

Online Appendix:  
Images that Matter: Online Protests and the  
Mobilizing Role of Pictures

Andreu Casas  
University of Washington

Nora Webb Williams  
University of Washington

## A Appendix: Image Labeling Procedures and Summary Statistics Table

This appendix presents the questions used to manually label images (Table 1, two sample images with their labeling scores (Figures 1 and 2, and a summary statistics table for the key model covariates (Table 2). For our top 1000 images, two undergraduate labelers tagged images as being of a street protest or not. For the remaining images, we labeled an image of being as a street protest if an annotator indicated that there were more than ten persons present in the image and at least one protest sign or slogan.

Table 1: Labeling Form for Images

Variable	Question	Options
<i>sign_slogan</i>	Is there a protest sign or slogan in the picture? (e.g. Black Lives Matter; Hands Up, Don't Shoot!)	(0,1)
<i>symbol</i>	Is there any symbol in the picture? (e.g. flags, logos)	(0,1)
<i>anger</i>	How much anger does the image incite in you? If none, select 0.	(0, 1, ..., 10)
<i>fear</i>	How much anger does the image incite in you? If none, select 0.	(0, 1, ..., 10)
<i>disgust</i>	How much anger disgust the image incite in you? If none, select 0.	(0, 1, ..., 10)
<i>sadness</i>	How much sadness does the image incite in you? If none, select 0.	(0, 1, ..., 10)
<i>enthusiasm</i>	How much enthusiasm does the image incite in you? If none, select 0.	(0, 1, ..., 10)
<i>people</i>	By your guess, how many people are in the picture? Leave blank if no people.	(number)

Figure 1: The Most Tweeted Image During the April 14 Protest



Research staff labeled this image as not being a protest, and having no symbols. On the emotions, the average scores were: anger: 2, fear: 1, disgust: 2, sadness: 3, enthusiasm: 1.

Figure 2: The Fifth Most Tweeted Image During the April 14 Protest



Research staff labeled this image as being a protest, but not having any symbols. On the emotions, the average scores were: anger: 2, fear: 1.5, disgust: 1, sadness: 1, enthusiasm: 2.5.

Table 2: Key Variable Summary Statistics

Variable	Minimum	Maximum	Mean	SD
Image	0	1	0.19	0.40
Protest*	0	1	0.03	0.18
Symbol*	0	1	0.02	0.14
Anger*	0	10	1.75	2.70
Disgust*	0	10	1.74	2.79
Enthusiasm*	0	10	1.51	2.41
Fear*	0	10	1.05	2.04
Sadness*	0	10	1.93	2.84
Number of Followers	0	5540545	4692.23	59339.65
Number of Friends	0	350644	1425.84	5198.41
Previous Tweets	0	1815	54.29	179.21

\*For these variables we provide summary statistics for the messages that have an image. The statistics for the other variables are based on the whole sample of original messages.

## B Appendix: Interrater Reliability, Evidence of Stable Emotions Labeling, and Turker Demographics

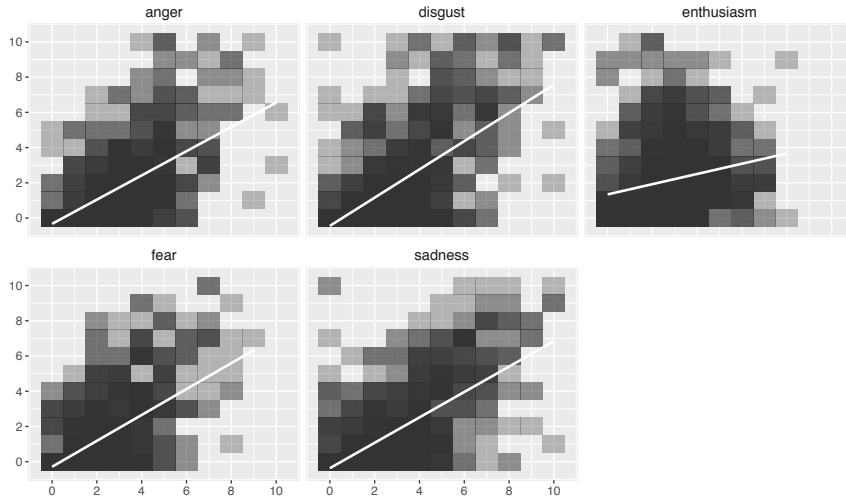
In this appendix we address interrater reliability concerns. Table 3 presents Cohen’s Kappa or one-way intraclass correlation coefficients for each of the seven independent variables of interest. These values were generated based on the ratings generated by our two undergraduate coders on the top 1000 most-tweeted images. The raters had generally good agreement, with the lowest agreement for the symbol and enthusiasm labels.

Table 3: Interrater Reliability Measures

Variable	Interrater Reliability	Cohen’s Kappa or one-way intraclass correlation coefficient (ICC)
Symbol	0.23	Kappa
Protest	0.78	Kappa
Anger	0.46	ICC
Fear	0.48	ICC
Disgust	0.55	ICC
Sadness	0.54	ICC
Enthusiasm	0.19	ICC

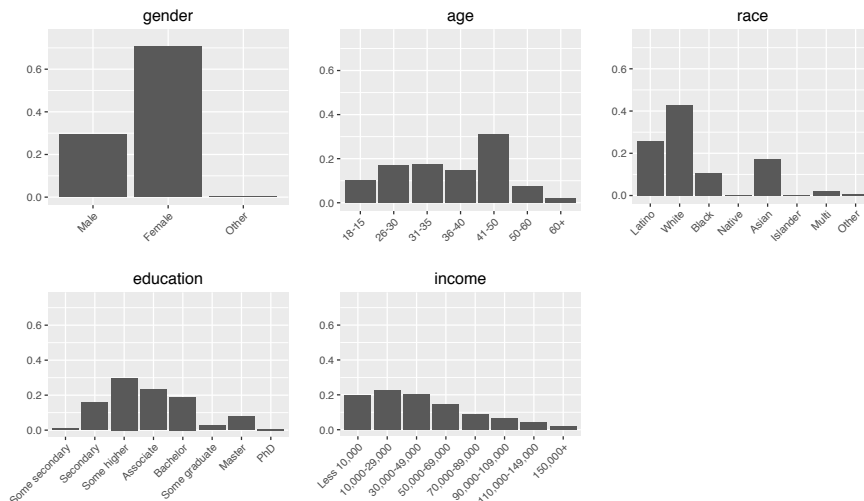
The emotions portion of labeling is particularly important for our purposes. Although emotions are subjective, and we expected a wide range of emotional responses, on average the top 1,000 images (which account for more than 50% of the messages with images) triggered very similar emotions to different people. When modeling the data we give each unique image a single score per emotion (on a 0-10 point scale). Each image has been labeled by five different people and for each image and emotion we averaged the scores given by the five individuals. In a first iteration two research assistants labeled the top 1,000 images. We had weekly meetings with them during the labeling process, they were aware of the substance and goals of the project, and they helped us improve other parts of the labeling form. Figure 3 shows the correlation between the emotional scores given by the two research assistants to the same images. The correlation is very strong in all the cases. The *enthusiasm* score shows the weakest correlation but it is still strong.

Figure 3: Correlation between the emotion scores given by 2 research assistants to the same images (top 1,000 images)



In a second iteration we used Mechanical Turk (MT) to label the top 1,000 images three more times. We decided to do so not only to have more emotion labels per image but also to get scores from people with different backgrounds, since our two research assistants were both undergraduate students, male, and white. We set it up so that only MT workers from the United States could participate and we also set it up so that workers could label more than one image but never the same image twice. Figure 4 presents summary statistics for the MT workers that participated in the labeling process. The figure shows how workers had a very diverse background.

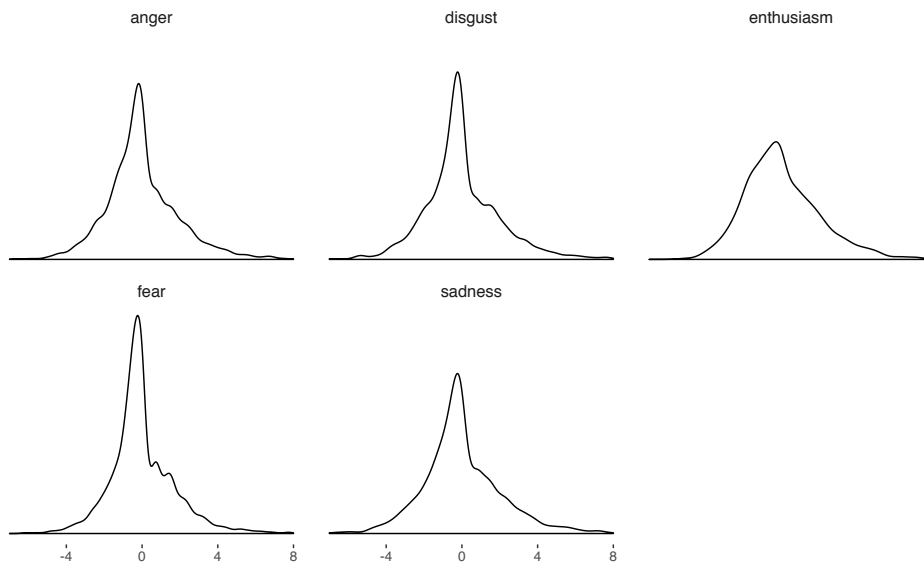
Figure 4: Summary of the socio demographic characteristics of Mechanical Turk workers that labeled the images



To see whether people gave very different emotion scores to the same images, for each image and emotion we calculated the average score given by the five annotators (the two research assistants and three people from MT), and then for each of the five scores we calculated the difference between them and the mean score. Figure 5 shows again that the same images triggered very similar emotions in different people, with most individual scores being around 1 or 2 points from

the five-scores mean. *Enthusiasm* is again the emotion that presents the most variation.

Figure 5: Distribution of the difference between emotions scores for the same top 1,000 images



## C Appendix: Regression Analysis Results Table

	Attention: BLM + ShutdownA14 (Number of Retweets)		Diffusion: ShutdownA14 (Retweets by New A14 Users)	
	Basic (1)	Mechanisms (3)	Basic (2)	Mechanisms (4)
Image	1.690*** (0.032)		1.513*** (0.097)	
Protest		0.996*** (0.062)		1.123*** (0.139)
Symbol		0.165** (0.072)		0.217 (0.193)
Anger		-0.026 (0.017)		-0.001 (0.048)
Fear		0.072*** (0.016)		0.121*** (0.045)
Disgust		0.046*** (0.016)		0.039 (0.044)
Sadness		-0.005 (0.012)		-0.171*** (0.033)
Enthusiasm		0.086*** (0.010)		0.105*** (0.030)
Number of Followers	0.00003*** (0.00000)	0.00002*** (0.00000)	0.00002*** (0.00000)	0.00000*** (0.00000)
Previous Tweets	-0.001*** (0.0001)	-0.001*** (0.0002)	-0.003*** (0.001)	-0.006*** (0.001)
Number of Friends	0.00004*** (0.00000)	0.00004*** (0.00001)	0.0001*** (0.00001)	0.00002 (0.00003)
Time(t2)	-0.377*** (0.053)	-0.871*** (0.099)	-0.998*** (0.214)	-1.558*** (0.362)
Time(t3)	-0.411*** (0.061)	-0.723*** (0.116)	1.102*** (0.242)	-0.796* (0.442)
Time(t4)	-0.293*** (0.043)	-0.593*** (0.083)	-0.263 (0.195)	-0.500 (0.328)
Time(t5)	-0.272*** (0.040)	-0.565*** (0.071)	0.329** (0.141)	0.108 (0.211)
Time(t6)	0.106*** (0.038)	-0.328*** (0.068)	1.161*** (0.121)	0.485** (0.192)
Constant	-0.319*** (0.030)	1.234*** (0.058)	-1.346*** (0.105)	0.069 (0.184)
Original Tweets (n)	49,345	8,706	7,502	2,078
Log Likelihood	-56,591.870	-18,147.520	-6,393.248	-2,814.994
Akaike Inf. Crit.	113,203.700	36,327.040	12,806.500	5,661.988

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01