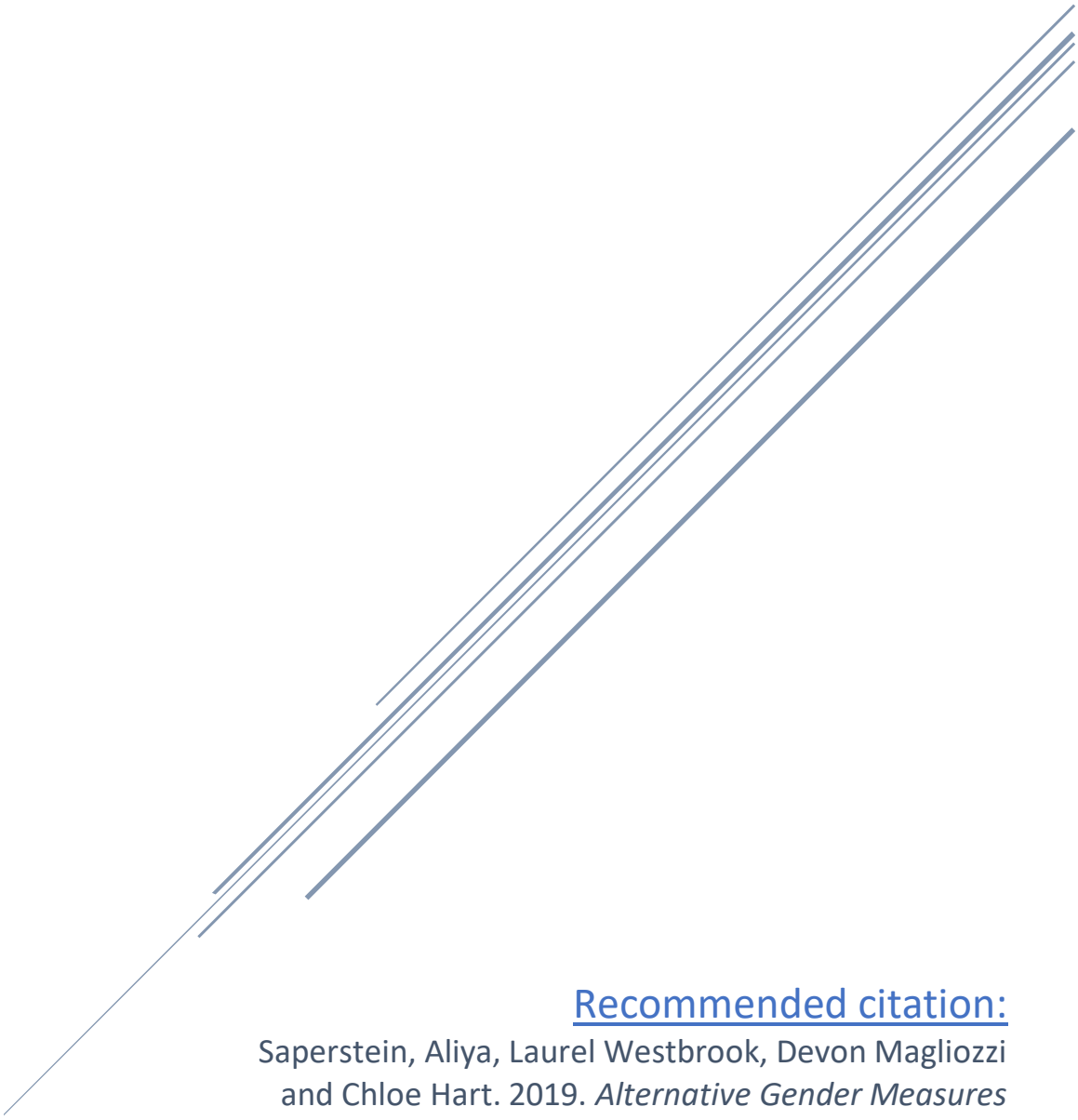
ALTERNATIVE GENDER MEASURES SURVEY

User's Guide and Codebook



Recommended citation:

Saperstein, Aliya, Laurel Westbrook, Devon Magliozzi and Chloe Hart. 2019. *Alternative Gender Measures Survey: User's Guide and Codebook*. Stanford, CA: Clayman Institute for Gender Research.

Table of Contents

Overview

The Alternative Gender Measures Survey (AGMS) was conducted on Amazon Mechanical Turk in November 2014 to pilot measures as part of a proposal for a new sex and gender module in the General Social Survey (GSS). In order to test the measures in a realistic context, a six-item sex and gender module was embedded in a questionnaire designed to approximate conventional questionnaires in the social sciences. Most of the items were drawn from the GSS and have the same question wording and variable names. All other items are noted in the codebook below.

The AGMS sex and gender module includes:

- a pair of feminine and masculine scales asked from the first-order perspective (how do you see yourself?)
- a pair of feminine and masculine scales asked from the third-order perspective (how do most people see you?)
- a two-step categorical measurement of sex at birth and current gender identity.

This two-step approach is the same one adopted by the GSS for its 2018 survey.

Variables derived from combinations of survey questions, other summary variables, and recodes are presented at the end of the codebook. Text in black indicates verbatim question wording from the survey; text in grey provides contextual information about the variables and describes variables that were derived from the survey items.

For more information about the design of the AGMS, please see the Appendix for our full set of proposals to the GSS, which include background on a previous pilot survey comparing sex and gender measures with binary and non-binary answer options. For additional details and published findings from the AGMS, see Magliozzi, Saperstein, and Westbrook (2016) and Hart, Saperstein, Magliozzi and Westbrook (2019).

Survey Design

Questions were asked in the order in which they are presented in the codebook, with the exception of the sex and gender items. The ordering of these items was randomly assigned depending on the participant's survey condition (as indicated by the variables **module_end**, **scales_first**, **self_first**, and **condition**).

As in the GSS, respondents were limited to U.S. residents aged 18 years or older. Responses were dropped from the dataset if: the IP address indicated they were answered from outside of the United States (N=31), the survey was submitted a second or third time from the same IP address (N=53) or MTurk ID (N=11), they were not correctly assigned to a single survey condition (N=59), or they dropped out of the survey before seeing the sex and gender module (N=4). In total, 142 cases were dropped. The survey yielded 1,537 valid responses.

SURVEY ADMINISTRATION VARIABLES

case_id
[ID number for each respondent]

surveytime
[How long it took each participant to complete the survey, in seconds]

module_end
[Placement of the sex and gender items within the survey, where 0 = displayed after cohort and 1 = displayed after sexsex18.]

scales_first
[Whether both sets of gender scales were shown before or after the sex and categorical gender items, where 0 = shown after and 1 = shown before.]

self_first
[Whether the first-order gender scales were shown before or after the third-order gender scales, where 0 = shown after and 1 = shown before.]

condition
[Which question order condition respondents were placed in, i.e., all combinations of the module_end, scales_first, and self_first conditions.]

sex_submit
[How long respondents spent on the page with the sex at birth item, in seconds.]

sex_clicks
[How many times respondents clicked around on the page with the sex at birth item.]

gender_submit
[How long respondents spent on the page with the current gender item, in seconds.]

gender_clicks
[How many times respondents clicked around on the page with the current gender item.]

firstord_submit
[How long respondents spent on the page with the pair of first-order gender scales, in seconds.]

firstord_clicks
[How many times respondents respondent clicked around on the page with the first-order scales.]

thirdord_submit
[How long respondents spent on the page with the pair of third-order gender scales, in seconds.]

thirdord_clicks
[How many times respondents clicked around on the page with the third-order scales.]

Survey Items

Items are shown in the order in which they were displayed in the survey

[natenviy-natfarey]
Note: These items were asked on the same page in a matrix format.
We are faced with many problems in this country, none of which can be solved easily or inexpensively. Some of these problems are listed below. For each one, please indicate whether you think we are spending too much money on it, too little money, or about the right amount.

	Spending too much (1)	Spending too little (2)	Spending the right amount (3)
	○	○	○

natenviy
The environment

natcrimy
Law enforcement

nateducy
Education

natarmsy
National defense

natsci
Scientific research

natfarey
Assistance for the poor

discaffm
What do you think the chances are these days that a man won't get a job or promotion while an equally or less qualified woman gets one instead:

Is this very likely, somewhat likely, or not very likely these days?

- ○ Very likely (2)
- ○ Somewhat likely (1)
- ○ Not very likely (0)

posslqy
Which of these statements applies to you?

- ○ I am married and living in the same household as my spouse (1)
- ○ I have a steady partner and we live in the same household (2)
- ○ I have a spouse or steady partner, but we don't live in the same household (3)
- ○ I don't have a steady partner (0)

[wrkstat_fulltime-wrkstat_other_text]
Note: these items are related to the GSS variable wrkstat; each response is recorded as a binary variable because respondents were allowed to select more than one answer.
Last week, were you working full time, part time, going to school, keeping house, or what? (Mark all that apply)

- ☐ Working full time (**wrkstat_fulltime**)
- ☐ Working part time (**wrkstat_parttime**)
- ☐ With a job, but not at work because of temporary illness, vacation, strike (**wrkstat_notatwork**)
- ☐ Unemployed, laid off, looking for work (**wrkstat_unemployed**)
- ☐ Retired (**wrkstat_retired**)
- ☐ In school (**wrkstat_school**)
- ☐ Keeping house (**wrkstat_keephouse**)
- ☐ Other (please specify) (**wrkstat_other**)
  _____ (**wrkstat_other_text**)

weekswrk
Now I'd like to ask you about last year. In 2013, how many weeks did you work either fulltime or parttime not counting work around the house? (Include paid vacations and paid sick leave.)

Number of weeks (if none, enter 0): _____

[fearsa-fearmurd]
Note: These items were asked on the same page in a matrix format and are adapted from Warr and Stafford (1983).
In your day to day life, how afraid are you of experiencing the following:

Not at all afraid (0)	1	2	3	4	5	Very afraid (6)
○	○	○	○	○	○	○

fear_sa
Being assaulted sexually

fear_ipv
Being physically attacked by a romantic partner

fear_aqnt
Being physically attacked by someone else you know

fear_strn
Being physically attacked by a stranger

fear_murd
Being murdered

[protect_alarm-protect_none]
Note: This question was based on OWNGUN in the GSS but expanded to cover other forms of protection.

Do you have in your home, or sometimes carry with you, any of the following items for your protection?
(Mark all that apply)

- ☐ Alarm or home surveillance system (**protect_alarm**)
- ☐ Firearm (**protect_firearm**)
- ☐ Knife (**protect_knife**)
- ☐ Pepper spray or Mace (**protect_mace**)
- ☐ Stun gun or Taser (**protect_taser**)
- ☐ Whistle (**protect_whistle**)
- ☐ Other (please specify) (**protect_other**)
  _____ (**protect_other_text**)
- ☐ I do not keep or carry anything for the purpose of protection (**protect_none**)

[viol_throw-viol-viol_none]

Note: This question and the follow-up below were adapted from the 2000 National Violence Against Women Survey.

In the past 12 months, has anyone: (Mark all that apply)

- ☐ Thrown something at you with the intent to hurt you (**viol_throw**)
- ☐ Pushed, grabbed, or shoved you (**viol_push**)
- ☐ Slapped, punched, or kicked you (**viol_hit**)
- ☐ Hit you with something with the intent to hurt you (**viol_hitobj**)
- ☐ Burned or scalded you on purpose (**viol_burn**)
- ☐ Assaulted you sexually (**viol_sa**)
- ☐ Choked or attempted to drown you (**viol_chokedrown**)
- ☐ Threatened you with a knife or gun (**viol_knifegun**)
- ☐ Stabbed or shot you (**viol_stabshot**)
- ☐ None of these (**viol_none**)

Who did this to you? (Mark all that apply)

[Note: this question was only displayed for those who did not select *viol_none*.]

- ☐ A stranger (**perp_stranger**)
- ☐ Someone from school (**perp_school**)
- ☐ Someone from work (**perp_work**)
- ☐ A neighbor (**perp_neighbor**)
- ☐ A friend (**perp_friend**)
- ☐ A relative or family member (**perp_relative**)
- ☐ A boyfriend, girlfriend, or lover (**perp_lover**)
- ☐ A spouse or partner (**perp_partner**)
- ☐ Other (please specify) (**perp_other**)
  _____ (**perp_other_text**)

class

If you were asked to use one of four names for your social class, which would you say you belong in: the lower class, the working class, the middle class, or the upper class?

- ○ Lower class (1)
- ○ Working class (2)
- ○ Middle class (3)
- ○ Upper class (4)

finalter
During the last few years, has your financial situation been getting better, getting worse, or has it stayed the same?

- Getting better (1)
- Getting worse (2)
- Stayed the same (3)

[hispanic_binary-hispanic_detail]
Note: The GSS variable *hispanic* is based on a two-step question. Both pieces of information are retained here.

hispanic_binary
Are you Spanish, Hispanic or Latino?

- Yes (1)
- No (0)

hispanic_detail
Which group are you from?
Note: this question was only displayed for those who chose "Yes" for the question hispanic_binary.

- Mexican, Mexican American, Chicano/Chicana (1)
- Puerto Rican (2)
- Cuban (3)
- Other (please specify) (4)
  _____ **(hispanic_other_text)**

[white-otherrace]
Note: this question is identical to the GSS item *racecen*, but we retained each response as a binary indicator.

What is your race? Indicate one or more races that you consider yourself to be.

- ☐ White (**white**)
- ☐ Black or African American (**black**)
- ☐ American Indian or Alaska Native (**amerind**)
- ☐ Asian Indian (**asianind**)
- ☐ Chinese (**chinese**)
- ☐ Filipino (**filipino**)
- ☐ Japanese (**japanese**)
- ☐ Korean (**korean**)
- ☐ Vietnamese (**vietnamese**)
- ☐ Other Asian (**otherasian**)
- ☐ Native Hawaiian (**hawaiian**)
- ☐ Guamanian or Chamorro (**guamcham**)
- ☐ Samoan (**samoan**)
- ☐ Other Pacific Islander (**pacisland**)
- ☐ Some other race (please specify) (**otherrace**)
  _____ (**otherrace_text**)

discaff
What do you think the chances are these days that a white person won't get a job or promotion while an equally or less qualified black person gets one instead:

Is this very likely, somewhat likely, or not very likely to happen these days?

- ○ Very likely (0)
- ○ Somewhat likely (1)
- ○ Not very likely (2)

marhomo
Do you agree or disagree with the following statement: Homosexual couples should have the right to marry one another.

- ○ Strongly agree (5)
- ○ Agree (4)
- ○ Neither agree not disagree (3)
- ○ Disagree (2)
- ○ Strongly disagree (1)

happy
Taking things all together, how would you say things are these days: would you say that you are very happy, pretty happy, or not too happy?

- Very happy (3)
- Pretty happy (2)
- Not too happy (1)

health
Would you say your own health, in general, is excellent, good, fair or poor?

- Excellent (4)
- Good (3)
- Fair (2)
- Poor (1)

mntlhlth
Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

Number of days (if none, enter 0): _____

prtypref
Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?

- Republican (1)
- Democrat (2)
- Independent (3)
- Other (please specify) (4)
  _____ **(prtypref_other_text)**

polviews
There is a lot of talk these days about liberals and conservatives. Below is a seven point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative.

Where would you place yourself on this scale?

- Extremely liberal (1)
- Liberal (2)
- Slightly liberal (3)
- Moderate, middle of the road (4)
- Slightly conservative (5)
- Conservative (6)
- Extremely conservative (7)

born
Were you born in the United States?

- Yes (1)
- No (0)

parborn
Were both of your parents born in the United States?

- Yes (1)
- No (0)

granborn
Were all of your grandparents born in the United States?

- Yes (1)
- No (0)

grade

Note: The GSS variable *educ* is derived from a combination of *grade*, *diploma*, and *coldegree*. (See the next section for our derived educ variable.)

What is the highest grade in elementary school or high school that you finished and got credit for?

- 1st grade (1)
- 2nd grade (2)
- 3rd grade (3)
- 4th grade (4)
- 5th grade (5)
- 6th grade (6)
- 7th grade (7)
- 8th grade (8)
- 9th grade (9)
- 10th grade (10)
- 11th grade (11)
- 12th grade (12)
- Have not had formal schooling (0)

diploma

Did you ever get a high school diploma or a GED certificate?

- Yes (1)
- No (0)

coldegree

Do you have any college degrees?

- Yes (1)
- No (0)

coldegree_type

Note: this question was only displayed for those who answered "Yes" to coldegree

What was the highest degree you earned?

- Associate or junior college degree (1)
- Bachelor's degree (2)
- Graduate or professional degree (3)

earnrs

How many persons in your family earned any money last year – 2013 – from any job or employment? (Include yourself and other people in your household who are related to you.)

Number of people (if none, enter 0): _____

income06

In which of these groups did your total family income, from all sources, fall last year – 2013 – before taxes?
(Total income includes interest or dividends, rent, Social Security other pensions, alimony or child support, unemployment compensation, public aid (welfare), armed forces or veteran's allotment.)

- Under $1,000 (1)
- $1,000 to 2,999 (2)
- $3,000 to 3,999 (3)
- $4,000 to 4,999 (4)
- $5,000 to 5,999 (5)
- $6,000 to 6,999 (6)
- $7,000 to 7,999 (7)
- $8,000 to 9,999 (8)
- $10,000 to 12,499 (9)
- $12,500 to 14,999 (10)
- $15,000 to 17,499 (11)
- $17,500 to 19,999 (12)
- $20,000 to 22,499 (13)
- $22,500 to 24,999 (14)
- $25,000 to 29,999 (15)
- $30,000 to 34,999 (16)
- $35,000 to 39,999 (17)
- $40,000 to 49,999 (18)
- $50,000 to 59,999 (19)
- $60,000 to 74,999 (20)
- $75,000 to 89,999 (21)
- $90,000 to 109,999 (22)
- $110,000 to 129,999 (23)
- $130,000 to 149,999 (24)
- $150,000 or over (25)

dwelown
Do you (or does your family) own your home, pay rent, or what?

- ○ Own or is buying (1)
- ○ Pays rent (2)
- ○ Other (please specify) (3)
  _____ **(dwelown_other_text)**

cohort
What year were you born? _____

SEX AND GENDER MODULE
Note: These were the alternative measures we piloted for the GSS. As noted above, the question order was randomized depending on survey condition. The version shown here corresponds to module_end = 0, scales_first = 1, self_first =1.

[firstord_fem-firstord_masc]
Note: These items were asked on the same page in a matrix format.
In general, how do you see yourself? Please answer on both scales below.

Not at all (0)	1	2	3	4	5	Very (6)
○	○	○	○	○	○	○

firstord_fem
Feminine

firstord_masc
Masculine

[thirdord_fem-thirdord_masc]
Note: These items were asked on the same page in a matrix format.
In general, how do most people see you? Please answer on both scales below.

Not at all (0)	1	2	3	4	5	Very (6)
○	○	○	○	○	○	○

thirdord_fem
Feminine

thirdord_masc
Masculine

sex
What sex were you assigned at birth? (For example, on your birth certificate.)

- Female (1)
- Male (2)
- Intersex (3)

gender
What is your current gender?

- Woman (1)
- Man (2)
- Transgender (3)
- A gender not listed here (please specify) (4)
  _____ **(gender_other_text)**

arrest
Were you ever picked up or charged by the police, for any reason whether or not you were guilty?

- Yes (1)
- No (0)

lockedup
Have you ever spent any time in prison or jail?

- Yes (1)
- No (0)

rhhwork
On average, how many hours a week do you personally spend on household work, not including childcare and leisure time activities?

Number of hours (if none, enter 0): _____

famfinan
Note: This question was only displayed to those who selected the following responses from variable posslqy "I am married and living in the same household as my spouse"; "I have a steady partner and we live in the same household"; or "I have a spouse or steady partner, but we don't live in the same household."
How do you and your spouse/partner organize the income that one or both of you receive? Please choose the option that comes closest.

- ○ I manage all the money and give my spouse/partner his/her share (1)
- ○ My spouse/partner manages all the money and gives me my share (2)
- ○ We pool all the money and each take out what we need (3)
- ○ We pool some of the money and keep the rest separate (4)
- ○ We each keep our own money separate (5)

earnshh
Note: As above, this question was only displayed to respondents who reported having a spouse or partner.
Considering all sources of income, between you and your spouse/partner, who has the higher income?

- ○ My spouse/partner has no income (1)
- ○ I have a much higher income (2)
- ○ I have a higher income (3)
- ○ We have about the same income (4)
- ○ My spouse/partner has a higher income (5)
- ○ My spouse/partner has a much higher income (6)
- ○ I have no income (7)

[cutahead-talkedto]
Note: These items were asked on the same page in a matrix format.
During the past 12 months, how often have you done each of the following things:

	At least once a week (3)	At least once a month (2)	At least once in the past year (1)	Not at all in the past year (0)
cutahead Allowed a stranger to go ahead of you in line	○	○	○	○
volchrty Done volunteer work				

givchrty
Given money to a charity

helpaway
Looked after a person's plants, mail or pets while they were away

carried
Carried a stranger's belongings, like groceries, a suitcase, or a shopping bag

talkedto
Spent time talking with someone who was a bit down or depressed

[hrdshp1-hrdshp7]
Note: These items were asked on the same page in a matrix format.
During the past 12 months, did you find yourself in any of the following circumstances?

Yes (1)	No (0)	Not applicable (-1)
○	○	○

hrdshp1
Fell behind in paying your rent or mortgage

hrdshp4
Unable to purchase needed food

hrdshp5
Unable to purchase needed medical care

hrdshp7
Had to temporarily live with others or in a shelter or on the street

partners
How many sex partners have you had in the last 12 months?

- No partners (0)
- 1 partner (1)
- 2 partners (2)
- 3 partners (3)
- 4 partners (4)
- 5-10 partners (5)
- 11-20 partners (6)
- 21-100 partners (7)
- More than 100 partners (8)

sexfreq
About how often did you have sex during the past 12 months?

- Not at all (0)
- Once or twice (1)
- About once a month (2)
- Two or three times a month (3)
- About once a week (4)
- Two or three times a week (5)
- Four or more times a week (6)

sexornt
Which of the following best describes you?

- Gay, lesbian, or homosexual (1)
- Bisexual (2)
- Heterosexual or straight (3)

sexsex18
Since the age of 18, have your sex partners been:

- Only men (1)
- Some women, but mostly men (2)
- Equally men and women (3)
- Some men, but mostly women (4)
- Only women (5)
- I have not had any sex partners (0)

SEX AND GENDER MODULE

Note: In the survey conditions where module_end = 1, the alternative sex and gender measures appeared here. This is similar to the placement of the two-step sex at birth and current gender measures on the 2018 GSS, which immediately follow *sexornt*.

Derived Variables

cisgender
[Derived from **sex** and **gender**. Cisgender = yes if sex is *female* and gender is *woman*, or if sex is *male* and gender is *man*.]

female
[Binary variable **sex** where 0 = *male* and 1 = *female*. (The sex variable also offered the option *intersex*, but no respondents selected this answer.)]

male
[Binary variable **sex** where 0 = *female* and 1 = *male*. (The sex variable also offered the option *intersex*, but no respondents selected this answer.)]

woman
[Variable derived from **gender** where 0 = *man, transgender,* or *a gender not listed* and 1 = *woman*.]

man
[Variable derived from **gender** where 0 = *woman, transgender,* or *a gender not listed* and 1 = *man*.]

gnlh
[Variable derived from **gender** where 0 = *woman, man,* or *transgender* and 1 = *a gender not listed here*.]

trans_direct
[Variable derived from **gender** where 0 = woman, man, or a gender not listed and 1 = transgender]

trans_indirect
[Variable derived from **gender** and **sex.** Trans_indirect = 0 if gender is *woman* and sex is *female*, if gender is *man* and sex is *male*, or if gender is *transgender* or *a gender not listed.* Trans_indirect = 1 if gender is *woman* and sex is *male* or if gender is *man* and sex is *female*.]

polar1
[Derived by taking the absolute difference between responses to firstord_fem and firstord_masc. This variable indicates the polarization of the respondent's gendered sense of self, where 0 reflects equal feminine and masculine responses and a 6 is consistent with traditional binary notions of gender.]

polar3
[Derived by taking the absolute difference between responses to thirdord_fem and thirdord_masc. This variable indicates the polarization of how the respondent thinks they are seen by others, where 0 reflects equal feminine and masculine responses and a 6 is consistent with traditional binary notions of gender.]

crossgen1
[Binary variable derived from the firstord_fem and firstord_masc items. When this variable is coded 1, it indicates that the respondent rates themselves more highly on the gender scale not stereotypically associated with their binary categorical gender (i.e., a woman who sees herself as more masculine than feminine, or a man who sees himself as more feminine than masculine).]

crossgen3
[Binary variable derived from the thirdord_fem and thirdord_masc items. When this variable is coded 1, it indicates that the respondent feels others would rate them more highly on the gender scale not stereotypically associated with their binary categorical gender (i.e., a woman who feels others see her as more masculine than feminine, or a man who feels others see him as more feminine than masculine).]

notconven1
[More expansive version of crossgen1. When this variable is coded 1, it indicates that respondents rate themselves *equally* or more highly on the gender scale not stereotypically associated with their binary categorical gender.]

notconven3
[More expansive version of crossgen3. When this variable is coded 1, it indicates that respondents feel others rate them *equally* or more highly on the gender scale not stereotypically associated with their binary categorical gender.]

married
[Binary variable derived from **posslqy** where 1 indicates that the respondent selected "I am married and living in the same household as my spouse" and 0 indicates all other responses.]

fear_avg
[Mean of **fearsa**, **fearipv**, **fearaqnt**, **fearstrn**, and **fearmurd**.]

protect_count
[Sum of **protect_alarm**, **protect_firearm**, **protect_knife**, **protect_mace**, **protect_taser**, **protect_whistle**, and **protect_other**. Values greater than 1 indicate the respondent reported using multiple items for protection.]

viol_count
[Sum of **viol_throw**, **viol_push**, **viol_hit**, **viol_hitobj**, **viol_burn**, **viol_sa**, **viol_chokedrown**, **viol_knifegun**, and **viol_stabshot**. Values greater than 1 indicate the respondent reported experiencing multiple forms of violence.]

race_count
[Sum of **white, black, amerind, asianind, chinese, filipino, japanese, korean, vietnamese, otherasian, hawaiian, guamcham, samoan, pacisland,** and **otherrace**. Values greater than 1 indicate the respondent selected multiple race responses.]

race_detail
[Summary variable that combines information from the binary indicators **white**, **black**, **amerind**, **asianind**, **chinese**, **filipino**, **japanese**, **korean**, **vietnamese**, **otherasian**, **hawaiian**, **guamcham**, **samoan**, **pacisland**, and **otherrace**. Respondents are designated with the single race they selected or, if they selected more than one category, are designated with the label "multiracial."]

race_gss
[Summary variable derived from **race_detail**. This variable was constructed to align with the categories of the GSS variable *race*. *White* = respondent selected race **white** and no other race; *Black* = respondent selected race **black** and no other race; *Other* = respondent selected a race other than white or black.]

goodhealth
[Derived from **health**. Fair/poor (0) = combined responses to *fair* and *poor*; good/excellent (1) = combined responses to *good* and *excellent*.]

health_3cat
[Derived from **health**. Fair/poor (1) = combined responses to *fair* and *poor*; good (2) = *good*, and excellent (3) = *excellent*.]

educ
[Continuous variable derived from **grade** and **coldegree_type**. For those without any college education, educ = highest year of school completed. For those who attended college, educ = 14 for an associate degree, educ = 16 for a bachelor's degree, and educ = 20 for a graduate or professional degree.]

inc_cont
[Derived using income category midpoints and top coding, divided by 10,000.]

.05 = $1,000
.20 = $1,000-$2,999
.35 = $3,000-$3,999
.45 = $4,000-$4,999
.55 = $5,000-$5,999
.65 = $6,000-$6,999
.75 = $7,000-$7,999
.90 = $8,000-$9,999
1.125 = $10,000-$12,499
1.375 = $12,500-$14,999
1.625 = $15,000-$17,499
1.875 = $17,500-$19,999
2.125 = $20,000-$22,499
2.375 = $22,500-$24,999
2.75 = $25,000-$29,999
3.25 = $30,000-$34,999
3.75 = $35,000-$39,999
4.50 = $40,000-$49,999
5.50 = $50,000-$59,999
6.75 = $60,000-$74,999
8.25 = $75,000-$89,999
10.0 = $90,000-$109,999
12.0 = $110,000-$129,999
14.0 = $130,000-$149,999
16.0 > $150,000

age
[Derived variable: 2014-**cohort**.]

region_census
[Aligns with the U.S. Census Bureau's 9-category regional divisions (see https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf). This variable was derived from the respondent's self-reported state of residence, which is not included in the AGMS publicly available data.]

- **Pacific (1):** Alaska, California, Hawaii, Oregon, Washington
- **Mountain (2):** Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
- **West North Central (3):** Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
- **East North Central (4):** Illinois, Indiana, Michigan, Ohio, Wisconsin
- **Middle Atlantic (5):** New Jersey, New York, Pennsylvania
- **New England (6):** Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
- **South Atlantic (7):** Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, Washington DC, West Virginia
- **East South Central (8):** Alabama, Kentucky, Tennessee, Mississippi
- **West South Central (9):** Arkansas, Louisiana, Oklahoma, Texas

region_gss
[Derived from region_census. The variable corresponds to the major regions provided in publicly available GSS data.]

- **Northeast (1):** Middle Atlantic and New England regions
- **Midwest (2):** West North Central and East North Central regions
- **South (3):** South Atlantic, East South Central, and West South Central regions
- **West (4):** Pacific and Mountain regions

References

Hart, Chloe Grace, Aliya Saperstein, Devon Magliozzi, and Laurel Westbrook. 2019. "Gender and Health: Beyond Binary Categorical Measurement." *Journal of Health and Social Behavior* 60(1):101–18.

Magliozzi, Devon, Aliya Saperstein, and Laurel Westbrook. 2016. "Scaling Up : Representing Gender Diversity in Survey Research." *Socius* 2:1–11.

Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen L. Morgan. 2019. "General Social Surveys, 1972-2018 Cumulative Codebook." National Data Program for the Social Sciences Series, no. 25. Chicago: NORC.

Warr, Mark and Mark Stafford. 1983. "Fear of Victimization: A Look at the Proximate Causes." *Social Forces* 61(4):1033–1043

Appendix

We first proposed alternative sex and gender measures to the General Social Survey in an open competition for items to be included in the 2016 survey wave. We submitted our initial proposal in March 2014 and responded to requests for additional materials through several waves of review. Our module was cut from consideration in September 2015 in part because the GSS received less funding than expected in its grant renewal from the National Science Foundation. (The GSS panel was also discontinued at that time.)

However, inspired by our efforts, the GSS Board formed a subcommittee to consider revisions to their core demographic items that eventually resulted in the two-step categorical items for sex at birth and current gender being included in the 2018 General Social Survey. The wording and answer options in the 2018 GSS are identical to the versions piloted in the AGMS.

To our knowledge, the 2018 GSS represents the first time that researchers can estimate population shares of transgender and nonbinary people from a nationally representative survey of all U.S. adults living in households.

We include the original text of our proposal submissions here because they provide additional background about the design of the AGMS and about an initial pilot of sex and gender measures with binary and nonbinary answer options. They also offer a successful example of how to propose items to the GSS for future open competitions.

Initial GSS Proposal

Including Measures of Gender Identity and Diversity

A proposal for the 2016 General Social Survey

Shelley Correll, Stanford University*
Cecilia Ridgeway, Stanford University
Aliya Saperstein, Stanford University
Laurel Westbrook, Grand Valley State University

*Proposers are listed alphabetically. Please direct correspondence to: Aliya Saperstein, Department of Sociology, Stanford University, asaper@stanford.edu

Including Measures of Gender Identity and Diversity

Executive Summary

The General Social Survey has become a staple of social science research and student training in survey methods. As such, it helps to set the tone for best practices in measurement and analysis of social science data. The GSS can further cement that role by being at the forefront of recent efforts to improve data on the diversity of gendered experience among American adults. We propose adding a series of four items to the GSS self-administered questionnaire that will: 1) provide more nuanced data on respondents' gender identity – including how they see themselves and are perceived by others on scales of masculinity and femininity – and 2) allow researchers to select the most conceptually appropriate measures of sex or gender (or both) for use in their analyses of people's attitudes, opinions and life outcomes. These new measures will be of use to researchers from a wide range of disciplines, will take little additional survey time, and can help bring survey measures of sex and gender more in line with both Americans' lived experience and contemporary social science theory.

Background and Significance

Concerns about existing sex and gender categories and calls for new response options for surveys and official data collection are growing around the world. Thus far, most of the attention has been focused on providing what have become known as "third gender" or "third sex" categories aimed to include a range of experiences from identifying as transgender to being born intersex. In 2011, Nepal became the first country to include a third gender in its national census; India soon followed (Bochenek and Knight 2012). By the end of 2013, a third sex option was available on passports in New Zealand, all "personal documents" in Australia, and the option of not specifying a child's sex is now allowed in German birth registries. In the U.S., federal hate crimes law was expanded in 2009 to protect transgender people, and 17 states and the District of Columbia currently prohibit discrimination based on actual or perceived gender identity in housing and both public and private sector employment (HRC 2014). Task forces in the U.S. and Europe advocate for alternative measures in surveys and official data collection not only to recognize the diversity of gendered experience, but also to monitor these previously under studied forms of gender discrimination (Balarajan, Gray and Mitchell 2011; HRW 2011; IOM 2011; GenIUSS 2013).

Our proposal builds on these efforts. However, as social scientists, we believe the primary goal in rethinking how sex and gender questions are asked in surveys should be to develop clear, consistent and valid measures that are informed by the rich interdisciplinary literature on how to conceptualize sex and gender (e.g., West and Zimmerman 1987; Connell 1995; Kessler 1998; Fausto-Sterling 2000; Valentine 2007; Ridgeway 2011; Spence 2011; Westbrook and Schilt 2014). From this perspective, simply adding new response options to existing questions will not solve many of the limitations of conventional approaches to measuring these concepts. In major

U.S. national surveys, biological sex and gender identity are generally conflated in a single question, the standard dichotomous answer options are often treated as internally homogenous, and the data is still recorded almost exclusively by interviewers, rather than also being directly asked of respondents (Westbrook and Saperstein 2014; see McCulloch 2012 for a similar review of telephone surveys). In this sense, survey measurement of sex and gender appears to lag behind measures of other key dimensions of difference, including race and ethnicity, sexual orientation and diverse family forms. In these arenas, national surveys now often employ multiple measures, allow for self-identification and recognize the changing – or previously hidden – demographic landscape in the United States (SMART 2009; Powell et al. 2010; Moore and Stambolis-Ruhstorfer 2013; Saperstein 2013).

There are several potential challenges to consider when implementing new measures of sex and gender in a national survey, including issues of comprehension, non-response and the relatively small size of populations with non-normative responses on the various items. Small-scale studies have not found non-response or comprehension to be significant impediments to using alternative measures of sex and gender (e.g., Redford and Van Wagenen 2012; GenIUSS 2013; Tate, Ledbeter and Youssef 2013), and we aim to confirm these findings on larger, less-targeted populations by conducting our own pilot surveys (described in detail below). It is difficult to estimate the population counts that will result from using alterative approaches – such as separate measures of biological sex and current gender identity, or by conceptualizing gender as a spectrum or scale rather than a set of mutually exclusive categories – precisely because these questions have not yet been asked in nationally representative studies. In a college sample, Spence and Buckner (2000) find that men on average scored a 3 on a five-point "masculine" scale, while women scored on average around 0.7; roughly the reverse was found for the "feminine" item. Although the distributions are highly skewed, they note the standard deviations ranged from .62-.85 suggesting a reasonable amount of intra-sex/gender variability. Typical estimates of the transgender population in the United States range from .1% to .5% (Gates 2011), while intersex births are estimated to occur with .1% to .2% frequency (Blackless et al. 2000). The latter estimates are on par with several of the racial categories currently enumerated in the GSS, including Vietnamese, Korean and Native Hawaiian. Despite their expected small size, growing social and legal attention to third categories of sex and gender increases the need to empirically assess their frequency, particularly in contexts, such as the GSS self-administered questionnaire, in which people have an anonymous opportunity to choose alternate responses. Once established, the basic base-rate information can better inform future surveys in building the over samples necessary to provide important comparisons between the standard experience of sex and gender and other alternatives. Should the GSS decide to make a more permanent change in its measurement strategy, populations could also be pooled across years to facilitate more detailed analysis. Finally, we note the formative role that surveys and other official data collection can play in shaping how Americans view each other and the world around them (Starr 1992; Zeruvabel 1996; Igo 2007), which further supports the importance of recognizing sex and gender diversity as legitimate lived experiences regardless of their frequency.

Proposed questions and wording

We propose a set of four questions that would be asked sequentially on a self-administered section of the GSS questionnaire. These questions are designed to tap into various dimensions of

sex and gender, including self-identity and perceptions of the respondent's masculinity and femininity (see Appendix for detailed wording and format options). To maximize observing variation across questions, we recommend including these items in both new interviews and any panel re-interviews in 2016. We discuss each item below, in order of priority.

Masculine and feminine scales. Perhaps most important in any effort to update the measurement of sex and gender is allowing for diversity within the categories of "men" and "women" (or males and females) that are too often treated as monolithic in social science research. Studies of health disparities have already demonstrated the utility of employing graded measures of masculinity and femininity and find, for example, that relying on traditional sex categories in analysis masks an association of self-identified femininity with decreased risk of death from heart disease among men (Hunt et al. 2007; see also Hammerstrom and Annandale 2012). Similar sociological studies that fruitfully combine measures of masculinity and femininity with dichotomous sex categories include work by Burke and colleagues (e.g., Burke, Stets and Pirog-Good 1988; Burke and Cast 1997) and McLaughlin, Uggen and Blackstone (2012).

To capture greater variation in gender identity and presentation, we recommend asking respondents where they would be placed by others, and where they would place themselves, on separate masculine and feminine scales. Each scale would run from "not at all" to "extremely," and could be composed of either 5 or 7 categories. (Both are common in previous literature.) The questions would be worded in parallel fashion. The question order could present the third order perception (e.g., In general, how do most people see you?) first followed by the first order perception (e.g., In general, how do you see yourself?) to allow more room for respondents to contradict the "reflected appraisal" of how they think other people see them, or it could be counterbalanced across the sample.

These items are most similar to the masculinity and femininity trait ratings that have been a staple of psychological survey measures for decades (Morawski 1985; Bem 1993; Auster and Ohm 2000; see also Stets and Burke 2000). Recent studies have suggested that asking only one question (e.g., how would you describe yourself) causes confusion among respondents in terms of whether they should answer based on other's perceptions or their own, and have further recommended assessing different dimensions of gender presentation – e.g., appearance, mannerisms, personality – separately (Clark, Armstrong and Bonacore 2005; SMART 2009; Wylie et al. 2010). We believe more general measures of first and third order perceptions are most appropriate given the purpose and time constraints of the GSS, but we do affirm the importance of asking these items using two separate masculine and feminine scales, rather than assuming that these concepts represent the poles of a single spectrum. (See Constantinople 1973 for the seminal statement on this issue.)

Self-identified gender. The current GSS variable SEX prompts interviewers to "SELECT GENDER OF CHOSEN RESPONDENT" at the beginning of each survey, and provides the categories "male" and "female." This measure conflates sex categories, presumably based on physical or biological differences, with gender presentation, based on appearance, name, mannerisms and other behavioral factors, which presents a challenge for researchers who seek to explain the causes and consequences of stratification and social inequality. A cleaner, more conceptually clear measure would simply be to ask respondents to identify their current gender.

Some survey administrators have expressed concern that respondents will experience being asked to clarify their gender as awkward or rude (see, e.g., DeBell et al. 2010). Placing the question on the self-administered portion of the GSS can alleviate this concern. Americans are regularly asked to report whether they are male or female on an array of forms and other surveys, so the basic idea should be familiar to them. However, we recommend that rather than providing the options "male" and "female," the categories "man" and "woman" should follow a question about self-identified gender. (Male and female are generally considered sex terms while man and woman are gender terms.) We further suggest that it would be appropriate to offer at least one alternative response option, for example, allowing respondents to specify "a gender not listed here." Cognitive interviews indicate respondents who do not identify with either of the standard gender categories find being asked to select "Other" offensive and stigmatizing (HRW 2011), and "a gender not listed here" is seen as preferable (Harrison et al. 2011).

Sex assigned at birth. We have argued for the importance of making a clear conceptual distinction between measures of sex and measures of gender to improve issues of interpretation in social science research. Although the current measure of interviewer-classified sex/gender might be useful to retain for other reasons, it is not an explicit measure of the respondent's biological sex but rather a measure of their perceived sex category (see West and Zimmerman 1987). Thus, our final recommendation is to ask respondents directly for the sex they were assigned at birth. This would be accompanied by the answer options "male" and "female" and interviewers could provide respondents with the prompt, "For example, on your birth certificate" in response to requests for explanation or clarification. Ideally, response options would also include "intersex," which is already recognized on birth certificates in several countries, though we expect reporting in this category would be rare, given the sample size of the GSS.

Targeted health and discrimination studies have developed an array of detailed measures to capture changes in sex and gender classification over time (e.g., Grant et al. 2011; Gonzalez et al. 2012), but we recognize that these kinds of instruments are not feasible for a general survey like the GSS. However, using a simple two-question method – self-identified gender and sex assigned at birth – can assist researchers in identifying the prevalence of gender fluidity among American adults, as one can compare respondents' answers for sex at birth with their current self-identified gender. Previous studies have also found that asking respondents what sex they were assigned at birth helps provide context for the current gender measure recommended above and thus reduces non-response (Tate et al. 2013; Deogracias et al. 2007; Melendez et al. 2006; Singer et al 1997).

In sum, whether used alone or in combination, we believe implementing these questions in the 2016 GSS will go a long way towards improving the match between theory and methods in studies of sex and gender in the United States.

Pilot survey plans

We are not aware of any studies that have tested this set of questions in its entirety or with our specific combinations of question wording and answer options. We have reviewed the relevant literature in sociology, psychology and several health disciplines but as this literature is in its

infancy, it would be premature to make a general statement regarding best practices for national social surveys. As such, we are collaborating with Robb Willer (Stanford University) to run at least two pilot studies to test the performance of our measures and any effects of question order or presentation of answer options on the pattern of responses. We will recruit participants through Amazon.com's Mechanical Turk (MTurk) to provide more representative data than previous small-scale cognitive interviews and group-specific studies (Buhrmester et al. 2011). We received funding from the American Sociological Association Fund for the Advancement of the Discipline to support the pilot surveys and have received human subjects exemption from Stanford's Institutional Review Board. Willer also has previous experience running survey experiments using the MTurk platform (see, e.g., Côté, Piff and Willer 2013).

In our initial survey, we plan to recruit 1,600 respondents who will be randomly distributed across four experimental conditions designed to test the effects of question order in combination with the order in which answer options are presented. We also expect that this pilot will provide evidence of respondents' willingness and ability to answer questions about their sex category, gender identity and gender presentation. It will further allow us to estimate the extent to which GSS respondents are likely to provide answers that vary across the set of measures. We expect to have these results available in time for the Fall 2014 GSS Board Meeting, pending continued interest in our proposal.

Once we have established the feasibility of asking these questions, and identified an appropriate question ordering, we plan to embed the new measures of gender identity and diversity in a scaled down version of the General Social Survey, also executed online through AMT. This "mini-GSS" will ask approximately 40 questions of a new set of approximately 1,400 respondents. The second pilot survey will use the question wording and rough ordering of the 2012 GSS to demonstrate the utility of these new measures in analyses of common survey questions. It will also provide further evidence of data quality in terms of item non-response. While the online format is not ideal in replicating the face-to-face aspect of GSS interviews, it will mimic the format of the self-administered portions of the survey. We are aiming to have these data available in time for the Spring 2015 Board Meeting.

The initial pilot conditions are illustrated in the Appendix below. Across each condition we will examine the effects of question ordering to determine whether respondents express more or less diversity across their responses when presented with some items before others. An example would be if asking for sex assigned at birth before current gender does make the distinction being measured in the second question clearer to respondents. Another possibility could be that asking the masculine and feminine scales last will yield responses that appear to be biased toward gender conformity. We also plan to test whether introducing non-binary sex category and gender identity answer options affects responses to the masculine and feminine scales. In particular, we are interested in whether respondents presented with a non-traditional view of sex and gender (i.e., more than two categories, or a reversed hierarchy of answer options) are more likely to select the extremes on the scale items to reassert traditional notions of sex differences and gender conformity (cf. Willer et al. 2013). We aim to provide information to the Board on how best to deploy new measures of gender identity and diversity without compromising data quality or responses to existing questions. We hope that you will support us in this effort towards improving the measurement of sex and gender in the General Social Survey.

References

Auster, Carol J. and Susan C. Ohm. 2000. "Masculinity and Femininity in Contemporary American Society: A Reevaluation Using the Bem Sex-Role Inventory." *Sex Roles* 43(7):499-528.

Balarajan, Meera, Michelle Gray and Martin Mitchell. 2011. *Monitoring Equality: Developing a Gender Identity Question*. Manchester, UK: Equality and Human Rights Commission.

Bem, Sandra L. 1993. *The Lenses of Gender: Transforming the Debate on Sexual Inequality.* New Haven: Yale University Press.

Blackless, Melanie et al. 2000. "How Sexually Dimorphic Are We? Review and Synthesis." *American Journal of Human Biology* 12(2):151–66.

Bochenek, Michael and Kyle Knight. 2012. "Establishing a Third Gender Category in Nepal: Process and Prognosis." *Emory International Law Review* 26:11-41.

Buhrmester, Michael, Tracy Kwang and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality Data?" *Perspectives on Psychological Science* 6(1): 3-5.

Burke, Peter J. and Alicia D. Cast. 1997. "Stability and Change in the Gender Identities of Newly Married Couples." *Social Psychology Quarterly* 60: 277-290.

Burke, Peter J., Jan E. Stets, and Maureen A. Pirog-Good. 1988. "Gender Identity, Self-Esteem, and Physical and Sexual Abuse in Dating Relationships." *Social Psychology Quarterly* 51: 272-285.

Clark, Melissa, Gene Armstrong, and Laura Bonacore. 2005. Measuring Sexual Orientation and Gender Expression Among Middle-aged and Older Women in a Cancer Screening Study. *Journal of Cancer Education.* 20(2): 108-112.

Connell, Raewyn. 2005. *Masculinities*. Berkeley, CA: University of California Press.

Constantinople, Anne. 1973. "Masculinity-Femininity: An Exception to the Famous Dictum?" *Psychological Bulletin* 80: 389-407.

Côté, Stéphane, Paul K. Piff and Robb Willer. 2013. "For Whom Do the Ends Justify the Means? Social Class and Utilitarian Moral Judgment." *Journal of Personality and Social Psychology* 104(3):490-503.

DeBell, Matthew, Jon A. Krosnick, and Arthur Lupia. 2010. "Recruitment Interviewer Training Materials." Appendix C in *Methodology Report and User's Guide for the 2008–2009 ANES Panel Study.* Palo Alto, CA, and Ann Arbor, MI: Stanford University and the University of Michigan.

Deogracias, Joseph J. et al. 2007. "The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults." *Journal of Sex Research* 44(4):370–79.

Fausto-Sterling, Anne. 2000. "The Five Sexes, Revisited." *Sciences-New York* 40(4): 18-25.

Gates, Gary J. 2011. "How Many People Are Lesbian, Gay, Bisexual and Transgender?" Los Angeles, CA: The Williams Institute. Retrieved February 2, 2014 (http://escholarship.org/uc/item/09h684x2).

GenIUSS Group (Gender Identity in U.S. Surveillance). 2013. "Gender-Related Measures Overview." Los Angeles, CA: The Williams Institute. http://williamsinstitute.law.ucla.edu/wp-content/uploads/GenIUSS-Gender-related-Question-Overview.pdf

Gonzalez, Cesar A., Walter O. Bockting, Linda J. Beckman and Ron E. Duran. 2012. "Agentic and Communal Personality Traits: Their Associations with Depression and Resilience among Transgender Women." *Sex Roles* 67(9-10): 528-43.

Grant, Jaime M., Lisa A. Mottet, Justin Tanis, Jack Harrison, Jody L. Herman, and Mara Keisling. 2011. *Injustice at Every Turn: A Report of the National Transgender Discrimination Survey.* Washington, D.C.: National Center for Transgender Equality and National Gay and Lesbian Task Force.

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(4): e34193.

Harrison, Jack, Jaime Grant and Jody L. Herman. 2011. "A Gender Not Listed Here: Genderqueers, Gender Rebels and OtherWise in the National Transgender Discrimination Survey." *LGBTQ Policy Journal* 2:13-24.

HRC (Human Rights Campaign). 2014. Maps of State Laws and Policies. Accessed online: https://www.hrc.org/resources/entry/maps-of-state-laws-policies

HRW (Human Rights Watch). 2011. Controlling Bodies, Denying Identities: Human Rights Violations against Trans People in the Netherlands. New York: Human Rights Watch.

Hunt, Kate, Heather Lewars, Carol Emslie and G. David Batty. 2007. "Decreased Risk of Death from Coronary Heart Disease Amongst Men with Higher 'Femininity' Scores: A General Population Cohort Study." *International Journal of Epidemiology* 36(3): 612-20.

Igo, Sarah. 2007. *The Averaged American: Surveys, Citizens, And the Making of a Mass Public.* Cambridge, MA: Harvard University Press.

IOM (Institute of Medicine). 2011. *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding*. Washington, D.C.: The National Academies Press.
Kessler, Suzanne J. 1998. *Lessons from the Intersexed*. New Brunswick, NJ: Rutgers University Press.
McCulloch, Susan K. 2012. *Effects of Acoustic Perception of Gender on Nonsampling Errors in Telephone Surveys*. Ph.D. Dissertation. University of Maryland, College Park. Accessed online: http://drum.lib.umd.edu/bitstream/1903/13391/1/KenneyMcCulloch_umd_0117E_13838.pdf
McLaughlin, Heather, Christopher Uggen and Amy Blackstone. 2012. "Sexual Harassment, Workplace Authority, and the Paradox of Power." *American Sociological Review* 77(4): 625-647.
Melendez, Rita M., Theresa A. Exner, Anke A. Ehrhardt, Brian Dodge, and et al. 2006. "Health and Health Care Among Male-to-Female Transgender Persons Who Are HIV Positive." *American Journal of Public Health* 96(6):1034–37.
Moore, Mignon R. and Michael Stambolis-Ruhstorfer. 2013. "LGBT Sexuality and Families at the Start of the Twenty-First Century." *Annual Review of Sociology* 39:491-507.
Morawski, J. G. 1985. "The Measurement of Masculinity and Femininity: Engendering Categorical Realities." Pp. 108-135 in Abigail J. Stewart and M. Brinton Lykes (Eds.), *Gender and Personality: Current Perspectives on Theory and Research*. Durham, NC: Duke University Press.
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.
Redford, Jeremy and Aimee Van Wagenen. 2012. "Measuring Sexual Orientation Identity and Gender Identity in a Self-Administered Survey: Results from Cognitive Research with Older Adults." Paper presented at the Population Association of American Annual Meeting, San Francisco, Calif., May 4.
Ridgeway, Cecilia. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.
Saperstein, Aliya. 2013. "Representing the Multidimensionality of Race in Survey Research" In *Mapping 'Race': Critical Approaches to Health Disparities Research*, eds. Laura E. Gomez and Nancy Lopez. New Brunswick, NJ: Rutgers University Press.
SMART (Sexual Minority Assessment Research Team). 2009. "Best Practices for Asking Questions about Sexual Orientation on Surveys." Los Angeles, CA: The Williams Institute. http://williamsinstitute.law.ucla.edu/wp-content/uploads/SMART-FINAL-Nov-2009.pdf
Singer, T. B., M. Cochran and R. Adamec. 1997. Final Report by the Transgender Health Action Coalition to the Philadelphia Foundation Legacy Fund Needs Assessment Survey Project (A.K.A. the Delaware Valley Transgender Survey). Philadelphia, PA: Transgender Health Action Coalition.
Spence, Janet T. 2011. "Off With the Old, On With the New." *Psychology of Women Quarterly* 35(3):504-9.
Spence, Janet T. and Camille E. Buckner. 2000. "Instrumental and Expressive Traits, Trait Stereotypes, and Sexist Attitudes: What Do They Signify?" *Psychology of Women Quarterly* 24(1): 44-53.
Starr, Paul. 1992. "Social Categories and Claims in the Liberal State." *Social Research* 59(2):263–95.
Stets, Jan E. and Peter J. Burke. 2000. "Femininity/Masculinity." Pp. 997-1005 in Edgar F. Borgatta and Rhonda J. V. Montgomery (Eds.), *Encyclopedia of Sociology, Revised Edition*. New York: Macmillan.
Tate, Charlotte, C., Jay N. Ledbetter, and Cris P. Youssef. 2013. "A Two-Question Method for Assessing Gender Categories in the Social and Medical Sciences." *Journal of Sex Research* 50(8):767–76.
Valentine, David. 2007. *Imagining Transgender: An Ethnography of a Category*. Duke University Press.
West, C., and D. H. Zimmerman. 1987. "Doing Gender." *Gender & Society* 1(2):125–51.
Westbrook, Laurel, and Aliya Saperstein. 2014. "Gender Counts: The Construction of Sex and Gender in U.S. Social Surveys." Paper to be presented at the American Sociological Association Annual Meeting, San Francisco, CA, August 18.
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(1): 32-57.
Willer, Robb, Christabel Rogalin, Bridget Conlon and Michael T. Wojnowicz. 2013. "Overdoing Gender: A Test of the Masculine Overcompensation Thesis." *American Journal of Sociology* 118(4): 980-1022.
Wylie SA, Corliss HL, Boulanger V, Prokop LA, Austin SB. 2010. Socially Assigned Gender Nonconformity: A Brief Measure for Use in Surveillance and Investigation of Health Disparities. *Sex Roles* 63(3-4): 264-276.
Zerubavel, Eviatar. 1996. "Lumping and Splitting: Notes on Social Classification." *Sociological Forum* 11(3):421–33.

Appendix: Pilot study instrument

The first stage of our pilot study will test different question orders and response options on a pool of 1,600 respondents randomly distributed across four conditions. The alternative measures of sex and gender were selected based on their performance in previous small-scale studies. We are testing to see how these questions perform in a larger, less-targeted population. We also seek to determine whether there are any question order effects when these measures are asked one right after the other, and whether this differs by the answer options provided.

Our four conditions are summarized by the following 2X2 table:

	Traditional Hierarchy of Answers	**Non-Traditional Hierarchy of Answers**
Binary responses	***Condition 1:*** Sex and gender are binary and answer orders are traditionally hierarchical (male/man/masculine first, female/woman/feminine second).	***Condition 2:*** Sex and gender are binary and answer orders are the reverse of the traditional hierarchy (instead, it is: female/woman/feminine first, male/man/masculine second).
Non-binary	***Condition 3:*** Sex and gender are non-binary and answer orders are traditionally hierarchical (male/man/masculine first, female/woman/feminine second, non-traditional categories last).	***Condition 4:*** Sex and gender are non-binary and answer orders do not follow the traditional hierarchy (instead it is: female/woman/feminine first, non-traditional categories second, and male/man/masculine third, followed by "a gender not listed here").

The questions/answer options for Condition 1 are:

1) What sex were you assigned at birth? (For example, on your birth certificate.)
    a. Male
    b. Female

2) What is your current gender?
    a. Man
    b. Woman

Please record your responses to the next set of questions using both of the scales provided.

3) In general, how do most people see you:
    Not at all masculine ----------> Extremely masculine
    Not at all feminine ----------> Extremely feminine

4) In general, how do you see yourself:
    Not at all masculine ----------> Extremely masculine
    Not at all feminine ----------> Extremely feminine

The questions for Condition 2 reverse the order of the answer options:

1) Sex
    a. Female
    b. Male

2) Gender
    a. Woman
    b. Man

3) How do most people see you
    Not at all feminine ----------> Extremely feminine
    Not at all masculine ----------> Extremely masculine

4) How do you see yourself
    Not at all feminine ----------> Extremely feminine
    Not at all masculine ----------> Extremely masculine

The questions for Condition 3 introduce non-binary answer options in the original order:

1) Sex
    a. Male
    b. Female
    c. Intersex

2) Gender
    a. Man
    b. Woman
    c. Transgender
    d. A gender not listed here
        i. Please specify _____

3) How do most people see you
    Not at all masculine ----------> Extremely masculine
    Not at all feminine ----------> Extremely feminine

4) How do you see yourself
    Not at all masculine ----------> Extremely masculine
    Not at all feminine ----------> Extremely feminine

The questions for Condition 4 have non-binary options & a non-traditional answer order:

1) Sex
    a. Female
    b. Intersex
    c. Male

2) Gender
    a. Woman
    b. Transgender
    c. Man
    d. A gender not listed here
        i. Please specify _____

3) How do most people see you
    Not feminine ----------> Extremely feminine
    Not masculine ----------> Extremely masculine

4) How do you see yourself
    Not feminine ----------> Extremely feminine
    Not masculine ----------> Extremely masculine

Within each condition, we will test four variations in the overall order of the questions:

- Sex first (this order is in some sense chronological)
    - Sex
    - Gender
    - Scales

- Gender first (prioritizing current gender, which some perceive as more respectful of transgender people)
    - Gender
    - Sex
    - Scales

- Scales first (prioritizing gender diversity)
    - Scales
    - Gender
    - Sex

- Sex, Scales, Gender (testing whether responses on the scales might be influenced as a result of the desire for current gender category conformity)
    - Sex
    - Scales
    - Gender

After answering these four questions – each presented on their own screen – the respondent will see a final screen asking for their feedback on the questionnaire:

We are interested in your feedback on the questions you just answered. If you have questions or concerns, or if there is any additional information you would like to share, please do so in the space below.

GSS proposal resubmission 1

Gender Identity and Diversity

Revision and updates

August 2014

Shelley Correll
Cecilia Ridgeway
Aliya Saperstein*
Laurel Westbrook

*Corresponding author. Contact: asaper@stanford.edu

Executive Summary

Our pre-test on Amazon Mechanical Turk in May was a success. Counter to some of the Board's stated concerns, nonresponse was negligible and very few people left feedback suggesting that they were offended or otherwise reacted negatively to the questions. The handful of negative comments were offset by respondents thanking us for providing more than the usual answer options or otherwise expressing support for the question design. Although a few people noted that they found the masculine and feminine scales difficult to answer, two hundred volunteered explanations of why they responded the way they did, and many seemed to enjoy the opportunity to think about or describe their gender in a new way. One percent of respondents selected a sex at birth and/or current gender that would not be captured using a standard static, binary, single-measure method, and just 15 percent responded to the masculine and feminine scales in a way that suggests the categories are diametrically opposed. Overall, we believe the pre-test provides strong evidence that including the full module, with non-binary answer options, not only would be feasible as part of the 2016 General Social Survey but also would provide important insights on the range of gendered experience among U.S. adults.

That said, the Board requested we prioritize among the items to clearly identify the four most important questions. Based on our pre-test results, and considering what would best complement existing measures in the GSS, we rank the items in the following order of priority:

1. First order masculine and feminine scales (How do you see yourself?)
2. Sex assigned at birth
3. Gender identity
4. Third order masculine and feminine scales (How do you think most people see you?)

We provide our justification for this ranking, and details on how each of the items performed during our pre-test, below. Our specific recommendations for the question wording and answer options appear, in priority order, in Appendix 1 – just as they would appear on a computer screen. We also note that we have a second pre-test planned for later this Fall, in which we plan to include our full module along with more than two dozen questions from the GSS core. This pilot study will demonstrate whether inclusion of the new gender identity and diversity items has any effect on either subsequent questions or overall survey response, and provide a first look at whether the previously unmeasured variation revealed by our module sheds new light on gender inequality or other attitudes and behaviors of interest to GSS researchers.

We welcome any recommendations the Board might have for which GSS items should be included on our next pilot study, as well as any other suggestions for split ballots or other variations in survey conditions that would further enhance our proposal and improve the chances that our items will be selected for the 2016 survey. (See Appendix 2 for a brief description of our planned study design and a list of the GSS items we are considering for the second pre-test.)

The pre-test

As outlined in our initial proposal, we fielded a seven-question survey on Mturk in order to assess respondents' willingness to answer detailed questions about their sex at birth, current

gender identity, how they perceived their masculinity and femininity, and how respondents thought others perceived them. We also tested variations in answer options, answer order and question order, and asked respondents to give us feedback on the questions they had answered. Respondents were offered 25 cents for their participation, and were only allowed to take the survey once. They were also required to be 18 years or older, U.S. residents and have an Mturk Human Intelligence Task (HIT) approval rating of at least 80 percent.

Our goal was to recruit 1,600 respondents, randomly assigned to each of our four major conditions: 1. binary answer options and conventional answer order (male, female/man, woman/masculine, feminine); 2. binary answer options and unconventional answer order (female, male/woman, man/feminine, masculine); 3. non-binary answer options and conventional answer order (male, female, intersex/man, woman, transgender, other/masculine, feminine); and 4. non-binary answer options with an unconventional answer order (female, intersex, male/woman, transgender, man, other/feminine, masculine). To speed up data collection, we reposted the HIT several times over the course of a three-week period, most often in the early morning. This resulted in responses arriving in batches that we flagged with unique indicators. All findings were checked for variation by batch. No significant differences were found. The final sample of 1,521 excludes all duplicate IP addresses and Mturk IDs, 40 respondents whose IP addresses indicated they submitted the survey from outside the U.S., and 12 respondents who completed the survey in less than 25 seconds. See Table 1 for basic descriptive statistics by survey condition, including average completion time and the regional distribution of responses (assigned based on the respondent's IP addresses). Two cases in our analytic sample are missing on region, one is missing on the first order scales and one is missing on sex at birth, but not the other measures.

Thirty-five percent of respondents (530) left some form of feedback in response to the final question on our survey. Twelve percent (183) simply responded with "none," "n/a" or something similar to indicate they had no questions or concerns. One way to interpret this result is that the vast majority of respondents did not perceive the survey itself or the non-binary answer options and unconventional answer order as unusual, confusing, or otherwise worthy of comment. The amount of feedback received did vary somewhat across conditions, with respondents in the non-binary answer option and unconventional answer order condition leaving the fewest comments, while respondents in the non-binary answer option and conventional answer order condition left the most (see Table 1). Explicit expressions of support for the survey were most common in the binary and conventional condition and decreased somewhat as the survey became more unconventional, while the few explicit expressions of discomfort or other negative reactions to the survey were relatively equally distributed. Indeed, it would seem that the negative responses we did receive were prompted more by the implication from the series of questions that one's sex and gender might differ at all than by the specific ways respondents were offered to express any such differences. None of the substantive feedback suggested that respondents felt threatened by the scale items inquiring about their levels of masculinity and femininity. (For examples of feedback received, see Appendix 3.) However, we did find significant differences in average responses to the masculine and feminine scales depending on whether respondents answered those items first or last. We discuss these results, and their implications, in more detail below.

Measuring masculinity and femininity

The masculine and feminine scales were presented with seven response categories, ranging from "not at all" to "extremely." Only the poles of the scales were given nominal labels, other options were marked with numbers. (We did not give nominal labels to all points to avoid possible negative connotations, such as describing oneself as "somewhat" masculine.) All respondents rated themselves on both masculine and feminine scales, from both the third-order (How do you think most people see you?) and first-order (How do you see yourself?) perspectives, for a total of four scale ratings per respondent. The scales were presented with the third-order item first, followed by the first-order item.

Average responses did not differ significantly between the first and third order scales (see Table 1). This suggests that the vast majority of respondents experience little dissonance between how they think other people perceive them and how they see themselves – or that they see little distinction between the two perspectives. We do plan to include both sets of scales in our second pre-test, this time counterbalancing their question order (between self and other perception) to see if that encourages greater differentiation on the two sets of scales. Otherwise, it would seem that asking only one perspective or the other will suffice to allow most people to express their current levels of masculinity and femininity. If only one set of scales were to be used, we recommend asking the first-order question (How do you see yourself?) because it provides the clearest counterpoint to the interviewer's classification of the respondent's sex/gender that occurs at the beginning of the GSS.

Respondents did make use of the full range of responses on the seven-point scales. Although few respondents selected the "not at all" response on the scale consistent with their sex or gender, they did not appear to avoid the "extremely" sex/gender typical points of the scales (see Table 2). The majority of males responded using 1 to 3 on the feminine scale (86 percent) and 5 to 7 on the masculine scale (87 percent), and vice versa for females. Despite this clear clustering, 15-24 percent of respondents chose responses either in the middle or at the opposite end of the spectrum consistent with their sex category (the frequency varied somewhat by sex and scale). In our second pre-test, we plan to substitute "very" at the high end of the scale instead of "extremely" (which might have a negative connotation for some respondents) to determine whether it provides similar results. For now, we continue to recommend a seven-point scale because of the distribution of responses on our first pre-test. Doing so offers more nuanced information about respondents' masculinity and femininity but does not preclude researchers from collapsing categories if that best fits their research design.

We also affirm the importance of maintaining the masculine and feminine scales as separate items, rather than combining them into a single scale with "extremely" (or "very") masculine at one pole and "extremely" (or "very") feminine at the other. Fifteen percent of respondents selected the sex-category-consistent poles on both scales, 20 percent selected an extreme sex-consistent response on one scale or the other but not both, while approximately two-thirds of respondents chose non-extreme responses on both scales (not shown). Females also reported higher levels of masculinity, on average, than males reported for femininity, a result that would be hidden on a single bipolar scale. Further, by gauging masculinity and femininity separately, researchers have the flexibility to include either or both measures as variables in their analysis. It

is also possible to calculate a "polarization" score from the two scale responses that allows the extent of separation between one's masculinity and femininity to become an empirical question. Comparing average gender polarization scores over time and across contexts could open intriguing new lines of research about trends in gender expression.

Other results of interest from our pre-test include finding significant variation in responses to the scales by region and by when the scales appeared in the survey (see Table 3). Gender polarization was the lowest in the West and highest in the South for both males and females. Polarization increased by 0.27 points when the scales appeared third compared to when they appeared first in the survey. (This difference was statistically significant controlling for region, sex at birth, survey condition and batch.) Although males gave more polarized responses overall, females were most likely to give extreme sex-consistent responses to both scales when the scales appeared third, following the sex at birth and current gender items. Extremely polarized responses (i.e., polarization scores of 6) were also more common in general in the Midwest and the South, statistically significantly so for females (not shown).

We interpret the variation in scale responses by question order to suggest that respondents reacted to seeing questions about their sex at birth and current gender by asserting more extreme, sex-category-consistent responses on the masculine and feminine scales. When the scale items appeared first (rather than second or third), respondents on average selected less extreme responses. It is possible that the scale items would be sensitive to other question order effects, as well, and thus we recommend that they do not follow directly after items assessing gendered beliefs or stereotypical behaviors (e.g., who does the housework).

A note on the third order scales. Although comparing average scores on the masculine and feminine scales between first-order and third order perspectives yielded no significant differences, the two sets of scales did not entirely replicate one another (see Table 4). Twenty-two percent of respondents gave distinct first- and third-order responses on both masculine and feminine scales, while 35 percent differed on one scale or the other; however, the vast majority of respondents differed by only one point. A lack of concordance between the scales was particularly noticeable for categorically gender variant respondents (not shown), and we expect that examining when first and third order perspectives diverge could open up interesting lines of research. We plan to further test and refine both sets of scales in our second pre-test and hope to have more encouraging results to report regarding the contribution of the third order scale in time for the Spring 2015 meeting, if the Board is still considering our proposal at that time.

Capturing categorical gender variance

In addition to the range of gender expression revealed by the scale items, one percent of our sample (15 respondents) could be described as categorically gender variant based on their responses to the item asking for their current gender and/or the sex they were assigned at birth (see Table 5). These respondents either selected gender identities outside the traditional man/woman binary, or gave a sex at birth and current binary gender identity that differed, suggesting they had transitioned from male at birth to woman (i.e., a trans woman) or female at birth to man (a trans man). None of our respondents identified as intersex, which might reflect not only their small numbers in the population but the lack of consensus on how to best identify

them in population surveys (see GenIUSS forthcoming). Of the four respondents who selected the alternative "a gender not listed here" response for their current identity, two specified "genderqueer," one wrote in "androgynous" and one did not specify. However, it is important to note that our results likely underestimate the number of people who would have offered non-binary gender responses because only half of the respondents in our pretest could explicitly identify as a gender other than "man" or "woman" (i.e., those who were assigned to one of the two non-binary conditions). Indeed, three respondents assigned to binary answer option conditions reported different categories for their sex at birth and their current gender, providing an indirect measure of their transgender status.

A frequency of transgender respondents between 0.3 and 0.7 percent (using either direct or indirect measurement) is in line with estimates based on samples in Massachusetts and California (Gates 2011). Should the GSS implement our recommended two-step approach of asking sex at birth and current gender identity, using non-binary answer options, we estimate that as many as 45 respondents will identify (or be identifiable) as categorically gender variant. Other methods and smaller GSS sample sizes will result in fewer gender variant cases (see Table 6). Pooling across years likely would be necessary for detailed multivariate or sub-group analysis of gender variant populations (e.g., comparing trans men to trans women), but that is the case for many other studies of interest with GSS data, including comparing outcomes across the youngest and oldest birth cohorts.

Although the GSS currently includes a measure of the respondent's sex/gender that is recorded by the interviewer at the beginning of the survey (the variable SEX), we stress the importance of using the additional two-step method as part of the self-administered questionnaire for several reasons. At the moment, the interviewers' appraisal of respondents captures some combination of their observation of the respondents' physical appearance with other cues such as the respondents' names and relationships to other members of their household. Interviewers are advised during training to ask respondents when they are unsure, but they are not explicitly required or given a script to do so (unlike they were, e.g., in the pre-2000 GSS race question). Thus, most of the time, the SEX variable will capture the sex/gender the respondent is currently presenting, and perhaps presents on a regular basis. It is not clear what the interviewer's classification would capture for respondents in the midst of transition or who otherwise do not conform to gender norms (e.g., our respondents who wrote in alternative gender identities). The two-step approach is also the only way to identify Americans who have transitioned but prefer to identify with a traditional gender category. Four categorically gender variant respondents (three trans men and a trans woman) identified this way, despite being assigned to a non-binary condition and thus having the option to explicitly identify as transgender.

If adding the two-step approach is not possible due to funding or space constraints, we have a slight preference for including the question regarding the respondent's sex at birth over the respondent's current gender. We make this recommendation with reservations related to the tradeoffs between the information in the two items. However, we assume that the respondent's sex at birth is information the GSS interviewer would be less likely to know simply from the household enumeration form and early interactions with the respondent (e.g., if he or she had transitioned). It is also the closest proxy to the respondent's biological sex, which is what many researchers think they are using when including the current SEX variable in their analyses.

Further, if the interviewer's classification of the respondent more closely approximates a binary measure of the respondent's current gender, then it would still be possible to compare the interviewer's classification with the respondent's reported sex at birth to indirectly identify some transgender respondents. That said, not all respondents who consider themselves "transgender" will be captured by this comparison. (See Schilt and Bratter forthcoming for a discussion of the range of possible transgender responses.)

We expect our second pre-test will offer additional evidence that including the entire gender identity and diversity module does not compromise overall survey response, and we believe that knowing each facet, not alone but in combination, provides important context for understanding Americans' experiences with sex and gender inequality. The masculine and feminine scales are an improvement over the existing SEX variable because they allow for a spectrum of gender expression, but they still require responses relative to the traditional sex/gender binary. Sex at birth, in combination with the current interviewer classification, will capture some changes over time in respondents' identities, but only asking respondents about their current gender, including at least one option beyond "man" and "woman," truly allows Americans to identify beyond the traditional binary if they so choose. It is not clear if a growing number of Americans are identifying their gender in more complex ways (Chalabi 2014), or if survey researchers simply have not been asking the right questions all along. Either way, it is important to capture this trend and track its impact on other domains in people's lives. In doing so, the GSS would be at the forefront of efforts to better understand gender identity and diversity in the United States.

References

Chalabi, Mona 2014. "Why We Don't Know the Size of the Transgender Population." *Five Thirty Eight Life*, July 29. http://fivethirtyeight.com/features/why-we-dont-know-the-size-of-the-transgender-population/

Gates, Gary J. 2011. "How Many People Are Lesbian, Gay, Bisexual and Transgender?" Los Angeles, CA: The Williams Institute. http://escholarship.org/uc/item/09h684x2

GenIUSS. Forthcoming. "Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys." Los Angeles, CA: The Williams Institute.

Schilt, Kristen and Jenifer Bratter. Forthcoming. "From Multiracial to Multigender? Assessing Attitudes toward a Transgender Category on the U.S. Census." *Transgender Studies Quarterly*.

Table 1. Sample descriptives, Gender Identity and Diversity Pre-test on Amazon Mechanical Turk

		Condition				
		Binary, conventional	Binary, unconventional	Non-binary, conventional	Non-binary, unconventional	Total
Sex at birth						
Male		63%	64%	58%	63%	62%
Female		37%	36%	42%	37%	38%
Intersex		-	-	0%	0%	0%
Gender identity						
Man		63%	64%	57%	63%	62%
Woman		37%	36%	41%	36%	37%
Trans		-	-	0.8%	0.3%	0.5%
Another gender (specify)		-	-	0.5%	0.5%	0.5%
First order masculine scale	(mean)	3.6	3.4	3.2	3.5	3.4
Third order masculine scale		3.5	3.4	3.2	3.4	3.4
First order feminine scale		2.3	2.4	2.6	2.4	2.4
Third order feminine scale		2.4	2.4	2.6	2.5	2.5
First order polarization		3.4	3.5	3.4	3.4	3.4
Third order polarization		3.4	3.5	3.5	3.5	3.5
Left any feedback		35%	35%	40%	29%	35%
Explanation of response		15%	13%	14%	11%	13%
Generic gratitude		3%	6%	6%	4%	5%
Explicit support		3%	2%	1%	0.8%	2%
Discomfort		1%	0.5%	0.8%	1%	0.9%
Region						
West		25%	25%	21%	24%	24%
Midwest		21%	23%	23%	19%	21%
Northeast		21%	18%	22%	21%	21%
South		33%	34%	34%	35%	34%
Total survey duration (secs)		80	80	84	77	80
N		383	382	378	378	1521
% of sample		25.2%	25.1%	24.9%	24.9%	100.0%

Note: Sample includes cases with U.S. IP addresses, unique IP addresses or Mturk IDs, and completed surveys with durations greater than 24 seconds. The distribution of the sample by sex (or gender) does not mirror the U.S. population, but is in line with the known skew among Mturk workers (e.g., Richey and Taylor 2012 http://themonkeycage.org/2012/12/19/how-representative-are-amazon-mechanical-turk-workers/; though see Berensky et al. 2011). The total percentage presented for non-binary genders reflects the percentage among respondents who were offered non-binary gender responses, which was one-half of our total sample. Masculine and feminine scales were recoded from 0 to 6 and averaged; scale polarization is the absolute value difference between the masculine and feminine response categories selected by the respondent. Region is derived from IP address, coded using detailed GSS categories, then collapsed.

Table 2. Distribution of first-order masculinity and femininity scale scores by reported sex at birth

			Proportions						
	Mean	*Standard deviation*	Not at all	1	2	3	4	5	Extremely
Male at birth									
Masculine scale	4.6	1.1	1%	1%	2%	8%	28%	39%	20%
Feminine scale	1.2	1.2	34%	33%	17%	9%	5%	1%	0.2%
Polarization	3.5	1.7	7%	8%	14%	16%	25%	16%	16%
Female at birth									
Feminine scale	4.4	1.3	1%	2%	5%	12%	30%	30%	21%
Masculine scale	1.5	1.4	32%	24%	20%	13%	8%	2%	1%
Polarization	3.2	1.9	10%	13%	17%	13%	18%	14%	16%

Note: N=1,519; 944 males and 575 females. Scales were recoded from 0-6. Polarization is the absolute value of the difference between the two scales.

Table 3. Polarization scores by region and scales order

	Males	Females
Region		
West	3.3	2.8
Midwest	3.6	3.2
Northeast	3.6	3.0
South	3.7	3.6
Total	3.5	3.2
Scales order		
First	3.4	3.1
Second	3.5	3.1
Third	3.6	3.4
Total	3.5	3.2

Note: Two-tailed tests indicate that polarization scores for males are significantly higher than for females overall, and in each region, except the South. Polarization scores are also significantly higher in the South compared to the West for both males and females. Differences in polarization by scales order are not statistically significant on average, but the difference between when the scales were presented first compared to third is statistically significant in an OLS regression controlling for sex at birth, condition, region and batch.

Table 4. Divergence between sex-category consistent first and third order scales

Male at birth

	First order masculine scale							
Third order	Not at all	1	2	3	4	5	Extremely	Total N
Not at all	0	0	0	0	0	0	0	0
1	4	0	1	1	0	0	0	6
2	0	6	0	13	3	0	0	22
3	2	0	7	0	26	4	0	39
4	1	0	5	23	0	76	5	110
5	1	0	1	8	83	0	51	144
Extremely	1	0	0	1	5	17	0	24
Total N	9	6	14	46	117	97	56	345

Female at birth

	First order feminine scale							
Third order	Not at all	1	2	3	4	5	Extremely	Total N
Not at all	0	0	0	0	0	0	0	0
1	2	0	5	0	0	0	0	7
2	2	3	0	2	2	0	0	9
3	3	0	5	0	11	2	0	21
4	0	0	7	33	0	41	3	84
5	0	0	3	4	60	0	23	90
Extremely	0	0	0	1	4	16	0	21
Total N	7	3	20	40	77	59	26	232

Table 5. Cross-tabulation of reported sex assigned at birth and current gender

	Self-identified gender				
	Man	Woman	Transgender	Another gender (specify)	Total
Sex assigned at birth					
Male	938	2	3	2	945
	99.3%	0.2%	0.3%	0.2%	100.0%
Female	5	567	1	2	575
	0.9%	98.6%	0.2%	0.4%	100.0%
Intersex	0	0	0	0	0
Total	943	569	4	4	1,520
	62.0%	37.4%	0.3%	0.3%	100.0%

Note: Row percentages may not sum to 100 due to rounding. Only half of the sample (756 respondents) was offered non-binary gender response options. Total N does not include one self-identified man who said he accidentally skipped the sex at birth question.

Table 6. Estimates of categorical gender variance by measurement approach for full GSS sample

	Percent of sample	*Estimated N*		
		Sample=2000	Sample=2800	Sample=4500
Measurement approach				
Current GSS interviewer classification	0%	0	0	0
One-step, gender identity, non-binary	0.6%	12	17	27
One-step, sex at birth (compared to interviewer classification)	unknown	unknown	unknown	unknown
Two-step, binary	0.5%	10	14	23
Two-step, non-binary	1.0%	20	28	45

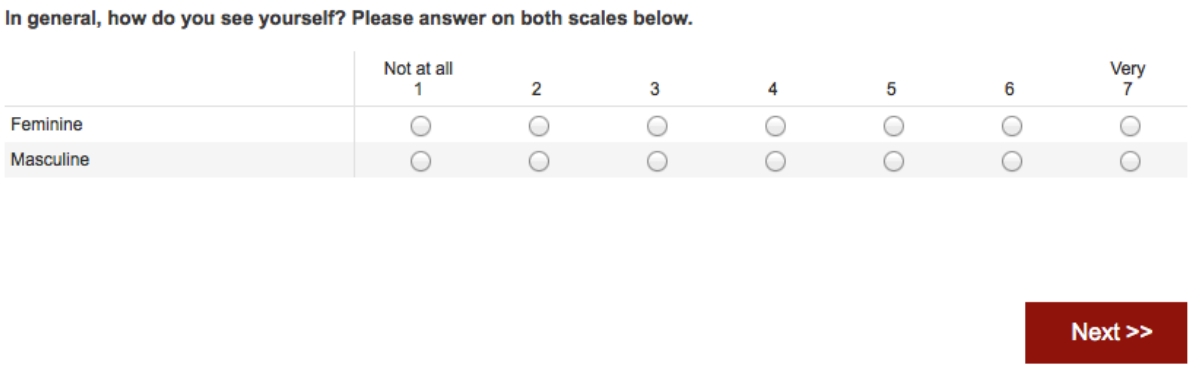
Note: Percentages draw on results from the Mturk pre-test. Estimated Ns represent a range of sample sizes available in recent GSS surveys. Two-step binary approach only identifies transgender respondents indirectly, by comparing sex at birth with current gender.

Appendix 1: Recommended items

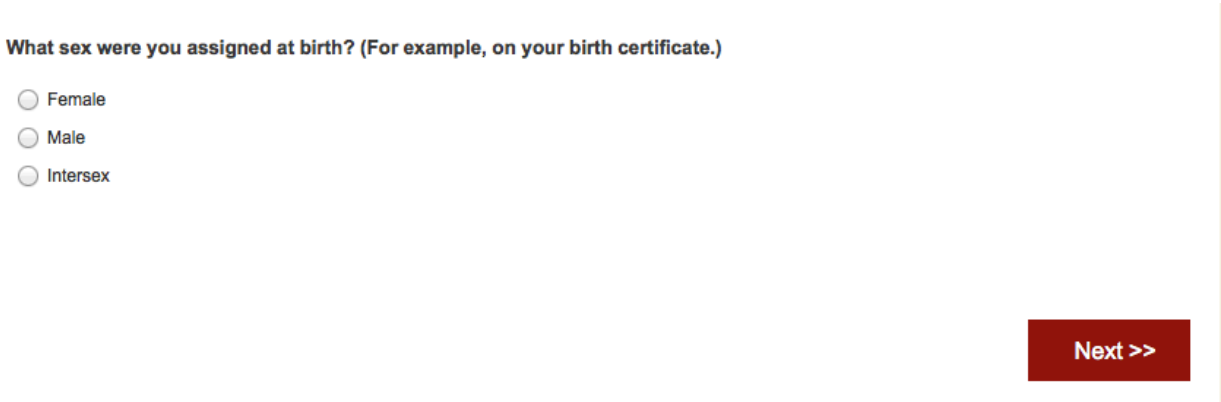
Our recommendations include an answer order that presents the responses in order of population size. The population size response ordering is something of a compromise between the conventional and unconventional response orders in our pre-test – neither of which drew any comment (positive or negative) in the feedback from our pre-test respondents. Each item appears on its own page.

Please note that we are not recommending that the first-order and third-order scales be presented at opposite ends of the module, as appears below. We are simply illustrating our priority ranking for the items here. However, we do recommend that if either or both sets of scales are included the scales should be presented to respondents before the sex at birth and gender identity items to avoid what appeared to be gender "overcompensation" in scale responses (cf. Willer et al. 2013).

First-order scales:



Sex at birth:



Gender identity:

What is your current gender? (How you describe yourself.)

- ○ Woman
- ○ Man
- ○ Transgender
- ○ A gender not listed here (please specify)

Next >>

Third-order scales:

In general, how do you think most people see you? Please answer on both scales below.

	Not at all 1	2	3	4	5	6	Very 7
Feminine	○	○	○	○	○	○	○
Masculine	○	○	○	○	○	○	○

Next >>

Appendix 2: The second pre-test

In late October we plan to recruit another 1,600 Mturk respondents to complete a 20-minute survey drawn from actual items employed by the General Social Survey. These include measures of socioeconomic status, health and well-being, as well as attitudes and behaviors that we expect are likely to vary along the different dimensions of sex and gender we are proposing for the 2016 survey round. The items will appear in a similar order as they appear in the actual GSS questionnaire and with the same question wording and answer options. Our goal is to mimic the original GSS as closely as possible – excepting the addition of our module.

This time, our survey conditions will allow for three key comparisons: 1) whether the sex and gender items appear at the beginning of the questions borrowed from the GSS self-administered questionnaire (SAQ) or at the end of the survey; 2) whether the third-order scale appears first or the first-order scale does; and 3) whether the scales both appear before the sex at birth and current gender items, or after them. The latter comparison is a necessary carry-over from our first pre-test because we plan to change the high-end scale response labels to "very," and we do not want to attribute changes in results to the new nominal label that actually reflect the question ordering or vice versa. By testing the placement of the gender identity and diversity module relative to other items in the survey, we do not mean to imply that we recommend the module be included at the end of the SAQ. In fact, we believe the end of the SAQ would not be a good place for the new items because they would follow questions on sex practices and sexual orientation, which is likely to affect responses, especially to the scale items. However, we want to include a small sample that answered all previous questions on the survey before seeing our proposed module in order to determine whether or not seeing the module earlier in the questionnaire is likely to affect responses on later items. Our expected sample sizes, by condition, appear below:

	First-order scale first	Third-order scale first
Module starts SAQ portion		
N= 1200		
Both scales first	300	300
Both scales last	300	300
Module ends survey		
N = 400		
Both scales first	100	100
Both scales last	100	100

The list that follows specifies the items from the GSS that we plan to include, in the order the questions will be asked. We specifically included items on employment status, economic hardships, health, and experiences of violence because previous research indicates that transgender Americans are particularly likely to be unemployed, marginally housed, subject to violence and lacking access to appropriate health care (see, e.g., GenIUSS forthcoming).
We also changed the format of some items, including the NAT- series that opens the survey (and the GSS), to make them more efficient in our fully self-administered Internet mode.

(NATENVIY, NATCRIMY, NATEDUCY, NATARMSY, NATFAREY & NATSCI modified)

We are faced with many problems in this country, none of which can be solved easily or inexpensively. Some of these problems are listed below. For each one, please indicate whether you think we are spending too much money on it, too little money, or about the right amount.

	Spending too much	Spending too little	Spending the right amount
The environment	☐	☐	☐
Law enforcement	☐	☐	☐
Education	☐	☐	☐
National defense	☐	☐	☐
Scientific research	☐	☐	☐
Assistance for the poor	☐	☐	☐

(DISCAFFM)

What do you think the chances are these days that a man won't get a job or promotion while an equally or less qualified woman gets one instead. Is this very likely, somewhat likely, somewhat unlikely, or very unlikely these days?

(POSSLQY, substitutes "spouse" for husband and wife)

Which of these statements applies to you?

I am married and living in the same household as my spouse
I have a steady partner and we live in the same household
I have a spouse or steady partner, but we don't live in the same household
I don't have a steady partner

(WRKSTAT, check all that apply; will recode according to GSS instructions after the fact)

Last week, were you working full time, part time, going to school, keeping house, or what?

(WEEKWRK)

How about last year? In 2013, how many weeks did you work either full-time or part-time?

(FEAR, modified)

Is there any area around your home–that is, within a mile–where you would be afraid to walk alone?

(FEARHOME, modified)

How about at home – do you feel safe and secure, or not?

(LAW2, modified)

In the past 12 months, were you beaten up, hit, or attacked with a weapon?

(CLASS)

If you were asked to use one of four names for your social class, which would you say you belong in: the lower class, the working class, the middle class, or the upper class?

(FINALTER)

During the last few years, has your financial situation been getting better, getting worse, or has it stayed the same?

(HISP1 & 2)

Are you Spanish, Hispanic or Latino?

(if yes, Which group are you from?)

(RACECEN)

What is your race? Indicate one or more races that you consider yourself to be.

(DISCAFF)

What do you think the chances are these days that a white person won't get a job or promotion while an equally or less qualified black person gets one instead? Is this very likely, somewhat likely, or not very likely to happen these days?

(MARHOMO)

Do you agree or disagree with the following statement: Homosexual couples have the right to marry one another.

(HAPPY)

Taking things all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?

(HEALTH)

Would you say your own health, in general, is excellent, very good, good, fair or poor?

(MNTLHLTH2)

Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

(PRTYPREF)

Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?

(POLVIEWS)

There is a lot of talk these days about liberals and conservatives. Below is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale?

(BORN)

Were you born in this country?

(PARBORN)

Were both of your parents born in this country?

(GRANBORN)

Were all of your grandparents born in this country?

(RGRADE)

What is the highest grade in elementary school or high school that you finished and got credit for? (List from "no formal schooling" to 12^{th} grade)

(RDIPLOMA)

Did you ever get a high school diploma or a GED certificate?

(RCOLDEG)
Do you have any college degrees? (if yes, What was the highest degree you earned?)

(ERNRS)
How many persons in your family – those people in your household who are related to you – earned any money last year – 2013 – from any job or employment (including yourself)?

(INCOME06, with categories A-Y)
In which of these groups did your total family income, from all sources, fall last year – 2013 – before taxes? (Total income includes interest or dividends, rent, Social Security other pensions, alimony or child support, unemployment compensation, public aid (welfare), armed forces or veteran's allotment.)

(DWELOWN)
Do you (or does your family) own your home, pay rent, or what?

(BIRTHYR, modified to comply with Mturk terms of service)
What year were you born?

[Scales and sex/gender self-identifications will most often be included here]

(ARREST2)
Were you ever picked up or charged by the police, for any reason whether or not you were guilty?

(LOCKEDUP)
Have you ever spent any time in prison or jail?

(RHHWORK)
On average, how many hours a week do you personally spend on household work, not including childcare and leisure time activities?

(ORGINC)
(If R has spouse/partner) How do you and your (spouse/partner) organize the income that one or both of you receive? Please choose the option that comes closest.

(HIGHINC)
Considering all sources of income, between you and your (spouse/partner), who has the higher income?

(CUTAHEAD, VOLCHRTY, GIVCHRTY, HELPAWAY, CARRIED, TALKEDTO)
During the past 12 months, how often have you done each of the following things:

	At least once a week	At least once a month	At least once in the past year	Not at all in the past year
Allowed a stranger to go ahead of you in line	☐	☐	☐	☐
Done volunteer work for a charity	☐	☐	☐	☐
Given money to a	☐	☐	☐	☐

charity				
Looked after a person's plants, mail or pets while they were away	☐	☐	☐	☐
Carried a stranger's belongings, like groceries, a suitcase, or shopping bag	☐	☐	☐	☐
Spent time talking with someone who was a bit down or depressed	☐	☐	☐	☐

(HRDSHP series, modified)

During the past 12 months, did you find yourself in any of the following circumstances?

	Yes	No	Not applicable
Fell behind in paying your rent or mortgage	☐	☐	☐
Unable to purchase needed food	☐	☐	☐
Unable to purchase needed medical care	☐	☐	☐
Had to temporarily live with others or in a shelter or "on the street"	☐	☐	☐

(PARTNERS)

How many sex partners have you had in the last 12 months?

(SEXFREQ)

About how often did you have sex during the past 12 months?

(SEXORNT)

Which of the following best describes you?

(Gay, lesbian or homosexual; bisexual; heterosexual or straight; don't know)

(SEXSEX18)

Since the age of 18, have your sex partners been:

Only men
Some women, but mostly men
Equally men and women
Some men, but mostly women
Only women
I have not had any sex partners

[In some conditions, the scales and sex/gender self-identification would appear here]

(VERADDR, modified)

Please enter your current state and zip code.

Thank you very much for your time and help. That is all the questions we have for you.

Appendix 3: Assessing Feedback

After completing the six sex and gender items, respondents were asked a follow-up question soliciting feedback on the survey. Just over a third of the sample took the time to type some response, though many were one word answers such as "none," "n/a" or simply "thanks!" It is important to note that, for serious Mturk workers, taking the time to respond to our feedback question would lower their pay rate relative to not answering. Thus, the responses we received are likely not representative of everything respondents might have had to say in a controlled cognitive interview setting. Nevertheless, they offer an important window into how American adults are likely to react to being asked questions about their gender expression and identity.

We were particularly interested in three types of feedback: positive responses to the survey, negative responses to the survey, and requests for clarification. These were the key distinctions we made when coding the comments for content. However, a category we did not expect – explanations for responses, especially on the scale items – may also be of interest to the Board and is summarized below.

Twice as many female respondents left explicitly supportive feedback, while male respondents dominated the discomfited responses. Four of the 15 categorically gender variant respondents left feedback: two expressed explicit support for the survey and two offered explanations to contextualize their responses. Respondents were more likely to leave comments, explanatory or otherwise, if they responded to the scale items last – immediately before being asked for their feedback. Respondents who gave more atypical, or less polarized, responses on the masculine and feminine scales also were significantly more likely to take the time to explain why they did so. (This result held net of sex at birth, region, survey condition and batch.)

Explicit support

Twenty respondents provided explicitly supportive feedback that communicated gratitude for non-binary options, support of non-binary genders, or other interest in the content of the study. Comments that were positive but vague were not included. For example, comments that said "interesting" but did not explicitly refer to an aspect of the survey's content left it unclear whether "interesting" was meant as praise. Also not included in this group are the 54 comments expressing generic support or gratitude ("neat survey!" "Thank you!"), or the 50 comments that described the survey as "easy," straightforward, or indicated the respondent had no concerns about the questions. The explicitly supportive comments included:

- Thank you for differentiating between sex assigned at birth, and gender identity.
- This is the best survey ever--I get so sick of surveys asking which "gender" I am and then providing only two options, neither of which are my gender identity.
- Adding a third option or something would really add to this -- plenty of people prefer identification that's a bit more complicated than male/female! *(from respondent assigned to a binary condition)*
- Thanks for recognizing that not all women are female!
- I think those questions do provide a sense of reflecting how a person feels about him or herself. I think many people today do feel different from how people see them. (*male respondent*)
- It was interesting to see that you actually differentiated between sex and gender(how we describe ourselves) and I have never thought of how I percive my sex gender vs how others view me.
- Liked the first one "what sex were you assigned at birth" nothing really (*presumably meaning no questions or concerns)*

- Most men and woman have masculine and feminine qualities which I think is perfect.
- I completely beleive people feel different than what their gender may be, probably due to different brain wiring. I am all for transgender, I just know that I myself, am in the correct body.
- No concerns, I hope the purpose is to help the general population be more understanding/accepting of those who do not identify with the gender assigned to them at birth.
- No concerns, I'd be curious to see the results!! We learned in my psychology class that gender is a spectrum and I'd be curious to see if these results supported that! (*male respondent*)
- That was really brief! I guess it would be interesting to see a breakdown of feminine and masculine characteristics (maybe not identified as such) and see how that's related to how people view themselves on the spectrum of femininity and masculinity.
- They were interesting questions about how we perceive ourselves
- Thank you for allowing for the possibility of being neither masculine nor feminine instead of presenting the two options as a single sliding scale! I don't have a problem being female, but I don't feel any particular attachment to my gender, or to the concept of gender in general.
- Although I am straight and have never had any gender issues, I feel like there is a femanine part of me that only I know about. That is why I answered the way people perceive me different than what I feel. Also I enjoyed participating in this study. (*male respondent*)
- It is interesting that the survey asked questions to do with femininity and masculinity, but it differentiated between the sex given at birth vs. what and how we identify ourselves as adults. Very interesting.
- I like that you ask our gender, most surveys just assume we are male or female and that's it. Seems like other surveys don't get the difference between sex and gender.

Discomfort or concern

Thirteen respondents left negative feedback about the content of the survey, broadly construed. Twelve of these comments came from male respondents. Included in this group are respondents who imply they did not have a problem with the survey, but suggest others might. All of these comments are reproduced below:

- I don't think the term "assigned at birth" is fair or accurate. Gender at birth is determined by anatomy, not by societal labels.
- It is really a shame that such questions need to be asked about one's gender.
- I don't have any questions. I think it was very simple to answer. I think some people might not be too happy with the question "What gender were you assigned at birth?" Especially extreme conservatives.
- I find it sad that we are asking questions like this in today's society. I do not understand why people want to change their gender. This is so sad.
- Interesting questions about whether I feel like a man or woman. I think this is a controversial topic because you have to identify as a male or female for most things.
- I don't know what kind of feedback you want, but here is what I have to say. I was born a male, and I have always lived as a male. I know there are people out there who feel they are confused on their gender role, but I am not one of them. God has a plan for my life, and that is as a male.
- I was not "assigned" a sex at birth. I'm a man. Science!
- These kind of questions tend to make me feel uncomfortable.
- These questions just make me sigh.
- I am a man! What kind of questions are these?
- Just seems like a normalish demographic survey considering the age we live in.
- What the hell is going on in this world--what is my CURRENT sex? Too bizarre. What is going on with all this "gender" obsessiveness is unnatural and sick.Wake up. *(female respondent)*
- Seems like natural questions to ask in the social justice climate of today.

Another two dozen respondents made comments indicating they thought some or all of the survey was "odd," "weird" or "strange." Some of these respondents may have meant their comments as a critique; some also seemed to be expressing surprise or curiosity about the goal of the study. Unlike the previous two feedback categories, male and female respondents were represented in this group in proportion to their distribution in the survey as a whole. Nevertheless, including these comments with the more clearly concerned respondents provides a generous estimate of those bothered by the survey at 2.5 percent of our sample. This broader group includes comments such as:

- The "what is your current gender" question threw me. I wonder how many transsexuals you get taking these surveys.
- Really weird survey, quick and easy though
- All little odd but no real questions
- Would love to know what on earth this is about
- Strange survey, but interesting
- I think they are odd questions.
- They were very simple questions and easy to understand. I don't really have any feedback. I was kind of surprised by the questions to be honest. I did not expect to be asked that, but I am not offended or anything. It's just not something I normally give much thought to.

We should also note that one respondent, assigned to the binary/unconventional condition, appeared to click through the entire survey but did not answer any questions. One man assigned to the non-binary/conventional condition skipped sex at birth, answered current gender and the scales, but then failed to click through the feedback page. It is not clear whether these respondents might have broken off their participation out of discomfort or failed to complete the Mturk HIT for other reasons. Two other men skipped one question each; both were assigned to the binary/conventional condition, but they skipped different questions (one sex at birth and one the first-order scales).

Requests for clarification

Five respondents indicated some difficulty answering questions, particularly the scale items. Most of these concerns were related to a lack of definitions and all of the respondents were male. It is worth noting that our scale items are not unlike questions about whether a respondent identifies as "liberal" or "conservative," which also do not include explicit definitions for the categories. We reproduce this feedback in its entirety:

- the terms masculine and feminine are somewhat open to interpretation. they mean different things to different people and the definitions have changed over time and continue to evolve.
- They were open ended, they need more specifics.
- I resisted the urge to treat the paired options as dependent (e.g., top value is X and the bottom value is 1 - X). Adding vertical space between the two items may help.
- I'm not totally sure what is meant by feminine and masculine; they seem like pretty loaded terms.
- What do people use to define masculinity? Visual? Interesting study.

Explanations of responses

Nearly 200 respondents (13 percent) took the time to explain how they interpreted the questions or some aspect of their answers. These comments are not positive or negative per se, though many imply that the respondent saw something interesting or self-relevant in the questions. For

example, many explanations state that the questions gave males an opportunity to express their feminine side, or that they let females show that they are tomboys. Others suggest some defensiveness or sensitivity in their responses. As a group these comments also reveal the criteria respondents used to arrive at their answers to the masculine and feminine scales, citing appearance, hobbies, sexuality, emotionality, and occupation. Responses include:

- Im a man I have a beard!
- No, I think I'm pretty standard.
- Uh, well, I'm not like LETS GO PLAY SPORTS so I didnt give myself full manliness.
- its just how i look
- was a tomboy. Still not afraid to get under a car hood or pick up a toad. But in my head I am very woman.
- I am a pretty manly man I guess though I don't look it. That is I work with my hands, build stuff, do wood working, chop wood, and such, but I have long hair and generally try to keep my appearance nice.
- I suppose I don't really see myself as too feminine because I don't have children and don't see myself as maternal. I've been single for so long, I hardly see myself as sexual. I don't often act in traditionally feminine ways, I suppose. I hate wearing make-up and I rarely dress in skirts. I just see myself as a human who happens to be a woman. I suppose I am a little masculine, but even that doesn't feel quite right since I have no desire to be male.
- I feel like a typical guy, so I don't think I can provide much context to this survey unfortunately.
- I'm a man and not a woman 100% and heterosexual. I sometimes act silly with my friends and be like I'm feminine but that's it.
- Definitely made me think. I have a sensitive side but am also very masculine.
- Sometimes I don't feel as feminine as other women.
- I have never been confused for a woman in my entire life. This is why I put 100% masculine, and not feminine at all. I am not very emotional and have very guy hobbies, like soccer, basketball, and chess.
- I gauge the strong male like a football player, oil rig worker etc... Id love to see myself that way, dont think I pull it off.
- I gave myself 5's instead of 7's because I am morbidly obese, and as such have a more boxy figure than the typical slender hourglass that is considered a beautiful feminine shape.
- It was very interesting to answer these questions. I have had an ex imply I was very masculine before.
- I was in the marine corp, and I think I come off as tougher then I am, but I don't worry about it.
- I grew up as a tomboy, and I'm an electrical engineer, a field that is almost all male. I also can have quite an assertive personality at times, so I feel somewhat masculine, but more feminine still.
- Gender is not important to me. I like being female. I think I'd enjoy being male as well.
- I only marked feminine a two because I shave most of my body hair.
- I don't feel like I have a gender. I have never identified as female, but I've never felt a need to identify as male, either.
- I am a fairly masculine homosexual. But I have my days. ;) Thanks.
- Female but not extreme feminine, as in not prissy or girly.
- I think I need to cut my hair and stop wearing v-necks now...

GSS proposal resubmission 2

Measuring masculinity and femininity:
Updated pre-test data and recommendations for the 2016 GSS

Shelley Correll
Cecilia Ridgeway
*Aliya Saperstein**
Laurel Westbrook

This report summarizes results from a second pre-test of gender identity and diversity items proposed for the 2016 General Social Survey. Based on preliminary findings from our November survey and those from our previous pilot study in May, we can confidently make several recommendations regarding the use of feminine and masculine scales:

1) Scale items assessing femininity and masculinity can be asked in the middle of the survey, as part of the self-administered questionnaire, without fear of question order or priming effects on subsequent items.
2) Greater differentiation between first- and third-order scale responses is encouraged by presenting the first-order scale first.
3) "Very" is an appropriate top-end label for the scales.
4) If the scale items appear as part of a series of questions about sex/gender identity and diversity, as we proposed previously, then the scale items should appear first in the module.

Below, we discuss and present data related to these recommendations, and respond to Board suggestions from the Fall meeting. We also provide information negating concerns about non-response, or other response bias, resulting from the inclusion of a sex/gender module in a national survey of attitudes and opinions. We conclude with a brief assessment of the extent to which results from this second pre-test should be considered representative or comparable to the GSS sample.

Summary of the study

Data collection for our second pre-test occurred from November 4 to 24 using the Amazon Mechanical Turk platform. The survey included more than 45 questions either adapted or drawn directly from the GSS core. Our analytic sample provides data on 1,522 respondents after excluding cases from non-US IP addresses, duplicate IPs or Mturk IDs, and those who completed the survey in less than 4.5 minutes. Respondents were paid $1.50 for their time. Those who responded to our first pilot study in May were ineligible for participation.

All respondents were U.S. residents, aged 18 and older, with an Mturk HIT approval rate of at least 90 percent. They were randomly assigned to eight different study conditions. (See Table 1 for valid sample sizes by condition.) These conditions varied the placement of the sex/gender module in the overall survey (middle vs. end), where the feminine-masculine scales appeared within the module (first vs. last), and the order in which the scales were presented (first-order vs. third-order first).

*Corresponding author: asaper@stanford.edu

Table 1. Sample sizes, by study condition

	Module middle		Module end		Total
	Scales first	Scales last	Scales first	Scales last	
1st-order first	287	280	94	97	758
3rd-order first	287	284	96	97	764
Sub-totals	574	564	190	194	
Total	1138		384		1522

Note: Respondents were randomly assigned to one of the eight conditions that represented the union of these three dimensions. Sample sizes were intentionally unbalanced between module placement in the middle or at the end of the survey because placement in the middle was the preferred survey design.

Scale presentation and randomization

At its Fall meeting, the GSS Board suggested randomizing the presentation of the first and third order scales, but stated a preference for offering the first-order (see self) before the third-order (others see you) scales. The Board also suggested randomizing the presentation of the feminine and masculine scales within the first- and third-order question sets. We address each of these recommendations in turn.

First- and third-order scale randomization

In our initial pre-test, we presented the third-order scales first across all conditions. We did so based on the assumption that it might be easier for people to "contradict" external perceptions if asked about them first and their own perceptions second. In the second pre-test, we randomized the order in which the two perspectives were presented. The key outcome in testing the accuracy of our initial assumption is the difference between the first- and third-order responses on the same gendered scale (i.e., the absolute value of the difference between the first-order feminine and third-order feminine responses).

The order of scale presentation does affect response variation, but not in the direction we expected. In the full sample, although the separate scale scores do not vary significantly, on average, the difference between respondents' first- and third-order scale scores does vary (see Table 2). There is significantly less differentiation between the two scales when the third-order scale appears first. This result appears to be driven by how people respond on the sex-atypical scales: when the first-order scales appear first, females exhibit significantly greater response differentiation between the two masculine scales and males exhibit marginally greater response differentiation on the feminine scales (not shown). That is, both sexes identify as more sex atypical than how they say others see them when asked to self-identify on the scales first. Presenting the third-order scales first may make people feel more accountable to social expectations rather than less. Thus, when asking both sets of scales, we recommend that the first-order scales should appear first.

Table 2. Average scale scores, by order of scale presentation

	1st order first	3rd order first	Total
1st order feminine	3.04	2.97	3.01
3rd order feminine	3.00	2.89	2.94
Difference	.45	.38*	.41
1st order masculine	2.84	2.86	2.85
3rd order masculine	2.77	2.80	2.78
Difference	.45	.40#	.42

Note: # p<.10, *p<.05, one-tailed test

There is some evidence that the difference between first- and third-order responses to the same scales is largest when the sex/gender module appeared in the middle of the survey, and when the scales appeared first in the module – though these differences were not statistically significant. Overall, the percentage of respondents who selected different categories on the first- and third-order scale pairs was similar to, if slightly lower than, what we found in the first pre-test. In the first pre-test, 22 percent of respondents gave distinct answers to both pairs of masculine and feminine scales, while 35 percent gave distinct answers on one pair of scales or the other (but not both). The corresponding figures for the second pre-test are 19 and 29 percent, respectively.

Feminine-masculine scale randomization

Due to other study priorities, we did not randomize whether respondents saw the feminine scale before the masculine scale, or vice versa, in the second pre-test. Evidence from our first pre-test, which did randomize the order in which the scales were presented, suggested there was little if any difference in results between the two designs. Males, for example, had average first-order gender polarization scores of 3.5 when the feminine scale appeared first (n=483) and 3.6 when the masculine scale appeared first (n=461), a difference that was not statistically significant. (Note: Polarization represents the absolute value of the difference between responses on the feminine and masculine scales; the highest possible polarization score is 6.)

Thus, we conclude that it is likely not necessary to randomize the order in which the feminine and masculine scales are presented. Of course, if it were not difficult or costly to implement in the GSS, there certainly would be no harm in doing so.

Response bias and non-response

A major concern in the second pre-test was to determine whether or not seeing and responding to the sex/gender module in the middle of the survey seemed to affect responses to the remaining questions, compared to answering the module at the end of the survey. We found no evidence of statistically significant differences in responses between the two conditions.

Some response patterns could be interpreted to suggest that the sex/gender module primed respondents to give more gender stereotypical subsequent responses. For example,

women did report spending more time on household work (10.6 vs. 10 hours) when the sex/gender module was in the middle of the survey, compared to the end – but the difference was not statistically significant, and men's responses on the household work item were nearly identical in the two conditions. A higher percentage of men reported having been arrested when the module appeared in the middle of the survey compared to the end (23% vs. 20%), but men also reported being somewhat more altruistic (on GIVCHRTY, HELPAWAY, etc.) when the module appeared in the middle of the survey. Again, none of the response differences between the two module placement conditions were statistically significant.

Missing data

Non-response has also been a concern. However, we can find no evidence that respondents were troubled by the inclusion of the sex/gender module. All questions included soft prompts for answers before respondents could advance to the next page, but payment was not predicated on completing the entire survey, and respondents were informed that they had the right to skip questions or withdraw at any time on the initial study information and consent page. Very few availed themselves of the right to not respond, and no one simply clicked through the survey without answering any questions.

Four respondents appeared to break off their participation in the midst of the survey, but all stopped responding before they saw the sex/gender module. Missing data was highest on variables like SEXFREQ (n=4), but there was just one missing case across the entire sex/gender module. (One respondent skipped the third-order feminine scale, but answered all the rest.) There also was no evidence that respondents who saw the sex/gender module in the middle of the survey took the survey any less seriously, as measured by their total completion time. Average completion time for the survey was just over 10 minutes (607 seconds). If anything, respondents appeared to spend slightly more time completing the survey when the sex/gender module appeared in the middle compared to when it appeared at the end (614 seconds vs. 588 seconds) – though the difference was not statistically significant.

Scale labels: Very vs. Extremely

In our first pre-test, the poles of the feminine and masculine scales were labeled "not at all" and "extremely." In the second pre-test, we changed the top-end label to "very," out of concern that "extremely" might have a negative connotation for some respondents.

Results from the second pre-test are consistent with that expectation. Comparing the distribution of responses from both studies indicates that males were more likely to use the highest response category on the masculine scale, and females were more likely to use the highest response category on the feminine scale, when it was labeled "very." However, the same was not true for sex-atypical scale responses (i.e., females were less likely to use the high end of the masculine scale in the second pre-test). As a result, polarization scores are higher on average in the second pre-test, particularly among females (see Table 3).

This result might vary by sex because labeling oneself as "extremely" feminine has greater negative connotations than being "extremely" masculine. In both pre-tests, the modal

response for males on the masculine scale was a 5, while the modal response for females on the feminine scale increased from a 5 when the highest category was labeled "extremely" to the highest category (6) when it was labeled "very." However, it is also possible that the samples from our two pre-tests are not directly comparable. Our first pre-test sample was 38 percent female, compared to 53 percent female in the second pre-test. The first pre-test not only might have underrepresented females numerically, but those who did respond could also be less representative of females in general. (We did not collect additional demographic data in the first pre-test that would allow us to confirm that speculation). That said, assuming respondents in general, and females in particular, avoided the highest category on sex-typical scales in the first pre-test because of the "extreme" label, then "very" should be the preferred wording.

Replicating findings

The change in response category wording likely had implications for other results, as well. For example, our finding from the first pre-test that males had significantly higher first-order polarization scores than females was not replicated in the second pre-test. Nor did we find in the second pre-test that polarization scores varied significantly by whether the scales appeared first or last in the sex/gender module. However, the finding that females in the South report significantly higher first-order gender polarization scores than females elsewhere does appear to be robust across the two pre-tests (see Table 4).

How representative is the second pre-test?

A final concern regarding our results is whether Mturk workers are comparable to GSS respondents. We use demographic data from the 2012 GSS for comparison.

Our second Mturk sample compares well with the GSS distribution for respondent's sex, and the regional and political party affiliation distributions are also quite similar between the two surveys (see Table 5). The biggest differences are in racial self-identification, educational attainment and age. The Mturk sample overrepresents whites and Asians. Our pre-test sample is also significantly younger, on average, and more likely to have a four-year college degree. This pattern is consistent with previous studies.[1]

Preliminary multivariate analysis suggests that older respondents and respondents who self-identify as black give more sex-typical (and polarized) responses, on average, on the feminine and masculine scales (not shown). Given that our Mturk sample underrepresents these types of respondents, our results could be underestimating sex/gender conformity relative to a sample that is more nationally representative on these characteristics. Nevertheless, our pilot tests demonstrate that the proportion of Americans who give gender neutral or otherwise non-conforming responses on the feminine and masculine scales is an important part of the U.S. adult population. Just 24 percent of our respondents give the extreme gender-typical responses (e.g., women who say they are "very" feminine and "not at all" masculine) that are implied by the standard categorical measures of sex/gender in major social surveys.

---

[1] Berinsky, Adam, Gregory Huber and Gabriel Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3): 351a 68.

Table 3. Comparison of scale responses between pre-tests, with different response labels

Panel A: Male at birth

			Proportions						
First pre-test	*Mean*	*Standard deviation*	Not at all	1	2	3	4	5	**Extremely**
Masculine scale	4.6	1.1	1%	1%	2%	8%	28%	39%	20%
Feminine scale	1.2	1.2	34%	33%	17%	9%	5%	1%	0.2%
Polarization	3.5	1.7	7%	8%	14%	16%	25%	16%	16%
Second pre-test			Not at all	1	2	3	4	5	**Very**
Masculine scale	4.7	1.2	1%	1%	1%	11%	24%	33%	28%
Feminine scale	1.1	1.2	41%	30%	15%	9%	3%	1%	1%
Polarization	3.7	1.8	7%	6%	13%	15%	21%	14%	23%

Panel B: Female at birth

			Proportions						
First pre-test	*Mean*	*Standard deviation*	Not at all	1	2	3	4	5	**Extremely**
Feminine scale	4.4	1.3	1%	2%	5%	12%	30%	30%	21%
Masculine scale	1.5	1.4	32%	24%	20%	13%	8%	2%	1%
Polarization	3.2	1.9	10%	13%	17%	13%	18%	14%	16%
Second pre-test			Not at all	1	2	3	4	5	**Very**
Feminine scale	4.7	1.2	1%	1%	2%	11%	25%	28%	32%
Masculine scale	1.2	1.2	36%	30%	17%	12%	4%	0.6%	0.3%
Polarization	3.7	1.9	8%	8%	11%	18%	18%	12%	24%

Note: Scales were recoded from 0-6. Polarization is the absolute value of the difference between the two scales.

Table 4. First-order polarization score comparisons

	First pre-test			*Second pre-test*	
	Males	Females		Males	Females
Region			*Region*		
West	3.3	2.8	West	3.8	3.6
Midwest	3.6	3.2	Midwest	3.6	3.4
Northeast	3.6	3.0	Northeast	3.8	3.4
South	3.7	3.6	South	3.7	3.9
Total	3.5	3.2	Total	3.7	3.7
Scales order			*Scales order*		
First	3.4	3.1	First	3.7	3.6
Second	3.5	3.1	Second	-	-
Third	3.6	3.4	Third	3.7	3.7
Total	3.5	3.2	Total	3.7	3.7

Table 5. Descriptive statistics comparison, GSS and Mturk second pre-test

				Module Placement	
		GSS 2012	Mturk Sample	Middle	End
Sex:	Female	55%	53%	54%	50%
	Male	45%	47%	46%	50%
	Intersex	-	0	0	0
Gender:	Woman	-	53%	54%	49%
	Man	-	47%	46%	50%
	Transgender (direct)	-	0.1%	0	0.5%
	Transgender (indirect)	-	0.2%	0.1%	0.5%
	All other responses	-	0.2%	0.1%	0.5%
Scales (mean)					
	1st order feminine	-	3	3.1	2.8
	1st order masculine	-	2.8	2.8	3.0
	1st order polarization	-	3.7	3.7	3.8
	3rd order feminine	-	2.9	3.0	2.9
	3rd order masculine	-	2.8	2.7	2.9
	3rd order polarization	-	3.9	3.8	3.9
	Gender neutral (%)	-	7%	7%	7%
	Gender non-conforming (%)	-	4%	4%	4%
Race:	White	70%	81%	81%	82%
	Black	13%	7%	7%	6%
	Asian	3%	5%	6%	4%
	Multiracial	7%	5%	4%	6%
	All other responses	7%	2%	2%	3%
Hispanic origin		14%	7%	8%	7%
4-yr degree		28%	58%	59%	56%
Region:	West	22%	25%	25%	25%
	Midwest	23%	21%	20%	21%
	Northeast	17%	17%	18%	16%
	South	38%	38%	38%	37%
Party affiliation:	Democrat	37%	41%	40%	42%
	Republican	23%	17%	19%	13%
	Independent	39%	38%	37%	42%
	All other responses	3%	4%	4%	3%
Age (mean)		48	35	35	34
N		1974	1522	1138	384

Note: Indirect transgender categorization is based on reporting a sex at birth different from one's current gender. Gender-neutral respondents reported first-order polarization scores of zero. Gender non-conforming respondents are either categorically gender variant or have responses to the first-order scales that do not match their reported gender category (e.g., women who see themselves as more masculine than feminine). Region is based on self-reported state of residence; state inferred from IP address provides similar results.