Culture Remains Elusive: On the Identification of Cultural Effects with Instrumental Variables

Winston Chou[a]

Abstract

I evaluate a method recently proposed in the *American Sociological Review* for identifying the causal effects of cultural values on individual behavior. I derive the assumptions of the method and explain why they are unlikely to hold, even when ideal data are available. When ideal data are unavailable, I show that the proposed alternative is not only unnecessary, but diminishes the transparency and validity of empirical research. I conclude with some suggestions regarding the conceptualization and measurement of cultural effects in future research.

Understanding the influence of cultural values on behavior remains a venerated if incomplete task in sociology. Unfortunately, progress in this domain has been marked by a wide gap between theory and practice. Sociologists benefit from a wealth of theoretical models of culture, many of which imply causal claims. For example, if a person were more benevolent, then she would donate more money (Miles 2015). If she were endowed with different cultural skills, then she would take greater advantage of structural opportunities (Swidler 1986). Although claims like these are essential to cultural theorizing, researchers typically avoid drawing equally strong inferences from observational data. Consequently, theoretical elaboration has far outpaced empirical certitude; whereas many theories of culture posit causal relationships, researchers are usually content to demonstrate the predictiveness of cultural variables, not estimate causal effects directly.

In a 2015 *ASR* article, Polavieja proposed a new method for estimating the causal effects of cultural values. The intuition behind Polavieja's method is simple. Imagine a researcher observes an immigrant who happens to be employed as well as culturally traditional. One challenge to inferring that the immigrant's traditionalism caused her employment is that the two variables may be jointly affected by a third variable, such as the availability of work (Gerson 1985). However, suppose the researcher observes a seemingly identical woman from the immigrant's country of origin who did not migrate. Logically, this woman's traditionalism cannot be affected by the availability of work in the receiving country.

[a]Princeton University

Corresponding Author:
Winston Chou, Department of Politics, Princeton University, 025 Corwin Hall, Princeton, NJ 08544
E-mail: wchou@princeton.edu

To the extent that the two women are genuinely identical, the non-immigrant's traditionalism can be viewed as an unconfounded version of the immigrant's traditionalism. Thus, it can be used to estimate the causal effect of traditionalism on labor force outcomes, free of bias from confounders such as the availability of work.

Polavieja mounts an interesting defense of this method, but does not adequately elucidate its assumptions. I evaluate Polavieja's method in a sympathetic but ultimately critical light. After deriving the "first-best" identification assumptions of the method (i.e., the necessary assumptions given ideal data), I explain why these are unlikely to hold. I then show that the "second-best" alternative proposed by Polavieja is not only unnecessary, but diminishes the transparency and validity of empirical research. Finally, I offer some suggestions regarding the conceptualization and measurement of cultural effects in future research.

CULTURE AND CAUSAL INFERENCE WITH INSTRUMENTAL VARIABLES

Theories of culture routinely specify causal claims. For example, in criticizing the "culture of poverty" thesis prominent in the 1970s, Swidler (1986:281) wrote,

> People do not readily take advantage of new structural opportunities which would require them to abandon established ways of life . . . not because they cling to cultural values, but because they are reluctant to abandon familiar strategies of action.

Two causal claims are embedded in this hypothesis. A person would not pursue new opportunities if only her *values* were changed. (Indeed, Swidler observed, many people seemingly mired in the culture of poverty appear to have middle-class values and aspirations.) However, she would pursue these opportunities if endowed with the necessary *tools*—for instance, knowing "how to dress,

talk in the appropriate style, or take a multiple-choice examination" (Swidler 1986:275).

"Tool kit" theory inspired an efflorescence of research, much of it examining the relative importance of these two forms of culture; however, scholars in this tradition typically avoid claiming to have measured their effects directly. For example, Vaisey (2009) argues that tool kit theory understates the importance of deep moral systems, such as "individualistic" systems that prioritize self-gratification and "theistic" systems prioritizing faith. Using panel data, he shows that these systems are correlated with behavior years after their measurement. However, despite statistically controlling for many variables, Vaisey (2009:1703) concludes only that "the choice of moral [system] in 2002 is a very good overall predictor of behavior in 2005." As he warns, the challenge to viewing this predictiveness as causation is that moral systems are not chosen at random. Rather, because other factors may be "previously implicated in producing different moral schemas" (Vaisey 2009:1703), the observed correlation may be a dubious indication of how an individual's behavior would vary if her moral system were somehow changed.

Of course, moral systems are hardly unique in this respect; most "treatments" of interest cannot be ethically or practically randomized by the researcher. Thus, Polavieja identifies the method of instrumental variables, a widely used tool for estimating causal effects from nonexperimental data, as a promising avenue for future research. The essence of the method of instrumental variables is to regard certain observed variables as if they had been randomized, for example, through a "natural experiment" (Rosenzweig and Wolpin 2000). Intuitively, if these instrumental variables (IVs) strongly affect the treatment, and if they affect the outcome only through the treatment, then they can be used to estimate causal effects, even when the treatment has not been randomly assigned. Note, however, that the resulting estimates have a causal interpretation only if strong (and usually debatable) assumptions are met.

These assumptions are commonly introduced in terms of structural diagrams or

models. For example, consider a standard linear equation:

$$Y = \alpha + \tau T + \varepsilon \qquad (1)$$

where Y is the probability of employment, T is a variable indicating traditionalism, and ε is an error term that captures other causes of labor force participation. The quantity of interest in this model is τ, the effect of traditionalism on labor force participation. As is well-known, τ cannot be consistently estimated if T is correlated with ε, as when it is affected by a "lurking" variable that also affects Y.

The IV solution involves an additional variable, Z, which is strongly correlated with T but not ε. More precisely, the necessary conditions for Z to be a valid instrument are as follows. First, Z must not covary with other causes of Y. This *exogeneity* condition is expressed by the following equation:

$$Cov(Z, \varepsilon) = 0. \qquad (2)$$

Exogeneity is sometimes taken to mean that the instrument is as good as randomly assigned, because it implies that the instrument is uncorrelated with unobserved causes of the outcome. The second condition is the *exclusion restriction*, which states that the instrument affects the outcome only through its effect on the treatment. Finally, the instrument must covary with the treatment, or

$$Cov(T, Z) \neq 0. \qquad (3)$$

The first two IV assumptions, exogeneity and the exclusion restriction, cannot be tested from the data (Morgan and Winship 2007). Even when the instrument is known to be randomly assigned, and hence exogenous, the exclusion restriction can only ever be established in theory. Of course, an issue arising in most nonexperimental work is that the instrument is not known to be randomly assigned, and so the assumption of exogeneity is also rarely *prima facie* credible (Rosenzweig and Wolpin 2000). I discuss this aspect of IV methods in greater detail below.

THE PROPOSED METHODOLOGY

Polavieja extends the method of IVs to address the following question: How does a woman's traditionalism affect her labor force outcome? The problem with simply comparing the labor force outcomes of highly and lowly traditional women is that "assignment" to traditionalism is nonrandom. Consequently, highly traditional women likely differ from lowly traditional women along many unobservable or hard-to-measure dimensions relevant to labor force participation, such as the availability of work or endowment of abilities. For example, Gerson (1985) interviewed many women whose expanded work opportunities and discovery of latent talents produced greater ambivalence toward traditional feminine roles. These women would not be a suitable comparison group for highly traditional women, since modifying the traditionalist views of the latter group would not endow them with the same distributions of work opportunities or talents.

To address this problem, Polavieja introduced the Survey-based Imputation of Synthetic Traits used as Exogenous Regressors (SISTER) method, consisting of the following steps:

1. *Sampling step.* First, the researcher draws a sample of migrating and non-migrating women born in the same country. For migrants, only traditionalism at destination is observed, whereas only traditionalism at origin is observed for non-migrants.
2. *Imputation step.* Second, observationally equivalent non-migrating women are used to generate imputed values of traditionalism at origin for migrants. These values represent the levels of traditionalism that would have been observed for migrants, had they not chosen to migrate.
3. *IV step.* Finally, these imputed values are used as instrumental variables for the actually observed values among migrants, ostensibly to identify the causal effect of traditionalism on labor force participation.
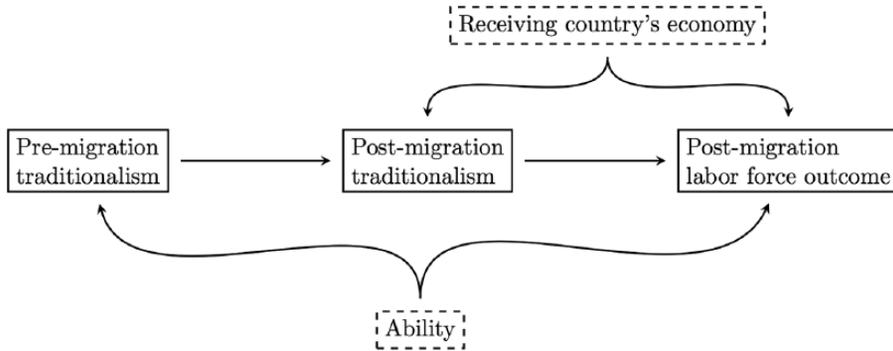
Figure 1. Pre-migration Traditionalism Is Not a Valid Instrument by Construction

The simple intuition behind this method is that the traits of non-migrating women can be viewed as unconfounded substitutes for the traits of migrating women. As Polavieja (2015:184) reasons,

> The fundamental property of synthetic traits is that they are *by construction* exogenous to the destination environment, because they are imputed using information from obser-vational-equivalent women who did not migrate.

Although compelling, this logic has several flaws. These are best illustrated by analyz-ing the imputation and IV steps of the SIS-TER method separately. Thus, the next section examines the "first-best" assump-tions of SISTER, assuming that the researcher has ideal data and does not need to impute any values. Specifically, I assume the researcher has collected data on migrants before and after migration, and believes that the pre-migration values would have remained the same without migration.[1] This eliminates the need to impute values based on *observationally* equivalent non-migrants, because the values of *exactly* equivalent non-migrants are observed. Because these values cannot be affected by the context of reception, and are therefore "exogenous" (in Polavieja's sense), the researcher may be tempted to use these pre-migration values as an IV. Unfortunately, as I will show, even these ideal data will rarely meet the IV assumptions.

FIRST-BEST ASSUMPTIONS

In the first-best scenario, the necessary assumptions are as follows. First, the exoge-neity condition states that the cultural traits measured prior to migration are unrelated to unobserved causes of labor force participa-tion after migration. Second, the exclusion restriction states that pre-migration traits do not affect post-migration labor force out-comes except through post-migration traits. Finally, pre-migration traits must covary with post-migration traits. For space, I address the first two assumptions only.

Unfortunately, even in the first-best case, the exogeneity assumption will rarely be credible. This is because women with differ-ent cultural values *before* migration may dif-fer in ways that influence their labor force outcomes *after* migration. For example, a conventional source of confounding bias is unobserved ability. Clearly, latent attributes like ability are not discarded in the process of migration. Although use of SISTER is moti-vated by the reasoning that pre-migration values will not be confounded by variables in the receiving country, pre-migration values may still be confounded due to attributes that follow migrants across borders, such as abil-ity (see Figure 1).

The second assumption, or exclusion restriction, will also tend to be implausible. If it is reasonable to posit a causal effect of tradi-tionalism on employment outcomes, it is also likely that traditionalism affects other behav-iors, such as childbearing and educational

Table 1. Hypothetical Distribution of Migrants' Observed Outcomes When Migration Is Random

		Employed after Migration	
		Yes	No
Traditionalism before Migration	High	25%*	75%
	Low	50%*	50%*

*Migrating cells when migration is nonrandom (see text).

Table 2. Hypothetical Distribution of Migrants' Observed Outcomes When Migration Is Selective

		Employed after Migration	
		Yes	No
Traditionalism before Migration	High	100%	–
	Low	50%	50%

attainment, known to affect employment independent of migration status. For example, consider a woman who would have a child if she were highly traditional prior to migration, and whose child would continue to affect her post-migration employment status. Then the exclusion restriction would not be satisfied, because pre-migration traditionalism affects post-migration labor force participation through a channel other than post-migration traditionalism. Unfortunately, there is no fix for this problem, as controlling for childbearing will induce post-treatment bias (Rosenbaum 1984).

A final assumption left implicit in this discussion is that individuals migrate at random.[2] This assumption is also implausible; for example, it will not be satisfied if some women migrate because they are sure to be employed in the receiving country, nor will it be satisfied if some women do not migrate because they are highly traditional. Polavieja (2015:187) states that the potential bias from selection into migration is unknown. In fact, nonrandom migration has the startling implication that, even if the exogeneity assumption is satisfied in the population, it may *not* be satisfied when migrants and non-migrants are analyzed separately, as in SISTER.

The following example illustrates. Suppose pre-migration traditionalism (Z) is randomly assigned in the sending country, so that Z is exogenous in the population. Further assume that, among individuals observed to have high pre-migration traditionalism ($Z = 1$), 25 percent would be employed after migration. By contrast, 50 percent of individuals with low pre-migration traditionalism ($Z = 0$) would be employed after migration. If a *random* sample of the population were to migrate, revealing their post-migration employment outcomes, a simple comparison of the average employment rate among high and low traditionalism migrants would be unbiased for the average effect of high traditionalism on employment, $\hat{\tau} = -.25$.

Tables 1 and 2 contrast the case of random migration with selective migration. Table 1 shows the expected employment rate among migrants when they represent a random draw from the sending country. However, suppose that individuals migrate only when they have low traditionalism or are sure to be employed in the receiving country. This is represented by asterisks in the corresponding cells in Table 1. This leaves Table 2. The same comparison of low- and high-traditionalism migrants results in a starkly biased estimate of the average treatment effect, $\hat{\tau} = .5$, despite random assignment of the instrument. This is a consequence of collider bias, which arises from the dual selection

into migration on the basis of employment and traditionalism (Elwert and Winship 2014).[3]

These examples illustrate how, even with ideal data, SISTER can entail unreasonable assumptions. In effect, the researcher assumes that pre-migration values are as good as randomly assigned, that they do not affect post-migration outcomes through alternative channels, and that the decision to migrate is random.

THE SECOND-BEST ALTERNATIVE

When researchers do not repeatedly observe the same individuals, they must resort to Polavieja's proposed version of SISTER. This version is "second-best" because it relies on missing data imputation, adding greater uncertainty and complexity. In fact, this section shows this is unnecessary, as the imputation step nearly always assumes that the researcher already has a valid, non-imputed IV. However, although the imputation step is unnecessary, it is not always innocuous. Not only does the imputation step reduce the transparency of the relevant assumptions, it can also inadvertently invalidate a valid instrument.

1. SISTER is unnecessary. In the second-best version of SISTER, Polavieja (2015:175) advises that the instrument be imputed using multiple regression. Specifically, suppose that, for each migrant i and non-migrant j, the researcher observes a set of covariates, such as race, gender, age, and so on. Denote these covariates (along with an intercept) by X_i for migrants and X_j for non-migrants. The proposed imputation methodology involves first modeling non-migrants' traditionalism (T_j) using multiple regression, or

$$\hat{T}_j = X_j^T \hat{\gamma}, \qquad (4)$$

where $\hat{\gamma}$ are multiple regression coefficients. Then the imputed value of traditionalism for migrants is estimated by multiplying their values of X by the estimated coefficients:

$$Z_i = X_i^T \hat{\gamma}. \qquad (5)$$

For simplicity, suppose the researcher observes only one covariate X. Then, for each migrant, the value of the instrument is $Z_i = \hat{\gamma}_0 + \hat{\gamma}_1 X_i$. However, using the linearity of covariances, it is easy to show that the validity of Z as an instrument implies the validity of X. First, if Z is to covary with T, it must be that $\hat{\gamma}_1 Cov(X,T) \neq 0$. But this implies that

$$\begin{aligned} 0 = Cov(Z,\varepsilon) &= Cov(\hat{\gamma}_0,\varepsilon) \\ &+ \hat{\gamma}_1 Cov(X,\varepsilon) \rightarrow Cov(X,\varepsilon) = 0. \end{aligned} \qquad (6)$$

Thus, the exogeneity of Z implies the exogeneity of X. Furthermore, because $Cov(X,T) \neq 0$ by the assumption that Z is valid, X covaries with the treatment. Finally, X must satisfy the exclusion restriction, since it is collinear with Z and cannot be included alongside Z in a regression. This establishes that the validity of Z as an instrument implies the validity of X.

In general, if the imputed instrument is valid, the researcher will nearly always have a valid, non-imputed instrument. A formal proof of this statement, generalizing to multiple imputation regressors, is provided in the Appendix (Proposition 1). Therefore, the imputation step is superfluous, as one can simply use the non-imputed variables as the instruments.

2. SISTER threatens transparency and validity. Although the imputation step is unnecessary, it is not necessarily innocuous. By transforming the non-imputed imputation regressors into a single variable, the imputation step hides the dependence of SISTER on *all* non-imputed regressors. For example, Polavieja (2015:175) uses "age, years of schooling, parental education, and religious denomination as predictors of traditionalism in the imputation regression." Equation 7 shows that all these variables must satisfy the exogeneity condition; it is not enough, as Polavieja (2015:175) suggests, to include "at least one regressor that is (arguably) orthogonal to the error term in the structural equation of interest." Misconceptions such as these are especially damaging to IV methods, which rely on assumptions that must be argued theoretically

rather than demonstrated using the data (Morgan and Winship 2007). Polavieja (2015:182) claims that regression diagnostics can test whether the instrument is exogenous, but this is a misunderstanding of the Wald exogeneity test, which examines whether the instrumented variable is endogenous on the assumption that the instrument is exogenous.

Unfortunately, the imputation step is not just obscurant, it can also diminish the validity of empirical estimates. To see this, suppose the instrument is now imputed with two variables, only one of which is exogenous, so that $Cov(X_1, \varepsilon) = 0$ but $\hat{\gamma}_2 Cov(X_2, \varepsilon) \neq 0$. But then

$$Cov(Z, \varepsilon) = Cov(\hat{\gamma}_0, \varepsilon) + \hat{\gamma}_1 Cov(X, \varepsilon)$$
$$+ \hat{\gamma}_2 Cov(X_2, \varepsilon) \neq 0. \tag{7}$$

This inequality demonstrates that, even when the researcher has a genuinely valid instrument, the imputed variable may still be invalid.

CONCLUSIONS: FROM UNMOVED MOVERS TO MOVABLE MOVERS

This comment evaluates a method recently proposed by Polavieja (2015) to estimate the causal effects of cultural values. I argue that Polavieja's SISTER method makes assumptions that are too strong for most applications, even when ideal data are available. When ideal data are unavailable, SISTER is an unnecessary and obscurant alternative to standard IV methods, and thus should not be used.

Are there more promising avenues for causal analysis in cultural sociology? An important task for future research—preliminary to the development of more elaborate methodological techniques—will be to sharpen the conceptualization of cultural effects. According to an increasingly prominent view, causation is best understood in terms of a comparison between potential states of the world, one in which a cause is applied to a unit and another in which it is not. For example,

If we say, "This boy has grown tall because he has been well fed," we are not merely tracing out the cause and effect in an individual instance; we are suggesting that he might quite probably have been worse fed, and that in this case he would have been shorter. (R. A. Fisher, quoted in Rubin [2005:323])

As Holland (1986) argues, thought experiments of this kind presuppose *manipulability*: in principle, one could design an experiment in which some boys were randomly chosen to be better fed than others. It is far less clear how one might similarly manipulate a person's cultural values without distorting many other fundamental characteristics relevant to outcomes of interest. This is not just a philosophical quibble, but bears, for example, on the specification of statistical models that purport to identify the effect of cultural values while "holding constant" variables such as parentage and context of upbringing. These models have no obvious interpretation, as no intervention could realistically induce such an effect (cf. Rosenbaum 1984).

A fruitful direction may therefore be to focus on the more manipulable aspects of culture, such as the habits and skills highlighted in tool kit theory. For example, cognitive and dual-process theories of culture posit that individuals can be induced to act in ways that are more or less motivated by their underlying moral systems (Vaisey 2009). Thus, Miles (2015) conducted an experiment in which individuals were randomly assigned numbers to memorize, increasing their cognitive load and thus their propensity of relying on deep-seated values. This design facilitates inference by manipulating the salience, rather than content, of values. Along similar lines, Sen and Wasow (2016) discuss research designs for estimating the influence of race, another "immutable" characteristic, by disaggregating it into its more flexible parts. Of course, in turning from the "unmoved movers" to the more movable aspects of culture, these designs arguably sacrifice some of classical theory's grand ambitions for greater clarity and credibility. How to manage these tradeoffs will be an important question as the causal analysis of culture progresses.

APPENDIX

Definition

Let

$$Y = \alpha + \tau T + \varepsilon \tag{A1}$$

be a population model for Y where $Cov(T, \varepsilon) \neq 0$. We say that a variable Z is a *valid instrument* for T or that Z *satisfies the instrumental variables (IV) assumptions* if $Cov(Z, \varepsilon) = 0$ and $Cov(Z, T) \neq 0$.

Proposition 1

Let $X_1, X_2, \ldots, X_k$ be variables in a multiple regression imputation function, so that

$$Z = \hat{\gamma}_0 + \sum_{j=1}^{k} \hat{\gamma}_j X_j. \tag{A2}$$

We further assume that

$$\hat{\gamma}_j Cov(X_j, \varepsilon) \neq 0 \rightarrow \hat{\gamma}_j Cov(X_j, \varepsilon)$$
$$\neq -\sum_{m=1}^{k} \hat{\gamma}_m Cov(X_m, \varepsilon), m \neq j.$$

(Assumption 1)

Assumption 1 rules out the exceptional event that the weighted covariances cancel each other out.

Then Z satisfies the IV assumptions only if X_j satisfies the IV assumptions for at least one j.

Proof of Proposition 1

1. Because Z is valid by assumption, $Cov(Z, T) \neq 0$. By definition of Z, this implies that

$$Cov\left(\hat{\gamma}_0 + \sum_{j=1}^{k} \hat{\gamma}_j X_j, T\right) \neq 0. \tag{A3}$$

2. By linearity of covariances, Equation A3 implies that

$$\hat{\gamma}_j Cov(X_j, T) \neq 0 \tag{A4}$$

for at least one j.

3. Because Z is valid by assumption, $Cov(Z, \varepsilon) = 0$. By definition of Z, this is the same as

$$Cov\left(\hat{\gamma}_0 + \sum_{j=1}^{k} \hat{\gamma}_j X_j, \varepsilon\right) = 0. \tag{A5}$$

4. By Equation A4, $\hat{\gamma}_j \neq 0$ for at least one j. By Assumption 1 and exogeneity of Z,

$$Cov(X_j, \varepsilon) = 0 \tag{A6}$$

for this j.

5. Equations A4 and A6 together imply that X_j is a valid instrument for T, *Q.E.D.*

Notes

1. I thank a reviewer for suggesting this organization.
2. This is an assumption of SISTER, not of IV methods more generally.
3. A simulation study demonstrating the effects of collider bias is given in the online supplement (http://asr.sagepub.com/supplemental).

References

Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40:31–53.

Gerson, Kathleen. 1985. *Hard Choices: How Women Decide about Work, Career, and Motherhood*. Berkeley: University of California Press.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–60.

Miles, Andrew. 2015. "The (Re)genesis of Values: Examining the Importance of Values for Action." *American Sociological Review* 80(4):680–704.

Morgan, Stephen, and Christopher Winship. 2007. *Counterfactuals and Causal Inference*. New York: Cambridge University Press.

Polavieja, Javier G. 2015. "Capturing Culture: A New Method to Estimate Exogenous Cultural Effects Using Migrant Populations." *American Sociological Review* 80(1):166–91.

Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable That Has Been

Affected by the Treatment." *Journal of the Royal Statistical Society, Series A* 147(5):656–66.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38(4):827–74.

Rubin, Donald. 2005. "Causal Inference Using Potential Outcomes." *Journal of the American Statistical Association* 100(469):322–31.

Sen, Maya, and Omar Wasow. 2016. "Race as a 'Bundle of Sticks': Designs that Estimate Effects of Seemingly Immutable Characteristics." *Annual Review of Political Science* 19:499–522.

Swidler, Ann. 1986. "Culture in Action: Symbols and Strategies." *American Sociological Review* 51(2):273–86.

Vaisey, Stephen. 2009. "Motivation and Justification: A Dual-Process Model of Culture in Action." *American Journal of Sociology* 114(6):1675–1715.

Winston Chou is a PhD candidate in the Department of Politics at Princeton University. His dissertation examines the causes of populist voting in France and the United States. His research also appears in *Social Forces*.