ONLINE SUPPLEMENTARY MATERIAL

for

Berry, Fording and Crofoot,
"Response to Lagodny, Jones Koch, and Enns,"
***State Politics and Policy Quarterly*, MONTH 2023: xxx-xxx.**

Lagodny et al.'s Measure of State Policy Mood: Substantially Different than Enns and Koch's Measure, But Raising the Same Validity Concerns

William D. Berry
Marian D. Irish Professor and Syde P. Deeb Eminent Scholar in Political Science
Department of Political Science
Florida State University
Tallahassee, FL 32306-2230
wberry@fsu.edu

Richard C. Fording
Marilyn Williams Elmore and John Durr Elmore Endowed Professor
Department of Political Science
University of Alabama
Tuscaloosa, AL 35487
rcfording@as.ua.edu

Justin K. Crofoot
Graduate Student
Department of Political Science
Florida State University
Tallahassee, FL 32306-2230
jcrofoot@fsu.edu

January 2023

Lagodny, et al. (2023) reassert Enns and Koch's (2015) claim that Enns and Koch's (2013) measure of state policy mood (to be abbreviated E&K's measure) is valid but Berry et al.'s (1998) measure (to be abbreviated BRFH's measure) is not. Lagodny et al. (2023, 2) also introduce a new measure of state policy mood (to be abbreviated LJKE's measure) that they claim has "even better properties than [E&K's measure]." In this paper, we undertake five tasks:

- We subject LJKE's new measure to the same battery of face validity and construct validity tests that Berry et al. (2015) applied to Enns and Koch's measure. The results show that the characteristics of E&K's measure that make us doubtful that it is valid are also present in LJKE's measure—making us doubtful that LJKE's measure is valid.
- We argue that the correlation between E&K's measure and LJKE's measure is low enough to imply that LJKE's (2023, 14) claim that both their measure and E&K's measure are valid and that their measure is "even better" cannot be correct; if one measure is valid, the other is not.
- We extend Berry et al.'s (2023, Table 2) replication of published studies—which operationalizes state policy mood using each of BRFH's measure and E&K's measure—to use LJKE's measure as well. We find that conclusions about the effect of state policy mood are frequently substantially different when using LJKE's measure than when using E&K's measure—thereby providing additional evidence that if one of these measures is valid, the other cannot be.
- We respond directly to a few assertions by LJKE (2023) that we believe are unjustified or wrong.
- We conclude with some observations about where the state politics field stands in its quest to construct a measure of state policy mood that could be widely accepted as valid.

LJKE's Measure Raises Similar *Face* Validity Concerns as Does E&K's Measure

Berry et al. (2015) identify three respects in which E&K's measure lacks face validity.[1] Our analysis of LJKE's measure leaves us with the same concerns about this new measure.

- Berry et al. (2015) contend that scores for E&K's measure are at odds with conventional wisdom about policy mood in the South in the sense that the measure indicates that the Deep South (defined as the 11 states of the Confederacy) is a liberal region of the country. Specifically, during the 51 years between 1960 and 2010, 44% of Deep-South state-years are among the 20 most liberal states in the year, while only 38% are among the 20 most conservative states.[2] With LJKE's measure, 60% of Deep-South state-years are among the 20 most liberal states in the year, while only 20% of Deep-South states are among the 20

[1] In each respect, BRFH's measure fares better.

[2] Unless stated otherwise, any empirical analysis reported in this paper is based on annual observations of the 50 states between 1960 and 2010, inclusive. We use this set of cases for most of our analyses because it is the largest pool of observations for which each of E&K's measure, LJKE's measure and BRFH's measure is available.

most conservative states.[3] Thus, LJKE's new measure paints the South as even more liberal than does E&K's measure.

- Berry et al. (2015) claim that E&K's measure shows less variation across regions than conventional wisdom dictates should be the case. This conclusion is based on a Tukey-Kramer HSD test for differences in the mean of E&K's measure across Census divisions (the results for which are reproduced in Figure S-1); this figure also shows results when the same test is applied to LJKE's measure. LJKE's measure exhibits a larger number of significant pair-wise regional differences than does E&K's measure. However, we believe that many of the regional differences lack face validity. For example, although E&K's measure reflects the conventional wisdom that policy mood is most liberal in the Northeast (i.e., the New England and Middle Atlantic Census regions), LJKE's measure indicates that policy mood in New England (CT, ME, MA, NH, RI, VT) is substantially more conservative than mood in the East South Central (AL, KY, MS, TN), and nearly equally conservative as mood in the East North Central (IL, IN, MI, OH, WI), in the South Atlantic (DE, FL, GA, MD, NC, SC, VA, WV) and in the West South Central (AR, LA, OK, TX).
- Berry et al. (2015) assert that E&K's measure lacks face validity in the sense that the time trend in E&K scores is more similar across states than seems plausible. Specifically, across all pairs of the 48 continental states between 1956 and 2010, the mean longitudinal correlation between E&K's score in one state and E&K's score in another state is 0.84.[4] For the same period, the mean correlation between LJKE's score in one state and LJKE's score in another state is substantially lower: 0.48.[5] When observing the 50 states for the period 1960-2010, the mean correlations is 0.51.[6] This quantity strikes us as more plausible. However, additional analysis presented below leads us to believe that the mean correlation of 0.51 is high enough to raise validity concerns.

LJKE's Measure Raises Similar *Construct* Validity Concerns as Does E&K's Measure

Berry et al. (2015) compute the cross-sectional correlation of each of BRFH's measure and E&K's measure with each of several indicators of state policy in each of four years spread between 1980 and 2009, and report that E&K's measure yields seven correlations (displayed in bold font) with a sign contrary to conventional wisdom. Table S-1 extends this analysis to include LJKE's measure, and shows that LJKE's measure produces thirteen correlations (also in bold) with this characteristic.

[3] LJKE's measure is computed for each state in each year between 1960 and 2020; using this longer period, 56% of Deep-South state-years are among the 20 most liberal states in a year, while 23% of Deep-South states are among the 20 most conservative states.
[4] When observing all 50 states between 1960 and 2010, the mean correlation remains 0.84.
[5] Figure S-2 shows the distribution of the correlation between mood in one state and mood in another state for both E&K's measure and LJKE's measure.
[6] For the same states and the period 1960-2020, the mean correlation is 0.46.

The Correlation Between LJKE's Measure and E&K's Measure Implies that LJKE's Claim that Both Measure are Valid is Wrong

LJKE (2023, 14) claim that E&K's measure is valid and that LJKE's measure "performs even better." Yet, LJKE (in section 5 of their "Supplementary/Online Materials") report that the mean (across the 50 states and D.C) of the over-time correlation (for the years between 1956 and 2010) between the two measures is "just above" 0.50.[7] Squaring a correlation of 0.50 yields a *coefficient of determination* (r^2) of 0.25—indicating that each measure explains just 25% of the variation in the other (Woodbridge 2006, 43). When one is considering the relationship between two variables that measure distinct concepts (e.g., a dependent variable, *Y*, and an independent variable, *X*, expected to affect *Y*), it is often reasonable to interpret an r^2 of 0.25 as "strong." However, when one is considering two variables each of which is hypothesized to be a valid measure of the same concept, we believe an r^2 of 0.25 between them is evidence that one's hypothesis is wrong—that the two variables are not measuring the same concept, and therefore at most one of them can be valid.

Results from Replications of Published Studies Suggest that LJKE's Claim that Both Their Measure and E&K's Measure are Valid is Wrong

Berry et al. (2023) report results when each of BRFH's measure and E&K's measure is substituted for the measure of state policy mood in models from seven published studies that estimate the effect of state policy mood on public policy.[8] Table S-2 extends this analysis to include LJKE's measure as well as Caughey and Warshaw's [hereafter, C&W] (2018) measure of mass economic liberalism. In this section we consider only the extent to which E&K's and LJKE's measures yield similar results. In our concluding section, we consider the implications of variation in results across all four measures.

In the majority of the seven studies—the four by Boehmke and Shipan (2015), Boehmke et al. (2015), Hawes and McCrea (2018), and Ojeda et al. (2019)—E&K's measure and LJKE's measure yield starkly different conclusions about the effect of state policy mood. Indeed, in the model replicated for each of these studies, one of the two measures (E&K's or LJKE's) has a coefficient statistically significant at the 0.05 level, and the coefficient for the other measure has the opposite sign and is not statistically significant.[9] We recognize that this finding is based on a small sample, but when the finding is considered alongside the low correlation between E&K's

[7] Our own analysis shows that the average over-time correlation is 0.49 when D.C. is excluded and the analysis is confined to the 50 states between 1960 and 2010. LJKE (2023) stress the importance for validation of analyzing *annual change* in mood measures; the average over-time correlation between *annual change* in E&K's measure and *annual change* in LJKE's measure is 0.26.

[8] Berry et al. (2023, ##) describe the process by which the sample of seven articles was chosen: the authors identified all published articles that meet a specific set of criteria, and then limited analysis to the subset of these articles for which replication data were publicly available.

[9] In the other three studies—those by Hannah and Mallinson (2018), Hayes and Dennis (2014), and Taylor et al. (2019)—the coefficient for E&K's measure has the same sign as the coefficient for LJKE's measure, and either (*i*) both coefficients are statistically significant at the 0.05 level or (*ii*) both are insignificant.

measure and LJKE's measure, we believe it justifies a conclusion that LJKE's claim that both LJKE's measure and E&K's measure are valid cannot be sustained. Either neither measure is valid, or one is valid and the other is not.

A Simulation to Place Observed Relationships in Perspective

Assume (*i*) one wants to measure state policy mood for a study, (*ii*) one accepts LJKE's claim that both measures are valid, and (*iii*) one also believes that Stimson's *national* mood measure is valid.[10] Assume further that one prefers LJKE's measure, but that one's data for this variable has been corrupted by the loss of the information that allows one to determine for any state which state is being observed; essentially, one has full over-time data for LJKE's measure for each state, but there is no way to match up any state's data with its correct state. We consider the question, *"In this hypothetical situation, which would be the best proxy for LJKE's state policy mood measure: E&K's valid measure, Stimson's valid national mood measure (used for every state), or a nonsensical 'shuffled LJKE' measure constructed by matching up each state's LJKE's mood scores with a randomly-chosen state?"* To clarify, a "shuffled LJKE" measure would have the same distribution of scores across all state-years as does LJKE's measure, but the state labels would be shuffled randomly so that, for example, the scores assigned to Texas would be equally likely to be the true LJKE scores for California, Massachusetts, Texas, or any other state.

To answer the above question, we do a simple simulation. We create 100 *shuffled* LJKE measures each involving a different random shuffling of the states. Next, for each of the 100 shuffled measures, we compute the average (across the fifty states) over-time correlation between LJKE's measure and the shuffled LJKE measure. Finally, we compute the average of these 100 average correlations—obtaining 0.52.[11] This "average of averages" of 0.52 is an estimate of the *expected* correlation between LJKE's measure and a "shuffled" measure in which the actual scores for each state are replaced by the LJKE scores for a *randomly*-chosen state. Keeping in mind that the average over-time correlation between LJKE's measure for a state and E&K's measure for the same state is 0.50, we turn to the question in italics in the previous paragraph. We can see that the average correlation between LJKE's measure and E&K's measure (0.50) is slightly smaller than the expected correlation between LJKE's measure and a randomly-shuffled LJKE measure (0.52), and slightly larger than the expected correlation between LJKE's measure and Stimson's national mood measure (0.45)—a variable that completely ignores any differences across states in how mood changes over time. This implies that both a nonsensical "shuffled" LJKE measure and Stimson's national mood measure used for each state would be almost equally as strong a proxy for the LJKE measure as the assumed-to-be-valid E&K measure.

[10] We follow LJKE and use the version of Stimson's mood measure that was updated through 2018. Although Stimson has released a new version updated through 2020, he refers to this version as "preliminary" and cautions against using it due to "dangerous assumptions" required because of changes in General Social Survey data collection during the pandemic (see https://stimson.web.unc.edu/data/).

[11] Equivalently this "average of averages" can be thought of as an average of 5,000 correlations, each of which characterizes the over-time relationship between a state's LJKE score and its score on a shuffled LJKE variable that substitutes scores for a randomly-selected state.

We also use the same simulation procedure to create and analyze 100 *shuffled* E&K measures. Here, we find that the average of the 100 average correlations is 0.84. This is an estimate of the *expected* correlation between E&K's measure and a "shuffled" measure in which the actual scores for each state are replaced by E&K scores for a *randomly*-chosen state. Thus, we find that the average correlation between E&K's measure and LJKE's measure (0.50) is much lower than the expected correlation between E&K's measure and a "shuffled" E&K measure (0.84), implying that a nonsensical "shuffled" E&K measure would be a substantially better proxy for E&K's measure than the assumed-to-be-valid LJKE measure.

We contend that these simulation results are consistent with our claim that E&K's and LJKE's measures of state policy mood lack face validity. Simply put, in our view it strains credibility to believe that both measures are valid when whichever measure we prefer, if it were not available, we would better off using a nonsensical measure created by randomly shuffling its scores across states than by using the other measure—even though this other measure is also presumed valid.

Our Responses to a Few Specific Claims Made by LJKE (2023)

LJKE (2023, 2) assert that their measure is a "direct" measure of state policy mood, as contrasted with BRFH's "indirect" measure of the concept.

Although we recognize in Berry et al. (2023) that BRFH's measure is "indirect" in several respects, and we write in Berry et al. (2015, 426) that "a measure relying on direct information about the public's issue preferences, *if demonstrated to valid and reliable* [emphasis added], would be preferable to the BRFH indicator," we do not believe that LJKE's measure should be conceived as "direct." LJKE's (2023) measurement methodology involves taking public opinion data from national surveys, and using a variety of statistical procedures (reflecting numerous stated and unstated assumptions) to *estimate* state policy mood scores—rather than observe them *directly*. As such, both LJKE's measure and BRFH's measure (as well as E&K's measure) have to make many assumptions to derive their mood scores from the data they use. The real issue is which set of assumptions is more appropriate.

LJKE (2023, 4) criticize Berry et al. (2023) by writing, "…despite their interest in over-time relationships, [Berry et al.] limit their analysis to comparing linear trends instead of evaluating year-to-year variation."

We are puzzled by this criticism since it was E&K (2015) themselves who claimed the relevance of examining the over-time trend of measures of policy mood by presenting Figure 2 in their article. E&K (2015) claim that the fact that the trend in Southern welfare benefit (decreasing) matches the trend in E&K's measure (decreasing) but not the trend in BRFH's measure (increasing) implies that E&K's measure is valid and BRFH's is not. In Berry et al. (2023), we did not present Figures 1, 2 and 3 because we believe that the variables plotted in the figures are high-quality valid measures of policy mood against which BRFH's and E&K's measures should be evaluated. The intent of these figures was merely to show that when one examines the trend in better proxies for mood than average welfare benefit (a policy output)—namely measures built from responses to GSS surveys—this trend looks like the trend in BRFH's measure (increasing) rather than the trend in E&K's measure (decreasing).

LJKE (2023, 4) write, "Not only does year-to-year variation offer a more rigorous test of relationships, but this is the concept of interest in time series analysis."
LJKE (2023) use this claim to justify examining the relationship between annual change in the "Stimson Items Index" of eleven GSS items and annual change in each of LJKE's and BRFH's measures, and ignoring the relationships that Berry et al. (2023) report among levels of the same variables. Assume we want to measure the validity of one measure (call it MEAS) of a concept, and we are willing to assume that another measure (call it TRUE) is perfectly valid. If MEAS also were perfectly valid, not only would MEAS be perfectly correlated with TRUE, but *any* transformation of MEAS would be perfectly correlated with the same transformation of TRUE. For this reason, we have no objection to analyzing change scores to assess whether a measure is valid. However, such analysis does not make analysis of levels irrelevant. Evidence that change in MEAS is highly correlated with change in TRUE does not indicate that MEAS is valid when accompanied by evidence that MEAS is weakly correlated with TRUE.

LJKE (2023, 4) contend that when observing the relationship between two measures of state policy mood, "to avoid the risk of identifying spurious relationships, scholars must remove or control for linear trends in their data."
We reject the argument that when observing the relationship between two measures of a concept, one should be concerned that the relationship observed may be spurious. To the contrary, when comparing two variables believed to measure the same concept, the simple correlation between the two variables is a highly relevant quantity of interest regardless of whether the correlation is due to a spurious relationship between the two. Indeed, in the typical measurement context, the relationship between two indicators of the same concept will be spurious—since the two indicators are related not because one causes the other, but instead because the concept affects each indicator.
LJKE (2023, 11) write that Berry et al. (2023) "critique the cross-sectional patterns of the Enns and Koch measure, and argue that including state fixed effects, which Enns and Koch recommend, is not sufficient."
We do not understand exactly what LJKE are asserting here about our position on "state fixed effects." In Berry et al. (2023), we do not question the common practice of including state fixed effects in models. We argue only (pp. 6-7) that except under conditions that seem highly unlikely in the case of E&K's measure, an indicator of state policy mood that is valid in longitudinal comparisons would also yield accurate cross-sectional comparisons.

Concluding Observations: Where Do We Stand?

Table S-2's extended replication analysis involves three indicators of state policy mood (those of E&K, LJKE, BRFH), as well as C&W's measure of mass economic liberalism, all of which measure each state over a long period of years.[12] To facilitate analysis of results in this table, we define two measures as producing a "similar inference" about the effect of state policy mood if and only if both measures yield (*i*) a coefficient for mood having the same sign, and (*ii*) the same binary result about whether the coefficient is statistically significant at the 0.05 level. Given this definition, Table S-2 yields the discouraging finding that *none* of the four measures yields a

[12] C&W (2018, 254-5) note that they seek to measure a concept "similar," but not identical, to policy mood.

similar inference to *any other* measure in a majority of the seven studies: for any pair of measures, a similar inference occurs in at most three of the seven studies.[13] This suggests that if one believes that *any* of the four measures is valid, one should be very skeptical of a claim that *any* of the other three measures is valid. This, in turn, leads us to believe that the goal of developing a measure of state policy mood that would be widely accepted as valid remains elusive.

Although LJKE (2023) report some analysis of C&W's measure, we chose not to consider this measure prior to this concluding section because doing so would take us beyond the scope of a reply to LJKE. However, when subjecting LJKE's measure to the battery of face validity tests Berry et al. (2015) use to evaluate E&K's measure, we applied the same tests to C&W's measure of mass economic liberalism. Two findings are notable: (*i*) Like E&K's and LJKE's measures, C&W's measure identifies the South as a liberal region[14], and (*ii*) the average over-time correlation between C&W's measure in one state and C&W's measure in another state is nearly as high (.77) as the average correlation of .84 between E&K's measure in one state and E&K's measure in another state.

We encourage scholars interested in state policy mood to reflect on these features of E&K's, LJKE's and C&W's measures. For example, does the fact that the three measures produce scores that deviate from the conventional wisdom that the South is conservative imply that (*i*) the measures are capturing a true feature of policy mood in the South, and the conventional wisdom about southern mood is wrong, or (*ii*) there is some shared element of the data on which the measures rely and/or some commonality in the methods underlying the measures that lead to a similar systematic overestimation of liberalness in the South? Similarly, is an average correlation of about 0.80 between a measure of mood in one state and the measure in another state (*i*) a signal of systematic measurement error, or (*ii*) an accurate reflection of the true nature of state policy mood because variation in each state's mood is largely driven by national forces?

REFERENCES

Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 1998. "Measuring Citizen and Government Ideology in the American States, 1960–93." *American Journal of Political Science* 42:327-48.

_____. 2015. "Assessing the Validity of Enns and Koch's Measure of State Policy Mood." *State Politics and Policy Quarterly* 15:425-35.

[13] We believe our crude operationalization of producing a "similar inference" actually overestimates the similarity of the results by counting two coefficients that are statistically significant and with the same sign as similar even if the point estimate for one coefficient is many times larger than the other. For example, in the Hayes and Dennis model, the E&K coefficient (6.97) is counted as similar to the C&W coefficient (0.84) despite the fact that the former is more than eight times the magnitude of the latter (when each of these mood measures has been linearly transformed to the same 0-1 range).

[14] Over the period 1960-2010, 41% of Deep-South state-years are among the 20 most liberal states in a year, while 26% of Deep-South states are among the 20 most conservative states.

Berry, William, D., Richard C Fording, Russell L. Hanson, and Justin Crofoot. 2023. "The Validity of the Enns and Koch, and Berry et al. Measures of State Policy Mood: Continuing the Debate." *State Politics and Policy Quarterly* 23:##-##.

Boehmke, Frederick J., Tracy L. Osborn and Emily U. Schilling. 2015. "Pivotal Politics and Initiative Use in the American States." *Political Research Quarterly* 68:665-77.

Boehmke, Frederick J. and Charles R. Shipan. 2015. "Oversight Capabilities in the States: Are Professionalized Legislatures Better at Getting What They Want?" *State Politics and Policy Quarterly* 15:366-86.

Caughey, Devin, and Christopher Warshaw. 2018. "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014." *American Political Science Review* 112:249-66.

Enns, Peter K., and Julianna Koch. 2013. "Public Opinion in the U.S. States: 1956-2010." *State Politics and Policy Quarterly* 13:349-72.

_____. 2015. "State Policy Mood: The Importance of Over-Time Dynamics." *State Politics and Policy Quarterly* 15:436-46.

Hannah, A. Lee, and Daniel J. Mallinson. 2018. "Defiant Innovation: The Adoption of Medical Marijuana Laws in the American States." *Policy Studies Journal* 46:402-23.

Hawes, Daniel P., and Austin Michael McCrea. 2018. "Give Us Your Tired, Your Poor and We *Might* Buy Them Dinner: Social Capital, Immigration, and Welfare Generosity in the American States." *Political Research Quarterly* 71:347-60.

Hayes, Thomas J., and Christopher Dennis. 2014. "State Adoption of Tax Policy: New Data and New Insights." *American Politics Research* 42:929-55.

Lagodny, Julius, Rebekah Jones, Julianna Koch, and Peter K. Enns. 2023. "Validation and Extension of State-Level Public Policy Mood: 1956 to 2020." *State Politics and Policy Quarterly* 23:##-##.

Ojeda, Christopher, Anne M. Whitesell, Michael B. Berkman, and Eric Plutzer. 2019. "Federalism and the Racialization of Welfare Policy." *State Politics and Policy Quarterly* 19:474-501.

Taylor, Jami K., Donald P. Haider-Markel and Benjamin Rogers. 2019. "Toward a New Measure of State-Level LGBT Interest Group Strength." *State Politics and Policy Quarterly* 19:334-50.

Woodbridge, Jeffrey M. 2006. I*ntroductory Econometrics*, 3rd edition. Mason, OH: Thompson South-Western.

Table S-1. The Correlation between Alternative Measures of State Policy Mood or Symbolic Ideology and Various Measures of State Policy

State Policy Indicator	E&K Policy Mood	BRFH Policy Mood	LJKE Policy Mood	EWM Public Opinion [i.e., symbolic ideology] (1976-88)
EWM Policy Liberalism				
1980	**-0.08**	0.82	0.03	0.83
Per Capita State Tax Collections				
1980	**-0.35**	0.31	0.07	0.38
1990	0.25	0.56	**-0.05**	0.52
2000	0.41	0.50	0.30	0.56
2008	0.28	0.44	0.23	0.39
State AFDC/TANF Benefit (Family of Four)				
1980	**-0.06**	0.66	**-0.07**	0.53
1990	0.26	0.69	**-0.22**	0.65
2000	0.36	0.64	**-0.02**	0.64
2008	0.50	0.62	0.12	0.63
Ratio of State AFDC/TANF Recipients to Number of Poor				
1980	0.27	0.72	0.23	0.76
1990	0.26	0.69	**-0.03**	0.73
2000	0.53	0.53	0.19	0.59
2008	0.30	0.60	0.10	0.61
State Medicaid Expenditures Per Beneficiary				
1980	**-0.09**	0.20	**-0.12**	0.08
1990	0.11	0.50	**-0.11**	0.34
2000	0.21	0.31	**-0.07**	0.38
2008	0.20	0.37	0.13	0.36
Ratio of State Medicaid Beneficiaries to Number of Poor				
1980	0.23	0.72	0.10	0.73
1990	0.24	0.66	**-0.03**	0.61
2000	0.59	0.41	0.29	0.42
2008	0.22	0.61	0.24	0.44
State Imprisonment Rate Per 100,000 Population				
1980	**0.02**	-0.38	**0.39**	-0.26
1990	**0.08**	-0.22	**0.31**	-0.13
2000	**0.01**	-0.27	**0.31**	-0.30
2008	-0.05	-0.34	**0.14**	-0.36

Note: When calculating correlations, state policy mood measures are observed in the same year as the state policy indicator. Each measure of policy mood or symbolic ideology is coded so that greater scores indicate greater liberalism. Each correlation with a sign contrary to our perception of conventional wisdom is printed in bold font. Data for the AFDC/TANF program, Medicaid beneficiary data, and state poverty and population data are from the University of Kentucky Center for Poverty Research (www.ukcpr.org). State tax collections are from the U.S. Census Bureau (*Annual Survey of State Government Tax Collections*, various years). State imprisonment data were obtained from the Bureau of Justice Statistics (*Prisoners in 19XX/200X*, various years).

Table S-2. The implications of the choice about which of four measures of state policy mood to use on the estimated impact of mood in seven published articles

Replicated article and model	Dependent variable[1]	Coefficient for **BRFH's** measure of policy mood (higher = more liberal) (with *p*-value in parenth)	Coefficient for **E&K's** measure of policy mood (higher = more liberal) (with *p*-value in parenth)	Coefficient for **LJKE's** measure of policy mood (higher = more liberal) (with *p*-value in parenth)	Coefficient for **C&W's** measure of mass economic liberalism (with *p*-value in parenth)
Boehmke & Shipan (2015): Model 3 in Table 2 (a pooled cross-sectional time-series negative binomial model that includes state and year fixed effects)	number of deficiencies found in nursing home inspections	0.032 (0.89)	0.358** (< 0.001)	-0.303** (0.003)	0.195 (0.29)
Boehmke et al. (2015): third model from the left in Table 3 (a longitudinal negative binomial model that includes a year count variable and its square)	number of citizen-sponsored initiatives (+)	1.14* (0.08)	-0.47 (0.42)	2.25** (0.001)	0.41 (0.55)
Hannah & Mallinson (2018): logit event history analysis model in Table 2 (that includes a year count variable)	whether a state adopts medical marijuana policy (+)	4.74* (0.07)	3.74 (0.36)	3.02 (0.28)	3.65 (0.24)
Hawes & McCrea (2018): Model 2 in Table 2 (a pooled cross-sectional time-series regression that includes state and year fixed effects)	state welfare generosity (+)	0.0082 (0.80)	-0.0722** (0.02)	0.0125 (0.55)	-0.1146** (0.002)

Table S-2, continued

Replicated article and model	Dependent variable	Coefficient for **BRFH's** measure (with *p*-value in parenth)	Coefficient for **E&K**'s measure (with *p*-value in parenth)	Coefficient for **LJKE's** measure (with *p*-value in parenth)	Coefficient for **C&W's** measure (with *p*-value in parenth)
Hayes & Dennis (2014): Cox proportional hazards model in Table 2	whether a state permits deduction of fed. income tax against state income tax (–)	-3.42** (<.001)	6.97** (<.001)	1.86** (<.001)	0.84** (.01)
Ojeda et al. (2019): random effects GLS model (that includes year fixed effects) for "Work Exemptions" in Table 5	number of TANF work exemptions	0.008 (0.91)	0.109** (0.05)	-0.052 (0.24)	-0.173** (.04)
Taylor et. al. (2019): Model 4 (Cox proportional hazards) in Table 4	whether a state adopts a sexual orientation employment discrimination law (+)	-3.47 (0.52)	2.59 (0.59)	-1.58 (0.51)	-12.87 (0.11)

** = statistically significant at the conventional threshold ($p < 0.05$) in political science research

* = not statistically significant at the conventional level, but would be significant using a slightly higher threshold ($p < 0.10$)

[1]If an article's author(s) offer an explicit hypothesis about the direction of the effect of policy mood on the dependent variable, the predicted direction (+ or –) is enclosed in parentheses after the dependent variable listed below.

Note: As do Berry et al. (2023), to facilitate comparison of results across models, we linearly transformed each of the four mood measures (BRFH, E&K, LJKE, and C&W) to the range between 0 and 1. Berry et al. (2023, ##) *incorrectly* state that the dependent variable in each study was similarly transformed; in our analysis—and in Berry et al.'s (2023) analysis—the dependent variable in each study is on the scale used by the authors of the article being replicated.

Figure S-1. Regional Variation in Average E&K and LJKE State Policy Mood Over Years Between 1960 and 2010

Mean **LJKE** State Policy Mood in **Top**-Margin Division
Relative to **Left**-Margin Division

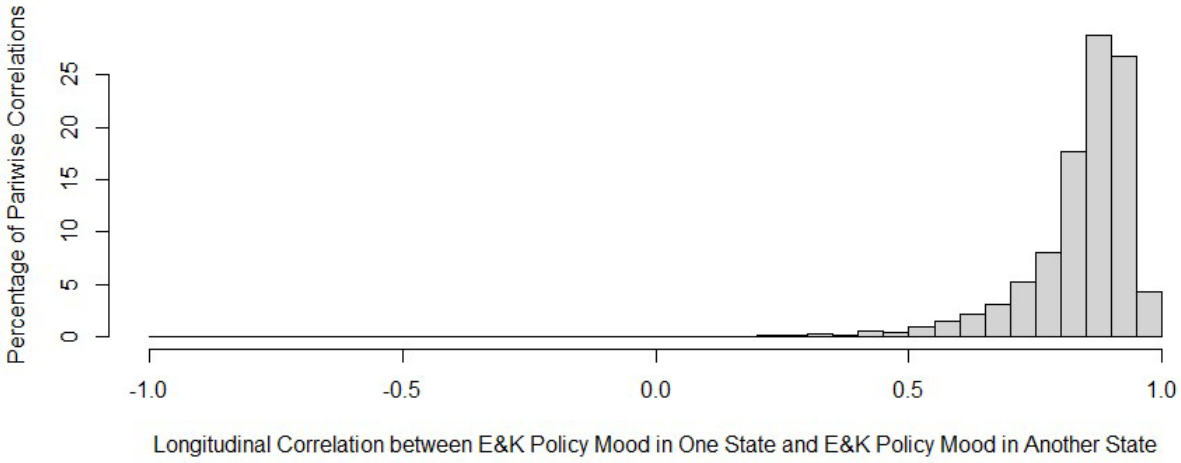
	New England	Middle Atlantic	E.N. Central	W. N. Central	S. Atlantic	E.S. Central	W. S. Central	Mountain	Pacific
New England	.455 46.2	**.038***	.003	**-.036***	.016	**.026***	-.001	**-.024***	-.005
Middle Atlantic	2.0	.493 48.2	**-.035***	**-.074***	**-.022***	-.012	**-.039***	**-.063***	**-.043***
E.N. Central	**-3.2***	**-5.2***	.458 43.0	**-.039***	.014	**.023***	-.004	**-.027***	-.007
W. N. Central	**-2.8***	**-4.8***	0.4	.420 43.4	**.052***	**.062***	**.035***	.012	**.031***
S. Atlantic	**-3.1***	**-5.1***	0.1	-0.3	.472 43.1	.010	-.017	**-.041***	**-.021***
E.S. Central	**-2.7***	**-4.7***	0.5	0.2	0.4	.481 43.5	**-.027***	**-.050***	**-.031***
W. S. Central	**-3.4***	**-5.4***	-0.2	-0.6	-0.3	-0.7	.455 42.8	**-.023***	-.004
Mountain	**-3.8***	**-5.8***	-0.6	-1.0	-0.7	-1.1	-0.4	.431 42.4	**.019***
Pacific	**-2.0***	**-4.0***	1.2	0.8	1.1	0.6	1.3	1.8	.451 44.2

Mean **E&K** State Policy Mood in **Left**-Margin Division
Relative to **Top**-Margin Division

Note: Values in the shaded region are based on Enns & Koch (2013) state policy mood scores; values in the unshaded portion of the figure are based on Lagody et al. (2023) state policy mood scores. Each value in a bold-bordered cell on the main diagonal is a mean policy mood score within a division. Each off-diagonal value is a difference in mean mood across two Census divisions; values in bold and marked with an asterisk (*) indicate that a difference in means is statistically significant at the .05 level based on a Tukey-Kramer HSD test.

Figure S-2. The Distribution (Across All Pairs of States) of the Longitudinal Correlation Between Policy Mood in One State and Policy Mood in Another State

(a) Measuring State Policy Mood with E&K's Indicator



(b) Measuring State Policy Mood with LJKE's Indicator