Received: 19 May 2021 | Accepted: 10 June 2021

DOI: 10.1111/nana.12756

REPLY

WILEY

Mapping culture with latent class analysis: A response to Eger and Hjerm

Bart Bonikowski | **Paul DiMaggio**

Department of Sociology, New York University, New York City, New York, USA

Correspondence
Bart Bonikowski, Department of Sociology, New York University, United States.
Email: bonikowski@nyu.edu

Abstract

Eger and Hjerm's methodological critique of our 2016 study of Americans' sentiments towards the nation asserts that the latent class (LCA) models employed in our paper did not fit the data and that consequently, the paper fails to demonstrate the existence of multiple varieties of American nationalism. We challenge E&H's analyses and argue that their conclusions stem from erroneous assumptions, both about our models and about best practices for applied LCA-based research. Based on a review of their results and additional analyses carried out with their preferred measures, we demonstrate that our model choices were justified and our 2016 findings are robust. In so doing, we offer a critique of unreflective adherence to inappropriate model fit criteria that ignores theory and concerns over the parsimony, interpretability, construct validity and external validity of model results.

> In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters.
>
> — Jorge Luis Borges, 'On Exactitude in Science'

1 | INTRODUCTION

In their research note 'Identifying Varieties of Nationalism: A Critique of a Purely Inductive Approach', Eger and Hjerm (2021) (hereafter E&H) reevaluate a subset of the analyses from our 2016 *American Sociological Review* article.

The authors claim to have conducted a 'thorough examination of the analytical strategy and exploratory methods' (p. 2) in our study and contend that due to our 'misinterpretation of [the] LCA analysis', there is 'no empirical basis for [our] claims' (p. 10). We appreciate the authors' engagement with our work and their suggestion that it 'could be pathbreaking for the field of nationalism' (p. 2), but we disagree with their conclusions and the premises on which those conclusions are based. Although E&H's overview of recent methodological research is diligent, the model evaluation strategy they advocate is neither reflective of the goals and standard practices of the substantive literature nor suitable for analysing the sort of complex attitudinal data we employ in our study. Moreover, they elevate a single approach to model fit over such generally accepted and equally important criteria as parsimony, interpretability, construct validity and external validity, all of which support our conclusions. For these reasons, as we will demonstrate, E&H's methodological critique fails to undermine our study's contribution.

Let us briefly recap our 2016 article. Guided by an interdisciplinary literature on American political culture, which spans historical analyses (Lieven, 2012; Smith, 1997) and survey-based research (Citrin et al., 1997, 2001; Schildkraut, 2010; Theiss-Morse, 2009), we posit that Americans hold heterogeneous dispositions towards their nation, that such dispositions cohere into discrete configurations and that the latter have implications for a range of social and political outcomes. Using latent class analysis (LCA)—a data reduction method that clusters respondents based on the similarity of their answers to multiple survey questions—we identify four such varieties of nationalist beliefs in data from the 2004 General Social Survey. Building upon insights from past research, we label the classes creedal, disengaged, restrictive and ardent. Our results demonstrate that respondents' class membership is systematically associated with their sociodemographic characteristics and opinions regarding ethnoracial group relations, economic redistribution, international cooperation and economically and culturally protectionist state policies. We go on to replicate these latent class results using two separate surveys (the 1996 General Social Survey and a 2012 GfK survey), confirming that the same four types of nationalism are present in each wave of data and that their distribution is relatively stable (with the exception of short-term deviations following the September 11th attacks). The approach developed in our article has been generative: in the years since our study's publication, similar configurations of nationalist beliefs have been observed in other data and shown to be associated with a range of outcomes, including voting preferences in the 2016 US presidential election (Bonikowski et al., 2019) and anti-immigrant sentiments and radical-right support in Germany and France (Bonikowski, 2017b). In addition, scholars have demonstrated that the country-specific distribution of these four types of nationalism has been shaped by nation-states' historical experiences with geopolitical threat to their sovereignty and territorial integrity (Soehl & Karim, 2021) and that in the United States, these beliefs have become increasingly sorted by party (Bonikowski et al., 2019).

We summarize our study's contribution and the research agenda it has inspired to illustrate the wide range of evidence supporting our fourfold typology's construct and external validity and theoretical utility. E&H consider none of this evidence in their critique. Instead, they hone in on a single question: whether the latent class models in our 2016 article adhere to their own preferred quantitative measures of fit—metrics that, as we will argue, are not useful for our study. By jettisoning theory and engaging in a mechanical pursuit of a single standard of model fit, not only do E&H end up with a hollow critique of our work but they also arrive at a substantive conclusion that is likely to puzzle anyone familiar with US political culture, namely, that there exists only one type of US nationalism. This argument is at odds with a voluminous literature in political science, history and sociology, on which our article explicitly builds. As it turns out, it is also demonstrably wrong based on E&H's own model selection strategy.

In contrast to E&H's analyses, our approach in the 2016 article is not purely inductive, as the title of E&H's piece alleges. In choosing our models, we were guided by carefully formulated theoretical priors, a thorough understanding of our data, extensive exploration of the full range of available modelling options, careful application of multiple validation criteria and purposeful replication of our results using multiple sources of data and configurations of variables. All of these considerations point to the four-class model as an appropriate representation of the data. Moreover, as we demonstrate below, had E&H implemented their own model selection strategy properly, they would have concluded that absolute model fit metrics point to the existence of no fewer than 29 dispositions towards the nation, which can be reasonably reduced to the four classes we described. In short, having considered E&H's evidence, we stand by our original results.

Our response proceeds as follows. We begin with a description of the analytical strategy we employed in the 2016 article, with a particular focus on our comprehensive approach to model selection and validation. The latter involved not only numerical measures of fit but also interpretive assessment of class content and prevalence and concern with the identification of robust and theoretically consistent associations with other variables in our data (i.e., construct validity). We again point out that comparable nationalism classes have been found in other data covering multiple countries and time periods, yielding empirical results relevant to a wide range of substantive research problems and demonstrating the external validity of our typology. Then, turning to model attributes in particular, we explain why relative rather than absolute goodness-of-fit criteria are more appropriate for clustering cultural dispositions—as reflected in substantive articles published in top sociology and political science journals—and reassert that in terms of relative fit, a four-class is optimal for our data.

In our view, these arguments sufficiently justify the conclusions of our 2016 article, but we go one step further and respond to E&H's challenges on their own terms. Using the narrow validity metrics favoured by E&H, we show that a 29-class model—not a one-class model—yields the most accurate reflection of the data. But here is the rub: the best fitting model is not the most useful one, as shown by a closer examination of the class output and comparison of goodness-of-fit criteria across models. Our four-class model is not only more interpretable but also outperforms the more complex solution in relative terms, based on commonly used fit measures that take into account model parsimony. This, in fact, is a common situation in attitudinal research, and it demonstrates why a theory-free search for perfect model fit is a poor strategy for identifying substantively useful typologies. Such an approach is bound to lead to either wrong conclusions, as in E&H's untenable claim that Americans share a single definition of the nation or useless ones that privilege complexity over parsimony and analytical relevance. In a final analysis, we show that the number of items used to measure each dimension of nationalism has no bearing on our model choice, but we do so using a more theoretically appropriate strategy than E&H's.

Having refuted E&H's methodological arguments, we offer a rejoinder to their theoretical disagreements with our paper, but do so briefly, because much of this critique is tangential to E&H's methodological points. We end our paper with a call for greater alignment of the methodological literature on LCA with the method's substantive application in actual research settings.

2 | LCA AND ATTITUDINAL RESEARCH

We begin with E&H's methodological critique of the LCA model selection strategy that yielded the identification of the four types of nationalism described in our article. For clarity, let us review E&H's main claims:

1. Verifying absolute model fit is a necessary and standard first step in LCA.
2. Because models with two to eight classes applied to our data do not satisfy criteria of absolute fit, 'there are not enough patterns [in the data] to identify more than one type of American nationalism' (Eger & Hjerm, 2021, p. 5).
3. Bootstrap likelihood ratio tests (BLRTs) are more appropriate tests of relative model fit (i.e., improvement in model fit based on the introduction of additional classes) when dealing with sparse tables than comparisons based on the Bayesian information criterion (BIC).
4. Relative fit criteria do not support the choice of a four-class model.
5. LCA of random combinations of four items drawn from our data do not support a four-class solution either.
6. By not reporting absolute fit statistics and choosing the four-class model in our article, we misinterpret our LCA results and build a theoretical framework without 'empirical basis' (Eger & Hjerm, 2021, p. 2).

We will demonstrate that claims 1, 2, 3, 4 and 6 are incorrect, whereas claim 5 and the analysis on which it is based are misguided.

2.1 | Absolute model fit criteria are unsuitable for clustering complex attitudinal data—and few applied scholars use them

At the heart of E&H's critique is the dual assumption—which appears to have been drawn solely from the simulation-based methodological literature on LCA—that establishing absolute model fit is a necessary first step in LCA and that models that do not satisfy this criterion should be discarded. This assumption, as we will argue, is mistaken: absolute fit is not in fact the most relevant model selection criterion for latent variable analyses of domain-specific attitudes, such as those employed in our article.

Let us clarify what is meant by 'absolute fit' and under what circumstances this metric may be useful. Absolute fit tests use the estimated model to produce expected values for all the observations on all the variables. If the expected values are statistically indistinguishable (at a standard significance threshold) from the observed data used to estimate the model, the model is said to fit well in absolute terms. In cases when tests based on standard goodness-of-fit statistics like the Pearson chi-squared yield unreliable p values, as when the data are sparse, an alternative is to estimate them from an empirical distribution generated via bootstrapping. The absolute fit approach may seem sensible at first, but fitting a model that faithfully reproduces the data while remaining simple enough to be useful is, as we shall argue, a tall order when modelling complex attitudes.

The problem is not only with the fit metrics themselves but also the methods used to evaluate them. Model fit criteria are typically assessed with simulation-based studies that follow a common strategy: they generate artificial data containing a known number of discrete classes and they then attempt to recover those classes using LCA models. The goodness-of-fit tests that most frequently identify the correct models are seen as the most reliable. Absolute fit criteria have often satisfied this requirement. The artificial data on which such simulation studies rely, however, bear little resemblance to the attitude surveys used by sociologists and political scientists (Dziak et al., 2019).

In what practical applications might one find the sorts of well-separated classes used in the simulation studies that endorse absolute fit metrics? One possible example is diagnostic patterns in medical research, a field that frequently utilizes LCA. Because physiological symptoms tend to occur in specific configurations for different diseases, they are likely to yield relatively distinct classes (though separation may be complicated by comorbidities). Symptom clusters should therefore be identifiable with LCA models that satisfy the types of fit metrics favoured by E&H.

In contrast to disease symptoms, variation in domain-specific cultural attitudes—when measured with more than a handful of items—is far more complex. This is so because individual beliefs are the product of heterogeneous sequences of social experiences shaped by a multitude of institutions, interactions and cultural milieus. As a result, in most cases, it is unrealistic to expect discrete classes of individual-level beliefs to be clearly demarcated in a population (Lanza & Rhoades, 2013). Instead, the objective of latent approaches to cultural measurement—whether using LCA or related methods, like relational class analysis (Goldberg, 2011) and concept-association-based schematic analysis (Hunzaker & Valentino, 2019)—is to reduce complex patterns of attitudinal variation to a manageable set of internally consistent classes (Hagenaars & Halman, 1989). Such classes, by definition, abstract from and distort a messy reality, but what they lose in descriptive precision they gain in interpretability and explanatory power. To quote a classic statistics adage, 'all models are wrong, but some are useful' (Box, 1979, p. 202).

Given the complex structure of attitudes, when dealing with sparse data (i.e., where a matrix of variable response categories by respondents features a large number of empty or low-count cells), such as our battery of 23 theoretically relevant items administered to nearly 1,000 respondents, the appropriate question is not 'can the model reproduce the data and therefore recover the 'true' classes in the sample?,' but rather, 'which of the various models is better than the alternatives at capturing the phenomenon of interest at a level of aggregation that is parsimonious and interpretable, has construct validity via theoretically expected associations with other variables, is reproducible in other data sets (i.e., has external validity), and helps illuminate a given research question?' As Dziak et al. (2019, p. 561) put it, 'the model desired is sometimes not the literally true model but simply the most useful model, a concept which cannot be identified using fit statistics alone but requires subjective judgement. Depending on the situation, the number of classes in a mixture model may either be interpreted as a true quantity needing to be objectively estimated, or else as a level of approximation to be chosen for convenience, like the scale of a map.' Thus, the choice of LCA models for complex attitude data necessarily depends on model *comparison*, which can be aided by goodness-of-fit measures that reveal whether the information gained from more complex models justifies their selection despite those models' reduced parsimony. This is the strategy we employ in our study.

Our approach, though different than E&H's, is entirely consistent with how LCA is used in empirical research published in highly selective peer-reviewed journals. In fact, we examined every LCA-based article ever published in *The American Sociological Review*, *The American Journal of Sociology*, *Social Forces*, *The American Political Science Review*, *The Journal of Politics* and *The American Journal of Political Science*(N = 34) and found only five instances of the use of absolute model fit measures, of which only two dealt with attitude data. In the other 29 studies, the authors chose models based on relative fit considerations (typically using the BIC, which penalizes less parsimonious models), along with theoretical and interpretability considerations. It is clear then that examining absolute model fit is far from 'the mandatory first step in evaluating LCA models' (p. 7), as E&H characterize it. On the contrary, it departs from mainstream practice in sociology and political science, the two social science disciplines that most frequently rely on LCA.

How do we explain the discrepancy between the methodological literature reviewed by E&H and prevalent norms in social science? Are the authors, reviewers and editors of these disciplines' most selective journals misguided? Not at all. Conventional practice follows from the common objective of these studies, which our 2016 article shares: to reduce complex attitudinal data to a small number of types that approximate overarching attitudinal dispositions and illuminate important substantive questions. Perhaps E&H wish to make a broader methodological intervention into this substantive literature and argue that absolute model fit should become a *new* first step in LCA, maybe even a 'mandatory' one. If so, such a critique would need to show not only that the models in dozens of extant studies exhibit poor fit but that the substantive findings based on those models are themselves incorrect and that such criteria as parsimony, construct validity and external validity are irrelevant to model evaluation.

2.2 | Relative fit, construct validity and external validity should guide LCA model selection

Setting absolute model fit aside for now, let us briefly explain how we selected the four-class model in the 2016 article and why we are confident that the model represents the data well. Consistent with the best practices in the literature, we rely on a combination of parsimony-prioritizing relative fit comparisons, theoretical priors, careful examination of the classes yielded by various model solutions and concerns with interpretability.[1] Moreover, we adhere to the standard principle in the substantive literature that 'identified classes can be validated by investigating whether there are expected relationships between classes and other variables' and 'conducting the same analysis in different samples, or with a subset of the same sample, to see if they are consistently found' (Petersen et al., 2019, p. 2). This ensures that the ideal types identified by LCA satisfy criteria of construct and external validity.

Specifically, we (1) compared BIC metrics across models, revealing that the four-class model strikes a reasonable balance between accuracy and parsimony (more on this later); (2) looked at whether additional classes identified by more complex models added substantively interesting information or were mere variations of previously identified

[1]Oddly enough, these are the practices recommended by E&H, even though their critique fails to engage with any of our validation evidence aside from quantitative model fit.

classes; (3) examined the sample proportions of additional classes to determine if they represented non-trivial numbers of respondents; (4) varied the number of indicators measuring each aspect of nationalism to ensure that our solution was not driven by imbalanced measures; and (5) experimented with a wide range of model specification options related to weights, direct effects, covariates and error variances. The four-class model emerged as optimal following all of these procedures.

Just as important, our study clearly demonstrates the four-class model's construct validity by systematically associating it with numerous demographic variables and a plethora of political attitudes, revealing patterns consistent with theoretical expectations (none of which were substantively altered by the model specification decisions enumerated above). This is something that an artifactual model simply could not accomplish. Finally, as stated earlier, the same four classes have been repeatedly replicated in distinct data sets from different years (Bonikowski & DiMaggio, 2016; Bonikowski et al., 2019) and countries (Bonikowski, 2013, 2017b; Soehl & Karim, 2021), attesting to the typology's external validity. Surprisingly, these are considerations that E&H choose to disregard in their critique.

3 | WHICH MODEL SHOULD WE SELECT?

Having explained our model selection strategy and shown it to be consistent with the norms of the substantive literature, we now turn to the model fit statistics themselves. First, we establish that the appropriate fit statistics that take into account parsimony do in fact favour a four-class model in relative terms. Second, we use E&H's preferred absolute fit tests to reveal errors in their interpretation of their own model results and to identify a model that fits the data in accordance with E&H's definition. Finally, we show that our four-class model is again preferable to the model yielded by a proper implementation of E&H's approach, based on relative fit, interpretability and theoretical utility.

Our first point is simple: Table 1 in E&H's research note contains all the information necessary to select a four-class model based on relative model fit. When E&H compare models with one through eight classes, the lowest BIC—the most commonly used metric and one repeatedly shown to be robust across various LCA specifications—corresponds to the model with four classes, clearly indicating it to be the model of choice.[2] It is odd that E&H describe our use of BIC as ‘reasonable’ (p. 7) but then use the work of Nylund et al. (2007) and Tein et al. (2013) to argue that BICs fare more poorly with sparse data compared to E&H's preferred relative fit metric, the BLRT.[3] This is surprising not only because the use of BIC without the BLRT is widespread in applications of LCA to sparse attitudinal data but also because Nylund et al. (2007) and Tein et al. (2013) never mention sparseness as a consideration in the comparison of these two metrics. In fact, these authors assert that both BIC and BLRT are excellent measures of model fit, particularly when dealing with a sample size comparable to ours. van Kollenburg et al. (2015, p. 78) concur when they state that ‘interpretation of [...] information criteria like the AIC and BIC does not change when the sample size is small or when contingency tables are sparse.’ Nonetheless, when choosing between the BLRT and BIC, we favour the latter because it places a greater premium on parsimony (Dziak et al., 2019).[4] Also, as we will demonstrate, the BLRT is not helpful for selecting an optimal model in our particular case.

[2]In our original study, the model fit results differed in minor ways from E&H's. The BIC initially favoured a five-class solution, but the inclusion of direct effects for pairs of indicators based on model residuals resulted in the four-class solution having the lowest final BIC. In the analyses presented here, we make use of both baseline LCA models and those with direct effects.

[3]The BLRT reports whether the log-likelihood of a k-class model is significantly different from that of a (k-1)-class model. A statistically significant result indicates that the absolute fit of the more complex model represents an improvement over the simpler model.

[4]The BLRT takes into account degrees of freedom—and therefore the number of parameters being estimated—but it is less sensitive to model complexity than the BIC, which includes a penalty term based on the product between the number of model parameters and logged sample size.

3.1 | Absolute fit criteria point to multiple varieties of nationalism and relative fit criteria favour our model

The issue, however, is not just whether our approach is more appropriate than E&H's—E&H's argument fails even by their own absolute-fit standards, because they misinterpret their own results in Tables 2 and 3 of their research note. E&H's Table 2 shows that *none* of the models estimated by the authors, ranging from one class to eight, satisfies their criteria of absolute fit. Yet the authors conclude that 'a one–class solution (K_1) fits the data well enough' and that 'there are not enough patterns among these 23 indicators to identify more than one type of American nationalism' (Eger & Hjerm, 2021, p. 5). Not only is the notion that there is one type of American nationalism theoretically untenable (see Citrin et al., 2001; Smith, 1997; Schildkraut, 2010), but it is directly at odds with E&H's evidence as evaluated by their own preferred metrics: the one-class model fits the data poorly in both absolute (their Table 2) and relative (their Table 3) terms. The reasonable conclusion is not that there is one type of nationalism, but that—according to metrics that assume a high degree of class separation and do not privilege parsimony—there are *many more* types of nationalism than the models in E&H's tables are able to identify.

To restate the above, if our four-class model allegedly did not fit the data, the onus was on E&H to find one that did—and yet, they did not pursue this course of action. Well, we have. Using E&H's strategy, we have identified a model for which parametric bootstrap procedures yield a significant L^2, a non-significant $\chi^2$and a non-significant Cressie–Read statistic (i.e., conditions that satisfy E&H's own preferred absolute fit criteria), as illustrated in Table 1.[5] That model contains 29 classes. In other words, what E&H should have concluded, according to their own analytical logic, is that there are 29 types of American nationalism.

Does this mean that we will abandon our existing theoretical agenda on nationalism in favour of a new perspective focusing on 29 varieties of the phenomenon? No. Let us list the reasons. First, a 29-class model is not useful. Interpreting regression models with 28 dummy variables measuring a single phenomenon does not make for very good social science. Second, the classes do not map onto sufficiently large segments of the sample (which is not surprising given their number relative to the sample size): nine of the classes comprise 10 or fewer respondents each and the two most prevalent classes comprise only 83 and 80 respondents, respectively. Third, the attitudinal profiles of the 29 classes are simply more complex variations on the four-class model, as illustrated in Figure 1, which aligns the class-specific item response means from the two models. The additional information generated by the complex model is therefore redundant. These considerations lead us to conclude that choosing the 29-class model over simpler models would be a mistake. As Healy (2017) argues, the quest for ever more nuanced typologies is a fool's errand—what we need are theories that maximize explanatory power. In fact, in many applied contexts, ours included, 'we lose information by adding detail' (Healy, 2017, p. 122).

But interpretation is not the sole basis on which we reject this model—the model also fails on the basis of the BIC, the most commonly used fit statistic that takes into account model parsimony (Nylund et al., 2007). The 29-class model's log-likelihood BIC is 37,688 compared to 35,147 for the four-class model, which reveals the four-class model to be preferable.[6,7] What about BLRT, the relative model fit test preferred by E&H? It consistently favours models with more complexity and mistakenly rejects the 29-class model for the 30-, 31-, and 32-class models, even though none of the latter meet E&H's absolute fit criteria—or our criteria of interpretability. In such circumstances, it is reasonable to favour a BIC comparison over the BLRT (Lanza et al., 2012).

[5]To ensure sufficient degrees of freedom, we estimate this model with all the nationalism indicators but without sociodemographic covariates. The total bivariate residuals of the model are still above the threshold specified by E&H, but this can be partly alleviated by including direct effects for indicator pairs.

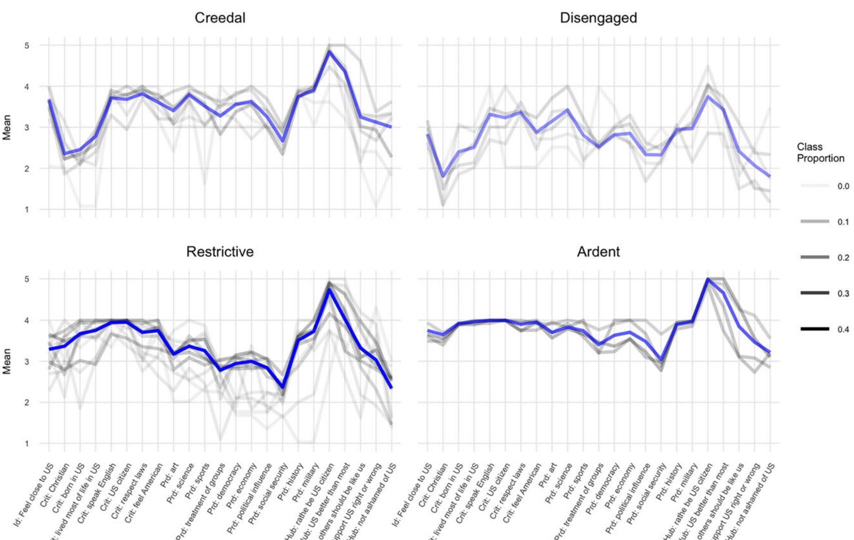
[6]For a similar example of selecting a model that fails an absolute fit test but has the lowest BIC and yields interpretable classes, see Lanza and Rhoades (2013).

[7]To best approximate the four-class model from our 2016 article, we relax the local independence assumption for 10 pairs of indicators with the highest model residuals via the inclusion of direct effects (Vermunt, 1997), a strategy for improving model fit that we had employed in our original analysis. This results in a lower BIC for the four-class model than for models with five through eight classes. A four-class model without direct effects has a BIC of 35,710, which is still well below that of the 29-class model.

TABLE 1 Model fit statistics for four-class model versus more complex alternatives

Model	LL	BIC(LL)	L^2	L^2 *p* value	χ^2	χ^2 *p* value	Cressie-Read	Cressie-Read *p* value
4-class (10 DEs)	−17,049	35,147	22,944	0.444	2.3E+17	0.000	7.5E+11	0.000
25-class	−16,448	37,265	21,741	0.018	7.5E+15	0.000	9.1E+10	0.000
26-class	−16,398	37,327	21,642	0.008	9.9E+13	0.032	6.4E+09	0.020
27-class	−16,415	37,522	21,676	0.010	7.0E+13	0.034	4.9E+09	0.028
28-class	−16,399	37,650	21,643	0.014	2.3E+13	0.054	2.8E+09	0.042
29-class	−16,337	37,688	21,519	0.002	9.5E+12	0.184	2.0E+09	0.080
30-class	−16,333	37,841	21,511	0.008	5.1E+14	0.006	1.3E+10	0.006

Note: All *p* values are bootstrapped using 500 iterations. The four-class model relaxes the local independence assumption via direct effects for 10 pairs of indicators with the largest residuals, in line with the modeling strategy from our 2016 *ASR* article.



Note: Line opacity is scaled to the sample proportions of the corresponding classes, with darker shades indicating more prevalent classes. Classes from the 29-class and 4-class models were classified into the four types of nationalism from our 2016 *ASR* article based on their overall response patterns. We omitted one low-prevalence class generated by the 29-class model because it did not appear to reflect substantively meaningful responses.

FIGURE 1 Class-specific item response means, four-class model (blue/black) versus 29-class (grey) model [Colour figure can be viewed at wileyonlinelibrary.com]

Taken together, the above evidence suggests that the model favoured by E&H's fit criteria is simply too complex. As in Borges' short story quoted in this article's epigraph, the 29-class model is akin to a map that represents the territory with such fidelity that it becomes useless. In contrast, our four-class model makes for a fine map, even if it does not capture every feature of the landscape.

3.2 | Correcting for item imbalance still yields a four-class model

E&H regard the imbalance in the number of indicators tapping different dimensions of national sentiment as a potential problem. We agree. For that reason, we had conducted robustness checks with subsets of the indicators in the course of our *ASR* article's peer review—analyses that confirmed the robustness of the four class model. We believe that our approach to this problem was appropriate to the data, whereas E&H's is not. Specifically, by iteratively selecting different sets of four indicators at random from our data, E&H disregard our theoretical arguments about the composition of nationalist beliefs. The resulting latent constructs are fundamentally incomparable with the analyses in our article, which intentionally examine the relationships among measures of attachment, identity, pride and hubris, and not just one or two of those elements at a time.

In the absolute best case, E&H's four-item models are able to select only one variable from each of the four subsets of nationalist attitudes (i.e., attachment, identity, pride and hubris). But this again confuses subsets of national sentiment with attitudes within them and ignores the latter's heterogeneity. Based on prior research, we make it clear in our 2016 article that national identity comprises inclusive and exclusionary attitudes (what E&H call civic and ethnic nationalism) and that national pride spans evaluations of national heritage and of national political, economic and cultural institutions.

A more appropriate alternative is to select variables for a reduced model in a theoretically purposeful manner that is attuned to the heterogeneity within the subsets of nationalist attitudes mentioned above. To determine whether item imbalance poses a problem for our study, we constructed the most sensible model specification guided by the above theoretical criteria; it includes two identity variables (importance of birth in the country and of subjective identification with the country as criteria of legitimate national membership), two pride variables (pride in the nation's accomplishments in the arts and in the state of its democracy) and two hubris variables (endorsing the claims that the world would be better if others were more like Americans and that 'America' is the greatest country in the world). We exclude national attachment because it is measured with only one item and has not proven to be particularly influential in our past analyses.

When comparing LCA model specifications using this subset of variables, as shown in Table 2, a similar pattern emerges to what we observed with the full set of 23 indicators: a four-class model has the lowest BIC but a significant bootstrapped χ^2, whereas a higher order solution—in this case, one with six classes—satisfies E&H's absolute model fit thresholds but is unnecessarily complex. The inclusion of direct effects for indicator pairs (Vermunt, 1997), an approach to improving model fit that we employed in our 2016 article, further reduces the BIC of the four-class model, well below that of the six-class model, and brings the former in line with E&H's absolute fit criteria.[8] Fit statistics aside, closer inspection of the attitudinal composition of the six-class model demonstrates its failure to provide additional useful information: the model directly reproduces the disengaged and ardent classes from the four-class model and yields minor variations of the latter's creedal and restrictive classes (based solely on differences in the two hubris indicators). Once again, the results of the more complex model are redundant. At the same time, the very fact that we can interpret a four-class model estimated with six survey items using the same conceptual categories as our original four-class model estimated with 23 items further attests to the robustness of our typology.

We focused on the above model in some detail, because this enabled us to once again demonstrate the importance of closely examining multiple model selection criteria. A sceptical reader could wonder, however, whether models based on other combinations of items would produce different results. To answer this question, we conducted a similar model permutation analysis to E&H's but did so in a manner that ensured the complete coverage of the relevant subsets of nationalism items. Specifically, each model drew one item per category from among measures of exclusionary national identity, inclusive national identity, pride in national heritage and pride in national institutions, as well as two items from among the hubris measures. We excluded from the analysis six items that did not

[8]This illustrates another situation in which absolute fit criteria may align with BIC-based model comparisons: when the number of indicators is small. Consistent with our earlier discussion of attitude measurement, however, when faced with a trade-off between maximizing the descriptive coverage of domain-specific attitudinal measures and absolute model fit criteria, we prefer and recommend the former option.

TABLE 2 Model fit statistics for a six-indicator model chosen based on theoretical priors

Model	LL	BIC(LL)	L^2	L^2 p value	χ^2	χ^2 p value	Cressie–Read	Cressie–Read p value
1-class	−6,167	12,471	1,988	0.000	7,647,933	0.000	92,276	0.000
2-class	−5,975	12,134	1,604	0.000	394,514	0.000	15,066	0.000
3-class	−5,910	12,052	1,473	0.000	36,584	0.000	4,806	0.000
4-class	−5,880	12,040	1,414	0.002	14,092	0.030	3,374	0.022
4-class (four DEs)	−5,835	11,978	1,324	0.006	7,322	0.132	2,745	0.050
5-class	−5,860	12,048	1,373	0.006	14,603	0.026	3,393	0.010
6-class	−5,836	12,049	1,327	0.036	7,666	0.114	2,729	0.064
7-class	−5,815	12,054	1,284	0.064	8,439	0.070	2,715	0.036
8-class	−5,802	12,076	1,258	0.032	7,798	0.108	2,658	0.040
9-class	−5,793	12,107	1,241	0.030	10,754	0.034	2,853	0.018
10-class	−5,788	12,145	1,230	0.028	13,866	0.040	3,138	0.016

Note: All *p* values are bootstrapped using 500 iterations. The four-class model with direct effects (DEs) relaxes the local independence assumption for four pairs of indicators with the largest residuals, in line with the modeling strategy from our 2016 *ASR* article.

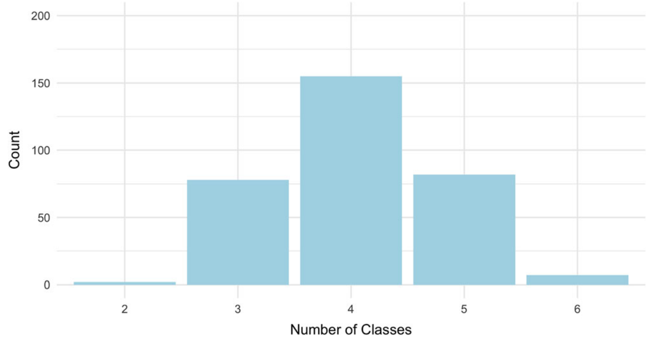


FIGURE 2 Count of best fitting six-item models based on all permutations of theoretically relevant items [Colour figure can be viewed at wileyonlinelibrary.com]

unambiguously fit into these five categories. In addition, national attachment was once again omitted to ensure balance in the number of items measuring each of the five subsets of attitudes. We ran models ranging from 1 to 10 classes for all possible combinations of the variables within the constraints of the five-category grouping and then noted which model in each of the 324 variable combinations yielded the lowest BIC. The results are reported in Figure 2. Four-class models clearly offer the optimal fit to the data for the vast majority of the item combinations (155, compared to 82 for the five-class models)—and this is without the inclusion of any direct effects based on model residuals. This exercise once again demonstrates a four-class model to be the correct choice for our purposes and bolsters the results of our 2016 article.

4 | THEORETICAL DISAGREEMENTS

Having refuted E&H's methodological critiques, we briefly discuss the theoretical challenges with which E&H's research note begins. These are tangential to the rest of their piece, but we address them here because they reveal further misunderstandings of our work.

At the heart of the matter appears to be the authors' distaste for our claim that people hold overarching dispositions towards the nation, comprising a range of specific attitudes. As stated earlier, we group those attitudes into the categories of attachment, identity criteria of authentic membership, pride and hubris. Frankly, we are puzzled by the authors' strong reaction to this conceptual framework, because we do not see our argument as particularly controversial. Ample research in sociology and political science, including sources that we and E&H both cite, demonstrates consistent and patterned correlations between survey items that tap attitudes towards the nation. Given extensive evidence from cognitive psychology and cultural sociology that cognitive representations of social domains are organized relationally (DiMaggio, 1997; Goldberg, 2011), it follows that people's conceptualization of the nation should take the form of a network of related beliefs that vary in some patterned way across the population.

What may appear more controversial is that we subsume the four subsets of attitudes under the common rubric of 'sentiments towards the nation' and classify them using our own preferred labels, rather than concepts favoured by political psychologists, such as patriotism or nationalism-*qua*-chauvinism. These semantic choices are theoretically principled, as we explain at length in our published work (Bonikowski & DiMaggio, 2016; Bonikowski, 2016, 2017a, 2017b), but they also do not affect the main empirical conclusions of our 2016 article or E&H's critique of it.

Nonetheless, E&H present two surprising misunderstandings of our argument: that we 'reject both the classic two-dimensional approach to national identity and the notion that it is distinct from other kinds of political sentiments' and that we 'conflat[e] theoretically distinct concepts under the umbrella of nationalism' (p. 3). We disagree with both charges: we neither reject the dual categories of civic and ethnic nationalism (though we view them as insufficient for capturing the full variation in nationalist sentiments) nor conflate distinct concepts, by which E&H seem to mean attachment, membership criteria, pride and hubris (to use our labels). Instead, we argue that (a) what unites the various attitudes included in our model is their common focus on the nation; (b) these attitudes are organized in patterned configurations that represent distinct conceptions of nationhood; and (c) the relational structure of the patterned configurations affects the meaning of the constitutive attitudes.[9] This in no way undermines the idea that national pride is analytically distinct from hubris or from criteria of national belonging. Our approach simply allows for heterogeneity within these categories and places their constitutive attitudes in relation to one another.

It bears repeating, however, that these disagreements are not put to an empirical test in E&H's research note. No amount of LCA model tweaking can determine the correct label for an analytical category or determine whether patterned associations between nation-related attitudes can be understood as approximations of cognitive dispositions. These are theoretical and conceptual questions that can only be adjudicated on theoretical and conceptual grounds. We would welcome such a debate, but it is outside the scope of the present exchange.

5 | CONCLUSION

It is ironic that E&H accuse us of espousing a 'purely inductive' approach. As we show, it is their focus on a single (and in our view misguided) standard of validity—to the exclusion of parsimony, interpretability, construct validity and external validity—that leads to the erroneous conclusions in E&H's critique. As Ram and Grimm (2009, p. 571) remind us in a passage that E&H quote, 'model selection is an art—informed by theory, past findings, past experience

[9]Contrary to E&H's claim, we see our arguments as largely consistent with Brubaker's work (Brubaker, 1992, 2004; Brubaker et al., 2004). As our article makes clear, we cite his work not to justify our rejection of the term 'patriotism,' but to support our claims about the nation as a cognitive and cultural category.

and a variety of statistical fit indices.' This does not mean that it is arbitrary or inscrutable but simply that it requires weighing multiple criteria and staying focused on the research's explanatory objectives, as we did in our study.

Ultimately, the point of analytical methods, LCA among them, is to help answer substantive research questions. The methodological choices at stake in the present exchange are important, but one must not lose sight of the ends to which we used LCA in the first place: to reduce the complexity in attitude data and to produce a typology of nationalist dispositions that could help us better understand Americans' social and political attitudes. After carefully reviewing and refuting E&Hs critique, we are confident that our 2016 *ASR* paper fulfilled that objective successfully. Moreover, research informed by our approach has led to an improved understanding of other political–cultural processes of interest to sociologists and political scientists, ranging from the relationship between the partisan sorting of nationalist beliefs and recent electoral outcomes in the United States (Bonikowski et al., 2019) to the effects of historical sovereignty crises on present-day social exclusion across a wide range of countries (Soehl & Karim, 2021).

Finally, this exchange reveals the need for greater scholarly attention to the disjuncture between the methodological literature on LCA and the method's use in substantive research. As we have argued, social scientists rarely use absolute model fit measures in applied settings—and when they do, these metrics often point towards different conclusions than relative fit criteria and concerns over model parsimony, interpretability and usefulness. Simulation-based studies, in contrast, have shown the potential promise of both absolute fit measures and bootstrapped likelihood-ratio tests for model selection when the goal is to identify objectively existing classes, such as diagnostic categories with physiological or morphological foundations. The contradictions between these two literatures should be reconciled by identifying clearer scope conditions for the use of different metrics for particular types of real-world data.

ACKNOWLEDGEMENTS

The authors are grateful for feedback on earlier versions of the manuscript from Sean Bock, Cristobal Young, Amir Goldberg, Yuval Feinstein, Kristina Bakkær Simonsen and Stuart Perrett.

ORCID

Bart Bonikowski https://orcid.org/0000-0003-0107-0644

REFERENCES

Bonikowski, B. (2013). Varieties of popular nationalism in modern democracies: An inductive approach to comparative research on political culture. Weatherhead Center for International Affairs, Harvard University.

Bonikowski, B. (2016). Nationalism in settled times. *Annual Review of Sociology*, *42*(1), 427–449. https://doi.org/10.1146/annurev-soc-081715-074412

Bonikowski, B. (2017a). Ethno-Nationalist populism and the mobilization of collective resentment. *The British Journal of Sociology*, *68*, S181–S213.

Bonikowski, B. (2017b). Nationhood as cultural repertoire: Collective identities and political attitudes in France and Germany. In Skey, M., & Antonsich, M. (Eds.), *Everyday nationhood: Theorising culture, identity and belonging after banal nationalism* (pp. 147–174). London, UK: Palgrave.

Bonikowski, B., & DiMaggio, P. (2016). Varieties of American popular nationalism. *American Sociological Review*, *81*(5), 949–980.

Bonikowski, B., Feinstein, Y., & Bock, S. (2019). The partisan sorting of 'America': How nationalist cleavages shaped the 2016 u.s. presidential election. SocArXiv https://osf.io/preprints/socarxiv/pmg95/

Box, G. E. P. (1979). Robustness in the Strategy of Scientific Model Building. In R. L. Launer & G. N. Wilkinson (Eds.) *Robustness in Statistics* (pp. 201–236). New York: Academic Press.

Brubaker, R. (1992). *Citizenship and nationhood in France and Germany*. Cambridge, MA: Harvard University Press.

Brubaker, R. (2004). *Ethnicity without groups*. Cambridge, MA: Harvard University Press.

Brubaker, R., Loveman, M., & Stamatov, P. (2004). Ethnicity as cognition. *Theory and Society*, *33*(1), 31–64.

Citrin, J., Green, D. P., Muste, C., & Wong, C. (1997). Public opinion toward immigration reform: The role of economic motivations. *The Journal of Politics*, *59*(3), 858–881.

Citrin, J., Wong, C., & Duff, B. (2001). The meaning of american national identity. In Ashmore, R. D., Jussim, L. J., & Wilder, D. (Eds.), *Social identity, intergroup conflict, and conflict reduction* (pp. 71–100). New York: Oxford University Press.

DiMaggio, P. (1997). Culture and cognition. *Annual Review of Sociology*, *23*(1), 263–287.

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2019). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics*, *21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Eger, M. A., & Hjerm, M. (2021). Identifying varieties of nationalism: A critique of a purely inductive approach. *Nations and Nationalism*. https://onlinelibrary.wiley.com/doi/abs/10.1111/nana.12722

Goldberg, A. (2011). Mapping shared understandings using relational class analysis: The case of the cultural omnivore reexamined. *American journal of sociology*, *116*(5), 1397–1436.

Hagenaars, J. A., & Halman, L. C. (1989). Searching for ideal types: The potentialities of latent class analysis. *European Sociological Review*, *5*(1), 81–96. http://www.jstor.org/stable/522517

Healy, K. (2017). Fuck nuance. *Sociological theory*, *35*(2), 118–127.

Hunzaker, M. B. F., & Valentino, L. (2019). Mapping cultural schemas: From theory to method. *American Sociological Review*, *84*(5), 950–981.

Lanza, S. T., Bray, B. C., & Collins, L. M. (2012). An introduction to latent class and latent transition analysis. In *Handbook of Psychology* (2nd ed., pp. 691–716). Hoboken, NJ: Wiley.

Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention science: The official journal of the Society for Prevention Research*, *14*(2), 157–168. https://pubmed.ncbi.nlm.nih.gov/21318625

Lieven, A. (2012). *America right or wrong: An anatomy of american nationalism*. New York: Oxford University Press.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Petersen, K. J., Qualter, P., & Humphrey, N. (2019). The application of latent class analysis for investigating population child mental health: A systematic review. *Frontiers in Psychology*, *10*, 1214. PubMed-not-MEDLINE.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, *33*(6), 565–576. https://doi.org/10.1177/0165025409343765, PMID: 23885133.

Schildkraut, D. J. (2010). *Americanism in the twenty-first century: Public opinion in the age of immigration*. New York: Cambridge University Press.

Smith, R. M. (1997). *Civic ideals: Conflicting visions of citizenship in us history*. New Haven, CT: Yale University Press.

Soehl, T., & Karim, S. (2021). How legacies of geopolitical trauma shape popular nationalism today. *American Sociological Review*, *86*(3), 406–429.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical power to detect the correct number of classes in latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *20*(4), 640–657. https://pubmed.ncbi.nlm.nih.gov/24489457

Theiss-Morse, E. (2009). *Who counts as an American?: The boundaries of national identity*. New York: Cambridge University Press.

van Kollenburg, G. H., Mulder, J., & Vermunt, J. K. (2015). Assessing model fit in latent class analysis when asymptotics do not hold. *Methodology*, *11*(2), 65–79.

Vermunt, J. K. (1997). *Log-linear models for event histories*, Thousand Oaks, CA: Sage.

How to cite this article: Bonikowski, B., & DiMaggio, P. (2022). Mapping culture with latent class analysis: A response to Eger and Hjerm. *Nations and Nationalism*, *28*(1), 353–365. https://doi.org/10.1111/nana.12756