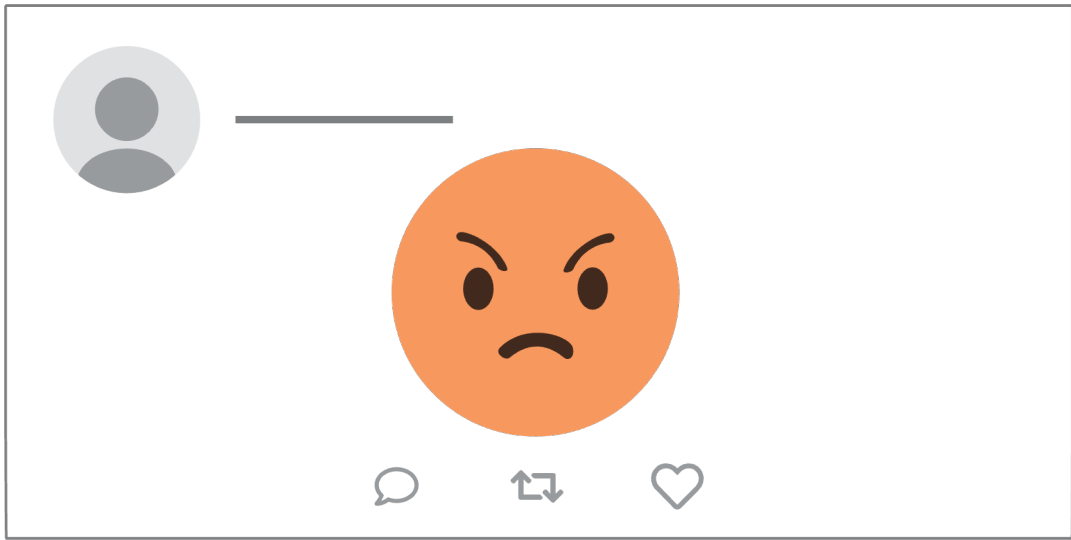
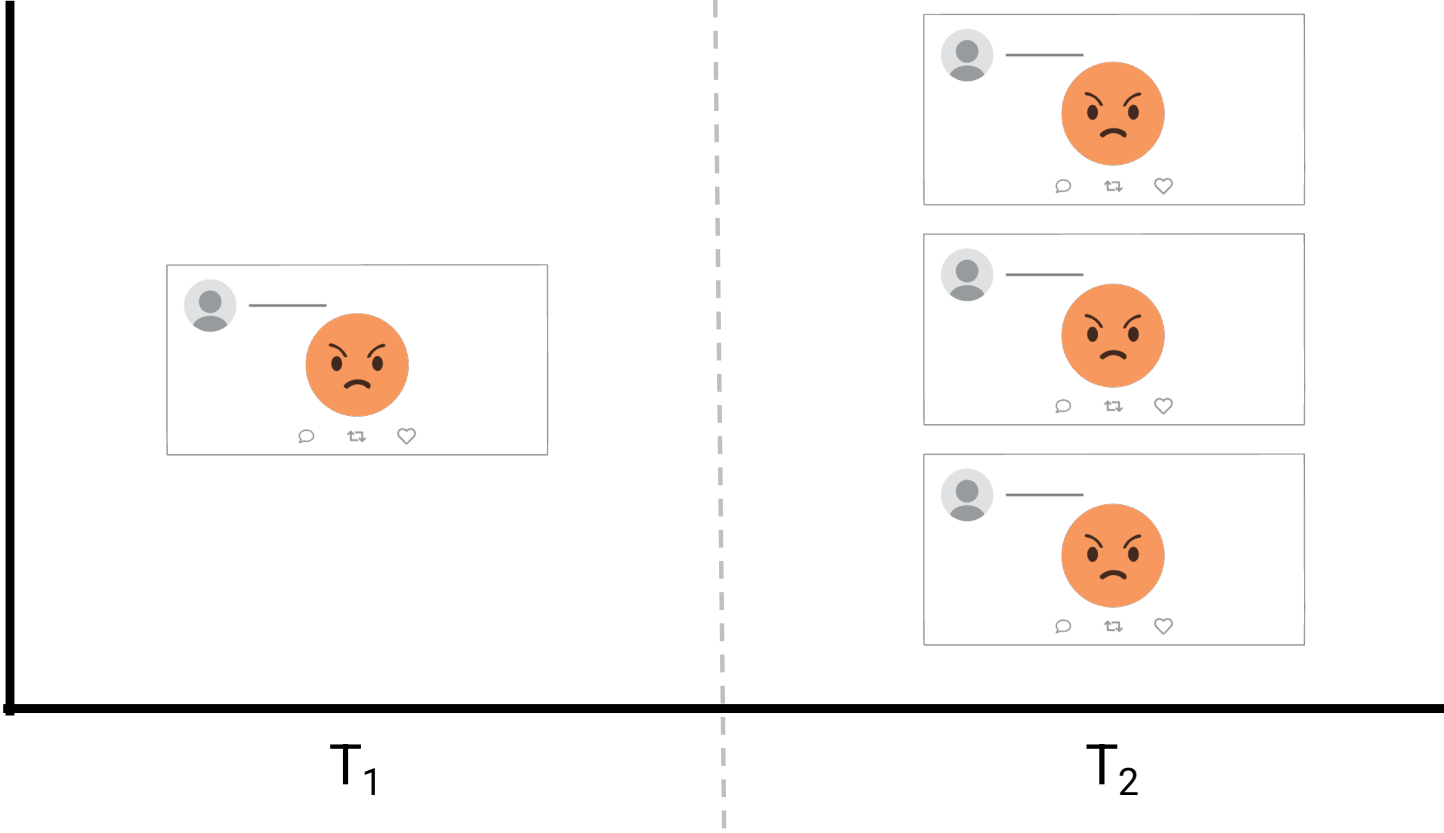


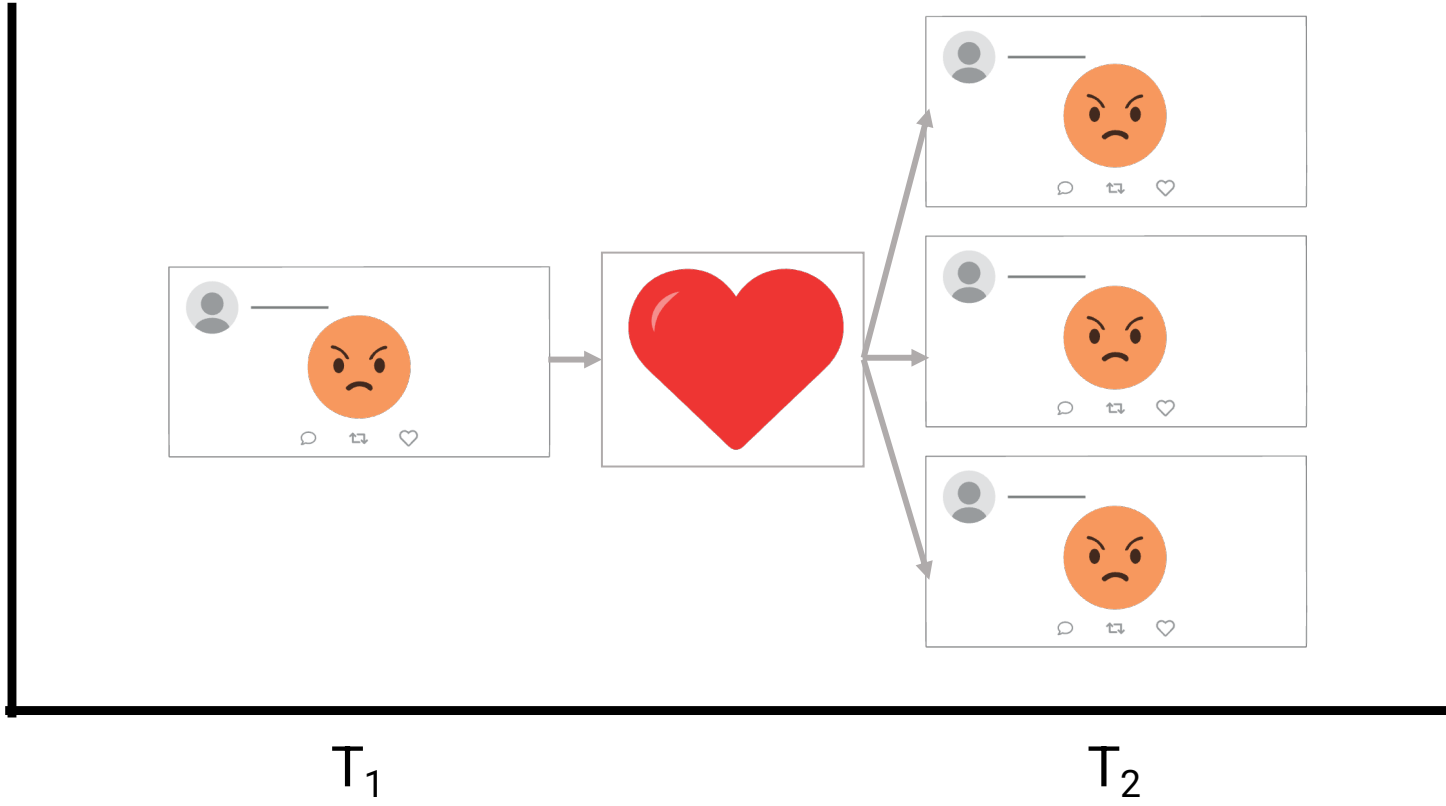
Multilevel models and Generalized estimating equations

Killian

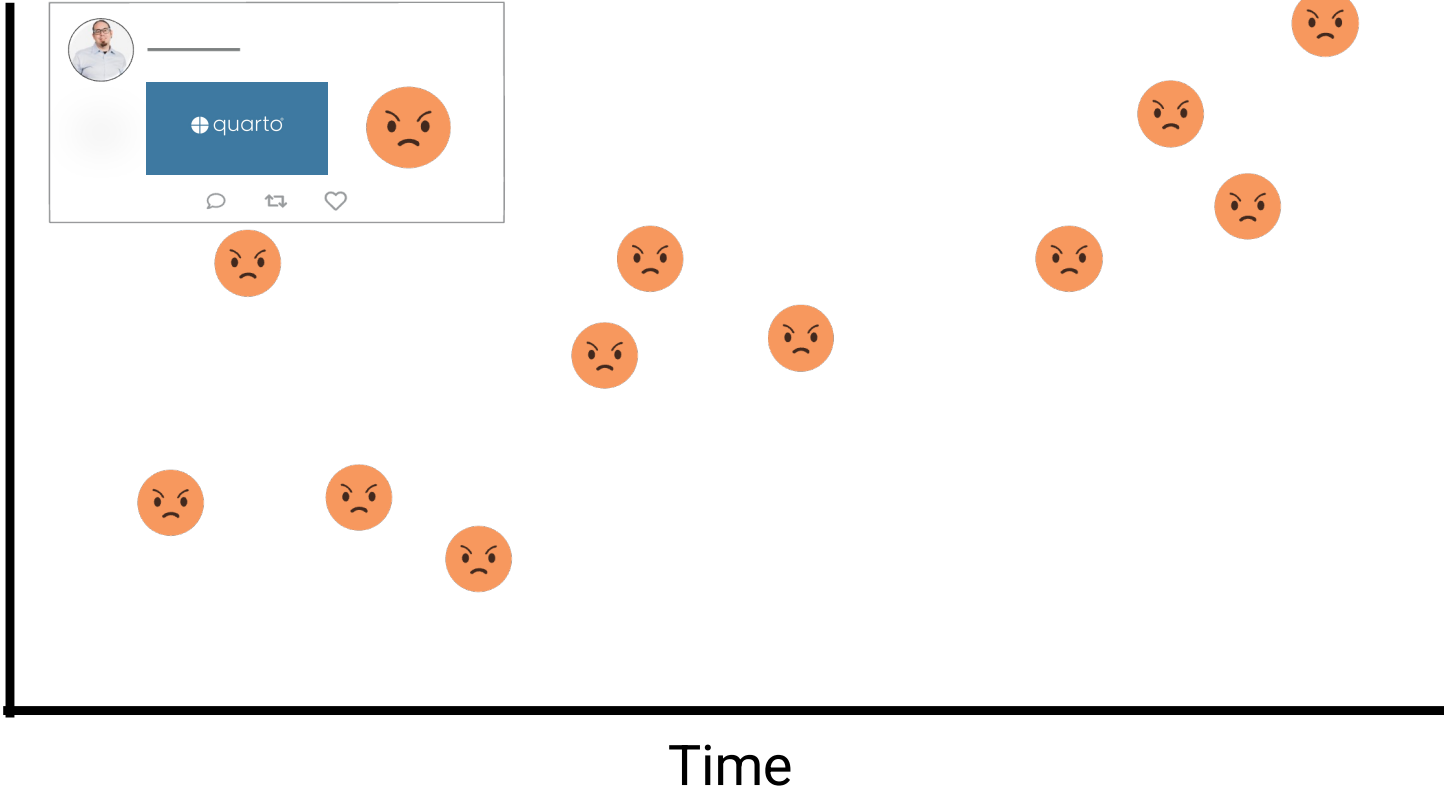
Apr 24

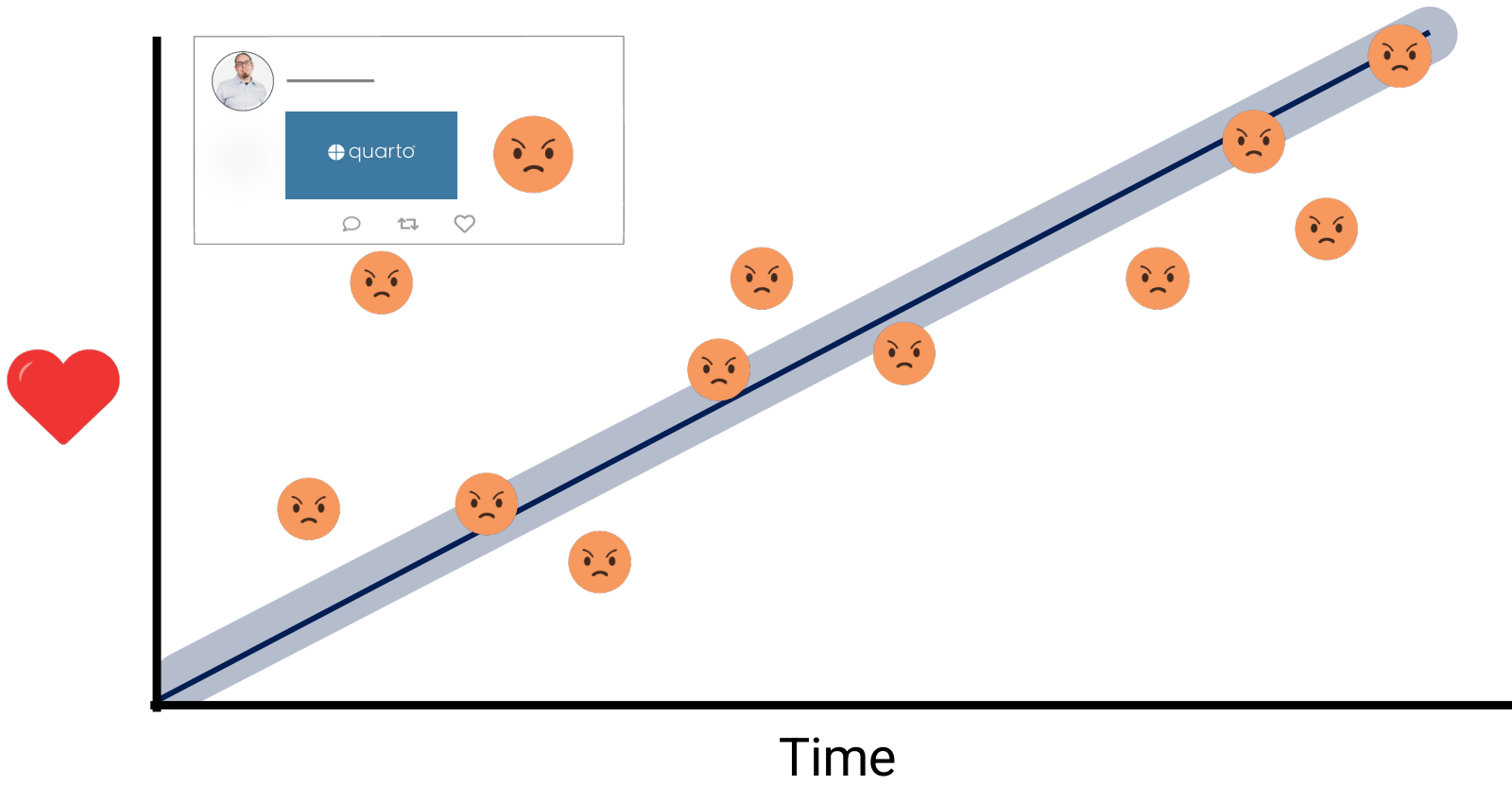


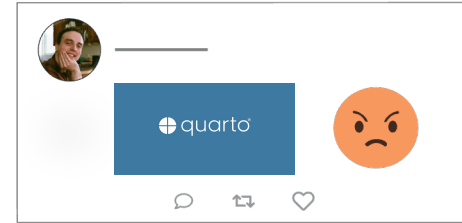
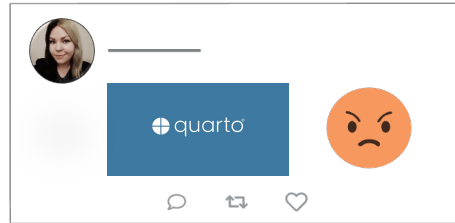


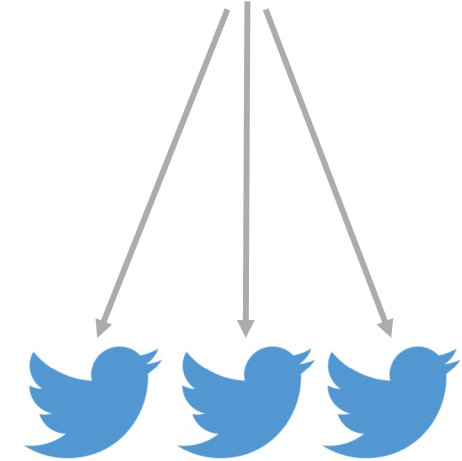
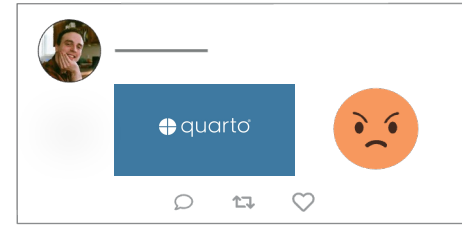
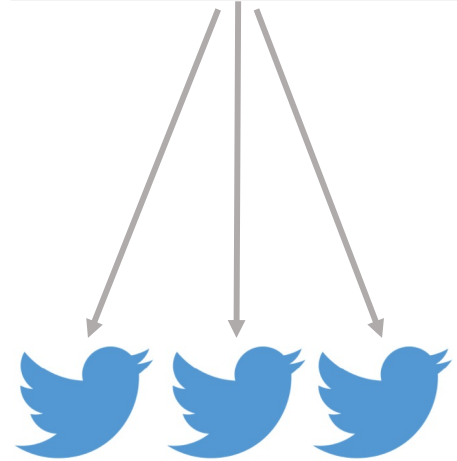
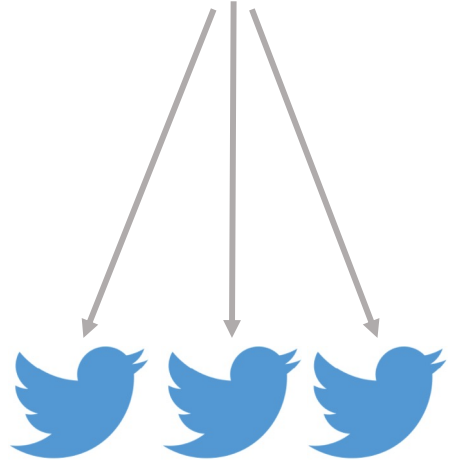
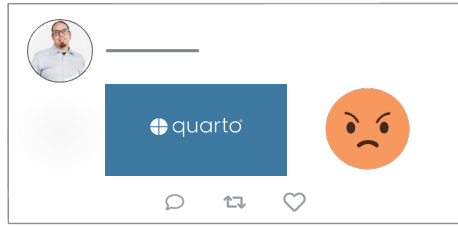


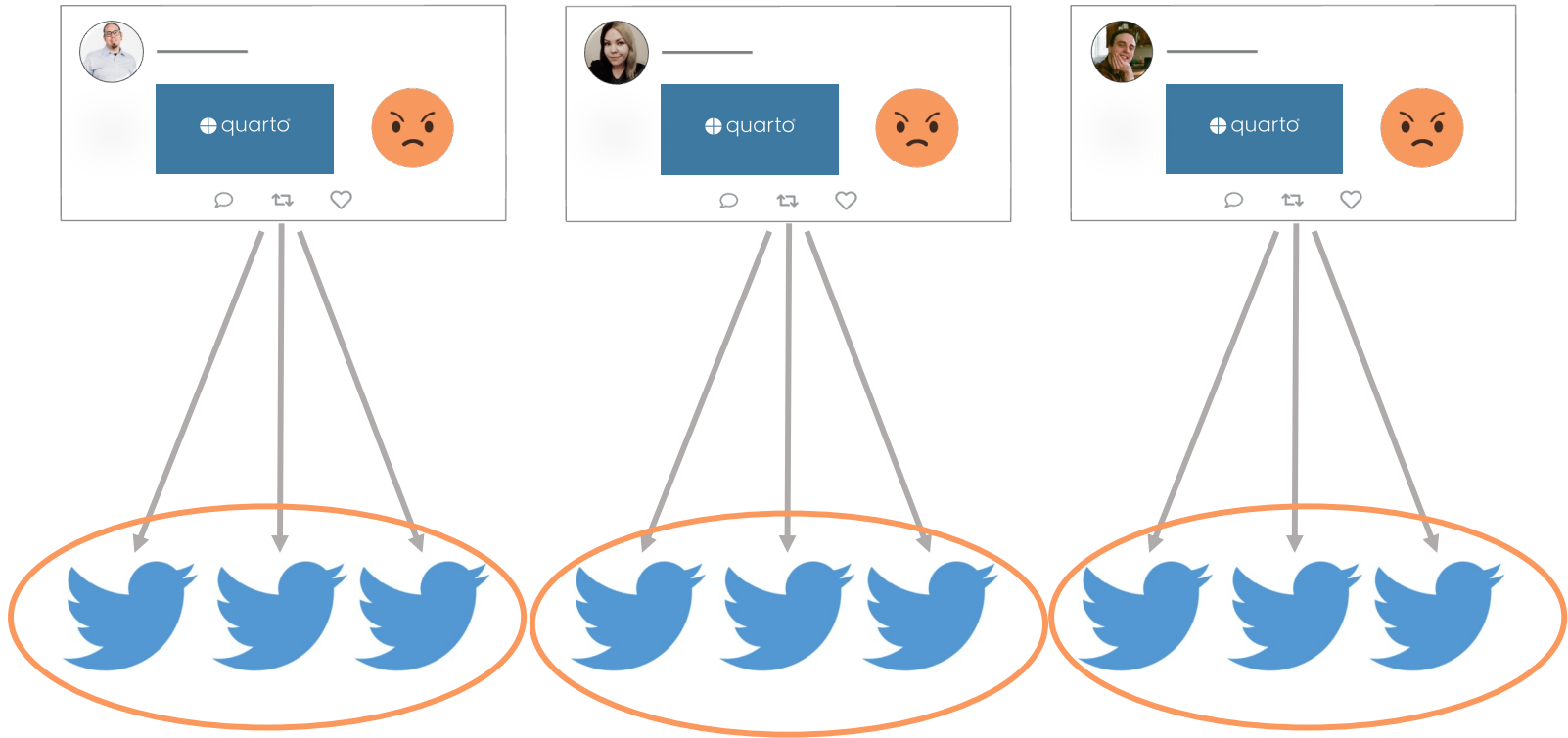


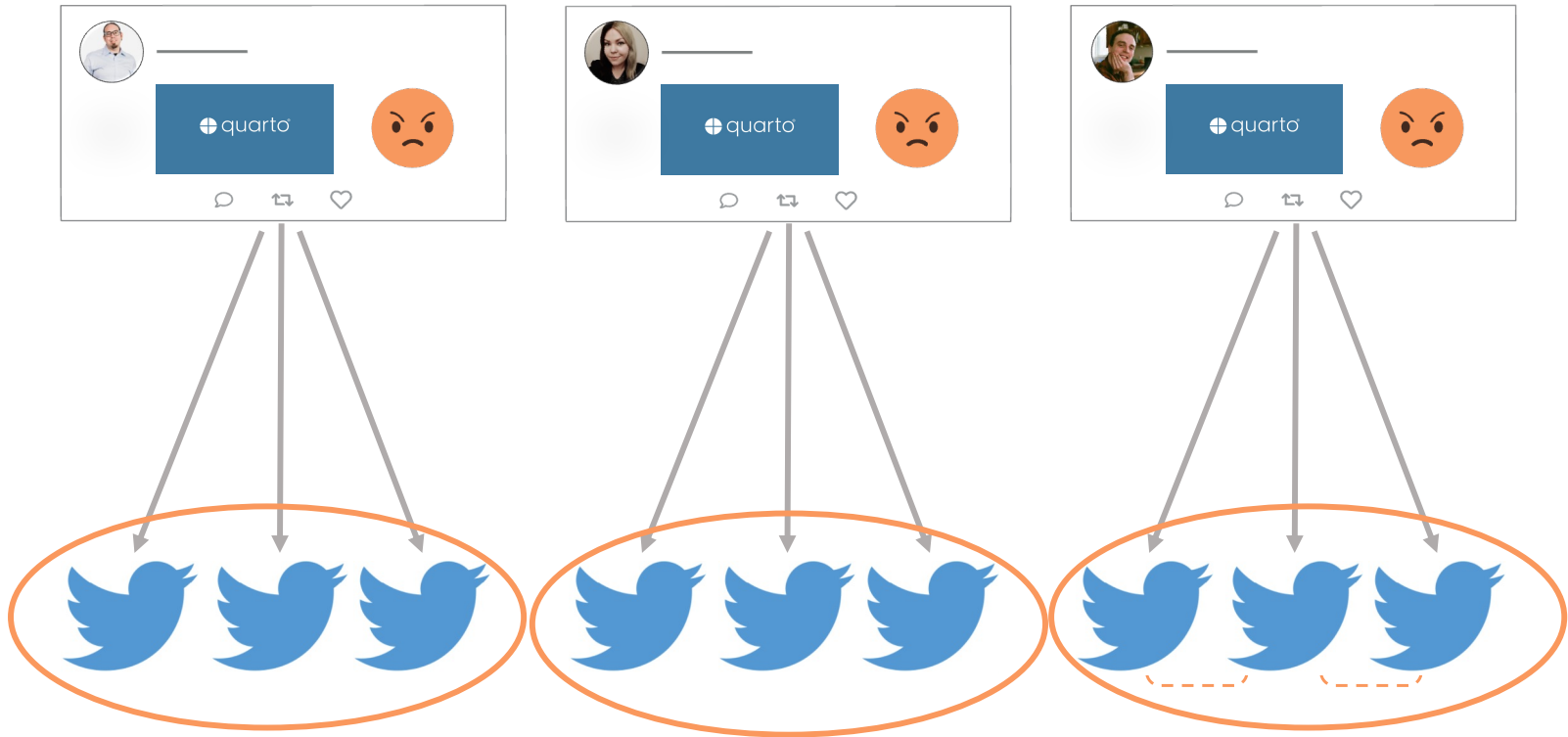






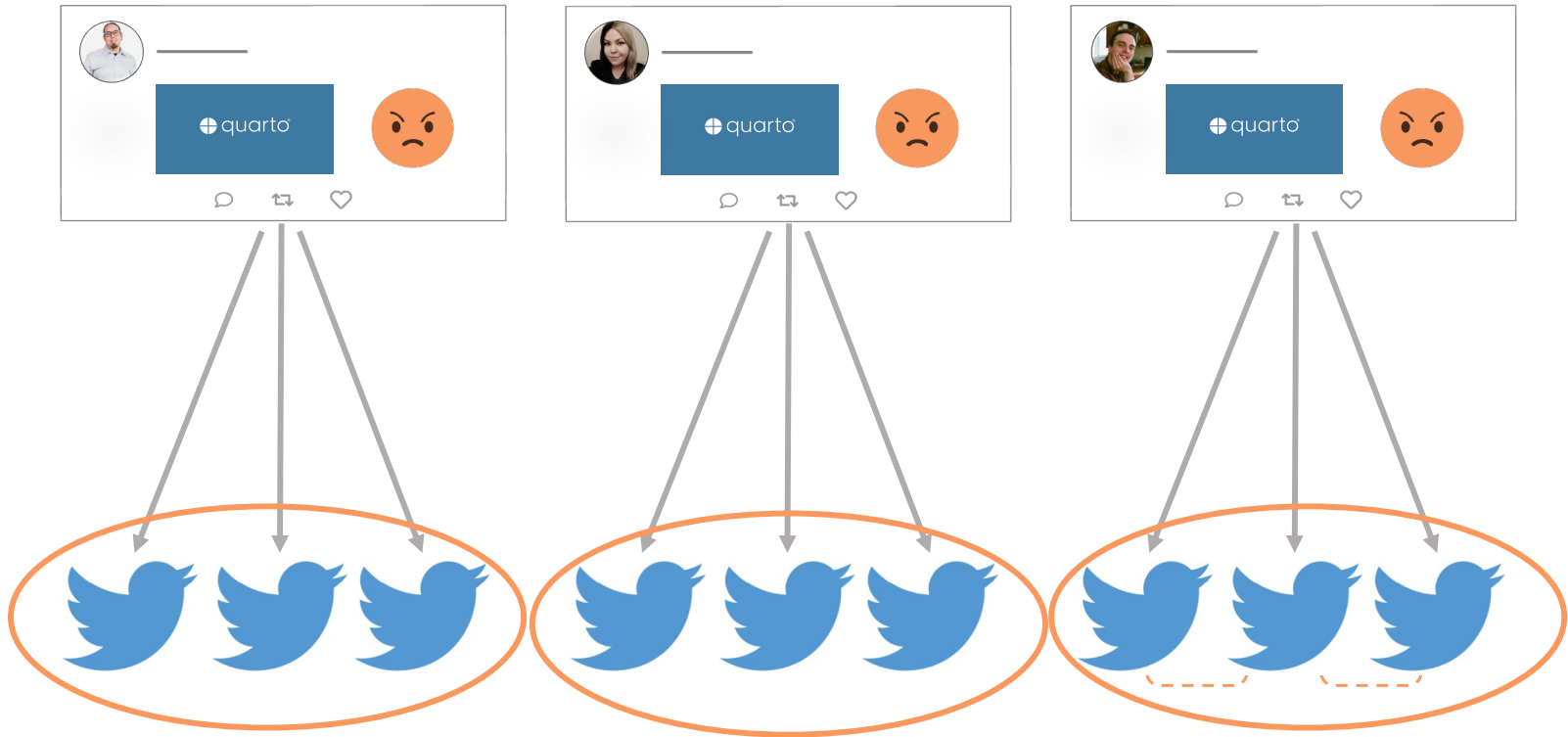






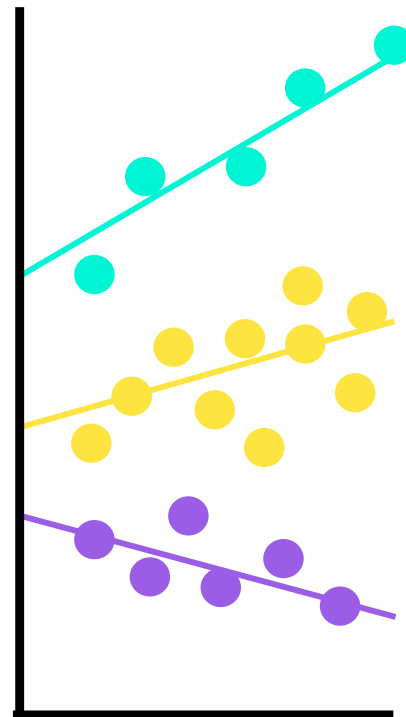
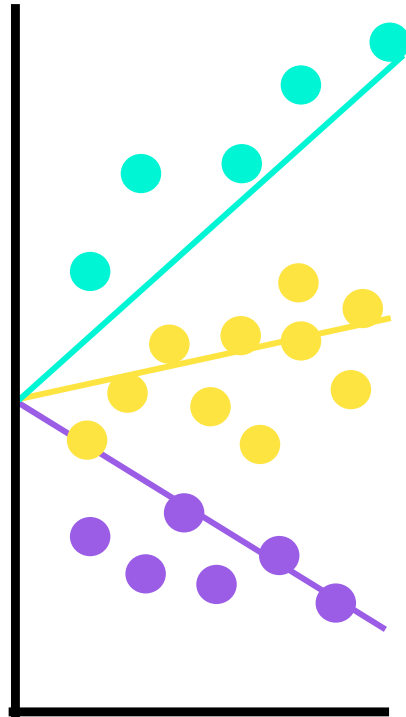
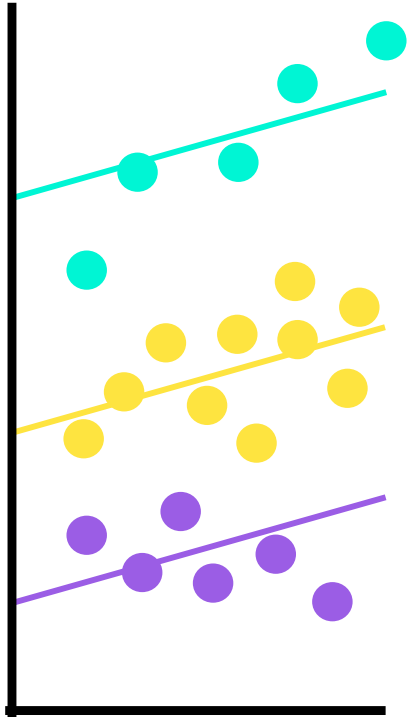
Random effects

Fixed effects



Random effects

Fixed effects



$$Y_i = \beta_0 + r_i$$

$$Y_{ij} = \beta_{0j} + r_{ij}$$

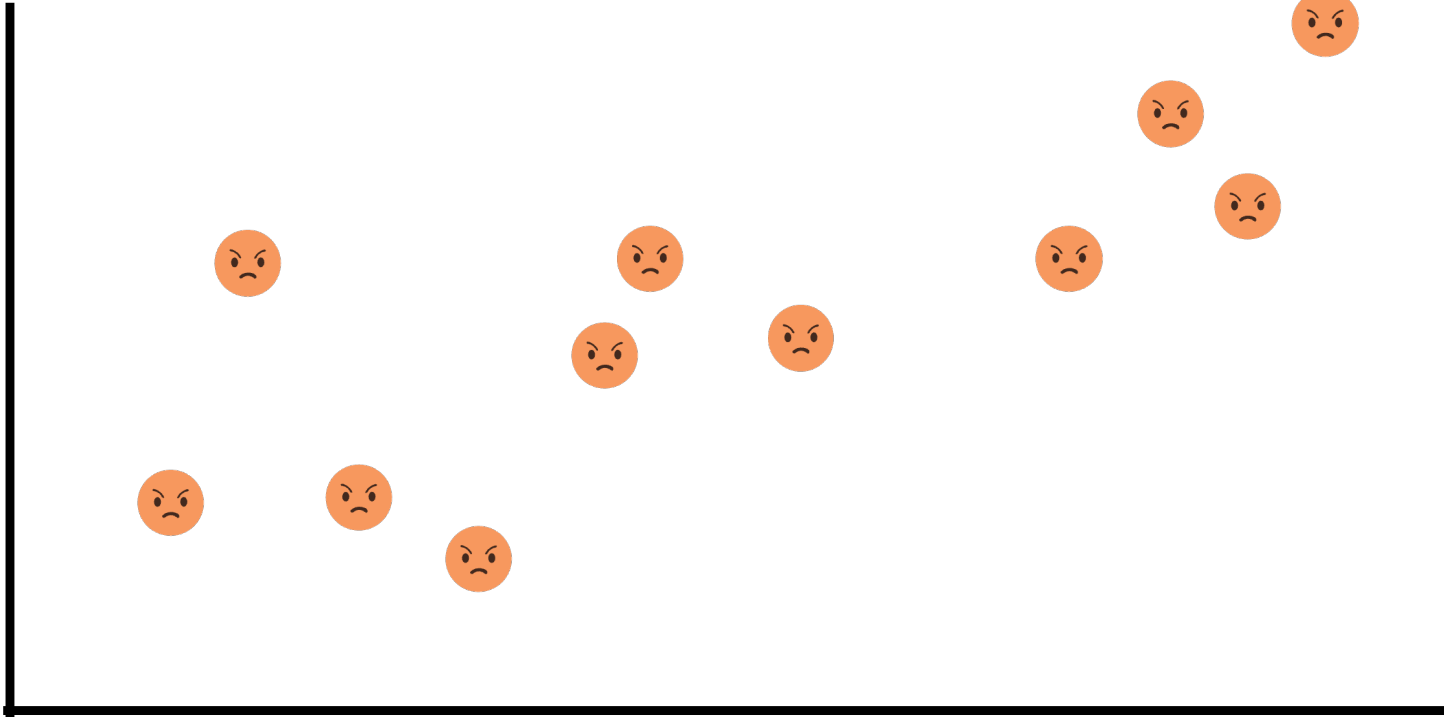
$$Y_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$Y_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$Y_{ij} = \beta_{0j} + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij}) + r_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$



Time

Generalized Estimating Equations

- Flexible
 - Can handle non-linear outcome variables
 - Can handle count, binary, and continuous outcomes
- Robust
 - Provides consistent parameter estimates even when the correlation structure is misspecified
- Performance curve
 - Parameter estimates are consistent and asymptotically normal: their properties improve as sample size increases

Generalized Estimating Equations

- Population-averaged
- Pre-Specified Correlation structure
- Quasi-likelihood estimation
- Missing data handling

Generalized Estimating Equations

To test the association between outrage and previously received social feedback, we used GEEs with robust SEs (observations, or tweets, were clustered by user) to estimate the population-level association between moral outrage expression and the amount of social feedback received on the previous day, with data aggregated at the level of days.

We modeled the sum of outrage expression as a negative binomial distribution with a log link function and an autoregressive correlation structure using PROC GENMOD in SAS 9.4. Decisions for modeling the outcome variable and correlation structure were based on the fact that the outcome variable was overdispersed count data and also on quasi-likelihood under independence model criterion (QIC) model fit statistics available in PROC GENMOD. To replicate the analyses in R 3.6.1 in a computationally efficient manner, we used the “bam” function in the package “mgcv” v1.8. SAS and R scripts used for data organization and model estimation described in this section are available at <https://osf.io/9he4n/>. Model specifications and variable formations listed below were preregistered at <https://osf.io/dsj6a>

The model predicting outrage expression from previous social feedback included as predictors the sum of feedback received when outrage was expressed for seven lagged days, previous outrage tweeting for seven lagged days, previous sums of non-outrage feedback for seven lagged days, user-level tweet history total, number of tweets containing URLs per day, number of tweets containing media per day, and the user follower count.

Results were robust to various model specifications including a model that included one previous day of outrage feedback, previous tweeting, and feedback for non-outrage tweets (i.e., including only one lag for each variable). Results were also robust when modeling the main lagged predictor variable as the difference between feedback received for outrage tweets versus non-outrage tweets (i.e., what is the effect of receiving more feedback for outrage expression compared to other tweets a user sent?).

Generalized Estimating Equations

Supporting our hypotheses, we found that daily outrage expression was significantly and positively associated with the amount of social feedback received for the previous day's outrage expression [study 1: $b = 0.03$, $P < 0.001$, 95% confidence interval (CI) = [0.03, 0.03]; study 2: $b = 0.02$, $P < 0.001$, 95% CI = [0.02, 0.03]].

For our model, this effect size translates to an expected 2 to 3% increase in outrage expression on the following day of tweeting if a user received a 100% increase in feedback for expressing outrage on a given day. For instance, a user who averaged 5 likes/shares per tweet, and then received 10 likes/shares when they expressed outrage, would be expected to increase their outrage expression on the next day by 2 to 3%. While this effect size is small, it can easily scale on social media over time, become notable at scale at the network level, or for users who maintain a larger followership and could experience much higher than 100% increases in social feedback for tweeting outrage content (e.g., political leaders).