# The Methodological Divide of Sociology - Online Supplementary Material

Carsten Schwemmer and Oliver Wieczorek

S1. Wordfish parameter distributions

The following figure shows distributions for parameters of interest from the wordfish model. Psi is a word-fixed-effect, controlling for some terms being used more frequently than others. Beta is the estimated weight for each term used to position the documents on the unidimensional scale. Theta is the estimate for each document (abstract) on the unidimensional scale.
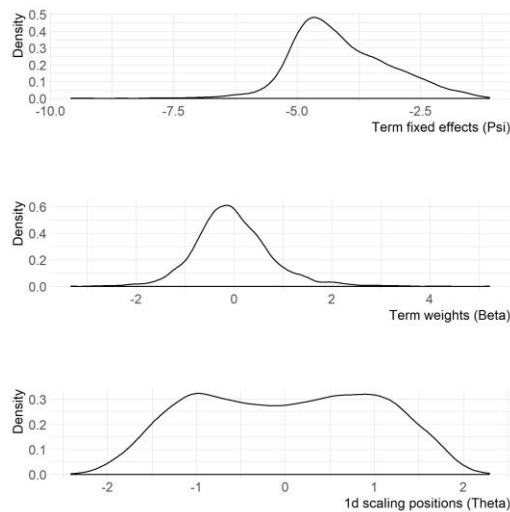
S2 Abstract examples

This section includes additional examples for abstracts with very low / high scaling estimates (θ). Abstracts with low scaling estimates include more (orange) words from

the left-hand side of Figure 3 in the main article, while abstracts with high scaling estimates include more (green) words from the right-hand side.

Abstract with low scaling estimate (qualitative):

"This article offers a critique of framing perspectives on collective action discourse and an alternative dialogic approach. The argument set forth is that the latter sees collective action discourse as a joint product of actors' agency and discourse dynamics, including its multivocal nature. Such discourse is a joint product of challengers' rational actions and the constraints of the discursive field. Challengers seek to appropriate and subvert the dominant discourses that legitimate power, creating discursive repertoires. To illustrate this, the contentious actions of English cotton spinners in the 1820s and 1830s are analyzed. The spinners produced a discursive repertoire drawing on mill owners' dominant discourses."

Steinberg, M. W. (1999). The Talk and Back Talk of Collective Action: A Dialogic Analysis of Repertoires of Discourse among Nineteenth-Century English Cotton Spinners. American Journal of Sociology, 105(3):736–780.

Abstract with high scaling estimate (quantitative)

"There has been minimal research on the pre-school enrollment of immigrant children. Using 1990 US Census data, this paper investigates pre-school enrollment of child immigrants, those who immigrated as children and the US-born children of immigrants. The analysis is conducted using probit analysis. Preschool enrollment is found to vary systematically with parental characteristics (income and education), immigrant generation, number of siblings, mother's labor supply, and country of origin. Among the foreign-born, differences in pre-school enrollment are analyzed by country of origin. Among the US born children of immigrants pre-school enrollment is greatest among those with both parents foreign born"

Chiswick, B. R., & DebBurman, N. (2006). Pre-school enrollment: An analysis by immigrant generation. Social Science Research, 35(1), 60–87.

S3 Regressions for scaling estimates

The following regression table includes OLS regression models for scaling positions of all abstracts in our sample. Corresponding journals are included in the model as

dummy variables with the *American Journal of Sociology* as reference category. Regarding the publication year, the first model includes a linear term, while the second and third model include polynomial splines with three and five degrees of freedom. As indicated by the RMSE and Adjusted R^2 values the flexible terms for publication years of abstracts do not substantially improve model fit. Figure 3 in the main article therefore depicts effect estimates from the linear model.

	Linear Model	Spline Model 1	Spline Model 2
(Intercept)	-11.29 (2.81)***	-0.30 (0.05)***	-0.26 (0.06)***
Acta Sociologica	0.07 (0.06)	0.07 (0.06)	0.06 (0.06)
American Sociological Review	0.28 (0.04)***	0.28 (0.04)***	0.28 (0.04)***
Annual Review of Sociology	-0.33 (0.06)***	-0.33 (0.06)***	-0.32 (0.06)***
British Journal of Sociology	-0.55 (0.05)***	-0.55 (0.05)***	-0.55 (0.05)***
European Sociological Review	0.85 (0.05)***	0.85 (0.05)***	0.85 (0.05)***
Social Forces	0.58 (0.04)***	0.58 (0.04)***	0.57 (0.04)***
Social Problems	-0.04 (0.05)	-0.04 (0.05)	-0.04 (0.05)
Social Science Research	0.93 (0.04)***	0.93 (0.04)***	0.93 (0.04)***
Sociological Forum	-0.21 (0.05)***	-0.20 (0.05)***	-0.20 (0.05)***
Sociological Quarterly	-0.13 (0.05)**	-0.13 (0.05)**	-0.13 (0.05)**
Sociological Science	0.42 (0.08)***	0.45 (0.08)***	0.45 (0.08)***
Sociology	-0.57 (0.04)***	-0.57 (0.04)***	-0.57 (0.04)***
Year	0.01 (0.00)***		
Spline Year: 1		0.07 (0.11)	-0.10 (0.20)
Spline Year: 2		0.26 (0.06)***	0.16 (0.27)
Spline Year: 3		0.09 (0.05)	0.22 (0.28)
Spline Year: 4			0.08 (0.12)
Spline Year: 5			0.08 (0.06)
Number of Observations	8737	8737	8737
R²	0.28	0.29	0.29
Adjusted R²	0.28	0.28	0.28
RMSE	0.85	0.85	0.85

Standard errors in parenthesis; *** p <0.001; ** p<0.01; * p<0.05

S4 Correspondence analysis

To examine whether our results are robust to another methodological approach, we used the same feature space - counts for 2.125 different terms within our 8.737 abstracts - and conducted a correspondence analysis. This allows us to project our textual data to more than one single dimension, e.g. to two dimensions as shown in the following figures. The results are very similar in comparison to our wordfish model.

   The following figure shows term level position from the correspondence analysis, with the same terms highlighted as in our main paper. It is obvious that the first dimension of the correspondence analysis represents the same entangled methodological divide as the wordfish model.

The second dimension of the correspondence analysis is harder to interpret. On the top end of second dimension we find terms that are often used in studies about gender, family, occupation and parenthood. On the bottom end, terms are related to ethnicity, segregation and protests become apparent. Data points on these ends correspond to empirical work on these issues, e.g. "It's about time and gender: spousal employment and health" at the top and "Hispanic Segregation in metropolitan America: Exploring the Multiple Forms of Spatial Assimilation" at the bottom. Ultimately, we retrain from assigning a clear label for the second dimension.
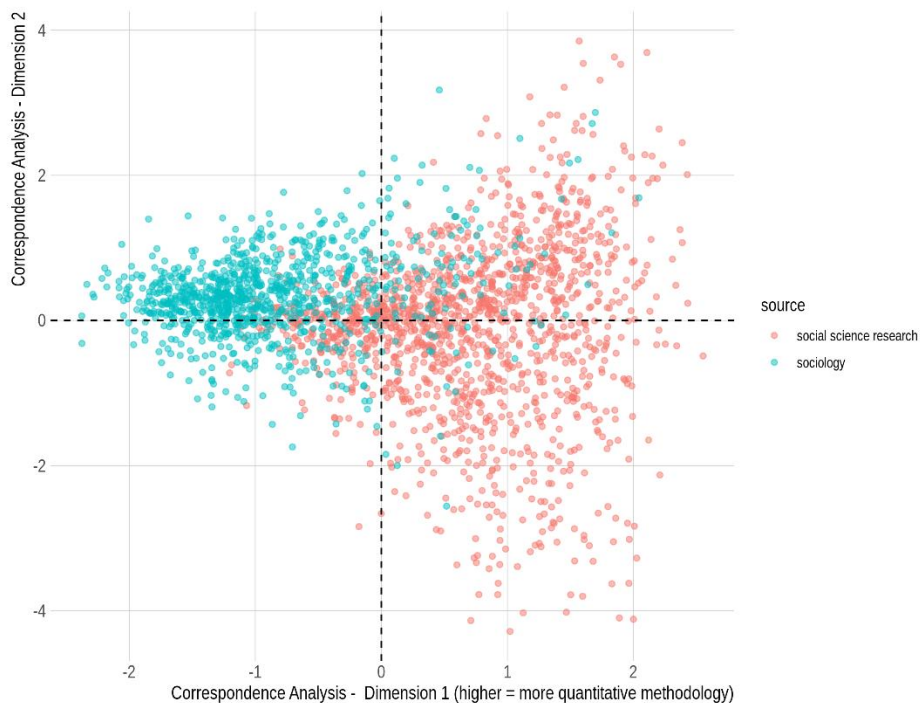

Differences for the first dimension across journal outlets are also robust to using a different measurement approach. The next figure shows document level positions from our correspondence analysis for two exemplary journals. These journals were identified as extremes in our main paper. *Social Science Research* predominantly publishes quantitative research and *Sociology* focuses on qualitative research, which is illustrated in the figure.

S5 Author Affiliations

Nationally embedded scholarly cultures have their own distinctions between different paradigms and over definitions of what counts as qualitative or quantitative research. We therefore examined whether our findings for journal and time effects on methodological preferences are robust when controlling for the country of each author's affiliation. Unfortunately, the data quality of affiliations is problematic: institutions are often spelled in a variety of ways and in many cases the corresponding country is not included. Moreover, for 248 articles in our sample, affiliation data is missing completely.

We therefore relied on an automated correction method for examining all publications, for which at least some affiliation data is available. We passed the corresponding affiliations on to the Google Geocoding Application Programming Interface (API), which infers countries based upon institution names or addresses even if the affiliation data is incomplete. While this approach is far from perfect, qualitative assessment of the country labels showed that it is sufficient for robustness checks of our main findings. The following map visualizes affiliation counts for our dataset by country:
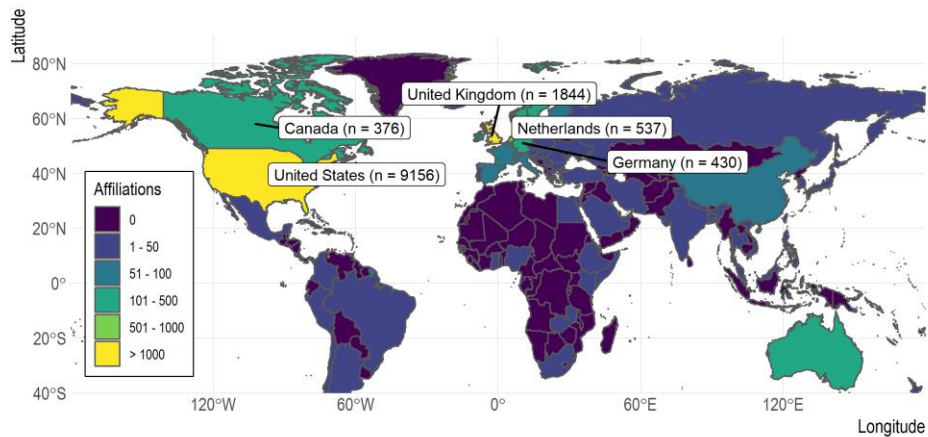
Colours indicate categories for affiliation counts, and the five countries with the most affiliations are depicted with labels. Overall, more than 75% of author affiliations are located either in the United Kingdom or the United States. After merging all geocoded affiliations with our main data, we created relative indicators for US, UK and international (neither US nor UK based) affiliations. To give one example: if an article has 2 US-based authors, 1 UK-based author and 1 international/European author, the indicators would be 0.5, 0.25 and 0.25. We then ran additional regression models for our scaling estimates, where we included these indicators together with time and journal effects. The following table shows results from a baseline model without affiliation covariates in comparison to a model including affiliation data. The indicator for international affiliations is omitted in order to avoid collinearity issues.

	Baseline Model	Affiliation Model
(Intercept)	-10.47 (2.92)***	-10.55 (2.94)***
Acta Sociologica	0.06 (0.06)	0.08 (0.07)
American Sociological Review	0.28 (0.04)***	0.28 (0.04)***
Annual Review of Sociology	-0.34 (0.06)***	-0.34 (0.06)***
British Journal of Sociology	-0.55 (0.05)***	-0.48 (0.06)***
European Sociological Review	0.86 (0.05)***	0.88 (0.05)***
Social Forces	0.57 (0.04)***	0.57 (0.04)***
Social Problems	-0.04 (0.05)	-0.05 (0.05)
Social Science Research	0.93 (0.04)***	0.94 (0.04)***
Sociological Forum	-0.21 (0.05)***	-0.21 (0.05)***
Sociological Quarterly	-0.14 (0.05)**	-0.14 (0.05)**
Sociological Science	0.42 (0.08)***	0.42 (0.08)***
Sociology	-0.57 (0.04)***	-0.48 (0.05)***
Year	0.01 (0.00)***	0.01 (0.00)***
Share of UK Authors		-0.12 (0.04)**
Share of US Authors		0.02 (0.03)
Number of observations	8489	8489
R^2	0.28	0.28
Adjusted R^2	0.28	0.28
RMSE	0.85	0.85

Standard errors in parenthesis; *** p <0.001; ** p<0.01; * p<0.05

Results from these models suggest that the origins of authors at best have a weak effect on methodological preferences. UK-based authors are slightly more likely to produce qualitative work, while no difference is apparent for US-based authors. The following figures further illustrate these effects, which are visualized while holding all other covariates at their empirically observed values.

(a)

(b)
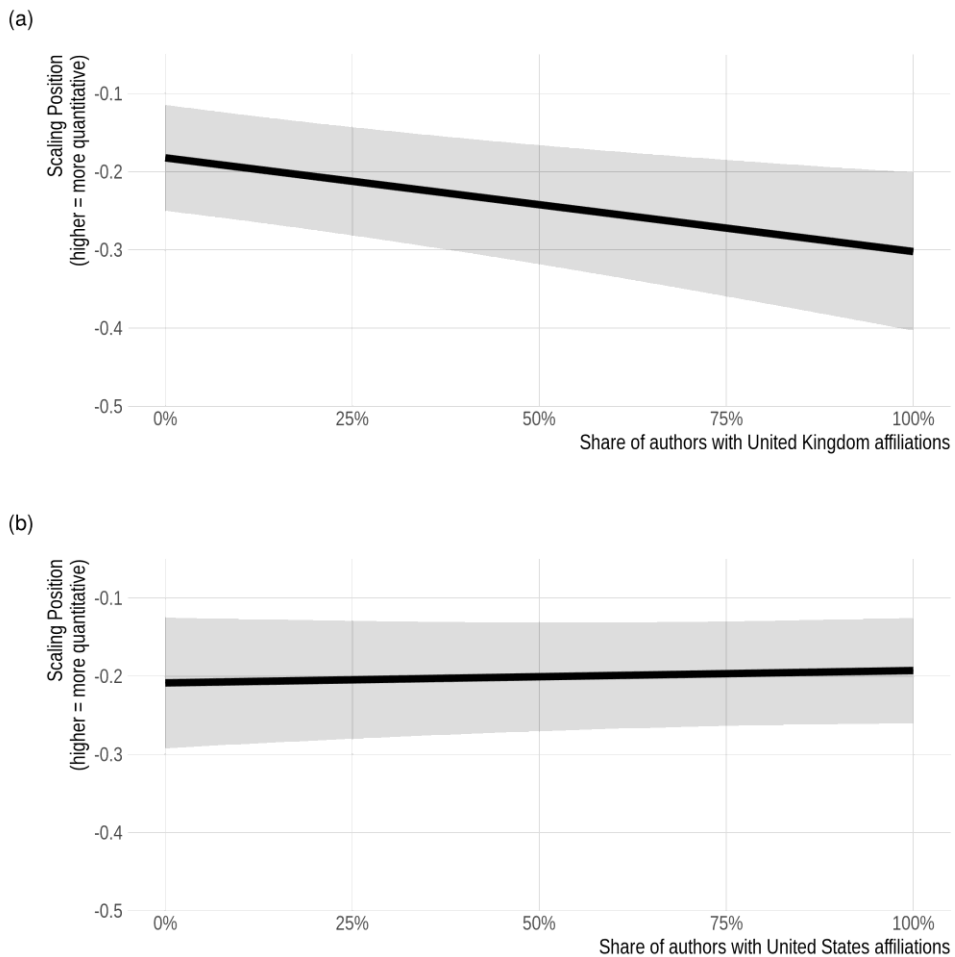
Our main findings for journal effects and time trends do not change after controlling for author affiliations.

S6 Methodology and coauthored work

In our main paper, we discuss literature stating that quantitative work is more likely to be coauthored than non-quantitative work. In this section, we examine whether this also applies to our data on Sociology publications.

A boxplot shows the comparison between single author and coauthored publications for our methodology scaling estimate:

The figure suggests that coauthored work is indeed more often positioned on the more quantitative extreme of our methodology scaling.

The same holds when treating the number of authors as a continuous variable and conducting a Pearson correlation test ($R = 0.22$, $p < 0.001$). This can also be visualised with a jittered scatter plot, where the blue line indicates a positive linear trend between quantitative methodology and the number of authors:

Publications with > 7 authors excluded for visualisation purposes