Reality Check: The Effects of Hiding Dislikes on YouTube's User Behavior

A triangulated design

Maggie Zhang, PhD Candidate in Communication, UIUC
mz44@illinois.edu

Margaret Ng, Assistant Professor of Journalism, UIUC
ymn@illinois.edu

Created on May, 2023

Table of Contents

Background

1 Introduction

- Background: Social media platforms have recently experienced a surge in hate attacks.
- YouTube's Response: On November 10, 2021, YouTube announced the decision to hide the dislike count to prevent "dislike mobs" from harassing creators or promoting a particular agenda.
- Question: What impact will it have on the platform ecosystem?
- Purpose of the Study: This study aims to examine the effects of dislikes and commenting sections on users' behavior.

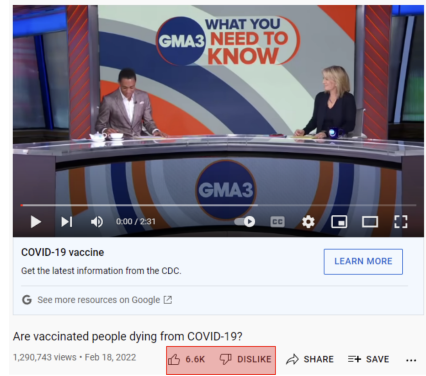


Table of Contents

2 Literature Review

Social Cue & Social Influence

Heuristic Processing: Bandwagon Heuristic

- *"If others think a story is good, then I should think so too"* (Sundar, 2008).
- Users are more likely to engage with content that has received a higher number of likes
- Desire to "join the crowd and behave like their fellows" (Hu Yao, 2021, p. 4).

Social Influence Bias - Herding Effect

- Popular content becomes more popular due to the irrational effect of past positive ratings - a *"rich-get-richer"* dynamic (Muchnik, Aral, Taylor, 2013).

Table of Contents

3 Research Questions

RQs

3 Research Questions

Research Question 1: How do aggregated ratings, the dislike number especially, impact people's expressions and reactions to media content?
Research Question 2: How does the valence of pre-existing comments impact people's expressions and reactions to media content?
Research Question 3: How does ideological alignment impact viewer reactions and expressions to media content?
Research Question 4: How do aggregated ratings and pre-existing comments interact to influence viewer reactions and expressions?

Table of Contents

4 Method

Method: a Triangulated Research Design

4 Method

Study 1: Experiment

3 × 3 factorial online experiment

- Dislike-to-like ratio
- Pre-existing comments valence

Study 2: Observational Study

analysis of real-world YouTube data

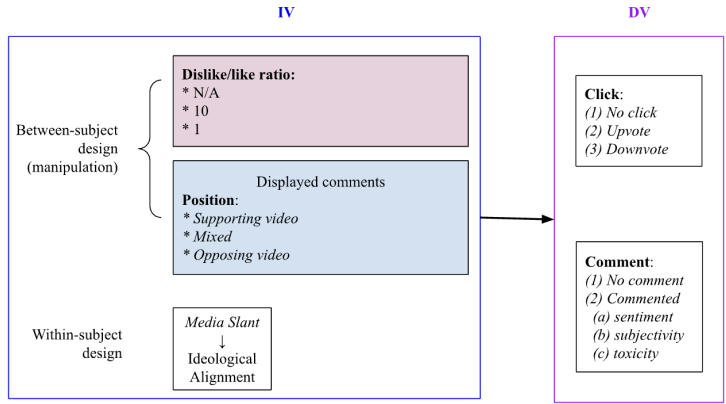
- Focus on news channels
- Interrupted Time Series (ITS) analysis

Table of Contents

Research Design

5 Study 1: Design

- Built a website that resembled the YouTube interface & customized two social feedback cues
- Chose the topic of gun control as the context of study: a highly debated and polarizing issue

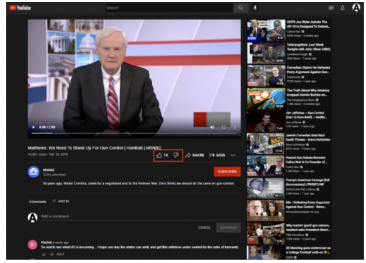


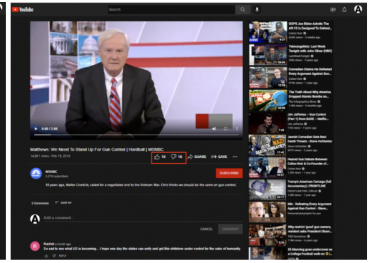
Manipulation 1: Dislike-to-like Ratio

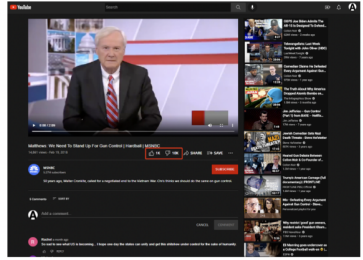
5 Study 1: Design

Dislikes Conditions

- Dislikes not shown
  1k likes
- 1k dislikes 1k likes
  (ratio = 1)
- 10k dislikes 1k likes
  (ratio = 10)



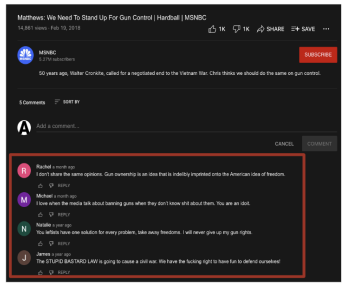


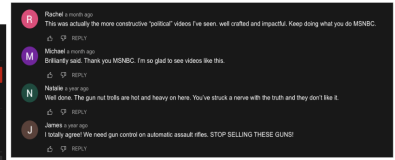


Manipulation 2: Pre-existing comments valence

Comments conditions:

- Agree with video: 4 positive comments
- Disagree: 4 negative comments
- Mixed opinions: 2 positive & 2 negative comments





Experiment Procedures

5 Study 1: Design

Participants: 700 participants with diverse attitudes towards gun control from MTurk in Jan 2023 - 606 passed the attention check

Procedures:
Measured pre-test attitude towards gun policy → randomly assigned to 1 of 9 conditions → watch 1st video, click & comment → watch 2nd video, click & comment → demographics → end experiment

Table of Contents

6 Study 1: Results

Group Differences in Clicks & Comments

6 Study 1: Results

Table 1: Descriptive Statistics, Chi-square, One-way ANOVA and T Tests for Participants' Clicks and Comments

	Interaction type: click				Interaction type: comment			Comment					
	Click dislike (N = 249)	No click (N = 539)	Click like (N = 424)	Chi-square	No comment (N = 293)	Comment (N = 919)	Chi-square	Sentiment (*N* = 919)		Subjectivity (*N* = 919)		Toxicity (*N* = 919)	
	Freq	*Freq*	*Freq*		*Freq*	*Freq*		*M(SD)*		*M(SD)*		*M(SD)*	
Dislike/like ratio									One-way ANOVA *F*		One-way ANOVA *F*		One-way ANOVA *F*
1	75	184	159		105	313		0.15		0.44		0.13	
10	96	166	132	7.29	76	318	8.53*	0.15	0.46	0.42	0.60	0.14	0.69
Dislikes hidden	78	189	133		112	288		0.13		0.43		0.15	
Display comment									One-way ANOVA *F*		One-way ANOVA *F*		One-way ANOVA *F*
Disagree	98	171	153		55	367		0.15		0.44		0.15	
Mixed	77	185	142	5.18	119	285	**44.05 *****	0.13	0.36	0.45	1.85	0.14	**3.12***
Agree	74	183	129		119	267		0.16		0.40		0.12	
Ideological alignment									T test		T test		T test
aligned	70	264	272	**81.90 *****	145	461	0.02	0.18	**3.75*****	0.42	-0.94	0.12	**-3.13****
not aligned	179	275	152		148	458		0.11		0.44		0.16	

* $p < .05$ ** $p < .01$ *** $p < .001$

On Participants' Reactions: Clicking Behavior

6 Study 1: Results

Table 2: Multinomial Logistic Regressions Predicting Clicking Behavior

	DV: click (reference category: no click)							
	Clicking dislike button (n = 249) (1)				Clicking like button (n = 424) (2)			
	B	(SE)	OR	95% CI	B	(SE)	OR	95% CI
Intercept (constant)	-1.41***	0.26	0.25	(0.15, 0.41)	-0.14	0.21	0.87	(0.58, 1.31)
Main Effect								
Dislike/like ratio (reference: dislike number hidden)								
1 (1k dislikes vs. 1k likes)	-0.07	0.33	0.94	(0.49, 1.79)	0.28	0.27	1.33	(0.78, 2.27)
10 (10k dislikes vs. 1k likes)	-0.04	0.34	0.96	(0.49, 1.86)	0.13	0.29	1.14	(0.65, 2.00)
Display comments (reference: mixed opinion)								
Agree	-0.11	0.35	0.89	(0.45, 1.75)	0.13	0.29	1.14	(0.61, 1.80)
Disagree	-0.02	0.32	0.98	(0.52, 1.84)	0.04	0.28	1.05	(0.65, 2.00)
Ideological alignment (reference: aligned)								
Nonalignment	**0.91*** **	0.17	2.48	(1.79, 3.43)	**-0.62*** **	0.13	0.54	(0.41, 0.70)
Interaction Effect								
Ratio 1 * Disagree comments	0.38	0.47	1.47	(0.59, 3.66)	0.28	0.39	1.32	(0.62, 2.82)
Ratio 10 * Disagree comments	0.65	0.46	1.92	(0.78, 4.70)	0.08	0.40	1.08	(0.50, 2.35)
Ratio 1 * Agree comments	-0.21	0.49	0.81	(0.31, 2.10)	-0.47	0.39	0.63	(0.29, 1.35)
Ratio 10 * Agree comments	0.46	0.48	1.58	(0.61, 4.05)	-0.14	0.41	0.87	(0.39, 1.94)
Nagelkerke R^2	0.10							

* $p < .05$ ** $p < .01$ *** $p < .001$

On Participants' Expressions: Comments

6 Study 1: Results

Table 3: Logistic Regression and Linear Regressions Predicting Comments

	Dependent variable: comment				
	whether comment (n = 1212) logistic (1)		sentiment (n = 919) OLS (2)	subjectivity (n = 919) OLS (3)	toxicity (n = 919) OLS (4)
	B (SE)	OR (95% CI)	B (SE)	B (SE)	B (SE)
Intercept	0.85*** (0.20)	2.33 (1.58, 3.50)	0.13*** (0.03)	0.45*** (0.03)	0.15*** (0.02)
Main effect					
Dislike-to-like ratio (reference: dislike number hidden)					
1	-0.12 (0.26)	0.89 (0.53, 1.47)	0.07 (0.04)	-0.03 (0.04)	**-0.05** (0.03)**
10	0.29 (0.28)	1.33 (0.77, 2.32)	0.06 (0.04)	0.02 (0.05)	-0.00 (0.03)
Display comments (reference: mixed opinion)					
Support	0.29 (0.28)	1.33 (0.77, 2.33)	0.07 (0.04)	-0.08* (0.05)	**-0.05* (0.03)**
Oppose	0.07 (0.26)	1.08 (0.64, 1.81)	0.05 (0.04)	-0.01 (0.04)	-0.00 (0.03)
Ideological alignment (reference: aligned)					
Non-alignment	-0.03 (0.14)	0.97 (0.74, 1.28)	**-0.07 *** (0.02)**	0.02 (0.02)	**0.04*** (0.01)**
Interaction Effect					
Ratio 1 * Oppose	**2.11 *** (0.50)**	8.25 (3.23, 23.45)	-0.07 (0.06)	0.05 (0.06)	0.04 (0.04)
Ratio 10 * Oppose	**1.93 *** (0.54)**	6.87 (2.53, 21.12)	-0.04 (0.06)	-0.07 (0.06)	-0.01 (0.04)
Ratio 1 * Support	-0.44 (0.38)	0.65 (0.31, 1.35)	-0.05 (0.06)	0.10 (0.06)	0.05 (0.04)
Ratio 10 * Support	-0.61 (0.40)	0.55 (0.25, 1.19)	-0.08 (0.06)	-0.02 (0.06)	0.01 (0.04)
R^2					

* $p < .05$ ** $p < .01$ *** $p < .001$.

Main Findings

6 Study 1: Results

RQ1: Aggregated ratings → User behavior

- Dislikes don't affect whether people click.
- However, if dislikes are hidden, people comment less.

RQ2: Existing comments → User behavior

- The tone of comments doesn't affect whether people click.
- But, when comments are mostly negative, new comments tend to be more toxic.

RQ3: Ideological alignment → User behavior

- If a video matches a user's beliefs, it gets more upvotes and fewer downvotes.
- These videos also inspire more positive and less toxic comments.

RQ4: Interaction: ratings & comments → User behavior

- The interaction between dislikes and comment tone doesn't affect clicking.
- However, these factors combined significantly affect commenting behavior, especially when negative comments are present.

Table of Contents

7 Study 2: Design

Research Design

7 Study 2: Design

- Data: videos posted by 49 YouTube media channels in 2 weeks prior to the intervention.
- YouTube API: 1,116 videos & all comments posted within 30 days of video release.
- ITS: the effect of hiding dislikes on comment sentiment, subjectivity, and toxicity.
- Only kept videos with at least 4 days of comments both before & after the interruption: 308 valid videos.
- Fit into optimal AutoRegressive Integrated Moving Average (ARIMA) model

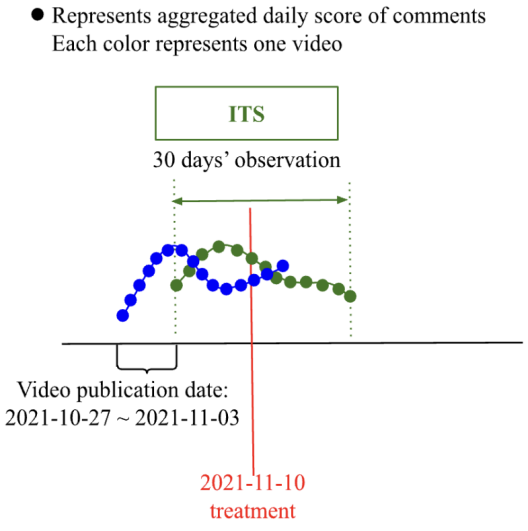


Table of Contents

8 Study 2: Results

ARIMA Results

8 Study 2: Results

- Each video's comments: one time series - 308 ARIMA results

Immediate effect

- The majority of videos (over 70%) did not have any significant change in the intercept of time series data of their comments.

Sustained effect

- Similarly, the majority of videos did not have any significant change.

Table 4. ARIMA model results for interrupted time series analysis

	Effect type	Significant Increase	Significant Desrease	No Significant Change
Sentiment	Immediate (step change)	11.18%	8.88%	79.93%
	Sustained (slope change)	9.87%	12.50%	77.63%
Subjectivity	Immediate (step change)	11.18%	11.51%	77.30%
	Sustained (slope change)	11.84%	10.53%	77.63%
Toxicity	Immediate (step change)	11.18%	8.88%	79.93%
	Sustained (slope change)	10.20%	11.51%	78.29%

Media leaning VS treatment effect

- Chi-square tests: no significant relationship, suggesting that no specific media channels benefited from YouTube's decision to hide the dislike counts.

Table of Contents

9 Discussion

Takeaways

9 Discussion

YouTube's dislike-hiding aimed at creator protection, but its effectiveness is questioned in this study. Our experiments and observational studies show:

- Aggregated ratings, as social cues, do not determine upvoting or downvoting.
- Pre-existing comments have a stronger influence on opinion expression than aggregated likes and dislikes.
- Ideological alignment has the most significant impact on people's reactions and expressions.

This finding aligns with previous research on confirmation bias, indicating a preference for information that reinforces existing beliefs.

Q&A

Thank you for listening!

Presented by: Maggie Zhang
Email: mz44@illinois.edu
Twitter: @CommZhang
Website: http://drmaggiezhang.com