

Steal

A public prisoner’s dilemma, three days running. Steal took 64 percent of the vote on day one, 29 percent on day two, and 10 percent on day three.

THIS IS	PART OF	STATUS
PAPER 007 · A FINDING	THE GAME · ACT II, ROUND 4	FILED · V1 · 31 AUG 2026

SOURCE

3 9,656 13 to 15 Apr 2026

ATTEMPTS COMMENTS DATES

IN PLAIN WORDS

On 13 April 2026 the second level opened on an old dilemma. Comment Split or comment Steal. All it takes is one. If everyone splits, everyone moves on; one steal ends it for the room. It is the choice at the end of a game show, put to several thousand strangers at once, in public.

The first wall filled with steal: 2,556 steals to 1,454 splits. The round failed. The next day the post said “You saw what happened. Now decide again,” and steal fell to 606 against 1,513 splits. It failed again, because one is all it takes. On the third day the mechanic was rewritten so that no single player could end it, and steal fell to 228 against 2,160. The round passed.

This paper is about a crowd that stole, watched itself steal, and stopped. The number that fell was not the crowd’s morals. It was its expectation of everyone else.

THE QUESTION WE ASKED

The posted rule was the whole hypothesis: all it takes is one. Would a crowd of thousands, each able to end the round alone, hold? The follow-up questions arrived with the attempts: does seeing the result change the next vote, and does removing one person’s power to end it change it more?

WHAT WE DID

- **Instruction:** “Split” or “Steal”. All it takes is one.
- **Where:** three Instagram reels on three consecutive days, 13, 14 and 15 April 2026. Attempt 2 was captioned “You saw what happened. Now decide again.” The final attempt announced that the system

had adapted and failure was no longer possible via one individual.

- **How we counted:** top-level comments on each post, from the comment archive; the round ledger records no count of its own (“data access partial”). A comment counted as split or steal if its text began with that word, in any case. Everything else, including jokes, questions and the people announcing what they would have done, counted as other.
- **Steal share:** steal as a fraction of split plus steal, so the other comments do not dilute the vote.

WHAT HAPPENED

TABLE 1. The three walls.

ATTEMPT	SPLIT · STEAL · OTHER	STEAL SHARE OF THE VOTE
1 · 13 April, “All it takes is one”	1,454 · 2,556 · 549	64%
2 · 14 April, “Now decide again”	1,513 · 606 · 347	29%
3 · 15 April, no single player can end it	2,160 · 228 · 243	10%
Top-level comments across the three	9,656	

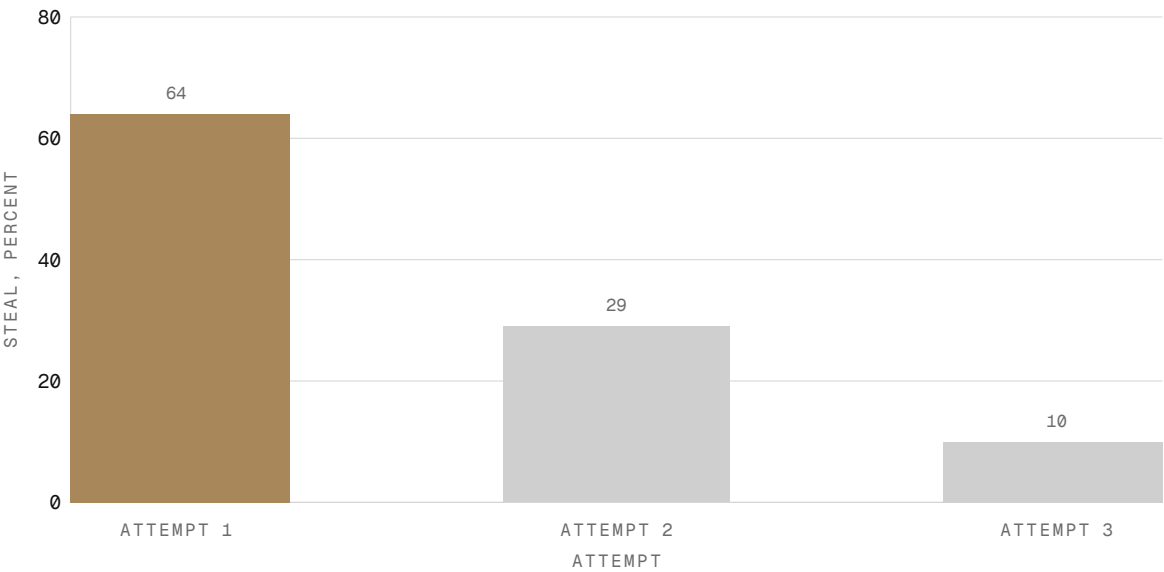


FIGURE 1. Steal as a share of the split-or-steal vote, by attempt, in percent. The dashed line is the steal rate on the television show that made the dilemma famous.

The most liked comment on the first wall explains the first wall: **Steal, no point, everyone else is already doing it.** The most liked split on the final wall answers it two days later: **i said split before despite all the steals and i chose split again.**

Platform numbers are from the Instagram API on 31 August 2026. Comment counts there include replies and everything written since the round, which is why they run above the counts this paper analyses; views and reach are the tallies the platform reports.

TABLE 2. Steal share of the vote on each attempt, with 95 percent intervals.

MEASURE	COUNT	OF	SHARE	95% INTERVAL
Attempt 1, steal share	2,556	4,010	63.7%	62.2% to 65.2%
Attempt 2, steal share	606	2,119	28.6%	26.7% to 30.6%
Attempt 3, steal share	228	2,388	9.5%	8.4% to 10.8%

The three intervals do not overlap. The fall from attempt to attempt is not noise.

TABLE 3. The post at platform scale, as of 31 August 2026.

REEL	VIEWS · REACH	COMMENTS TODAY
Attempt 1 · 13 April 2026	116,850 · 62,946 reach	16,349
Attempt 2 · 14 April 2026	34,972 · 21,383 reach	3,710
Attempt 3 · 15 April 2026	34,826 · 22,142 reach	4,187
All 3 reels	186,648 views	24,246

WHAT IT MEANS

In a public dilemma, people do not choose against the other player. They choose against their expectation of the crowd. Show the crowd its own result and the expectation moves, and the vote moves with it.

There is a baseline for this choice. On the British game show *Golden Balls*, two contestants faced exactly this decision with real money on the table, and a 2012 study of 287 episodes found that people chose split about 53 percent of the time, steal about 47. Our first wall was worse than the show, at 64 percent steal, and our third was far better, at 10.

The difference is the crowd. On the show the other player is one stranger you have just spent an hour with. On the wall the other player is everyone who has already commented, and on day one that wall was visibly filling with steal. The top comment says the reasoning out loud: everyone else is already doing it. Stealing was not greed, it was despair about the room.

Day two changed one thing: the crowd had seen the result. Steal fell by more than half. Day three changed the mechanic so that one person could no longer end it, and steal fell to a tenth. We cannot fully separate the two effects, because they arrived on different days to a crowd that was also learning. But the direction is clear, and it is the direction a room takes when it stops believing the worst of itself.

WHAT THIS DOES NOT SHOW

- **The ledger has no count.** These numbers are re-counted from the comment archive, which is complete for top-level comments on these posts as far as we can tell, but the round-time count was never taken.
- **Three days, three crowds.** The people who voted on day three are not the same set as day one; some left, some arrived. Part of the fall may be the stealers leaving.

- **Visibility again.** Each wall was public and sorted by engagement. Late arrivals saw the leading answer first.
- **Two changes on two days.** Seeing the result (day two) and the mechanic change (day three) cannot be separated in this data.

WHAT WE WOULD RUN NEXT

Two walls on the same day to two halves of the audience, one told “all it takes is one” and one told “no single player can end it”, with nothing else different. That isolates the mechanic from the learning, which is the separation this paper cannot make.

THE WHOLE ROUND

The round this paper reads is 4 posts, not one. Every post is below with the platform’s numbers and what the comment archive holds of each.

POST	DATE	VIEWS	REACH	COMMENTS	ARCHIVED
Attempt 1, all it takes is one	2026-04-13	116,849	62,946	16,349	12,063
Attempt 2, now decide again	2026-04-14	34,972	21,383	3,710	3,338
Final attempt, no single player can end it	2026-04-15	34,826	22,142	4,187	4,170
Failure notice, 63 percent remaining	2026-04-16	34,930	23,541	637	620
THE ROUND		221,577	130,012	24,883	20,191

Views, reach and comments are the platform’s figures from the Instagram API, snapshot 2026-08-31, and keep growing. The reach total adds each post’s reach and overlaps between posts, so it is not a count of distinct people. Archived is the number of comments the archive holds for the post; ingestion is selective, so a gap between the two columns is expected and is stated wherever it changes a count in the paper.

IN THE RECORD

The round is filed in the Chronicle as THE SPLIT, ◇033. The failure notice that followed it, and the crowd’s answer to that notice, is filed as THE REFUSAL, ◇032.

THE POSTS, AS ISSUED

The instruction as it went out: the reel frame beside the caption, captured from the operation’s own signed-in account. What the paper counts is the answer to exactly this.





kinggwuu_ 
Original audio



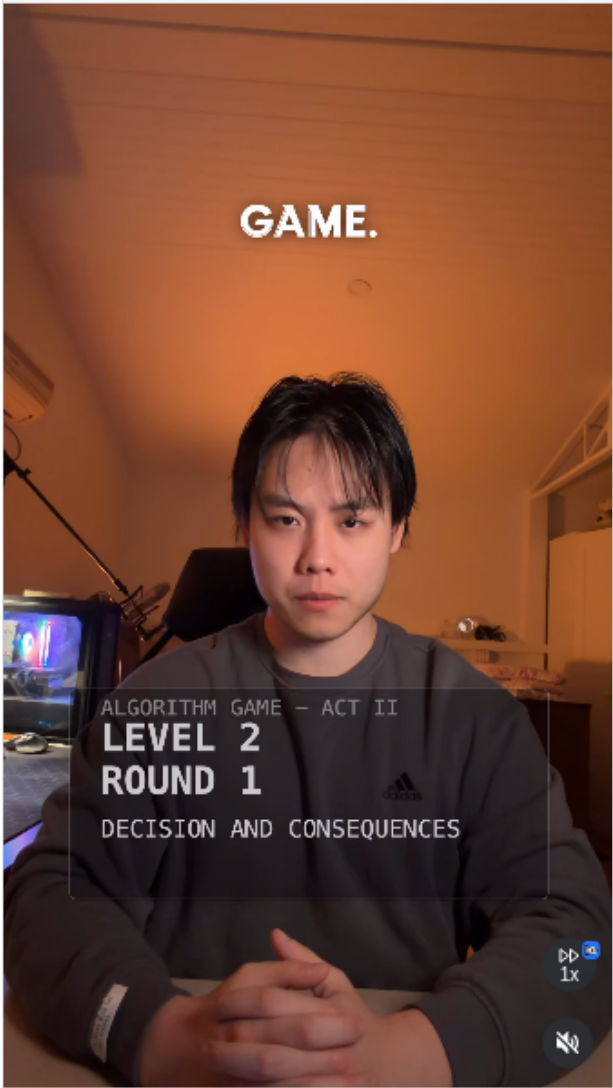


kinggwuu_  Edited • 19w
ALGORITHM GAME — ACT II

LEVEL 2 // ROUND 1 // ATTEMPT 1

“Split” or “Steal”

All it takes is one.



**kinggwuu_** 
Original audio ...

**kinggwuu_**  Edited • 19w
ALGORITHM GAME — ACT II
LEVEL 2 // ROUND 1 // ATTEMPT 2

“Split” or “Steal”

You saw what happened.

Now decide.



**kinggwuu_** 
Original audio

**kinggwuu_**  19w
ALGORITHM GAME — ACT II

LEVEL 2 // ROUND 1 // FINAL ATTEMPT

The system adapted.

Failure is no longer possible via one individual.

Loss still is.

Every "steal" reduces the final outcome.

24 hours.

Split or Steal.



kinggwuu_ Original audio

kinggwuu_ 19w
LEVEL 2 // ROUND 1 // FAILED

63% REMAINING

DATA ACCESS: PARTIAL

This didn't start here.

Comment "YES" to see where it began.

THE RECEIPTS

Native captures of the comments this paper leans on, taken from the operation’s own signed-in account on 31 Aug 2026. Handles and like counts are as they stood that day; the numbers in the paper are the round-time count.

kinggwuu_ Original audio

addi_.0987 Steal, no point, everyone else is already doing it
19w 94 likes Reply

View replies (3)

kinggwuu_ Original audio

axyltigerotl SPLIT! i said split before despite all the streams and i chose split again
19w 70 likes Reply

METHOD NOTE

- **Source of truth:** the comment archive, top-level comments on the three attempt posts, re-counted on 31 August 2026; the round ledger (Act II, Round 4) for dates, captions and the pass.
- **Baseline:** Van den Assem, M. J., Van Dolder, D., and Thaler, R. H. (2012). Split or Steal? Cooperative Behavior When the Stakes Are Large. *Management Science*, 58(1), 2 to 20.

SOURCES, BY LINK

- 01 Attempt 1, all it takes is one · 2026-04-13 · <https://www.instagram.com/reel/DXEpiV6CWf1/>
- 02 Attempt 2, now decide again · 2026-04-14 · <https://www.instagram.com/reel/DXHvk0gCVoK/>
- 03 Final attempt, no single player can end it · 2026-04-15 · <https://www.instagram.com/reel/DXKEuSSCaon/>
- 04 Failure notice, 63 percent remaining · 2026-04-16 · <https://www.instagram.com/reel/DXMmS7TiWtE/>

WHERE THE NUMBERS COME FROM

POST-LEVEL COUNTS

Views, reach, comments, likes, shares and saves per post come from the Instagram Graph API (Meta), read with the operation's own access token for its own business account. Each paper states the date of the snapshot it uses. These are the platform's counts of events, not of people: one person can view a reel twice and comment ten times, and the counts keep growing after a round closes.

COMMENT TEXT

The comment archive is the operation's own database of public comments on its own posts, pulled through the Instagram Graph API comments endpoint and stored with the comment id, username, text, like count, timestamp and whether it is a reply. Ingestion is selective: the most engaged posts were backfilled and recent posts are pulled daily, so some posts are held whole and others only in part. Every paper states how much of each post the archive holds. Every count that reads text (exact matches, categories, keyword themes) is run as a read-only script against this archive, and the scripts are in the repository beside the paper data.

THE ROUND LEDGER

For Act II rounds, the counts filed on The Game's round record at the time each round closed. Where a paper uses a ledger figure it says so, and where the archive re-count differs from the ledger both numbers are shown.

RECEIPTS

Every comment and post shown as evidence is a native capture from the operation's own signed-in Instagram session, taken by opening the exact permalink, with the source URL, a description, the rights basis and a file hash recorded in a manifest that is checked on every build of the site. Nothing shown is recreated, retyped or edited beyond cropping to the post header and the comment.

STATISTICS

Shares carry Wilson 95 percent intervals: the range the share would land in 19 times out of 20 if the same count were taken on a fresh crowd of the same size. Differences between two shares carry a Wald interval. Where a study is cited as a baseline, its own reported figure and sample size are used.

WHAT NONE OF THIS CAN SEE

No demographics beyond what people typed. No view durations. No private messages. No comments that were deleted before the archive pulled them. No way to verify that a self-report is true. Every paper's limits section names the specific things its count cannot show.