NEW CATEGORIES ARE NOT ENOUGH:

Rethinking the Measurement of Sex and Gender in Social Surveys

LAUREL WESTBROOK
Grand Valley State University, USA
ALIYA SAPERSTEIN
Stanford University, USA

Recently, scholars and activists have turned their attention toward improving the measurement of sex and gender in survey research. The focus of this effort has been on including answer options beyond "male" and "female" to questions about the respondent's gender. This is an important step toward both reflecting the diversity of gendered lives and better aligning survey measurement practice with contemporary gender theory. However, our systematic examination of questionnaires, manuals, and other technical materials from four of the largest and longest-running surveys in the United States indicates that there are a number of other issues with how gender is conceptualized and measured in social surveys that also deserve attention, including essentialist practices that treat sex and gender as synonymous, easily determined by others, obvious, and unchanging over the life course. We find that these understandings extend well beyond direct questions about the respondent's gender, permeating the surveys. A hyper-gendered world of "males" and "females," "brothers" and "sisters," and "husbands" and "wives" shapes what we can see in survey data. If not altered, surveys will continue to reproduce statistical representations that erase important dimensions of variation and likely limit understanding of the processes that perpetuate social inequality.

Keywords: *survey design; knowledge production; sex; gender; transgender*

The categories offered on surveys and official documents to capture key individual characteristics are often taken for granted and seen as natural distinctions, despite the fact that they, like all classification schemes, are socially constructed (Berger and Luckmann 1967; Zerubavel 1996). The question "Are you male or female?" is a case in point (Spade 2011). Most Americans are so used to seeing some version of this question on forms that we rarely stop to think about what purpose it serves, how we know which box to check, and why there are only two answer options.

Recently, in response to calls for recognition of sex and gender diversity and for more research on its correlates and consequences (e.g., Balarajan, Gray, and Mitchell 2011; Human Rights Watch 2011; Institute of Medicine 2011), specialty surveys have begun to test alternative measures of sex and gender. Such studies find that transgender and gender-variant Americans face considerable discrimination and material hardship, demonstrating that standard measures fail to capture important aspects of inequality (GenIUSS Group 2014).

This new body of work has focused on how best to ask survey respondents about their sex and gender—and what answer options to provide—but has not yet considered that assumptions about these concepts could affect other aspects of survey design. Sex and gender are primary, almost automatic, categorical distinctions through which we make sense of the world and evaluate others, with important consequences for the maintenance of social inequality (Ridgeway 2011). Thus, we might expect surveys to not only prioritize recording such information for research purposes, but also be unaware of how traditional understandings of sex and gender have seeped into measurement practices. We cast a wider net than previous research and scrutinize surveys in their entirety through a systematic analysis of questionnaires, manuals, and other technical materials from four of the longest-running surveys in the United States. Our examination of how sex and gender have been conceptualized and measured spans the six decades that large-scale social surveys have chronicled American life (Igo 2007) and evaluates survey practices against both contemporary gender theory and lived experience.

The results reveal potential problems in the survey measurement of sex and gender beyond an unnecessarily limited set of answer options. We find that large social surveys generally conflate sex and gender and treat the resulting conceptual muddle as a starkly dichotomous, biologically fixed, and empirically obvious characteristic. This treatment is not restricted to questions recording the respondent's sex or gender;

instead, it permeates the survey documents. These practices have changed little over time and have important consequences for sociological knowledge about sex and gender disparities. Knowledge construction is a cyclical process: Beliefs about the world shape how surveys are designed and data are collected; survey research findings, in turn, shape beliefs about the world, and the cycle repeats (Bowker and Star 2000). If major U.S. surveys continue to both essentialize and dichotomize sex and gender, survey research will continue to produce findings and reproduce beliefs that are disconnected not only from current social science theory but also from the diversity of gendered experiences in American society.

ALIGNING THEORY AND METHOD

Creating a survey questionnaire is both science and art (Schaeffer and Presser 2003). Writing a single question requires as many as a dozen separate decisions, from whether it should be closed- or open-ended to the number of response categories and what they should be called (Schaeffer and Dykema 2011). Some guidelines are commonly accepted, such as that response options should be both exhaustive and mutually exclusive, while others, such as the optimum length of a rating scale or whether to offer a "don't know" option, are still in dispute or vary depending on the question being asked (Krosnick and Presser 2010).

Methodologists also differ in the extent to which they advocate keeping the same questions, even if they are known to be suboptimal, in order to maintain a continuous time series on a particular issue (Krosnick and Presser 2010). Some argue survey measures should constantly be refined and evaluated for their fit to both theoretical concepts and empirical observation (Hox 1997). In line with Hox, we aim to consider not only how major U.S. social surveys have measured sex and gender over the past six decades, but also the extent to which their operationalization of the concepts aligns with contemporary social science theory.

In doing so, we do not mean to imply that there is unanimity in how to conceptualize sex and gender; theories of the "sex/gender system" (Rubin 1975) continue to evolve across the scholarly community. However, a broad consensus currently exists, at least among sociologists, on several key dimensions: (1) Although related, "sex" and "gender" are best understood as distinct concepts; (2) there are more than two sexes and more

than two genders; (3) how people identify in terms of sex or gender may not "match" how other people perceive and classify them; and (4) both identities and classifications can change over a person's life course.

Many contemporary gender scholars see sex and gender as distinct, though the concepts are often conflated in both academic and nonacademic contexts (Cowan 2005; Hammarstrom and Annandale 2012; Pryzgoda and Chrisler 2000; Wickes and Emmison 2007). "Sex" refers to the division of bodies into categories based on genitals, chromosomes, and/or hormone levels (West and Zimmerman 1987). Although this schema draws on biological criteria, research demonstrates that the specific distinctions are neither natural nor stable; our beliefs about sex have varied widely over time, differ between cultures, and tend to erase naturally occurring differences in physical development such as "intersex" people (Fausto-Sterling 2000; Lorber 1993; Nicholson 1994). "Gender" typically refers to behaviors associated with membership in a sex category (West and Zimmerman 1987). However, gender is not determined by sex, and people can "do gender" in a range of ways and/or identify with gender categories other than the one associated with the sex they were labeled at birth. Thus, there is significant gender diversity within both cisgender and transgender categories (Connell 1995; Halberstam 1998; Lucal 1999; Valentine 2007),[1] gender determinations made by others may not align with how people see themselves (Westbrook and Schilt 2014), and a person's sex category and/or gender identity can change over time (Meyerowitz 2002).

The ability of survey research to capture this complexity has long been questioned as part of a larger feminist critique of quantitative methods (e.g., Stacey and Thorne 1985). Consequently, as Sprague (2005) and others have noted, many gender scholars dismiss statistical representations out of hand, and work detailing feminist methodology often does not describe quantitative research at all (Undurraga 2010). This has left the measurement of sex and gender in surveys largely unexamined, despite critical attention to the conceptualization of many other categories of difference in survey research, such as race, ethnicity, and sexuality (e.g., Powell and Bolzendahl 2010; Snipp 2003), and numerous articles that offer new measures on other key population distinctions, from age structure (Settersten and Mayer 1997) to families and households (Randall, Coast, and Leone 2011).

Some scholars maintain that survey research can be "rehabilitated" because it is not inherently antithetical to feminist goals (e.g., Oakley 1998). From this perspective, careful attention to survey methods is necessary because the resulting research is vital for expanding our understanding

of the gender system, working to reduce gender inequalities, and providing entry into policy debates (e.g., Williams 2006). However, feminists working to improve surveys have thus far focused on issues such as the patriarchal definition of household "heads" (Oakley and Oakley 1979) or how to incorporate intersectional perspectives in survey research (Harnois 2013) rather than on the measurement of sex and gender itself.[2]

The recent push for new sex and gender measures has come largely from activists and academics outside the social sciences, including legal scholars and those in public health fields. A clear sense of best practices for data collection has yet to coalesce (but see GenIUSS Group 2014), with recommendations ranging from adding intersex, transgender, genderqueer, or open-ended response options to asking separate questions for sex assigned at birth, current gender identity, and/or gender expression (Cruz 2014; Harrison, Grant, and Herman 2011; Schilt and Bratter 2015). Most of this work takes issue with the fact that people outside current sex and/or gender norms cannot select a response that reflects their identities and lived experiences but stops short of presenting a broader critique of the conceptualization of sex and gender in surveys. This research also tends to focus on capturing diversity within the category of "transgender," but leaves the traditional categories of "male" and "female" unquestioned.

We agree that asking separate questions about sex and gender and offering more than two answer options are important steps toward better aligning survey research practice with gender theory and lived experience (see Saperstein, Westbrook, and Magliozzi 2015). However, a systematic investigation of how major social surveys conceptualize sex and gender is also necessary to determine if survey research methods are simply in need of a little renovation, or whether a more thorough rethinking of measurement practices is in order.

METHODS

Our data come from a larger study of the materials pertaining to the measurement of sex, gender, race, ethnicity, and sexuality from four of the largest, longest-running, and most influential social surveys with national sampling frames in the United States: the American National Election Study (ANES), Panel Study of Income Dynamics (PSID), General Social Survey (GSS), and 1979 National Longitudinal Survey of Youth (NLSY). Long-running and well-regarded surveys are important to analyze because, in addition to the wide dissemination of their findings, they serve as

models for newer surveys and are often used to train students in statistical methods. As the National Opinion Research Center advertises on its website, "more than 22,000 journal articles, books, and PhD dissertations are based on the GSS; and about 400,000 students use the GSS in their classes each year."[3]

We collected all of the questionnaires, codebooks, interviewer instructions, and user's guides produced by each survey and made available online for every round of interviews from their inception to 2011, representing a total of 132 survey-years. The documents for each survey-year ranged from 37 to more than 3,000 pages, with most being more than 200 pages. They capture all of the information the typical researcher would have access to when using these surveys for secondary data analysis. Thus, our investigation considers both what researchers can know about how the major U.S. social surveys conceptualize and measure categories of difference and what they cannot know because the information was either not written down or not made publicly available. See Table 1 for more details on each survey.

This study combines techniques from textual and discourse analysis with those from traditional content analysis (Mills 1997; Neuendorf 2002; Titscher et al. 2000). To explore the operationalization of sex and gender, each document was examined in its entirety for passages and questions of interest. We also searched each document for 30 keywords, including: sex, gender, fe/male, wo/man, wo/men, masculine, feminine, boy, girl, son, daughter, intersex, transgender, transsexual, and gendered pronouns (s/he, him, her, etc.). To provide a sense of how common these terms were in each survey-year, we kept counts for each of our gendered search terms and 16 additional gender-neutral terms (including spouse, partner, parent, and child). Table 1 provides the average number of each type of term per page. Our counts demonstrate that all four surveys are packed with gendered language, with some (PSID) more strongly gendered than others (GSS). This quantitative variation might suggest the four surveys differ in their conception of sex and gender, but our qualitative analysis finds little difference in measurement practices over time and across surveys. Although the frequency of gendered terminology varies by survey, how it is deployed does not.

Text was extracted for qualitative coding (in one or more pages) when it related to the measurement of sex and/or gender. For the purposes of this analysis, "measurement" includes any moment where respondents or others were put in a sex or gender category and counted. We also

TABLE 1:   Characteristics of Studied Surveys

	American National Election Study (ANES)	Panel Study of Income Dynamics (PSID)	General Social Survey (GSS)	National Longitudinal Survey of Youth (NLSY)
Years active	1948–present	1968–present	1972–present	1979–present
Survey design	Cross-sectional, pre- and post-election interviews; occasional multiyear panel studies	Longitudinal	Cross-sectional; three-wave panel studies post-2006	Longitudinal
Sample size	About 1,500 U.S. eligible voters	Began with 4,800 households	About 1,500 U.S. adults living in households	Began with 12,686 14-to-22 year olds
Research impact*	>6,300	3,760	>22,000	>6,000
Survey-years	39	39	30	24
Pages of documentation	35,470	17,539	8,689	8,462
Sex/gender excerpts	454	436	326	145
Gendered terms per page (average)	0.93	4.2	0.85	1.26
Gender-neutral terms per page (average)	0.90	1.41	0.60	2.81

*Research impact is defined as the number of books, journal articles, book chapters, dissertations, and reports listed on each survey's website.

collected content that indicated the range of sex and gender categories in each survey-year, such as when the questionnaire provided examples of possible pronouns or included terms such as "husband/wife" and "brother/sister." Beliefs about gendered behavior are also evident in the survey materials (such as what jobs men or women should have, who should do housework, etc.), but we did not analyze aspects of the surveys that reflected gendered beliefs without directly affecting the measurement of sex or gender because they have been documented elsewhere (e.g., Brewster and Padavic 2000). We instead focus on what often gets taken for granted even

in such studies—the implicit first step in any gendered analysis: What are the available categories and how do surveys decide who belongs in them?

To capture trends in how each survey conceptualized sex and gender, we extracted one example of each type of gendering present in a given survey-year.[4] For example, in 1993, the NLSY included more than 300 questions about the respondent's "husband/wife," but we extracted only the first example for coding because the number of categories and the criteria for membership in those categories were the same for each question. If, in the same year, the same survey also used a different conceptualization, such as referencing the respondent's "husband/wife/partner," we collected that example as well to capture the range of conceptualizations that existed in the same survey-year. Following the parameters described above, we collected and analyzed 1,361 extracts—an average of about 10 distinct types of gendering per survey-year.

Gendering takes diverse forms in these surveys. About one-fifth of the excerpts relate to the direct determination of the respondent's sex or gender. The remaining four-fifths consist of more hidden gender categorizations.[5] These include the use of gendered pronouns and gendered answers to questions about with whom the respondent lived as a child (e.g., aunt or uncle).

Each question, variable description, or text excerpt (all of which will be referred to as "excerpts" in this article) was coded using Atlas.ti, a qualitative data analysis software program ideal for handling projects with a large number of primary documents. Codes were developed inductively to respond to patterns that emerged from the data. We generated a total of 217 codes, clustered into 25 questions, including whether the terms "sex" or "gender" were used explicitly, who determined the sex or gender of the person in question (i.e., respondent, interviewer, etc.), how many categories were given (i.e., male and female only, or were there other options), and how much (if any) instruction was provided for how to make the categorization.

Our data are limited in the sense that we cannot speak directly to the motivations of the survey designers unless explanations were provided in the available materials. Nevertheless, it is equally if not more important to consider the consequences of such design decisions in terms of the surveys' abilities to reflect current gendered experiences and gender theories. Our analysis of the methodological decisions made by four major U.S. social surveys as to which aspects of sex and gender to measure, how to measure them, and which gendered ways of being to acknowledge illuminates a broader set of issues that should be taken into account by those seeking to improve social surveys.

MEASURING SEX AND GENDER
IN SOCIAL SURVEYS

The surveys we examined all operationalize sex and gender as binary, composed of males and females only, and as givens rather than socially constructed systems of classification. The answer options for direct determination of the respondent's sex or gender are always "male" and "female." There are never additional sex options (such as "intersex") or additional gender options (such as "transgender"). Although such findings are not particularly surprising, they stand in contrast to other commonly asked demographic questions. Characteristics from race to political affiliation are no longer counted as binary distinctions, and possible responses often include the category "other" to acknowledge the difficulty of creating an exhaustive preset list of survey responses (Krosnick and Presser 2010). By always providing the same two-answer options, surveys imply that the categories "male" and "female" cover all possible ways of being.

The limited answer options should be a concern for those seeking to improve surveys, but we find additional ways in which these surveys reflect understandings that are not in line with contemporary gender theory or lived experiences. All four surveys conflate sex and gender, do not distinguish between self-identified and other-determined gender, treat sex and gender as obvious, and do not allow a person's gender to change over time. These essentialist conceptions are not limited to direct questions about the respondent, but instead are found throughout the documentation. Binary categorization is both ubiquitous and hidden, emphasizing the need for a careful examination of entire surveys as an important step on the path to rehabilitation.

Conflating Sex and Gender

Unlike gender theories that treat sex and gender as analytically distinct, the surveys we examined tend to treat sex and gender as synonymous. Even when surveys say they are measuring *gender*, they are actually measuring *sex* or some unstated combination of the two. Questions that reference the respondent's gender always use the sex terms "male" and "female" as response categories (see Figure 1), rather than gender terms such as "man," "woman," "masculine," "feminine," "cisgender," and "transgender." Surveys also conflate the two concepts by using the word "sex" in the question wording but "gender" in the variable name and vice versa, or using the terms interchangeably in the same paragraph. The excerpt in Figure 1 is typical, where the variable is called "SEX," but the

SEX: Categorical (Single)
SELECT GENDER OF CHOSEN RESPONDENT.

Categories:
{male}   MALE
{female}   FEMALE

FIGURE 1:   Conflating Sex and Gender
Source: GSS, 2008, Ballot 1, p. 23.

interviewer is told to "SELECT GENDER OF CHOSEN RESPONDENT."
We use the term "sex/gender" in the remainder of the article to refer to the
surveys' fuzzy operationalization of these concepts.

Further evidence of conceptual conflation is the absence of questions
that capture the separate concepts of sex and gender; the surveys we
examined only ever ask about a person's sex or their gender without
acknowledging the potential difference between the two. This is a prob-
lem because it means researchers cannot disentangle which concept is
actually being captured by the variable. For example, although most
health research has focused on biological difference (sex), social expecta-
tions for behavior (gender) also shape health outcomes, and failing to
distinguish between the two results in missed opportunities for under-
standing health disparities (Bird and Rieker 2008). Whether sex and gen-
der have similar or different effects on a particular outcome is an important
empirical question, but that question cannot be answered with a single
measure of sex/gender.

Encouraging the practice of distinguishing between sex and gender
(while recognizing both as socially constructed) is an important step
toward capturing population diversity and facilitating the use of survey
data to study sex and gender as social systems. One such method is to ask
separate questions about a respondent's sex and gender, a practice that has
become common among surveys seeking to identify transgender respond-
ents (GenIUSS Group 2014).[6] Depending on the purpose of the survey,
this could include questions about sex at birth, current gender identity,
and/or assessments of how respondents rate themselves on scales of mas-
culinity and femininity.

Essentializing the Respondent's Sex/Gender

In addition to conflating sex and gender, these surveys treat the
resulting mash-up of sex/gender as a natural or essential characteristic

```
                        INTERVIEWER'S SUPPLEMENT

D25. Respondent's sex is:      /1. Male/    /2. Female/

D26. Respondent's race is:     /1. White/   /2. Negro/   Other:_____

D27. Relationship of R to head: /1. Is head/ /2. Is wife/  Other:_____
```

FIGURE 2: Interviewer Determination of Respondent Characteristics
Source: ANES, 1966, Post-election Questionnaire, p. 42.

of all people. We highlight three essentialist practices that are typically overlooked in previous research: Respondents are rarely allowed to self-identify, sex/gender is assumed to be "obvious," and there is little to no acknowledgement that an individual's sex/gender can change over time.

Who Determines Gender? Gender scholars highlight distinctions between self-identification of gender and other people's determinations of a person's gender. "Determining gender" is the corollary to "doing gender." We all "do gender" so as to be read as members of a gender category and others "determine" our gender, placing us into the category they perceive to be most appropriate (Westbrook and Schilt 2014). However, self-identified gender and other-determined gender do not always "match." For example, transgender women and masculine cisgender women both identify as women but might be seen as men by others in public restrooms (Cavanagh 2010). Ideally, to align with contemporary gender theory, research should distinguish between these two methods of categorization by measuring both self-identified gender and how others perceive the respondent. The surveys we examined do not do so.

The interviewer determines the sex/gender of the respondent without asking in nearly every survey-year we examined. This count includes instances when it is explicitly stated that the interviewer should code the respondent's sex/gender "by observation," when the interviewer classifies the respondent after completing the interview in a special section titled "interviewer remarks" or "interviewer's supplement" (Figure 2), and survey-years where the respondent's sex/gender was carried forward from a previous year when the past classification was done by observation. Only the 2008–2009 ANES allowed all respondents to self-identify in response to a direct sex/gender categorization question (Figure 3). However, this does not appear to represent a long-term change in practice.[7] If anything, the measurement modification was prompted by a

```
[Y2]              Are you male or female?
                  __ Male
                  __ Female
```

FIGURE 3:   Self-Identified Sex/Gender on an Internet Survey

Source: ANES, 2008–2009, Questionnaire (Wave 1), p. 53.

change in interview mode. The ANES typically employs in-person and some telephone interviews, but respondents completed the portion of the 2008–2009 survey that included their sex/gender categorization online. Interviewer classifications cannot be collected in a self-administered Internet survey—a point that has important implications for research on sex and gender as other surveys make the switch to online modes over traditional telephone and face-to-face interviews.

The current reliance on interviewer-determined sex/gender raises several other issues for those seeking to rehabilitate surveys and makes it clear that adding answer options to sex/gender questions will not fully address the limitations of current survey practices. When interviewers record their perceptions without asking, respondents are unlikely to know that they are being categorized and so cannot challenge the interviewer's determination or opt-out of the practice of sex/gender classification. Also, if only new answer options are added, it is the interviewer's perception of the respondent's intersex or transgender status would be recorded, rather than the respondent's self-identification.

Sex/Gender as Obvious. Recording respondents' sex/gender based solely on interviewer observation implies it is an obvious characteristic. Interviewers are expected to be able to look at respondents, or listen to their voices over the telephone, and easily categorize them as male or female. Further evidence of this essentialist conception of sex/gender is the general lack of instruction regarding how to decide whether a person is male or female. Sex/gender is treated as obvious to everyone: to the interviewer, who is rarely given criteria to use in labeling the respondent; to the respondent, who is given no instructions for how to determine the gender of others; and to the researcher, who is seldom told how sex/gender is operationalized under any circumstances. As a result of these omissions, researchers interested in studying the effects of sex or gender on other processes have no way of knowing which criteria were used to make the categorization (i.e., secondary sex characteristics, gendered behavior, voice timbre, clothing, etc.).

```
Q. 11:  For this question, if the person's sex is obvious, you simply
        enter an "M" or "F" in column 11 without asking. In most
        cases the sex of the person should be obvious, either because
        the person is present or because the sex is indicated by the
        relationship to the Householder. Do not infer a person's sex
        on the basis of the person's name alone.
```

FIGURE 4:   Instructions for Determining Sex
Source: NLSY, 1978, Interviewer Instructions, Household Screening, p. 93.

In most direct sex/gender questions, the interviewer is simply instructed to record the respondent's sex/gender "by observation." In rare cases, this is accompanied by the prompt to "ask if not obvious." The most extensive instructions appeared in the field manual for the NLSY79 screening interview (Figure 4). Yet even this relatively detailed passage tells us little more than the criteria that should NOT be used (person's name), giving no indication of what factors the interviewers used to determine a person's sex/gender. The passage makes explicit the understanding of sex/gender we found throughout these surveys—if a person is present, sex/gender "should be obvious." This treats sex/gender as visible and easily knowable, making sex/gender categorization seem straightforward, static, and natural rather than complex, variable, and socially constructed.

Out of all the documents we examined, interviewers were instructed to ask about an individual's sex/gender if it was "not obvious" in just 14 of 132 survey-years. Most of these were related to determining the gender of someone who was not present for the interview. The implication that sex/gender is not necessarily obvious first appears in the GSS (1975), followed by the NLSY (1979), and the ANES (2000), but it does not appear regularly or follow any apparent pattern. Although infrequent, the inclusion of the possibility that sex/gender categorization might not be obvious is notable because it demonstrates some alignment with contemporary gender theory.

The ANES is the only survey to include an "ask if not obvious"–type instruction for a direct question about the respondent's sex/gender, and it did so just twice (in 2002 and 2008). The 2008 version occurred in a validation survey follow-up conducted by telephone, and is worth examining in detail (see Figure 5). In this excerpt, the ANES implies sex/gender is usually obvious, even over the telephone. At the same time, interviewers were instructed to tell respondents they are "required to ask" for a sex category response, despite the fact that they do so only if they were unsure of the person's sex/gender. We cannot speak to the specific motivations for this choice of question wording, but it appears to result from thinking that asking about sex/gender is "awkward" (e.g., ANES 2008–2009 user's guide, Appendix C, 239). The belief that asking about a person's sex/

Q12
(INTERVIEWER: ONLY ASK IF YOU CAN'T TELL FROM THE NAME AND SOUND OF
THE VOICE)

I'm required to ask, are you male or female?

    1 Q NOT ASKED, CAN TELL R OBVIOUSLY MALE
    2 Q NOT ASKED, CAN TELL R OBVIOUSLY FEMALE
    3 R ANSWERS MALE
    4 R ANSWERS FEMALE

FIGURE 5: Rare "Ask If Not Obvious"–Style Question
Source: ANES, 2008–2009, User's Guide (Appendix E), p. 266.

gender is "awkward" likely stems from a conviction that everyone's sex/
gender should be obvious, and the perceived implication that there might
be something wrong with a person if it is not.[8]

Notably, Figure 5 presents the only example in our data where survey
answer options explicitly indicate whether the sex/gender responses were
the result of interviewer determination or self-identification. By typically
not distinguishing between these two different measurements of gender,
surveys treat them as equivalent. In contrast, measuring them separately
allows researchers to explore the relationship between self-identified and
other-determined gender, an analytical strategy that has been fruitful in
studies of racial inequality (Saperstein 2012).

Sex/Gender as Static Across the Life Course. As Garfinkel (1967) noted,
before what were then called "sex-change operations" became available
in the 1950s, Americans conceived of the world as comprised solely of
people whose current sex category was the same as the sex they were
assigned at birth (see Figure 6). Since then, Americans have begun to
recognize the possibility, and the legitimacy, of changing sex and gender
(Meyerowitz 2002; Schilt and Westbrook 2009).

Despite this cultural shift, the surveys we examined continue to reproduce
the same two-by-two table understanding that Garfinkel debunked more than
half a century ago. Not only are transgender possibilities not provided in
moments of gender determination, the words "transgender" or "transsexual"
do not appear anywhere in their documentation. Moreover, three of the four
surveys explicitly state in their materials that any change in sex/gender cat-
egorization from one interview to the next is considered an error. The NLSY
user's guide stated that sex/gender categorizations of respondents "are sub-
ject to a small degree of error from erroneous interviewer observation and/or
recoding and data entry error. Therefore, when using this series of variables,
a small number of respondents may appear to 'change' sex across surveys"

		At time$_2$	
		Male	Female
At time$_1$	Male	1.0	0.0
	Female	0.0	1.0

FIGURE 6:   A Static Understanding of Sex
Source: Garfinkel (1967), p. 117.

(NLSY 2000, 153). The user's guide for ANES 2008–2009 similarly noted that "data on the respondent's sex, age, country of birth, and other 'invariant' characteristics have been found to be inconsistent after repeated measures" (ANES 2010, 88). Thus, for the ANES, a person's sex/gender should be as fixed as one's year or country of birth. To avoid these "errors," panel surveys often require interviewers to double-check that the sex/gender recorded for each member of the respondent's household is the same at time 2 as it was at time 1. When a discrepancy is found, the old data appears to be treated as erroneous while the new statement of sex/gender is treated as "true." This further demonstrates the treatment of sex/gender as obvious; rather than recognize that different interviewers could categorize someone's sex/gender differently, or that changes can occur, the inconsistent codes are assumed to result from an error at time 1.

Panel studies also treat sex/gender as static over the life course by simply carrying forward sex/gender categorizations from previous years, as three of the four surveys do at various points in their history. The PSID stopped recording the respondents' sex/gender in 1989, the NLSY stopped its annual categorizations in 1998, and the ANES sometimes carries forward the respondent's sex/gender in its panel studies and sometimes does not. To justify implementing a "summary" variable, the PSID said that it had "encountered only one legitimate sex change case throughout the history of the study" (PSID 1992 manual, 419).

Thus, the world constructed by these surveys is cisnormative (Bauer et al. 2009), meaning that it assumes that everyone is cisgender (e.g., that people do not change sex/gender over time). This practice not only stands in contrast to current theories of gender, it also does not reflect lived experiences. Transgender people of all sorts exist in the United States, but they do not exist in the picture of society painted by the surveys we examined and so cannot exist in analysis and published research that draws from these data. When all changes over time are considered errors that must be "fixed," transgender experiences are literally erased from survey data collection.

One explanation for why sex/gender changes are not explicitly recorded in these surveys is that, if such changes are very rare, including that information could compromise the anonymity of the respondents. Of course, treating all changes in categorization as coding mistakes that need to be corrected and failing to recognize sex/gender diversity in the first place (e.g., by offering intersex and transgender response options or asking separate questions about sex at birth and current gender identity) also guarantee that large social surveys will continue to find sex/gender diversity to be rare or nonexistent.

A related objection to including questions that would identify sex/gender variant respondents is that, given the sample size of most surveys, the number of people identified would be too small for multivariate analysis (GenIUSS Group 2014). However, without including such questions in national surveys, it is difficult to estimate the size of a population in the first place (Gates 2011). Moreover, with a national estimate, small populations can be intentionally oversampled in future studies and, in cross-sectional surveys with repeated data collection (such as the ANES and GSS), data can be pooled across years to facilitate analysis. Surveys also regularly offer categories for which counts will be too small for most research purposes, such as American Indians and Native Hawaiians, or Buddhists and Muslims, when doing so is seen as important for representation. Given that statistical classification schemes often become taken-for-granted depictions of the social world (Urla 1993), excluding intersex and transgender possibilities has consequences well beyond the practice of a single survey.

The Hyper-Gendered World of Social Surveys

Gendering moments in surveys are not confined to direct categorizations of the respondent's sex/gender. Direct categorization of the respondent—when the respondent's sex/gender is being determined—accounts for just 22 percent of the 1,361 gendering moments we examined.[9] The remaining 78 percent represents a wide variety of other forms of sex/gender categorization, including the gendering of household residents or family members, sexual partners, the interviewers themselves, and hypothetical people in vignettes. Some of these are phrased as direct questions about the other person's sex/gender (i.e., "What sex is [hh member name]?"). However, most sex/gender determination is not done through direct categorizations, nor is it signaled by the term "sex" or "gender." Instead, gendering in these surveys generally occurs in more subtle ways.

AS3. [J36] CHILD OF: [CYAQRTH, CYNAMF CYNAML, age: AGEIWDATE]
What was [his / her] relationship to [you and your wife / you and HEAD/WIFE/"WIFE" / HEAD and WIFE/"WIFE" / you / your wife / HEAD / WIFE/"WIFE"] before the adoption?

96. None			
301. Son	302. Daughter	331. Stepson	332. Stepdaughter
381. Foster son	382. Foster daughter	401. Brother	402. Sister
471. Brother-in-law	472. Sister-in-law	481. Brother of co-habitor	482. Sister of co-habitor
601. Grandson	602. Granddaughter	651. Great-grandson	652. Great-granddaughter
701. Nephew of Head	702. Niece of Head	711. Nephew of Wife	712. Niece of Wife
740. Cousin of Head	750. Cousin of Wife	831. Son of boyfriend/girlfriend	832. Daughter of boyfriend/girlfriend
950. Relative of Head	960. Relative of Wife	970. Relative of co-habitor	

FIGURE 7:   Gendered Relationship Terms
Source: PSID, 2011, Questionnaire, p. 182.

```
Variable        Tape
Number        Location              Content

  315           569            H60.  Who is that? (1970 question)
 (1415)        (2569)
                                IN ORDER OF PRIORITY

                    1.7         2.  Wife
                    0.9         3.  Child, step-children
                    0.4         4.  Brother or sister
                    0.1         5.  Mother or father
                    0.0         6.  Grandchild; great grandchild
                    0.1         7.  Inlaw or other relative
                    0.0         8.  Non-relative
                    0.0         9.  N.A. relation
                   96.6         0.  Inap., no to H59; one-person family
                   99.8
```

FIGURE 8:   Mixed Gendered and Gender Neutral Terms
Source: PSID, 1970, Study Design, Procedures, Available Data, p. 245.

Gendering through Relationship Terms. Surveys often record information about other people in the respondent's life, and these people are typically labeled using gendered terms. For example, rather than count how many grandchildren a person has, surveys will ask about grandsons and grand-daughters. This occurs for all relatives and other relationships for whom there is a gendered term in English, as illustrated in painstaking detail by the 2011 PSID (Figure 7). This use of gendered relationship terms is a tool for recording a person's sex/gender without having to ask (which, as noted earlier, might be perceived as awkward or rude), but surveys often use gendered relationship terms even when the data cannot be disaggregated by sex/gender (Figure 8). Although the PSID codebook could have used

"sibling" for category 4 and "parent" for category 5, it used instead a pair of gendered terms for each. It is easy to miss the gendering happening here compared to a more explicit sex/gender categorization question, and these moments are rarely, if ever, critiqued by those seeking to improve surveys. Yet, these hidden moments employ the same conception of sex/gender as the more explicit survey items, including being binary, determined by someone other than the person being gendered, seen as obvious, and not allowed to change over time.

This is not to say that there are never options included beyond binaries. Gender-neutral terms for some relatives or other relationships occur frequently, but they are often included in a list of categories along with pairs of gendered terms. For example, in Figure 8, siblings and parents are referred to in gendered terms, while children are not (i.e., the category for children is not called "son or daughter"). It is unclear why the surveys include lists with mixtures of gendered and gender-neutral terms, as they follow no perceptible pattern. Gendering of category labels is unnecessary if the underlying data are undifferentiated, because it cannot contribute to knowledge about any relationship between gender and other life outcomes. Moreover, with the exception of the term "partner" (discussed below), gender-neutral terms are never added to gendered pairs (i.e., there is never a "brother, sister, sibling" nor "son, daughter, child" answer option). Instead, gender-neutral options stand alone, presumably functioning as umbrella terms for pairs of gendered relationship terms.

Gendering through Pronouns. The conception of gender as binary is also scattered throughout these surveys in the form of gendered pronouns. Often, interviewers (and, increasingly, computers) are expected to supply a pronoun based on what is known about a given individual. For example, the 2008 computer-assisted GSS inserted either masculine or feminine pronouns depending on the category recorded for the respondent in the variable "SEX" (Figure 9).[10] Even when the surveys ask about topics unrelated to gender or household composition, respondents and other survey actors are constantly gendered through these binary pronoun fills. Although the practice likely exists to help interviewers and respondents follow the referents in the survey questions, it has the consequence of constraining responses and reinforcing traditional categories. Gendering through pronouns reproduces an understanding that sex/gender is binary and the world is composed of only "hes" and "shes."

```
If SEX = {female} Then
  _spfill1 = {husband}
  _spfill2 = {he}
  _spfill3 = {his}
  _spfill4 = {him}
Elself SEX = {male} Then
  _spfill1 = {wife}
  _spfill2 = {she}
  _spfill3 = {her}
  _spfill4 = {her}
```

FIGURE 9:   Assigning Gendered Pronouns
Source: GSS, 2008, Ballot 1, p. 24.

The frequency of using gendered pronouns varies across surveys, but all are peppered with them, because every real or hypothetical person mentioned in the third person by the interviewer needs, following English grammar rules, some referent. ANES, the longest-running survey, used gendered pronouns more than 22,000 times in the documents we examined—an average of more than 560 per survey-year. This occurs in ubiquitous yet subtle ways, as most interviewers, respondents, and researchers are so used to hearing and using gendered pronouns and gendered relationship terms that they are unlikely to notice this reinforcement of the norm. Thus, overall, these national surveys reflect and produce a hypergendered world—in both explicit and implicit ways, binary, essentialist gendering is everywhere.

Change Over Time

Current question wording often reflects understandings of a concept from the year the survey began, because surveys tend to reuse questions in order to maintain comparability across years, unless subject to external pressure. Over the past several decades, changes in the survey measurement of race and ethnicity, sexuality, and family structure have occurred in response to public challenges (e.g., DaCosta 2007; Powell, Bolzendahl, Geist, and Steelman 2010; Presser 1998). Yet, there has been far less change in national surveys when it comes to measuring sex and gender, perhaps because similar concerted efforts are just beginning to organize. Most of the changes we found were differences in terminology, and many were superficial in the sense that they were not employed consistently and did not affect actual measurement practices. Nevertheless, they are important to highlight, both to recognize the steps surveys have taken and to help point the way to future improvements.

One promising change is the introduction of the term "gender," which signals some attention to contemporary gender theory. The word first appeared in the 1978 ANES and was later included in the PSID (1982), NLSY (1993), and GSS (1996).[11] However, *gender* appears sporadically even in the most recent surveys and is almost always conflated with *sex*. An important step toward the rehabilitation of surveys would be consistently incorporating *gender* and aligning its operationalization with contemporary gender theory.

Another area of improvement is the use of sex/gender-neutral terms, which have the potential to challenge normative binary distinctions. "Spouse" first appeared early in the history of the surveys (ANES 1952), and "partner" was added to all four surveys in the 1970s. This formulation is typical: "During 1978, did (your husband/wife/partner) receive any unemployment compensation?" (NLSY 1979, 155). The most dramatic shift in terminology in this area occurred in the NLSY: As late as 1993, "husband" and "wife" each appeared 388 times in the survey materials; by 2010, they appeared three and five times, respectively, replaced by 129 uses of "partner" and 261 of "spouse."[12] Although the switch to "partner" may have been prompted more by the rise of cohabitation after 1970 (Smock 2000) than a commitment to sex/gender neutrality per se, the use of "partner" or "spouse," particularly when used instead of, rather than in addition to, "husband" or "wife," has the potential to destabilize both sex/gender binaries and the heteronormativity previously reinforced by survey terminology.

A third important change in language involves the use of gendered pronouns: shifting from primarily masculine pronouns to a mix of masculine and feminine pronouns, and occasional attempts to provide sex/gender-neutral alternatives. In the 1948 ANES, across 62 pages of documentation, there were 63 masculine pronouns and just three feminine pronouns—each referring to Germany or Russia as "her" (see pre-election questions 7 and 32). By 2006, the ANES had a more even distribution of masculine and feminine pronouns at 206 and 165, respectively.[13] The GSS was the only survey to explicitly note changing its gendered terminology. In 1984, the GSS reworded a question about traits "most desirable for a child." The old question prompted respondents to think only of a boy (e.g., "that he is honest"), while the new version offered "gender-neutral" language (GSS 2011, 3314). Yet, even in the revised version, normative understandings of sex/gender remained; item 9 continued to gender the hypothetical child as masculine and item 7 asked about gender conformity in binary terms (see Figure 10).

HAND CARD I		Most Desirable			Least Important	
		A. Three Most (CODE THREE)	B. One Most (CODE ONE)		C. Three Least (CODE THREE)	D. One Least (CODE ONE)
1.	that a child has good manners.	2	1		4	5
2.	that a child tries hard to succeed.	2	1		4	5
3.	that a child is honest.	2	1		4	5
4.	that a child is neat and clean.	2	1		4	5
5.	that a child has good sense and sound judgment.	2	1		4	5
6.	that a child has self-control.	2	1		4	5
7.	that he acts like a boy or she acts like a girl.	2	1		4	5
8.	that a child gets along well with other children.	2	1		4	5
9.	that a child obeys his parents well.	2	1		4	5
10.	that a child is responsible.	2	1		4	5
11.	that a child is considerate of others.	2	1		4	5
12.	that a child is interested in how and why things happen.	2	1		4	5
13.	that a child is a good student.	2	1		4	5

FIGURE 10:   Partial Gender Neutrality
Source: GSS, 1984, Questionnaire, p. 31.

It is clear from these examples that surveys can and do change. However, the changes thus far seem sporadic and ad hoc rather than concerted attempts to align surveys with contemporary gender theory and lived experiences. Indeed, some changes actually widen the gap between theory and methods, as when the PSID and NLSY both switched from recording annual categorizations of sex/gender to treating sex/gender as static by carrying it forward from a single previous year. As surveys continue to evolve, a systematic effort should be made to assess whether the incorporation of sex and gender, in both direct measurement and terminology, is justified by specific research goals. We do not advocate removing sex and gender from surveys in their entirety, as that would make it impossible to study their role in inequality; rather, we encourage an acknowledgment of sex and gender fluidity

and diversity within, as well as beyond, traditional categories. The occasional moments of sex/gender-neutrality we observed demonstrate how easy it would be to change binary language in these texts when its inclusion does not further knowledge. Change—even to large-scale, long-running surveys—is possible.

CONCLUSION

Gender scholars cannot simply dismiss survey research, as many have in the past; a thorough examination of the operationalization of sex and gender in U.S. social surveys is overdue. Even the most recent scholarship, focused on inclusion for transgender respondents, critiques only direct questions about the respondent's sex and gender. These efforts point to important avenues for improvement, including asking separate questions about sex at birth and current gender identity, expanding the number of answer options beyond "male" and "female," and acknowledging diversity in gender expression (GenIUSS Group 2014). However, our findings indicate that, while these changes will help surveys better represent the diversity of lived experience, they do little to close the gap between contemporary gender theory and survey practices. To do so, surveys must also consistently distinguish between the concepts of sex and gender, incorporate self-identified sex and gender, provide clear criteria or instructions for how to determine sex and gender, acknowledge change in sex and gender over the life course, and rethink the necessity of employing binary sex and gender categories throughout the survey materials.

Making these additional changes should allow for a clearer understanding of the sex/gender system and improve the effectiveness of research working to reduce inequality. For example, recording variation over time for both sex and gender will facilitate research into both the prevalence and correlates of sex and gender fluidity over the life course. We hesitate to outline more specific guidelines for the rehabilitation of surveys, in part because we aim to encourage rather than constrain creativity in designing new methods of measurement. New ways of operationalizing sex and gender also need to be thoroughly tested before being widely adopted.

We do not expect survey measures to be perfect, nor do we hold them to a higher standard than we would other research methods. But the influence of data from large-scale social surveys in undeniable, and the gap

between our theory and our methods yawns more widely for sex and gender than for many other core social science concepts. We will never know what we may have sacrificed, in terms of deeper understanding of inequality and the sex/gender system, until we address these enduring problems.

NOTES

1. Cisgender people identify with the same sex and gender categories to which they were assigned at birth (Schilt and Westbrook 2009).

2. There is a long-running related debate in psychology about how to measure "masculinity" and "femininity" (Auster and Ohm 2000; Constantinople 1973), but an underlying assumption of binary sex categorization remains, because these traits are generally considered characteristic of people who are easily distinguished as males and females (cf. Spence 2011).

3. See http://www.norc.org/research/projects/pages/general-social-survey.aspx (accessed May 17, 2014).

4. We use "gendering" for all instances when sex and/or gender categories, including pronouns, are applied to the respondent or others within the survey.

5. Given our strategy for selecting excerpts, we expect that these figures underestimate the hidden gendering in these surveys, because we coded every direct categorization of the respondent's sex or gender but extracted only unique examples of other genderings (i.e., we did not extract and code every reference to he/she, brother/sister, or husband/wife in the text).

6. Surprisingly, these specialty surveys also use wording that conflates sex and gender. For example, the GenIUSS Group (2014, v.) recommends using "male" and "female" as answer options for "current gender identity" questions.

7. The 2010 ANES was a panel recontact study that did not remeasure the respondent's sex/gender. In 2012 (which was not part of our original data set), the ANES reverted to interviewer determination of the respondent's sex/gender "by observation"; although interviewers could mark "DK" (don't know) if they were "unsure." If the interviewer did not enter a sex/gender classification, respondents were later asked whether they were "male or female" in the computer-assisted self-interviewing portion of the survey.

8. This logic quickly becomes circular: the question will seem "awkward" as long as surveys avoid asking it and, when they do ask, apologize for doing so.

9. This number (299 of 1,361) is greater than the total number of survey-years because we collected variable frequencies along with questionnaires, and respondents are occasionally gendered more than once within a survey (such as when

interviewers are instructed to skip a series of questions for particular types of respondents, such as pregnancy histories for males).

10. Note, too, the heteronormativity in this excerpt; in it, only females have husbands and only males have wives.

11. The first use in the ANES described a code for open-ended responses given to a question about reasons for liking or disliking a political candidate ("0461. Gender e.g. 'She's a woman'").

12. By the mid-1980s, "partner" is used frequently in all of the surveys except for the PSID, which uses the term 12 times total.

13. Interestingly, in the 1980s and 1990s, the PSID included many more feminine pronouns, as it assumed a male respondent and had questions about the respondent's wife or heterosexual partner written in the third person.

REFERENCES

American National Election Study. http://www.electionstudies.org/studypages/download/datacenter_all_NoData.php.

Auster, Carol, and Susan Ohm. 2000. Masculinity and femininity in contemporary American society. *Sex Roles* 43:499-528.

Balarajan, Meera, Michelle Gray, and Martin Mitchell. 2011. *Monitoring equality: Developing a gender identity question*. London: Equality and Human Rights Commission.

Bauer, Greta R., Rebecca Hammond, Robb Travers, Matthias Kaay, Karin M. Hohenadel, and Michelle Boyce. 2009. "I don't think this is theoretical; this is our lives": How erasure impacts health care for transgender people. *Journal of the Association of Nurses in AIDS Care* 20:348-61.

Berger, Peter, and Thomas Luckmann. 1967. *The social construction of reality*. New York: Doubleday.

Bird, Chloe, and Patricia Rieker. 2008. *Gender and health*. Cambridge, UK: Cambridge University Press.

Bowker, Geoffrey, and Susan Star. 2000. *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Brewster, Karin, and Irene Padavic. 2000. Change in gender-ideology, 1977–1996. *Journal of Marriage and Family* 62:477-87.

Cavanagh, Sheila. 2010. *Queering bathrooms*. Toronto, Ontario, Canada: University of Toronto Press.

Connell, Raewyn. 1995. *Masculinities*. Oakland: University of California Press.

Constantinople, Anne. 1973. Masculinity-femininity: An exception to the famous dictum? *Psychological Bulletin* 80:389-407.

Cowan, Sharon. 2005. Gender is no substitute for sex. *Feminist Legal Studies* 13:67-96.

Cruz, Taylor. 2014. Assessing access to care for transgender and gender nonconforming people. *Social Science & Medicine* 110:65-73.

DaCosta, Kimberly. 2007. *Making multiracials*. Stanford, CA: Stanford University Press.

Fausto-Sterling, Anne. 2000. *Sexing the body*. New York: Basic Books.

Garfinkel, Harold. 1967. *Studies in ethnomethodology*. Upper Saddle River, NJ: Prentice Hall.

Gates, Gary. 2011. *How many people are lesbian, gay, bisexual and transgender?* Los Angeles: The Williams Institute.

General Social Survey. http://www3.norc.org/gss+website/publications/documentation.

GenIUSS Group. 2014. Best practices for asking questions to identify transgender and other gender minority respondents on population-based surveys. Los Angeles: Williams Institute.

Halberstam, Judith. 1998. *Female masculinity*. Durham, NC: Duke University Press.

Hammarstrom, Anne, and Ellen Annandale. 2012. A conceptual muddle: An empirical analysis of the use of "sex" and "gender" in "gender-specific medicine" journals. *PLoS One* 7:e34193.

Harnois, Catherine. 2013. *Feminist measures in survey research*. Thousand Oaks, CA: Sage.

Harrison, Jack, Jaime Grant, and Jody Herman. 2011. A gender not listed here. *LGBTQ Policy Journal* 2:13-24.

Hox, Joop. 1997. From theoretical concept to survey question. In *Survey measurement and process quality*, edited by Lars Lyberg, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin. Hoboken, NJ: Wiley-Interscience.

Human Rights Watch. 2011. *Human rights violations against trans people in the Netherlands*. New York: Human Rights Watch.

Igo, Sarah. 2007. *The averaged American*. Cambridge, MA: Harvard University Press.

Institute of Medicine. 2011. *The health of lesbian, gay, bisexual, and transgender people*. Washington, DC: National Academies Press.

Krosnick, Jon, and Stanley Presser. 2010. Question and questionnaire design. In *Handbook of survey research*, edited by Peter Marsden and James Wright. Bingley, UK: Emerald Group.

Lorber, Judith. 1993. Believing is seeing: Biology as ideology. *Gender & Society* 7:568-81.

Lucal, Betsy. 1999. What it means to be gendered me. *Gender & Society* 13:781-97.

Meyerowitz, Joanne. 2002. *How sex changed*. Cambridge, MA: Harvard University Press.

Mills, Sara. 1997. *Discourse*. Florence, KY: Routledge.

National Longitudinal Survey of Youth, 1979. https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation.

Neuendorf, Kimberly. 2002. *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Nicholson, Linda. 1994. Interpreting gender. *Signs* 20:79-105.

Oakley, Ann. 1998. Gender, methodology and people's ways of knowing. *Sociology* 32:707-31.

Oakley, Ann, and Robin Oakley. 1979. Sexism in official statistics. In *Demystifying social statistics*, edited by John Irvine, Ian Mills, and Jeff Evans. London: Pluto Press.

Panel Study of Income Dynamics. https://psidonline.isr.umich.edu/guide/default.aspx.

Powell, Brian, Catherine Bolzendahl, Claudia Geist, and Lala Carr Steelman. 2010. *Counted out: Same-sex relations and Americans' definitions of family*. New York: Russell Sage Foundation.

Presser, Harriet. 1998. Decapitating the U.S. Census Bureau's "head of household." *Feminist Economics* 4:145-58.

Pryzgoda, Jayde, and Joan Chrisler. 2000. Definitions of gender and sex. *Sex Roles* 43:553-69.

Randall, Sara, Ernestine Coast, and Tiziana Leone. 2011. Cultural constructions of the concept of household in sample surveys. *Population Studies* 65: 217-29.

Ridgeway, Cecilia. 2011. *Framed by gender*. Oxford: Oxford University Press.

Rubin, Gayle. 1975. The traffic in women. In *Toward an anthropology of women*, edited by Rayna Reiter. New York: Monthly Review Press.

Saperstein, Aliya. 2012. Capturing complexity in the United States: Which aspects of race matter and when? *Ethnic and Racial Studies* 35:1484-1502.

Saperstein, Aliya, Laurel Westbrook, and Devon Magliozzi. 2015. Gender gradations: Measuring sex and gender diversity in surveys. Paper presented at the Population Association of America Annual Meeting, San Diego, May 1.

Schaeffer, Nora, and Jennifer Dykema. 2011. Questions for surveys. *Public Opinion Quarterly* 75:909-61.

Schaeffer, Nora, and Stanley Presser. 2003. The science of asking questions. *Annual Review of Sociology* 29:65-88.

Schilt, Kristen, and Jenifer Bratter. 2015. From multiracial to transgender? Assessing attitudes toward expanding gender options on the U.S. Census. *Transgender Studies Quarterly* 2:77-100.

Schilt, Kristen, and Laurel Westbrook. 2009. Doing gender, doing heteronormativity. *Gender & Society* 23:440-64.

Settersten, Richard, Jr., and Karl Mayer. 1997. The measurement of age, age structuring, and the life course. *Annual Review of Sociology* 23:233-61.

Smock, Pamela. 2000. Cohabitation in the United States. *Annual Review of Sociology* 26:1-20.

Snipp, C. Matthew. 2003. Racial measurement in the American census. *Annual Review of Sociology* 29:563-88.

Spade, Dean. 2011. *Normal life: Administrative violence, critical trans politics, and the limits of the law*. Boston: South End Press.

Spence, Janet. 2011. Off with the old, on with the new. *Psychology of Women Quarterly* 35:504-9.

Sprague, Joey. 2005. *Feminist methodologies for critical researchers*. Lanham, MD: AltaMira Press.

Stacey, Judith, and Barrie Thorne. 1985. The missing feminist revolution in sociology. *Social Problems* 32:301-16.

Titscher, Stefan, Michael Meyer, Ruth Wodak, and Eva Vetter. 2000. *Methods of text and discourse analysis: In search of meaning*. Thousand Oaks, CA: Sage.

Undurraga, Rosario. 2010. How quantitative are feminist research methods textbooks? *International Journal of Social Research Methodology* 13:277-81.

Urla, Jacqueline. 1993. Cultural politics in an age of statistics. *American Ethnologist* 20:818-43.

Valentine, David. 2007. *Imagining transgender*. Durham, NC: Duke University Press.

West, Candace, and Don Zimmerman. 1987. Doing gender. *Gender & Society* 1:125-51.

Westbrook, Laurel, and Kristen Schilt. 2014. Doing gender, determining gender: Transgender people, gender panics, and the maintenance of the sex/gender/sexuality system. *Gender & Society* 28:32-57.

Wickes, Rebecca, and Michael Emmison. 2007. They are all doing gender but are they all passing? *The Sociological Review* 55:311-30.

Williams, Christine. 2006. Still missing? Comments on the twentieth anniversary of "the missing feminist revolution in sociology." *Social Problems* 53:454-58.

Zerubavel, Eviatar. 1996. Lumping and splitting: Notes on social classification. *Sociological Forum* 11:421-33.

Laurel Westbrook is an associate professor of sociology at Grand Valley State University. Her research focuses on the inner workings of the sex/gender/sexuality system, including how knowledge production, gendered violence, and social movements both maintain and alter the system. She is currently completing a book on "identity-based anti-violence activism."

Aliya Saperstein is an assistant professor of sociology at Stanford University whose work examines how categories of difference are operationalized in survey research and the consequences of these methodological decisions for studies of stratification. Previous research highlighted how fluidity in racial categorization reinforces racial inequality in the United States.