The Stewart Retractions: A Quantitative and Qualitative Analysis

Justin T. Pickett[1]

LINK TO ABSTRACT

This study analyzes the recent retraction of five articles from three sociology journals—*Social Problems*, *Criminology*, and *Law & Society Review*. Analyzing the retractions is important for several reasons:

- The retraction notices are vague, providing little information about what went wrong.
- The authors have continued to promote their retracted findings in print, insisting that "the main substantive results are correct" (*Law & Society Review* 2020).
- Other articles by the authors have some of the same irregularities (e.g., Mears et al. 2013; Mears et al. 2017; Stewart and Simons 2010; Stewart et al. 2006; Stewart et al. 2009), but thus far only one of these has been corrected and none have been retracted.
- Examining how coauthors and journal editors respond to learning about irregularities in articles sheds light on the sociology of science.

The retracted articles' titles and abstracts are provided later; the authors of each article are as follows:

1. Johnson, Brian D., Eric A. Stewart, Justin Pickett, and Marc Gertz, 2011.

---

1. State University of New York at Albany, Albany, NY 12222.

2. Stewart, Eric A., Ramiro Martinez, Jr., Eric P. Baumer, and Marc Gertz, 2015.
3. Stewart, Eric A., Daniel P. Mears, Patricia Y. Warren, Eric P. Baumer, and Ashley N. Arnio, 2018.
4. Stewart, Eric A., Brian D. Johnson, Patricia Y. Warren, Jordyn L. Rosario, and Cresean Hughes, 2019.
5. Mears, Daniel P., Eric A. Stewart, Patricia Y. Warren, Miltonette O. Craig, and Ashley N. Arnio, 2019.

The only coauthor on all five retracted articles was Dr. Eric Stewart. He was the data holder and analyst for each article. I coauthored one of the retracted articles, Johnson et al. (2011), but here I analyze all five. I organize my analysis of the quantitative and qualitative data into three sections: (1) what happened in the articles, (2) what happened among the coauthors, and (3) what happened at the journals. Everything—data, code, emails, text messages, Excel files, drafts, and university documents—needed to verify my claims is provided online (**link**).

The correspondence discussed herein is public record under law, per the Freedom of Information Act. The present study received Institutional Review Board approval from the author's university. The timeline is as follows. In February 2019, Drs. Nick Brown and James Heathers first emailed Dr. Mears to raise concerns about one of the articles. In May 2019, someone calling himself "John Smith" sent a list of irregularities in the five articles to the coauthors, journal editors, and administrators at Florida State University (FSU) (**link**). That email instigated a misconduct inquiry at FSU (see Pickett 2020). The retractions were announced in November 2019.

What happened in the five articles

Consistently incorrect means and standard deviations

The means and standard deviations for binary variables are connected mathematically (Heathers and Brown 2019; Schumm et al. 2019). For any specific sample size N, a binary variable with a given mean P—the proportion of the sample coded "1"—will have a standard deviation equal to:

$$SD = \sqrt{\frac{N}{N-1} \times P(1-P)}$$

In Johnson et al. (2011), the standard deviations are wrong for nine binary

variables, given the listed means. As Table 1 shows, six of these discrepancies are very large, too large to have resulted from rounding. Although these discrepancies could be due to error, it would not be a one-time error. All of the retracted articles have impossible standard deviations: there are discrepancies for six binary variables in Stewart et al. (2015), and five of them are large. In Stewart, Johnson et al. (2019), there are nine discrepancies, and six are large. In Stewart et al. (2018), there are four discrepancies and all four are large. Those are just the retracted articles. Impossible standard deviations also appear in many of Dr. Stewart's other published articles. There are six in Mears et al. (2013, 706), three in Stewart and Simons (2010, 604), three in Stewart et al. (2009, 865), one in Stewart (2003, 591), and one in Stewart et al. (2006, 17). There were five in Mears et al. (2017, 230), and they were corrected in an erratum.

TABLE 1. Johnson et al. (2011): Nine discrepancies in means and standard deviations (large discrepancies are in boxes)

Variables	Published article		
	Mean	SD	Correct SD
White	.86	.41	.35
Black	.10	.33	.30
Hispanic	.04	.22	.20
Married	.59	.31	.49
Education level (college graduate)	.42	.31	.49
Political conservative	.43	.31	.50
Own home	.78	.33	.41
Southwest	.17	.41	.38
South	.44	.39	.50

Notes: Large differences are those exceeding five points.

Dr. Stewart and his coauthors have offered different explanations for the discrepancies, none of which are credible. Dr. Mears told an editor they included "incorrect standard deviations for binary measures" because "the ones that we presented were mistakenly based on the formula for continuous measures."[2] But that is obviously untrue, for two reasons. First, there is not a separate formula for calculating standard deviations for continuous measures. Second, even if there was, modern statistical programs would not apply the wrong formula. In an email, Dr. Johnson wrote: "the standard deviations were wrong in some cases because they were based on categorical rather than binary measures (e.g., gender coded male, female, unknown)."[3] That explanation also cannot be true. If it was, the standard

2. Email from Dr. Mears to Dr. Sterett, May 30, 2019.
3. Email from Dr. Johnson to me, October 31, 2019.

deviations would have been identical for the different racial groups (Whites, Blacks, and Hispanics) and regions (East, West, Northwest, and South). They were not.

There are other conceivable explanations for such discrepancies (e.g., misreporting sample size, unreported imputation). One possibility is that "the descriptive statistics have simply been fabricated" (Heathers and Brown 2019, 7). Fabrication can lead to such discrepancies when researchers are either unaware or forget that binary variables' means and standard deviations are connected (Schumm et al. 2019).

Non-uniform terminal-digit distributions

One method for identifying fabricated numbers is to test whether the distribution of terminal (or rightmost) digits in reported statistics differs significantly from uniform (Diekmann 2007; Mosimann et al. 2002). The U.S. Office of Research Integrity has used this method to identify several cases of scientific fraud (Mosimann et al. 1995; 2002). It is based on Benford's law, which describes the logarithmic distribution of the first significant (nonzero) digit, with the implication "that the distribution of higher-order digits increasingly approximates the uniform distribution" (Diekmann 2007, 323). For example, the probability of observing a specific number (0–9) in the second digit (d_2) is given by the formula:

$$Prob(D_2 = d_2) = \sum_{d_1 = 1}^{9} \log\left(1 + \frac{1}{d_1 d_2}\right); \qquad d_2 \in \{0, \ 1, \ \ldots, \ 9\}.$$

Working through this formula reveals the second digit's expected distribution: slightly more zeros (12 percent) than nines (9 percent), with the percentages of the other numbers (1–8) falling in between. By the third digit, each number (0–9) should appear roughly 10 percent of the time. Specifically, the expected distribution is: 0 = 10.18 percent, 1 = 10.14 percent, 2 = 10.10 percent, 3 = 10.06 percent, 4 = 10.02 percent, 5 = 9.98 percent, 6 = 9.94 percent, 7 = 9.90 percent, 8 = 9.86 percent, and 9 = 9.83 percent (Nigrini 2012).

Regression coefficients and standard errors should be Benford-distributed (Diekmann 2007; Günnel and Tödter 2009). For three-decimal regression coefficients and standard errors, the distribution of terminal digits should be approximately uniform, especially with rounding from the unreported fourth digit.[4] This means that approximately 10 percent of reported terminal digits should be zeros.[5]

---

4. The exception being coefficients and standard errors with less than two significant digits (e.g., b = .000 or .001).

5. The terminal-digit distributions for coefficients and standard errors with only two significant digits (e.g., b = .020) should have slightly more zeros (10–12 percent, with rounding).

Fabricators, however, have difficultly generating the expected distributions for all but the first digit (Diekmann 2007). Mostly, this reflects their inability to create uniform distributions. For example, they tend to avoid ending numbers with zero, which results in terminal-digit distributions lacking the expected number of zeros (Mosimann et al. 1995).

In our article, Johnson et al. (2011), less than 2 percent of the regression coefficients and standard errors end with zero, and the terminal-digit distribution differs significantly from uniform ($\chi^2 = 26.18$, p = .002). If the true underlying distribution is uniform, we would expect to see such an extreme sample distribution by chance roughly 1 in 500 times. The numbers in the second article using the 2008 data are just as improbable. Less than 2 percent of the regression coefficients and standard errors in Stewart et al. (2015) end with zero, and the terminal-digit distribution differs significantly from uniform ($\chi^2 = 31.22$, p < .001).[6] In each of Dr. Stewart's more recent articles (Mears et al. 2019; Stewart et al. 2018; Stewart, Johnson et al. 2019), 2 percent or less of the coefficients and standard errors end with zero, and the terminal-digit distribution differs significantly from uniform ($\chi^2 = 61.00$, p < .001; $\chi^2 = 113.20$, p < .001; $\chi^2 = 43.70$, p < .001).[7] In fact, in Stewart, Johnson et al. (2019), none of the coefficients or standard errors end with zero.

What does this mean? To answer this question we need to know whether the distribution of terminal digits "from a sample of non-manipulated articles is in accordance with Benford's law" (Diekmann and Jann 2010, 398). Accordingly, I searched for articles that examined criminal justice topics and used similar methodologies—specifically, that had comparably large (or larger) samples, used multilevel modeling, estimated a series of stepwise models building from a baseline specification, and reported coefficients and standard errors to three decimals. To increase the likelihood that the comparison articles were "non-manipulated," I only included articles that used data available to outside researchers. I also excluded articles that involved Dr. Stewart or his coauthors on the five questionable articles. I used the first ten articles I found that met these criteria.[8]

Dr. Stewart's five articles reported a total of 1,582 coefficients and standard errors. To generate comparison groups with a similar number of articles and statistics, I block-randomized (on the basis of the number of reported statistics)

6. Odds ratios and *t*-statistics are excluded because they are calculated from the coefficients and standard errors.

7. After Dr. Stewart was notified of the unusual terminal-digit distributions in these articles, he corrected two, one between 'online first' and print publication (Mears et al. 2019) and the other after print (Stewart et al. 2018). The numbers I report are for the original articles. The 'corrected' articles also have non-uniform distributions, though.

8. The articles were Brauer 2009; Hagan, Shedd, and Payne 2005; Kirk 2008; Kirk and Matsuda 2011; Maimon and Browning 2010; Sampson, Morenoff, and Raudenbush 2005; Savolainen et al. 2017; Slocum et al. 2010; Wilcox, Madensen, and Tillyer 2007; and Xie, Lauritsen, and Heimer 2012.

the ten articles I found into two comparison groups of five articles each. The first comparison group reported 1,332 coefficients and standard errors; the second comparison group reported 1,232. Figure 1 shows the terminal-digit distributions for Dr. Stewart's five questionable articles and both comparison groups.

Less than 1 percent of the terminal digits in Dr. Stewart's articles are zeros, whereas 10 percent are in the first comparison group, and 9.4 percent are in the second. Unsurprisingly, panel A in Figure 1 shows that the terminal-digit distribution in Dr. Stewart's articles differs significantly from uniform ($\chi^2 = 252.07$, $p < .001$). 'Differs significantly' is an understatement; the p-value is incredibly small ($p = 3.65 \times 10^{-49}$). By contrast, in neither comparison group does the terminal digit distribution differ significantly from uniform (see panels C and E).

One possible explanation is that the standard errors in Dr. Stewart's articles are far more stable (across models) than in the comparison articles, which means that the same terminal digits are being counted multiple times. To examine this possibility, I restricted the analysis to terminal digits in full models, defined as the most complete specification in a given model set estimated with a specific sample and outcome variable. The full-model-only, terminal-digit distributions are in panels B, D, and F. The conclusion is the same. Neither comparison group has a terminal-digit distribution that differs significantly from uniform. By contrast, the distribution for Dr. Stewart's articles is extremely non-uniform ($\chi^2 = 87.73$, $p < .001$). If the true underlying distribution is uniform, we would expect a sample distribution as extreme as reported in Dr. Stewart's articles by chance about 1 in two hundred trillion times ($p = 4.64 \times 10^{-15}$).

Here is Dr. Stewart's explanation for the low frequency of terminal-digit zeros in his articles: "Although there generally weren't a lot of zeros in the 3rd decimal place, I round 3rd place zeros either up or down. For example, if a coefficient was .000007, I would round the value to .007 or .007x10-3."[9] (None of the articles reported rounding or scientific notation.) Unfortunately, this explanation cannot withstand empirical scrutiny. It applies only to statistics with two leading zeros (e.g., $b = .007$), yet even if we exclude all such statistics, the terminal digit distribution in Dr. Stewart's articles is extremely non-uniform ($\chi^2 = 242.22$, $p < .001$). Additionally, even if Dr. Stewart rounded 3rd place zeros, the distribution of non-zero (1–9) terminal digits should still be approximately uniform, but it is not ($\chi^2 = 98.17$, $p < .001$). By contrast, in the two comparison groups, the distribution of non-zero terminal digits is approximately uniform (group 1: $\chi^2 = 12.10$, $p = .147$; group 2: $\chi^2 = 5.85$, $p = .663$). Regardless of how we restrict the analysis, then, the statistics in Dr. Stewart's articles stand out for their improbability, given real data.

---

9. Memo from Dr. Stewart to Dr. Thomas Blomberg, May 28, 2019.

Figure 1. Third decimals in five Stewart articles and in two comparison groups, each including five articles by other authors
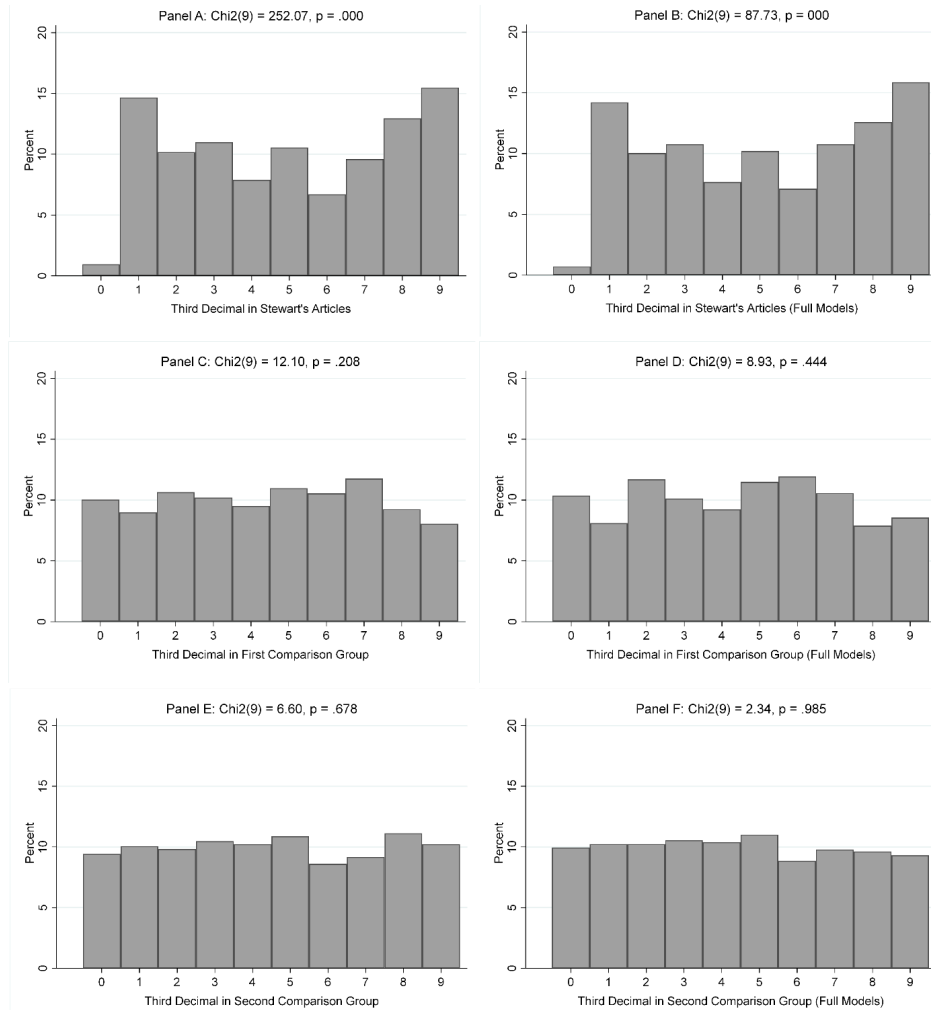
Notes: The Figure shows the percentage of third decimals in each numerical category for the regression coefficients and standard errors in five articles for which Dr. Stewart did the analysis versus two comparison groups of five articles each. The comparison articles did not involve Dr. Stewart, but all had similar methodologies (used multilevel modeling with large samples, and estimated stepwise models building from a baseline equation).

Unverifiable surveys

Three of the retracted articles reported using data from a large ($N = 2,736$), nationally representative, dual frame (landlines and cellphones) telephone survey conducted in 2013, with a 60.8 percent response rate (Mears et al. 2019; Stewart

et al. 2018; Stewart, Johnson et al. 2019). The response rate is surprising, because it greatly exceeds the typical rate of less than 10 percent obtained by professional polling organizations (Keeter et al. 2017). It is also surprising that none of the articles listed a funding source, because a survey of this size should cost in excess of $100,000 (Guterbock et al. 2018).

Perhaps most surprising, however, is that none of the articles named the survey organization that conducted the 2013 survey. When asked via email in 2018 about the survey organization, Dr. Stewart wrote:

> I teamed up with some of my grad school buddies on their survey. I helped them with some analysis on a few of their projects. In turn, they agreed to add some questions for me on their broader telephone survey. They used their students to make the calls because many of the vendors were charging so much. They did a fairly good job for their first and only telephone survey. Have you ever considered training and using undergraduate students and renting time in a CATI lab at Albany? I remember Gertz did this some when he had his research outfit at FSU.[10]

According to Dr. Stewart, then, the 2013 survey was done by his friends, using their students as interviewers, and it was their first survey. The last sentence is also important. It mentions Dr. Marc Gertz, a former professor at FSU, and explains that he did similar surveys *before* he closed his polling firm, The Research Network (TRN).

Later, Dr. Stewart changed his story. In July 2019, he told FSU's Inquiry Committee that it was Dr. Gertz and TRN staff who conducted the 2013 survey, but wrote: "I do not have the correspondence from Dr. Gertz and/or the Research Network staff providing the 2013 data files because they were given to me on a jump drive."[11] There are three problems with Dr. Stewart's new explanation. First, TRN closed in 2010. Second, when asked if he conducted the 2013 survey, Dr. Gertz wrote: "Not me, wish it were."[12] Third, the former TRN director, Jake Bratton, wrote that he never provided Dr. Stewart data for any survey conducted after 2009.

Two of the retracted articles reported using data from a large ($N = 1,184$ to 1,379), nationally representative, dual frame telephone survey conducted in 2008

---

10. Email from Dr. Stewart to me, March 27, 2018.
11. Inquiry documents from Dr. Stewart to Inquiry Committee, July 2019. Dr. Stewart did not provide the committee any documentation or data from the 2013 survey. Instead, he provided a copy of a 2017 email from Mr. Bratton, which showed a data attachment. The data attachment was for surveys conducted by other researchers on various topics between 2000 and 2009. In the email, Mr. Bratton told Dr. Stewart, "not all of these have anything to do with racial typicification [*sic*], so you'll have to sort through them."
12. Email from Dr. Gertz to me, March 27, 2018.

by TRN, with a 54.8 percent response rate (Johnson et al. 2011; Stewart et al. 2015). The articles each reported one survey, but Dr. Stewart told his coauthors and FSU's Inquiry Committee that TRN ran multiple surveys for him in 2008 and he combined them. To support this claim, he provided copies of two emails he received from Mr. Bratton in 2008 that each showed data attachments. He did not provide the raw data to the Committee. However, Mr. Bratton disputed Dr. Stewart's account repeatedly in writing. In one email, he wrote:

> That survey in the article and those questions are N = 500. The second file sent per my email you cite was a match file of census data to merge on respondent ID based on self-report zipcode, none of the original data was included. I have no record or recollection of asking that dependent variable in a following survey and TRN was closed in 1Q 2010.[13]

Identical statistics after changes in…everything else

TRN finished the survey for Johnson et al. (2011) in January 2008. When my coauthors and I presented our findings over a year later, in November 2009, we reported 868 respondents. In late 2010, when we were putting the final touches on our manuscript before submitting it, we still reported 868 respondents. However, the sample size reported in our published article is 1,184. There are two problems with this. First, the source of the 316 new respondents is a mystery. "I didn't notice differences from earlier to later versions of the paper in terms of sample sizes," Dr. Johnson wrote after I pointed out the mysterious new respondents in 2019.[14]

Second, the addition of 316 new respondents had almost no effect on any of the reported statistics—means, standard deviations, regression coefficients, or standard errors. Approximately 90 percent of the statistics reported in the presentation, manuscript draft, and published article are identical to the third decimal place. To illustrate, Table 2 presents the regression results from the first three models in the manuscript draft and published article. Numbers that change are in boxes, those that do not are unboxed. Although the article has 316 more respondents than the draft, and includes an additional county-level variable (Percent Republican), almost all of the coefficients and standard errors are identical.

Years later, the same problem—sample size growth without other chan-ges—happened again. Stewart et al. (2015) analyzed data from the same 2008 sur-vey we used in Johnson et al. (2011), and reported the same 54.8 percent response rate, the same 96 percent completion rate, the same 10 percent verification rate

---

13. Email from Mr. Bratton to me, November 11, 2019.
14. Email from Dr. Johnson to me, June 6, 2019.

**TABLE 2. Johnson et al. (2011) manuscript draft vs. published article:
Mostly identical statistics (numbers that change are in boxes)**

	Manuscript draft						Published article					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
Variables	b	SE	b	SE	b	SE	b	SE	b	SE	b	SE
Criminal threat	-	-	-	-	.229*	.084	-	-	-	-	.232*	.084
Econ. threat	-	-	-	-	.327*	.111	-	-	-	-	.328*	.111
Political threat	-	-	-	-	.056	.107	-	-	-	-	.053	.107
Pct. Hispanic	-	-	−.104	.662	−.083	.669	-	-	−.104	.662	−.083	.669
Hispanic gr'th	-	-	.288*	.102	.267*	.101	-	-	.288*	.102	.267*	.101
Homicide rate	−.016	.033	−.013	.033	−.009	.033	−.016	.033	−.013	.033	−.009	.033
Concentr'd dis.	.052	.086	.051	.086	.047	.086	.052	.086	.051	.086	.047	.086
Percent Repub.	-	-	-	-	-	-	.001	.005	.001	.005	.001	.005
Percent Black	−.043	.173	−.038	.173	−.036	.173	−.043	.173	−.038	.173	−.036	.173
Pop. structure	−.053	.173	−.051	.174	−.056	.185	−.053	.173	−.051	.174	−.056	.185
Black	−.639*	.248	−.628*	.248	−.597*	.248	−.639*	.248	−.628*	.248	−.597*	.248
Hispanic	−.997*	.359	−.982*	.359	−.959*	.359	−.997*	.359	−.982*	.359	−.959*	.359
Age	.016*	.005	.016*	.005	.016*	.005	.016*	.005	.016*	.005	.016*	.005
Male	.148*	.056	.148*	.056	.148*	.056	.164*	.056	.164*	.056	.164*	.056
Married	.208	.201	.208	.201	.208	.201	.208	.201	.208	.201	.208	.201
Education level	−.099	.198	−.099	.198	−.099	.198	−.099	.198	−.099	.198	−.099	.198
Family income	−.026	.061	−.026	.061	−.026	.061	−.026	.061	−.026	.061	−.026	.061
Employed	−.191*	.082	−.191*	.082	−.188*	.082	−.191*	.082	−.191*	.082	−.188*	.082
Political con.	.374*	.146	.367*	.146	.355*	.146	.374*	.146	.367*	.146	.355*	.146
Own home	−.137	.246	−.137	.246	−.101	.246	−.137	.246	−.137	.246	−.101	.246
Southwest	−.111	.274	−.111	.274	−.111	.274	−.111	.274	−.111	.274	−.111	.274
Northeast	−.183	.281	−.183	.281	−.183	.281	−.183	.281	−.183	.281	−.183	.281
Midwest	.054	.230	.054	.230	.054	.230	.054	.230	.054	.230	.054	.230
West	.082	.264	.082	.264	.082	.264	.082	.264	.082	.264	.082	.264
Gen'l punitive	.191*	.062	.191*	.062	.187*	.062	.191*	.062	.191*	.062	.187*	.062
Intercept	−.832*	.102	−.858*	.105	−.867*	.119	−.832*	.102	−.858*	.105	−.867*	.119
Variance Explained	10%		15%		28%		10%		15%		28%	
N	868						1,184					

Notes: Numbers that change between manuscript draft and published article are in boxes. *p < .05 (two-tailed).

("supervisors reviewed 10 percent of completed interviews for accuracy"), and the same 98 percent agreement rate ("between supervisors and interviewers"). However, whereas Johnson et al. (2011) reported 1,184 respondents, Stewart et al. (2015) reported 1,379 respondents. The 2008 sample thus grew by a total of 511 respondents after the survey finished—from 868 to 1,184 respondents (between

2008 and 2011), and then to 1,379 respondents (between 2011 and 2015). Where these 511 respondents came from is unclear. Remarkably, the response, completion, verification, and agreement rates all remain unchanged after such substantial growth in the sample size.[15]

More baffling still are the descriptive statistics. The samples in these articles differed in many ways. Whereas the Johnson et al. (2011) analytic sample was racially diverse, Stewart et al.'s (2015) analytic sample was racially homogenous, including only non-Latino Whites. The Johnson et al. (2011) respondents lived in 91 counties, and those in Stewart et al.'s (2015) lived in 88 counties. In sum, these articles had different total sample sizes, different analytic sample sizes at both levels (individual and county), and analytic samples with different racial compositions. As Table 3 shows, despite all of these differences, most of the descriptive statistics in the two samples are identical (unboxed).

TABLE 3. Johnson et al. (2011) vs. Stewart et al. (2015):
Mostly identical descriptive statistics (numbers that change are in boxes)

	Johnson et al. (2011)		Stewart et al. (2015)	
Variables	Mean	SD	Mean	SD
Dependent variable				
Use of ethnicity in punishment	.31	.46	-	-
Punitive Latino sentiment	-	-	10.75	4.73
Independent variables				
Perceived Hispanic/Latino threat				
Hispanic/Latino criminal threat	4.93	1.66	4.93	1.66
Hispanic/Latino economic threat	1.72	1.13	1.72	1.13
Hispanic/Latino political threat	4.38	1.41	4.38	1.41
Aggregate Hispanic/Latino threat				
Percent Hispanic/Latino	.12	.11	.16	.14
Hispanic/Latino growth	.26	1.53	.03	.13
County characteristics				
Homicide rate (per 100,000)	3.96	4.37	3.96	4.37
Concentrated disadvantage	1.09	1.53	1.48	1.39
Percent Republican	53.04	13.02	.53	.13
Percent Black	.10	.14	.19	.15
Population structure	5.39	.70	5.39	.70

15. In their National Science Foundation proposal, Stewart and Martinez (2010) listed the same 2008 survey with the same rates, but claimed it had a total sample size of 929, instead of 1,184 or 1,379.

TABLE 3 (continued). Johnson et al. (2011) vs. Stewart et al. (2015):
Mostly identical descriptive statistics (numbers that change are in boxes)

	Johnson et al. (2011)		Stewart et al. (2015)	
Variables	Mean	SD	Mean	SD
Demographic characteristics				
White	.86	.41	-	-
Black	.10	.33	-	-
Hispanic	.04	.22	-	-
Age	47.12	19.72	47.12	19.72
Male	.47	.50	.47	.50
Married	.59	.31	.59	.31
Education level (college graduate)	.42	.31	.42	.31
Family income	$62,700	$14,210	$62,700	$14,210
Employed	.46	.50	.46	.50
Political conservative	.43	.31	.43	.31
Own home	.78	.33	.78	.33
Southwest	.17	.41	.17	.41
Northeast	.15	.35	.15	.35
Midwest	.24	.43	.24	.43
West	.17	.38	.17	.38
South	.44	.39	.44	.39
General punitive attitudes	6.84	2.16	6.84	2.16
Level 1: Full Sample N	1,184, mixed-race sample		1,379 mixed-race sample	
Level 1: Analytic N	1,184, mixed-race sample		1,186 non-Latino Whites	
Level 2: Analytic N	91 counties		88 counties	

In December 2010, after a colleague read a draft of our manuscript and suggested we control for county political climate, Dr. Stewart asked me to collect county voting percentages and sent me an Excel file to use. It was an individual-level file without any variables that only included case numbers and geographic identifiers. He sent this file just a few weeks before we submitted our manuscript to *Criminology*, and long after he analyzed the data and produced results mostly identical to those in the published article (see above), so it should have included the right number of respondents and counties. It did not. The file included 1,000 respondents in 292 counties, not 1,184 respondents in 91 counties (more on this shortly).

The discrepancy in sample size has an important implication: the descriptive statistics for the variable I collected should differ from those in the published article. This is true regardless of the explanation for the sample size discrepancy. In the file I sent back, the mean percent voting Republican was 53.04 with a standard deviation of 13.02. This variable (Percent Republican) had the same mean (53.04)

and standard deviation (13.02) in our published article, even though the sample size at both levels changed (from $N_1 / N_2 = 1,000 / 292$ to $N_1 / N_2 = 1,184 / 91$). Although provided in proportion rather than percentage format, the variable also had the same mean (.53) and standard deviation (.13) in Stewart et al. (2015), even though the sample size at both levels changed again in that article. This kind of stability in statistics despite changes in sample size does not happen with real data.

Inexplicable sample sizes and statistics

After we received two emails in 2019 identifying data irregularities in our article and in four others coauthored by Dr. Stewart, I asked my coauthors to send me the full data for Johnson et al. (2011). Two things happened. First, Dr. Johnson told me he did not have a copy and had never seen the data. Second, I encountered difficulties getting the data from Dr. Stewart (more on this later). Consequently, I examined the limited Excel file I already had, which Dr. Stewart had sent in December 2010, shortly before we submitted our article. I discovered the file had only 1,000 respondents, not the 1,184 reported in the article, and that only 500 of those respondents were unique; the other 500 were duplicates. I informed my coauthors about the duplicates and other issues. Dr. Gertz then contacted the former TRN director, who confirmed that the survey he ran for us included only 500 respondents.

On June 10, 2019, Dr. Stewart finally shared with Dr. Johnson and me a copy of the data for our article. At that time he admitted "there are 300+ county units and 500 individuals."[16] Dr. Stewart wrote me to explain how the duplication happened: "I thought I was merging files from two different surveys for which I had questions. I merged the wrong the file."[17] This explanation is problematic, because Johnson et al. (2011) and Stewart et al. (2015) both described one survey, not two. Ten days later, Dr. Stewart gave Dr. Johnson the same explanation for the duplicates, "I received multiple files, but mistakenly merged the incorrect one," but he added something new: "I found the correct data files that should have been merged."[18] Several things here are problematic.

First, the former director of the TRN has said repeatedly that there was only one sample, and it included only 500 respondents. He did send Dr. Stewart two files, but they were for the *same* survey and included the *same* respondents. In one email, Mr. Bratton wrote: "The second file sent per my email you cite was a match file of census data to merge on respondent ID based on self-report zipcode, none

---

16. Email from Dr. Stewart to me, June 10, 2019.
17. Email from Dr. Stewart to me, June 10, 2019.
18. Email from Dr. Stewart to Dr. Johnson, June 20, 2019.

of the original data was included."[19] In another email, he wrote:

> Dr. Stewart was sent exactly 500 records that January 2008 … [and] in May of 2008 I sent another SPSS file that only had the names/census info of the counties for the first ~420 and I think he had to [look] up the county names by zip code in the file for the other 80 by hand to do his analysis.[20]

Second, accidentally doubling the sample would not have yielded the numbers reported in our published article. Dr. Stewart claimed he accidentally included the 500 duplicates in his analysis for our article because he "merged the wrong the file."[21] He clarified that "the correct data files that should have been merged" were not.[22] This is important, so it bears repeating. Dr. Stewart has said that the duplicates were included in the analysis for our published article and that the explanation for the duplicates is accidental doubling. However, accidentally doubling the full sample of 500 respondents, would have resulted in 1,000 respondents, not the 1,184 reported in our article, nor the 1,379 reported in Stewart et al. (2015).

Figure 2. Johnson et al. (2011): Published article vs. shared data
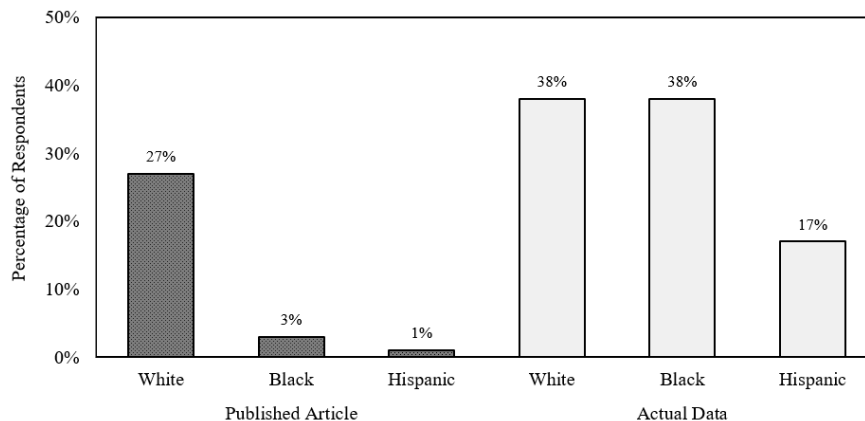
Additionally, the descriptive statistics in the published article should match those in the data, even if the 500 respondents were accidentally doubled. Doubling the sample would increase the sample size, but it would not change its composition. However, the descriptive statistics in the published article differ substantially from the shared data. The outcome variable in our analysis is public support for the use

---

19. Email from Mr. Bratton to me, November 11, 2019.
20. Email from Mr. Bratton to Dr. Gertz, June 7, 2019.
21. Email from Dr. Stewart to me, June 10, 2019.
22. Email from Dr. Stewart to Dr. Johnson, June 20, 2019.

of defendants' ethnicity in sentencing decisions. The distribution of the outcome variable by respondents' race is shown on page 419 of Johnson et al. (2011). Figure 2 shows how it compares to the shared data ($N = 500$). Doubling the sample and then adding another 184 respondents (to get the reported sample size of 1,184) cannot explain the discrepancies. For example, even doubling the sample to 1,000 and then adding 184 Black respondents who *all* oppose ethnicity-based sentencing would reduce the percent of Blacks supporting it from 38 percent to 13 percent, not to the article's 3 percent.

TABLE 4. Johnson et al. (2011): Descriptive statistics in published article vs. shared data
(significant differences are in boxes)

Variables	Published article		Shared data		p-value for difference
	Mean	SD	Mean	SD	
Use of ethnicity in punishment	.31	.46	.37	.48	.007
Hispanic criminal threat	4.93	1.66	2.85	2.67	.000
Hispanic economic threat	1.72	1.13	1.64	1.13	.137
Hispanic political threat	4.38	1.41	1.95	1.59	.000
Percent Hispanic	.12	.11	9.52	11.61	.000
Hispanic growth	.26	1.53	3.18	3.11	.000
Homicide rate (per 100,000)	3.96	4.37	3.92	4.25	.819
Concentrated disadvantage	1.09	1.53	1.40	.94	.000
Percent Republican	53.04	13.02	52.56	13.07	.415
Percent Black	.10	.14	11.63	11.98	.000
Population structure	5.39	.70	5.11	1.00	.000
White	.86	.41	.85	.36	.606
Black	.10	.33	.10	.30	1.000
Hispanic	.04	.22	.05	.21	.494
Age	47.12	19.72	46.41	16.98	.352
Male	.47	.50	.46	.50	.591
Married	.59	.31	.61	.49	.275
Education level (college graduate)	.42	.31	.42	.49	.902
Family income	$62,700	$14,210	$61,196	$22,593	.137
Employed	.46	.50	.55	.50	.000
Political conservative	.43	.31	.70	.46	.000
Own home	.78	.33	.78	.41	.829
Southwest	.17	.41	.16	.37	.721
Northeast	.15	.35	.15	.36	.802
Midwest	.24	.43	.24	.43	.917
West	.17	.38	.17	.38	.812
South	.44	.39	.43	.50	.787
General punitive attitudes	6.84	2.16	4.69	3.57	.000

There are other noteworthy distributional differences. Table 4 compares all of the descriptive statistics in the published article to those in the shared data. Some of the differences are simply impossible. In the article, for example, the 1,184 respondents had a mean of 4.38 on the Hispanic political threat index. In the shared data, the 500 respondents have a mean of only 1.95. Working through the math reveals that the 684 additional respondents included in the article (but not in the data) must have had an *average* score of 6.16 on the index. The problem is that this average score is higher than the highest possible value on the index, which is 6.

Similarly, the published article claims that 43 percent of respondents are political conservatives. In the shared data, 70 percent are political conservatives. Even doubling the sample to 1,000 and then adding 184 liberals would only drop the percentage of conservatives in the sample to 59 percent, not to the 43 percent reported in the article. Additionally, the mean for Hispanic criminal threat is almost two points higher (mean = 4.93 vs. 2.85), and the mean for general punitive attitudes is over two points higher (mean = 6.84 vs. 4.69), in the published article than in the shared data. Even doubling the sample and then adding 184 respondents with the highest possible value (a value of 9) on these two variables would only increase their means to 3.81 and 5.36, respectively—both still a point lower than in the article.

TABLE 5. Johnson et al. (2011):
Interaction effects in published article vs. shared data
(notable differences are in boxes)

	Published article			Shared data		
	b	SE	Exp(b)	b	SE	Exp(b)
Variables						
Perceived Hispanic threat						
Criminal threat	.183[*]	.079	1.201	.212[***]	.048	1.236
Economic threat	.272[*]	.111	1.312	.632[***]	.124	1.881
Political threat	.008	.116	1.009	−.045	.072	.956
Aggregate Hispanic threat						
Percent Hispanic	−.089	.766	.815	.032[*]	.015	1.032
Hispanic growth	.334[**]	.127	1.396	−.121	.064	.886
Interactions						
Criminal × His. Growth	.126[*]	.051	1.134	.026	.017	1.026
Economic × His. Growth	.175[*]	.073	1.191	.015	.055	1.015
Political × His. Growth	−.101	.087	.904	.007	.022	1.007
Intercept	−.848[***]	.119		−.776[***]	.121	
Notes: [*]p < .05; [**]p < .01; [***]p < .001 (two-tailed).						

Accidentally doubling the sample of 500 respondents would leave the regression coefficients unscathed—they would be identical if all respondents were

duplicated. However, the regression results in the published article differ substantially from those in the shared data. Most notably, the main findings in the article—the interaction effect of perceived Hispanic threat (criminal and economic) and Hispanic growth—do not emerge with the shared data. Those findings are reported on page 423 of Johnson et al. (2011). In the shared data, none of the coefficients for the interaction terms are statistically significant and they are all much smaller. I show the comparison in Table 5.

TABLE 6. Johnson et al. (2011): Model 2 in published article vs. shared data (notable differences are in boxes)

	Published article			Shared data		
	b	SE	Exp(b)	b	SE	Exp(b)
Variables						
Percent Hispanic	−.104	.662	.901	.021	.020	1.021
Hispanic growth	.288**	.102	1.334	−.070	.053	.933
Homicide rate (per 100,000)	−.013	.033	.987	−.031	.030	.970
Concentrated disadvantage	.051	.086	1.052	.033	.144	1.034
Percent Republican	.001	.005	1.000	.009	.009	1.009
Percent Black	−.038	.173	.963	.003	.012	1.003
Population structure	−.051	.174	.950	.068	.126	1.070
Black	−.628*	.248	.534	.949**	.361	2.582
Hispanic	−.982**	.359	.375	−.752	.547	.472
Age	.016**	.005	.016	.004	.006	1.004
Male	.164**	.056	1.178	−.111	.199	.895
Married	.208	.201	1.231	.425*	.210	1.529
Education level	−.099	.198	.906	.215	.226	1.240
Family income	−.026	.061	.974	.062	.096	1.064
Employed	−.191*	.082	.826	−.058	.202	.944
Political conservative	.367*	.146	1.443	.360	.218	1.434
Own home	−.137	.246	.872	.158	.281	1.171
Southwest	−.111	.274	.895	.721	.369	2.057
Northeast	−.183	.281	.832	.522	.334	1.686
Midwest	.054	.230	1.055	.102	.291	1.107
West	.082	.264	1.085	−.350	.362	.705
General punitive attitudes	.191**	.062	1.210	.175***	.031	1.191
Intercept	−.858***	.105		−.640***	.097	

Notes. *p < .05; **p < .01; ***p < .001 (two-tailed).

The main effect of Hispanic growth also fails to replicate in the shared data; indeed, the coefficient is in the opposite direction. This is the case even when the interaction terms are removed from the model. In the published article, the

main effect of Hispanic growth is shown in Model 2 on page 420, and is positive and statistically significant ($b = .288$, $p < .01$). In the shared data, the coefficient is negative and non-significant. Table 6 compares the estimates in Johnson et al. (2011) to those from the shared data. The differences are striking, extending to many other variables besides Hispanic growth. For example, the coefficient for Black in the published article is negative and significant ($b = -.628$, $p < .05$), but it is positive and significant in the actual data ($b = .949$, $p < .01$). Again, the regression results would be identical if the sample was accidentally doubled. They are not, so it is impossible that the only change to the data was accidental doubling.

Unreported, implausible county clusters

In the data Dr. Stewart shared, there are 500 respondents and they are nested in 326 counties. In Johnson et al. (2011), however, we claimed to have 1,184 respondents nested in 91 counties. Similarly, Stewart et al. (2015) used the same data and claimed the respondents in their analytic sample lived in 88 counties. Dr. Stewart has acknowledged that the county numbers reported in both articles are wrong. The explanation he gave to FSU's Inquiry Committee is that he created "county units" by clustering the 326 counties. There are at least three problems with the explanation.

First, there is little justification for grouping together so many counties in either Johnson et al. (2011) or Stewart et al. (2015), given the time period of the studies. It is not desirable to group counties, because it throws away geographic detail and creates "meaningless socio-political entities" (Hagen et al. 2013, 770). Typically, researchers only group counties when their boundaries change during measurement years. Therefore, county clusters normally appear only in historical studies that examine data over a large number of decades or across centuries, and even in those studies the researchers only create county clusters for those specific counties that have boundaries that changed during the time period examined (e.g., King et al. 2009; Messner et al. 2005). To see if there was any justification for grouping counties in Johnson et al. (2011) or Stewart et al. (2015), I investigated how the specific counties in our data changed during the measurement years for the aggregate variables. For these studies, boundary changes would justify the creation of exactly one cluster, made from four counties in Colorado: Adams County, Boulder County, Jefferson County, and Weld County (**link**). That would leave 322 separate counties and one cluster, or 323 county units.

Second, neither article mentioned county clusters or reported any grouping together of counties. FSU's Inquiry Committee noted this in its final report:

Dr. Stewart acknowledged that, although the paper referred to counties and

reported descriptive statistics for counties, in performing the analyses, he aggregated the counties into clusters. He provided no explanation for this decision, but committee members note that aggregating the counties allowed Dr. Stewart to use what was then an emerging statistical approach that could not have been supported by the county-level data.

Figure 3. Dr. Stewart's description of samples in FSU inquiry document

---

Article 1:

Johnson, Brian D., Eric A. Stewart, Justin Pickett, and Marc Gertz. 2011. "Ethnic Threat and Social Control: Examining Public Support for Judicial Use of Ethnicity in Punishment." *Criminology,* 49(2):401-441.

Related data file(s):

- Final Dataset 010808 (500 cases)
- Fsua050108 (343 cases)
- Fsua121407 (164 cases)

Log(s) for the data file(s) - (See Appendix B)

Article 5:

Stewart, Eric A., Brian D. Johnson, Patricia Y. Warren, Jordyn Rosario, and Cresean Hughes. 2019. "The Social Context of Criminal Threat, Victim Race, and Punitive Black and Latino Sentiment." *Social Problems, 66(2), 194-221.*

Related data file(s):

- Final Dataset 010808 (500 cases)
- Fsua050108 (343 cases)
- Fsua121407 (164 cases)
- Final Fsu 08 (425 cases)
- RACE FINAL 2013 (1079 cases)

Log(s) for the data file(s): (See Appendix B)

Preliminary analyses are available. No corrections have been submitted to the journal yet. (Appendix C)

Actually, the situation is quite a bit worse, as Figure 3 shows. Specifically, in the documents Dr. Stewart submitted to FSU, he claimed he included the same data used in our 2011 article in each of his later articles. Of course, that differs from what the 2018 and 2019 articles reported; they all reported a single 2013 survey of 2,736 respondents: "The telephone surveys were conducted during the spring, summer, and fall of 2013" (Stewart, Johnson et al. 2019, 200). Nevertheless, Dr. Stewart told the Committee the 2013 sample only included 1,079 respondents (see Figure 3), and he combined that smaller sample with the data he used in

Johnson et al. (2011). The problem is that one of the files Dr. Stewart supposedly combined—"Final Dataset 010808 (500 cases)"—included 326 counties *by itself*. However, the five articles only reported between 79 and 168 total counties. That leaves two possibilities: Dr. Stewart clustered the counties down to a smaller number in all five articles but failed to report it in any of them, or he fabricated the number of counties in each article.

Figure 4. Research Network data: The 326 counties clustered liberally into 176 'county units' (clusters and counties)
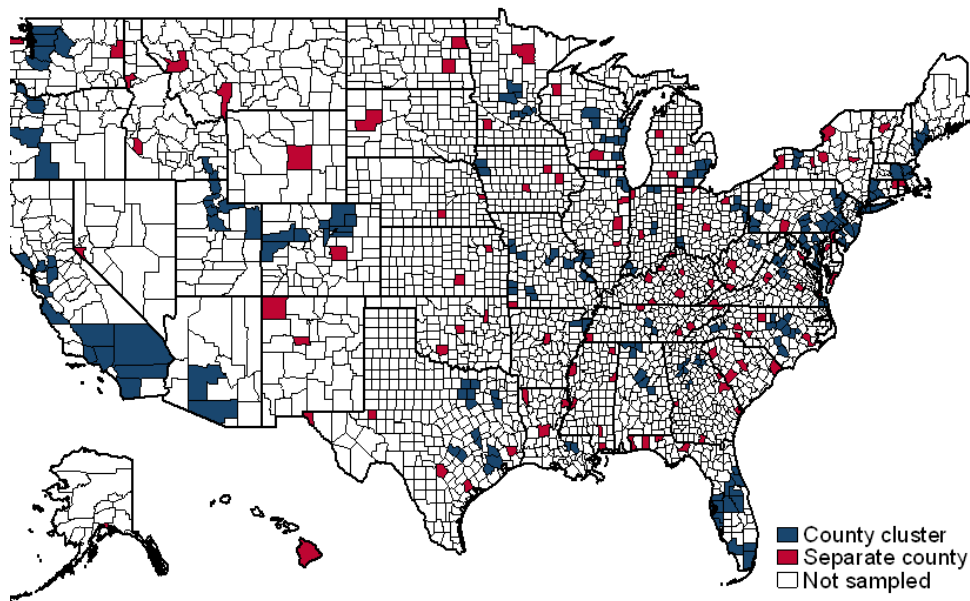
Notes: Counties clustered with adjacent counties in the same state are in blue ($N = 67$). Counties isolated from other counties in the state are in red ($N = 109$).

Finally, there does not appear to be any reasonable clustering method that would yield the number of counties reported in the articles. For example, consider the article I coauthored, Johnson et al. (2011), which reported 91 counties. Dr. Stewart has not provided the code for generating county clusters or any cluster-level data. However, I have tried different clustering methods, and none yield anywhere close to 91 county units (clusters and separate counties). The most liberal method, which yields the smallest number of county units, is to group together any counties in the data that are within the same state and share a border, either directly or indirectly (via another county). Figure 4 shows the results of this clustering method. County clusters are in blue, and separate counties—counties that are not adjacent to other sampled counties in the same state—are in red. Even this method

fails to come close to producing 91 county units. Instead, it yields 109 separate counties and 67 county clusters, for a total of 176 county units.

What happened among the coauthors

The coauthors of the five retracted articles include two past editors of *Criminology*, the flagship journal of the American Society of Criminology (ASC), as well as three ASC Fellows and two ASC vice presidents. Two coauthors, Brian Johnson and Eric Baumer, have written articles about the importance of research ethics (e.g., "What Scholars Should Know about 'Self-Plagiarism'" (Lauritsen et al. 2019); "Salami-Slicing, Peek-a-Boo, and LPUS: Addressing the Problem of Piecemeal Publication" (Gartner et al. 2012)). Thus, how these scholars responded to learning about irregularities in their own articles is likely to be informative about the research norms and practices in criminology and sociology.

To my knowledge, none of the coauthors have spoken publicly about what happened in the retracted articles, except to insist in the retraction notices that the irregularities resulted from "coding mistakes" and "transcription errors" (*Law & Society Review* 2020; *Criminology* 2020a; b), and to defend the accuracy of the retracted findings (*Law & Society Review* 2020). Additionally, several of the coauthors also coauthored other papers with Dr. Stewart that have irregularities (e.g., Mears et al. 2013; Stewart et al. 2009).

Drs. Eric Stewart and Brian Johnson

On May 28, 2019, when I first asked Dr. Stewart for the data for Johnson et al. (2011), he said to wait a week. A week later, he delayed again, claiming: "I have to wait until after I talk to Research Board."[23] I asked him, "why won't they let you share the data with your coauthor?"[24] He ignored my question. So, I contacted FSU's Office of Research (OR), who disputed his claim: "Sharing copies of the data is perfectly okay. Please just do not alter the original data set in any way."[25] I forwarded the OR's email to all of my coauthors, including Dr. Johnson, but Dr. Stewart still withheld the data. It was only after I identified the 500 duplicates in the limited file that I already had that Dr. Stewart shared a copy of the data. He explained, "I had to recreate the data file with no duplicates."[26] In response,

---

23. Text message from Dr. Stewart to me, June 6, 2019.
24. Text message from me to Dr. Stewart, June 6, 2019.
25. Email from Ms. Diana Key to me, June 6, 2019.
26. Email from Dr. Stewart to me, June 9, 2019.

I asked for the original data with duplicates included, so I could try to replicate the published results. Dr. Stewart replied saying he had destroyed the original file: "corrections have been made and saved."[27] He did this despite the OR's directive not to alter the data in any way.

Later, when Dr. Johnson reached out to Dr. Stewart—"I haven't heard from you in a bit … have you had any luck locating the final dataset with the same sample size used in the 2011 paper?"[28]—Dr. Stewart replied saying that he found another 425 respondents from a second survey. But neither he nor Dr. Johnson told me. When Dr. Stewart eventually sent Dr. Johnson the new output for the combined sample ($N = 925$), he told Dr. Johnson he was not going to share it with me. Eventually, I heard about the new developments from someone who was not a coauthor, and I emailed Dr. Stewart. Dr. Stewart refused to answer my questions about the second survey. So I contacted Dr. Johnson, explained that I could not get a response from Dr. Stewart, and asked him to send me the new output. "Sorry I'm just getting back to you, I needed a few days to consider your request," Dr. Johnson replied after several days, "Eric asked that I not share additional output…[and] I think it is best under the circumstances that any data/output comes directly from Eric."[29]

Dr. Johnson asked Dr. Stewart for the combined data in June, "if you don't mind sending me a copy of the updated 925 data I'd like to look through it to see if I can refute any of Justin's claims."[30] Dr. Stewart replied to him immediately, "I will send it. Let me pair [*sic*] it down to just the variables we use. I will send in a bit."[31] However, he never followed through. "Eric never did send me the data but he did allow me to remotely access his computer while he was running the new models," Dr. Johnson wrote months later.[32] Despite never receiving the data, despite knowing that Dr. Stewart had deleted data against the ORI's directive, and despite being the lead author on our article, Dr. Johnson went along with Dr. Stewart's request to withhold information from a coauthor (me) for four months.

Dr. Marc Gertz

Dr. Marc Gertz coauthored two of the five retracted articles. After I told Dr. Gertz there were hundreds of duplicates in Johnson et al. (2011), he emailed Jake Bratton, the former TRC director, to double-check. Mr. Bratton confirmed to him

---

27. Email from Dr. Stewart to me, June 10, 2019.
28. Email from Dr. Johnson to Dr. Stewart, June 19, 2019.
29. Email from Dr. Johnson to me, July 1, 2019.
30. Email from Dr. Johnson to Dr. Stewart, June 27, 2019.
31. Email from Dr. Stewart to Dr. Johnson, June 27, 2019.
32. Email from Dr. Johnson to me, October 31, 2019.

that there was only one 2008 sample and that it only had 500 respondents, writing: "Basically the choices are inexperience in blending data resulting in loading the data twice and not noticing or more sample size was needed to get to p<.05."[33] What did Dr. Gertz do after learning from two different sources—Mr. Bratton and me—that an article he coauthored reported over twice as many respondents as were actually interviewed? He emailed Dr. Stewart a letter of support. In it, he wrote:

> At a minimum, Eric Stewart, my colleague, received data sets from me in 2008, 2009, 2013, 2017, and 2018 investigating various topics around the influence of racism in our culture. It is possible he had access to other data sets from me as well. Anyone who claims Eric had access to only one national survey is clearly incorrect.[34]

The letter did two important things. It seemingly disputed my claim that there was only one survey done for our 2011 article, and it implied that the 2013 data came from Dr. Gertz. Dr. Gertz wrote this letter despite what Mr. Bratton told him, and despite having previously written to me that he did not conduct the 2013 survey ("Not me, wish it were"). Dr. Stewart then provided the letter to his coauthors, the journals, and FSU to support his claims about both the 2008 and 2013 surveys. The letter became a key piece of evidence in their investigations. As late as September, one editor wrote: "the only response I have from the authors is an email from someone confirming he did the survey,"[35] to which another editor replied "we received the same letter from the person who supposedly conducted the surveys in question."[36]

Dr. Daniel Mears

Dr. Mears was the lead author on the 2019 *Law & Society Review* article and a coauthor on Stewart et al.'s (2018) *Criminology* article. After being contacted by Drs. Brown and Heathers in February, Dr. Mears coauthored corrections to both articles that were mathematically impossible. For example, the reported age difference between the original and corrected samples required the 140 dropped respondents to have a negative mean age (Polyacantha 2019). When alerted to these impossibilities, Dr. Mears sent an email to Dr. Sterett, the editor of *Law & Society Review*, disputing them and attacking Drs. Brown and Heathers. He stressed "the lack of clear credentials that these individuals have for leveling post-hoc critiques,"

---

33. Email from Mr. Bratton to Dr. Gertz, June 7, 2019.
34. Email from Dr. Gertz to Dr. Stewart, August 6, 2019.
35. Email from Dr. Sterett to Dr. Linders, September 25, 2019.
36. Email from Dr. Linders to Dr. Sterett, September 25, 2019.

and said "'data thugs appear intent on maligning researchers, journal editors, programs, and universities under the guise of 'advancing' science."[37]

Months later, after I posted a preprint on SocArXiv describing the problems in the 2008 data (Pickett 2019), Dr. Mears wrote another email to Dr. Sterett disputing what I wrote. He also provided Dr. Gertz's letter of support to prove he conducted the 2013 survey, writing: "there was no funding source … The survey data were freely provided to Dr. Stewart by Dr. Marc Gertz."[38] Over a month later, Dr. Mears emailed Dr. Sterett and once again defended Mears et al. (2019), claiming that the 61 percent response rate he reported "was comparable to well-conducted national studies using random-digit dialing techniques (Czajka and Beyler 2016)."[39] The article he cited focused on government surveys; it said the National Immunization Survey (NIS) and the Behavioral Risk Factor Surveillance System, two highly-funded telephone surveys, achieved 2013 response rates of 62 percent and 46 percent, respectively. NORC's five-year contract to administer the NIS is valued up to $163,658,456 (**link**). Well conducted, indeed.

Dr. Mears did these things—coauthored the corrections, disparaged "data thugs," impugned me, repeatedly reassured the editor—apparently without analyzing the data himself, or even verifying the accuracy of the data description in the articles. For example, Dr. Stewart told FSU that in both articles he combined samples from different years, but failed to report it.[40] However, Dr. Mears said nothing about this in his emails to the editor, which means he either did not know about it or he withheld the information. Recall, too, that Dr. Mears gave the editor a false explanation for the mean/SD discrepancies, so he apparently did not verify the information that he himself communicated, either.

Drs. Eric Baumer and Patricia Warren

Drs. Baumer and Warren each coauthored two of the retracted articles, and also coauthored corrections to their articles that were mathematically impossible, apparently without verifying the data themselves. On June 10, I alerted them, and five other coauthors, about the problems in the 2008 data (e.g., inclusion of duplicates, county number discrepancy, differences in findings, etc.). I reminded them the information "has direct implications for the 2015 Social Problems article (Stewart, Martinez, Baumer, and Gertz), which uses the same data, and indirect implications for all of the other articles." My email included one, seemingly

---

37. Email from Dr. Mears to Dr. Sterett, May 30, 2019.
38. Email from Dr. Mears to Dr. Sterett, August 7, 2019.
39. Email from Dr. Mears to Dr. Sterett, September 29, 2019.
40. Dr. Stewart said he combined 1,432 respondents interviewed in 2007 and 2008 with 1,079 interviewed in 2013. That sums to a total sample size of 2,511, which is 225 less than the 2,736 reported in the articles.

noncontroversial recommendation: "I am sending you this because, if you haven't yet, you may want to request a copy of the data from the article you are on, and examine it yourself."

Dr. Baumer replied to everyone, writing: "my own view is that it is important to give Eric Stewart a chance to address these questions before drawing conclusions or taking further steps."[41] His clear suggestion was that none of the coauthors should ask for the data, which is the only thing that I had recommended they do. Shortly thereafter, Dr. Baumer and I talked on the phone and he told me he refused to request the data from Dr. Stewart, who he emphasized was his close friend. He also said that although he had never seen the data, he had nonetheless consulted with an editor about the data irregularities, and had also offered to travel to FSU to meet with university officials on Dr. Stewart's behalf.

After receiving my June 10 email about the data problems, Dr. Warren also replied to everyone, and seemingly suggested that it would be unprofessional or even disrespectful to ask for the data. She wrote: "I plan to operate with full professionalism and scholarly respect. I too choose to give Eric Stewart time to work through all the issues."[42] Almost two months later, she contacted journal editors to defend Dr. Stewart and, apparently, to blame me. "The email we just received from Patricia Warren…ends up shifting the blame to the whistle blower instead," the editors of *Social Problems* wrote on August 6.[43]

Later in August, Dr. Warren filed a complaint about me with the police. It was about my June 10th reply to her and Dr. Baumer's emails, where I had written, "Let me stop this before it turns into an assault." I meant 'an assault on me,' and I apologized for my email's tone shortly after sending it. To say I was shocked when the police contacted me in late August would be an understatement. The last communication I had with Dr. Warren was two months earlier, when she responded to my apology: "Thank you for your apology. All will be well."[44]

In response to publicity surrounding my preprint and Dr. Stewart's articles, the American Society of Criminology held a Forum on Scientific Integrity at its annual meeting in November 2019. At the forum, despite having coauthored two of the articles, having emailed editors, and having contacted the police about my email, Dr. Warren told the ASC executives, "I am not necessarily attached to the incident." Her question for them was about how they planned to deal with it when "there is a public war…attached to what's going on."[45]

---

41. Email from Dr. Baumer to me and six other coauthors, June 10, 2019.
42. Email from Dr. Warren to me and six other coauthors, June 10, 2019.
43. Email from Dr. Linders to Dr. Wright, August 6, 2019.
44. Email from Dr. Warren to me, June 12, 2019.
45. Dr. Warren's comments start at 39:53 in **this video**.

What happened at the journals

The Committee on Public Ethics (COPE) says that journal editors should investigate when "a published article is criticised via direct email," regardless of whether the sender is anonymous, and emphasizes that "it is important not to try to 'out' people who wish to be anonymous" (**link**). An analysis of emails obtained under the Freedom of Information Act reveals that all of the editors were alerted in May about the data irregularities in all five of Dr. Stewart's articles, either by an anonymous sender—"John Smith"—or by another editor.

Law & Society Review, regarding Mears et al. 2019

After receiving the anonymous email, Dr. Susan Sterett, the journal's editor, shared it with the authors, writing: "it seems to imply pretty egregious misconduct—points 4 and 8 especially."[46] Then she emailed the editors of the other journals and tried to get them all on the same page. She wrote: "I would like to have a coordinated response, *including possibly ignoring the email*" (my emphasis).[47] She also explained that she tried to discover the source's identity: "I asked 'John Smith' to give me more information about himself and he would not."[48]

In July, after I posted my preprint, Dr. Sterett contacted the other editors again to reiterate her position, "I am not interested in asking for a response from the authors to an anonymous email. However, to my mind it's worth knowing that the issue isn't going away."[49] But she also explained that if any of the other editors ever decided to do anything, she wanted to be included: "I'd appreciate knowing, and I'd appreciate doing something together."[50] In late August, she contacted the other editors again, and forwarded them a discussion by Dr. Jeremy Freese of the mathematical impossibilities in the articles. In the same email, Dr. Sterett noted that she had received Dr. Gertz's letter of support, and once more reiterated her stance on the data irregularities: "I want to treat the issue as closed unless someone wants to question the survey in detail."[51] Months after she closed the issue, the article was retracted at the authors' request.

---

46. Email from Dr. Sterett to Dr. Mears, May 29, 2019.
47. Email from Dr. Sterett to Drs. Linders and Wright, May 31, 2019.
48. Email from Dr. Sterett to Drs. Linders and Wright, May 31, 2019.
49. Email from Dr. Sterett to Drs. McDowall, Linders, and Wright, July 15, 2019.
50. Email from Dr. Sterett to Drs. McDowall, Linders, and Wright, July 15, 2019.
51. Email from Dr. Sterett to Drs. McDowall, Linders, and Wright, August 26, 2019.

Social Problems, regarding Stewart et al. 2015 and Stewart, Johnson et al. 2019

Drs. Annulla Linders and Earl Wright, the journal's co-editors, received an email in May from Dr. Stewart listing some of the accusations and irregularities, and, in relation to Dr. Brown, Dr. Heathers, and Mr. Smith, asserting that "data thugs…demand data and if they do not receive it, they contact editors and universities and threaten to write blogs and tweets about the errors uncovered."[52] Drs. Linders and Wright also received emails in May and July from Dr. Sterett about Dr. Stewart's articles. The May email included a full list of the irregularities in all five articles. The co-editors did not investigate.

Two weeks after they got Dr. Sterett's second email, and two months after they received Dr. Stewart's email, Drs. Linders and Wright received an email from a reporter, Thomas Bartlett, asking if they were looking into the irregularities in Stewart et al. (2019). They replied, "no question concerning this paper has been brought to our attention."[53] Before replying to the reporter, however, Dr. Wright wrote to Dr. Linders: "a writer from the Chronicle of Higher Education is sniffing around. Is the paper he cites below the one inquired into by the 'data thugs?' Of course, I won't respond until we get a plan together."[54]

When I found out what Drs. Linders and Wright had told the reporter, I emailed them my preprint. It turns out they already had it. "Earl just stumbled on the (damaging) information below," Dr. Linders wrote about the preprint when they first found it online.[55] Dr. Wright had circulated the preprint before I emailed them; he noted that "one of the author's 'outted' is Stewart,"[56] to which Dr. Linders responded: "Ok, so more damage."[57] Still, it apparently took a direct email from a non-anonymous source to get them to investigate. "So now we have a formal complaint to justify an investigation," Dr. Linders wrote after receiving my email.[58]

Criminology (regarding Johnson et al. 2011 and Stewart et al. 2018)

Criminology's co-editors—Drs. Brian Johnson, Janet Lauritsen, David McDowall, and Jody Miller—adopted a comment-and-reply response to the anony-

---

52. Email from Dr. Stewart to Drs. Linders and Wright, May 30, 2019.
53. Email from Dr. Wright to Mr. Bartlett, July 31, 2019.
54. Email from Dr. Wright to Dr. Linders, July 31, 2019.
55. Email from Dr. Linders to Mr. Michael Blong and several others, August 2, 2019.
56. Email from Dr. Wright to Dr. Linders, August 2, 2019.
57. Email from Dr. Linders to Dr. Wright, August 2, 2019.
58. Email from Dr. Linders to Dr. Sterett, August 5, 2019.

mous email. They "invited 'John Smith' to submit a comment or comments about the articles in question" (Johnson et al. 2019; Stewart et al. 2018), but demanded he reveal his identity before they would move forward.[59] One co-editor, Dr. Johnson, kept Dr. Stewart in loop. When the co-editors first learned of the data irregularities, Dr. Johnson told Dr. Stewart: "I'm not sure it will amount to anything but thought you'd want to know. If anything comes down the pipeline I'll keep you informed."[60] When the co-editors eventually offered 'John Smith' the comment-and-reply opportunity, Dr. Johnson let Dr. Stewart know, explaining that:

> One of the provisions of the invitation is that the critique could not be anonymous and it would also be subject to external reviews. I think if we end up submitting a correction first that would make the comment pointless, and I think David et al. just felt like they had to provide some type of formal response to the email.[61]

Later, a professor from another discipline, Dr. Walter Schumm, who also had analyzed the irregularities in Dr. Stewart's articles, did send *Criminology* a non-anonymous critique, but the co-editors denied the professor the comment-and-reply opportunity. Dr. McDowall, the lead editor, replied to Dr. Schumm, writing: "I do not think additional commentary will be useful at this point, but I appreciate your offer to contribute."[62]

In June, I emailed the co-editors and asked them to retract Johnson et al. (2011). I sent them evidence that neither the findings nor sample reported in the article existed, and I told them that Drs. Stewart and Johnson were refusing to share data or even output with me. The co-editors replied saying they were going to give my coauthors a few months to work through their reanalysis. During that time period, my coauthors were free to withhold data, output, and even basic information from me. That is so unbelievable it bears repeating: for several months, the co-editors let Drs. Stewart and Johnson refuse to share data and output with a coauthor. About the evidence I sent, Dr. McDowall told the *Chronicle of Higher Education* that he "didn't read it in great depth," and that he thought it was "pretty hostile for Justin to start making these claims" (Bartlett 2019).

After Stewart, Mears et al.'s (2019) corrigendum was published, accusations quickly surfaced that it was mathematically impossible. Mr. Bartlett, the *Chronicle* reporter, asked Dr. McDowall about those accusations, and Dr. McDowall responded by outlining his negative views of the accusers:

---

59. See email from Dr. Johnson to Dr. Stewart, June 20, 2019.
60. Email from Dr. Johnson to Dr. Stewart, May 29, 2019.
61. Email from Dr. Johnson to Dr. Stewart, June 20, 2019. Dr. Johnson was referring to his co-editors.
62. Email from Dr. McDowall to Dr. Schumm, September 27, 2019.

> If regular circumstances prevailed, I imagine that the corrigendum would have passed without notice after it appeared. Given the current situation, I was surprised that it took five days before the trolls on www.socjobrumors discovered it and began savaging it. I also know, of course, that a Stanford sociologist with a perhaps unjustifiably high sense of self-esteem has tweeted disparagingly about the correction. From my point of view, some of the socjobrumors postings offer better and more thoughtful criticisms than did the high self-esteem Stanford sociologist.[63]

Dr. McDowall also discussed the timeline of the corrigendum and explained why it was not the journal's responsibility to ensure its accuracy:

> Eric Stewart asked sometime around February if he could submit a correction to his 2018 article. This was well before the appearance of "John Smith" or Justin Pickett. It may have been after the "data thugs" contacted him, but he did not mention that as a motivation … The document does not address any specific criticism that the journal has or will publish, and it is not itself an original peer reviewed contribution. It is simply an author's attempted correction to a set of results, and it is unnecessary and out of place for me to offer a defense of it.[64]

But there is more to the story than that. The Stewart, Mears et al. (2019) corrigendum was published in mid-August. The anonymous email Dr. McDowall and his coeditors received three months before, in May, did more than just list the irregularities in the original Stewart et al. (2018) article. It also explained that Dr. Stewart and his coauthors had sent a correction for the same 2013 sample to another journal, after receiving outside criticism in February; that the correction appeared to cover up the original irregularities, rather than explain them; and that the correction had many new irregularities. The May email listed those irregularities, which included the same mathematical impossibilities that Dr. Freese (the Stanford sociologist) and others later pointed out in the *Criminology* corrigendum. The co-editors either did not read the anonymous email or ignored its content.

Dr. McDowall also expressed to Mr. Bartlett disapproval of the criticism directed at Dr. Stewart and provided a potential explanation for the irregularities. Even though the corrigendum listed a single coding error and reported using the same 2013 data as the original article, Dr. McDowall wrote:

> I will nevertheless suggest the outlines of a defense, since I think that Stewart

---

63. Email from Dr. McDowall to Mr. Bartlett, August 27, 2019.
64. Email from Dr. McDowall to Mr. Bartlett, August 27, 2019.

has been treated unfairly about it … the descriptive results are not possible if the original and corrected versions used exactly the same data … I have not worked through whether variations in missing data patterns could in fact account for the different summary statistics. It seems to be a reasonable possibility, however, and I offer the matter to the trolls and Stanford University professors … in fairness please note that Stewart does not represent the data as being identical in both samples. Again, this would not be a substantial issue absent the lynch mob atmosphere that was only beginning to emerge as he completed the correction document.[65]

Mr. Bartlett asked whether the co-editors would request Dr. Stewart's data. Dr. McDowall said they would not. He explained, "We will not be rushed into one-sided decisions to satisfy the demands of internet bullies or Stanford University professors, no matter how high their apparent self-esteem."[66] Both of the *Criminology* articles were eventually retracted, at the authors' request. Afterward, the co-editors published a statement claiming that science was coming "under growing attacks from…those who are trying to establish themselves as self-appointed guardians (and often entrepreneurs) of science" (McDowall et al. 2020).

Conclusion and recommendations

The articles

Scientific fraud occurs all too frequently—approximately 1 in 50 scientists admit to fabricating or falsifying data (Fanelli 2009)—and I believe it is the most likely explanation for the data irregularities in the five retracted articles. Dr. Stewart's current claim about the source of the 2013 survey differs from his previous claim and from what the survey firm's owner and director have said. His claim about the number of 2008 samples also differs from the director's account. When asked for the 2008 data, Dr. Stewart claimed he destroyed the original file, even though FSU officials said not to change it. More generally, many aspects of the data and findings are impossible, and others are so implausible or improbable as to be preposterous.

Knowing whether the retracted articles are fraudulent is important because Dr. Stewart has several other articles with irregularities (e.g., Mears et al. 2013; Mears et al. 2017; Stewart 2003; Stewart and Simons 2010; Stewart et al. 2006;

---

65. Email from Dr. McDowall to Mr. Bartlett, August 27, 2019.
66. Email from Dr. McDowall to Mr. Bartlett, August 27, 2019.

Stewart et al. 2009). The retraction notices say honest error, not fraud, is the explanation. Fortunately, if that is true, Dr. Stewart could easily prove it: recreate the original sample ($N = 1{,}184$) that produces the findings in Johnson et al. (2011) and then publicly explain how he did it. Dr. Stewart claims he got from $N = 500$ to $N = 1{,}184$ through accidental duplication, but then dropped the duplicates when I asked for the data. Because dropping duplicates is reversible, Dr. Stewart should be able to duplicate his way back to the same sample again, if he is telling the truth. I have made the sample of 500 respondents available publicly. All that is needed now is to know which of the 500 respondents to duplicate, and how many times, to recreate the original sample ($N = 1{,}184$) that produces the findings in our published article. Dr. Stewart could also post code showing how he clustered the 326 counties in the data I released ($N = 500$) down to 91 county units.

Coauthors and data

How did a group of competent researchers end up publishing five unsound, unsalvageable articles? Monopolization of the data seems to be part of the answer. "I never worked with (or even saw) any version of the data," Dr. Johnson wrote about Johnson et al. (2011).[67] "Dr. Stewart had conducted the analyses and created the tables for all five papers," Dr. Mears explained.[68] To my knowledge, none of Dr. Stewart's coauthors ever analyzed, or even laid eyes on, the full data for any of the five articles, including those they first-authored. As a consequence, they took a passive role in validating their articles, even while they took an active role in defending them.

Most how-to-improve-science lists include open data policies. Having more eyes on the data reduces the survival rate for honest errors. Having more eyes on the data reduces the opportunity for fabrication or falsification. But many authors are reluctant to share data publicly, and sometimes there are legitimate privacy concerns or externally imposed restrictions. Sharing data with coauthors, however, should be uncontroversial and feasible. Yet without institutional support, coauthors may feel uncomfortable requesting data. For example, once irregularities were identified in their articles, Dr. Stewart's coauthors were reluctant to press him for the data, probably because of concerns related to friendship and loyalty. (There certainly is no scientific justification for refraining from requesting data.) Therefore, one recommended reform is that, short of an open data policy, journals should at least require authors submitting articles to confirm that all of their coauthors have a copy of the data.

---

67. Email from Dr. Johnson to me, June 6, 2019.
68. Email from Dr. Mears to Dr. Sterett, May 30, 2019.

Editors and COPE

None of the editors followed COPE's guidelines when alerted to the irregularities in Dr. Stewart's articles. One editor seemingly tried to coordinate a collective response of ignoring the allegations, even though she recognized their potential seriousness. At two journals, the editors sought the whistleblower's identity. I believe that it is possible that one or more of the editors would have revealed the whistleblower's identity had they discovered it. For instance, email correspondence reveals that Dr. Johnson kept Dr. Stewart up to date on what his co-editors knew and were doing, even after officially recusing himself from the matter. Not a single editor started an investigation in response to the anonymous allegations. Dr. Linders explained her reluctance to take those allegations seriously:

> At this point, especially since the person complaining would not come forward, I assumed this was something along the lines of the scientific version of complaints about 'fake news' (now 'fake science'). If you cannot verify the credibility of the source, how can you trust the information?[69]

It appears two journals' editors ignored COPE guidelines because they were unfamiliar with them. Once they learned about the guidelines from a publisher's representative, they seemed committed to following them. For example, Dr. Sterett wrote to Dr. Linders: "I think the [COPE] flow chart I sent separately is far and away the most valuable document, except for the point that journals need policies."[70] One recommendation, then, is to require all editors to review COPE guidelines before taking on editorial responsibilities, and to follow them if they receive allegations about an article published in their journal.

At *Criminology*, what seems to have driven how the co-editors responded was sympathy for some of the authors and a low opinion of critics. Connections between the co-editors and authors are likely to blame; Dr. Johnson, the lead author of one article, was a co-editor, and Dr. Stewart, the lead author of the other, was to become a co-editor. The obvious recommendation is to avoid such conflicts of interest. When authors of questioned articles have relationships (professional or personal) with editors, journals should use independent investigators to investigate scientific irregularities.

Before closing, let me emphasize that many journals do require authors to post their data publicly, and some, like the *American Journal of Political Science*, go so far as to replicate reported results before publishing articles (**link**). But even at journals that do not, editors can still request authors' data if concerns about

---

69. Email from Dr. Linders to Mr. Blong and four others, August 1, 2019.
70. Email from Dr. Sterett to Dr. Linders, September 27, 2019.

findings emerge (see Miyakawa 2020). The Stewart scandal took place over five months, and it required considerable time and effort from the editors involved. The editors corresponded extensively with each other and with other parties, including the authors, reporters, officials of their academic societies, publisher representatives, and university administrators. And the *Criminology* co-editors wrote multiple public statements about the steps they were taking to address the problems. Why did they not simply ask Dr. Stewart for his data? It would have saved a lot of time. Is there any good reason for editors not to verify data when someone, much less a coauthor, provides credible evidence of potential fraud?

The titles and abstracts of the five retracted articles

Here are the titles and abstracts of the five articles, as well as the Google Scholar citation counts as of February 27, 2019:

"Ethnic Threat and Social Control:
Examining Public Support for Judicial Use of Ethnicity in Punishment"
(Johnson et al. 2011, *Criminology*, citations: 80)

Research on social inequality in punishment has focused for a long time on the complex relationship among race, ethnicity, and criminal sentencing, with a particular interest in the theoretical importance that group threat plays in the exercise of social control in society. Prior research typically relies on aggregate measures of group threat and focuses on racial rather than on ethnic group composition. The current study uses data from a nationally representative sample of U.S. residents to investigate the influence of more proximate and diverse measures of ethnic group threat, examining public support for the judicial use of ethnic considerations in sentencing. Findings indicate that both aggregate and perceptual measures of threat influence popular support for ethnic disparity in punishment and that individual perceptions of criminal and economic threat are particularly important. Moreover, we find that perceived threat is conditioned by aggregate group threat contexts. Findings are discussed in relation to the growing Hispanic population in the rapidly changing demographic structure of U.S. society.

"A Legacy of Lynchings: Perceived Black Criminal Threat Among Whites"
(Mears et al. 2019, *Law & Society Review*, citations: 3)

This article examines the legacy of lynchings on contemporary whites' views of blacks as criminal threats. To this end, it draws on prior literature on racial animus to demonstrate the sustained influence of lynching on contemporary America. We

hypothesize that one long-standing legacy of lynchings is its influence in shaping views about blacks as criminals and, in particular, as a group that poses a criminal threat to whites. In addition, we hypothesize that this effect will be greater among whites who live in areas in America where socioeconomic disadvantage and political conservatism are greater. Results of multilevel analyses of lynching and survey data on whites' views toward blacks support the hypotheses. In turn, they underscore the salience of understanding historical forces, including the legacy of lynchings that influence contemporary views of blacks, criminals, and punishment policies.

"The Social Context of Latino Threat and Punitive Latino Sentiment"
(Stewart et al. 2015, *Social Problems*, citations: 46)

Prior research on the racial threat perspective and social control typically relies on aggregate-level demographic measures and focuses on racial, rather than on Latino group, composition. This predominant focus in research on racial threat and social control makes it unclear whether the assumed linkages are confined to one subordinate group or whether other groups, such as Latinos, are viewed as threatening and elicit heightened social control reactions as well. In the current study, we use data from the Punitive Attitudes Toward Hispanic (PATH) Study, a national sample of U.S. residents to investigate the influence of macro- and micro-level measures of Latino group threat on punitive Latino sentiment. More specifically, we use multilevel models to detect direct and interactive relationships between Latino presence and perceived Latino threat on punitive sentiment. The findings show that Latino population growth and perceived Latino criminal and economic threat significantly predict punitive Latino sentiment. Additionally, multiplicative models suggest that the effect of perceived criminal threat on punitive Latino sentiment is most pronounced in settings that have experienced recent growth in the size of the Latino population.

"Lynchings, Racial Threat, and Whites' Punitive Views Toward Blacks"
(Stewart et al. 2018, *Criminology*, citations: 8)

Disparities in historical and contemporary punishment of Blacks have been well documented. Racial threat has been proffered as a theoretical explanation for this phenomenon. In an effort to understand the factors that influence punishment and racial divides in America, we draw on racial threat theory and prior scholarship to test three hypotheses. First, Black punitive sentiment among Whites will be greater among those who reside in areas where lynching was more common. Second, heightened Black punitive sentiment among Whites in areas with more pronounced legacies of lynching will be partially mediated by Whites' perceptions of Blacks' criminality and of Black-on-White violence in these areas. Third, the

impact of lynching on Black punitive sentiment will be amplified by Whites' perceptions of Blacks as criminals and as threatening more generally. We find partial support for these hypotheses. The results indicate that lynchings are associated with punitive sentiment toward Black offenders, and these relationships are partially mediated by perceptions of Blacks as criminals and as threats to Whites. In addition, the effects of lynchings on Black punitiveness are amplified among White respondents who view Blacks as a threat to Whites. These results highlight the salience of historical context for understanding contemporary views about punishment.

"The Social Context of Criminal Threat, Victim Race,
and Punitive Black and Latino Sentiment"
(Stewart, Johnson et al. 2019, *Social Problems*, citations: 2)

A well-established body of research focuses on the relationship between criminal threat and the exercise of formal social control, and a largely separate literature examines the effects of victim race in criminal punishment. Despite their close association, few attempts have been made to integrate these related lines of empirical inquiry in the sociology of punishment. In this article, we address this issue by examining relationships among criminal threat, victim race, and punitive sentiment toward black and Latino defendants. We analyze nationally representative survey data that include both subjective and objective measures of criminal threat, and we incorporate unique information on victim/offender dyads to test research questions about the role victim race plays in the formation of anti-black and anti-Latino sentiment in the criminal justice system. The results indicate that both subjective perceptions of criminal threat and minority population growth are significantly related to punitiveness among whites, and that punitive sentiment is enhanced in situations that involve minority offenders and white victims. Moreover, we show that aggregate indicators of racial threat strongly condition the effect of victim race on punitive attitudes. Implications of these findings are discussed in relation to racial group threat theories and current perspectives on the exercise of state-sponsored social control.

Data, code, and documentation

All data, code, and documentation related to this research is available from the journal website (**link**).

References

Bartlett, Tom. 2019. The Criminologist Accused of Cooking the Books. *Chronicle of Higher Education*, September 24. **Link**

Brauer, Jonathan R. 2009. Testing Social Learning Theory Using Reinforcement's Residue: A Multilevel Analysis of Self-Reported Theft and Marijuana Use in the National Youth Survey. *Criminology* 47(3): 929–970.

Criminology. 2020a. Retraction Statement: Lynchings, Racial Threat, and Whites' Punitive Views Toward Blacks. *Criminology* 58: 189. **Link**

Criminology. 2020b. Retraction Statement: Ethnic Threat and Social Control: Examining Public Support for Judicial Use of Ethnicity in Punishment. *Criminology* 58: 190. **Link**

Czajka, John L., and Amy Beyler. 2016. Declining Response Rates in Federal Surveys: Trends and Implications. June 15. U.S. Department of Health and Human Services (Washington, D.C.). **Link**

Diekmann, Andreas. 2007. Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics* 34(3): 321–329.

Diekmann, Andreas, and Ben Jann. 2010. Benford's Law and Fraud Detection: Facts and Legends. *German Economic Review* 11(3): 397–401.

Fanelli, Daniele. 2009. How Many Scientists Fabricate or Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLoS ONE* 4(5): e5738. **Link**

Gartner, Rosemary, D. Wayne Osgood, and Eric Baumer. 2012. Salami-Slicing, Peek-a-Boo, and LPUS: Addressing the Problem of Piecemeal Publication. *The Criminologist* (American Society of Criminology) 37(5, Sept.–Oct.): 23–25. **Link**

Günnel, Stefan, and Karl-Heinz Tödter. 2009. Does Benford's Law Hold in Economic Research and Forecasting? *Empirica* 36(3): 273–292.

Guterbock, Thomas, Grant Benson, and Paul Lavrakas. 2018. The Changing Costs of Random Digital Dial Cell Phone and Landline Interviewing. *Survey Practice* (American Association for Public Opinion Research, Washington, D.C.) 11(2): 1–10. **Link**

Hagan, John, Carla Shedd, and Monique R. Payne. 2005. Race, Ethnicity, and Youth Perceptions of Criminal Injustice. *American Sociological Review* 70(3): 381–407.

Hagen, Ryan, Kinga Makovi, and Peter Bearman. 2013. The Influence of Political Dynamics on Southern Lynch Mob Formation and Lethality. *Social Forces* 92(2): 757–787.

Heathers, James, and Nicholas J. L. Brown. 2019. DEBIT—A Simple Consistency Test for Binary Data. Working paper. **Link**

Johnson, Brian D., Janet Lauritsen, David McDowall, and Jody Miller. 2019. Statement from the Co-Editors of *Criminology*. September 28. American Society of Criminology (Columbus, Ohio). **Link**

Johnson, Brian D., Eric A. Stewart, Justin Pickett, and Marc Gertz. 2011. Ethnic Threat and Social Control: Examining Public Support for Judicial Use of Ethnicity in Punishment. *Criminology* 49(2): 401–441.

Keeter, Scott, Nick Hatley, Courtney Kennedy, and Arnold Lau. 2017. *What Low*

*Response Rates Mean for Telephone Surveys*. Washington, D.C.: Pew Research Center.

King, Ryan D., Steven F. Messner, and Robert D. Baller. 2009. Contemporary Hate Crimes, Law Enforcement, and the Legacy of Racial Violence. *American Sociological Review* 74(2): 291–315.

Kirk, David S. 2008. The Neighborhood Context of Racial and Ethnic Disparities in Arrest. *Demography* 45(1): 55–77.

Kirk, David S., and Mauri Matsuda. 2011. Legal Cynicism, Collective Efficacy, and the Ecology of Arrest. *Criminology* 49(2): 443–472.

Lauritsen, Janet, Brian D. Johnson, David McDowall, and Jody Miller. 2019. What Scholars Should Know About "Self-Plagiarism." *The Criminologist* (American Society of Criminology) 44(1, Jan.–Feb.): 13–14. **Link**

Law & Society Review. 2020. Retraction. *Law & Society Review* 54: 313. **Link**

Maimon, David, and Christopher R. Browning. 2010. Unstructured Socializing, Collective Efficacy, and Violent Behavior Among Urban Youth. *Criminology* 48(2): 443–474.

McDowall, David, Charis Kubrin, and Jody Miller. 2020. Editor's Corner. *The Criminologist* (American Society of Criminology) 45(1, Jan.–Feb.): 9. **Link**

Mears, Daniel P., Eric A. Stewart, Sonja E. Siennick, and Ronald L. Simons. 2013. The Code of the Street and Inmate Violence: Investigating the Salience of Imported Belief Systems. *Criminology* 51(3): 695–728.

Mears, Daniel P., Eric A. Stewart, Patricia Y. Warren, Miltonette O. Craig, and Ashley N. Arnio. 2019. A Legacy of Lynchings: Perceived Black Criminal Threat Among Whites. *Law & Society Review* 53(2): 487–517.

Mears, Daniel P., Eric A. Stewart, Patricia Y. Warren, and Ronald L. Simons. 2017. Culture and Formal Social Control: The Effect of the Code of the Street on Police and Court Decision-Making. *Justice Quarterly* 34(2): 217–247.

Messner, Steven F., Robert D. Baller, and Matthew P. Zevenbergen. 2005. The Legacy of Lynching and Southern Homicide. *American Sociological Review* 70(4): 633–655.

Miyakawa, Tsuyoshi. 2020. No Raw Data, No Science: Another Possible Source of the Reproducibility Crisis. *Molecular Brain* 13: 24.

Mosimann, James E., John E. Dahlberg, Nancy M. Davidian, and John W. Krueger. 2002. Terminal Digits and the Examination of Questioned Data. *Accountability in Research* 9(1): 75–92.

Mosimann, James E., Claire V. Wiseman, and Ruth E. Edelman. 1995. Data Fabrication: Can People Generate Random Digits? *Accountability in Research* 4(1): 31–55.

Nigrini, Mark J. 2012. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Hoboken, N.J.: Wiley.

Pickett, Justin T. 2019. Why I Asked the Editors of *Criminology* to Retract Johnson, Stewart, Pickett, and Gertz (2011). Working paper. **Link**

Pickett, Justin T. 2020. How Universities Cover Up Scientific Fraud. *Areo Magazine* (London), February 20. **Link**

Polyacantha, Centaurea [pseud.]. 2019. Comment on "Corrigendum: Lynchings, Racial Threat, and Whites' Punitive Views Toward Blacks" by Eric A. Stewart, Daniel P. Mears, Patricia Y. Warren, Eric P. Baumer, and Ashley N. Arnio. September 29.

PubPeer.com (PubPeer Foundation, San Francisco). **Link**

Sampson, Robert J., Jeffrey D. Morenoff, and Stephen Raudenbush. 2005. Social Anatomy of Racial and Ethnic Disparities in Violence. *American Journal of Public Health* 95(2): 224–232.

Savolainen, Jukka, Samantha Applin, Steven F. Messner, Loriine A. Hughes, Robert Lytle, and Jane Kivivuori. 2017. Does the Gender Gap in Delinquency Vary by Level of Patriarchy? A Cross-National Comparative Analysis. *Criminology* 55(4): 726–753.

Schumm, Walter R., Duane W. Crawford, and Lorenza Lockett. 2019. Using Statistics from Binary Variables to Detect Data Anomalies, Even Possibly Fraudulent Research. *Psychology Research and Applications* 1(4): 112–118.

Slocum, Lee A., Terrance J. Taylor, Bradley T. Brick, and Finn-Aage Esbensen. 2010. Neighborhood Structural Characteristics, Individual-level Attitudes, and Youths' Crime Reporting Intentions. *Criminology* 48(4): 1063–1100.

Stewart, Eric A. 2003. School Social Bonds, School Climate, and School Misbehavior: A Multilevel Analysis. *Justice Quarterly* 20(3): 575–604.

Stewart, Eric A., Eric P. Baumer, Rod K. Brunson, and Ronald L. Simons. 2009. Neighborhood Racial Context and Perceptions of Police-Based Racial Discrimination Among Black Youth. *Criminology* 47(3): 847–887.

Stewart, Eric A., Brian D. Johnson, Patricia Y. Warren, Jordyn L. Rosario, and Cresean Hughes. 2019. The Social Context of Criminal Threat, Victim Race, and Punitive Black and Latino Sentiment. *Social Problems* 66(2): 194–221.

Stewart, Eric A., and Ramiro Martinez Jr. 2010. Collaborative Research Proposal: Local Context and Latino Growth: A Multilevel Investigation of Punitive Attitudes. National Science Foundation Proposal 1023333. National Science Foundation (Alexandria, Va.)

Stewart, Eric A., Ramiro Martinez Jr., Eric P. Baumer, and Marc Gertz. 2015. The Social Context of Latino Threat and Punitive Latino Sentiment. *Social Problems* 62(1): 68–92.

Stewart, Eric A., Daniel P. Mears, Patricia Y. Warren, Eric P. Baumer, and Ashley N. Arnio. 2018. Lynchings, Racial Threat, and Whites' Punitive Views Toward Blacks. *Criminology* 56(3): 455–480.

Stewart, Eric A., Daniel P. Mears, Patricia Y. Warren, Eric P. Baumer, and Ashley N. Arnio. 2019. Corrigendum: Lynchings, Racial Threat, and Whites' Punitive Views Toward Blacks. *Criminology* 57(3): 474. **Link**

Stewart, Eric A., Christopher J. Schreck, and Ronald L. Simons. 2006. 'I Ain't Gonna Let No One Disrespect Me' Does the Code of the Street Reduce or Increase Violent Victimization Among African American Adolescents? *Journal of Research in Crime and Delinquency* 43(4): 427–458.

Stewart, Eric A., and Ronald L. Simons. 2010. Race, Code of the Street, and Violent Delinquency: A Multilevel Investigation of Neighborhood Street Culture and Individual Norms of Violence. *Criminology* 48(2): 569–605.

Wilcox, Pamela, Tamara D. Madensen, and Marie Skubak Tillyer. 2007. Guardianship in Context: Implications for Burglary Victimization Risk and Prevention. *Criminology*

45(4): 771–803.

Xie, Min, Janet L. Lauritsen, and Karen Heimer. 2012. Intimate Partner Violence in U.S. Metropolitan Areas: The Contextual Influences of Police and Social Services. *Criminology* 50(4): 961–992.

About the Author
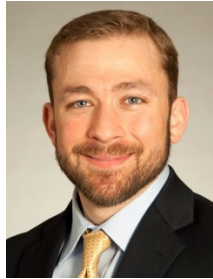
Justin T. Pickett is an associate professor in the School of Criminal Justice at the State University of New York at Albany. He received his Ph.D. in criminology in 2011 from the College of Criminology and Criminal Justice at Florida State University, where he was Eric Stewart's teaching assistant. He is the 2015 recipient of the American Society of Criminology's Ruth Shonle Cavan Young Scholar Award. His research interests include public opinion, survey research methods, theories of punishment, and police-community relations. His email address is jpickett@albany.edu.

Discuss this article at Journaltalk:
https://journaltalk.net/articles/6005/