Response to Irregularities in LaCour and Green (2014)

Michael J. LaCour
University of California, Los Angeles

May 29, 2015

Abstract

In a critique of LaCour and Green (2014), David Broockman, Joshua Kalla, and Peter Aronow, posted a paper online. In this essay, I introduce evidence uncovering discrepancies between the timeline of events presented in Broockman et al. (2015) and the actual timeline of events and disclosure. I argue that Broockman et al.'s failure to replicate LaCour and Green (2014) is likely the result of a failure to follow the respondent-driven sampling procedure in LaCour and Green (2014). However, the failure of Broockman et al. to describe the sampling procedure utilized in their "replication efforts" makes it impossible to evaluate the study's scientific merit and thus has no bearing on LaCour and Green (2014). I show that the results presented in LaCour and Green (2014) withstand criticism of Broockman et al. Most problematic for the claim that the data in LaCour and Green (2014) are "statistically indistinguishable" from CCAP data is the fact that Broockman et al. selected the incorrect variable from CCAP (2012), they then further manipulate this variable to make the distribution look more like that in LaCour and Green (2014). When the correct variable is used, the distributions between the CCAP thermometer and the LaCour and Green (2014) thermometer are statistically distinguishable. Selecting the incorrect variable may have been an oversight, but further manipulating that variable to make the distribution look more like LaCour and Green (2014) is a curious and possibly intentional "error." Broockman et al. (2015) also notably omit the primary analyses reported in the main text of LaCour and Green (2014), which challenges their hypothesis – the within person correlation between the baseline wave and the 9-month follow-up. Finally, a replication experiment, that does not rely on surveys, conducted independently of the parties involved, reproduces the main finding reported in LaCour and Green (2014).

Claim I. LaCour and Green (2014) Met The Replication Standard

Broockman et al. (2015) critique the validity and authenticity of the data in LaCour and Green (2014). The data in question were collected two years ago, May - August 2013. The replication standard requires that researchers create a replication data set that includes all information necessary to replicate empirical results reported in a study (King 1995). The replication dataset in LaCour and Green (2014), have been publicly archived on the ICPSR website since November 11, 2014. Although, LaCour and Green (2014) met the replication standard, below I outline the present controversy.

The information Professor Green indicates he requested in the letter he sent to Gilbert Chin, an editor at *Science* magazine (see Appendix in Broockman et al. (2015)) is the raw survey data including respondents' private/confidential information. After data collection was completed in January 2014, the raw survey data obtained from Qualtrics was destroyed, as required by institutional policy, "Protecting Privacy and Maintaining Confidentiality in Research," which states, "Protocols should be designed to minimize the need to collect and maintain identifiable information about research participants. If possible, data should be collected anonymously or the identifiers should be removed and destroyed as soon as possible..." On LaCour's IRB application, shown in Figure 1, he indicates "The personal identifying information will be removed and destroyed....I will use the unique ID's to merge the canvassing contact information to the survey information. After the data has been merged I will delete the unique ID's." This protocol is not only required, it is common. King (1995) states in his replication article, "to maintain confidentiality, survey organizations do not normally release the names and addresses of respondents. In these and related cases, authors relying on such information can comply with the replication requirement by releasing a subset of their data, stripped of identifying information. However, in some instances, providing any data would be inappropriate." The data in LaCour and Green (2014) are highly sensitive in nature, including respondents geographic location, IP addresses, contact information, and friends information collected through respondent-driven sampling. The request by Broockman et al. (2015) for the raw data is inconsistent with the best practices in the field, and not permitted under the institutional rules at UCLA.

Broockman et al. (2015) cited their inability to replicate the response rate in LaCour and Green (2014) as the impetus for their investigation. On July 17, 2014, Professor Green was asked complete the form, "Science/AAAS Authorship Form and Statement of Conflicts of Interest," to state whether he had, "personally checked all the original data that was generated by my lab or group." My understanding is that Professor Green answered "Yes," to this question, but I do not have access to Professor Green's Science Authorship Form. On May 19, 2015, I was asked to provide the raw survey data, that Professor Green ostensibly claimed to have personally checked on July 17, 2014. I refused to furnish this information because the data reported in LaCour and Green (2014) were destroyed due to privacy/confidentiality requirements as stated above.

I note that Broockman et al. (2015)'s decision to not present the lead author with the critique directly, by-pass the peer-review process, privately investigate data collection activities without knowledge or consent of the author, demand confidential identifying information from respondents in a study without grounds or standing to do so, publicize unsubstantiated allegations and hearsay prior to a formal investigation, is unprecedented, unethical, and anomalous in the relevant literature.

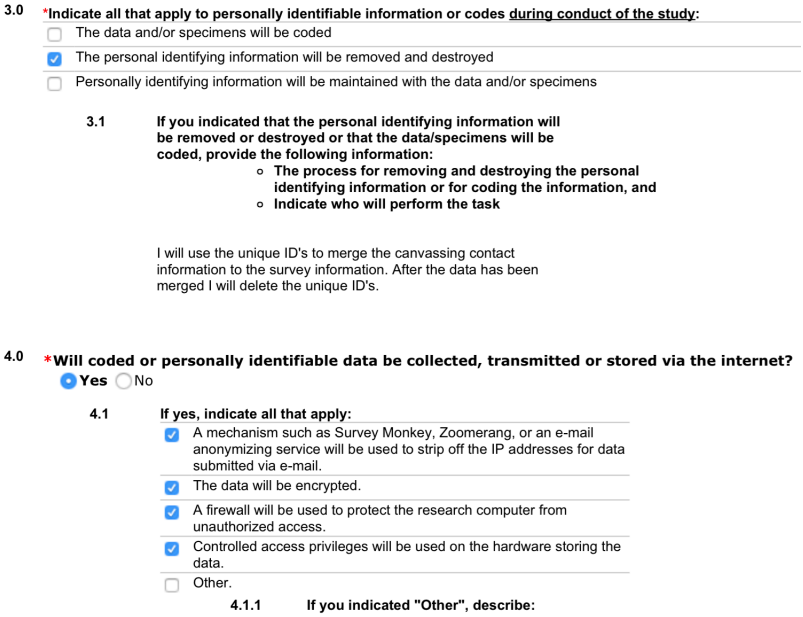
I take full responsibility for errors in the design, implementation, and data collection regard-

ing the field experiments and panel survey reported in LaCour and Green (2014). I also take full responsibility and apologize for misrepresenting survey incentives and funding in LaCour and Green (2014). In fact, I received a grant offer from the Williams Institute, but never accepted the funds, the LA GLBT received funding from the Evelyn and Walter Haas Jr. Fund., and the Ford Foundation grant did not exist. Instead, I raffled raffled Apple computers, tablets, and iPods to survey respondents as incentives. I located some of the receipts, Link here. Some of the raffle prizes were purchased for a previous experiment I conducted. I take full responsibility for destroying data in the interest of institutional requirements.

Claim II. Broockman et al. (2015) Did Not Present An Accurate Timeline And Failed To Disclose Their Critique To LaCour

- May 7, 2013: Pilot study begins with mail invitation to participate in online survey, pilot survey link here. LaCour sends IRB application to Professor David O. Sears for faculty assurances. Screenshot here. Final IRB application is here.
- May 2013: Survey conducted with email and mail invitations to participate in online survey. I set up an email account for the surveys, uclasurvey2013@gmail.com. I located records for ten survey respondents that called a google number I listed as an alternative number to call for questions about the survey. Link here. I tried to upload the surveys back into Qualtrics on May 22, 2015, unsuccessfully, but a screenshot shows the survey was created May 4, 2013. Screenshot here. I deleted the original surveys and emails, but I recreated a small example survey linked here. I recreated the email invitation Link here. A panel I set up through Qualtrics includes a screenshot here.
- July 29, 2013: LaCour submits a Pre-Analysis Plan on Experiments in Governance and Politics (EGAP) website for a replication study 2. Link here.
- November 2, 2013: LaCour emails replication data, files, and R code to Peter Aronow (co-author of Broockman et al. (2015), and Professor Green. Link here.
- December 18, 2013: Professor Green and LaCour have a lengthy email discussion of point three and point five in Broockman et al. (2015), "High Reliability of the Feeling Thermometer" and "Follow-up Waves exhibit different heaping than Wave 1." Link here.
- November 20, 2013: LaCour submits a grant application to the Williams Institute at UCLA to measure the 9-month follow-up results.

Figure 1: **LaCour's UCLA Institutional Review Board Application Pledging to Control Access Privileges, Remove, and Destroy Respondent's Personal Identifying Information**



3.0 *Indicate all that apply to personally identifiable information or codes during conduct of the study:

- ☐ The data and/or specimens will be coded
- ☑ The personal identifying information will be removed and destroyed
- ☐ Personally identifying information will be maintained with the data and/or specimens

3.1 If you indicated that the personal identifying information will be removed or destroyed or that the data/specimens will be coded, provide the following information:

- The process for removing and destroying the personal identifying information or for coding the information, and
- Indicate who will perform the task

I will use the unique ID's to merge the canvassing contact information to the survey information. After the data has been merged I will delete the unique ID's.

4.0 *Will coded or personally identifiable data be collected, transmitted or stored via the internet?

◉ Yes ○ No

4.1 If yes, indicate all that apply:

- ☑ A mechanism such as Survey Monkey, Zoomerang, or an e-mail anonymizing service will be used to strip off the IP addresses for data submitted via e-mail.
- ☑ The data will be encrypted.
- ☑ A firewall will be used to protect the research computer from unauthorized access.
- ☑ Controlled access privileges will be used on the hardware storing the data.
- ☐ Other.

4.1.1 If you indicated "Other", describe:

- March 2014: After collection of the 9-month follow-up data, the responses reported in LaCour and Green (2014) were obtained from Qualtrics. LaCour then destroyed the survey responses, as required by UCLA institutional policy and as agreed to on page 23 of the UCLA IRB Application, see Figure 1. LaCour states that, "The personal identifying information will be removed and destroyed....I will use the unique ID's to merge the canvassing contact information to the survey information. After the data has been merged I will delete the unique ID's." The data collected in Qualtrics was highly sensitive in nature: including respondents email address, geographic location, IP addresses, contact information, and social network referral contact information collected through respondent-driven sampling. Qualtrics representative Derek Johanson established conclusively to LaCour during phone conversations on May 20, 2015 that the online survey responses are no longer available. Matthew Amsden, CEO of ProofPilot (a survey firm that conducts randomized controlled trials with respondent-driven sampling) corroborates this information in conversations with LaCour on May 25, 2015, telling me that "medical RCT's are not run through Qualtrics because their data are not fully audible" and "ProofPilot does not work with UCLA because of their institutional policy of destroying identifying information."

- March 13, 2014: LaCour is offered a 2014 Williams Institute Grant, but does not turn in paperwork because data collection is complete. Link here. ➦

- August 4, 2014: LaCour and Green conduct canvasser study for reviewers. A screen-shot of pilot responses are shown at the Link here. ➦

- November 11, 2014: The data in LaCour and Green (2014), with personal identifiers removed, again as required by UCLA institutional policy, "Protecting Privacy and Maintaining Confidentiality in Research," available, states, "Protocols should be designed to minimize the need

to collect and maintain identifiable information about research participants. If possible, data should be collected anonymously or the identifiers should be removed and destroyed as soon as possible" have been publicly archived on the ICPSR website since November 11, 2014.

- December 20, 2014: Several of the points raised in Broockman et al (2015) are posted online at political science rumors. Link here. Professor Green emails LaCour to tell him to post the following to the website: "If you are using earlier versions of the gay equality canvassing dataset, please contact me at mjlacour@gmail.com to compete a Terms of Use agreement, stating that you will not use the data for commercial purposes and that you will not use the data to identify or contact respondents." Link here.

- Janurary 1, 2015: Broockman emails LaCour, "Questionnaire design advice question", using language closely resembling the critiques in Broockman et al. (2015). LaCour discusses the issue over the phone and provides Broockman with a version of the survey instrument used in LaCour and Green (2014), questions drawn from the CCAP 2012 survey. Link here.

- January 15, 2015: LaCour writes response letter to UCLA IRB regarding the timeline of submission. Screenshot here. Link here.

- January 22, 2015: LaCour receives letter from UCLA IRB. Link here.

- January 23, 2015: Broockman emails LaCour regarding online panel survey recruitment. Link here.

- March 2015: Brian Calfano contacts LaCour and Green about conducting a replication of LaCour and Green (2014) in light of an upcoming ballot initiative in Springfield, Missouri.

- March 15, 2015: Brian Calfano conducts randomization for the replication experiment partnering with the gay advocacy group, *Human Rights Campaign*. Calfano emails randomization list to Professor Green and LaCour. Link here.

- April 3, 2015: Broockman calls LaCour to discuss details regarding online panel survey recruitment.

- April 8, 2015: Brian Calfano's replication experiment reproduces the main finding in LaCour and Green (2014), with no survey needed. Professor Green participates in the April 8 email exchange about the replication experiment, Link here, indicating knowledge of the fact that a major finding from LaCour and Green (2014) has been replicated by an independent researcher (Brian Calfano), using a different partnering organization (Human Right Campaign), a different sample (Springfield, Missouri), unit of analysis (precinct), and outcome measure (actual voting behavior) – all safeguarded from the survey criticisms outlined in Broockman et al (2015). This exchange is significant because it contradicts Professor Green's public statements about the prospects for replication, Link here.

- April 9, 2015: Josh Kalla emails LaCour to meet at the 2015 Midwest Political Science Conference.

- April 22, 2015: Professor Green relinquishes authorship on the Calfano replication study, stating an email, "I suggested to Michael that you and he should write a short PS article about this – of course, I am happy to be involved, but I want you two to get the by-line."

- May 6, 2015. David Broockman and Joshua Kalla email LaCour asking for the recruitment letter and email used in LaCour and Green (2014), which LaCour provides.

- May 13, 2015. David Broockman and Joshua Kalla email LaCour asking to set up a meeting and state that they did a pilot study in Miami and the response rate is "lower than what they hoped." Link here. Broockman et al. (2015) state, "The response rate of the pilot study was notably lower than what LaCour and Green (2014) reported." Broockman and Kalla email LaCour a document with scattered notes about the pilot study, which they have since taken down. Broockman and Kalla seemed to have not followed the sampling protocol in LaCour and Green (2014), namely respondent-driven sampling and distribution through email and mail.

- May 17, 2015: Lynn Vavreck emails LaCour to set up a meeting. LaCour is not notified of the purpose or content of the meeting.

- May 18, 2015: Professor Vavreck meets with LaCour. LaCour is told there are "accusations" against him, but provided no specifics regarding the identity of the accusers or the substance of the allegations. LaCour is told to turn over his computer and personal survey account information to Lynn Vavreck and Jeff Lewis. LaCour refuses to turn over his computer, and his personal account information, as agreed to in his UCLA IRB requirement, "A firewall will be used to protect the research computer from unauthorized access" and "controlled access privileges will be used on the hardware storing the data." Professor Vavreck emails a colleague of LaCour's, Colleen Carpinella, with whom LaCour shared an survey account other than the one used to collect the data at issue here, to ask her for login and password information.

- May 18, 2015: David Broockman and Joshua Kalla cancel their scheduled meeting with LaCour about boosting "response rates." Stating, "Hey, sorry to do this but we're stuck on a call with someone else giving us advice on mail and i don't think will be ready in time for the 1p PT call today. and with your diss due let's just delay – sorry to have to reschedule." Link here.

- May 19, 2015: In his May 19, 2015, letter to Science's Gilbert Chin, (see Appendix in Broockman et al. (2015)), Professor Green writes, "He (LaCour) claimed he deleted the source file accidentally, but a Qualtrics service representative who examined the account and spoke with UCLA Political Science Department Chair Jeffrey Lewis reported to him that she found no evidence of such a deletion." This assertion by Professor Green is implausible, because when LaCour spoke with Qualtrics representative Derek Johanson on May 20, 2015, he had to supply a username on the account and the userid for the survey he had conducted. LaCour never shared these identifiers with anyone.

- May 19, 2015: In his May 19, 2015, retraction request, Professor Green writes, "On Tuesday, Professor Vavreck asked Michael LaCour for the contact information of survey respondents so that their participation in the survey could be verified, but he declined to furnish this information." LaCour notes that furnishing this information would be a direct violation of institutional policy, "Protecting Privacy and Maintaining Confidentiality in Research," ("data should be collected anonymously or the identifiers should be removed and destroyed as soon as possible..."). Note Professor Green's May 19, 2015, letter to Science's Gilbert Chin (see Appendix in Broockman et al. (2015)) contradicts his emailed instructions to LaCour, dated December 20, 2014, that those who wish to use the dataset must sign a Terms of Use agreement stating that "you will not use the data to identify or contact respondents."

LaCour notes that furnishing this information would be a direct violation of institutional policy, "Protecting Privacy and Maintaining Confidentiality in Research," ➦, stating, "data should be collected anonymously or the identifiers should be removed and destroyed as soon as possible..." Note Professor Green's retraction request memo contradicts his emailed instructions to LaCour, dated December 20, 2014, that those who wish to use the dataset must sign a Terms of Use agreement stating that "you will not use the data to identify or contact respondents."

- May 20-Present, 2015: Professor Green and David Broockman conduct media interviews publicizing their allegations. Acting on the advice of his attorney, Michael J. Deniro, LaCour refrains from speaking with anyone about the allegations.

Claim III. Broockman and Kalla's "Pilot Study" Is Not Evidence of A Failure To Replicate – The Study Has Not Been Made Available In Writing

Broockman et al's inability to describe their "pilot replication efforts" or produce a written document of exactly what they did, creates a barrier to evaluating the replication effort and thus has no bearing on the results in LaCour and Green (2014). At a minimum, the interested readers need to know what they did before assessing their claims about failure to replicate. I encourage readers to apply the same level of scrutiny applied to LaCour and Green (2014) to unpublished allegations of "failing to replicate."

LaCour notices two substantial deviations from the sampling protocol used in LaCour and Green (2014). First, Broockman and Kalla appear to have only sent mail invitations to take an online survey, while LaCour and Green (2014) sent mail and *email invitations* to participate in an online survey. Second, LaCour and Green (2014) used respondent-driven sampling or a snowball sample, stating, 'In an effort to impanel multiple voters per household, individuals were offered additional financial incentives to refer their friends and family to participate in the survey panel. If the referred individual was a registered voter who met the above criteria, he or she was invited to participate in the study....This recruitment procedure enables us to estimate how experimentally induced treatments diffuse through voters' social networks Bond et al. (2012); Gile and Handcock (2010). Response rates are opaque with respondent-driven sampling, because the sampling denominator is fluid. In sum, Broockman et al. did not follow the sampling protocol in LaCour and Green (2014), namely respondent-driven sampling and distribution through email and mail.

The failure of Broockman et al. to describe the sampling procedure utilized in their "replication efforts" makes it impossible to evaluate the study's scientific merit and thus has no bearing on LaCour and Green (2014).

Claim IV. Broockman et al (2015) Use The Incorrect CCAP Variable, Then Manipulate The Variable to Make It Look Similar To LaCour and Green (2014)

Broockman et al (2015) argue that flaws in the research data cast doubt on the findings in LaCour and Green (2014). Most relevant to these Responses is that 6 out of the 8 irregularities raised by Broockman, Stanford and Aronow (2015) are not mentioned in main text of LaCour and Green (2014). Six of the 8 points raised apply *solely* to Panel D of Table 13 in the Supplemental Appendix.

Comparison to Other Surveys

Broockman et al. (2015) state that their most compelling evidence of irregularity in LaCour and Green (2014) is that the data are similar to CCAP data. In their critique, Broockman et al. (2015) compare the baseline pre-treatment distribution of the feeling thermometer to five other surveys, shown in Table 1. I cannot discuss first two studies in Table 1 as I am unable to locate the details of the study, working paper, or pre-registration planning documents involved for them. The LaCour and Green (2014) survey was modeled after the CCAP survey and is the most similar across the dimensions listed in Table 1. Thus, it is not surprising that the CCAP data and LaCour and Green (2014) are the most similar – as they have the most in common – in that they are identical across survey mode, feeling thermometer measurement, and were collected closest in time. I now discuss the main critiques of LaCour and Green (2014).

Table 1: **Comparing LaCour and Green (2014) to Other Surveys**

Study	*Year*	N	Mode	Therm Measurement
Philadelphia	2015	142	Unknown	Unknown
Miami	2015	127	Unknown	Unknown
ANES	2000-2002-2004	1807	Face-to-Face	Interviewer Card
CCAP	2012	43,998	Online Survey	Sliding Scale
LaCour & Green (2014) S1	2013	9,507	Online Survey	Sliding Scale
LaCour & Green (2014) S2	2013	2,441	Online Survey	Sliding Scale

Figure 2: **Correction To Broockman et al. (2015) Feeling Thermometer Distribution**

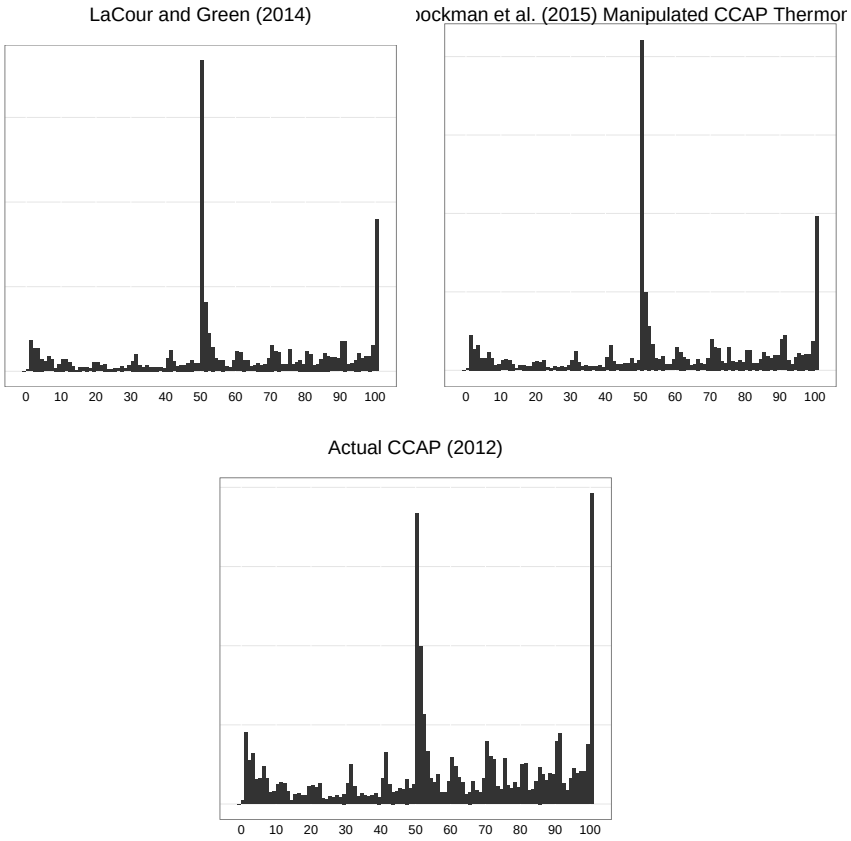
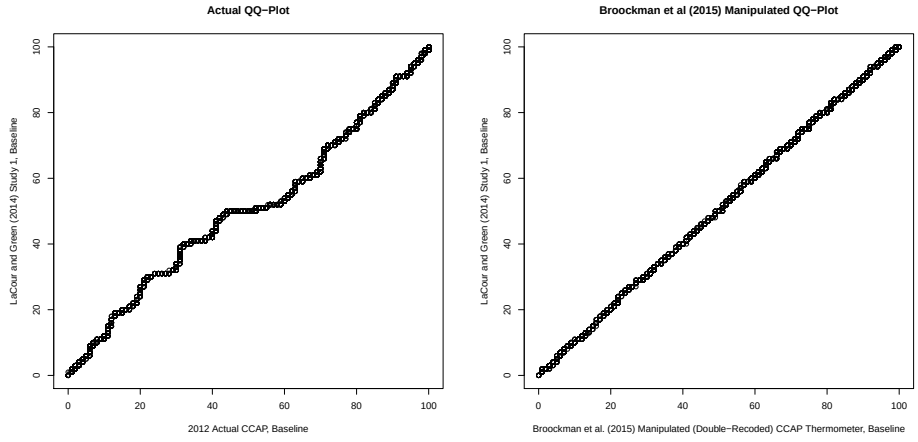


Figure 3: **Correction To Broockman et al. (2015) QQ-Plot**



Correction in R Code and Response to Critiques in Broockman et al. (2015)

Michael J. LaCour, Ph.D. Candidate, University of California, Los Angeles,
mlacour87@ucla.edu

May 29, 2015

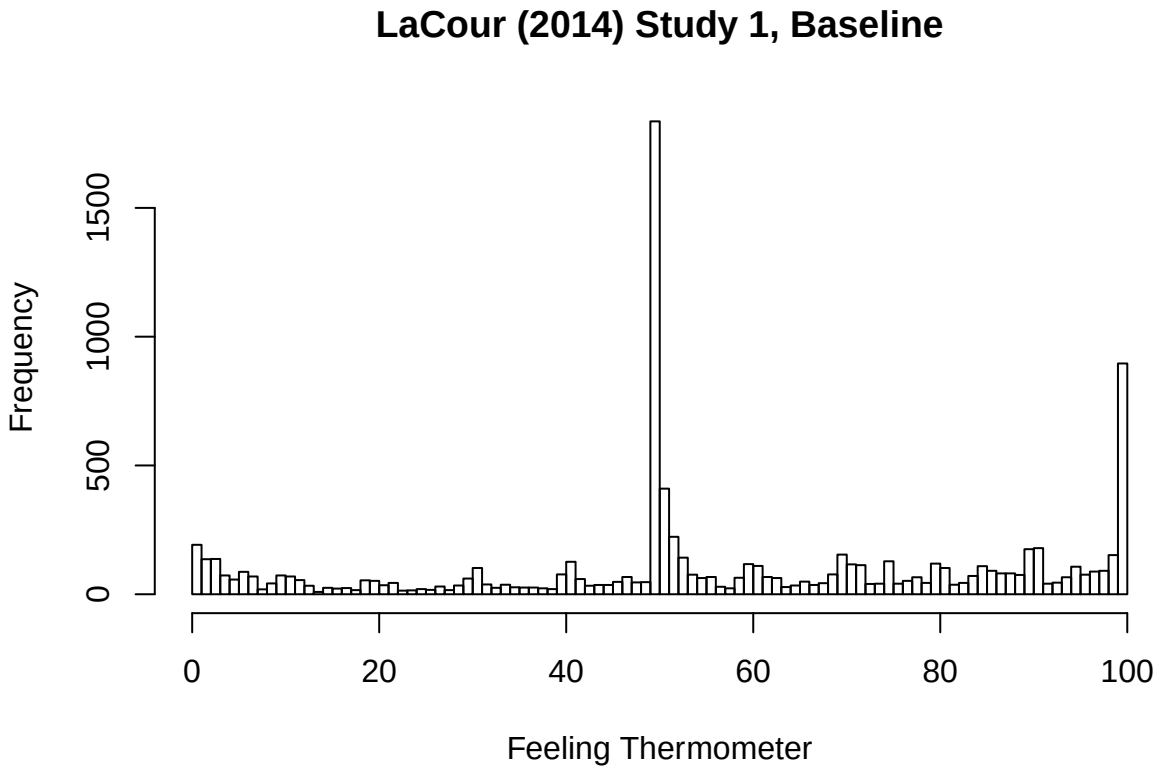
Load Packages and Read in Data

```
set.seed(63624) # from random.org
lacour <- read.csv('/Users/michaellacour/Dropbox/LaCour_Data/broockman_kalla_aronow_lg_irregularities_r
# https://www.openicpsr.org/repoEntity/show/24342
lacour.therm <- subset(lacour, wave == 1)$Therm_Level
lacour.reshaped <- reshape(lacour, v.names = c('Therm_Level',
      'Therm_Change', 'SSM_Level', 'SSM_Change'), idvar =
      'panelist_id', timevar = 'wave', direction = 'wide')
```

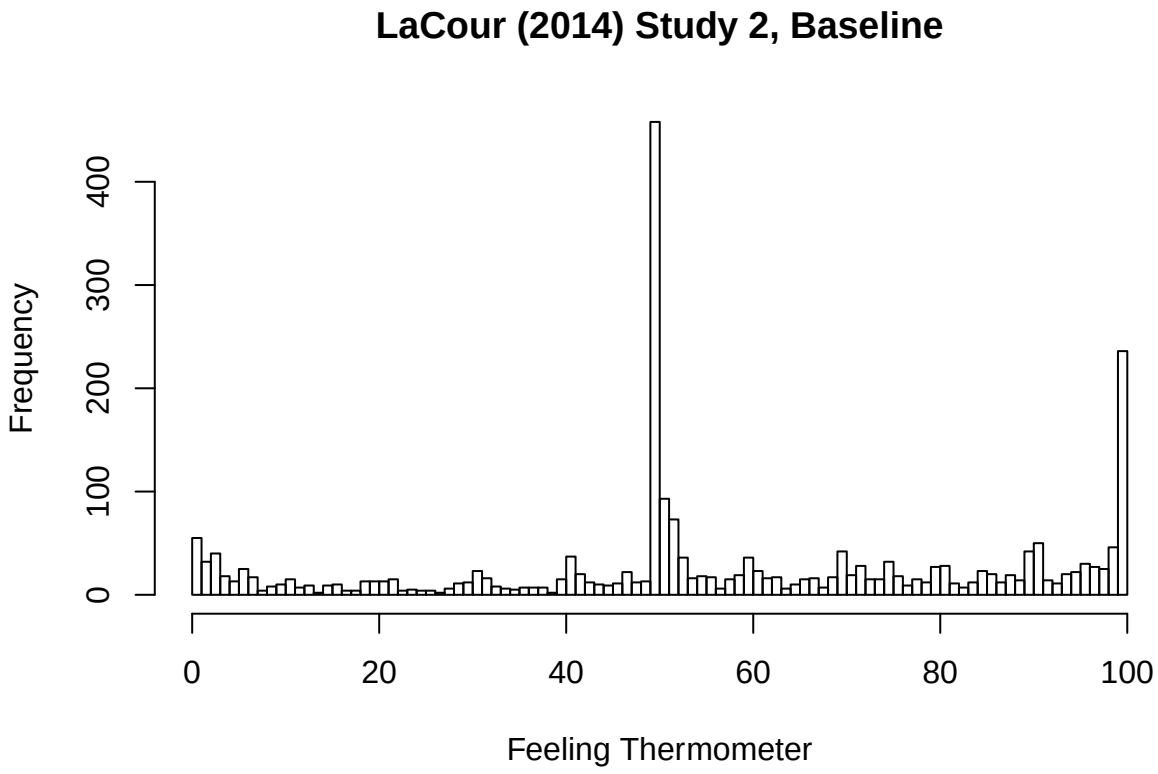
1. Similarity with 2012 CCAP Data

Most problematic for the authors' claim that the Feeling Thermometer from CCAP (2012) and LaCour and Green (2014) are "statistically indistinguishable" is the fact that the authors' selected the incorrect variable from CCAP (2012). Broockman et al (2015) use the CCAP vari- able,"gaytherm", which has already had the NA's recoded at 50 (which is not the conventional method) they manipulate the variable further, by recoding, "skips" as 50. The raw variable the authors intended to select is, "pp gays t", (pp stands for panelist profile), this is the baseline feeling thermometer on CCAP, with no recoding. Selecting the incorrect variable may have been an over- sight, but further manipulating that variable to make the distribution look more like LaCour and Green (2014) is curious and possibly intentional "error." The authors' appear to be manufacturing the CCAP distribution to make their case.

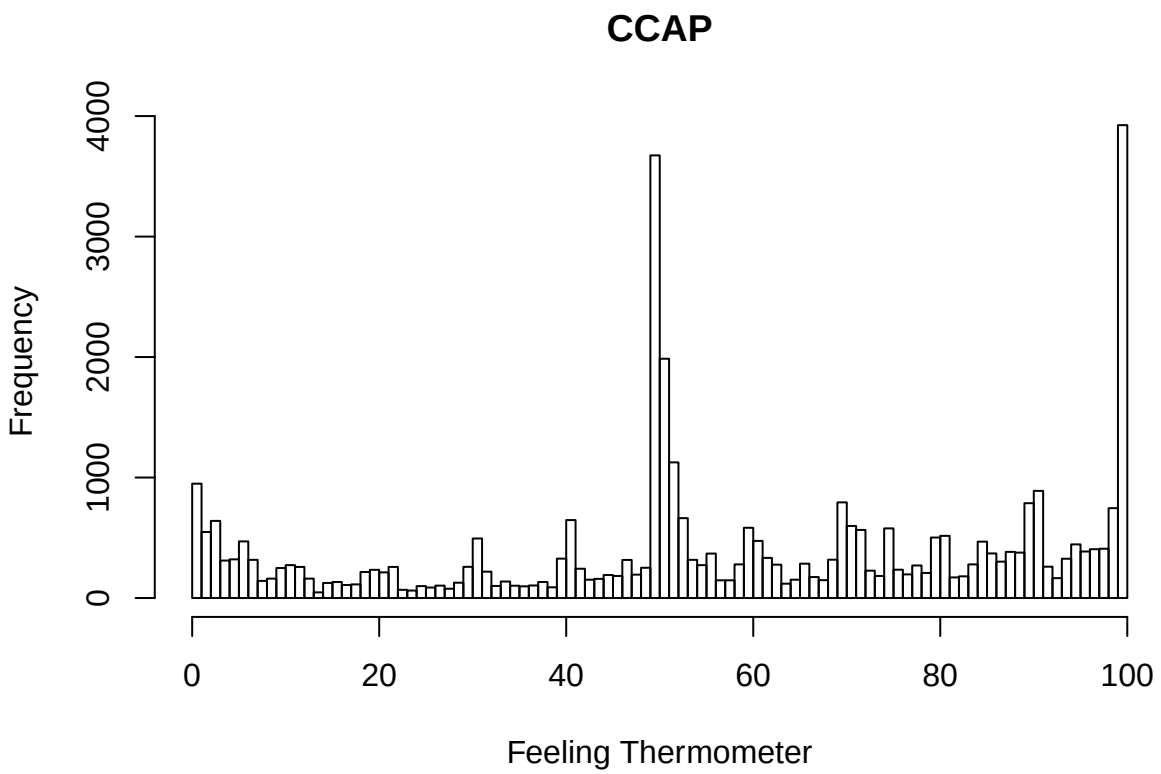
```
ccap.data <- read.table('/Users/michaellacour/Dropbox/LaCour_Data/broockman_kalla_aronow_lg_irregularit
# 1. LaCour note: Broockman et al. (2015) Select The Incorrect Variable,
# This Variable Has Already Been Recoded
# ccap.therm <- ccap.data$gaytherm

# 2. LaCour note: Broockman et al. (2015) state above that it is general
# convention to recode NA's to 50 on thermometer scales, there is no evidence to support this claim.
# ccap.therm[is.na(ccap.therm)] <- 50

# 3. LaCour note: The correct variable for the CCAP baseline
# Feeling Thermometer Toward Gays and Lesbiasns is ``pp_gays_t""
ccap.therm <- ccap.data$pp_gays_t[ccap.data$pp_gays_t<101] # omit skips and refusals

lacour.therm.study1 <- subset(lacour, wave == 1 & STUDY == "Study 1")$Therm_Level
lacour.therm.study2 <- subset(lacour, wave == 1 & STUDY == "Study 2")$Therm_Level

hist(lacour.therm.study1, breaks=101, xlab="Feeling Thermometer",
     main = "LaCour (2014) Study 1, Baseline")
```




```
hist(lacour.therm.study2, breaks=101, xlab="Feeling Thermometer",
     main = "LaCour (2014) Study 2, Baseline")
```



```r
hist(ccap.therm, breaks=101, main="CCAP", xlab="Feeling Thermometer")
```



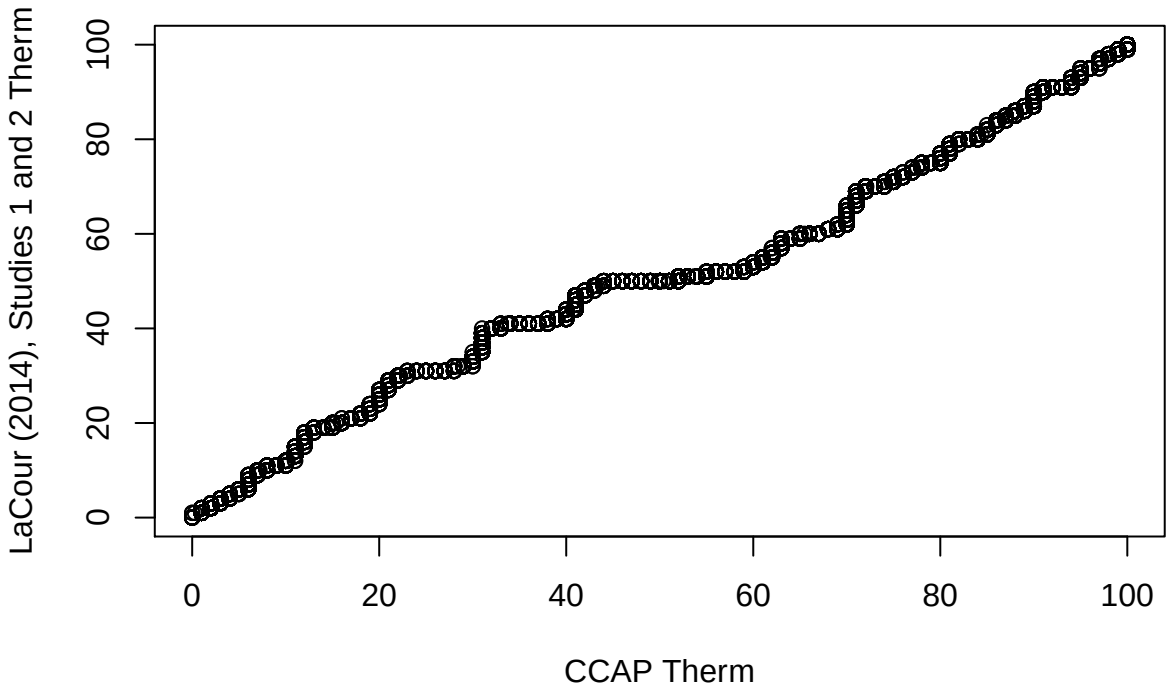
When the correct variable is used, the distributions between the CCAP thermometer and the LaCour and Green (2014) thermometer are quite different – notice the modal spike at 100 in the CCAP data, no such spike exists in LaCour and Green (2014).

When the correct variable is used, a Kolmogorov-Smirnov test finds differences between LaCour (2014) and the CCAP data.

```r
ks.test(lacour.therm, ccap.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  lacour.therm and ccap.therm
## D = 0.067086, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

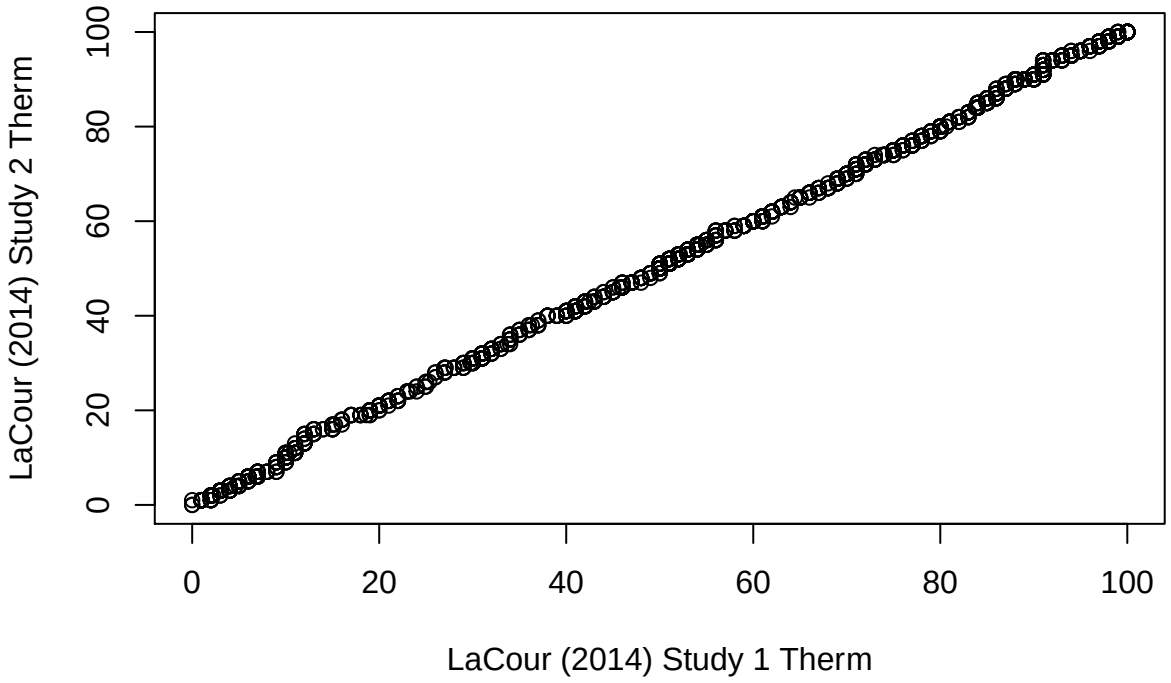
```r
qqplot(ccap.therm, lacour.therm, ylab = "LaCour (2014), Studies 1 and 2 Therm", xlab = "CCAP Therm")
```




```
ks.test(lacour.therm.study1, lacour.therm.study2)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  lacour.therm.study1 and lacour.therm.study2
## D = 0.013922, p-value = 0.8458
## alternative hypothesis: two-sided
```

```
qqplot(lacour.therm.study1, lacour.therm.study2, xlab = "LaCour (2014) Study 1 Therm",
       ylab = "LaCour (2014) Study 2 Therm")
```



Difference Between LaCour (2014) and Reference Distributions Using Correct Variable

```
#Philadelphia, 2015
philly <- read.csv('/Users/michaellacour/Dropbox/LaCour_Data/broockman_kalla_aronow_lg_irregularities_r
philly.nas.recoded <- philly
philly.nas.recoded[is.na(philly.nas.recoded)] <- 50

#Miami, 2015
miami <- read.csv('/Users/michaellacour/Dropbox/LaCour_Data/broockman_kalla_aronow_lg_irregularities_re
miami.nas.recoded <- miami
miami.nas.recoded[is.na(miami.nas.recoded)] <- 50

#ANES, 2000-2002-2004 Panel Study
library(foreign)
anes <- read.dta('/Users/michaellacour/Dropbox/LaCour_Data/broockman_kalla_aronow_lg_irregularities_repl
anes2000 <- anes$M001321
anes2002 <- anes$M025067
anes2004 <- anes$M045035

ks.test(anes2000, lacour.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  anes2000 and lacour.therm
## D = 0.24706, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```r
ks.test(anes2002, lacour.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  anes2002 and lacour.therm
## D = 0.27815, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```r
ks.test(anes2004, lacour.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  anes2004 and lacour.therm
## D = 0.25505, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```r
ks.test(philly, lacour.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  philly and lacour.therm
## D = 0.2757, p-value = 3.861e-08
## alternative hypothesis: two-sided
```

```r
ks.test(philly.nas.recoded, lacour.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  philly.nas.recoded and lacour.therm
## D = 0.2106, p-value = 7.846e-06
## alternative hypothesis: two-sided
```

```r
ks.test(miami, lacour.therm)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  miami and lacour.therm
## D = 0.16558, p-value = 0.01079
## alternative hypothesis: two-sided
```

```r
ks.test(miami.nas.recoded, lacour.therm)
```

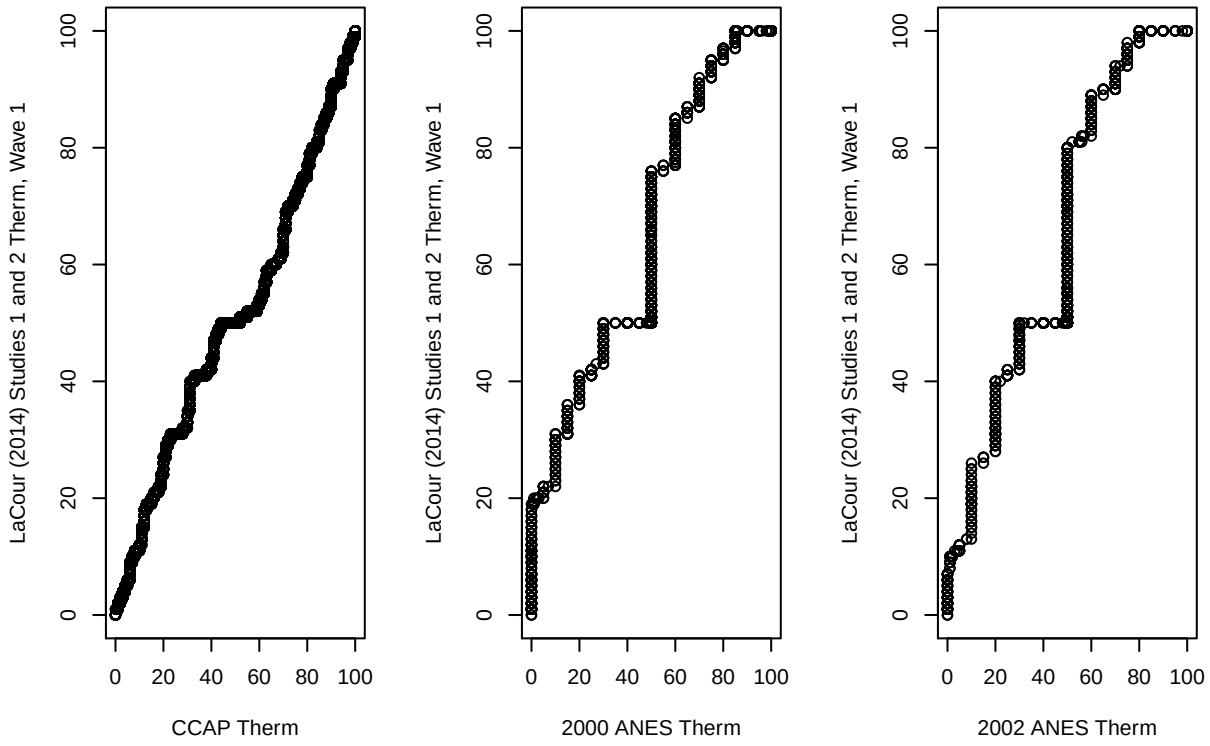
```
##
##  Two-sample Kolmogorov-Smirnov test
```

```
##
## data:  miami.nas.recoded and lacour.therm
## D = 0.24522, p-value = 5.462e-07
## alternative hypothesis: two-sided
```

```
## CCAP, California only
ccap.therm.ca <- ccap.therm[ccap.data$inputstate==6]
ks.test(ccap.therm.ca, lacour.therm)
```

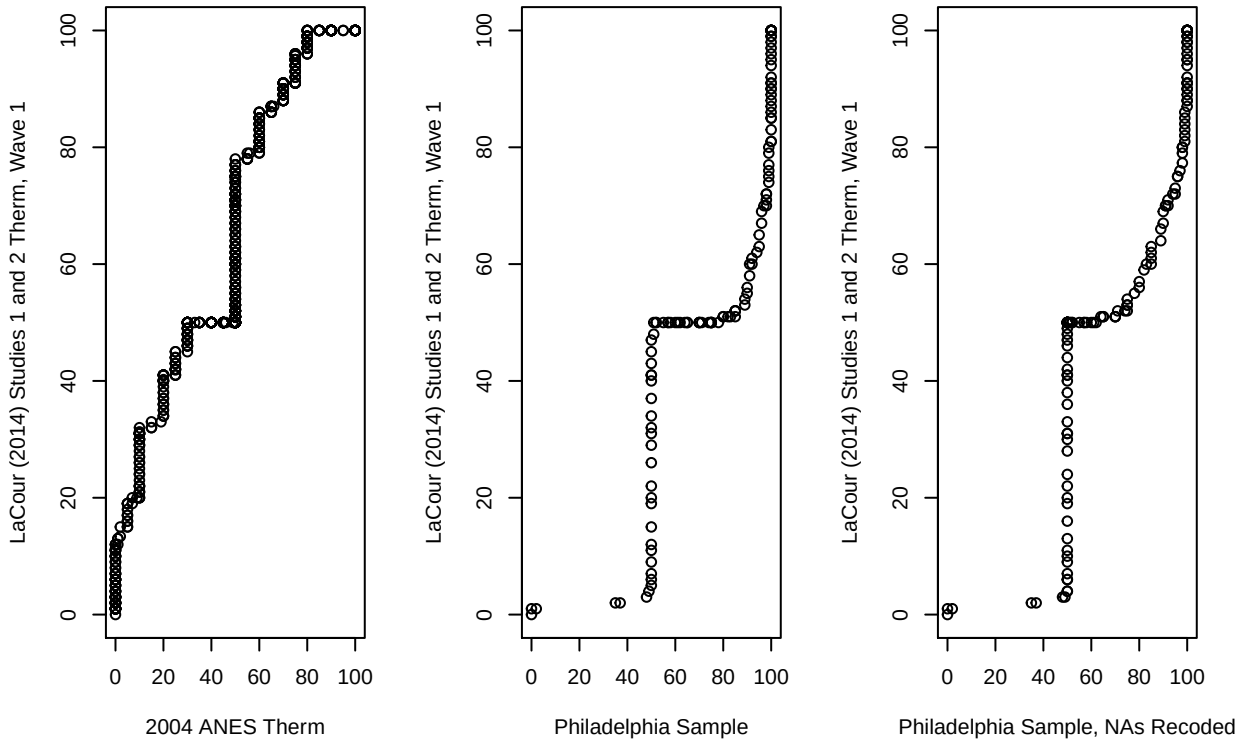
```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  ccap.therm.ca and lacour.therm
## D = 0.071129, p-value = 2.078e-13
## alternative hypothesis: two-sided
```

```
par(mfrow=c(1,3))
qqplot(ccap.therm, lacour.therm, xlab = "CCAP Therm",
       ylab = "LaCour (2014) Studies 1 and 2 Therm, Wave 1")
qqplot(anes2000, lacour.therm, xlab = "2000 ANES Therm",
       ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
qqplot(anes2002, lacour.therm, xlab = "2002 ANES Therm",
       ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
```

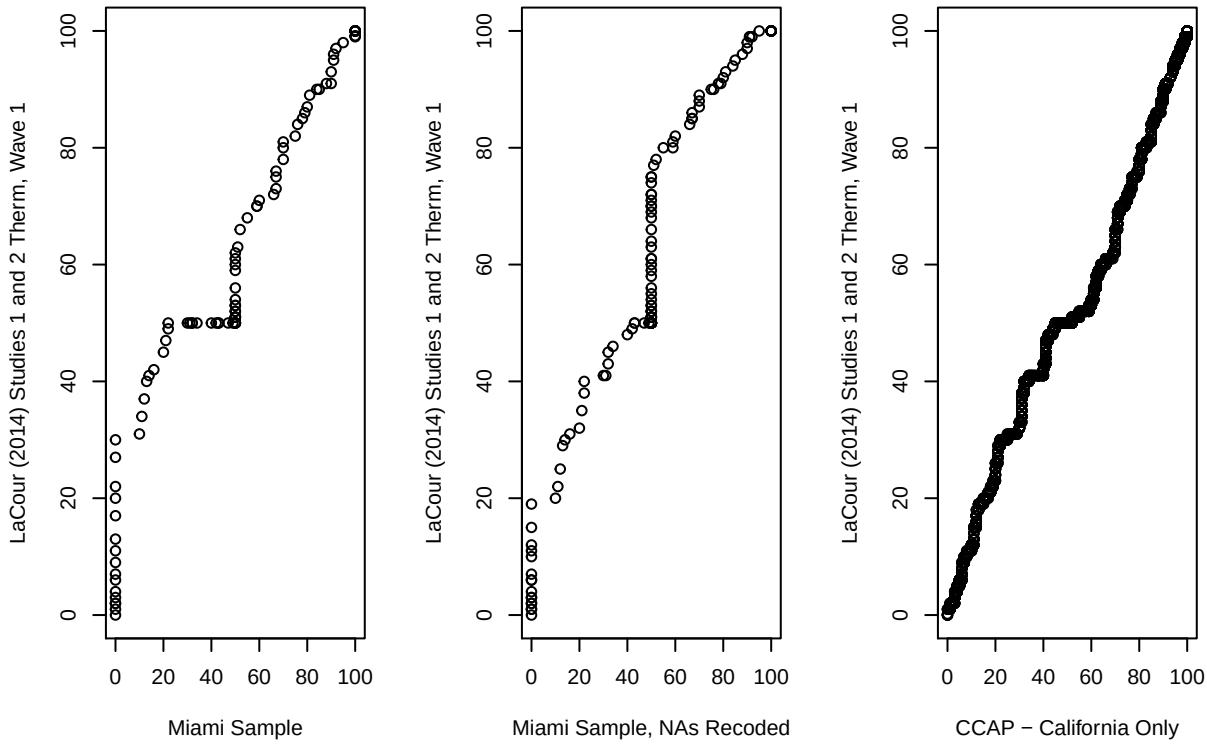



```
par(mfrow=c(1,3))
qqplot(anes2004, lacour.therm, xlab = "2004 ANES Therm",
       ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
qqplot(philly, lacour.therm, xlab = "Philadelphia Sample",
```

```
       ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
qqplot(philly.nas.recoded, lacour.therm, xlab = "Philadelphia Sample, NAs Recoded",
       ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
```




```
par(mfrow=c(1,3))
qqplot(miami, lacour.therm , xlab = "Miami Sample", ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
qqplot(miami.nas.recoded, lacour.therm , xlab = "Miami Sample, NAs Recoded",
       ylab="LaCour (2014) Studies 1 and 2 Therm, Wave 1")
qqplot(ccap.therm.ca, lacour.therm, xlab = "CCAP - California Only",
       ylab = "LaCour (2014) Studies 1 and 2 Therm, Wave 1")
```



2. High Reliability of the Feeling Thermometer

Contrary to Broockman et al. (2015)'s assertion that Feeling Thermometers are notoriously unreliable, feeling thermometers toward social groups typically exhibit high within-person correlations (~.70) as they are global measures of attitudes toward groups. As LaCour and Green (2014) note, "Given the high degree of over-time stability in attitudes toward social issues observed in longitudinal studies (Ansolabehere, Rodden and Snyder 2008) in tandem with theories stressing the persistence of social attitudes that are acquired early in life (Sears and Funk 1999) and that serve to justify existing hierarchies (Wellman 1993), it is not surprising that scholars summarizing the literature on prejudice tend to stress the difficulty of changing individuals' attitudes (Bobo, Kluegel and Smith 1997)."

Within-person correlations are higher when measurements are collected closely in time. Broockman et al. (2015) choose to restrict the analyses to thermometer measurements that were collected closest apart from one another. Because this is a randomized experiment, any panel bias we find, we would expect to be consistent across treatment groups. What is more common in panel studies is to have months elasped between survey waves.

Broockman et al (2015) criticize the level of within person correlation on the feeling thermometer ratings. Broockman et. al restrict their analysis to the outcome measure not presented in the main text of study 2, and focus on the within person correlations measured closely in time. We should expect higher within-person correlations when less time has elapsed between survey measurements. As Professor Green states in his email to LaCour a year prior to the study being published, "Is the reliability maybe TOO strong? Or maybe that's what you expect given a 2-week period, since the persistence would be raised to the 26th power to get to a 52 week (year) equivalent."

Broockman et al. (2015) notably omit the primary analyses reported in the main text of LaCour and Green (2014), the within person correlation between the baseline wave and the 9-month follow-up – a curious and possibly intentional "ommission." This long-term follow-up comparison is the most central to the LaCour and Green (2014) findings.

Lets look at the full correlation matrix.

```
lacour.study2.controlgroup <- subset(lacour.reshaped,
        STUDY == "Study 2" & Treatment_Assignment == "No Contact")
# LaCour Note: We need to look at the full correlations table, so let's include the test-retest correla
lacour.study2.therms <- lacour.study2.controlgroup[,c('Therm_Level.1',
        'Therm_Level.2', 'Therm_Level.3', 'Therm_Level.4','Therm_Level.7')]
cor(lacour.study2.therms, use = 'complete.obs')
```

```
##               Therm_Level.1 Therm_Level.2 Therm_Level.3 Therm_Level.4
## Therm_Level.1     1.0000000     0.9711471     0.9570280     0.9694395
## Therm_Level.2     0.9711471     1.0000000     0.9269083     0.9397330
## Therm_Level.3     0.9570280     0.9269083     1.0000000     0.9329295
## Therm_Level.4     0.9694395     0.9397330     0.9329295     1.0000000
## Therm_Level.7     0.8260524     0.7990056     0.7854040     0.8101274
##               Therm_Level.7
## Therm_Level.1     0.8260524
## Therm_Level.2     0.7990056
## Therm_Level.3     0.7854040
## Therm_Level.4     0.8101274
## Therm_Level.7     1.0000000
```

```
cor(lacour.study2.therms, use = 'pairwise.complete.obs')
```

```
##               Therm_Level.1 Therm_Level.2 Therm_Level.3 Therm_Level.4
## Therm_Level.1     1.0000000     0.9734449     0.9594085     0.9709017
## Therm_Level.2     0.9734449     1.0000000     0.9308287     0.9436621
## Therm_Level.3     0.9594085     0.9308287     1.0000000     0.9343249
## Therm_Level.4     0.9709017     0.9436621     0.9343249     1.0000000
## Therm_Level.7     0.8273645     0.8091700     0.7874356     0.8171640
##               Therm_Level.7
## Therm_Level.1     0.8273645
## Therm_Level.2     0.8091700
## Therm_Level.3     0.7874356
## Therm_Level.4     0.8171640
## Therm_Level.7     1.0000000
```

When we look at the full correlation matrix, we see within person correlations that decrease with time. The within person correlation between the baseline and ninth-month follow-up thermometer is 0.80 for study 1 and 0.82 for study 2. The within person correlation between the baseline and ninth-month follow-up support for same-sex marriage policy item is 0.80 for study 1 and 0.75 for study 2.

Changes over time are correlated if you include wave 7. Thermometer and Policy Item changes are also correlated over-time.

```
# LaCour Note: Include wave 7
lacour.study2.therm.changes <- lacour.study2.controlgroup[,c('Therm_Change.2',
      'Therm_Change.3', 'Therm_Change.4','Therm_Change.7')]
cor(lacour.study2.therm.changes, use = 'complete.obs')
```

```
##                Therm_Change.2 Therm_Change.3 Therm_Change.4 Therm_Change.7
## Therm_Change.2     1.00000000    -0.02418791    -0.02955057     0.02460002
## Therm_Change.3    -0.02418791     1.00000000     0.07187646     0.01910751
## Therm_Change.4    -0.02955057     0.07187646     1.00000000     0.06053096
## Therm_Change.7     0.02460002     0.01910751     0.06053096     1.00000000
```

```
cor(lacour.study2.therm.changes, use = 'pairwise.complete.obs')
```

```
##                Therm_Change.2 Therm_Change.3 Therm_Change.4 Therm_Change.7
## Therm_Change.2    1.000000000   -0.003088789   -0.009142816     0.04134108
## Therm_Change.3   -0.003088789    1.000000000    0.065957305     0.02377887
## Therm_Change.4   -0.009142816    0.065957305    1.000000000     0.11131035
## Therm_Change.7    0.041341077    0.023778868    0.111310349     1.00000000
```

```
lacour.study2.outcome.changes <- lacour.study2.controlgroup[,c('Therm_Change.7','SSM_Change.7')]
cor(lacour.study2.outcome.changes, use = 'complete.obs')
```

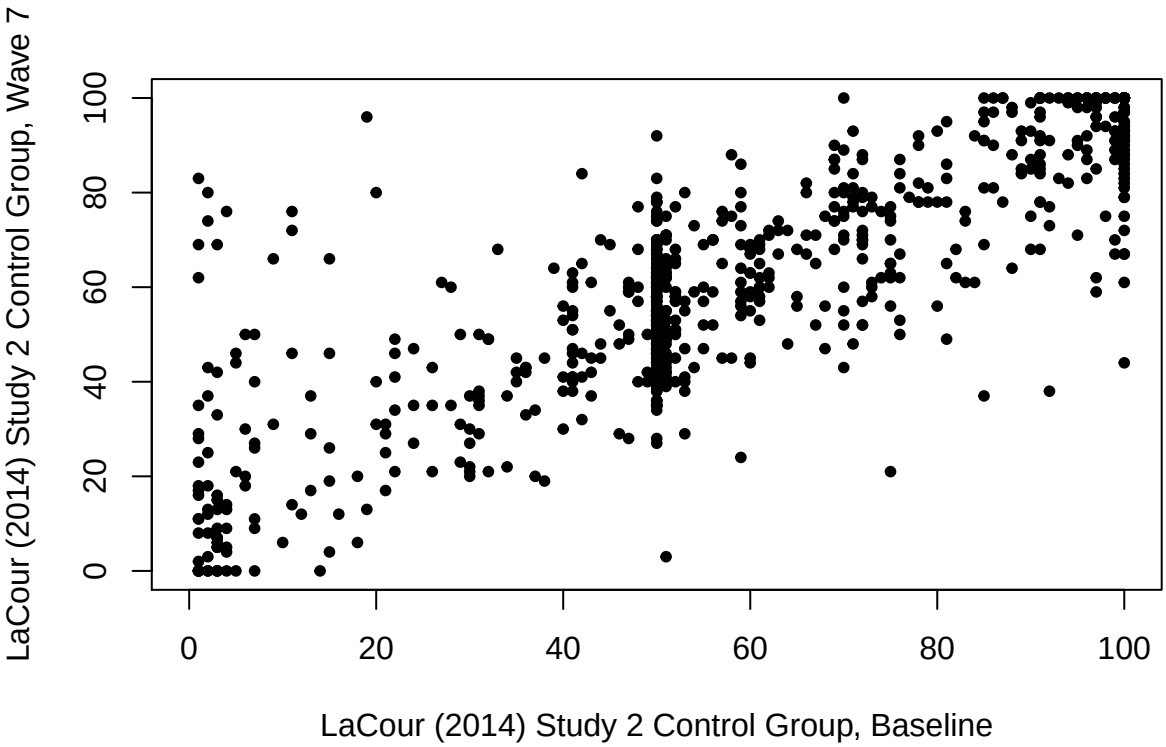
```
##                Therm_Change.7 SSM_Change.7
## Therm_Change.7     1.00000000   0.08411706
## SSM_Change.7       0.08411706   1.00000000
```

3. Distributions of Changes in Feeling Thermometers Highly Regular

I agree with Broockman et al. that the normally distributed errors in the study 2 waves 2, 3, and 4 presented in Table 13 of the supplemental appendix in LaCour and Green (2014) are irregular, and I cannot rule out the possibility that I mistakenly mixed up simulated thermometer waves generated to gauge statistical power, prior to the launch of study 2. As correspondence to Professor Green indicates, there were many issues with the thermometer scale in study 1. Note that LaCour and Green did not present the early waves of the thermometer from Study 1 and no thermometers for Study 2 in the main text, so Broockman et al.'s point regarding the distribution of changes in thermometers in the early waves is largely moot – these were not presented in the paper, only in Panel C and D of Table 13 of the supplemental appendix in LaCour and Green (2014).

Finally, I want to focus on the distribution of feeling thermometer changes between pre-treatment measurement and the ninth-month follow-up presented in LaCour and Green (2014).

```
# LaCour Note: Again let's not restrict the analysis to just the data patterns that conform to our hypo
plot(lacour.study2.controlgroup$Therm_Level.1,
     lacour.study2.controlgroup$Therm_Level.7, pch=20,
     xlab = "LaCour (2014) Study 2 Control Group, Baseline",
     ylab = "LaCour (2014) Study 2 Control Group, Wave 7")
```



The pattern is not normally distributed.

4. Follow-up Waves exhibit different heaping than Wave 1

One explanation for this "heaping"" is a panel effect, in that taking a survey repeatedly affects respondents answers. In the absence of a counterfactual group with which to compare, the heaping criticism is speculative.

5. Endogenous takeup of treatment appears completely random

Broockman et al. (2015) criticize the fact that there are no significant differences in LaCour and Green (2014) between door-answering behavior and attitudes toward gays and lesbians on a feeling thermometer scale. Broockman et al. (2015) cite a get-out-the-vote experiment by Professor Green as evidence that we should expect a relationship. In the absence of evidence there is a expect a relationship between attitudes toward gay men and lesbians as measured on a feeling thermometer scale and door-answering behavior this criticism is meritless. LaCour and Green (2014), addressed this issue in their paper, writing, "We also verified that contact rates are not significantly predicted by the interaction between assigned treatment and the covariates listed in Table S5 and obtained an F-statistic of 0.985, with a p-value of 0.547. By the same token, subjects in households that were not contacted by canvassers express opinions on the key outcome measures that do not vary significantly according to whether subjects were assigned to treatment or placebo conditions. These insignificant p-values suggest that canvassers did not make a special effort to contact subjects with

References

LaCour, Michael J. 2014-12-21. Political Persuasion and Attitude Change Study: The Los Angeles Longitudinal Field Experiments, 2013-2014: LaCour & Green (2014) Replication Files. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. http://doi.org/10.3886/E24287V1.

LaCour, Michael J. and Donald P. Green. 2014. When contact changes minds: An experiment on transmission of support for gay equality. Science 346(6215): 1366-1369.

Tesler, Michael. 2014. "Replication data for: Priming Predispositions and Changing Policy Positions", http://dx.doi.org/10.7910/DVN/26721, Harvard Dataverse, V2

Replication Experiment Reproduces Finding in LaCour & Green (2014) – No Survey Needed

The findings reported in LaCour & Green (2014) are replicated with a new field experiment conducted by Brian Calfano in partnership with the *Human Rights Campaign*. This field experiment took place in Missouri in April 2015, randomizing contact with canvassers advocating for gay equality at the precinct level. Because contact with canvassers advocating for gay equality was randomized at the precinct level, (and outcomes were assessed with vote share for the gay equality ballot measure, no public opinion survey was needed. Randomized contact with canvassers produced a 4.78% point increase in support for a gay equality ballot measure at in the Springfield, Missouri local election, as shown in the Table below constructed using data Brian Calfano reported to Professor Green and me. The treatment effect in LaCour and Green (2014) after 3 days was a 6% increase in support for same-sex marriage.

Changing Voting Behavior in Missouri on Gay Rights: A Randomized Field Experiment

Treatment Assignment	*N*	Precinct Vote Share
Same-Sex Marriage Script by Gay Canvasser	13	43.06%
No Contact	14	47.81%
Average Treatment Effect		4.78%

References

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam DI. Kramer, Cameron Marlow, Jaime E. Settle and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489(7415):295–298.

Broockman, David, GSB Stanford and Peter Aronow. 2015. "Irregularities in LaCour (2014).".

Gile, Krista J. and Mark S. Handcock. 2010. "Respondent-Driven Sampling: An Assessment of Current Methdology." *Sociological Methodology* 40(1):285–327.

King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics* 28(03):444–452.