

Everyday Concepts and Classification Errors: Judgments of Disability and Residence

Roger Tourangeau¹, Frederick Conrad, Zachary Arens, Scott Fricker,
Sunghee Lee, and Elisha Smith

This article examines two neglected sources of misinterpretations of survey questions. Respondents may misunderstand the questions because the survey uses an everyday term in a technical way that differs from the everyday sense (and respondents fail to recognize the difference); in addition, respondents may have trouble applying the concept to borderline cases, situations that do not map neatly onto whichever sense of the concept they are trying to use. We call these two problems *misalignment* of the concepts and *imperfect fit* between concept and instance. We examined two everyday concepts – residence and disability – that figure prominently in surveys. Our initial experiment gave respondents definitions for residence and disability and asked them to classify vignettes describing concrete instances. We constructed one definition that reflected the everyday concept (the *everyday definition*) and a second one that departed from it (the *technical definition*). The vignettes varied in how well the instance depicted matched the requirements of the two definitions (for example, some vignettes clearly met the requirements of one definition but clearly did not meet the requirements of the other). Participants who got the technical definition for residence seemed able to keep it distinct from the everyday definition but those who got the technical definition for disability seemed to fall back on their everyday concept. In addition, participants were better at classifying vignettes that closely matched the definition (*central* instances) than ones that did not match it so well (*peripheral* instances). In our second experiment, we tried to encourage participants to pay more attention to the definition of the concepts by giving the concepts unfamiliar labels (e.g., calling a residence an *enumeration unit*). Although the use of unfamiliar terminology did increase the proportion of respondents who consulted the definition of the concept as they judged the vignettes, there was strong evidence (especially for disability) that participants still relied on their everyday sense of the concepts. Respondents were again more accurate in classifying central than peripheral instances. People have difficulty using concepts in unfamiliar ways and, when they have to, they often make errors. Even with concepts used in their everyday senses, people have trouble classifying borderline instances, those that only partially match the definition of the concept. Category membership is graded, rather than all or none, and people have trouble dealing with cases that are near the concept's boundary. We suspect that both misalignment and imperfect fit often produce substantial measurement error in surveys.

Key words: Reporter errors; conceptual misalignment; definitions; concepts; categories.

¹ Corresponding author: University of Maryland, Joint Program in Survey Methodology, 1218 LeFrak Hall, College Park, MD 20742, U.S.A. Email: rtourang@survey.umd.edu

Acknowledgments: We are very grateful to Molly McNeeley for her help in collecting the data for Study 1. Study 3 was supported by a grant from the National Science Foundation (NSF) (SES0106222) to Roger Tourangeau, Fred Conrad, Mick Couper, and Reg Baker. We gratefully acknowledge NSF's support. We also thank Mick Couper and Reg Baker for their helpful suggestions on Study 3 and Lance Rips for his helpful comments on an earlier draft of this article.

1. Introduction

Survey researchers have long been concerned about the possibility that respondents may misunderstand the questions in surveys. Belson (1981) reported that respondents frequently misunderstood even familiar everyday terms like “you” or “children,” adding considerable fuel to these worries about comprehension problems. Buttressing Belson’s argument, Schober and Conrad (1997) demonstrated that there can be dramatic improvements in the accuracy of survey responses when interviewers are trained to clarify the questions. In discussing the process of understanding survey questions, Tourangeau, Rips, and Rasinski (2000) distinguish seven problems that can lead to misunderstanding of questions, ranging from the use of unfamiliar technical terms in the questions to overly complicated syntax; the sheer number of distinct problems further suggests that misunderstanding can be a serious source of error in surveys.

Our research focuses on two relatively overlooked reasons that respondents may misunderstand survey questions. One problem is that respondents think that they know what the question means but in fact apply a different definition for a key term from the one used by the questionnaire. Surveys often use everyday terms in specialized ways, and, when they do, respondents may fail to adjust to the new definition that they are supposed to apply in the survey. We refer to this problem as *misalignment* between concepts – the definition the respondents apply does not fully line up with the definition they are supposed to use. The second problem that can lead to errors is that respondents may have trouble applying whichever definition they do adopt to the instances in question. Cognitive psychologists have argued for some time that concepts are graded in the sense that some instances are more clearly included in the concept than others (e.g., Rosch and Mervis 1975; Smith, Shoben, and Rips 1974). For example, a chair is more clearly an example of *furniture* than a TV or a lamp is. Respondents may be uncertain about whether a given potential instance falls within the boundaries of the concept. Schober and Conrad (1997) describe what they call “complicated mappings” between the survey question and the respondent’s situation and show that these can lead to high levels of error. We argue that these complicated mappings reflect the graded structure of concepts and the resulting difficulty that respondents have in dealing with borderline cases. We refer to this second source of comprehension problems as *imperfect fit* between the definition and the instance.

Both problems seem to reflect basic features of the structure of everyday concepts. Everyday concepts often have multiple, sometimes conflicting organizations. For example, Ross and Murphy (1999) argue that the real-world category *food* follows at least two important structural principles – we can readily classify foods by what they are (bread, meat, fruit) and by how we eat them (main courses, breakfast food, snack food). Both structural principles are clearly useful, and similar rival structures may be found with many natural concepts. (We follow Medin and Coley (1998) in referring to the mental representation of a category as a concept and the collection of instances picked out by a concept as a category.) In some cases, the different organizational principles for a concept lead to different classifications of specific instances. For example, Hirschfeld (1994) has argued that there are both biological and social senses for kin terms (like *father*, *mother*, and *brother*). The biological sense rests on the exact reproductive facts, whereas the social sense reflects characteristic patterns of family interaction. This latter framework allows

children to use kin terms accurately even though they lack the relevant biological knowledge. Although the two senses generally point to the same classification of instances, it is possible for the biological conception to clash with the social one, forcing us to use terms like “biological mother” to indicate which framework applies in a given case (see also Medin 1989, on “essentialism”).

Another reason that everyday concepts may have multiple senses is that the terms used to describe them are appropriated for official use. Bureaucracies often use everyday terms in ad hoc or technical ways for various official purposes. We focus here on two concepts that have both everyday and official senses, the concepts of residence and disability.

The concept of residence is crucial in a number of official transactions with government agencies, ranging from payment of property taxes and registering to vote to getting a driver’s license. In the United States, one specific transaction that involves the concept of residence is filling out the questionnaire for the decennial census. The first item on the Census 2000 short form asked respondents “How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?” Although the U.S. Census Form does not use the term residence, the instructions it gives to clarify the question (“include . . . people living here most of the time while working, even if they have another place to stay”) suggest that each person’s usual residence is the relevant concept. In addition, people seem to base their answers to the census question on their everyday concept of residence (Gerber 1994; Gerber, Wellens, and Keeley 1996).

For a variety of reasons, the concept of residence used in the U.S. Census and other surveys differs from the everyday concept of residence. In the first place, bureaucracies have a much lower tolerance for ambiguity than everyday concepts often exhibit. In everyday usage, it rarely matters whether someone who spends time at two homes counts the one place, the other place, or both as his or her usual residence. For a census, though, these distinctions are critical. Accordingly, the U.S. Census Bureau follows a number of residence rules in resolving ambiguous situations (e.g., students living away from home at college) and to assure that each person is counted once and only once. Aside from the attempt to impose greater precision, official uses of concepts like residence differ from everyday uses in a second way – they tend to disregard much of the social or emotional framework that is often a key structural component of the everyday concepts. For example, our results indicate that the everyday concept of residence involves notions like living with one’s family or sharing the costs of maintaining an apartment with a roommate. These notions play no direct role in the official definitions of residence for the purposes of the census or other surveys.

Gerber and her colleagues at the U.S. Census Bureau have explored the everyday concept of residence in an effort to improve the accuracy of responses to the census questions (Gerber 1994; Gerber, Wellens, and Keeley 1996). This work has shown that the census concept of “usual residence” is unfamiliar to many respondents, who think instead in terms of the concepts of *home* or *permanent address*. In some ways, these results parallel Hirshfeld’s findings about kin terms. There is one sense of residence (home) that is largely social in nature and a second sense (permanent address) based on more objective criteria; in the case of residence, the objective criteria involve official transactions (such as where people get their mail or where they vote). Partly to deal with the discrepancy between the everyday and official senses of residence, the Census 2000 questionnaire gave

some guidance as to who should be counted (e.g., “INCLUDE . . . foster children, roomers, or housemates”) and who should be left out (“DO NOT INCLUDE . . . college students living away while attending college”). Unfortunately, the research suggests that people often disregard such definitional information. When the definitions are intuitive, respondents do not need them; when they are counterintuitive, they do not follow them (Gerber et al. 1996).

Disability is another example of an everyday concept with several potentially conflicting senses. Our findings (described later) suggest that the everyday sense of this concept is organized both around specific disabling conditions (e.g., Alzheimer’s disease) and the impairments these conditions produce (e.g., inability to work). In addition to the everyday sense, several researchers (e.g., Jette and Badley 2000) have proposed more formal definitions for the concept and these have provided the underpinnings for the measurement of disability in surveys and academic studies. Apart from its everyday and research uses, the term disability is often used in official senses as well. Most economically developed countries have systems to support workers who become unable to perform their jobs due to illness or injury, leading to detailed official definitions of eligible disabilities. In the U.S., workers must convince the Social Security Administration that their condition prevents continued employment in a recent job or a comparable one. Because of the importance of the concept of disability, the U.S. census long form included a battery of disability questions, and similar questions appear on many surveys in the U.S.

One of our central hypotheses is that the presence of multiple senses of a concept can make classification errors more likely, leading to measurement error. When there are multiple senses of a concept, people may apply the wrong one. In particular, when they are supposed to use an unfamiliar technical sense of the concept, respondents may fall back on their everyday sense of the concept instead. Or, in some cases, they may not realize a concept is being used in a specialized sense and apply their everyday sense in making classification judgments. Both of these problems involve misalignment of different senses of a concept; the definition the respondents apply does not line up with the one they are supposed to use. Such classification errors should become more likely the more that the two senses of the concept differ – that is, the more instances they classify differently. Classification errors due to misalignment between concepts are hardly unique to surveys. Jurors may find it difficult to map their judgments of a case onto the relevant verdict categories (Pennington and Hastie 1992); similarly, clinical psychologists may have trouble aligning their concepts for a mental disorder with the official definitions set forth in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; Kim and Ahn 2002).

A second source of difficulties reflecting the structure of concepts is that people may be unsure how to classify specific instances. When survey designers impose technical meanings on everyday concepts, they often provide definitions or other instructions for making the relevant classifications but people may still have trouble classifying borderline instances that do not clearly meet or fail to meet the definitions. In the case of the 2000 U.S. Census, those who filled out the forms seem confused about whether to report people (such as college students) who have two plausible candidates for their usual residence (their dorms and their family homes). Similarly, people may be unsure whether persons who are unable to work because of drug addiction “really” are disabled. As we noted earlier, one way that technical definitions for a category can differ from everyday concepts

is in imposing explicit boundaries in an effort to deal with the hard-to-classify instances. Here, for instance, is the technical definition of “meats and poultry” used in a U.S. Federal Survey on consumer purchases (from Schober and Conrad 1997); the category includes “beef, lamb, pork, game; organ meats, such as kidney, sweetbreads, chitterlings, heart, tongue; sausages and luncheon meats; poultry, such as chicken, turkey, pheasant, goose, duck; [and] canned ham,” but not “other canned meats and canned poultry, or any prepared meats and poultry.” The definition covers the easy-to-classify instances – like beef and lamb – but mostly it deals with borderline cases – like organ meats (cf. Nosofsky, Palmeri, and McKinley 1994, on central instances vs peripheral instances vs exceptions). When the evidence regarding category membership is mixed under whatever definition of the concept people are applying, that is, when there is a partial or imperfect fit between the concept and the instance, respondents may be uncertain about whether a specific instance belongs in the category or should be excluded.

Surveys sometimes include definitions for key concepts, but there are several reasons why these may not align the concepts people actually use in completing the forms with the ones they are supposed to adopt. In the first place, space is often at a premium on such forms and the definitions provided are likely to be highly abbreviated. For example, the U.S. census residence rules cover some 31 ambiguous situations, but the Census 2000 short form itself mentioned only 11 of these. Second, even if the definitions were adequate in principle, people may not be willing to consult them. The study by Gerber and her colleagues found that instructions about who to count as residents had little effect on how participants classified the people described in a series of vignettes. Similarly, in a study by Conrad and Schober (1999), the participants interacting with a computerized, self-administered questionnaire could have obtained potentially helpful definitions by clicking on the question text but they rarely did so unless they were explicitly told to use that feature (see also Conrad, Couper, Tourangeau, and Peytchev 2006).

Our studies explored three issues. First, we attempted to create misalignment between concepts by varying the definitions we gave respondents for the concepts of residence and disability. One definition corresponded to the everyday sense of the concept, but the other definition imposed a specialized use of the term. Second, we attempted to vary the fit between the concept and the instances respondents were asked to judge. We constructed vignettes that described different living situations or persons with different potentially disabling conditions; the vignettes systematically varied in how well they fit the definitions. Finally, we explored how to improve the effectiveness of definitions in surveys. In particular, we varied whether the target concept was labeled by an unfamiliar term (“enumeration unit”) rather than a more familiar one (“residence”). People are not likely to use definitions if they do not think they need them (Gerber et al. 1996). We thought that the use of unfamiliar terms for the key concepts would increase respondents’ motivation to use the definitions.

2. Pretest: Mapping the Everyday Senses of Concepts

The purpose of our pretest was simply to map out the everyday concepts of residence and disability. We borrowed methods (described below) from Ross and Murphy’s (1999) study of food concepts to investigate our two concepts. These methods are a somewhat more

quantitative alternative to the focus groups and one-on-one interviews often used to investigate concepts in survey settings.

2.1. Method

Sixty participants from the College Park, MD area answered a series of questions about both disability and residence. They were recruited via flyers and advertisements in local newspapers and received \$25 for taking part. A computer program administered the questions on each topic, with the residence questions administered first.

The questions for both concepts included five types of items: 1) An item eliciting the typical features of category members; 2) one asking about major subcategories; 3) several items on inferences that might be drawn from category membership; 4) several vignettes; and 5) items asking about specific living arrangements or potentially disabling conditions.

The initial item on each questionnaire asked the respondent to list typical features of category members. For usual residence, we asked, "For the typical person, what things make a place his or her usual residence?" The next item asked them to list major subtypes ("Are there different types of living arrangements? What are the most common types?"). The next set of items on each concept asked about possible inferences based on category membership. For residence, we asked whether knowing that someone lives at a place indicated anything about how often he or she slept there and ate meals there, and about whether he or she contributed to the rent or mortgage payment, helped with the housework, received mail at the address, and was related to the other people living there. Just after the inference items, we gave participants brief vignettes describing hypothetical persons and situations and asked respondents to rate the extent each the vignette met the requirements for the concept on a five-point scale. Vignettes are a familiar fixture of research on concepts (beginning with Coleman and Kay 1981). The vignettes systematically varied different features of the concept. For example, the residence vignettes varied whether the person depicted was related to others at the residence. Aside from the participants' judgments of the vignettes, the computer program recorded their response times in rating the vignettes. The final set of residence items asked whether most people in a given living arrangement (e.g., a person in a nursing home) would be considered living there; similarly, the final disability items asked whether most people with a given condition (e.g., blindness) would be considered disabled.

2.2. Results

2.2.1. Main Findings

In the open-ended listing of features of a usual residence, the three most commonly mentioned were that the person sleeps or has a bed there (30 participants out of 60, or 50%, listed this), kept his or her clothes and other belongings there (50%), and got telephone, mail, or other messages there (48.3%). These same characteristics showed up in our inference items as well. The four most common inferences based on the fact that a place was someone's usual residence were that he or she slept there most nights or every night (80% of the participants), kept his or her clothes there (91.7%), received mail there (78.3%), and helped pay the rent or mortgage (73.3%). The two living arrangements most

commonly seen as representing the person's primary residence were roommates or housemates (91.7%) or family members living with the owner (86.7%).

In the open-ended listing of features of the typical disabled person, the most common features involved the need for assistance (17 participants out of 60, or 28.3%), blindness or vision problems (23.3%), the need for a wheelchair or other assistive devices (21.7%), and deafness or hearing problems (20.0%). Fewer than 5% of the participants inferred anything about the person's age, race, or sex from the fact he or she had a disability. On the other hand, that someone was disabled did indicate something about his or her abilities – to work outside the home (36.7%), to get around outside (40.0%), to get around inside (31.7%), or to dress and bathe (31.7%). The participants agreed that several conditions are likely to produce a disability. These included deafness (95.0% indicated that most people with deafness are disabled), Alzheimer's (93.3%), paralysis (93.3%), and mental retardation (93.3%). By contrast, the participants were less likely to see various forms of addiction as producing a disability. Only 48.3% of the sample agreed that most alcoholics were disabled; even fewer considered those addicted to prescription (40.0%) and illicit drugs (43.3%) to be disabled.

2.2.2. Vignette Judgments and Reaction Times

We carried out a final set of analyses that examined the relation between the judgment for the vignette and the reaction times. Our hypothesis was that vignettes that described persons or situations that clearly fit the concept (or clearly did not fit it) would generate both extreme ratings and fast reaction times. For both residence and disability, we plotted the average reaction time as a function of the difference between the mean rating of the vignette and the midpoint of the rating scale. For both topics, there was a clearly negative relationship between the reaction time and extremity of the ratings. The correlation was $-.738$ for residence and $-.580$ for disability – $p < .02$ for both. Figure 1 shows the residence data.

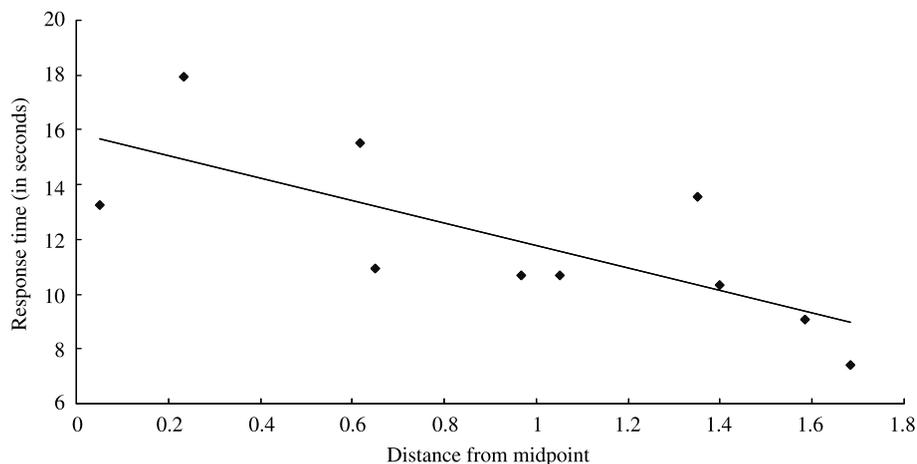


Fig. 1. Mean reaction times in seconds for the residence vignettes as a function of the distance of the mean ratings from the scale midpoint

2.3. Discussion

The residence results suggest there may be two prototypical living situations. One involves a family members living together in their home, where they sleep every night (or nearly every night), keep their possessions, and share expenses. The other involves roommates or housemates living together and sharing the rent or mortgage payments. (The roommate scenario may have been especially salient to our participants, who live in or near a college town.) The very high levels of agreement about these scenarios suggest that there is at least a shared core in the everyday concept of residence. The residence vignettes that were hardest to classify – yielding average ratings closest to the scale midpoint – involved situations in which the person could be linked to two residences (a businessman on an extended trip away from his family) or had no real place to call home (a person looking for a place to live).

The prototypical disability seems to involve a condition, especially a physical condition, that cannot be treated and renders it difficult or impossible for the person to get around or work at a job. The participants showed considerable agreement about the causes and consequences of disability. More than 90% of the participants saw several conditions as rendering most people with the condition disabled; these ranged from deafness and Alzheimer's disease to paralysis. This high degree of consensus again suggests there is at least some level of agreement across respondents in their everyday concept of disability. In addition, participants tended to see disabled persons as having difficulties in working and getting around without help. The eligibility rules for Social Security Disability payments exhibit a similar composite character; claims can be based either on specific medical diagnoses or on evidence that the person is no longer able to pursue his or her previous career (or one that pays as well).

The reaction time data suggest that, for both concepts, category membership is graded – some vignettes describe situations or people that are clearly included in the category, some describe cases that are clearly excluded, and some describe cases on the borderline. The degree of fit between concept and vignette clearly varies across the vignettes, and those that are near the category boundary are the hardest and slowest to classify. Draisma and Dijkstra (2004) also present results suggesting that longer response times signal greater uncertainty about a judgment (see also Bassili and Scott 1996). Our results document a specific type of uncertainty, one that reflects the proximity of an instance to the category boundary.

3. Experiment 1: Definitions and Centrality

Many everyday words have multiple senses. Consider the word *paper*, for example. It can refer to a particular material, the content printed on that material, an institution (such as a newspaper publisher) printing that content, and so on. With such familiar polysemous words, the different senses seem to be stored separately and there is little interference between them (Klein and Murphy 2001; 2002). It could be that different senses of complex concepts like residence and disability are kept distinct in the same way, with the context suppressing the inappropriate senses. However, questionnaires often impose new definitions of everyday terms (rather than using the terms in their familiar conventional senses), and we thought that, in such cases, respondents would have trouble suppressing the everyday meaning of the term. The resulting misalignment between the everyday sense

of the concept and the definition imposed in the questionnaire would lead respondents to make systematic classification errors.

We carried out an experiment to investigate the effects of misalignment; the study also investigated whether our ad hoc, technical versions of the disability and residence concepts would exhibit the same sort of graded membership apparent with naturally-occurring concepts such as furniture. Are more complex social concepts like residence and disability graded in the same way? The results in Figure 1 suggest that they are. We suspected that both concepts would display a gradient of fit between instances and concept, with an imperfect fit leading to a higher rate of classification errors. Our experiment presented participants with definitions of residence and disability that were based either on the everyday concepts identified in our pretest or on technical senses derived from official uses of these terms. We asked participants to use these definitions in classifying the situations described in a series of vignettes. The vignettes varied systematically in how well they fit the definitions.

3.1. Method

Participants first studied a definition of either residence or disability and then, based on that definition, judged whether the person described in each of twelve vignettes should be included in the concept. This process was then repeated for the second concept. We developed two different definitions for the concepts; we refer to them as the *everyday* and *technical* versions, respectively. Each definition consisted of a set of brief instructions about what should be included in or excluded from the category. (Appendix 1 shows the two definitions for residence.) The everyday definition for residence was primarily about the place one considers to be home, where one sleeps, keeps belongings, and contributes to the rent or mortgage. The technical definition was concerned primarily with the address listed on official documents like a driver's license. The everyday sense of the disability concept was concerned primarily with physical, nontreatable conditions; the technical sense, modeled on the definition of the U.S. Social Security Administration, had mostly to do with receipt of payments because of the inability to work.

3.1.1. Design of Vignettes

The vignettes were constructed so that the character or situation it depicted was (1) included under one definition of the concept but not the other, (2) included under both definitions, or (3) included under neither. Although we did not tell the respondents this, we constructed each vignette in such a way that it explicitly met at least one of the inclusion instructions in the definition or at least one of the exclusion instructions; no vignette met both an inclusion instruction and an exclusion instruction for a given definition. Thus, a single match with one of the instructions was sufficient to qualify or disqualify the character for inclusion in the category. For example, in the following vignette, the spouse should be included as "living here" by the everyday residence definition because he or she regards this as home (which is the fifth inclusion criterion in the everyday definition) and excluded under the technical definition because another address appears on his or her driver's license (which is the fifth exclusion criterion in the technical definition of residence).

Your spouse works in a distant city and rents an apartment there. Your spouse's driver's license lists the address in the distant city. On weekends, your spouse comes here and stays with you and your children and regards this as home. Should you list your spouse as living here?

When the vignette described a person who was included by one but not the other definition, the fit to the definition was stronger for some vignettes than for others. A *central* instance matched three of the inclusion instructions in the definition of the concept while a *peripheral* instance matched just one (as in the example above). Here is an example of a vignette depicting a central instance of the technical definition of residence:

Your niece lives in a dorm at her college. However, she uses your address as her official residence. Her driver's license lists your address, and she pays taxes here. Should you list your niece as living here?

Overall, then, each of the vignettes fit the two definitions in one of six ways: 1) It met the everyday definition only and depicted a central instance of that version of the concept; 2) it met the everyday definition only but depicted a peripheral instance of it; 3) it met the technical definition only and depicted a central instance of that version of the concept; 4) it met the technical definition only but depicted a peripheral instance; 5) it met both definitions of the concept equally well; or 6) it met neither definition. We refer to these as the six types of vignette. We created two vignettes of each type, for a total of twelve vignettes for each concept. More specifically, we constructed the different types of vignettes by varying the number of inclusion and exclusion instructions the character depicted in the vignette met for each definition. For instance, the "both" vignettes met two of the inclusion criteria for each definition. Table 1 below displays the combination of features defining each type of vignette; Appendix 2 gives the full set of vignettes.

3.1.2. Participants and Procedure

We recruited 60 participants from the College Park, MD community by placing an advertisement in the local newspaper and posting flyers. They were paid \$25 for the session, which lasted about a half hour.

Participants studied a definition of the concept displayed on the computer screen until they decided they were ready to begin answering questions about the vignettes. The definitions were not available once the task of rating the vignettes began. The vignettes were displayed one at a time on a computer screen and participants pressed one of five

Table 1. Number of inclusion and exclusion instructions met, by definition and type of vignette

Type of vignette	Everyday definition	Technical definition
Both	2 Inclusion	2 Inclusion
Everyday only, central instance	3 Inclusion	1 Exclusion
Everyday only, peripheral instance	1 Inclusion	1 Exclusion
Technical only, central instance	1 Exclusion	3 Inclusion
Technical only, peripheral instance	1 Exclusion	1 Inclusion
Neither	2 Exclusion	2 Exclusion

buttons to indicate how much they agreed or disagreed that the vignette described someone included by the concept. Each scale point was labeled.

We counterbalanced the order in which the two concepts were tested and the assignment of the different senses of the concepts to participants. As a result, there were four combinations of order and definitions: (1) residence–everyday, disability–technical; (2) residence–technical, disability–everyday; (3) disability–everyday, residence–technical; and (4) disability–technical, residence–everyday. Fifteen participants were assigned randomly to each combination. Two participants, who took copious notes on the definitions before rating the vignettes, were dropped from the study.

3.2. Results

3.2.1. Ratings of the Residence Vignettes

Overall, participants' classification judgments reflected the fit between the vignette and the definition (that is, the type of vignette). Average ratings of the vignettes by the two definitions are presented in Figure 2. Participants agreed most strongly that the vignette fit the concept when the vignette met both definitions of residence; they agreed somewhat less strongly for central vignettes (under both definitions); their ratings were fairly neutral for peripheral vignettes (under both definitions); and they rejected vignettes that matched neither definition. The main effect of the type of vignette was highly significant: $F(5, 270) = 76.5; p < .001$. The higher agreement for central than peripheral vignettes suggests that, like other concepts, the concept of residence is graded. It is also possible to treat the middle two panels graphed in Figure 1 as a 2×2 design (central/peripheral examples of the everyday/technical definition). This would entail discarding a third of the data. Instead, we included all six types of vignettes in the analysis, relying on special contrasts to examine relevant differences within the middle two panels.

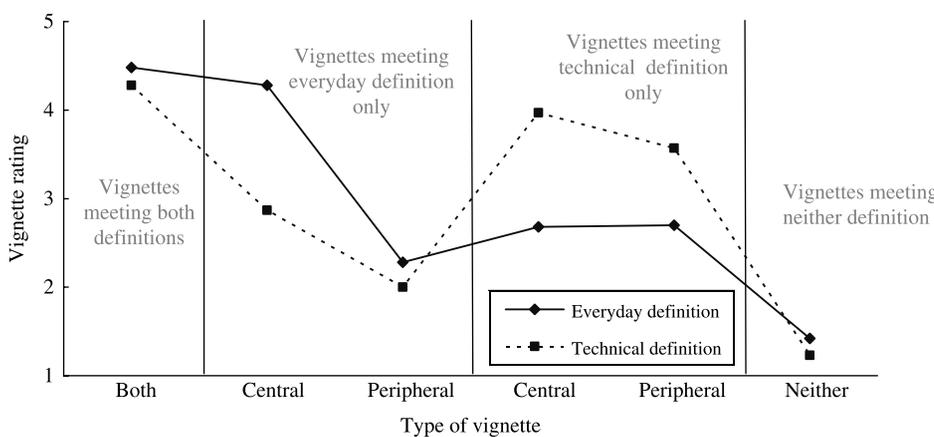


Fig. 2. Average ratings for the residence vignettes, by definition and type of vignette. The different types of vignette described instances that fit both definitions, only the everyday definition, only the technical definition, or neither definition. The vignettes that met only one definition either satisfied three of the inclusion criteria for that definition (central) or one (peripheral). Higher ratings indicate that the vignette depicted an instance of the concept; lower ratings indicate that the vignette did not depict an instance of the concept

In addition, the particular definition of the concept (technical or everyday) that participants received had an overall effect on their ratings. The participants given the technical definition agreed more strongly that the vignettes that met that definition were instances of the concept (see the third panel of Figure 2) and tended to reject the vignettes that met only the everyday definition (see the second panel of the figure). This pattern was essentially reversed when the everyday definition was used. The interaction of type of vignette and definition was significant: $F(5, 270) = 13.5, p < .001$. The overall pattern of ratings for the respondents given a definition modeled on the everyday concept differed considerably from the pattern for the participants who were given the technical definition, suggesting that they were able to separate the different concepts of residence in making these judgments.

As we expected, under both definitions for residence, respondents judged the central vignettes to be better instances of the relevant concept than the peripheral vignettes. A comparison of the ratings for the central and peripheral vignettes that met the everyday definition among those who got that definition of residence was significant: $F(1, 26) = 104.7; p < .001$. Among those who got the technical definition of residence, the contrast between the vignettes depicting central instances and those depicting peripheral instances was somewhat weaker but still marginally significant: $F(1, 28) = 3.32; p < .10$.

3.2.2. Ratings of the Disability Vignettes

Like residence, disability appears to be a graded concept, with some vignettes fitting the definition better than others (see Figure 3). Vignettes that matched both definitions elicited the strongest agreement; those that matched neither elicited the weakest; and those that matched one definition but not the other produced agreement somewhere in between. As with the residence ratings, the main effect of the type of vignette was again highly

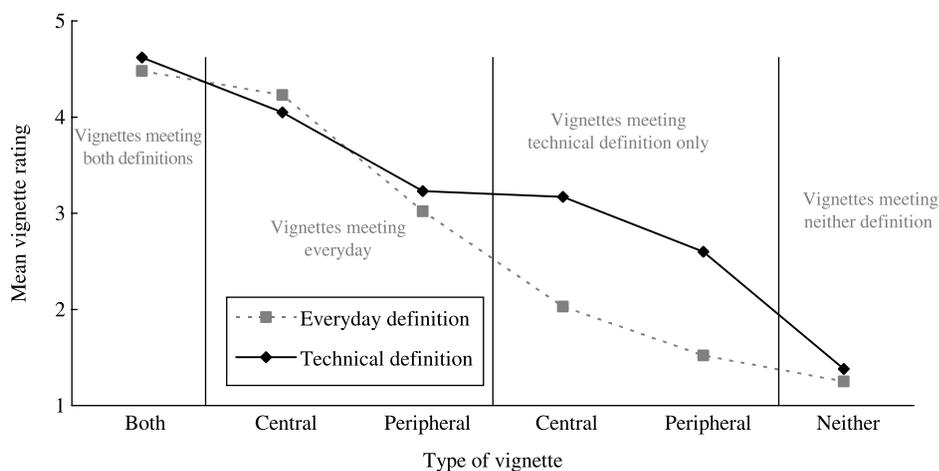


Fig. 3. Average ratings for the disability vignettes, by definition and type of vignette. The different types of vignettes described instances that fit both definitions, only the everyday definition, only the technical definition, or neither definition. The vignettes that met only one definition either satisfied three of the inclusion criteria for that definition (central) or one (peripheral). Higher ratings indicate that the vignette depicted an instance of the concept; lower ratings indicate the vignette did not depict an instance of the concept

significant for disability: $F(5, 270) = 134.0$; $p < .001$. Moreover, central vignettes produced stronger agreement that the vignettes should be included in the category than the corresponding peripheral vignettes, though agreement was generally stronger for the vignettes that met the everyday definition than for those that met the technical definition. The contrast between the central and peripheral vignettes that met the everyday definition (see the second panel of Figure 3) was significant for the respondents who got that definition: $F(1, 28) = 27.8$; $p < .001$. Similarly, the contrast between the central and peripheral vignettes meeting the technical definition was significant for those who got the technical definition (see the third panel of the figure): $F(1, 26) = 6.51$; $p < .05$.

The ratings of the disability vignettes were affected by the definition participants received, as was the case for residence; again, the interaction of vignette type and definition was significant: $F(5, 270) = 5.90$; $p < .001$. But the pattern (displayed in Figure 3) differed sharply from the pattern for residence. The definition participants received for the disability vignettes had much less effect on their agreement ratings than the residence definitions did (compare Figures 2 and 3). In Figure 2, the patterns in the second and third panels are mirror images; in Figure 3, the pattern in those panels is nearly parallel. In fact, the groups getting the different definitions of disability showed significant differences in their ratings of the vignettes only for the vignettes that met only the technical definition (those in the third panel of Figure 3). For those vignettes, the participants who got the technical definition were more likely than those who got the everyday definition to agree that the character described in the vignette was disabled. Overall, the vignette ratings suggest that participants who were asked to use the technical definition nonetheless leaned heavily on their everyday sense of the disability concept in making their judgments.

3.3. Discussion

Although respondents' ratings of the vignettes were affected when we gave them technical definitions of the concepts, our evidence suggested that they found it difficult to ignore the everyday senses of the concepts in making their classification decisions. For example, we observed stronger agreement that the vignettes that met both definitions of the concept should be included in the concept than those designed as central instances of the definition they actually received, even though the latter vignettes matched more of the inclusion criteria (three versus two).

With disability, a preference for the everyday sense of the concept is quite apparent. The participants who got the technical definition of disability hardly seemed to use it in making their judgments. We reclassified the ratings for the vignettes according to whether the vignette actually met the definition of the concept the participant had gotten. For vignettes that met the definition, we counted "Definitely yes" and "Probably yes" ratings as correct; similarly, for vignettes that did not meet the definition, we counted "Probably not" and "Definitely not" ratings as correct. (We also counted all responses at the scale midpoint – "Neither" – as correct.) Despite this liberal scoring, participants who got the technical definition of the disability concept were correct a little less than half of the time in their classification of the disability vignettes (48.6%). By contrast, those who got the everyday definition classified 73.5% of the disability vignettes correctly. A logit analysis indicates

that the difference is significant ($\chi^2 = 7.01$, $df = 1$, $p < .01$). There is no corresponding difference in accuracy in the ratings of the residence vignettes. For disability at least, there is clear evidence for misalignment between the technical definition of the concept presented to the respondents and the concept they actually applied – with large consequences for the accuracy of the answers.

4. Experiment 2: Definitions and Terminology

Experiment 1 demonstrated some interference between the everyday senses of residence and disability and the technical definitions respondents were supposed to use, particularly for the latter concept. Still, that experiment left a number of issues open. The results differed somewhat for the two concepts, and we wondered whether these differences would replicate. We did not let respondents review the definition they were supposed to apply after the initial study phase and thus they may have been forced back on their everyday sense of the concept as their memory for the definition faded. The vignettes in Experiment 1 differed considerably from one another in length, preventing us from comparing reaction times for the different types of vignette in a meaningful way. Furthermore we created only two vignettes of each type and their peculiarities may have affected the results. We therefore decided to replicate our first experiment, using four vignettes of each type and allowing the respondents to consult the definitions as they rated each vignette. In addition, we rewrote the vignettes so that their lengths were about equal, facilitating analyses of the reaction time data.

We added two other refinements to the design. First, we added a group of respondents who did not get any definition at all. Including this group allows us to test our assumption that the everyday definitions we created actually reflect the respondents' preexisting concepts. Second, we gave some respondents vignettes that used unfamiliar terms for each concept ("enumeration unit" in place of "residence" and "designated impairment" in place of "disability"). If respondents pay little attention to the definitions because they think they already understand the relevant concepts, the unfamiliar terminology may improve the accuracy of their judgments by giving them a reason to attend carefully to the definitions.

We also examined one other variable. For half of the respondents who got definitions, the definition included a general statement summarizing the more specific instructions. For example, the summary for residence stated that an enumeration unit (or residence) "is the place where a person stays most of the time." We hoped that including a summary would make it easier for respondents to apply the definitions, but this variable had little apparent effect on the results and we discuss it only briefly below.

4.1. Method

Respondents to a web survey classified the characters described in vignettes, using the same five-point scale as in Experiment 1. Each respondent received a definition and vignettes for only one of the concepts. The definition was displayed on a web page for the respondent to study. After they had studied the definition, the respondents received 12 of the 24 vignettes for that concept. By clicking on the highlighted term in each vignette (e.g., "enumeration unit"), respondents could return to the definition page and review the definition they had received. About one respondent in five took advantage of this feature.

4.1.1. Experimental Design

We varied three characteristics of the definitions for each concept. As in the previous experiment, approximately half the respondents who got a definition got one designed to mirror the everyday sense of the concept; the remainder got a definition that deviated markedly from the everyday notion. We again refer to the two definitions as the *everyday* and *technical* definitions for the concepts. The two definitions for each concept were very similar to those used in Experiment 1; each one listed four grounds for including a potential instance and four grounds for excluding one. We also varied the terms used to convey these concepts. In the *familiar terminology* group, the concepts were labeled “disability” and “residence” in the definitions and those terms were used in the vignettes (e.g., each disability vignette ended by asking whether the person depicted “had a disability”). In the *unfamiliar terminology* group, we replaced these familiar terms with the terms “designated impairment” and “enumeration unit” in the definition and vignettes. Finally, the definitions varied in whether or not they included an overall summary. For instance, the everyday definition of residence included this summary: “A residence is the location where a person stays most of the time” as the first part of the definition. Crossing these three variables yielded a total of eight definition conditions for each concept. We added a ninth group that did not receive any definition for the concept at all. Respondents were randomly assigned to a concept (either residence or disability) and to one of the nine definition conditions for that concept.

As in Experiment 1, we created vignettes in which the characters depicted met either both definitions for the concept, neither definition, or one definition but not the other. In addition, central vignettes matched three of the inclusion criteria mentioned in the relevant definition; peripheral vignettes met just one. (The vignettes that qualified under both definitions met two of the inclusion criteria for each one; similarly, the vignettes that met neither definition matched two of the exclusion criteria for each definition.) Thus, the six types of vignettes were defined in terms of the same number of inclusion and exclusion criteria shown in Table 1. We created a total of 24 vignettes for each concept, four of each type. Each vignette was between 49 and 63 words in length. To reduce the time needed to complete the vignette portion of the survey, participants classified instances of only one concept (either residence or disability). In addition, we subdivided the 24 vignettes for each concept into two nonoverlapping sets of twelve (Set A and Set B) by randomly selecting two vignettes from each of the six types of vignette; respondents got the vignettes in just one of the sets. Each respondent got the twelve vignettes in a different random order.

4.1.2. Sample and Procedure

Market Strategies, a web survey firm, conducted the study, and a commercial vendor provided the sample of respondents from two sources – an “opt-in” web panel of almost one million persons who have signed up on-line to receive survey invitations and a second list of millions of e-mail addresses from opt-in lists from a wide variety of web sites. A total of 39,217 invitations to participate were sent via e-mail; nonresponders received a second, reminder e-mail. The e-mail provided the URL and a unique PIN for the survey (to prevent respondents from completing the survey more than once). Overall, 3,179 persons

began the survey, 2,722 getting all the way through the questionnaire. Our analysis focuses on the complete respondents.

The sample respondents were a heterogeneous group of web users, ranging in age from under 18 to over 65. Of those who completed the survey, 22.4% were men and 77.6% were women; 3.5% were Hispanic, 5.0% were African-American, 89.1% were White, and 5.9% were Asian, Native American, or some other race; 28.8% had a high school diploma or less education, 45.4% had some college, and 25.8% had graduated from college. This overrepresentation of women, Whites, and the highly educated is characteristic of web survey samples (Couper 2000). The larger, more heterogeneous sample allows us to assess the generality of the results from Experiment 1, a much smaller study done in a laboratory setting.

The questions were administered in a web survey, and both the answers and response time data were collected. The program administering the survey also recorded whether the respondent clicked on a term in order to see the definition again. The questionnaire included a number of items on health, diet, and other lifestyle issues, unrelated to the present research. The vignettes began at the 25th question on the survey, about halfway through the questionnaire. Each asked the respondent to rate the degree that the character in the vignette should be included in the concept (e.g., “Does Josh have a disability/designated impairment?”) on a five-point scale ranging from “Definitely yes” (= 1) to “Definitely no” (= 5).

4.2. Results

We examined the ratings of each vignette, the percentage of correct judgments, and response times. The response times encompass the time from the display of the item to the recording of the response and, thus, include download and reading times. In addition, we explored when respondents clicked on the key term in order to review the definition and how clicking affected their answers. For each concept, each set of vignettes represents a complete replication of the experiment, with different respondents rating different vignettes. In analyzing the results, we examined the data from each set of vignettes separately and report those findings that were at least marginally significant in both sets. Most of the analyses were 2 (everyday vs technical definition) \times 2 (familiar vs unfamiliar terminology) \times 6 (type of vignette) ANOVAs or logit analyses, with the type of vignette treated as a repeated measures factor. The levels of the type-of-vignette variable depended on whether the vignette met both definitions of the concept, only the everyday definition, only the technical definition, or neither definition; for the vignettes that met only one definition, we distinguished vignettes that depicted central instances of that definition from those that depicted peripheral instances.

4.2.1. Ratings of Vignettes

Figure 4 displays the average ratings of the vignettes by definition and type for each concept. The means for each type of vignette average across both sets of respondents and vignettes. Higher numbers again indicate that the respondents thought the person described in the vignette more clearly fit the concept.

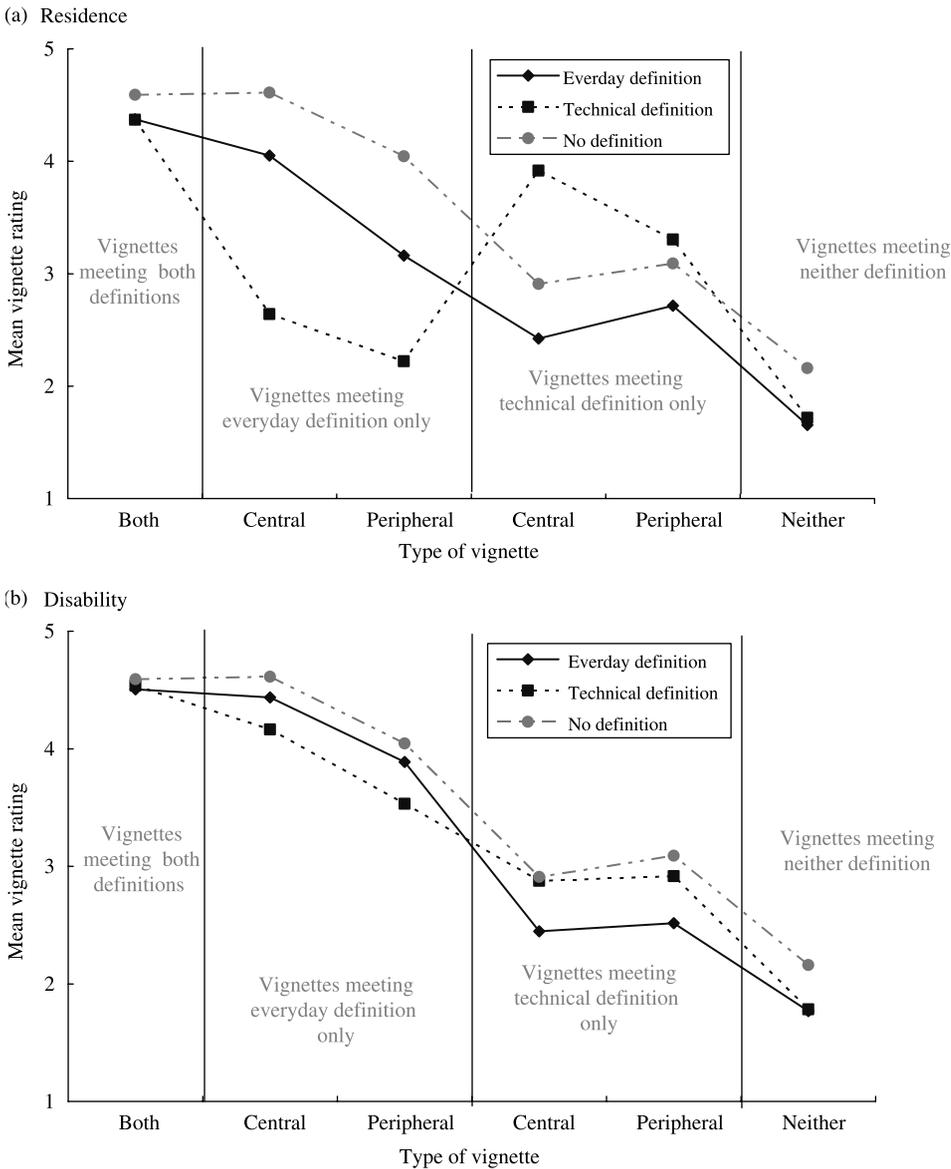


Fig. 4. Average ratings for residence vignettes (top graph) and disability vignettes (bottom graph), by definition and type of vignette. The different types of vignette described instances that fit both definitions, only the everyday definition, only the technical definition, or neither definition. The vignettes that met only one definition either satisfied three of the inclusion criteria for that definition (central) or one (peripheral). Lower ratings indicate that the vignette depicted an instance of the concept; higher ratings indicate that the vignette did not depict an instance of the concept

As in Experiment 1, for residence, there are large differences in the ratings by the type of vignette as well as a marked type by definition interaction. The type main effect is significant for both sets of vignettes $-F(5, 586) = 693.5 (p < .001)$ for Set A, and $F(5, 600) = 536.7 (p < .001)$ for Set B – as is the vignette type by definition (everyday vs technical) interaction $-F(5, 586) = 85.7 (p < .001)$ for Set A, and $F(5, 600) = 89.6$

($p < .001$) for Set B. The ratings for the group that got no definition at all clearly parallel those for the group that got the everyday definition (though the ratings of the no definition group are, on average, somewhat lower – that is, the no definition respondents were more likely to count the person depicted in the vignette as an instance of the relevant concept), and the pattern by type of vignette for both of those definition groups differs sharply from the pattern for the group that got the technical definition of residence.

The disability ratings results also parallel those from Experiment 1 – the two groups that got definitions have similar profiles across the six types of vignettes. The main effect of vignette type is highly significant for both sets of disability vignettes (both F 's > 500 and both p 's $< .001$); despite the similarities of the profiles for the two definition groups, the type of vignette by definition interaction was significant: $F(5, 621) = 9.87$ ($p < .001$) for Set A, and $F(5, 592) = 26.6$ ($p < .001$) for Set B. In addition, the respondents in both groups that got definitions gave ratings similar to those of the respondents that did not get any definition for the concept; this finding suggests that members of all three groups relied at least to some extent on their everyday definitions in rating the disability vignettes, just as the respondents in Experiment 1 did.

We had expected that the unfamiliar terms would increase respondents' use of the definitions we provided (and reduce their reliance on their preexisting concepts). There was some evidence for this with the residence vignettes; the three-way interaction of vignette type by definition by terminology was marginally significant for Set A ($F(5, 590) = 2.19$, $p < .06$) and significant for Set B ($F(5, 600) = 3.37$, $p < .01$). The differences are small, though, and do not produce significant differences in accuracy classifying the vignettes.

4.2.2. Proportion Classifying Vignettes Correctly

As in Experiment 1, we scored the ratings of the vignettes correct or incorrect according to the definition of the concept the respondent had received. For vignettes that met the definition, we counted “definitely yes” and “probably yes” responses as correct; similarly, for vignettes that should have been excluded, we counted “definitely no” and “probably no” responses as correct. (Once again, we used a liberal scoring system that counted responses at the midpoint as correct.)

Figure 5 displays the major results from this analysis – large differences for both concepts in the percent correct by type of vignette and definition. We carried out logit analyses examining the proportion of respondents who correctly classified both of the vignettes they rated of each type. (We also analyzed the proportion who classified neither vignette of a given type correctly, and the results are quite similar to those we present here.) These analyses show that for both concepts the main effect of the type of vignette and the vignette type by definition (everyday vs technical) interactions are significant (p 's $< .01$ for both sets of vignettes). Contrasts indicated that for both residence and disability respondents were more accurate at classifying the vignettes that met both definitions (the “Both” vignettes) or neither definition (the “Neither” vignettes) than the other four types of vignettes (all p 's $< .001$). On the average, respondents classified 92.4% of the Both and Neither residence vignettes correctly and 90.9% of the Both and Neither disability vignettes; the corresponding figures for the other four types of vignettes were

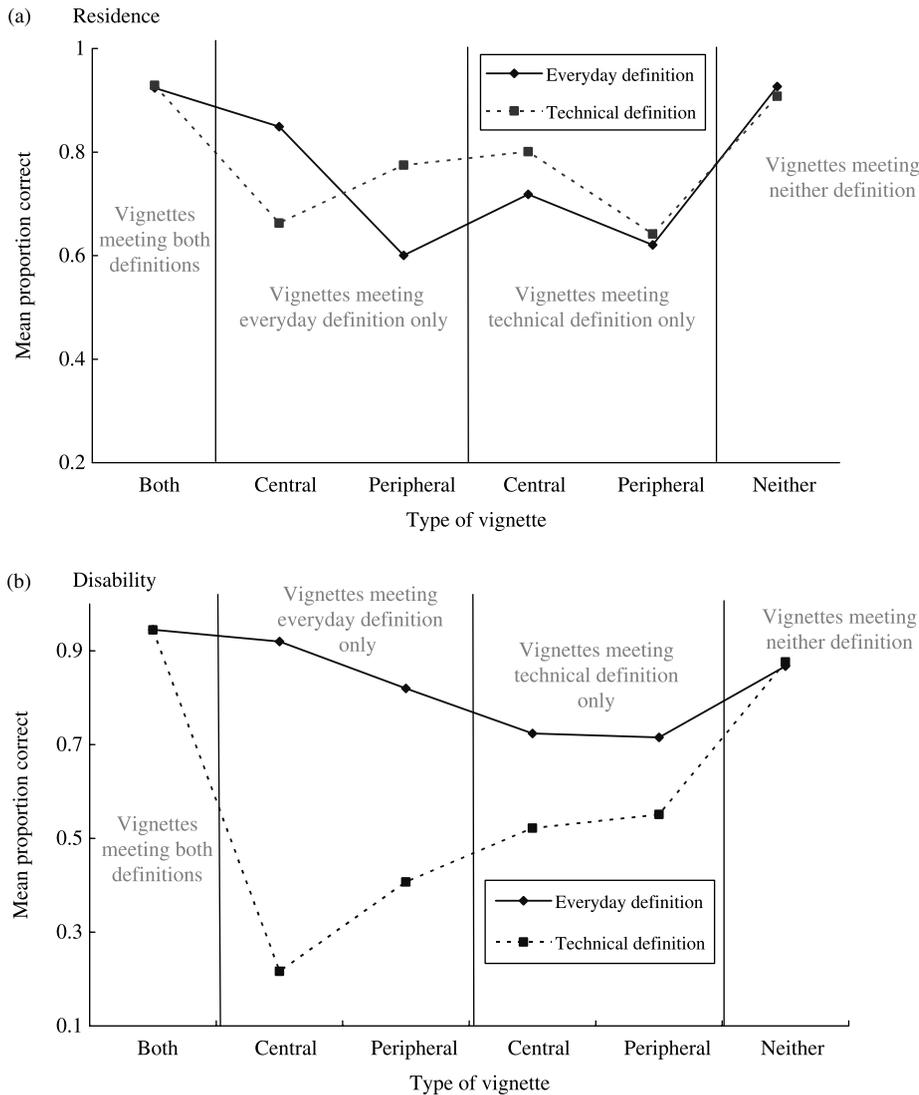


Fig. 5. Proportion correctly classifying residence (top graph) and disability vignettes (bottom graph), by definition and type of vignette. The different types of vignette described instances that fit both definitions, only the everyday definition, only the technical definition, or neither definition. The vignettes that met only one definition either satisfied three of the inclusion criteria for that definition (central) or one (peripheral)

71.4% for residence and 61.5% for disability. The Both and Neither vignettes have the advantage that they are classified the same way under either definition.

In addition, for both concepts, respondents who got the everyday definitions were more accurate with the vignettes describing central instances of that definition than with those describing peripheral instances of that definition. Respondents who got the everyday definition of residence classified 84.9% of the vignettes depicting central instances of that definition correctly, but only 60.0% of those depicting peripheral instances; similarly, for disability, those who got the everyday definition correctly classified 92.0% of the central

vignettes meeting that definition but only 82.0% of the peripheral vignettes. Contrasts confirming this pattern were significant for both residence and disability (all p 's < .01). The respondents who got the technical definition for residence show a corresponding increase in accuracy for the vignettes depicting central instances of that definition over those depicting peripheral instances of the technical definition (80.1% correct vs 60.0%). For the disability vignettes, there is no advantage within the group that got the technical definition of disability for the vignettes depicting central instances of that definition over the vignettes depicting peripheral instances. This seems to reflect their reliance on the everyday concept in making their judgments. Whether a vignette depicted a central or peripheral instance of the technical definition of disability, respondents who got that definition tended to classify these vignettes at chance levels of accuracy.

The other striking effect in Figure 5 is the sharply reduced performance for the respondents who got the technical definition of disability (see the bottom panel of Figure 5); overall, these respondents classified only 58.6% of the vignettes correctly (versus 83.1% for those who got the everyday definition). The logit results confirm that there is a significant main effect for definition for both sets of disability vignettes ($\chi^2 = 209.0$, $df = 1$, $p < .001$ for Set A; $\chi^2 = 139.8$, $df = 1$, $p < .001$ for Set B). The respondents in Experiment 1 had similar difficulty in applying the technical definition of disability.

4.2.3. Response Times

For both residence and disability, the response time data (not shown) exhibited large differences by type of vignette (all p 's < .001). Response times for both concepts were significantly faster for the Neither vignettes than for the vignettes of the other types (all p 's < .001); in addition, for disability, respondents rated the vignettes that met both definitions (Both) about as quickly as those that met neither definition (Neither).

For the residence response times, there also was a main effect of the definition respondents received ($F(5, 580) = 28.0$, $p < .001$ for Set A; $F(5, 591) = 17.4$, $p < .001$ for Set B), as well as a vignette type by definition interaction ($F(5, 580) = 4.47$, $p < .001$ for Set A; $F(5, 591) = 2.64$, $p < .08$ for Set B). Overall, the respondents who got the technical definition of residence took an average of two seconds longer to rate the vignettes than those who got the everyday definition (27.5 vs 25.4 seconds). (These means are based on both correct and incorrect judgments.) The respondents who got the technical definition were especially slow in rating the vignettes depicting central instances of the everyday definition (taking an average of 29.5 seconds); these vignettes, presumably, created the most conflict for them, since these respondents were supposed to disregard their everyday notion of residence in evaluating the vignettes and the characters depicted in these vignettes clearly fit the everyday definition. Their judgments of these vignettes were also often incorrect (about 34% of the time), further supporting the idea that respondents had trouble ignoring the match between the vignette and their everyday understanding of residence.

We also carried out an additional set of analyses relating the extremity of the ratings for the vignette and the average response time, replicating the analyses of the pretest vignettes presented in Figure 1. Once again, for both concepts, the difference between the mean rating for the vignette and the scale midpoint was negatively related to the mean response

time. The correlation between the two variables was $-.362$ for the residence vignettes ($p < .10$) and $-.758$ for disability ($p < .001$), both based on 24 vignettes. Across both studies, the closer the fit between an instance and a concept, the faster (and easier) it is to make the classification judgment.

4.2.4. Clicking for the Definition

One difference between this experiment and the previous one was that the respondents had access to the definition while they rated the vignettes. To review the definitions, respondents merely had to click on the highlighted term. We analyzed the proportion of respondents who clicked to access the definition as a function of the definition and terminology variables. Because relatively few respondents clicked – 20.1% of those who got the residence vignettes and 19.5% of those who got the disability vignettes – the data were too sparse to examine clicking by type of vignette.

The terminology used in the definition and vignettes had a consistent effect on the proportion of respondents who clicked to access the definition. The respondents who got the definition and vignettes about “enumeration units” were more likely to click for a definition than those who got the definition and vignettes about “residence” (23.4% across all 24 vignettes vs 15.6%; $p < .05$ for both sets); similarly, the respondents asked about “designated impairments” were more likely to click for the definition than those asked about “disability” (27.7% across all 24 vignettes vs 12.5%; $p < .001$ for both sets).

We carried out some exploratory analyses on the effects of clicking to access the definition. For both concepts, those who clicked for the definition took longer on average to rate the vignettes and were more likely to classify them correctly. For residence, those who clicked took an additional 47.4 seconds on average to rate the twelve vignettes; for disability, those who clicked took 45.2 more seconds to make their ratings. (We reran the main response time analyses excluding those who clicked to review the definitions; the main findings regarding the effects of vignette type, definition, and the interaction between the two are unchanged.) The data do not allow us to tell whether the respondents used all this time to examine the definition or were struggling to make their judgments and decided to consult the definition as a result. Clicking for the definition seemed to improve performance in the rating task. Those who clicked for a definition correctly classified 85.5% of the residence vignettes versus 75.4% among those who did not click. Similarly, those who clicked for the definition of disability classified 74.0% of the disability vignettes correctly compared to 70.3% among those who did not click. For both concepts and both sets of vignettes, this represented a significant improvement (all p 's $< .001$).

4.3. Discussion

Even when they were free to reexamine the definitions, respondents showed some difficulty in applying them. With disability, the definition respondents got had relatively little effect on their ratings of the vignettes (bottom graph of Figure 4); instead, they seemed to rely on their everyday sense of disability regardless of which definition we gave them. As a result, those who were supposed to be applying a more technical definition only classified 58.6% of the vignettes correctly (see the bottom graph of Figure 5), even though we used a rather liberal criterion in scoring the judgments. Respondents who seemed to

pay more attention to the definitions (those who clicked to review the definition after they had begun rating the vignettes) classified them more accurately. The same sort of interference from the everyday concept that we observed for the disability judgments was also apparent for the residence data, though it manifested itself in somewhat subtler ways. The respondents who got the technical definition for residence were much slower and less accurate (top panel of Figure 5) in assessing the vignettes that described central instances under the everyday definition. Whichever definition respondents got, the fastest and easiest vignettes were those that clearly met both definitions and those that clearly did not meet either.

5. General Discussion

The results of our two experiments indicate that respondents have difficulty classifying instances when they are supposed to apply an unfamiliar technical sense of the concept and when the instances are borderline cases. In our terms, both misalignment between the intended and everyday senses of a concept and poor fit between the concept and the instances produced serious classification errors.

5.1. The Effect of Misalignment

In both of our experiments, respondents had difficulty applying the technical definition for disability. In Experiment 1, respondents who got the technical definition classified only 49% of the vignettes correctly; by contrast, respondents given a definition modeled on their everyday sense of the concept classified more than 73% of the vignettes correctly. Similarly, in Experiment 2, respondents who got the technical definition classified only 59% of the disability vignettes correctly versus 83% for those who got the everyday definition. These differences in accuracy reflect misalignment – the concept respondents applied in judging the vignettes seemed to reflect their everyday notion of disability rather than the technical version of the concept that they were supposed to use.

One possible explanation for the disability results in the two experiments is that respondents did not pay enough attention to the definitions to realize that their everyday concept differed from the one they were supposed to apply. More generally, our classification judgments may be wrong because we do not understand the intended concept but think that we do (see Conrad and Schober 2000; Schober and Conrad 1997; Schober, Conrad, and Fricker, 2004; and Suessbrick, Schober, and Conrad 2000). With the disability concept, our respondents did not seem to pay much attention to the definitions in either experiment – the ratings of the disability vignettes are quite similar regardless of which definition we gave them (see Figure 3 and the bottom panel of Figure 4). The participants in our experiments may have been disinclined to pay much attention to the definitions since they thought they already knew what the concept was; after all, both residence and disability are familiar everyday concepts. Another possibility is that they may have rejected the technical definition. Something like that appears to have happened in Census 2000 with Hispanics and the race question on the census questionnaire. Many Hispanics appeared to reject the official six-category system used in federal surveys in the U.S. since 1997; that system treats Hispanic origin as a separate question from race. In the 2000 U.S. Census, many Hispanics either left the race question blank or reported their race

as Hispanic (del Pinal, Bennett, Cresce, and Martin 2002). Still, another possibility is that our definitions were just too complex for respondents to absorb and apply easily. As is apparent from Appendix 1, the definitions listed a number of criteria and in some instances the criteria may have conflicted with each other. Still, we believe that our definitions are no more complicated than those that underlie the concepts in many U.S. surveys. (Compare our definitions with the definition of meat and poultry cited earlier or with the residence rules used in the 2000 Census. These are available at www.census.gov/population/www/-censusdata/resid_rules.html.) Whatever the exact reasons for it, misalignment between the intended concept and the ones respondents actually applied led to high levels of error in our experiments.

5.2. *The Effect of Imperfect Fit*

Our experiments also demonstrated that the fit between instances and concepts affected the accuracy of respondents' judgments. In Experiment 1, most of the judgments of the peripheral instances were incorrect – overall, about 55% of the respondents misclassified the peripheral vignettes. (Given our scoring criteria, the respondents would have done better had their ratings been completely random.) In Experiment 2, in which respondents could review the definition, they did somewhat better, but still classified about a third of the vignettes depicting peripheral instances incorrectly. The peripheral instances were the ones we designed to have an imperfect fit to the definition so they were neither clearly included in nor clearly excluded from the concept. The response time results give a similar picture; in both our pretest (see Figure 1) and Experiment 2, respondents were slowest to classify instances that fell near the borderline of the concept.

It has been known for almost 30 years that some instances of natural categories are harder to classify than others (Rosch and Mervis 1975; Smith, Shoben, and Rips 1974). Our studies demonstrate this is also true for the socially defined concepts that appear in surveys. We suspect that respondents were generally unsure about whether the peripheral vignettes cases met the definitions we provided, because some of their features suggested that they were instances of the concept and some suggested that they were not. The easiest cases were those that met both definitions of the concept and those that met neither definition. With these vignettes, every feature pointed to the same answer and it did not matter which definition respondents were trying to apply.

5.3. *Differences Across Concepts*

In both our experiments, the results differed somewhat for the two concepts. Respondents seemed to have an easier time applying the technical definition of residence than the technical definition of disability. We suspect this is because the concept of an official or permanent address is already one sense of the everyday concept of residence. People may have more practice in thinking about their residence in several senses; in interacting with government agencies, they certainly acquire considerable experience in listing addresses as their residence for various official purposes. Thus, participants' intuitions may still have provided a reasonably accurate basis for their residence judgments even when they got the technical definition. Most people probably do not have comparable experience in using the

disability concept for official purposes. We used real-world concepts in our experiments and it would not be surprising if they brought some real-world baggage with them.

As an operational matter, we constructed vignettes that varied quantitatively in their fit to the definitions we provided (see Table 1), but our results do not shed much light on the theoretical question of how everyday concepts are represented cognitively. The results in Figure 4 – in particular, the similarity between the judgments of those who got the everyday definition and those who got no definition at all – suggest that our definition of the everyday concepts must have been reasonably consistent with the *content* of participants' underlying representations of them, but not necessarily with their format. It is possible that the conceptual cores of everyday categories are represented as sets of inclusion and exclusion rules like the ones we used (a view consistent with the model of Nosofsky et al. 1994) or that conceptual cores may take a different format whose essence can nonetheless be captured by such rules.

5.4. Practical Implications

Our results raise the practical question of how to get people to use definitions when their everyday sense of a concept does not apply to the task at hand. In our first experiment, we took the definitions away from the participants before they made their judgments about the vignettes, but other evidence suggests that people may not look carefully at definitions even when they are accessible (Conrad and Schober 1999; Conrad et al. 2006; Lind et al. 1999). The length and complexity of our definitions may have reduced the likelihood that the participants would apply them correctly. Each definition included 8 to 10 rules and the principle behind those rules may not have been easy to discern. Participants are generally better at learning a category when the category members follow some overall theme or principle, provided that there are not too many exceptions (Murphy and Kaplan 2000; see also Nosofsky et al. 1994). This implies that a simpler definition consisting of a few rules plus an integrating theme might have had more effect on participants' classification judgments. Although our attempt to provide such an integrating theme in Experiment 2 had little effect on respondents' judgments, we think this is a promising avenue for future research.

Another strategy for getting participants to pay attention to definitions is to increase their motivation to look at them. If participants do not bother to look at the definitions because they do not think they need to, then it may help to point out that the current use of the concept differs from its everyday use. In Experiment 2, we used unfamiliar terms in the hope of getting respondents to examine the definitions more carefully. There was some evidence that this worked; respondents who got the term "enumeration unit" rather than "residence" were more accurate in classifying the residence vignettes and more likely to review the definitions.

Still, there is clearly room for improvement in our participants' use of technical senses for everyday concepts. In our final study, the participants who did not get any definition did nearly as well overall as those who got the everyday definitions. If we score their judgments according to the same criteria we used for the respondents who got the everyday definitions, they correctly classified 80% of the residence vignettes (versus 83% for those who got the everyday definition) and 71% of the disability vignettes (versus 76% for those who got the

everyday definition for disability). Actually providing the respondents with a definition did not improve their performances very dramatically. Unless the concept endorsed by the survey sponsors differs from respondents' everyday understanding, definitions are unlikely to be important. However, when such differences do exist, it may take an interviewer to offer effective clarification of the intended meaning of the question (see Schober and Conrad 1997). Clearly, respondents prefer to rely on their preexisting senses of concepts rather than applying the new definitions imposed in survey questionnaires.

Appendix 1. Everyday and Technical Definitions for Residence

Technical Concept of Residence

Inclusions

- 1R** Count all people who list this address as their official residence.
- 2R** Count people who are registered to vote here.
- 3R** Count people whose driver's license lists this address.
- 4R** Count people who receive utility bills at this address.
- 5R** Count people who pay their taxes here.

Exclusions

- 6R** Do not count people who stay here, but are registered to vote elsewhere.
- 7R** Do not count people who pay taxes in other states.
- 8R** Do not count people who are only staying here temporarily.
- 9R** Do not count roomers or boarders if they have another place that is their official residence.
- 10R** Do not count people who stay here but who list another address on their driver's license.

Everyday Concept of Residence

Inclusions

- 1R** Count all people who sleep here most of the time.
- 2R** Count all other relatives who have no other residence.
- 3R** Count other people who stay here and contribute to the rent or mortgage payment.
- 4R** Count people who receive their mail here and not elsewhere.
- 5R** Count all people who stay here frequently and think of this as home.

Exclusions

- 6R** Do not count children who visit once or twice a week but stay most of the time with their other parent.
- 7R** Do not count relatives who live most or all of the time in a nursing home or other institution.
- 8R** Do not count students who live elsewhere while attending school.
- 9R** Do not count people who are visiting.

Appendix 2: Vignettes*Experiment 1: Disability Vignettes*

John is almost blind even when he wears corrective lenses. Because of modifications to his home and workplace, he is able to do his job and carry out other activities of daily living. Should you list John as disabled?

Susan lost her foot as a result of diabetes. Although she wears an artificial foot, she has difficulties getting around and was forced to switch her career from a school bus driver to a technical writer. Should you list Susan as disabled?

Last month, David received a severe head injury in a truck accident. Despite treatment he often sees aliens and has other hallucinations. He is unable to return to work, and the doctor does not expect that he will recover any time soon. Should you list David as disabled?

Kevin has suffered from bipolar disorder for many years and has not responded to treatment. Last year it was discovered that he has glaucoma and is now legally blind. Although he lost his job at a local gas station, he has been able to support himself by playing music with a surprisingly successful group. Should you list Kevin as disabled?

Jane has severe arthritis and hears voices. She is not able to wash herself and receives weekly payments for her treatment from the federal government. Should you list Jane as disabled?

Mike has suffered from Parkinson's Disease for the last ten months and now is losing his eyesight. Because of his condition, he finds it increasingly difficult to get around. He recently received a notification from the Department of Motor Vehicles saying his driver's license was revoked, due to his condition. Should you list Mike as disabled?

Jill works at NASA. Three years ago, she was accidentally exposed to dangerous chemicals that made her very sensitive to light. She currently receives worker's compensation payments and will remain eligible even if she gets a new job. Recently she has found a special pair of glasses that have enabled her to return to work. Should you list Jill as disabled?

Bill broke his leg while working as a roofer. His company pays for the therapy and medical expenses. The doctor says he will eventually recover. Should you list Bill as disabled?

Adam's medical record classifies him as having attention deficit disorder and multiple sclerosis, and he is unable to take part in extracurricular activities after school, due to this condition. Although he has multiple sclerosis, he has shown only minor symptoms so far. The insurance company pays for his medicine, which alleviates his symptoms. Should you list Adam as disabled?

Ana acquired serious asthma while working for an auto body shop where she sprayed paint. She now uses an inhaler as her doctor recommended and rarely shows symptoms. The doctor sent the company her medical records so that they can pay for her treatment. Should you list Ana as disabled?

Because of an allergic reaction, George has come down with a serious but treatable case of acne, which he says makes it impossible for him to work. According to his doctor, this condition will last only for a few weeks. His employment counselor assures him that he

has the needed skills to work. Although he has had several job opportunities, he has declined to take them, because he was not interested in working. Should you list George as disabled?

Molly broke her arm on a skiing trip. She was forced to stay out of work for a few weeks and to use her vacation time while she was recovering. Should you list Molly as disabled?

Experiment 1: Residence Vignettes

Your spouse works in a distant city and rents an apartment there. Your spouse's driver's license lists the address in the distant city. On weekends, your spouse comes here and stays with you and your children and regards this as home. Should you list your spouse as living here?

Your cousin is looking for a job and a new home in your city. For two weeks, your cousin stays with you. Should you list your cousin as living here?

You and your roommate share a house. You both sleep there most nights and split the rent. Both of you are registered to vote elsewhere and regard this as home. Should you list your roommate as living here?

Your son is away for a week with your ex-spouse. He sleeps most nights in your home, and he considers your house to be his home. He receives his subscription to *Time Magazine* at your address. He just registered to vote at your ex-spouse's home. Should you list your son as living here?

Your brother uses your address as his voting address, and his driver's license lists this address. He sleeps here most of the time, and he contributes to the rent. Should you list your brother as living here?

Your son is in the military, and although his wife has her own apartment, she sleeps here most nights and pays her taxes here. She contributes regularly towards the rent and votes here. Should you list your daughter-in-law as living here?

Your daughter pays her taxes from this address, but she is currently a recovering alcoholic and is living at a recovery center for an extended period of time. Should you list your daughter as living here?

Your grandson was recently released from the juvenile detention center. His parents got divorced while he was in the detention center. While they are getting resettled, he lists your address as his official address. Although he stays most of the time with his mother, he visits you often. Should you list your grandson as living here?

Your niece lives in a dorm at her college. However, she uses your address as her official residence. Her driver's license lists your address, and she pays taxes here. Should you list your niece as living here?

Your nephew receives his utility bill at this address and is registered to vote and pays his taxes at this address. He visits your house once or twice a week and spends the night. Should you list your nephew as living here?

You are having two Japanese students come and visit for two weeks during a summer internship. Should you list the students as living here?

Your mother stays with you for a week before she returns to a nursing home, which is her permanent residence. Should you list your mother as living here?

Experiment 2: Disability Vignettes

While in the Peace Corps in Africa, Julia contracted a form of leprosy. Although she is currently free from symptoms, the disease is thought to be highly contagious. As a result, Julia is unable to resume her work in Africa and can't find work in the U.S. Does Julia have a disability?

Walter is a 60-year-old retiree who lives alone. Six weeks ago he began experiencing significant memory loss and episodes of severe disorientation. When he went to the hospital last week for a check-up, Walter's doctors diagnosed him as having a form of dementia. Does Walter have a disability?

Kevin is blind and uses a seeing-eye dog to help him get around. Because of his condition, he has not been able to keep a regular job for pay. However, he has been able to do normal household activities independently because the house has been altered to accommodate his needs. Does Kevin have a disability?

A pre-natal infection caused Stacey's limbs to be slightly deformed and left her with moderate mental retardation. The group house where she lives now as an adult is specially equipped and staffed to enable her to carry out basic daily activities, but she has never been able to hold a regular job. Does Stacey have a disability?

Jill was accidentally exposed to dangerous chemicals at her job that damaged her respiratory system. She is being treated by her doctor, and receives worker's compensation payments that she'll be eligible for even if she gets a new job. Since she started receiving compensation she has been unmotivated to return to work. Does Jill have a disability?

Marcus is a professional football player who was placed on injured reserve when he tore ligaments in his knee during a game. Although the injury is not permanent, the league and the player's union decided that Marcus should continue to be paid for the remaining 5 games of the season. Does Marcus have a disability?

Adam's medical record classifies him as having multiple sclerosis and a mild form of depression, both of which are authorized medical conditions under the health guidelines set forth by his employer. His insurance company pays for his medications, which alleviate his symptoms for both conditions. Does Adam have a disability?

According to Tom's employee medical records, he developed diabetes when he was in his mid-forties and occasionally also suffers from temporary loss of color vision. He receives monthly checks from the U.S. Social Security Administration for his condition, but continues to work part-time at his job. Does Tom have a disability?

Mike has suffered from Hodgkin's Disease for the last ten months and now is losing his eyesight. He finds it increasingly difficult to get around without a walker, which is available to him through his company's Employee Benefits Program. He also receives money for treatment from the state's disability insurance fund. Does Mike have a disability?

Brianna was granted a medical discharge from the Navy after it was discovered that she had bone cancer. She is receiving chemotherapy and is unable to work at all because the treatments make her extremely nauseous and weak. In order to go anywhere she requires the use of a wheelchair. Does Brianna have a disability?

Because of an allergic reaction, George has come down with a serious but treatable case of acne on his face and back in the last two weeks. He went to his doctor who prescribed

some medication and assured him that his acne will heal soon. Does George have a disability?

Over the years, Carla has experienced episodes of fatigue and mild depression and occasionally has missed work as a result. Recently, she saw a doctor about her condition and he prescribed medication that has been effective in eliminating her symptoms. As a result, she has been able to maintain a consistent work schedule. Does Carla have a disability?

John has Lou Gehrig's disease. He is increasingly unstable on his feet, and he requires a part-time nurse. Because of recent modifications to his home and workplace, he is still quite able to do his job and carry out other activities of daily living. Does John have a disability?

Susan lost her foot as a result of diabetes. Although she has a prosthetic foot, she has difficulties getting around and was forced to change her career. After driving a school bus for years, she went to work as a technical writer for a computer company. Does Susan have a disability?

David received a head injury in an accident two months ago. The injuries he suffered in that accident left him paralyzed below the waist and with significant brain damage. In the last two months he has shown little improvement. He has been unable to work, and the doctors do not expect him to recover. Does David have a disability?

For 12 years, Carol has suffered from Parkinson's disease and an anxiety disorder while working as a schoolteacher. No effective treatment exists for either condition, but Carol gets around with the help of a part-time nurse. Recently, she quit her job because of the mental and physical stress involved, but has found work as a computer programmer. Does Carol have a disability?

Bill severely injured his spine while he was on active duty with the Army. The Army paid for his physical therapy and disability benefits. Bill was expected eventually to recover most of his mobility, but was nonetheless discharged from service because he could not pass the minimum aptitude test required for his sergeant's exam. Does Bill have a disability?

In eighth grade, Candice was diagnosed with Attention Deficit Disorder (ADD) and began taking Ritalin to help her concentrate in class. Because she is in a foster care program and a ward of the state, the school filed a disability insurance form on her behalf to help defray the cost of the medication. Does Candice have a disability?

After years of working on a factory production line, Ana developed arthritis and decided to quit her job. She receives state workman's compensation and is eligible for health insurance payments from her employer. Although she has received offers from several other companies, Ana can't seem to get motivated to apply for these jobs. Does Ana have a disability?

Kelly contracted viral pneumonia while working as a school nurse. To avoid possible contamination of students, the school placed Kelly on administrative leave until she recovers. The principal at her school provided her with a special medical leave form that allows her to draw her salary and receive Medicare payments for her illness. Does Kelly have a disability?

Jane has carpal tunnel syndrome in her wrists. She also suffers from recurring schizophrenic episodes that cause delusions and hallucinations. Because of her condition,

her employer has agreed to place her on medical leave, and she receives weekly payments for her treatment from the federal government. Does Jane have a disability?

Josh was born with cerebral palsy and has moderate mental impairment as well as no fine motor control. Although he has tried before, he is unable to work at any job. In addition, Josh receives funds to help cover his living expenses from the federal government which officially recognizes his condition. Does Josh have a disability?

While water skiing on vacation in the Caribbean Islands, Molly broke her right arm. As a result of her injury, she could not work at her job as an aerobics instructor for a few weeks, and she had to use her vacation time while she was recovering. Does Molly have a disability?

Pauline has minor hearing loss in both ears as a result of a childhood infection. She works as a dispatcher in a moving company. Although she occasionally misses a few words in a conversation, she can do her job adequately because her company provides her with a telephone and intercom that have powerful speakers. Does Pauline have a disability?

Experiment 2: Residence Vignettes

Linda's cousin, Dwight, is looking for a job and a new home in the city where Linda lives. He sold his house and, for two weeks, Dwight is staying with Linda. He continues to pay taxes in the state where he owned his old house. Should Linda list Dwight as a member of her residence?

Martin's fiancée has just moved into his apartment with him. She sleeps at Martin's apartment and, since they plan to live there after they are married, she helps to pay the rent there. She still has her old apartment, where she is registered to vote. Should Martin list his fiancée as a member of his residence?

Tom and his roommate share a house and they split the rent. They also split the cost of furnishing the house and have separate rooms where they keep personal belongings. Tom's roommate is listed at another address for tax purposes but regards this as home. Should Tom list his roommate as a member of his residence?

Dorothy's son usually sleeps in his room at Dorothy's house. He considers Dorothy's house to be his home and he keeps his clothes there and helps to pay the rent. This week, he is staying at Dorothy's ex-husband's house, where he is registered to vote. Should Dorothy list her son as a member of her residence?

Jane's grandson was recently released from the juvenile detention center. His parents got divorced while he was in the detention center. While they are getting resettled, he lists Jane's address as his official address. Although he stays most of the time with his mother, he visits Jane often. Should Jane list her grandson as a member of her residence?

Jim has a friend who travels frequently around the country as part of his job. When he returns to town, he stays with Jim at Jim's house. In all of his correspondence, he lists Jim's house as his official address. Should Jim list his friend as a member of his residence?

Joan's daughter, Mary, is still in high school and stays most of the time with Joan's former husband. When Mary's father is away on business trips, she stays with Joan. Joan's house is Mary's official address, and her driver's license, tax forms, and voter registration card list Joan's address. Should Joan list Mary as a member of her residence?

Adam's nephew is schizophrenic and receives care at the state mental hospital where he stays. Adam's house, which is near the hospital, is his nephew's official address. It is also where Adam's nephew is registered to vote and where his driver's license lists him as living. Should Adam list his nephew as a member of his residence?

John's brother uses John's address as his voting address, and his driver's license lists this address. He sold his house a month ago, so now he sleeps at John's most of the time, and he contributes to the rent. Should John list his brother as a member of his residence?

Mike splits the rent and utilities for his apartment with his roommate. Mike's roommate is registered to vote at their apartment and he sleeps at the apartment, which he considers his home. Sometimes Mike's roommate stays with his girlfriend at her apartment. Should Mike list his roommate as a member of his residence?

Paul and Shannon are members of an international exchange program in which they act as a host family. This summer, a Japanese student has come to visit for two weeks during an internship. The exchange program pays Paul and Shannon a small fee for acting as hosts. Should Paul and Shannon list the student as a member of their residence?

Frank's mother stays in a nursing home, where the staff can help care for her. When Frank can arrange it, his mother stays with him for a week or two. Eventually, Frank has to return his mother to the nursing home, where she is registered to vote. Should Frank list his mother as a member of his residence?

Eric's wife, Amy, works in the city and rents an apartment there. Her driver's license lists the address for her apartment in the city. On Thursdays through Mondays, she stays with Eric and their children and regards that as her home. Should Eric list Amy as a member of his residence?

Julio's mother frequently stays with him and his wife. She has her own room in Julio's house, where she keeps most of her clothes and personal items. She is registered to vote at her apartment, where she stays when she is not at Julio's house. Should Julio list his mother as a member of his residence?

Sally's best friend, Donna, officially lives at her parent's house. But most nights Donna sleeps at Sally's parent's house, which she thinks of as her home. Donna has her own room where she keeps some of her belongings and in exchange she pays Sally's parents a small monthly rent. Should Sally's parents list Donna as a member of their residence?

Monica's sister just moved to this country from Germany and sleeps at Monica's house because she has no other place to stay. She has unpacked and stored her possessions in a spare bedroom. She thinks of Monica's house as home but still pays taxes in Germany. Should Monica list her sister as a member of her residence?

Todd's daughter's tax forms list her living at Todd's address. However, she is currently a recovering alcoholic and enrolled herself in a recovery program. For the last six months she has been living at a nearby clinical center, receiving treatment for her condition. Should Todd list his daughter as a member of his residence?

Louis lives in a condo that his father bought for him. His father pays for the maintenance and utilities at the house. Louis' father will often visit Louis to check up on everything but stays most of the time in another house he owns. Should Louis list his father as a member of his residence?

Allison's nephew lives in a dorm at his college. However, he uses Allison's home as his official address for important matters. The address on his driver's license is Allison's

address, and his tax forms also list this address. Should Allison list her nephew as a member of her residence?

Ben's nephew lives most of the time at a residential hotel. But he receives his phone bill at Ben's address and is registered to vote there. Also, his driver's license lists Ben's address. He visits Ben's house once or twice a week and spends the night. Should Ben list his nephew as a member of his residence?

Steve's son Nick is in the military, and although Nick's wife has her own apartment, she sleeps at Steve's house most nights and pays her taxes there. She keeps half of her belongings in her room there. She contributes towards the rent and her driver's license lists Steve's address. Should Steve list his daughter-in-law as a member of his residence?

Ellen and her spouse recently purchased a condo for retirement and split the mortgage and utilities. Their belongings are moved in and the condo is their official address. Ellen's spouse has not fully retired and sometimes sleeps at their son's house in the city during the week. Should Ellen list her spouse as a member of her residence?

Susan's niece attends college in the same town where Susan lives. Susan's niece lives in the dorm, but she stays with Susan during the weekends and short breaks. Her driver's license lists her parent's address where she stays during the summer break. Should Susan list her niece as a member of her residence?

Simon's ex-wife has full custody of their daughter since they were divorced. Simon gets to see his daughter for a few weekends each year. His daughter has just gotten her driver's license which lists Simon's ex-wife's house on the address. Should Simon list his children as members of his residence?

6. References

- Bassili, J.N. and Scott, B.S. (1996). Response Latency as a Signal to Question Problems in Survey Research. *Public Opinion Quarterly*, 60, 390–399.
- Belson, W.A. (1981). *The Design and Understanding of Survey Questions*. Aldershot: Gower.
- Coleman, L. and Kay, P. (1981). Prototype Semantics: The English Word Lie. *Language*, 57, 26–44.
- Conrad, F.G., Couper, M.P., Tourangeau, R., and Peytchev, A. (2006). The Use and Nonuse of Clarification Features in Web Surveys. *Journal of Official Statistics*, 22, 245–269.
- Conrad, F.G. and Schober, M.F. (1999). A Conversational Approach to Text-based Computer-Administered Questionnaires. *Proceedings of the Third International ASC Conference*. Chesham, UK: Association for Survey Computing, 91–102.
- Conrad, F.G. and Schober, M.F. (2000). Clarifying Question Meaning in a Household Telephone Survey. *Public Opinion Quarterly*, 64, 1–28.
- Couper, M.P. (2000). *Web Surveys: A Review of Issues and Approaches*. *Public Opinion Quarterly*, 64, 464–494.
- del Pinal, J., Bennett, C., Cresce, A., and Martin, E. (2002). Overview of Hispanic and Race Reporting in Census 2000. Paper presented at the Joint Statistical Meetings, American Statistical Association, New York, New York, August 15.

- Draisma, S. and Dijkstra, W. (2004). Response Latency and (Para)linguistic Expressions as Indicators of Response Error. In S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds), *Methods for Testing and Evaluating Survey Questionnaires*, 131–147. New York: John Wiley and Son.
- Gerber, E. (1994). *The Language of Residence: Respondent Understandings and Census Rules*. Final Report of the Cognitive Study of Living Situations. Report for the Center for Survey Methods Research. Suitland, MD: U.S. Bureau of the Census.
- Gerber, E., Wellens, T., and Keeley, C. (1996). Who Lives Here? The Use of Vignettes in Household Roster Research. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 962–967. Alexandria, VA: American Statistical Association.
- Hirschfeld, L.A. (1994). Is the Acquisition of Social Categories Based on Domain-specific Competence or Knowledge Transfer? In L.A. Hirschfeld and S.A. Gelman (eds), *Mapping the Mind: Domain Specificity in Cognition and Culture*, 201–233. Cambridge: Cambridge University Press.
- Jette, A.M. and Badley, E. (2000). Conceptual Issues in the Measurement of Work Disability. In N. Mathiowetz and G. Wunderlich (eds), *Survey Measurement of Work Disability: Summary of a Workshop*, 4–27. Washington, D.C.: National Academy Press.
- Kim, N.S. and Ahn, W.-K. (2002). Clinical Psychologists' Theory-based Representations of Mental Disorders Predict Their Diagnostic Reasoning and Memory. *Journal of Experimental Psychology: General*, 131, 451–476.
- Klein, D.E. and Murphy, G.L. (2001). The Representation of Polysemous Words. *Journal of Memory and Language*, 45, 259–282.
- Klein, D.E. and Murphy, G.L. (2002). Paper Has Been My Ruin: Conceptual Relations of Polysemous Senses. *Journal of Memory and Language*, 47, 548–570.
- Medin, D.L. (1989). Concepts and Conceptual Structure. *American Psychologist*, 44, 1469–1481.
- Medin, D.L. and Coley, J.D. (1998). Concepts and Categorization. In J. Hochberg and J.E. Cutting (eds), *Handbook of Perception and Cognition* (2nd ed.): Perception and Cognition at Century's End, 403–439. San Diego: Academic Press.
- Murphy, G.L. and Kaplan, A.S. (2000). Feature Distribution and Background Knowledge in Category Learning. *Quarterly Journal of Experimental Psychology*, 53A, 962–982.
- Nosofsky, R.M., Palmeri, T.J., and McKinley, S.C. (1994). The Rule-plus-exception Model of Classification Learning. *Psychological Review*, 101, 53–79.
- Pennington, N. and Hastie, R. (1992). Explaining the Evidence: Testing the Story Model of Jury Decision Making. *Journal of Social and Personality Psychology*, 62, 189–206.
- Rosch, E. and Mervis, C.B. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573–605.
- Ross, B.H. and Murphy, G.L. (1999). Food for Thought: Cross-classification and Category Organization in a Complex Real-world Domain. *Cognitive Psychology*, 38, 495–553.
- Schober, M.F. and Conrad, F.G. (1997). Does Conversational Interviewing Reduce Survey Measurement Error? *Public Opinion Quarterly*, 60, 576–602.

- Schober, M.F., Conrad, F.G., and Fricker, S.S. (2004). Misunderstanding Standardized Language. *Applied Cognitive Psychology*, 18, 169–188.
- Smith, E.E., Shoben, E.J., and Rips, L.J. (1974). Structure and Process in Semantic Memory: A Featural Model for Semantic Decisions. *Psychological Review*, 81, 214–241.
- Suessbrick, A., Schober, M.F., and Conrad, F.G. (2000). Different Respondents Interpret Ordinary Questions Quite Differently. *Proceedings of the American Statistical Association, Section on Survey Methods Research*, 907–912. Alexandria, VA: American Statistical Association.
- Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The Psychology of Survey Response*. New York: Cambridge University Press.

Received December 2003

Revised September 2005