ENGAGING THE DEBATE OVER
“BIOLOGICAL LIMITS OF GENDER CONSTRUCTION”:
APPLYING CONCEPTS FROM RESEARCH METHODS TO PUBLISHED
SOCIOLOGICAL RESEARCH*

Michael Francis Johnston, Melinda Allen, M. Quinn Ottaway, Jennifer Lutrick,
Katie Massey, Latasha McCrary, and Daphne Ohlson
University of California, Los Angeles
Johnston@ucla.edu

Abstract. Enlivening research methods classes is a challenge faced by many instructors. Theory from the Sociology of Teaching and Learning (SoTL) suggests that incorporating published sociological debates into course content provides a useful, practical method of enlivening such classes. In this article, use of a controversial debate published in the American Sociological Review making the claim that prenatal hormone exposure influences adult female gendered behavior is discussed. Concepts from an introductory research methods course are used to frame and assess methodological issues in the sociological debate. Student self-evaluations suggest that the exercise improved their familiarity with and ability to employ methodological concepts to assess the strengths and weaknesses of quantitative survey research evidence.

* Michael Francis Johnston conceptualized, drafted, and revised the paper. Melinda Allen, Jennifer Lutrick, Latasha McCrary, Katie Massey, Daphne Ohlson, and M. Quinn Ottoway (listed alphabetically) contributed written parts to the paper and assisted in revisions. We benefited from discussions with classmates: Nori Bryant, Kantra Hall, Karen Honderich, Morgan Morgan, Jason Pitman, Lori Rounsavall, and Jennifer Taylor. We especially thank Elaine Pruett for stimulating us to think more thoroughly about the debate. Please direct all correspondence to: Michael Francis Johnston, David Geffen School of Medicine, University of California, Los Angeles, johnston@ucla.edu.
Successful completion of a course in research methods is required of most undergraduate sociology majors (Schutt, Wagenaar, and Mulvey 1987; Longmore, Dunn, and Jarboe 1996). Arguably, instructors should pay extra attention to survey research in these classes, since this is sociologists’ primary methodological tool (Singleton, Straits, and Straits 1993). And yet, if an instructor introduces too many survey-oriented research concepts in lectures and assigned readings, students may “glaze over” in trying to respond to the overwhelming detail. One solution is to structure the content on survey research so that students begin at the level of learning in terms of memorizing facts and then progress to learning at the level of application (Bloom et al. 1956). Our purpose in this paper is to describe a sociological debate that provides a very useful forum for applying textbook survey research concepts.

In a 2000 issue of the *American Sociological Review*, Udry articulated the sociobiological claim that prenatal hormone experience influences women’s adult gendered behavior and then provided statistical analyses of longitudinal data that he claimed supported the argument (Udry 2000). The next year, the *American Sociological Review* published critiques from several scholars (Miller and Costello 2001; Kennelly, Merz, and Lorber 2001; Risman 2001), along with responses by the author and editor (Udry 2001; Firebaugh 2001). In a Spring 2005 research methods class for undergraduates at the University of Alabama in Huntsville, an assistant professor and students applied textbook concepts to this attention-grabbing sociological debate concerning sampling, survey research, inquiry, paradigms and social research, the politics of research, and research design and conceptualization (Babbie 2004).

To use this material in class, the instructor organized discussions about various methodological critiques, lectured about the strengths and weaknesses of each, and then asked students to write a series of four position papers on various aspects of the sociological debate. (See Appendix A for a sample assignment). Although students found the material challenging in the beginning, they self-reported that the exercise improved their familiarity with and ability to employ methodological concepts to assess the strengths and weaknesses of quantitative survey research. In this paper, we review the article and debate, and then illustrate topics to which concepts from research
methodology may be fruitfully applied to assess the debate. Subsequently, we return to student learning outcomes in the conclusion.

OVERVIEW OF UDRY’S ORIGINAL ARTICLE AND RESULTING DEBATE

A Biosocial Interaction Model portrays gendered behavior (behavior that differs by sex) as being influenced by two distinct mechanisms: biological predisposition and socialization (Udry 2000:443). Biological predispositions originate from exposure to hormones during the second trimester of the prenatal stage. The claim is that female fetuses exposed to above-average androgen levels will later display more masculine behavior. The model’s second mechanism is socialization, whereby parents, peers, and other members of society condition women to act more feminine than men.

In “Biological Limits of Gender Construction,” Udry (2000:443) develops and tests the hypothesis that women’s prenatal hormonal exposure limits the impact of socialization on gendered behavior. He finds support for his hypothesis, stating: “the models predicting gendered behavior show that a high prenatal androgenization of females not only masculinizes their gendered behavior predispositions at later ages, but immunizes them against socialization toward typical feminine behavior” (Udry 2000:452).

Udry’s article provoked a wave of commotion among sociologists of gender. Critics’ published essays challenged Udry’s conceptualization and operationalization, his use of outdated theories, his comparisons of humans to animals, and the sociopolitical implications he drew from his findings (Miller and Costello 2001; Kennelly et al. 2001; Risman 2001). Udry (2001) responded by clarifying his purpose, measures, and findings, and then characterized most of the uproar as a result of misunderstanding and terminological differences. Firebaugh (2001), the editor who made the decision to publish “Biological Limits of Gender Construction,” provides the last published words in the debate by describing the review process and why he decided to publish the manuscript.

Even a cursory reading of the debate shows that participants and manuscript reviewers had very different opinions about the quality of Udry’s data and methods of research and analysis. Sorting through these issues was somewhat daunting for the
undergraduates, especially because this published debate among sociologists was carried out in the *American Sociological Review* and involved sophisticated statistical models. In class, we distilled and expanded upon the themes presented in the debate at a level more consistent with that of an undergraduate research methods class, using Babbie’s (2004) research methods textbook as a reference. In the next several sections, we discuss these themes in terms of topics conventionally presented in a research methods class: sample quality, research design, data collection and measurement, conceptualization, and operationalization.

**ASSESSING SAMPLE QUALITY**

Sample quality is usually assessed according to how representative the sample is. One important mechanism for increasing the likelihood that a sample is representative is to use probability sampling. Babbie (2004:189-191) describes probability sampling as drawing a sample in which all members from a precisely specified population have an equal chance of being selected.

Udry does not clearly specify his study population. He explains that he began with a collection of respondents from the Child Health and Development Study (CHDS): “Pregnant women presenting for prenatal care at Kaiser Plan facilities in the San Francisco Bay area” from 1960 to 1963 (Udry 2000:446-447). He clearly states that nonwhites were excluded. At one point he suggests that participating women (that is, the adult daughters of these then-pregnant women) had parents of a higher socioeconomic status, though shortly thereafter he asserts that participating and nonparticipating eligibles did not differ from each other (Udry 2000:447). Our best guess about what constitutes his study population is white women born in the San Francisco Bay area from 1960 to 1963.

Udry’s sample was collected using a type of nonprobability sampling known as “reliance on available subjects.” Babbie (2004) points out: “Clearly, this method does not permit any control over the representativeness of a sample. It’s justified only [in certain circumstances]. Even when this method is justified on the grounds of feasibility, researchers must exercise great caution in generalizing from the data. Also [researchers]
should alert readers to the risks associated with this method” (p. 183). Nowhere in the article does Udry talk about the limitations of his sample for making generalizations.

Another important mechanism for increasing the likelihood of representativeness has to do with the response rate: the higher the response rate, the higher the likelihood of representativeness (Babbie 2004:261). Udry’s sample has a surprisingly small response rate (Udry 2000:447). The number of individuals who provided necessary data is only 35 percent of the 470 people who fit sample eligibility requirements. According to the standards for acceptable response rates as discussed by Babbie (2004:261-263), percentage is well below the minimum 50 percent rate that would be adequate for report and analysis.

Despite the fact his survey data was collected via “reliance on available subjects” and the response rate was only 35 percent, Udry writes as if he is convinced that his sample is representative, at least of women “living in the same culture at a particular time” (see Udry 2001:611). According to the principles commonly taught in undergraduate research methods concerning sampling, we conclude there is no reason to assume that the sample of 163 women is representative of adult white women born in the San Francisco Bay area between 1960 and 1963.

ASSESSING RESEARCH DESIGN

Udry argues that prenatal hormonal levels explain some of the variation in gendered behavior among women, which is clearly an explanatory research problem (Babbie 2004:89-90). Udry’s approach is that of nomothetic causality, which is premised on three criteria: “(1) the variables must be correlated, (2) the cause takes place before the effect, and (3) the variables are nonspurious” (p. 90).

Udry’s data were collected via a longitudinal study design at three points in time. First, blood draws from pregnant women were taken to provide proxy measures of fetal prenatal androgen exposure and levels of the inhibitor sex hormone-binding globulin (Udry 2000:446). Second, when these women’s daughters had become teenagers (15-17 years old), the daughters were asked whether time with their families would be important to them ten years later (p. 451-452). Third, when these daughters had become adults (27-
30 years old), they completed a range of questions about their gendered behavior and provided blood samples that were assayed for androgens (p. 446).

Udry (2000:447-452) shows correlations between the prenatal hormonal measures and adult gendered behavior in a series of tables. Clearly, the cause (hormones) takes place before the effect (gendered behavior). Further, it is difficult to dismiss this time-ordered correlation as spurious. The agreed-upon method of establishing spuriousness is to include other variables in an analysis that changes the original relation from correlation to noncorrelation. Udry does include variables such as adult hormonal measures and childhood socialization measures in further analyses, yet the original correlation between prenatal hormone exposure and adult gendered behavior remains.

According to principles taught in an undergraduate research methods class, Udry’s research design enables a claim that hormones exert a causal influence on the gendered behavior of women. The force of Udry’s writing provides an opportunity to explore in greater depth issues of causal sociological claims that have to do with reductionism (Babbie 2004:101) and the fallacy of complete causality (p. 92). This plays out in the critique by Miller and Costello (2001:592-594), who paint Udry’s study as one of “neuroendocrinological determinism,” and the response by Udry, who acknowledges that prenatal hormonal exposure may only explain a small amount of variance in gendered behavior (Udry 2001:614-617).

ASSESSING DATA COLLECTION AND MEASUREMENT

The independent variables are mostly biological data, which are very objective. From Udry’s (2000:446) description, this particular set of biological data appears to be of high quality, with absolutely no reason to expect that it is biased so as to produce predicted correlations with gendered behavior. However, one of the key independent variables, female gender socialization (p. 450), is based on respondent reports about their childhoods. Reports about behavior that occurred years ago are generally not considered to be very reliable because of recall bias (Babbie 2004:102).

The dependent variable, gendered behavior, was constructed from data collected in a high-quality fashion. Survey interviews have a tendency to be biased because of social desirability effects in the data collection process (Babbie 2004:250). Respondents
answer questions in ways that they think will please the interviewer (pp. 256-263, especially 273). However, for this research, respondents filled out the questionnaires in a private room alone without an interviewer present (Udry 2000:447). Such an approach thereby decreases the social desirability effect. Another reason that the data used to construct the dependent variable are of high quality is that respondents answered standardized questionnaires about their gendered behavior, including the Bem Sex Role Inventory, the Adjective Check List, and the Personality Research Form (p. 448). There is one glaring exception to the overall theme of high-quality data being used to construct the dependent variable, however: interviewers rated respondents according to the femininity of their appearance. Several of the critics questioned the subjectivity of this technique. In a word, the debate over the dependent variable provides an excellent opportunity to discuss the reliability of survey questions (Babbie 2004:274).

ASSESSING UDRY’S CONCEPTUALIZATION AND OPERATIONALIZATION OF GENDER

Udry theoretically deduces his Biosocial Interaction Model and deductively operationalizes the hormonal measures. However, he uses a statistical technique (principal-components factor analysis) to inductively develop one (continuous) dependent variable from ten different scales and several additional questionnaire items (Udry 2000:448, 613). Although his dependent variable includes measures of attitudes, personality, and behavior, Udry opts for the simple label of “gendered behavior” (p. 611). Contrasting the way that Udry conceptualizes and operationalizes the hormonal and gendered behavior variables provides an excellent opportunity to review the elements of inductive and deductive theory construction (Babbie 2004:44-56).

One of the strengths of Udry’s inductive approach is that he uses a wide variety of indicators (Udry 2001:611). This approach accords well with basic principals from research methods: composite measures based on indicator variables. Babbie (2004) writes: “concepts are subject to varying interpretations—each with several possible indicators. In these cases, you’ll want to make several observations for a given variable. You can then combine the several pieces of information you’ve collected, creating a composite measurement of the variable in question” (p. 138).
Udry’s inductive approach sorts gendered behavior into four dimensions: importance of home, feminine interests, job status, and masculinity/femininity (Udry 2000:448). Each dimension consists of multiple indicators. Many of the gender scholars (Miller and Costello 2001; Kennelly et al. 2001; Risman 2001) took issue with Udry because his conceptualization of gender did not include a dimension explicitly accounting for gender stratification in terms of power, status, and inequality. They also took issue with several of his specific indicators within the four dimensions, such as importance of family and job status. One way to rephrase these critiques is to state that Udry’s conceptualization insufficiently covers the range of social factors that constitute gender, that is to say, there is a problem with content validity (Babbie 2004:145).

Tensions between Udry and critics raise the issue, in a form slightly different than that described in the textbook, of: “Who decides what’s valid?” On the one hand, gender scholars argue that by failing to include power and inequality as a dimension of gender in his inductive conceptualization, Udry substantially distorts his composite measure of gendered behavior. On the other hand, Udry claims to be engaged in explanatory (not exploratory) research, for which precise agreed-upon definitions are less important (Babbie 2004:129-131), and his measure of gendered behavior draws upon a very wide variety of indicators (p. 138).

DISCUSSION

Two weeks following conclusion of the several class periods devoted to the debate on biological limits of gender construction, students self-reported mastering a range of methodological concepts. Thirteen of the 15 registered students completed an individual, anonymous survey questionnaire while the instructor was outside the classroom (two were absent). The questionnaire asked students to indicate their familiarity with survey research concepts on a scale of 1 to 5, with the values assigned the following meanings: (1) never heard of the concept; (2) heard of the concept, but don’t know what the concept means; (3) know what the concept means, but not how to use the concept in a conversation; (4) could use the concept in a conversation but not confidently; and (5) could confidently use the concept in a conversation. Students reported for the concept of “representativeness (of sample)” an average score of 4.7; for
the concept of “random sampling” an average score of 4.9; for “response rate” an average score of 4.5; for the “social desirability effect” an average score of 5; for “reliability (of survey questions)” an average score of 4.5; for “indicators (for operationalizing a concept)” an average score of 4.1; for “longitudinal study” an average 4.2; for “causality” an average score of 4.3; and for “spurious relationships” an average score of 3.7. In sum, student self-evaluations suggest that the module facilitated mastery of a diverse range of methodological concepts. This was corroborated by instructor assessment of classroom discussion and student position papers on the debate.

Although students enthusiastically engaged material in the module, the instructor had to work to convert this energy to productive and analytical thinking. Many of the students were predisposed to look unfavorably on the research because they had completed a course in gender that critiqued similar material. Another obstacle was that the articles more than fulfilled Light’s (2001) call for “substantive content”—students were overwhelmed by the vast opportunities to develop critiques. In fact, some student papers merely listed multiple methodological concepts that could be used to assess Udry’s research, without actual assessment. For this reason the instructor found it necessary to repeatedly focus students’ attention on developing two to three methodological critiques in their position papers. Similarly, in initial classroom discussions, students were so excited by the opportunity to engage Udry that they limited themselves to naming critiques without applying them to the published research. To deal with this, the instructor found it very helpful to identify specific critiques with reference to the textbook (i.e., let’s open up the textbook and talk more in depth about this idea) and then collectively and professionally discuss them in depth.

The survey results imply that students increased their ability to critically apply, in both written and oral mediums, textbook information about research methodology concepts to a controversial sociological debate. In addition, student coauthors observed that the exercise strengthened their skills in identifying complex methodological issues in sociological debates, understanding the ways in which methodological problems shape substantive outcomes, framing evidential problems in professional language, and discerning the relative strength of different sociological critiques. A more thorough testing of the sophisticated outcomes implied by coauthors’ observations remain to be
examined in a more rigorous format that includes a pretest, survey questions testing knowledge, and accumulation of results over several semesters, with attention to the extent to which students are able to apply skills learned in this debate to other studies.

APPENDIX A: EXAMPLE OF A POSITION PAPER ASSIGNMENT

Now that you have finished reading the original Udry article, critiques, and rejoinder, please state your position on one of the overarching issues raised by Udry. Identify and then describe two or three methodological reasons that inform your stance. In doing so, reference specific page numbers from the articles and textbook. When linking the methodological reason to the textbook, do not stop at merely listing a page number; instead, describe the methodological reason and why you find this to be applicable and persuasive.

Here’s a short example based upon our discussion today:

Udry argues that prenatal hormonal experience dampens the effects of socialization on gendered behavior (Udry 2000:443). I am not persuaded by his argument because I consider the data he uses to be flawed in three ways, that there are problems with construct validity (Babbie 2004:144), representativeness (p. 189), and interviewer-induced bias (p. 264). [Then go on to describe each of these issues in a separate paragraph.]

Please do feel free to argue that you are persuaded to take his conclusions seriously or that you find some methodological reasons persuasive but others problematic. In any case, the paper should be no longer than two double-spaced pages.

REFERENCES


