Original Article

Sociological Methodology
2021, Vol. 51(2) 253–269

DOI: 10.1177/0081175020982632
http://sm.sagepub.com

Reconsidering the Reference Category

Sasha Shen Johfre[1] and Jeremy Freese[1]

Abstract

Social scientists often present modeling results from categorical explanatory variables, such as gender, race, and marital status, as coefficients representing contrasts to a "reference" group. Although choosing the reference category may seem arbitrary, the authors argue that it is an intrinsically meaningful act that affects the interpretability of results. Reference category selection foregrounds some contrasts over others. Also, selecting a culturally dominant group as the reference can subtly reify the notion that dominant groups are the most "normal." The authors find that three of four recently published tables in *Demography* and *American Sociological Review* that include race or gender explanatory variables use dominant groups (i.e., male or white) as the reference group. Furthermore, the tables rarely state what the reference is: only half of tables with race variables and one-fifth of tables with gender variables explicitly specify the reference category; the rest leave it up to the reader to check the methods section or simply guess. As an alternative to this apparently standard practice, the authors suggest guidelines for intentionally and responsibly choosing a reference category. The authors then discuss alternative ways to convey results from categorical explanatory variables that avoid the problems of reference categories entirely.

Keywords

quantitative, ethics, scientific communication, guidelines, regression

Social science research often relies on categorical conceptions of human difference. Many of the most commonly used independent variables in social scientific analysis, including race, gender, and marital status, are treated as categorical. Even ostensibly continuous variables such as age, income, and health status are commonly operationalized as categories.

These categorizations are generally intended to be symmetric: no one category has a special standing vis-à-vis the others. "Male" and "female," for example, are considered different sides of the same coin; one is not more intrinsically or theoretically fundamental than the other. Yet when analysts include categorizations as explanatory variables in regressions, the standard approach is to present results as contrasts to a single designated reference (or "omitted") category. Using a reference group solves the algebraic problem in which a model with an intercept and terms for every category

[1]Stanford University, Stanford, CA, USA

Corresponding Author:
Sasha Shen Johfre, Stanford University, 450 Serra Mall, Building 120, Room 160, Stanford, CA 94305, USA.
Email: sjohfre@stanford.edu

is underidentified: the right-hand-side $\mathbf{X}$ matrix is rank deficient, and so, inter alia, $\mathbf{X}'\mathbf{X}$ cannot be inverted.

Selection of the reference category does not affect formal results and so in this sense is purely arbitrary. However, formal arbitrariness does not necessarily imply cognitive neutrality. Using one category as the reference inherently introduces an asymmetry upon categories. Even when contrasts between categories other than the reference can be calculated by simple arithmetic, the required mental operations make some results more readily available than other. Moreover, the standard errors, p values, and confidence intervals for contrasts not involving the reference typically cannot be obtained from information in a conventional table.

Looking at what analysts do in practice makes plain that they do not treat the choice of reference category as a random matter. As we show shortly, for some variables, the reference category disproportionately corresponds to a socially dominant group. Using "male" and "white" as the reference category may seem conventional, but this convention may encourage the idea that dominant groups are "baseline" and marginalized groups are deviations. Language practices that position dominant groups as the default have been shown to reinforce existing hierarchies (Bem 1994; Murray 1973; Ridgeway 2011), and reducing such practices has been an active concern on other fronts of scientific communication (Chestnut and Markman 2018). In that context, an unreflective convention of assigning a dominant group as the reference category is hard to justify and worth reconsidering.

What should social scientists do? In this article, we start by articulating concerns about categorization in social scientific research generally and the reference category specifically. We use data on how recent publications in *American Sociological Review* (*ASR*) and *Demography* present reference categories for race and gender variables (spoiler: they don't do a great job). Then, we offer guidance for (1) choosing the reference category and (2) ways to display results from categorical predictor variables that avoid omitted categories entirely.

BACKGROUND

Researchers use categorical predictors in their models with good intentions, but this use has been recognized as fraught (American Sociological Association 2003; Zuberi and Bonilla-Silva 2008). In this section, we first note some epistemological reservations that have been raised about using categorical independent variables in the social sciences. We contribute to this conversation an additional challenge that is often overlooked and yet highly addressable: formulating categorical difference using a reference category.

Categorization

The axes of variation that get turned into categorical variables in everyday life and in research are often much more complex than their treatment as discrete might suggest. Boundaries may be fuzzy, definitions may be contingent and variable, and membership may be fluid. As such, using categorical measures for constructs such as race, sex,

gender, class, sexual orientation, labor market experience, education, health, and immigration status may obscure and flatten true social processes.

Existing research suggests at least three primary ways this flattening can happen. First, using finite categorical options when collecting data can nonrandomly exclude participants' experiences, such as when an intersex person tries to answer a binary sex question (Westbrook and Schilt 2014). If such people are not counted in the first place, scientists lose the ability to learn about them. Second, separating gender, race, and other variables of social difference may collapse variation within groups (e.g., variation among women) and make it more challenging to understand the interconnectedness of multiple axes of variation (Collins 2002; Roth 2016). Third, using a single variable for a system of difference such as gender or race can make multilevel processes appear to operate solely at an individual level, treating these systems as characteristics of people rather than social structures (Martin and Yeung 2003; Sprague and Zimmerman 1993; Zuberi 2000).

Yet researchers may want to use categorical variables precisely because they are understood by real-world social actors as meaningful. Sociologists and others have provided strong theoretical arguments for why inequality-generating social processes so often operate in categorical terms, including how they are pervasively institutionalized in laws, regulations, and norms (Barad 1996; Bonilla-Silva 1999; Massey 2016; Ray 2019; Ridgeway 2011; Tilly 1998). If employers in an industry discriminate against women qua women, then operationalizing gender variation through only continuous measures of masculinity and femininity would obscure that, just as measuring education only as years of schooling obscures the rewards for the categorical milestones of completing high school and college.

Conceptualizing human variation through discrete terms will likely remain a commonplace feature in quantitative social research. How can scientists analyze human variation with categorical variables in the most scientifically appropriate and responsible ways? One solution, put forth by those in the constructivist tradition, is to encourage inquiries into categorizations as outcomes rather than predictors of social processes (Morning 2011; Nagel 1995; West and Zimmerman 1987; Wimmer 2008). Another is to develop more valid and useful measures of generally taken-for-granted categorical variables, such as the General Social Survey's recent change from one to two separate gender identification questions (Magliozzi, Saperstein, and Westbrook 2016). A third solution, which we address here, is to consider systematically the ways categorical variables are and should be used in research (often as independent variables), including whether and how researchers choose and convey a reference category.

The Purpose of Reference Categories

Reference categories are useful algebraic solutions to a ubiquitous challenge of model specification. If a linear model has an intercept term and unconstrained coefficients for each category of a categorical variable, it is underidentified. Take the simplest case in which a categorical variable is the only explanatory variable. The mean of the outcome for members of category m is $\overline{y_m} = a + \beta_m$. If there are k categories, we can write k such

Table 1. The Reference Category in Recent Issues of *ASR* and *Demography*

	Percentage with Dominant Group as Reference		Percentage Not Listing the Reference Category	
	Race Variables	Gender Variables	Race Variables	Gender Variables
ASR (2014–2019)	89	82	75	87
Demography (2017–2019)	95	68	28	67
Combined	92	76	50	79

Note: The unit of analysis is the table. We examined every publication in *American Sociological Review* (*ASR*) from January 2014 to June 2019 and in *Demography* from January 2017 to June 2019. *ASR* had 65 tables with a race variable and 141 tables with a gender variable. *Demography* had 67 tables with a race variable and 89 tables with a gender variable.

equations with k β coefficients. However, because of a, there are $k + 1$ unknown parameters for our k means, so there is no unique solution for our parameters without an additional constraint. The reference category approach is to set β for one category to 0. If *ref* is the reference category, then

$$\overline{y_m} - \overline{y_{ref}} = (\alpha + \beta_m) - (\alpha + 0) = \beta_m.$$

In the linear regression model, the coefficient β_m is thus the difference in the expected value of the outcome between members of category m and members of category *ref*.

Which category is used as the reference does not matter for the substance of the resulting estimates: contrasts between any pair of groups can be calculated, with the contrast between categories m and n being $\beta_m - \beta_n$. Even so, as we will discuss, the choices researchers make in the way they code and convey information about categorical variables may matter for how results are understood and used.

Current Standard Practice

To better understand how researchers currently use and convey categorical variables in research, we examined every table in the past several years of two highly regarded social science journals: *ASR* (January 2014 to June 2019) and *Demography* (January 2017 to June 2019).[1] We first identified every published table that included a variable for either race (U.S.-focused) or gender as a predictor in a quantitative model; in most cases, this meant results from a regression model.[2] This resulted in a sample of 65 tables with a race variable and 141 with a gender variable from *ASR* and 67 tables with a race variable and 89 with a gender variable from *Demography*. Finally, we coded each table on the basis of what group was used as the reference category and how that information was displayed.

As shown in Table 1, among the 132 tables across both publications that included U.S.-based race or ethnicity variables, more than 92 percent used a reference group that included "white." Among the 230 tables that included gender or sex variables, 76 percent used "male" or "man" as the reference category. Taken together, 82 percent of

tables with race or gender variables across the two journals used the culturally dominant group as the reference category.

Furthermore, only half of tables in our analysis (49 percent) listed what the reference category was for race variables, and fewer than a quarter of tables (21 percent) did so for gender variables (either in the column of variable names or in table notes).[3] In the tables we examined, estimates for "black" and "Latino" were sometimes presented without it being readily discernible whether they were coded as exclusive categories.[4] The apparently standard practice of not fully describing unordered variable categories in tables seems to pointlessly obscure methods in ways that can greatly hinder accurate interpretation of results.

Some of these tables might code dominant groups as the reference because the authors are theoretically interested in disadvantaged groups and therefore convey results compared with dominant groups. However, if this were the case, wouldn't researchers want to be clear about the contrast, rather than leave it to readers to puzzle out what the reference is on the basis of their own assumptions? Indeed, when we were coding these tables, we had to make exactly such assumptions; for example, if a table simply had a row labeled "female," we assumed that "male" was the reference. We suspect that this pattern partly reflects researchers following practices they are accustomed to seeing without much further reflection. Following standard practices is not itself bad; however, as we describe next, we believe that in this case it can introduce unnecessary confusion and even harm.

The Problem of Using Dominant Groups as the Reference

Even if simply done out of a sense of convention, using social dominance as grounds for reference category selection is a bad principle for at least two reasons. First, using dominance to identify reference categories reinforces the common practice of linguistically treating dominant groups as "unmarked": a category that may seem like it does not have any characteristic, such as a white race or a male sex (Brekhus 1998; Frankenberg 2001). Prior research shows that this type of normalization process contributes to the construction of social difference and inequality (Bem 1994; England 2005). Importantly (for well-meaning researchers), this can even occur when the content is ultimately intended to reduce inequality. For example, reading the sentence "girls are as good as boys at math" has been found to actually *increase* gender stereotyping because it normalizes boys as those with math ability (Chestnut and Markman 2018). By identifying dominant groups as the reference category, scholars may be sustaining notions of dominance as baseline in ways that reify existing power relations. This consequence may be directly contrary to researchers' goals, if they are interested in examining and ultimately reducing inequality.

Second, even though a reference category might seem neutral from a mathematical standpoint, it often implies asymmetries in the relative ease of interpreting results for different groups. There is an irony here in that sociological theorizing often focuses on the plight of disadvantaged groups, yet using the advantaged group as the reference category means results for disadvantaged groups are often less available than those of

the advantaged group. For example, for polytomous categorizations, standard errors are usually presented only for the contrast of each category to the reference category, which means uncertainty estimates cannot be recovered for contrasts between other groups. Furthermore, when the intercept of the model has a meaningful interpretation, it is an interpretation in terms of the absolute level of the outcome for the reference group.

Interpretation becomes even more asymmetric and cumbersome once interaction terms are involved. Consider a model in which a two-category gender variable is interacted with a multiple-category race/ethnicity variable, in which "male" and "white" are used as the reference categories. If the other variables are centered, then the intercept provides the expected value of the outcome for white men. To obtain the expected value for white women or nonwhite men, one arithmetic operation is needed, whereas for nonwhite women, two arithmetic operations are needed. For any contrasts involving nonwhite women, an additional arithmetic operation is required, and neither standard errors nor p values can be simply calculated. For social scientists whose work focuses on disadvantage, it seems suboptimal that in the prevailing practice, results are systematically hardest to extract for the most disadvantaged groups.

How a researcher chooses and conveys results from categorical predictor variables therefore has subtle but important effects on both scientific readability and broader social norms. As we have described, the implications may be particularly problematic when a dominant group is used as the reference.[5] Rather than determining the reference category by a sense of convention or software package defaults, we believe that a more principled approach is warranted.

CHOOSING THE REFERENCE CATEGORY

Toward developing best practices for selecting a reference category, we offer a set of ordered principles that together form a decision tree: we recommend simply using the first principle that applies. Figure 1 conveys a summary of our decision tree, and Table 2 provides a summary of recommendations based on these guidelines for commonly used categorical explanatory variables.

Using these guidelines may involve recalculating estimates after deciding on the reference category to be used in a presentation or report. Just as researchers take the time to polish the font, borders, and table notes before presenting results to an audience, they should also consider their selection of reference categories and adjust as necessary for maximal clarity. Regardless of which reference group is selected, it is also crucial to make clear what the omitted category is in both tables and text.

1. Is There a Theoretically Fundamental Reference Group?

Sometimes researchers have a theoretically justified reason to consider one group as the "baseline," warranting use of this group as the reference category. This is clearest in truly asymmetric cases in which one category is best understood as the default condition of a categorical variable, and all other categories represent departures from the default. The prototype is the control group in an experiment: control groups are given

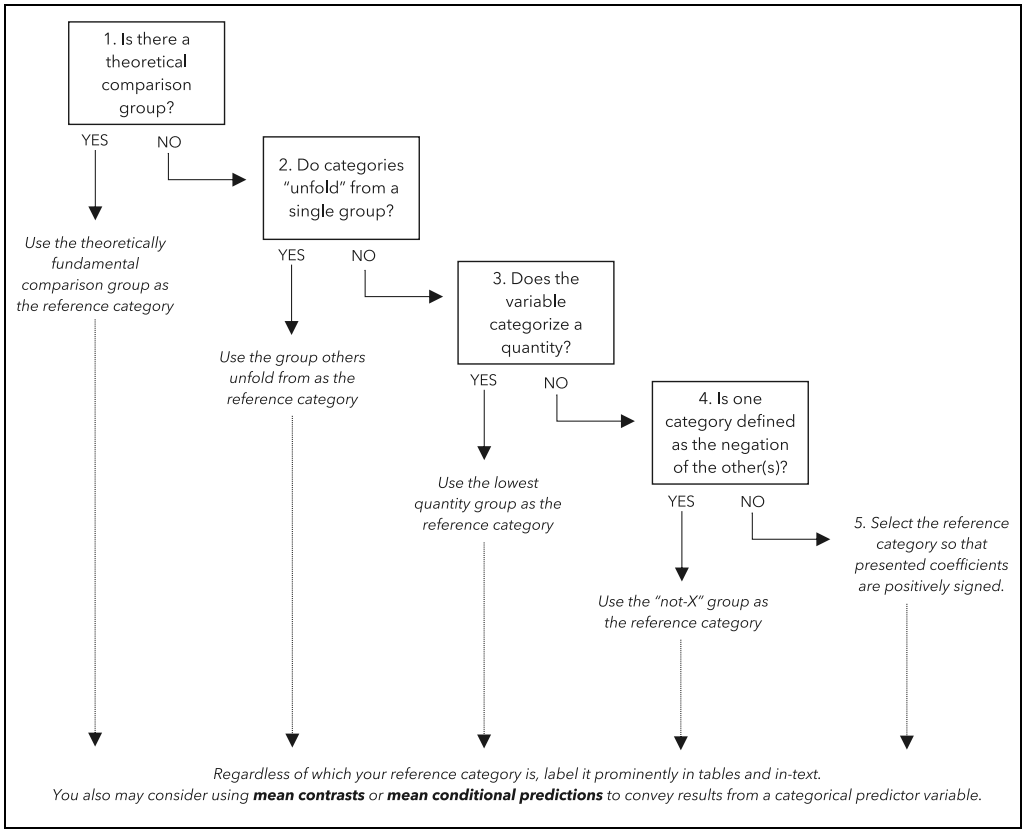

Figure 1. Five guidelines for choosing a reference category.

the most neutral treatment, or not intervened with at all, and their purpose is to provide a standard against which treatments are compared. For example, in their recent *ASR* publication, Flores and Schachter (2018) presented results of a conjoint experiment of perceptions of likelihood of "illegality" among Hispanic Americans. A control group of "no record" was contrasted with several treatment conditions involving different types of criminal records.

This logic also extends to quasi-experimental and observational studies that are structured theoretically around an intervention. As an example, de Vaan and Stuart (2019) asked whether the introduction of opioids into a household increases the likelihood of a subsequent opioid prescription for another household member. Having opioids introduced into the household is not an experimental intervention, but is theoretically akin to being the "treatment" group, and so not having opioids introduced into the household is the appropriate reference category.

There may also be times when a category is not an absolute default, but given a well-described framework, it is theoretically clearest to consider it as such. For example, if a researcher is interested in the experience of immigrants within a predominantly native-born population, it may make sense to code a binary nativity variable with "native born" as the reference and "foreign born" as 1.

Table 2. Examples of Common Categorical Variables and the Best Choice for a Reference Category

Variable	If Categorized as . . .	Recommended Reference Group	Guideline Number Used for Recommendation
Sex	Male Female	The one with a lower predicted value, so coefficient in table is positive (varies by table)	5. No clear reference
Race	White Black Asian Latinx Other	The one with a lower predicted value, so coefficient in table is positive (varies by table), except "other" presented last	5. No clear reference
Marital status	Married Not married	Not married	4. Negation
Marital status	Married Never married Divorced Widowed	Married	2. "Unfolding" categories
Degree earned	Less than high school High school College Postgraduate	Less than high school	3. Quantity
Employment	Employed Not employed	Not employed	4. Negation
Receiving an intervention	Received intervention Did not receive intervention	Did not receive intervention	1. Fundamental reference
Class	Lower class Working class Middle class Upper class	Lower class	3. Quantity
Age	18–40 41–65 >65	18–40	3. Quantity

Still, when this guideline may suggest coding a dominant group as the reference, we urge researchers to be particularly conscientious about their motivations in order to maximize clarity of interpretation. It may be helpful to consider whether what seems like an asymmetry between groups arising from theory actually is. For example, many accounts of wage gaps among different categories (e.g., gender, race, parental status) involve a combination of different sorts of advantages for the higher earning group and disadvantages for lower earning groups. As such, although it may seem plausible to

assert that men are the theoretically fundamental group for a study of the gender gap in wages and therefore should be used as the reference category, we believe that it would be more responsible to avoid the implications of this decision by either choosing a reference so that coefficients are positively signed (see point 5 below) or presenting results without a reference category entirely (see "Alternatives to Using a Reference Category").

2. Do Categories "Unfold" from a Single Group?

Some polytomous variables involve multiple categories that are all a change in state from the same category. Marital status, for example, is often measured as "married," "divorced," "widowed," and "never married." The latter three categories can all be defined as transitions from or to married, and not to or from one another. Consequently, comparisons with "married" are typically of more direct interest, so this should be the reference category. By coding "married" observations as 0, all other categories can be most directly interpreted as the difference between those in that group versus those who are married.

3. Does the Variable Categorize a Quantity?

Some variables categorize quantities, such as chronological age groups. Others are less explicitly numerical but still plainly distinguish less from more, as in any categorization of "low," "medium," and "high." For continuous explanatory variables, positive coefficients indicate that higher values of the variable are associated with higher values of the outcome. We suggest analysts handle ordered categorical variables in a similar way: treating the lowest category as the reference category. For example, a discrete measure of self-rated health status would best be coded with "poor" as the reference group, and "fair," "good," and "excellent" as nonreference categories.

4. Is One Category Defined as the Negation of the Other(s)?

Sometimes one category is fundamentally defined through negation. For example, some studies use a binary indicator of whether a respondent is "black." The other category here has no terminological alternative than some variant of "nonblack." Similarly, some studies use a dichotomous marital status variable with the categories of "married" and "not married" (unlike the polytomous operationalization of marital status we described earlier).

Following again the principle of quantitative variables that positive values mean "more," it is likewise more straightforward to associate positive values with positively defined categories. Hence, the negation should be used as the reference category. Note that this implies not only using "nonblack" as the reference category when research dichotomizes race by whether a respondent is reported as black but also using "nonwhite" as the reference when the options are white versus nonwhite. Similarly, in a dichotomous marital status variable, "not married" would be best understood as the reference category.

The exception to this principle is polytomous variables with a "none of the above" category that is peripheral to the meaning of the paper. Despite this category being a negation, it would be a bad choice for the reference category because it would foreground contrasts that are not pertinent to the research questions. For example, race variables sometimes include an "other" category (e.g., Mollborn, Lawrence, and Root 2018); in such cases, "other" should not be the reference group, and it should be listed last among race groups in tables.

5. Otherwise, Select the Reference Category so That Presented Coefficients Are Positively Signed

When the categorical variable is truly symmetric and the meanings of categories do not provide a rationale for choosing a reference, we recommend relying on the values of the coefficients themselves to make a decision. Given that positive numbers are cognitively simpler than negative values, the reference category can be chosen such that the presented coefficients are positive. For example, for a binary sex variable, whether "male" or "female" is communicated as the reference category should depend on what provides a positively signed coefficient. For binary variables with reasonably simple labels, both category labels can be readily presented as the label for the row: "male (vs. female)" or, in additive models, "male – female." For polytomous variables, we suggest picking the reference such that coefficients for the presented categories are positive.

We recognize the value of consistency for clarity; if one category is not consistently lowest across multiple models or tables, we suggest researchers select the category with respect to (1) the first presented model, (2) the most important model, or (3) what results in the largest proportion of positively signed coefficients across all tables.

What Order Should the Rest of the Categories Be In?

For polytomous variables, an analyst then must decide the most effective order for the remaining categories. This decision is simple for ordered variables, as their order should always be preserved when presenting results. For a set of chronological age categories, for example, categories representing consecutive age groups should be consecutive in the table. However, if the category order is otherwise arbitrary, categories should be ordered either alphabetically by their labels or ranked numerically by their results. For example, for a categorical variable representing respondents' region of residence, we suggest ordering categories alphabetically or so that the region with the largest coefficient is first and the region with the smallest is last.

ALTERNATIVES TO USING A REFERENCE CATEGORY

Just because software usually presents regression estimates using a reference category does not mean researchers have to present their results this way. We present two alternatives that avoid the reference category problem entirely. The first, *mean contrasts*, is especially useful for variables with several categories and little substantive rationale

for treating one as the reference. The second, *conditional predictions*, is useful for a model's key explanatory variable when one wants to clarify what results imply in terms of the expected value(s) of the outcome. As an online supplement to this article, we provide a .do file that produces all output described here automatically in Stata, as well as a Stata package for calculating binary contrasts.

Mean and Binary Contrasts

For variables with more than two categories, the reference category approach means that significance tests of contrasts that involve the omitted category are emphasized above contrasts that do not. For example, Reher et al. (2017) included father's occupation as a number of categories: "unskilled workers," "semiskilled workers," "skilled workers," "middle class: farmers," "middle class," "elite," and "no information." The authors used "unskilled" as the reference category, so the hazard ratios and significance tests compare other occupational categories with "unskilled." Yet substantively, comparisons with any single category may not be the most immediately informative; often it is only for the categories with extreme coefficients that one can answer the most basic question of whether category membership is positively or negatively associated with the outcome. (This otherwise requires information on the weighted proportion of observations in each category, and even then requires multiple arithmetic steps to determine.)

We suggest an alternative: present contrasts for each category relative to the overall mean, so that the sign of coefficients indicates an increase or decrease in the outcome compared with the average. The magnitude of differences between categories remains the same, but all categories are included, with the weighted mean of coefficients for all categories equal to 0.

To show how to compute the mean contrast, we consider a categorical variable x, in which p_i indicates the proportion of the sample in category i, β_i indicates the regression coefficient for category x in a model fit using a reference category (so $\beta_i = 0$ if i is the reference), and $\widehat{\boldsymbol{\beta}_x}$ is a column vector of all these estimated regression coefficients for x. We construct a $k \times k$ symmetrical contrast matrix $\mathbf{R}$, where k is the number of categories of x, in which

$$\mathbf{R}[m,n] = -p_n \text{ if } m \neq n$$
$$\mathbf{R}[m,n] = 1 - p_m \text{ if } m = n.$$

The vector of mean contrasts for categorical variable k, β_x^{MC}, is then the product $\widehat{\boldsymbol{\beta}_x}\mathbf{R}$. The mean contrast is sometimes called "weighted effect coding" (te Grotenhuis et al. 2017; Sweeney and Ulveling 2012); it can be obtained postestimation in Stata using the *contrast* command or in R using the add-on *wec* package (Nieuwenhuis, te Grotenhuis, and Pelzer 2017).

A disadvantage of mean contrasts is that coefficients for more frequent categories will tend toward zero as a result of having greater influence on the overall mean. An alternative, which we call the *binary contrast*, is to present the contrast of each category versus the weighted mean of the other categories. That is, coefficients for a

Table 3. Coefficients for Regression on Subjective Self Social Rank (1–10), 2012 to 2016 General Social Survey

	(A) Lowest Category as Reference Category	(B) Mean Contrast	(C) Binary Contrast
Respondent race			
Asian/Pacific Islander	.297	–.078	–.081
Black	.380**	.005	.006
Latinx	0	–.375**	–.445**
Native American	.138	–.236	–.239
White	.474**	.099**	.280**

Source: Data from 2012 to 2016 General Social Survey.
Note: Outcome is perceived social rank on a "ladder" scale (10 = highest). Race variable combines responses to separate questions about Hispanic/Latino ethnicity and mutually exclusive racial identification; we coded any participants who indicated that they were "Hispanic or Latino" as Latinx and not as any other race.
**$p < .001$.

polytomous race measure can be presented such that, for example, the coefficient for Asian Americans indicates the contrast with those who are not Asian American, as opposed to the contrast with a single other group or to the overall mean. Binary contrasts may be especially advantageous for unordered polytomous variables for which categories vary substantially in their relative frequency.

We show how to compute the binary contrast following the same notation as above for mean contrasts. For the binary contrast, the contrast matrix $\mathbf{R}$ remains a $k \times k$ symmetric matrix, but with

$$\mathbf{R}[m,n] = -\left(\frac{p_n}{1-p_m}\right) \text{if } m \neq n$$
$$\mathbf{R}[m,n] = 1 \text{ if } m = n.$$

The vector of binary contrasts for variable x, β_x^{BC} is then the product $\widehat{\boldsymbol{\beta}_x}\mathbf{R}$. We have written a Stata package to make it simple to compute binary contrasts (Freese and Johfre 2020).

Table 3 illustrates these alternative contrasts for a model in which the outcome variable is respondent ratings of their perceived social position on a 1 to 10 scale. Column A uses the standard reference category approach, with Latinx as the reference category so that all presented results in the table are positively signed. Column B uses mean contrasts, so that each category is contrasted with the (weighted) overall mean in the sample. The significant coefficient in column A for black respondents indicates that the difference between black and Latinx is statistically significant, and the nonsignificant coefficient for black in column B indicates that the mean for black respondents is not significantly different from the overall mean. The coefficients in column B all differ from those in column A by a constant (because Latinx was the reference category for column A, that constant is the difference from the mean for Latinx respondents presented in column B).

Column C in Table 3 presents results for the binary contrast. For white respondents the difference between columns B and C is especially large. Because white is the largest group in the sample and has the highest perceived rank, the overall mean is strongly influenced by the mean for white respondents. Thus, whereas the results in column B compare white respondents with the overall mean, column C shows the difference between white and nonwhite respondents and is therefore bigger. Binary contrasts will always be larger in magnitude than the corresponding mean contrast, specifically by a factor of $1/(1-p_m)$ for category m. Because p_m varies across categories, the absolute difference between column C and either column A or column B is not simply a constant. Consequently, if results are presented as column C, the contrasts of any two groups cannot be recovered by simply adding or subtracting.

Mean Conditional Predictions

Analyses using logit and other nonlinear models for categorical outcomes often illustrate results via changes in the predicted probability of the outcome. Such predicted probabilities are conditional on some value(s) of the explanatory variable(s). One common approach in presenting results is to use the predicted outcome conditional on explanatory variables being held to their mean (using the proportion for each nonomitted category of a categorical explanatory variable). Another is to generate predictions for each observation in the sample using its values for the explanatory variables, and then reporting the mean of these predictions. Adapting terminology sometimes used for analogous discussions of marginal effects (Long and Freese 2014), we can refer to these quantities as the *conditional prediction at the mean* and the *mean conditional prediction*, respectively.

Either way, the same logic of generating conditional predictions can be used to present results in tables for categorical independent variables in terms of predicted outcomes. In the familiar linear model, the difference between the conditional predictions for the reference category and any other category corresponds to each category's coefficients. That is,

$$\hat{y}(k=m,\boldsymbol{x}) - \hat{y}(k=n,\boldsymbol{x}) = \hat{\beta}_m - \hat{\beta}_n,$$

where $\hat{\beta}_m$ and $\hat{\beta}_n$ are the coefficient estimates for categories m and n of categorical variable k, with $\hat{\beta}_{ref} = 0$ for the reference category and $\boldsymbol{x}$ being any vector of values for the other independent variables. Significance tests can be reported for the difference in conditional predictions for any pair of categories or between a category and the mean. In Stata, the *margins* command provides a versatile means for working with conditional predictions; key elements of this functionality have been implemented in R as the package *margins* (Leeper, Arnold, and Arel-Bundock 2018).

More elaborately, differences can also be explicitly reported along with the conditional means. Table 4 illustrates this approach by juxtaposing the reference category approach for a binary independent variable (top panel) with the same results presented as conditional means of each group and their difference (bottom panel). That is, in the bottom panel, results for the dichotomous marital status variable are presented as the

Table 4. Comparison of Results Presented Using Reference Category (Top) versus Conditional Means (Bottom)

	Model 1	Model 2
Results presented using reference category		
Married (vs. not married)	.313**	.033
Household income		.012**
Results presented using conditional means		
Conditional mean if married	6.404	6.270
Conditional mean if not married	6.091	6.237
Difference	.313**	.033
Household income		.012**

Source: Data from 2012 to 2016 General Social Survey.
Note: Outcome is perceived social rank on a "ladder" scale (10 = highest).
**$p < .001$.

conditional means for respondents who are married and those who are not married, as well as the difference. We use an example in which an unconditional difference by marital status (model 1) is almost entirely accounted for when income is included in the model (model 2), so the example illustrates changes in conditional means and difference across models.

If predicted outcomes involve a nonlinear transformation of the linear predictor, as with predicted probabilities from a logit or probit model, then differences in conditional predictions no longer mirror the coefficients of the model. This has led to presentation of conditional predictions being relatively common for results from these models, with the circumvention of a reference category as a salutary side effect.

Many software packages make it simple to generate conditional predictions after fitting a model, but it is worth noting that the relationship between conditional prediction and the constant term in a regression model depends on how categorical variables are operationalized. Generally, the constant term can be interpreted as the predicted outcome when all explanatory variables are 0. When an omitted category is used, this would be the predicted outcome for members of the reference group when all other explanatory variables are 0, or members of the conjunction of all reference groups if there are multiple categorical variables in the model. The usefulness of the intercept for interpretation depends on the substantive interest in the case in which other explanatory variables are 0. Centering the other explanatory variables makes the intercept the predicted outcome when explanatory variables are held to their respective means.

We introduced the reference category by noting how a model with a free constant term and a coefficient for each category is underidentified. The reference category approach fixes the parameter of one category to 0 by omitting it from the model. The conditional prediction at the mean solved this another way: omit the constant (thus constraining it to 0) such that each category's coefficient is the predicted outcome when other explanatory variables are 0. If all explanatory variables are centered, the coefficients for each category will then be the conditional prediction at the mean for that category.

CONCLUSIONS

Assigning a reference category for categorical independent variables in regression models might seem both mathematically necessary and ultimately arbitrary. As we discuss, however, this formulation is in truth unnecessary, and the apparent formal arbitrariness belies a convention of designating dominant groups as the reference category. This practice is concerning because it tacitly reinforces the status quo and makes interpretation of magnitude and significance of results more difficult for nonrandomly distributed portions of the population. If researchers choose to present results using a reference category, they should do so intentionally; we therefore offer guidelines for selecting a reference category in a principled way. We also describe ways to present modeling results that avoid using a reference category in the first place. Regardless of how results are conveyed, it is good practice to make tables self-contained: if using a reference category, this means explicitly stating what the reference is. Being more intentional about how to present estimates for categorical regressors is an important and relatively easy way to be more scientifically responsible.

Our discussion and examples focused on person-level variables. Here the issues about what conventions might imply or reify about social hierarchies are most obvious and acute. However, the basic principles can be usefully applied regardless of the level of analysis of one's study, particularly given the potential increase in clarity. Ultimately, we seek to encourage social scientists to be more intentional and reflective in their consideration of categorical variables. Relatively small, but principled, shifts in practice can promote clearer and more effective communication between social scientists and their audiences.

ORCID iD

Sasha Shen Johfre https://orcid.org/0000-0001-5036-7204

Notes

1. We use different time windows for the two journals to get a better balance in terms of numbers of articles across the two journals. Across these time windows, *Demography* published 242 articles in 15 issues, and *ASR* published 278 articles in 33 issues.
2. We focus on tables because we believe that they should convey all information necessary to understand them. We therefore do not examine the in-text description of tables but simply the tables themselves and any associated notes. We focus on race and gender because they are two particularly common and important variables in research and U.S. society and therefore are useful cases for examining how categorical variables about human populations are currently presented.
3. More tables specified the reference category in *Demography* (57 percent) than in *ASR* (18 percent).
4. If a regression table had a coefficient for "black" and "Latino," or simply for "black," we infer that the reference included white and should be treated as an example of the dominant group used as the omitted category.
5. Many of these critiques also apply to the (related) logic of assigning the largest group to the reference. For polytomous variables, the standard errors for the contrasts presented in the table would be smaller than for the contrasts that are not presented. Also, this approach sustains the conflation of what is common with what is normative, suggesting minorities in a population are deviations from baseline.

References

American Sociological Association. 2003. *The Importance of Collecting Data and Doing Social Scientific Research on Race*. Washington, DC: American Sociological Association.

Barad, Karen. 1996. "Meeting the Universe Halfway: Realism and Social Constructivism without Contradiction." Pp. 161–94 in *Feminism, Science, and the Philosophy of Science*. New York: Springer.

Bem, Sandra Lipsitz. 1994. *The Lenses of Gender: Transforming the Debate on Sexual Inequality*. New Haven, CT: Yale University Press.

Bonilla-Silva, Eduardo. 1999. "The Essential Social Fact of Race." *American Sociological Review* 64(6):899–906.

Brekhus, Wayne. 1998. "A Sociology of the Unmarked: Redirecting Our Focus." *Sociological Theory* 16(1):34–51.

Chestnut, Eleanor K., and Ellen M. Markman. 2018. "'Girls Are as Good as Boys at Math' Implies That Boys Are Probably Better: A Study of Expressions of Gender Equality." *Cognitive Science* 42(7):2229–49.

Collins, Patricia Hill. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

de Vaan, Mathijs, and Toby Stuart. 2019. "Does Intra-household Contagion Cause an Increase in Prescription Opioid Use?" *American Sociological Review* 84(4):577–608.

England, Paula. 2005. "Separative and Soluble Selves: Dichotomous Thinking in Economics." Pp. 32–56 in *Feminism Confronts Homo Economics*, edited by M. A. Fineman and T. Dougherty. Ithaca, NY: Cornell University Press.

Flores, Rene D., and Ariela Schachter. 2018. "Who Are the 'Illegals'? The Social Construction of Illegality in the United States." *American Sociological Review* 83(5):839–68.

Frankenberg, Ruth. 2001. "The Mirage of an Unmarked Whiteness." Pp. 72–95 in *The Making and Unmaking of Whiteness*, edited by R. Rasmussen, E. Klinenberg, I. J. Nexica, and M. Wray. Durham, NC: Duke University Press.

Freese, Jeremy, and Sasha Johfre. 2020. "Binary Contrasts for Unordered Polytomous Regressors." *SocArXiv*. Retrieved December 15, 2020. http://osf.io/preprints/socarxiv/uk2jx.

Leeper, Thomas, Jeffrey Arnold, and Vincent Arel-Bundock. 2018. "Margins: Marginal Effects for Model Objects." R Package Version 0.3.23. Retrieved December 15, 2020. https://rdrr.io/cran/margins/.

Long, J. Scott, and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press.

Magliozzi, Devon, Aliya Saperstein, and Laurel Westbrook. 2016. "Scaling Up: Representing Gender Diversity in Survey Research." *Socius* 2. Retrieved December 15, 2020. https://journals.sagepub.com/doi/full/10.1177/2378023116664352.

Martin, John Levi, and King-to Yeung. 2003. "The Use of the Conceptual Category of Race in American Sociology, 1937–99." *Sociological Forum* 18(4):521–43.

Massey, Douglas S. 2016. *Categorically Unequal*. New York: Russell Sage.

Mollborn, Stefanie, Elizabeth Lawrence, and Elisabeth Dowling Root. 2018. "Residential Mobility across Early Childhood and Children's Kindergarten Readiness." *Demography* 55(2):485–510.

Morning, Ann. 2011. *The Nature of Race: How Scientists Think and Teach about Human Difference*. Berkeley: University of California Press.

Murray, Albert. 1973. "White Norms, Black Deviation." Pp. 96–113 in *The Death of White Sociology*, edited by J. Ladner. Arbutus, MD: Black Classic Press.

Nagel, Joane. 1995. "American Indian Ethnic Renewal: Politics and the Resurgence of Identity." *American Sociological Review* 60(6):947–65.

Nieuwenhuis, Rense, Manfred te Grotenhuis, and Ben Pelzer. 2017. "Weighted Effect Coding for Observational Data with Wec." *R Journal* 9(1):477–85.

Ray, Victor. 2019. "A Theory of Racialized Organizations." *American Sociological Review* 84(1):26–53.

Reher, David Sven, Glenn Sandström, Alberto Sanz-Gimeno, and Frans W. A. van Poppel. 2017. "Agency in Fertility Decisions in Western Europe during the Demographic Transition: A Comparative Perspective." *Demography* 54(1):3–22.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. Oxford, UK: Oxford University Press.

Roth, Wendy D. 2016. "The Multiple Dimensions of Race." *Ethnic and Racial Studies* 39(8):1310–38.

Sprague, Joey, and Mary K. Zimmerman. 1993. "Overcoming Dualisms: A Feminist Agenda for Sociological Methodology." Pp. 255–80 in *Theory on Gender, Feminism on Theory*. New York: Aldine de Gruyter.

Sweeney, Robert E., and Edwin F. Ulveling. 2012. "A Transformation for Simplifying the Interpretation of Coefficients of Binary Variables in Regression Analysis." *American Statistician* 26(5):30–32.

te Grotenhuis, Manfred, Ben Pelzer, Rob Eisinga, Rense Nieuwenhuis, Alexander Schmidt-Catran, and Ruben Konig. 2017. "When Size Matters: Advantages of Weighted Effect Coding in Observational Studies." *International Journal of Public Health* 62(1):163–67.

Tilly, Charles. 1998. *Durable Inequality*. Berkeley: University of California Press.

West, Candace, and Don H. Zimmerman. 1987. "Doing Gender." *Gender and Society* 1(2):125–51.

Westbrook, Laurel, and Kristen Schilt. 2014. "Doing Gender, Determining Gender: Transgender People, Gender Panics, and the Maintenance of the Sex/Gender/Sexuality System." *Gender and Society* 28(1):32–57.

Wimmer, Andreas. 2008. "The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory." *American Journal of Sociology* 113(4):970–1022.

Zuberi, Tukufu. 2000. "Deracializing Social Statistics: Problems in the Quantification of Race." *Annals of the American Academy of Political and Social Science* 568:171–85.

Zuberi, Tukufu, and Eduardo Bonilla-Silva. 2008. *White Logic, White Methods: Racism and Methodology*. Lanham, MD: Rowman & Littlefield.

Author Biographies

Sasha Shen Johfre is a PhD candidate in sociology at Stanford University and a fellow at the Stanford Center on Longevity. Her current focus of research is on the social construction of age. She is broadly interested in processes of social order and social inequality, including the production of systems of social difference, quantitative and computational methods, and cultural sociology.

Jeremy Freese is a professor of sociology at Stanford University. His research includes work on health inequalities, social science genomics, and survey methodology. He is a coauthor of books on regression models for categorical outcome variables and reproducible research practices in social science. He is co–principal investigator of the General Social Survey and of Time-Sharing Experiments in the Social Sciences.