variations

We next examined the impact of varying the level of supervisory independence or financial autonomy (the baseline vignettes and the four S and F varieties). Adding information to indicate lack of supervision (more worker autonomy and control), or to indicate a lack of any involvement of others in generating earnings or with the worker's financial management (more financial independence,) created "supervisory-positive" (S+) and "financial-positive" (F+) varieties respectively. Adding information to indicate some degree of supervisory control, or of some kind of external intervention to control the job-holders' work-related financial freedom created the counterpart supervisory- and financial-negative varieties (S- and F-). The added information in the S- and F- varieties (indicating more supervision or less financial independence) are predicted to *decrease* participants' choice of self-employment (SE) for a vignette. These varieties should *pull* the vignette away from self-employment and toward the intended 3-option category. Conversely, S+ or F+ should *push* the classification choice toward self-employed relative to the Baseline version. The prediction was made that these variations would affect participants' frequency of selecting the self-employed option in both the 4- and 2-option conditions. The further prediction was made that the probability that a Baseline vignette would be classified as SE lies somewhere between that for the S+/F+ and the S-/F- versions.

Table 4 shows the joint effects of vignette variation in additional supervisory or financial independence cues on the pattern of 4-option classifications of work only, and thus includes 540 vignette classifications. Expected cell frequencies would have been unacceptably small in a table that tabulated every Baseline, S and F Variation separately, so certain variations are combined in Table 4 as indicated in the row labels.

The SE row shows that category's use for the entire remaining set of S+ and F+ vignettes collapsed across all intended 3-option classes. It thus includes all those vignettes for which the combination of added S+ and F+ information were hypothesized to *maximize* the probability of respondents' choice of SE.

The cell entries show that out of 181 S+/F+ vignettes collectively, about 20% of these were classified as G, 24% as P, and a little over 14% as N, yielding a total of over 57% that were distributed across the 3 options other than self-employed. Still, fully 42% of these S+/F+ vignettes were classified as SE, and compared with an overall frequency of about 38% SE classification in Table C, this 4% increase provides modest support to the prediction that the added S+/F+ information would increase self-employed classification frequency.

Table 4. Rates of Appropriate G, P, N, or SE Classifications for Different Financial and Supervisory Vignette Variations^a

Intended Classes	Participants' Classifications (row %s)			
	Government	Private	Non-Profit	SE
Government (B, S-, F-)	67.4% (60)	7.9% (7)	1.1% (1)	23.6% (21)
Private (B, S-, F-)	2.8 (5)	41.3 (74)	3.4 (6)	52.5 (94)
Non-Profit (B, S-, F-)	9.9 (9)	14.3 (13)	61.5 (56)	14.3 (13)
SE (All S+, F+ varieties)	19.9 (35)	23.8 (43)	14.4 (26)	42.0 (76)
Total (N=540)	110	137	89	204

^a Data from the 4-option treatment group only (n=540)

The bolded main diagonal in Table 4 shows that, once again, majorities of the combined B, S-, and F- vignettes were assigned to their intended 3-option classes for two of those classes—Government and Non-Profit. In the Private category, however, participants placed 52% of the vignettes in the self-employed category despite added cues that were designed to make the work appear to be less self-employed in nature. In this condition, only 41.3% of participants chose the intended P class. By comparison, vignettes intended as Government were categorized as self-employed in 23.6% of cases; Non-Profit vignettes were classified as self-employed or as Private in nearly 15% of these cases. This pattern of classifications seems once more to show that vignettes intended to be classified as Private are most readily perceived as self-employed—even in the presence of cues designed to work against such a perception.

Table 5. Percentages of Appropriate Vignette Classifications for Baseline and Varieties in the 4-Option Condition

Vignette Varieties	Appropriate Classification (GPN only) (Row % and N)				
	Government	Private	Non-Profit	SE	Total
Baseline	18.3% (33)	28.9% (52)	15.6% (28)	37.2% (67)	180
S+ and F+ combined	19.9 (36)	23.8 (43)	14.4 (26)	40.3 (73)	181
S- and F- combined	22.9 (41)	23.5 (42)	19.6 (35)	34.1 (61)	179
Total (n=540)	110	137	89	204	

The bolded numbers in Table 5 show that adding information about worker supervisory or financial autonomy to the baseline vignettes did have a modest effect on SE classification rates in the 4-option subsample. These effects were in the predicted directions: the combination of S+ and F+ variant information added 3.1% to SE choice rate (compared to the Baseline), and the combination of S- and F- information caused an equivalent decrease in SE choices (3.1%).

Summary of Analysis of Debriefing Questions

The study participants completed a debriefing form after finishing the work classification task. Initially the debriefing form asked only about participants’ feature-based definitions of each of the COW categories; that is, it asked participants to list a few of the work characteristics mentioned in a vignette that led them to place it in one of the four categories. The study team member who transcribed the debriefing form data as they were collected noticed that the participants were having a difficult time generating such definitions of some of the categories, and so the form was modified about half-way through the study to ask for both definitions and examples of the four *classes* of employment (examples of the types of work that the respondent spontaneously thought of for each COW category).

Following this modification, participants showed an interesting and surprising difference in responses to these two work-type definitional and examples questions. Specifically, they generally provided fitting examples--frequently multiple examples--of work situations in the three dependent employment categories, but were able to generate examples of self-employed work *types* much less frequently. Conversely, participants were able to provide fewer adequate feature-based definitions of the three dependent classes of work while a greater number were able to provide fairly clear feature-based definitions of “self-employed” work.

We transcribed the debriefing responses and rated the quality of examples *and/or* definitions which each of the participants provided for each of the COW classes, using a 5-point scale anchored at 0 on the lower end. Table G shows the frequencies with which participants furnished poor examples or definitions (ratings of 0-2) versus frequency of relatively good examples or definitions (ratings of 3-5).

Table 6. Frequency of “Poor” vs. “Good” Ratings of Examples and Definitions for COW Classes

COW Class	Definitions (n=90)		Examples (n=47)	
	“Poor”	“Good”	“Poor”	“Good”
Government	75	15	5	42
Private	80	10	11	36
Non-Profit	69	21	13	34
Self-Employed	42	48	19	28

Note in Table 6 that, overall, participants were rated as providing “good” definitions of for SE more frequently than for the three dependent employment classes, although even for SE “good” definitions were provided in fewer than 50% of cases. Among the other classes, non-profits were somewhat more frequently adequately defined than the other two. The differences in proportions of “good” ratings for the examples are less starkly contrasting when the three dependent options as a group are compared with SE. Still, it is apparent that these participants were better at generating examples of work types for the other three categories than for SE. This was particularly true for governmental work, not surprising since the participants live in or near the Nation’s Capital where government agencies, and hence examples of such work, abound. The Discussion section will examine these ratings from a more theoretical perspective.

6. Discussion

Analysis of response patterns to the three alternative response option sets revealed that when respondents were offered four distinctive classification options, they generally assigned each vignette to the category corresponding to its job-related cues that indicated the type of work. When the distinctiveness of the three different work types was obscured by collapsing them into a single undifferentiated type (“Not Self-Employed”), a much higher proportion of the vignettes were now classified as “Self-Employed.” In a control condition, the respondents placed most of the vignettes into the one of the three dependent categories that its cues indicated was the intended categorization. The basis of the classification had shifted strongly from job type (explicit in each vignette) to worker type (indeterminate in all the vignettes.) This shift in response pattern was slightly affected by the introduction of additional verbal cues designed to suggest indirectly that the job/worker was more or less self-employed in nature.

The initial theoretical framework for the study was the prototype theory of concepts, as developed by Rosch and her colleagues beginning in the 1970s. Rosch’s 1978 book chapter provides a good summary of key studies in this tradition. A likely source of conceptual classification response error such as we were seeking to demonstrate, according to this well-established body of research on the cognitive structure of concepts, is the graded nature of membership in conceptually-based categories. This theoretical tradition views categories as including valid members with different grades of membership. Degree of membership is based on the idea of a central mental construct of a prototype that is the theoretical “best example” of the concept. Other perfectly valid conceptual instances can differ quite a lot from the prototype before the gradient of similarity disappears at the concept boundary (Rosch, 1978/1999).

Employing a Roschian framework, the concepts underlying the dependent employment COW categories would all be typified in terms of such features as: a long-term employment contract, job performance required to take place in an assigned space owned or rented by the employing organization at least part of the time, provision by the employing organization of necessary equipment and supplies to accomplish work tasks, a well-defined pay rate tied to a specified time-based or other unit of work, and so on. Above all, each of the categories is typified—and indeed distinguished—by the goals, mission, and revenue basis for the employing unit (an individual employer in a small business, or a larger more formal organization). Self-employed work, on the other hand, has fewer such obviously typical characteristics. The self-employed worker may or may not work consistently at one location. And if she does, she will normally be responsible for buying or renting that space herself, and paying for it out of earnings. The self-employed may supply all necessary equipment and supplies, but they may borrow or rent equipment as needed. Their rate of pay and hours of work may be erratic, including seasonal. In other words, it is much more difficult to describe a *typical* “self-employed” worker than a *typical* dependent one.

When all four COW response options, including self-employment, are presented as possible job classes, any of the hypothetical jobs described in our vignettes will have less tendency to be evaluated as “self-employed” than as one of the other categories. This is because the job descriptions all included cues that clearly associate the job with one of the 3-option dependent employment classes, and the notion of typicality-based similarity seems congruent with these categories. The information given in the vignettes also made

it *possible* for respondents to categorize it as SE—but such a decision would require a much more thorough cognitive processing of the work descriptions. It seems unlikely that different participants' interpretations would converge on the same set of features as "typical" of self-employed work and hence lead them to classify a vignette as SE. The responses to the debriefing questions discussed above indicate that the participants indeed found it difficult to list any set of common features to define or typify "self-employed" work. In the case of the COW categories, it appears that the three dependent classes of work are likely to have fewer *dimensions* that vary than "self-employment," which appears to be high-dimensional in nature.

When the respondents are only permitted to decide whether the job is SE or not, by presenting this single option as a choice contrasted with an undifferentiated "otherwise employed" option, the respondents must cognitively process the job description much more thoroughly, in order to find cues or clues that they associate with their personal concepts of "self-employed" on which to base their decision. But what are those cues? It turned out that there were no well-specified, agreed-upon features or attributes of self-employed workers' jobs. So the study team recognized that, within a prototype theoretical framework, binary decisions about self-employment or its absence might turn out to be so individually distinctive that the data would contain a large degree of essentially random variation from the standpoint of a strictly prototype theory framework.

Two well-known specific cognitive influences on survey response likely played a role in our main study findings: respondent satisficing (Krosnick et al., 1996) and response order effects (Bishop and Smith, 1997). Response order effect refers to respondents' tendency to choose options that occur at the beginning or end of a closed-ended list that they read, or that is read to them in the same order. Given this vagueness of the "self-employed" concept, workers who belong in this class may find it easier—and therefore tempting—to choose the dependent employment category that is most clearly congruent with the nature of their work, rather than make the cognitive effort to decide whether they "if" the "self-employed" class—they may satisfice. And, because our COW question offered its options in the same order to all respondents (following the convention used in the actual CPS COW item), satisficing may have led to response order effects.

We may further speculate that the four COW employment categories represent a small concept system. Our data and our interpretation of them indicate that it is not the nature of each of the categories taken alone but the relationships among them (especially the difference between all three dependent employment categories and self-employment) that influenced their relative attractiveness as choices for proxy job classifications in our study. Tables 1 through 6, and especially Table 3, demonstrate these contextual effects. This is particularly striking in the patterns of SE categorization when the presence of the Private dependent alternative varies. If the Private category is available it seems to hold a strong attraction as a category for work descriptions that will be placed in the self-employed category when Private is not offered as a choice. Note however that this is an intra-question context effect, and thus different from the inter-question context-setting that survey research usually studies.

If authors of survey questions are not aware of distinctions among types of concepts, they may generate a set of response options that appear to be clearly distinct from one another, and indeed may meet formal logical criteria of mutual exclusion and exhaustion, but that will not be seen as equivalent and distinct alternatives by the typical survey respondent. Their distinctiveness may be lost on respondents if the options mingle categories divided

between two different concept types. Further, the COW categories themselves are always in flux as traditional forms of employment disappear and are replaced with new and often complicated work arrangements. Some of these further blur the distinctions between dependent employment and —working for one’s self.” This ongoing change process could make future studies of the conceptual structure of employment even more challenging—but timely and even necessary.

References

Abraham, K., Haltiwanger, J., Sandusky, K., and Spletzer, J. (2007). —Comparing household administrative measures of self-employment.” Paper presented at the Princeton Data Improvement Initiative conference, February.

Abraham, K., Haltiwanger, J., Sandusky, K., and Spletzer, J. (2009). —Exploring differences in employment between household and establishment data.” NBER Working Paper No. 14805, March.

Bishop, G. F., and Smith, A.E. (1997) —Response-order effects in public opinion surveys: The plausibility of rival hypotheses,” *JSM Proceedings*, Survey Methods Research Section. Alexandria, VA. American Statistical Association. 1041-1046.

Bowler, M. and Morisi, T. (2006). —Understanding the employment measures from the CPS and CES survey.” *Monthly Labor Review*, February, 23-38.

Schober, M.F. & Conrad, F.G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 61, 576-602.

Krosnick, J. A., Narayan, S., and Smith, W.R. (1996) —Satisficing in Surveys: Initial Evidence.” In M.T. Braverman & J. K. Slater (Eds.), *Advances in Survey Research*. San Francisco: Jossey-Bass, 29-44

Kurtz, K. and Gentner, D. (2001). Kinds of kinds: Sources of category coherence. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Erlbaum, 522-527

Rosch, E. (1978). *Principles of categorization*. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Erlbaum. Reprinted in: Margolis, E. and Laurence, S. (Eds.) (1999). Concepts: Core readings. Cambridge, MA: MIT Press.