Preregistration-plan

Alberto Acerbi

16 July, 2020

Preregistration plan for the project "From storytelling to Facebook. Does the medium of transmission change which content is successful?"

Aims

While the exact extent and consequences of the problem are debated (Altay, Hacquin, and Mercier 2019; Grinberg et al. 2019), widespread worries concern the possible dangers related to the circulation of online misinformation, or "fake news". Online misinformation can be successful because of its content. Not being constrained by reality, misinformation can be manufactured (consciously or not) to appeal to common cognitive preferences for certain "culturally attractive" contents, such as threat, disgust, or gossip (Acerbi 2019a). Identifying this content can help to recognise questionable sources and articles.

A tool used by cultural evolution researchers to understand what content is more likely to be transmitted is the methodology of transmission chain experiments (Mesoudi and Whiten 2008). Transmission chain experiments are a controlled laboratory equivalent of the game of Chinese whispers. In this way, researchers can detect which content survives, which content is modified (and how it is modified), when passing through chains of participants.

However, online transmission of information – think a social media "share" – is profoundly different. It requires, differently from cultural evolution experiments, the decision of whether to transmit something or not, and it does not require retention and reproduction. My objective is to explicitly compare the results of standard transmission chain experiments with new, modified, versions, that resemble what happens with online transmission. Specifically, my research question is: does the different medium change which content is favoured? Online sharing may, for example, reduce the selection of memorable or easy to reproduce content, as transmission does not require these steps (Acerbi 2019b).

The study will involve two main experiments. In experiment 1, I will use standard transmission chain experiments to test whether particular content is favoured. I will focus on content that had already been shown as being retained and transmitted successfully in such experiments, in particular: negative information (Bebbington et al. 2017), information eliciting disgust (Eriksson and Coultas 2014), and threat-related information (Blaine and Boyer 2018). This experiment is essential both to assess the robustness of previous results of in a web-based experiment, and to test the material that will be used in the second stage.

In experiment 2, I will use the same material to mimic a digital, online, scenario, where transmission does not need retention and reproduction, but simply the decision to share it or not. Participants will be prompted with the hypothetical possibility of sharing (or not) material online. This experimental stage will involve two conditions: in the first, it will be suggested to participants that their sharing will be anonymous, in a large social network (such as, for example, reddit); in the second, it will be suggested they will be sharing with their name in a social network where they know most of their contacts (such as Facebook).

Experiment 1

The first experiment is an online study that will investigate whether specific attractive content (negative, disgust-eliciting, and threat-related) tend to be better remembered and reproduced in transmission chain experiments than a neutral counterpart. For convenience, I will use below the terminology "attractor" for the three types of content (negative, disgust, or threat) and "content" for whether the content is neutral or attractive.

1.1 Participants N=540 participants recruited through Prolific (180 for each condition).

1.2 Procedure For each attractor, I will run 60 independent chains of transmission, each including three participants each (three iterations is standard in transmission chains experiments (see e.g. Stubbersfield, Tehrani, and Flynn 2015)). Thirty chains will involve the attractive content, and the other 30 chains will involve a neutral version of a similar content. For each transmission chain, participant 1 reads online the original text, a short story of five/six sentences (after collecting basic demographic data). Then, they rewrite the story as if retelling it to a friend. The text generated is then provided to participant 2 who does the same and pass it to participant 3, who repeats. The experiment will be realised with the software *Qualtrics*.

1.3 Material

Negative *Attractive*: A woman in Oregon hit the $8 million jackpot at a Lucky Eagle Casino slot machine. Veronica Castillo took her mother to the casino in Rochester, Washington. She put $100 in a machine and hit the jackpot. The casino staff came over and told her the machine malfunctioned. The woman attempted a legal battle, but she could not claim her money.

Neutral: A woman in Oregon hit the $8 million jackpot at a Lucky Eagle Casino slot machine. Veronica Castillo took her mother to the casino in Rochester, Washington. She put $100 in a machine and hit the jackpot.The casino staff came over and told her the machine malfunctioned. After a legal battle, the woman was able to claim her money.

Disgust *Attractive*: A major outbreak of salmonella has been reported in the US. The outbreak is the largest in the last 25 years. The outbreak had its origin in Saint Rika's hospital. More than 500 cases were identified in the hospital only. The likely source of the outbreak is contact with contaminated feces in the hospital's toilets.

Neutral: A major outbreak of chickenpox has been reported in the US. The outbreak is the largest in the last 25 years. The outbreak had its origin in Saint Rika's hospital. More than 500 cases were identified in the hospital only. Chickenpox can be transmitted through the air over time and distance by small particles.

Threat *Attractive*: The FBI reported increases in the violent crime rate between 2017 and 2019. Violent crime includes offenses such as rape, robbery and assault. Most of the crimes that are reported to police are not solved. This is based on an FBI measure known as the "clearance rate." In 2018, police nationwide cleared less than half of violent crimes that were reported to them.

Neutral: The FBI reported increases in corporate bribery between 2017 and 2019. Corporate bribery means giving a reward to influence someone's behaviour. Most of the bribery cases that are reported are not solved. This is based on an FBI measure known as the "clearance rate." In 2018, police nationwide cleared less than half of bribery cases that were reported to them.

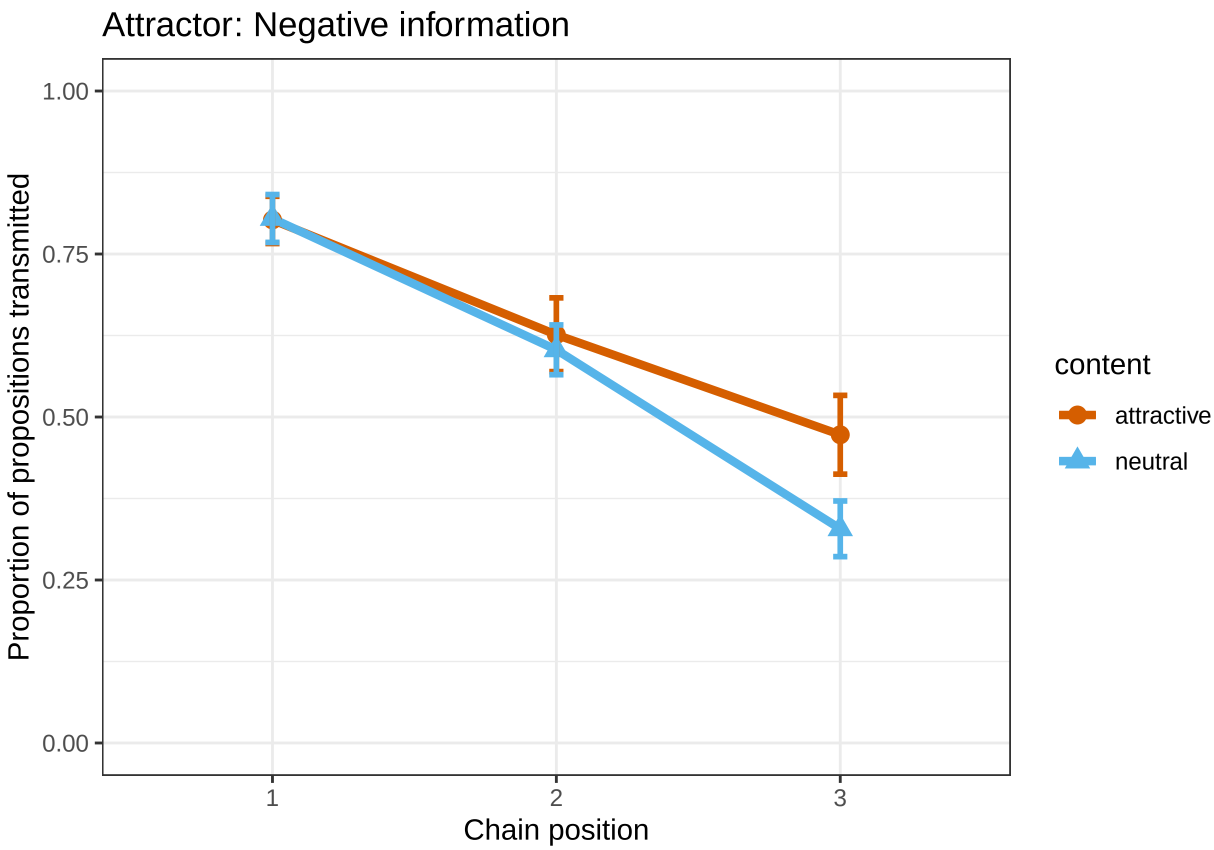
NOTE: The texts will be tested before the experiment, and may be subjected to changes.

1.4 Prediction In each condition, i.e. for each attractor, the attractive content will be better transmitted than the neutral content.

1.5 Analysis The texts produced by the participants at each steps of the chains will be analysed by two independent coders using propositional analysis, the output of which is the proportion of sentences correctly retained with respect to the target, original, text. The three resultant datasets (one for each attractor) will be analysed using linear mixed models, with position in the chain and content (attractive/neutral) as fixed effects, and each chain ID as random effect. The analysis will be performed in R, using the lme4 package (Bates et al. 2015).

See below the code for the analysis and the plots with simulated data (only for one condition, this will be replicated three times for each attractor). For the simulated data, I simply assume that proportions are normally distributed (and truncated between 0 and 1). The mean decreases along the chain position (i.e. participants reproduce less sentences) but there is *no* difference between contents (to test with difference, uncomment the pertinent lines in the code below).

```
library(tidyverse)
library(lme4)
library(lmerTest)

# data structure:
negative <- tibble(content = factor(rep(rep(c("attractive", "neutral"), each = 30), 3)),
                   proportion = NA,
                   chain_position = factor(rep(1:3, each = 60)),
                   chain_ID = factor(rep(1:60, 3)))

# simulated data:
negative$proportion[1:60] <- rnorm(60, 0.8, .3)
negative$proportion[61:120] <- rnorm(60, 0.6, .3)
negative$proportion[121:180] <- rnorm(60, 0.4, .3)

# test with difference:
# negative <- negative %>%
#  mutate(proportion = replace(proportion, content=="attractive", proportion[content=="attractive"] + 0

negative <- negative %>%
  mutate(proportion = replace(proportion, proportion > 1, 1)) %>%
  mutate(proportion = replace(proportion, proportion < 0, 0))

# palette friendly to colorblind:
cbPalette <- c("#D55E00", "#56B4E9", "#E69F00", "#009E73", "#999999", "#CC79A7")
# plot the proportion of sentence correctly retained:
negative %>%
  ggplot(aes(x = chain_position, y = proportion, colour = content)) +
    stat_summary(fun = mean, geom = "line", size = 1.5, aes(group = content, colour = content)) +
    stat_summary(fun = mean, geom = "point", size = 3, aes(shape = content)) +
    stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.05, size = 1) +
    theme_bw() +
    ylim(0,1) +
    scale_colour_manual(values = cbPalette) +
    labs(x = "Chain position", y = "Proportion of propositions transmitted", title = "Attractor: Negati
```



Here the output of the model:

```
negative_model <- lmer(proportion ~ content + chain_position + (1|chain_ID), data = negative)
coef(summary(negative_model))
```

```
##                    Estimate Std. Error        df   t value     Pr(>|t|)
## (Intercept)      0.83092905 0.03779116 155.57782 21.987392 1.368076e-49
## contentneutral  -0.05507088 0.03780394  58.22524 -1.456750 1.505574e-01
## chain_position2 -0.18878914 0.04626887 118.22602 -4.080262 8.203053e-05
## chain_position3 -0.40276002 0.04626887 118.22602 -8.704773 2.233176e-14
```

As expected, with no difference between content, there is no effect of content (β = -0.0550709, SE = 0.0378039, df = 58.2252407, p = 0.1505574).

Experiment 2

The second experiment is again an online study, and it will test whether the same content analysed in experiment 1 is also favoured when participants are not required to remember and reproduce the content, but simply deciding which content to share, as it happens in social media.

2.1 Participants N=600 participants recruited through Prolific (300 for each condition).

2.2 Procedure After collecting basic demographic information, each participant will be presented with three texts, one for each of the attractors analysed in experiment 1 (negative, disgust-eliciting, and threat-related). At random, each text will be either the story with the attractive content present or its neutral counterpart (the order of presentation of the texts will be also randomised). In the first condition, participants will be asked, for each story, whether they would share it or not anonymously in a large social network. In the second condition, they will be asked if they would share the story to their friends, in their favourite social network. The experiment will be realised with the software *Qualtrics*.

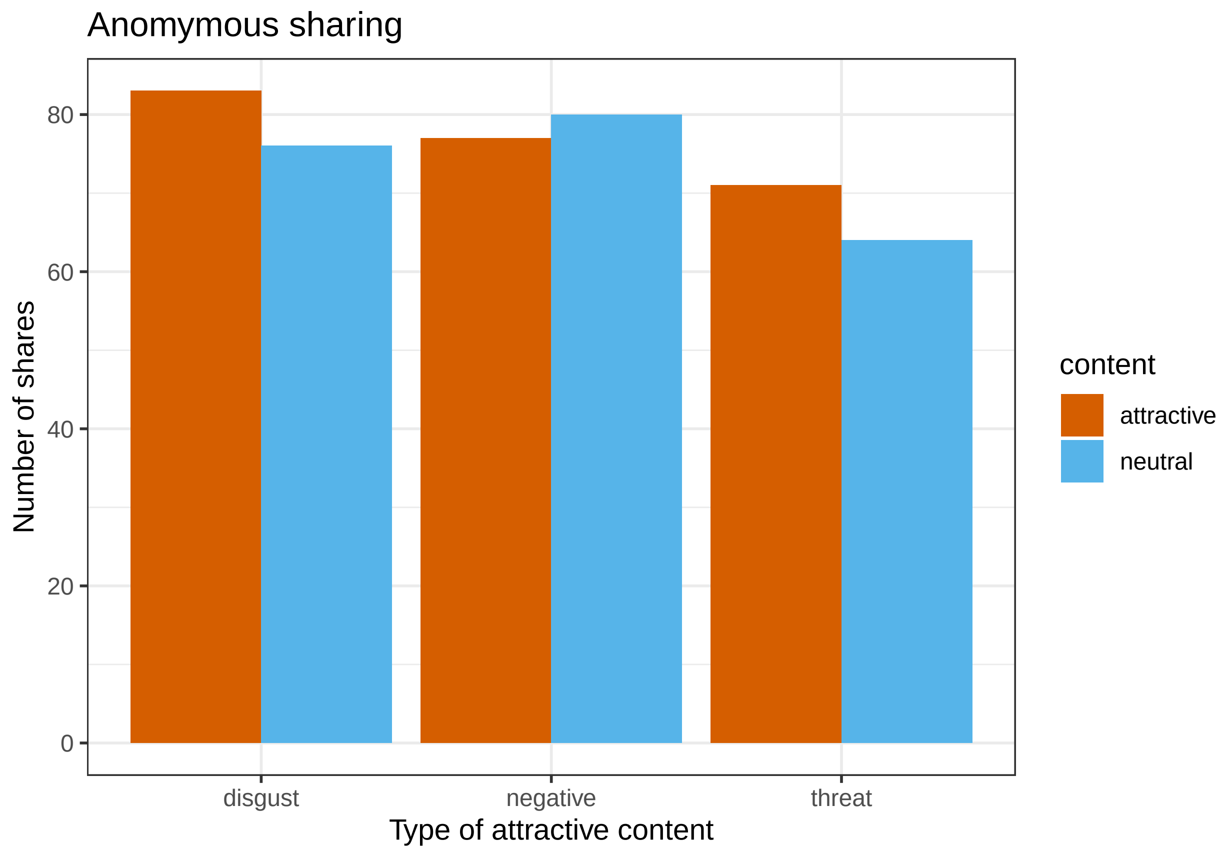
2.3 Materials The same six texts used in experiment 1.

2.4 Prediction There is no specific prediction, but the research question is whether in each condition (anonymous/friends), and for each attractor, the attractive content will be shared more than the neutral content.

2.5 Analysis Each of the two resultant datasets (one for condition) will be split in three sub-datasets (one for each type of attractor) and they will be analysed separately using general linear mixed models (binomial) with content (attractive/neutral) as fixed effect and order of presentation and participants ID as random effect. The analysis will be performed in R, using the lme4 package.

See below the code for the analysis and the plots with simulated data (only for one condition "anonymous sharing", this will be duplicated for "sharing with friends"). For the simulated data, I assume that choice of sharing is fully random, so there is *no* effect of content. As for experiment 1, to test with difference, uncomment the pertinent lines in the code below.

```r
# data structure:
anonymous <- tibble(attractor = NA,
                    content = NA,
                    share = NA,
                    order_of_presentation = factor(rep(1:3, 300)),
                    participant_ID = factor(rep(1:300, each = 3)))

# simulated data:
# randomise content type for each participant:
anonymous$attractor <- as.vector(replicate(300, sample(c("negative", "threat", "disgust"))))
# randomise whether attractive or not:
anonymous$content <- sample(c("attractive","neutral"), replace = TRUE, size = 900)
# randomise participant choices:
anonymous$share <- factor(sample(c(0,1), replace = TRUE, size = 900))

# test with difference:
# for(i in sample(1:900, size = 500)){
#   if(anonymous$content[i] == "attractive"){
#     anonymous$share[i] <- 1
#   }
# }

# plot the number of time a content is shared
anonymous %>%
  filter(share == 1) %>%
  ggplot(aes(x = attractor, fill = content)) +
    geom_bar(position = "dodge") +
    theme_bw() +
    scale_fill_manual(values = cbPalette) +
    labs(x = "Type of attractive content", y = "Number of shares", title = "Anomymous sharing")
```



Here the model (as before, only a single example, for negative content). The same model needs to be replicated three times for each attractor:

```
anonymous_negative <- anonymous %>%
  filter(attractor == "negative")
negative_model <- glmer(share ~  content + (1|participant_ID) + (1|order_of_presentation), family = bin
coef(summary(negative_model))
```

```
##                   Estimate Std. Error    z value  Pr(>|z|)
## (Intercept)      0.2166710  0.1706446  1.2697208 0.2041841
## contentneutral -0.2290935  0.2323044 -0.9861783 0.3240456
```

As expected, there is no effect of content (β = -0.2290935, SE = 0.2323044, Z(300) = -0.9861783, p = 0.3240456).

References

Acerbi, Alberto. 2019a. "Cognitive Attraction and Online Misinformation." *Palgrave Communications* 5: 15. https://doi.org/10.1057/s41599-019-0224-y.

———. 2019b. *Cultural Evolution in the Digital Age*. Oxford, New York: Oxford University Press.

Altay, Sacha, Anne-Sophie Hacquin, and Hugo Mercier. 2019. "Sharing Fake News Is Bad for Your Epistemic Reputation." Preprint. PsyArXiv. https://doi.org/10.31234/osf.io/82r6q.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bebbington, Keely, Colin MacLeod, T. Mark Ellison, and Nicolas Fay. 2017. "The Sky Is Falling: Evidence of a Negativity Bias in the Social Transmission of Information." *Evolution and Human Behavior* 38 (1): 92–101.

https://doi.org/10.1016/j.evolhumbehav.2016.07.004.

Blaine, Timothy, and Pascal Boyer. 2018. "Origins of Sinister Rumors: A Preference for Threat-Related Material in the Supply and Demand of Information." *Evolution and Human Behavior* 39 (1): 67–75. https://doi.org/10.1016/j.evolhumbehav.2017.10.001.

Eriksson, Kimmo, and Julie C. Coultas. 2014. "Corpses, Maggots, Poodles and Rats: Emotional Selection Operating in Three Phases of Cultural Transmission of Urban Legends." *Journal of Cognition and Culture* 14 (1-2): 1–26. https://doi.org/10.1163/15685373-12342107.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake News on Twitter During the 2016 U.S. Presidential Election." *Science* 363 (6425): 374–78. https://doi.org/10.1126/science.aau2706.

Mesoudi, Alex, and Andrew Whiten. 2008. "The Multiple Roles of Cultural Transmission Experiments in Understanding Human Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1509): 3489–3501. https://doi.org/10.1098/rstb.2008.0129.

Stubbersfield, Joseph M., Jamshid J. Tehrani, and Emma G. Flynn. 2015. "Serial Killers, Spiders and Cybersex: Social and Survival Information Bias in the Transmission of Urban Legends." *British Journal of Psychology* 106 (2): 288–307. https://doi.org/10.1111/bjop.12073.