

THE BEHAVIORAL LENS ON ENGINEERING THE COMMONS.

Engineering Attention

A Design for Holding a City's Attention Without
Taking It

Joe Hamilton
Cityverse, St. Petersburg, Florida
Version 9.1 · September 2026

Hamilton, J. (2026). *Engineering Attention: A Design for Holding a City's Attention Without Taking It*. Cityverse.

Free to share and quote with attribution. cityverse.life/papers/

Abstract. Engineering the Commons described how a city's attention is assembled: what is built, what is given away, and why the gathering pays for the giving. It also set the condition the whole design depends on. A shared layer for a city creates value only when people show up to it every day. This paper is about the showing up. It describes what happens once a person has arrived at Cityverse: the machinery that brings them back, holds them, and lets them go, and what separates a version of that machinery that serves the person from one that consumes them. That is a question about behavior, and a reader is entitled to ask it the moment a design admits it is engineering habits at all.

This paper answers it. The machinery of habit is the same machinery that runs the national platforms a city lost its attention to, and no platform holds attention without it. The two differ in what the machinery is aimed at and in who keeps what it produces. Aimed at what a person watches, it extracts. Aimed at what a person does, it serves. Doing has two halves here, and the paper weighs them equally: what a person does in the city, at the meeting and in the stands, and what a person does on the platform, in the rooms and tools built around the things they already care about. Cityverse uses the machinery in the open, says so, and points it at the one thing a national platform cannot supply: a city, and a person's own part in it.

1. What the feed is for

A woman sits down after dinner to check whether the school board moved the vote. She opens the application on her phone. Forty minutes later she puts the phone down. She does not know whether the vote moved. She has watched a stranger's kitchen renovation, read two hundred words of an argument between people she has never met about a policy in a state she has never visited, seen four photographs of a former colleague's new baby, learned that a restaurant she cannot name is opening in a city she does not live in, and felt, for about ninety seconds in the middle, something close to anger at a man she will never encounter. She is tired in a way she was not tired when she sat down. She will do it again tomorrow.

Ask her what the feed is for and she will say it is for keeping up with people. Ask her what she kept up with and she will have trouble answering. That gap is the product working as designed.

Engineering the Commons described what a newspaper actually sold. It sold the assembly: a town's attention, gathered in one place, priced by the half page. The reader received something in exchange for being assembled. She received the forty facts, the box scores, the obituaries, the zoning notice, a common set of building blocks for a day lived somewhere specific. The assembly was the product, and the reader was paid for her part in it with information she could act on.

The feed runs the same trade with one term removed. The assembly is still the product. The reader is still assembled and her attention is still sold. What she receives in return has changed. She receives the sensation of having kept up, which is not the same as having kept up, and she receives it in a form engineered so that the sensation runs out a little before the session does, so that there is always one more thing to look at before she is done. The gap is measurable. When researchers compared what an engagement ranker surfaces against what the same users say they value, the ranker amplified anger and out-group hostility, and the users rated that content lower than the content they would have chosen. The machine shows people what they react to, which is a different thing from what they want.

The feed is for being returned to. Everything else it does, the connection and the news and the entertainment and the argument, is instrumental to that one end, because return is the only thing the advertiser is buying. The newspaper needed you

back tomorrow and gave you a reason. The feed needs you back in eleven minutes and has found that a reason is optional, because it can biologically engineer your return.

Look at what the feed actually contains once you strip away the framing. On the legacy platform most residents of a mid-sized American city still use, the content sorts into two piles that have almost nothing to do with each other. One pile is the reason people say they are there: the baby pictures, the graduation, the new job, the old friend's dog. The other pile is drama. Most of it is people fighting, in the specific register of strangers who have discovered they can say to each other what they would never say across a table, but the pile also holds the stranger's grief, the neighbor's medical fund, the breakup narrated in installments, the missing dog, anything interpersonal enough to need a next episode. The fight is the largest and cheapest part of the pile, and this paper will often let it stand for the whole. The pile is drama.

Look closer at the second pile and it turns out to be a serial. The same names appear in it week after week. The argument that started over a school board vote in March is still running in June with the same six combatants, and each new post is an episode: a reply that lands, a reply to the reply, a third party arriving to take a side. Most people reading it would not call it entertainment, and it works on them exactly as entertainment does, as a program they cannot stop watching. Every episode opens a loop and promises that the next one will close it, and a person who has been handed an open loop will stay to see it shut. The need to finish a story that has been started is one of the most reliable compulsions a person has. A cliffhanger works on everyone. An argument between known parties with no resolution in sight is a cliffhanger renewed every day, and the reader tunes in for the next episode the way they would tune in for any other. Nobody at the platform wrote the serial. The platform built the conditions that produce rage, the rage produced the fighting, the fighting produced a story with recurring characters and no ending, and the story produced the forty minutes. The drama is manufactured from the bottom up, and the viewer cannot look away from a thing that was built so they could not.

The first pile is the platform's stated purpose. It is why it exists and why anyone signed up. The second pile is why anyone stays past the third baby picture. Remove the fighting and you are left with a pleasant, low-frequency service that most people would check on Sundays. The platform knows this. It is why the fighting rises, and why it is allowed to run for months.

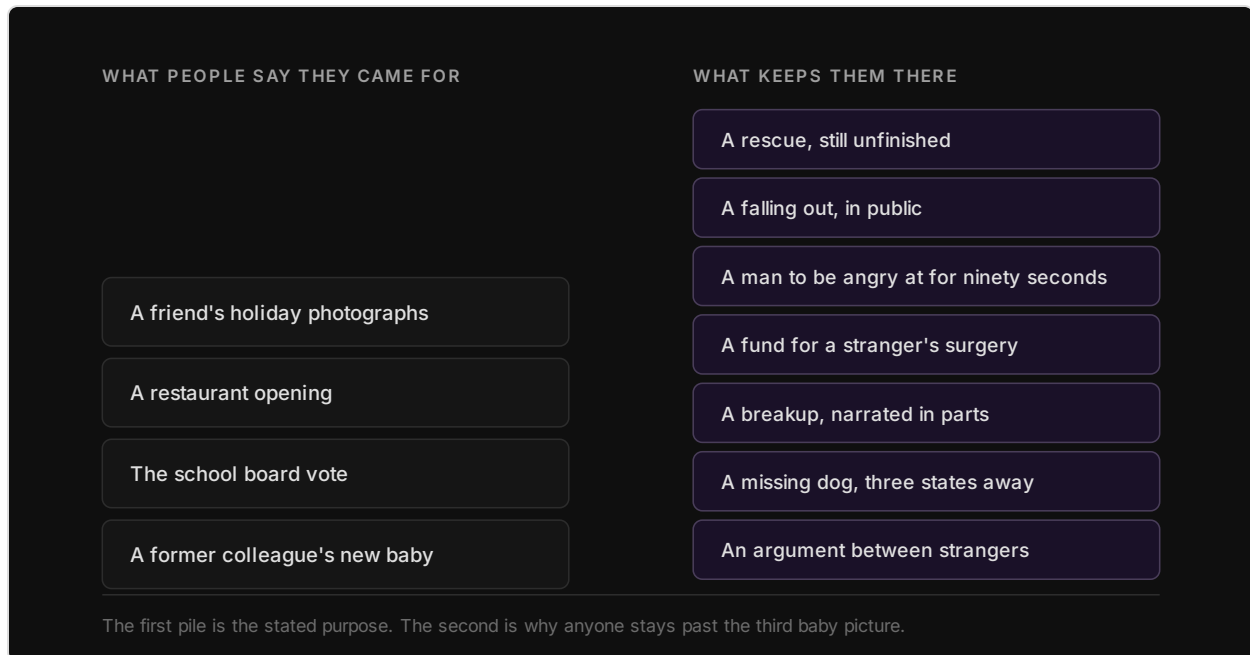


Figure 1. The two piles. On the left, what people say they came for. On the right, what keeps them there.

That observation is the one this paper is built on, and it is better put as a question, because the rest of Part One exists to answer it. Would the baby pictures alone hold the hour? If the honest answer is no, and we think it is no, then every platform that intends to hold a city's attention, this one included, has to decide what it will put in the second pile. The incumbents put fighting there because fighting is cheap, self-supplying, and reliable. The second pile will exist. The question is what goes in it, who decides, and who is better off at the end of the forty minutes.

2. One in ten

Picture a hunting party that succeeds one time in ten. Nine days they walk back to camp with nothing. On the tenth day there is meat, and the feeling attached to that day is euphoria, out of all proportion to the meat, and the reason is the nine days that preceded it. The success is measured against the misses, and the misses are what make it a success at all. A party that found game every morning would feel nothing in particular about game. Within a season the daily kill would be the floor, and only a large one would register.

That story is close to how the mechanism came to be. In the animal world, the pull of an uncertain reward looks like an adaptation to food that was scarce and unpredictable. An animal that keeps foraging through a run of failures eats, and one

that gives up starves, so the failures themselves came to sharpen the drive rather than dull it. Patrick Anselme and colleagues have argued on this evidence that the attraction to gambling is a foraging instinct running in an environment it was never built for. The mechanism itself is well documented. Wolfram Schultz and his colleagues showed in the 1990s that the neurons which release dopamine fire for reward that was not predicted and stay quiet for reward that was. Give an animal a treat it expected and the signal is flat. Give it one it did not expect and the signal spikes. Give it a cue that reliably predicts a treat, and within a few trials the spike moves to the cue and the treat itself produces almost nothing. The system tracks surprise, and surprise is a relative quantity. It exists only against a baseline of what was expected.

B. F. Skinner had found the behavioral half of the same fact forty years earlier. Of all the schedules on which a reward can be delivered, the one that produces the most persistent behavior, the hardest to extinguish, is the variable ratio: a reward that arrives after an unpredictable number of attempts. Fixed schedules produce steady work and quick abandonment when the reward stops. Variable schedules produce something closer to compulsion. The slot machine runs on a variable ratio. So does the content feed.

Now put a person in front of a feed. Nobody at the platform decides that this post will be a miss and that one a hit. Nobody has to. Most of what any person posts is of no interest to most of the people who might see it, and a platform carries more posts in an hour than a person could read in a year. The raw material is noisy on its own. What the platform contributes is the machine that sorts it, and the machine has one instruction: show whatever produces the most reaction. It learns, from a billion thumbs, that a certain rate of things a person cares about, arriving among things they do not, keeps the thumb moving longest, and it settles there. Nobody told it what a hit is and nobody designed the ratio. The ratio is found, the way water finds a level, by a system rewarded for time and supplied with an endless mix of the relevant and the irrelevant.

The person is not a bystander to this. Someone who has learned what the feed can deliver wants the biggest version of it, the way a gambler wants the jackpot and not the small payout, and adjusts, without ever deciding to, what counts as one. Show him excellent things all day and by evening excellent is ordinary and only the extraordinary registers. Show him nothing for a week and a mediocre thing feels like a gift. The floor is set by the person's own recent experience, and the person keeps moving it upward in pursuit of the next real hit. This is the part the hunting

party leaves out. The hunters had no say in their odds. The person with the phone, in a real sense, sets his. He is chasing the feeling, the feeling requires misses to exist against, and so he tolerates, and even seeks, a stream that is mostly misses, because that is the only kind of stream in which a hit still feels like one. The people on a subway car staring into their phones resemble the people in front of the machines in Las Vegas, pulling the lever for hours, and for the same reason.

And he is running fifty hunts at once. He follows boxing and he follows what his friends are doing and he is single and looking and he cares about his sports team and his neighborhood and a dozen musicians and thirty things he could not list if asked. Each is its own hunt with its own odds and its own kind of hit, and the feed carries them all at once, interleaved. The five fights he passes on the way to the one knockout are, in the same scroll, the five faces on the way to the one he would want to ask out, and the four dull updates on the way to the friend he actually wanted to hear from. *No single thread pays off often enough to become predictable. Together they pay off constantly.* The misses in one hunt are the hits in another, and the person is never bored long enough to put the phone down. That is the infinite scroll, and it required no one to design it. It required volume, a ranker rewarded for reaction, and a person who wants the jackpot.

None of this changes in a city. A city produces more in a day than any one person can take in, exactly as a national feed does: the council met, the team won or did not, the permit was filed, the restaurant opened, the road closed, a neighbor died, and a hundred smaller things beside them. Most of it will be a miss for most people, and the person's floor will move as it moves anywhere, because that is how a person works and no design changes it. What a city changes is what the hits are made of. On a national feed the hit is a stranger's knockout, a face, a post from someone three states away: real to the feeling and inert to the life. Here the hit is something that happened two miles away and can be acted on. It is the euphoria of the tenth day, and a reason to go outside.

The city has many hunts too. Sports, the arts, the council, the market, the school: each is its own thread with its own real successes, and chapter seven describes how Cityverse carries them side by side without grinding them into one stream. The layering that keeps a person on a national feed is available here, and here every thread is a thread of real things. The supply of them is not fixed, either. Where the city has a thing worth caring about and no place for it to live, Cityverse builds the place. The birding club that had a mailing list and nothing else gets a room, a calendar, a record of who showed up and what they saw, and that is a new thing

worth knowing, made where there was none. The infinite scroll keeps its ratio alive by never running out of the almost-right. Cityverse keeps its ratio alive by adding to the supply of the real, and lets the odds find their own level among things a person can act on. The task is to make sure that when a real hit lands, it lands well, and that it lands on a person who can do something about it.

3. Out of the tube

A reader who has followed the argument this far may be forming an objection, and it is the right one. If the machinery of the feed is a ratio managed to produce compulsion, then the decent thing to do is not to run it. Build the pleasant, low-frequency service. Let people check it on Sundays. Refuse the second pile.

This chapter exists to explain why that option is not available, and it can be short, because the argument rests on two facts the reader already has.

The first fact is in the reader's own hand. The baseline described in the previous chapter has already moved. Anyone who has used a phone in the last fifteen years carries a calibration set by the most heavily engineered products ever built, and that calibration does not reset when they open something gentler. An unengineered service reads as flat. The tenth-day euphoria has been distributed across every idle minute of the day, and a product that offers less of it is experienced as boring, and boring is abandoned, quietly, without a decision. That is the mechanism from chapter two applied to a whole population, and it has nothing to do with character.

The claim needs one boundary before it is used. Products people open because they must, or because the product does a job, hold attention without any of this: a bank, a weather forecast, a calendar, the tool a person's work requires. They are not competing for the evening. Cityverse carries utility of that kind too, and chapter ten returns to it. The point here is narrower. A platform that wants the discretionary hour, the one a person spends because they want to and could spend anywhere, is competing with the most heavily engineered products ever built, and that is the hour this paper is about.

The second fact is about the hour itself. There is one evening. It is finite, it is contested, and the competitors for it are engineered to the limit of what is legal and measurable. A platform that declines to engineer for it forfeits the hour to the

products that did not decline, and forfeits the people with it. Unilateral disarmament in an attention market produces a smaller share for the party that disarmed and no change whatsoever in the behavior of everyone else.

Engineering the Commons made a version of this point about the newspapers. The diagnosis of theft implied a remedy of restitution, and no restitution was coming. The diagnosis of product loss implied the harder and more useful answer: build a better product. This paper adds the corollary the earlier one did not have to state. The better product will also be engineered. There is no unengineered option that anyone will use. *The choice a builder actually has is where to aim the machinery, for whom, and whether to say so.*

So Cityverse runs the machinery, and it says so. That is the stance: transparency about a set of tools that every competitor uses and none describes. Abstinence is not on the menu, and the mechanism is not the harm, so there is nothing to apologize for. The description is treated as part of the product. A resident who understands why the platform is arranged the way it is has been given something the incumbents withhold as a matter of policy, and understanding a mechanism is the beginning of being able to judge who it serves.

The rest of the paper is that description.

4. Who stays home

Engineering the Commons made an argument about free speech, and it bears repeating here.

The argument was this. Imagine a store that announces that on Wednesdays, food is free. Consider who eats. The fastest arrive first. The largest carry the most. The most willing to push get to the front, and after a few weeks the people who dislike being shoved stop coming. The store is still, technically, giving away food to anyone. In practice it is giving away food to a particular kind of person, and the fact that the policy is open to everyone has become irrelevant to who benefits.

That is what an unmoderated platform is. The right to speak is distributed equally, and the fact of speech is not, because the resource being handed out is attention and the people best at taking it are the loudest, the most extroverted, and the most comfortable with conflict. The earlier paper put the cost in one line: unmoderated access is a subsidy to the most aggressive, paid for by the silence of everyone else.

The professor with the most accurate information about the water table stays home. She has read the replies other people get and has correctly concluded that the exchange is not worth it.

The earlier paper drew the economic conclusion. Gravity is mass, a platform that drives out the careful has less mass, and moderation of conduct is therefore a supply mechanism rather than a moral position. This paper accepts that conclusion and asks a different question. Why does she stay home? As a matter of mechanism, what, specifically, is the machinery of the feed doing to her, and what is it doing to the fifty people who would have read what she wrote?

The sociology has a name for the first half. Elisabeth Noelle-Neumann called it the spiral of silence. People carry a continuous, largely unconscious estimate of what opinions are safe to hold in public, and they update it from what they see expressed around them. Those who believe their view is losing ground go quiet, which removes their view from the visible sample, which convinces the next person that the view is rarer than it is, which quiets them too. The spiral runs on a distorted sample of what people think, and it runs faster the more distorted the sample becomes.

Now look at what the feed does to the sample. It does two things, and they both point the same way. This paper will call them the door and the ranking, because it will need to refer to them again.

The door comes first. A platform where the reply to a careful comment is a pile-on selects, before anything is posted, for people who can absorb being attacked. That is a small subset of the city: the combative, the loud, and those with nothing to lose. The professor is not in it. Neither is the retired engineer, the nurse who has seen the emergency room numbers, or the parent who sat through the last four school board meetings. They have plenty to say. They have looked at the room and decided, correctly, that it is not built for them. The technical freedom they enjoy to speak does not produce actual speech, and that is the problem this design is solving for.

Then there is the ranking. Chapter two described a ranker that finds, on its own, the mix of hits and misses that keeps a thumb moving. Conflict is the cheapest hit there is. It requires no production, it supplies itself, and it generates the two things a ranking system reads as success: comments and time. A fight in the replies is, to the machine, indistinguishable from a conversation that went well, and it is far more common. So the ranking learns to surface it, and it learns something more specific

than conflict. It learns that ongoing conflict between known parties outperforms a single flare-up, because a serial produces a session every day and a flare-up produces one. The ranker knows nothing about mini-series. It knows that the thread with the same six names in it for three months is the most reliable source of reaction on the platform, and it puts that thread at the top. The feed wants people reacting, and a fight that never ends is the most reliable way to get it.

Two filters, the door and the ranking, both selecting for the same temperament. The result is a room that appears to be the city talking and is in fact a small, self-selected fraction of the city talking loudly, amplified by a mechanism that rewards volume. The professor looks at that room, applies Noelle-Neumann's arithmetic, and concludes that her view is rarer and less welcome than it is. She stays home. The fifty people who would have read her never see the water table numbers. They see the fight instead, and some of them join it.

The earlier paper's conclusion covers the door. Moderating conduct is a supply mechanism; it keeps the professor's contribution inside the system. The ranking is the filter that argument does not reach, and it is the one a platform is most tempted to leave alone, because a ranker that rewards reaction is a ranker that reports growth. A platform civil at the door and hungry for reaction in the ranking has fixed half the problem and will not notice, because the half it left is the half its dashboard measures.

So Cityverse closes both. At the door, every post from every member, comments and everything else, is screened at the moment it is written, for conduct rather than viewpoint: the attack on a person, the sweeping claim about a class of people, the pile-on. Most of what is caught is a clause rather than a contribution, and the author is shown the clause and offered to drop it while the rest publishes untouched. Every action can be appealed to a person. In the ranking, the fight is not allowed to be the cheapest hit. Cityverse's feed is called the Content Flow, the Flow for short, and it moves differently from the feeds the reader knows. A person advances through it one piece at a time, with a deliberate slide between pieces, instead of the bottomless scroll. The Flow's objective is whether members say an item was worth their time, with a floor under civic content so the council vote cannot be buried by whatever was loudest that evening, and with a lift for content that earns approval from people who otherwise disagree. And a dispute the city is actually having is gathered into one conversation, which Cityverse calls a Super Thread. Chapter eleven describes how it works. What matters here is what it does to the serial. The conversation carries a running account of where the city stands, and when the city lands

somewhere, it records that. The argument may go on, as arguments in a city do. The record of it is kept in one place, and a person can leave it and come back without losing the thread or being made to re-fight it. The answer to the serial is less an ending than a destination: the conversation points at the vote, the meeting, the thing a person can do about it, and chapter eight describes what a person earns for crossing that bridge. Conflict still happens here. It is no longer paid, and it is not renewed every morning for the ratings.

One detail of the design is Noelle-Neumann's finding turned into a control. A member who wants less of something can say so, privately, and the platform adjusts their own Flow. What that member said is never shown to the author, and never aggregated into a count the author can see, because it is the perception of disapproval that silences, and a hidden tally of "fewer of these, please" would rebuild the spiral this design exists to break. What protects the professor is a room in which the disapproval of strangers cannot reach her as a number.

That is the answer to who stays home, in outline. Part Two takes each piece in turn, and one chapter is about her specifically: what has to be true of a room before a person who has decided it is not built for her walks back in. Chapter four names the people who stay home. Chapter eleven is about who will speak. Chapter twelve names the hardest version of the problem, which is that a room built differently will still be judged, on its first day, by people who assume it is the room they already know.

5. The same machine

Here is the argument in one paragraph.

The loop that runs a slot machine runs a feed, and it runs the small reward badge Engineering the Commons described for a parent who has attended four high school games, the one that brings her to a fifth. The structure is the same in all three. A cue arrives. An action follows. A reward comes. Something accumulates that makes the next cycle more likely. Nir Eyal described it as trigger, action, variable reward, investment, and it is an accurate description of every product that has ever been called addictive and of a great many that have not. One setting differs, and it differs on purpose. In the slot machine and the feed the reward arrives on a schedule that is not quite predictable, and the unpredictability is the engine. In the badge it arrives exactly when it was earned. That is a large change to make to a mechanism whose

power comes from surprise, and it needs a reason. The reason is where the surprise already lives. The game has a score nobody knows in advance. The meet has a race. The council vote can go either way. An act in the world carries its own uncertainty, and a reward attached to that act does not have to manufacture any. What the reward has to do instead is be reliable, because a person who cannot predict whether the thing they came to do will count does not do it. So the badge is exact, the surprise stays in the world and in the few places chapter eight describes where a surprise is a gift, and the loop loses nothing. The neuroscience is blunt about which part matters. Brady and colleagues showed in 2021 that moral outrage online is learned the way any behavior is learned, by being rewarded, and that the reward signal, more than the belief, is what escalates it. Change the reward and the behavior follows. The products differ in two things outside the loop: what it is aimed at, and who keeps what it produces.

Consider the aim first. A loop can be pointed at what a person watches or at what a person does, and doing has two halves that this paper weighs equally. The first half is off the platform: the parent who comes to the fifth swim meet, the resident who votes, the neighbor thanked, the shop walked into. The second half is on the platform, and it is easy to miss, because the incumbents have trained everyone to think of a screen as a place for watching. A Rays fan on Cityverse can attach a Fan Bar to her Homespace, which is her page on the platform. The Fan Bar is a set of tools. She uploads a photo and becomes an avatar in a Rays jersey. She keeps her season tickets and her all-time favorite player there. She builds a trivia quiz that other fans take, and she curates a list of what she has to trade. When the team runs a clubhouse night she is on the stage with the players, asking her question. Every one of those is a thing done, and none of them is a thing watched. Both halves keep value with the person and with the city, and both halves are where Cityverse aims the loop. Attention, on this platform, is the discovery-and-fuel layer: how a person finds the thing worth doing and is drawn back to do the next one. Chapter ten gives the experience a name, civic efficacy, and explains why it does not wear out.

Return matters to both loops, for opposite reasons. A platform aimed at watching wants the person back because the next session is the product. Cityverse wants the parent back too, and not only because Thursday's game is coming. She comes back to close the loop on having been in the stands: to have the fifth game recognized on her page, to see who else was there, to find out what is next. The application is

where the doing gets counted and where the next doing gets found. The difference is that the return is in service of the act, and the act is in the world or in a room she is building something in.

The two aims produce opposite designs from identical parts. Aimed at watching, the surprise is turned up until it approaches compulsion, because compulsion is long sessions. Aimed at doing, the surprise is kept away from the act itself, because a person who cannot predict whether the thing they came to do will count does not do it. Aimed at watching, the cue is manufactured, because a manufactured cue is another session. Aimed at doing, the cue is tied to a real event, because the event is the thing to do. It is the same loop, and the opposite product.

This is also what separates the design from a casino, which wants its customers back tomorrow as much as any civic platform does. A casino is aimed at hours in a chair. So is a cigarette, and so is a game built to bill a child's parents. None of them is aimed at anything the person does, in the world or on the screen. A person building a quiz for other fans is doing something. A person pulling a lever is having something done to them. That, rather than return, is the line.

Now the second thing, which is who keeps the value. Every one of these loops produces something. The question is where it goes.

A loop aimed at hours produces attention, and the attention is sold. The person who supplied it did get something, and the honest version of this argument has to say so plainly. The hour was compelling. Parts of it were enjoyable. Between the fights there was news of a friend, a joke, something useful. The trouble is in how the enjoyment was built. The feed is engineered to open loops and close them, on its own schedule: an argument that continues, a thread that promises a reply, a count that might have moved. The compulsion to close a loop someone else opened is what keeps a person there, and the platform builds the loops because the compulsion, and nothing else, is what it sells. So the accounting is lopsided in a specific way. The value the loop produced left the person and went to the operator, and what the person kept was the tiredness and a sensation of having kept up. That is what extraction means here: the person got less than they put in, and the difference went elsewhere.

A loop aimed at doing leaves the value where the doing happened. The parent who came back for the fifth swim meet has a record of having been there, a daughter who saw her in the stands, and a city that now has one more person who knows the name

of the team. The fan who built the quiz has a quiz with her name on it that other fans are taking, a standing in the room that she earned, and a team that knows she exists. The badge is small and what it attached to is large. The platform profits from the attention too, exactly as Engineering the Commons described, but the profit is downstream of a person having gotten something real, rather than instead of it. The recognition stays on her page. The relationship stays in the stands, or in the room. The city keeps the parent and the fan. Nothing left the person on the way to the operator.

The person is not the only one who keeps something. The parent in the stands raises the temperature of the meet. A gymnasium with people in it is a gymnasium other people walk into, the way a restaurant with a full parking lot draws the car that would have passed an empty one. Her presence is worth something to the swimmers, to the team, to the school's athletic program, and to a city that now has one more full room on a Tuesday night. Set that against the same hour spent watching a celebrity's morning routine. That hour produced a session for a company in another state and nothing for anyone within a hundred miles of her couch. The loop aimed at doing leaves value with the person and with everyone in the room around her. The loop aimed at watching leaves it with the operator.

Cityverse fixes this in one place. Paid placement does not enter the ranker at all. It is fixed in position on a set cadence, outside the algorithm, with no personalization, so the machine that decides what a resident sees never learns what sells. A ranker that cannot see the money has no way to bend toward it. That is the difference between a platform that promises the Flow is not for sale and one that has removed the mechanism by which it could be.

This is the test the reader should carry into Part Two and apply to every mechanic described there. The question is never whether something is behavioral engineering. It is, all of it, and so is everything the reader has used today. The test is: what is it aimed at, and who keeps the value? A badge that brings someone back to a real event they are glad they attended passes. A streak that punishes someone for missing a day fails, because the value it produces is the person's anxiety and the operator's session count. A notification that the council voted passes. A notification that someone you barely know posted for the first time in a while fails. A quiz a fan built for other fans passes. A count she checks to see whether it moved fails. A conversation that gathers a dispute and records where the city landed passes. A

thread kept alive for three months because the same six people cannot stop fails, and is the single most profitable object on the platforms that host it. The mechanism is the same in each pair. The test distinguishes them anyway.

Engineering the Commons argued that a platform which drives out the careful has less mass, and that the mass it keeps is the least useful kind. Put that in hours. A city should prefer eighty hours of attention from executives, officials, and active residents to a hundred hours from people watching each other fight, because who supplies the hour matters more than the hour. The incumbents each live off a slice of people: friends, neighbors, shoppers, the audience of a celebrity. Cityverse lives off the engaged citizen, and the hour that citizen gives is worth more to a city than the hour of someone watching strangers fight. This paper has tried to show the mechanism behind that preference. The hundred hours are what you get when you aim the machine at watching. The eighty are what you get when you aim it at what people do. They are produced by the same loop, they are measured on the same dashboard, and only one of them leaves the city with more than it started with.

The manipulation objection, then, in its strongest form: this design admits to running the same machinery as the platforms it criticizes, and dresses it in civic language. The answer is that the objection assumes an alternative, a neutral platform that runs no loop and shapes no behavior, and chapter three showed that there is no such platform and never will be. Every product that holds attention runs the loop. The only builders who can truthfully claim not to are the ones nobody uses. So the objection, correctly stated, is a question of how rather than whether, and how is a question with an answer. Aim it at what a person does. Keep the value with the person. Say what you are doing. The chapters that follow describe what that looks like, one mechanic at a time.

PART TWO

A note on status. Part One rests on established psychology. Part Two describes mechanics, and the mechanics are hypotheses about how that psychology behaves inside one platform. Chapter twelve says what would prove them wrong.

6. What calls you back

Every habit begins with a cue, and the incumbents learned early that the cue is the cheapest part of the loop to manufacture.

Consider what arrives on a phone in an ordinary evening from a platform aimed at hours. Someone you met once has posted for the first time in a while. A person you do not follow has commented on a thread you do not remember. Your streak is at risk. There are new posts from people you may know. Twelve people reacted to something. None of these is false, exactly, and none of them is nothing. Each is a small episode of the serial from chapter one, delivered to a pocket: something happened, someone is involved, and you are in it. They exist because a notification is a session and a session is revenue, and a growth team that can produce an episode on demand will do so until a metric tells it to stop. The metric never does, because the metric is sessions.

It would be comforting to say that this wears off, that a person learns the cues are empty and stops answering them. The evidence runs the other way, and has for a long time. The platforms have run this cue for most of two decades, and the people who complain about it most are still checking. What has happened is that the platform invented a new category of event and taught a population to feel it as one. Someone reacted to your post. That did not exist before, it costs the platform nothing to produce, and it now lands in a person's evening with the weight of a real occurrence. Call it manufactured, and it is. Call it nothing, and the feeling disagrees. It is real to the feeling and inert to the life, exactly as chapter two described the national hit, and it works for the same reason. The cue opens a loop. Not knowing what was said is worse than knowing, and the only way to close the loop is to go and look. The platform has injected its serial into a person's day, and the person, compelled to see the episode, walks back in.

There is fatigue, and it pays to be exact about what causes it. People tire of the volume and the intensity, of forty cues a day where four would do, and the platforms know it, because they throttle. What people do not do is stop. Ask someone who has turned off notifications for an application they still open twelve

times a day. The cue did not lose its power when it was silenced. It moved. Chapter two described what Schultz found: give an animal a signal that reliably predicts a reward and the spike migrates from the reward to the signal. Here the signal is the phone itself. After enough years of episodes arriving through it, the object in the hand is the cue, and it fires without a notification, in an elevator, at a red light, in the pause after a sentence. That is the endpoint the manufactured cue was working toward all along. The platform no longer has to send anything. The person carries the trigger.

So the cue is not the product. The serial it opens is the product, and a platform's serials are made of whatever that platform can reliably produce. Here the national platforms have a problem that has nothing to do with the supply of real events. They have every real event in the world. Games end on them by the hundreds every night, votes are called, storms make landfall, and none of it is theirs. The platforms create no content of their own, and their utility is thin, a marketplace and an events tab. What they can produce, reliably and on demand, is what their members make for one another, and the most reliable thing members make for one another is drama: the fight first, because it costs nothing to produce and renews itself every morning, but also the grief, the confession, the falling out, the rescue, anything interpersonal enough that a person has to know what happened next. So the serial a national cue opens is, most of the time, the drama loop of chapter one, and it is that way for a structural reason. Drama is the one product the platform manufactures itself. Everything else passes through.

Cityverse is built on the other footing, and this is a fact about the platform. St. Petersburg has no more real events than any other place its size, and every one of them is available to the national platforms too. What Cityverse has that they do not is a newsroom that writes the council vote and the game recap, and a set of tools that carry the calendar, the Fan Bar night, the market, the club. Those are the platform's own products, produced on a cadence the world sets and the newsroom keeps. So the serial a Cityverse cue opens is made of things that actually happened, because the platform itself made the account of them, and it can be relied on to make the next one. Engineering the Commons noticed this about the sports chain, that the attention it produced "recurs on a schedule set by the athletic calendar." Generalize it. The city's real calendar is the notification schedule, and Cityverse never has to invent an episode because it produces the real ones.

That is the whole difference, because both kinds of cue compel a return. A national cue opens an episode of someone else's drama, injected into your day so that you will come back and close a loop the platform opened and the platform owns. A Cityverse cue opens an account of something that happened, and closing the loop means learning the fact and, often, having something to do about it. The test from chapter five applies directly. A cue that answers to a real event passes, because what it leaves with the person is the fact, and the fact stays whether or not they open anything. A cue that answers to nothing but the need for a session fails, because what it leaves is an open loop and a longer evening, and the value went to the operator. The pull is the same pull. It opens different stories.

A real event is a necessary condition for a cue and it is a long way from a sufficient one, because reality supplies its own sensations. A shooting, a scandal, a fire three neighborhoods over are all real, and they produce stronger reactions than a zoning update or a school play. A ranker fed real events and rewarded for reaction would rediscover the incumbents' feed using nothing but the truth. Two things in the design stand in the way. The Flow's objective is whether a person says the item was worth their time, and the reaction an item produced is never the quantity being maximized. And the anchor of the day is set by a newsroom, whose editors decide what the city needs to know and in what order, with the ranker working beneath that judgment and never above it. Real is where the cue starts. Worth a person's time is what it has to pass.

There is a second kind of cue in this design, and it points the other way, at the person about to write rather than the person about to read. Every post is screened as it is composed, and when the screen finds something it would act on, the author is told before publishing. A clause that attacks a person rather than an idea is highlighted and the author is offered to drop it while the rest publishes untouched. If the conduct in a post means it will be shown to fewer people, the author learns that at the moment of writing, with the reason, so that the outcome is known and editable before anything goes out. For a reshare of something the person has not opened, a prompt asks whether they would like to read it first.

These are cues in the strict sense: a signal at the moment of decision that changes the decision. The evidence for them is modest and consistent. A field experiment comparing moderation before and after publication found that intervening before cut toxicity by about a quarter with no loss of participation, for a reason that is obvious once stated: people write differently when they know something is reading over their shoulder before the post goes out, and differently again when what reads

over their shoulder is offering an edit rather than a verdict. Platforms that have tried the reconsider prompt report a third or so of people shown it changing or withholding what they wrote. The read-first prompt, where it has been tried, increased how often people opened the thing they were about to share by about forty percent.

None of these is large on its own. Stacked, and applied to everyone rather than to new members only, they do a job that no amount of moderation after the fact can do, which is to move the intervention to the one moment when the author still owns the outcome. A cue at composition treats the author as someone who might prefer to be better, and most people, given the option before rather than after, are.

There is a third kind of cue, and it is the room itself. Toxicity reaches a comment section two ways. The toxic person posts something nasty. Then the reasonable person, who cannot always help it, answers, and now there are two. What breaks the sequence is what the reasonable person sees before they type. The St. Pete Catalyst has moderated its comments for most of ten years, and the pattern its editors watched was slow and then unmistakable. For about a year the work was batting away troll-style comments one at a time. Then fewer arrived, because the environment had informed the commenter before anyone had to. A thread of civil comments makes it safer for the sixth person to post, and the sixth person, posting, makes it safer for the seventh. Readers learned over those years that they could write on the Catalyst without being attacked, and most of them could not have said whether anyone was moderating at all. They had learned what kind of room it was. That is the cue at its most powerful and least visible. The room teaches. Cityverse is arranged to teach the same lesson faster: the screen at the door, the private controls, the ranking that will not pay for a fight, all of them are ways of shaping what a person sees so that what they write is already different.

7. The echo

Cityverse invents no new kind of content. It carries the four kinds people already use, and names them. A Zap is a short post, limited in length, with no image. A Frame is an image with a caption. An Epic is a full piece, the length of an article. A Vizzle is a short video. The four were built as direct parallels to the places Cityverse means to earn its traffic from, so that a person arriving from the platform of short

posts, or of photographs, or of short video, finds the mode they already know how to use. People take each form differently: a Zap in a glance, an Epic sitting down, a Frame with a look, a Vizzle with the sound on or off.

The design keeps the forms separate and lets a person move among them. Then it does the thing this chapter is named for. A piece of content, usually an Epic, is parsed into the other forms automatically. The newsroom writes the story of the council vote; the platform derives a Zap from it, the vote in two sentences, and a Frame, the photograph from the chamber with a caption, and a Vizzle if the story carried video. Each derived piece wears a small mark, the echo mark, which says this is one form of something fuller. A person sliding through Zaps who wants more touches the mark and the full Epic opens over the Zap. When they have finished, they close it and are back on the Zap, sliding to the next one. Nothing about their place in the Zaps has moved.

That is the front of the mechanism. The back is what the platform remembers. A person who has read the Epic, whether they opened it from the Zap or found it among the Epics, has experienced that piece of content, and it will not resurface for them in any other form. Browse the Frames later and the chamber photograph is gone from the new ones. Browse the Epics and the council story is already read. What a person has seen travels with them across every form, as one fact about the person rather than four. Nothing is missed, because every form of a piece points to the fullest form of it, and nothing repeats, because the platform carries one record of what the person has experienced across all of them. That is the echo.

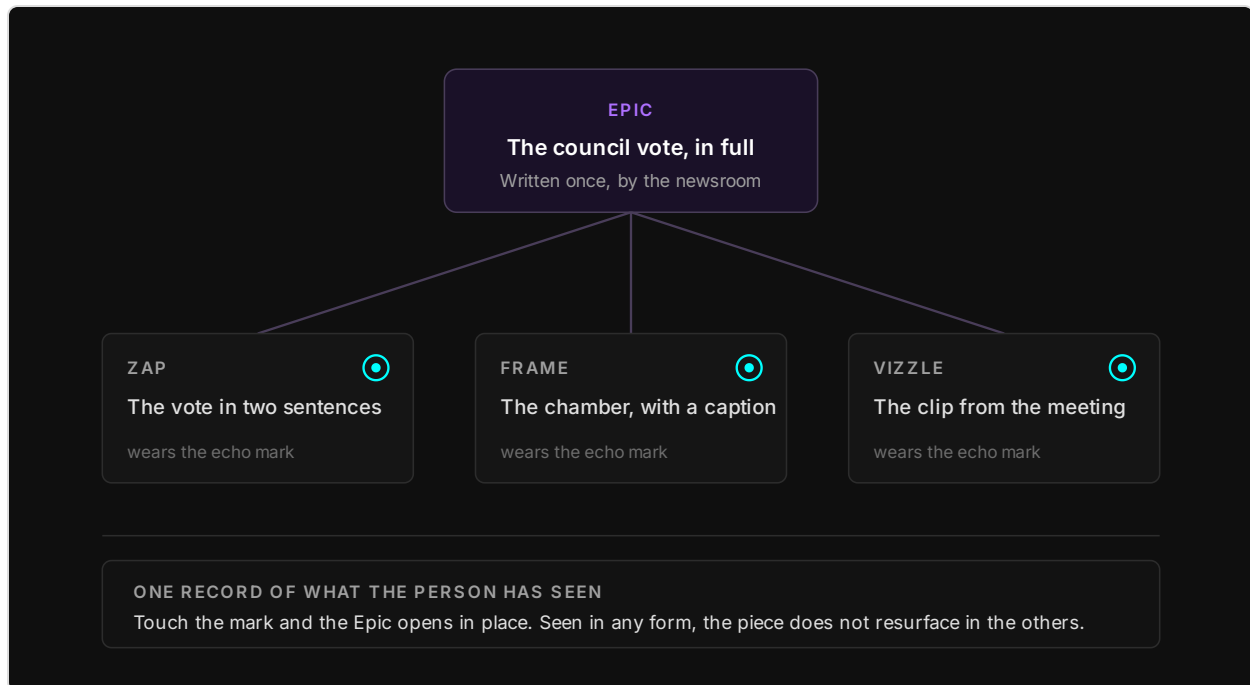


Figure 2. The echo. One Epic is parsed into a Zap, a Frame, and a Vizzle, each wearing the echo mark. Touching the mark on the Zap opens the Epic in place; closing it returns the person to the Zap. Whichever form the person experienced, the piece does not resurface in the others.

It is also how the fifty hunts of chapter two are carried without being ground into one scroll. Sports, the arts, the council, the market each keep their own place, each with its own real hits and real misses, and the person moves among them knowing where they stand in each. The layering that keeps a person on a national feed is available here, and every place is a place of real things.

What this removes is a driver, and the incumbents depend on it.

The fear of missing something is one of the strongest cues an attention platform has, and it is entirely manufactured. A single infinite feed guarantees that a person cannot know what they have not seen, because the feed has no bottom and no memory of the person's position beyond the current session. There is always more. The person cannot finish. The anxiety that produces is the cue for the next session, and it costs the platform nothing to maintain because the structure of the feed generates it automatically. Repetition works the same way from the other side: the same event surfaces as a post, then as a reshare, then as a comment on the reshare, then as a screenshot of the comment, and the person cannot tell whether they are seeing something new until they have looked, so they look.

The echo ends both, and it was built to. It is a deliberate answer to the thin, anxious version of not wanting to miss anything, and it works by giving the person what the infinite feed withholds: an edge. A person who knows what they have seen can finish. A person who will not be shown the same thing twice can trust that what they are shown is new. That is a smaller, quieter product, and it fails the test a platform aimed at hours would apply, because it produces shorter sessions. It passes the test from chapter five, because what it produces is a person who got what they came for and knows it, and who therefore has a reason to come back that does not depend on anxiety. The value it removes from the platform is the manufactured worry. The value it leaves with the person is the finished feeling, which the incumbents have spent fifteen years making sure nobody has.

Two further properties follow from the same design. The first is that the composition of what a person sees is a decision made in the open. There is a hierarchy of what anchors the day, news first, then the platform's own coverage, then what local creators have made, then what neighbors have posted, and a floor under civic content so that a quiet news day does not let the council vote be buried by whatever was most reacted to. No two consecutive items come from the same source. Time of day changes the mix. Every item in the Flow can say why it is there, in words, because a person who can see why they are being shown something is a person who can disagree with it, and a feed that cannot be argued with cannot be trusted.

The second is that the echo is how a person gets carried from one part of the city to another without deciding to travel. Engineering the Commons described a parent who arrived through the swim team and ended up reading the city council coverage. The echo is what makes that ordinary. She came for the meet results. The Flow that carried them also carried the council story, in the form she was already reading, and she took it in the same motion. Nothing asked her to go somewhere else. The city's other rooms were already in the one she was standing in.

8. The medal

A like is the incumbents' great invention, and it deserves a fair hearing before it is replaced. It is instant, it is social, it is cheap to give, and receiving one produces a small, real, measurable pleasure. It is the most efficient reward ever built into a piece of software.

It has one flaw, and the flaw is the mechanism of chapter two. A like counts agreement, and agreement is not scarce, so the count inflates. A post that would once have drawn ten now draws a hundred, and a hundred feels like ten did, and the person needs a thousand. The person is chasing the feeling the first likes gave them, and the feeling has moved. This is why the like, of all the mechanics, is the one that most reliably converts a habit into a compulsion. It has no ceiling.

There is a second flaw, less visible. A like rewards being agreed with, and being agreed with is easiest inside a group that already agrees. So the like, aggregated, rewards the poster for speaking to their own side, and it rewards the platform's ranker for surfacing what each side already believes. Paying anyone by attention received is the mechanism that manufactured the outrage the incumbents spent the last decade apologizing for. The reward signal was the lever, and it was pointed at applause.

Cityverse keeps the social reward and changes its unit.

One thing is kept from the incumbents, and kept deliberately. Cityverse has a thumbs up and a thumbs down. What it does not have is a published count. Nobody's post carries a number, so there is nothing to inflate and nothing to come back and check on. The signal still exists, and it is used: to understand what members find worth their time, to tune the Flow, to learn who the members are. The reward has been removed and the signal has been kept. There is an honest experiment inside that choice. The like's worst effects were measured in rooms built to reward reaction and to sort people into sides. Whether a plain up or down behaves the same way in a room with a civil floor and no silos is an open question, and Cityverse is one of the few places it can be tested.

The unit is the medal. A medal is a named track for a specific kind of real local activity: attending, volunteering, buying from the shop you are standing in, knowing the city, making things, bridging a disagreement, and the work a person does inside the platform's own rooms, running a Fan Bar night, building a quiz other members take, keeping a club's calendar current. Each has four tiers, and each tier is harder than the last in a way that is concave, so that early progress is accessible and mastery is rare. A medal is earned by a verified act, never by a reaction, and the verification is what makes it mean something. Nobody can give you a medal for attending four games. You attend four games.

What is rewarded is chosen. Cityverse continues the tradition of social reinforcement and points it at the two things the like cannot reach: earning approval from people who otherwise disagree with you, and demonstrating that you know what you are talking about. Content that draws support across groups that ordinarily oppose one another is lifted, in the Flow and in standing, more than content that draws applause from one side. Sustained contribution in a single domain over time, attested by someone other than the contributor, earns a recognition the design calls depth. These are harder than being agreed with. That is the point. A reward aimed at the hard thing produces more of the hard thing.

The medals compose into standing, and standing is where the design's most consequential choices sit. Every member is a Cityzen for as long as they are a member. Their standing rises through eight named levels, from Neighbor to Elder, and the level is what the public sees; the number underneath is available to anyone who wants to look, but it does not lead. Named tiers hold their meaning where numbers inflate, because a Builder is a Builder and a thousand of anything is a thousand of anything. Advancing requires medals spanning at least three unrelated families, and no single family may carry more than about forty percent of the way, so that no one kind of activity, and no amount of money, can take a person to the top alone. There are no city-wide leaderboards. A leaderboard at scale demotivates everyone below the top, and the standing this design confers is meant to signal belonging, the way membership in a long-running club does, and never victory.

Three commitments protect the mechanism from becoming the thing it replaces.

The first is that standing does not decay. Every precedent for decay in a reward system is a private, redeemable currency, where a falling balance is a nudge. A public, identity-linked standing that silently drops reads to everyone who sees it as a verdict, this person became less trustworthy, with no charge and no appeal, and the research on intrinsic motivation predicts that it suppresses exactly the behavior the system wants. So nothing earned is ever taken away. Freshness is carried by a separate seasonal meter that resets and is multiplied by permanent standing, and a person returning after months away is met with a welcome rather than a penalty. A streak, where the design uses one, has a grace period from the first day. There is never a bare cliff.

The second is that the surprise is quarantined. Every verified civic action produces reliable, immediate feedback, and that feedback is never randomized. A person who attends the meeting knows the meeting will count. The surprise, the bonus, the

unexpected recognition, all of which the design uses because they work, are confined to cosmetic and discovery rewards: a badge nobody knew existed, a bonus day, a partner's exception to a cap. The rule fits in one sentence, and it is the sentence that separates Cityverse from a slot machine. Variable reward never gates a civic action behind chance. The loop is present. It is not permitted to touch the thing the person came to do.

The third is about money, and it is where most reward systems fail. Buying a ticket and attending the play is civic participation with a receipt attached, and it counts in full, because presence authenticates it. A mailed check to a charity the donor has never visited is the only case of buying standing rather than earning it, and it is capped, at a small number of actions a month and at a small share of anyone's total. Sponsors may fund a medal and name it and promote it. They may never hand one out, and sponsored medals do not count toward the level gate, so that paid standing can never inflate earned standing. The airlines put paid and earned status on one ladder and spent a decade being accused of selling the ladder. Cityverse does not make that available.

Conduct touches standing in one way only. A member's civility is measured, from conduct only, never from popularity or agreement or viewpoint, and it is kept entirely backstage. There is no public conduct score. There is no visible tier. A member in poor standing for conduct sees their advancement pause, with a private explanation of what clears it, and nothing more. Conduct is never subtracted from the public number and never published as a negative. It is withheld momentum, felt by the person and invisible to everyone else. The design gets the consequence without the shaming, and it gets it without building the thing the incumbents' critics most fear, a reputation score that follows a person around.

The test from chapter five, applied. A medal is aimed at doing: it brings a person to the meeting, the game, the shop, the shift, the room she runs. The value it produces stays with the person as standing they keep, and with the city as the meeting attended and the shift worked. Nothing about it requires the person to be online longer. A like is aimed at watching: it brings a person back to check whether the count moved, the count moves regardless of anything real, and the value it produces is the session. It is the same social reward in both cases, and the same loop. The medal points it at the act and the like points it at the checking, and that is the whole difference.

9. The record

The medal is one thing a person accumulates on Cityverse, and chapter eight was about it. This chapter is about everything else that accumulates, and about a question that turns out to have an unusually literal answer: who does it belong to?

Every habit loop ends with something accumulating that makes the next cycle more likely. The incumbents call this investment and they are candid, in their internal documents, about what it is for: it is the switching cost. A person who has spent years building a profile, a following, a photo library, a graph of connections, does not leave, because leaving means abandoning the record. This is the least glamorous mechanic in the loop and, over a long enough horizon, the most powerful one.

It is also the one where the difference between aiming at watching and aiming at doing is most visible.

On a platform aimed at hours, the record is the platform's. The photographs are hosted there, the connections are stored there, the history exists in a form the person cannot take with them and often cannot even see in full. The switching cost is real, and it is a hostage. The person stays because the record is held.

Cityverse builds a record deliberately, and builds it to be the person's.

Begin with what accumulates. A resident's Homespace carries the medals they have earned, which are permanent. It carries the standing those medals compose into, which does not decay. It carries a garden. A Bloom is a piece of appreciation one resident sends another, and the garden is the Blooms a person has chosen to display. Engineering the Commons made the garden the one place on a page where another person's words appear, and it put a gate on it: a Bloom is published only when the recipient agrees, so what appears there is warm and public and consented to. That gate belongs to the garden alone. The rest of the record needs no one's permission, because the rest of the record is a record of things the person did. It carries a person's participation in the city's conversations, which persist as an account of how the community discussed a question and where it stood when the record was made. It carries the rooms a person built and ran: the Fan Bar with its quizzes and its trades, the club with its calendar. It carries, over years, the shape of a person's presence in their own city: the meetings, the games, the shifts, the shops, the arguments they took part in and the positions they held.

Some of what accumulates never happened on a screen at all. Chapter ten describes how a ticket bought on a theater's own website, or a minute recorded at a kiosk on the way out of a museum, lands on this page as a thing the person did. The record reaches past the application, because the doing does.

Now look at who this belongs to. None of it can be taken away by the platform. Nothing earned is ever revoked, and the design commits to that as a rule rather than a setting. None of it is a hostage, because all of it is a record of things that happened: the meeting was attended, the shift was worked, the neighbor was thanked, the quiz was built. The platform holds the ledger. The person holds the life it describes. The switching cost is real here too, and it is the honest version, the cost of walking away from a record of your own participation in the place where you live, which is the same cost a person pays when they move away from a city they have been part of for twenty years. They can leave. They have built something, and that is what leaving costs.

Engineering the Commons described the same mechanism at the scale of an institution. The school's coaching software holds the school's own history, a season of records, rosters and results, and "the school's own history accumulates inside it, so a season of records, rosters and results is lost or has to be migrated if they leave." That is a bond, and it holds because what accumulates is the school's, not the vendor's. The record on a resident's Homespace is the same bond at the scale of a person.

One more property makes this a chapter about behavior rather than about data. A person who can see their own record sees the shape of their own participation, and the shape changes what they do next. The parent who has been to four games has a page that says so. The fifth game is a continuation of something visible, and people continue what they can see themselves having started. This is the investment mechanic doing exactly what it does everywhere, making the next cycle more likely by making the previous cycles present. The incumbents use it to make leaving hard. Cityverse uses it to make continuing natural, and the difference is entirely in what the record is a record of.

10. What doesn't wear out

Every mechanism in this paper so far has been a mechanism of reward, and reward habituates; chapter two said why. A design built entirely on external reward is a design that must escalate forever to stand still, and escalation is the road to the slot machine. So the question a behavioral design has to answer, if it intends to last, is whether there is anything a person wants that does not wear out.

There is, and the research on it is among the most replicated in psychology. Edward Deci and Richard Ryan spent forty years establishing that human motivation divides into two kinds that behave differently. Extrinsic motivation, the pursuit of a reward external to the activity, is powerful and it habituates and, in a well-documented set of cases, it corrodes: pay a child to draw and the child draws less once the pay stops, because the payment taught them the drawing was for something else. Intrinsic motivation, the pursuit of an activity for what it is, does not habituate in the same way, and it rests on three needs that are stable across cultures and across lifetimes. Autonomy, the sense that one's actions are one's own. Competence, the sense of being effective at something that matters. Relatedness, the sense of being connected to people who count. Satisfy those and a person continues without escalation. Frustrate them and no schedule of reward will hold them for long.

A city is one of the few environments that can supply all three to an ordinary person on an ordinary day, and Cityverse is arranged to let it, in the city's rooms and in the platform's own.

Autonomy comes first, because it is the one attention platforms most reliably violate. A feed aimed at hours is experienced, correctly, as something done to the person. They did not choose what they saw, they cannot see why they saw it, and the controls they are offered change little. Cityverse gives the controls back in specific places. Every item can say why it is there. A person who wants less of something says so privately and their own Flow changes, with no consequence to anyone else. A conversation is joined by invitation and never by force; a person whose thread is drawn toward a Super Thread is asked, and may decline. Appreciation sent to a person is published only when that person decides. Reach is disclosed before posting so that the author, and never the machine, makes the final call about what goes out. None of these is a large feature. Together they produce the sense, rare on a platform, that the person is operating the machine and the machine is not operating them.

Competence is what the medal is for, understood correctly. The like rewards being approved of. The medal rewards having done something: attended, contributed, built, known, bridged, and each of those is a competence. A medal for demonstrated knowledge of the city, or for sustained work in one domain attested by others, is a competence signal in the strict sense, a record that this person is effective at something that matters here. The fan whose quiz forty people took, and eleven passed, has the same kind of signal. The research predicts that a reward framed as information about one's competence sustains motivation where a reward framed as a prize corrodes it, and the design's choice to lead with named standing rather than a score follows directly. A Keeper is told what they are. A number is told what they have.

Relatedness is the one a national platform cannot supply at all, and the reason is structural rather than a matter of effort. Relatedness is connection to people who count, and on a national platform almost nobody in the feed counts, in the specific sense that nothing they do can touch the person's life and nothing the person does can touch theirs. The argument about a policy in another state is with people who share no consequences. Here, every participant is a verified neighbor. The person on the other side of the disagreement about the shelter will live near the shelter. The parent in the stands is a parent whose child goes to the same school. The fan trading cards in a Fan Bar will be in the same stadium on Friday. Engineering the Commons made this the first of Ostrom's design principles, bounded membership with shared stakes, and treated it as the condition for a durable commons. It is also, read through this lens, the condition for relatedness, and relatedness is the intrinsic need that most reliably brings a person back to a place without anything having to be dangled in front of them.

There is one place on the platform where all three arrive at once, and it is the clearest case of what this paper means by doing on the screen. Cityverse calls it an Affinity Bar. It is a strip of tools a person attaches to their Homespace around a thing they already care about, and a person can stack several. There is a Fan Bar for a team, a Faith Bar for a congregation, a Gig Bar for a side business, and a Club Bar for the birders, the cyclists, and the chess players. Each one signals the affinity in public and carries the utility that goes with it, and every one of them is included in Cityzen membership, along with the ability to make a new one.

The Fan Bar is the one to watch, because a team is the largest affinity most cities have. It works in three moves. The first is to absorb what already exists. A Rays fan already buys tickets, already argues about the lineup, already posts from the stands.

The Fan Bar catches those acts as they happen and gives them a home: the season tickets, the all-time favorite player, the games attended, all on one strip of her page. The second move is to create what did not exist. Fanhood today lives in a Facebook group or a message board, flat, with its utility limited mostly to talking. The Fan Bar makes it public and useful. The fan uploads a photo and becomes an avatar in a Rays jersey. She builds a trivia quiz other fans take and sees who beat it. She keeps a list of what she has to trade and finds the fan across town who has the card she wants. None of that had a place before, and the team had no way to see it. The third move follows from the first two. A room full of visible fans, doing visible things, draws in fans who were not in the room, and deepens the ones who were.

The clubhouse is where the three moves meet, and it deserves an exact description, because it is also where a careful reader will look for a velvet rope. The team runs a live event on the platform, a warm-up before a series, say, with players present. Everyone can attend. Everyone hears every question asked and every answer given, and none of it costs anything, because content on Cityverse is free and reading it is free. The people with a Fan Bar are on the stage. Their avatars are up front, in their jerseys, and they are the ones who put the questions to the players, and who see one another there, and trade, and take one another's quizzes while they wait. The people without one are in the back, and the back is a good seat: the whole event, in full. What the back cannot do is go up and interact with a player. That is a service feature, and it belongs to membership plus one act, the photo that made the avatar. Nothing is bought. The step from the back to the stage is one act away, and a fan in the back who watches the stage for twenty minutes usually knows what she wants to do next. The back is where the next Fan Bars come from, and the room grows the way a room should, by people deciding to join it.

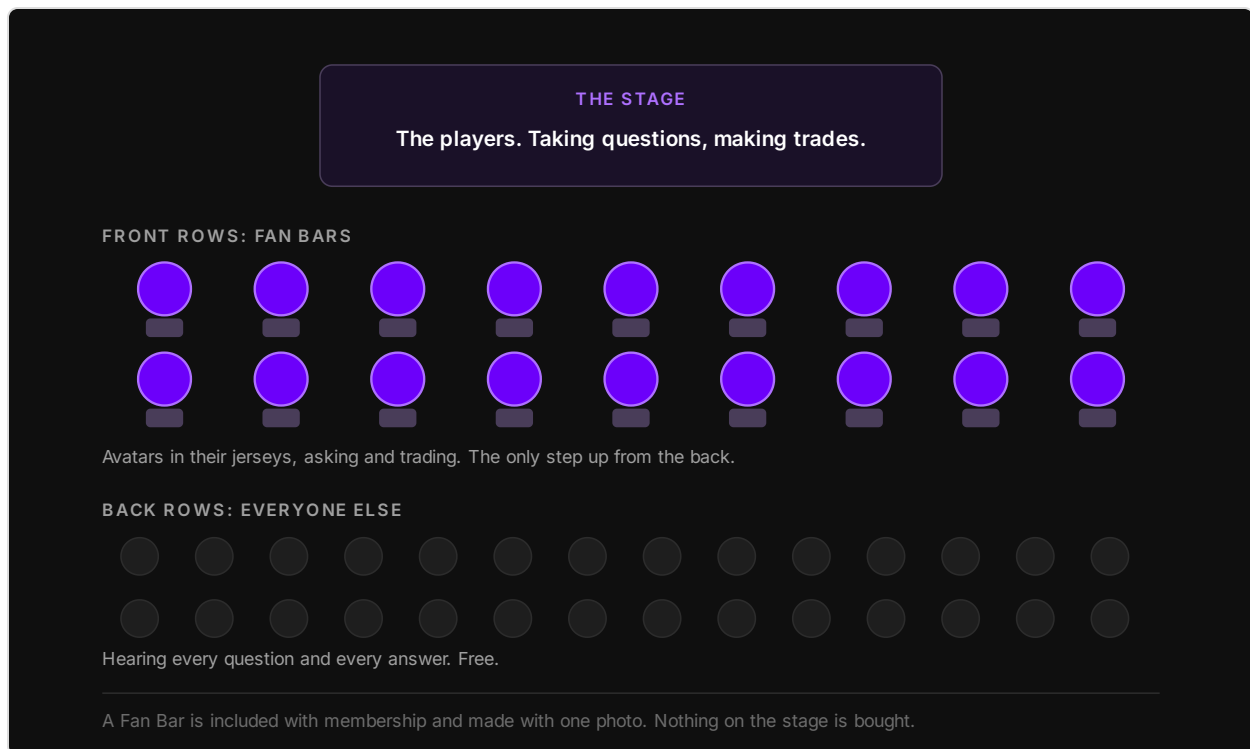


Figure 3. The clubhouse. Players on the stage. Fan Bar avatars up front, in their jerseys, asking and trading. Everyone else in the back, hearing every question and every answer, free. The only step from the back to the stage is a Fan Bar: included with membership, made with one photo.

Read the room against the three needs. Autonomy: nobody is pushed onto the stage, and the fan chose the team, the quiz, and the trades. Competence: the quiz that beat forty people, the standing earned in the room, the player who read her question aloud. Relatedness: the other avatars are people who will be in the same stadium on Friday. It is the swim meet with the platform as the building.

The doing does not stop at the edge of the application, and this is the part of the design no incumbent has attempted. A national platform ends at the screen. Cityverse is built to reach into the city's actual rooms. A community calendar, networked across the city and shaped to each organization, is offered as a plug-in that any local website can carry, so that the theater's site and the church's site and the running club's site are each more useful and all connected. The plug-in carries a purchase pixel, so that when a member buys a ticket to the orchestra on the orchestra's own website while logged in, the act lands on her Homepage as a thing she did, and counts. And there are the ports. Cityverse calls them Idea Portals: a podcast studio, a kiosk, a recording booth, any physical place where the city's life can enter the platform. At the exit of the Dalí Museum a visitor is asked one question, how seeing this work changed the way they see the world, and can record

a minute of video that goes straight to the museum's Createspace, its page on the platform, where the next visitor and the whole city can find it. The same port fits a nonprofit's lobby, a school's front hall, and in time a street corner. Walk into a museum, a church, a club, and there is a way into Cityverse from inside the room. That is the physical form of everything Part Two has described: attention converted into doing, on the platform and off, with the record of it kept in one place. It is also the reason the incumbents cannot follow. Their machinery ends where the screen does. This one is bolted to the city.

Put the three needs together and a specific thing appears, which is civic efficacy: the experience of having acted, in a place where the action mattered, among people who noticed. It is competence and relatedness in one act, exercised autonomously. It arrives in the stands and it arrives on the stage of the clubhouse, and the design treats those as the same thing in two places. The incumbents cannot produce it, because their platforms are not attached to any place where a person's action could land. Cityverse can, because a city is exactly such a place, and the design exists to reduce the distance between a resident and the point where their action lands, whether that point is a gymnasium on a Tuesday night or a quiz forty neighbors took. A person who has felt that once wants it again, and the wanting does not habituate, because what is wanted is a condition of being somewhere, and the design's real work is to keep supplying it.

11. Who will speak

Chapter four described a room the professor looked at and declined to enter. This chapter is about what has to be true of the room before she comes back, and the answer is that the room is built differently, in ways she can verify before she says anything.

Start with the arithmetic of who speaks at all. On the platforms most people use, the share of members who ever post is small, and the share who post regularly is smaller, and that fraction is the subset that chapter four's two filters, the door and the ranking, selected. A survey in the first years of the current platforms found that most people would discuss a contested public question with someone in person, and fewer than half would post the same view online. The gap between those two numbers is the professor, multiplied by the city. The quiet majority is quiet in a specific kind of room, and it will speak in a different one.

The room is different in its subject before it is different in its mechanics. The national platforms organize around national questions, and a national question is, for almost everyone reading about it, theoretical: nothing the reader does will move it, so the only thing left to do with it is take a side, and the platforms sort people into sides accordingly. That is where the tribes come from. A city's questions cut differently. What to do with the park, whether the noise ordinance changes, where the shelter goes: these are questions a resident can affect, and the sides they produce do not line up with anyone's party. The Super Thread organizes the city around those questions on purpose. The point of gathering everyone into one conversation is to hold a disagreement that can actually be resolved, among people who will all live with the result.

Cityverse's first move is to lower the cost of the first word to almost nothing, without lowering its meaning. When a question the city is actually arguing about reaches a certain density, the scattered threads about it are invited into a single conversation, the Super Thread described in chapter four, one room for one topic. Nobody's thread is taken; each author is asked whether to join and may keep their own. Inside the room, the state of the discussion is summarized at the top, continuously, and beneath the summary are the two to four positions residents have actually taken, surfaced from the conversation itself. A person who has never typed a public word about anything can read those positions and touch the one closest to their own. That is a contribution. It is counted, it moves the visible shape of where the city stands, and it required nothing of the person that the pile-on could punish. The professor can be present with one tap, and the design is built around the observation that most people who will never type will tap.

The second move governs what she sees when she does. The count of who holds each position is hidden until she has chosen, or until she has read the conversation to its end. A visible tally does real damage. It anchors the vote, and a person who sees that seventy percent hold one view is measurably more likely to hold it too, or to say nothing, which is the spiral of chapter four running inside a single screen. Hiding the count until the person has committed removes the anchor. Unlocking it for anyone who reads to the bottom rewards the person who took the conversation seriously, and it is the one place in the design where the reward for informed participation is built into the mechanics rather than the medal.

The third move is the one that makes the room safe for a name. Each position shows how many hold it. It does not show who, unless the person chooses to be shown. The named roster is private by default. A permanent public list of which residents

held which view on a contested local question would rebuild, at the level of a single screen, the exact loudness the tap was designed to remove, and it would make the quiet voter permanently loud in the one way she least wants to be. So the room lets her be counted without being catalogued.

The fourth move is about the negative, and it is the design's most careful piece of work, because there is no un-weaponizable button. Any signal that touches another person's content can be turned into a pile-on by enough people pressing it together. Cityverse builds an architecture in which abuse does not do anything.

The negative control opens into two branches, and they are two different jobs. The first is "see less." It is private and it is the person's alone: it changes what that person is shown and it is invisible to the author and to everyone else. A brigade of ten thousand people pressing it accomplishes exactly nothing to the target, because the signal never leaves the presser's own Flow. This is the only negative that is safe by construction, and its aggregates are never shown to authors, for the reason chapter four gave: it is the perception of disapproval that silences.

The second branch is "needs more perspective," and this one does reach the author's content, so it is defended three ways. It costs something: to register it, the person is shown the other positions already taken on the item and asked to point to the one that is missing. The objection cannot be registered without engaging an actual alternative view. An attacker's effort becomes a contribution. It is gated: its weight comes from being used by people who otherwise disagree, so a faction cannot drive it by agreeing with itself. And it is routed toward expansion rather than burial: what it does to the content is attach the missing perspective, and on news, surface the other coverage, rather than push the item down. The consequential negative, in this design, is a contribution or it is nothing. There is no way to make a stranger's disapproval land on the professor as a number, and there is no way to make it land as a burial. There is only a way to make it land as another view, which is the one thing she was hoping someone would offer.

Above all of this sits the ranking, and it completes the answer. Content that earns approval across groups that ordinarily disagree is lifted, and nothing is pushed down for being divisive. The sharp minority view that turns out to be right cannot be suppressed by the mechanism, because the mechanism only boosts and never penalizes. The professor's post about the water table, which nobody's side wanted to hear and which several sides quietly found persuasive, is exactly the content this ranking is built to find. Cityverse is doing two things to the groups at once. It

measures the bridges between them, and, through the Super Thread, it dissolves the silos that made them groups in the first place. Early on that means the bridging signal has little to see, which chapter twelve counts as a limit. It also means the design is shrinking the problem while it measures it.

The room does one more thing the incumbents' threads never do. It keeps its own account. When the city has landed, the conversation records where it landed, and the argument the six combatants would have run for months has a place where it can be left. The need to finish a story is still there, in everyone who enters. Here it is pointed at a question that can actually be settled, by the people in the room.

So the room she walks back into has these properties, and she can check each before she speaks. She can be present without writing. She can see where the city stands only after she has decided where she stands. She can be counted without being named. Nothing anyone presses can reach her as a number or bury what she said. And the only way a stranger can object to her is to offer another view. It is a room in which the machinery that drove her out has been taken apart and the pieces put to different use. It is the same machinery. It is aimed at the professor speaking, and the professor keeps what she said.

12. What this does not solve

The floor

Suppose the ambitious version does not work. Suppose the medals are earned by a small minority and ignored by the rest, the echo is appreciated and changes nobody's habits, the clubhouse fills slowly and the professor stays home anyway. What is left?

A local news platform, modernized. Every post is screened for conduct before it exists. The Flow is ranked by whether members found an item worth their time, with a floor under civic content and paid placement kept out of the ranker. The interface competes with the national platforms on experience. And the operation keeps more of the value it creates than a newspaper ever could, and can earn from it in ways a newspaper never had. That is a better place to read about a city than anything a resident has available today, whether or not a habit ever forms around it. Engineering the Commons described its own floor as a local news operation with a

modern audience, and called it a real business. The behavioral floor is the same operation with a civil comment section and a feed that does not reward the fight. It is smaller than what this paper describes. It is not a failure.

The limits

Here is what Cityverse does not do.

The room will be judged as the room people already know. This is the hardest limit in the paper and the one most likely to decide the outcome. Public perception of a platform is set before anyone reads a word about it. A person who sees a feed and a profile, a Flow and a Homespace, classifies what they are looking at in under a second: local social media. They are partly right, and they will then do what they already do. The people who make noise on the national platforms will arrive first, because arriving first is what they do, and they will be motivated to make the same noise here. The quiet people, seeing them, will settle back into their quiet.

Everything in chapters four and eleven is built to reverse that, and none of it is visible on the first day, because a screen at the door and a ranking that will not pay for a fight are things a person notices over weeks, and a pile-on is a thing a person notices in a minute. Worse, the ordinary growth playbook points the wrong way. The standard advice for a new platform is to win the social mavens, the people with the most followers and the loudest presence, and let them bring the room to life. That is precisely the life this design does not want, and a platform that takes the advice will have imported the second pile before the first pile has arrived.

Dissuading the noisy, without becoming the kind of place that shames anyone, while persuading the quiet that this room is different from every room they have declined to enter, is a problem of concerted communication and of patience, and the design does not solve it. It can make the room actually different, say so plainly, and hold the line long enough for the difference to be noticed. The paper claims nothing more than that, and the reader should weigh this limit heavily.

It does not escape the baseline. Chapter three argued that no one can opt out of the engineered attention market, and that argument cuts against this design as well as for it. A product that shows fewer things, tells a person when they have seen everything, keeps its surprises away from the things that matter and refuses to manufacture a cue, will be experienced by some residents as slower than what they are used to, and some of them will leave for that reason and not come back.

Cityverse accepts this. It is the same trade Engineering the Commons made with

verification: volume given up for properties the rest of the design depends on. The answer is the one the earlier paper gave. Do not respond by turning up the ratio. Respond by raising what the person gets until the trade is obvious.

The quarantine will be tested. The rule that variable reward never gates a civic action is the load-bearing commitment of chapter eight, and it will come under pressure at the moment growth is slowest, from someone with a dashboard showing that a small randomized bonus on check-ins would lift the number. The lift would be real. Engineering the Commons described the same life cycle for the instruments at the door: cutting them improves a metric immediately and the damage arrives quarters later in a different line. This is the behavioral version. The defense is the same, a standing measurement of the thing the quarantine protects, which is whether members say the time was worth it, so that the argument for breaching it has to be made against a number rather than against a principle.

The classifier can encode the wrong norms. A system that screens every post for conduct is a system whose definition of conduct is doing enormous work, and the documented risk is that automated civility systems flag the speech of some communities more often than others for reasons that have nothing to do with conduct. Cityverse's answer is to build the classifier rather than license it, to target the structure of an attack rather than the register of speech, to route uncertain cases to people primed for dialect and context, and to measure disparity across groups continuously and publish it. Those are the right answers and they are not a guarantee. A civility system that turns out to be a class filter has failed the thesis of chapter four, because it has driven out a different set of careful people. The audit is not optional.

Bridging needs density. Lifting content that earns approval across disagreeing groups requires enough participation to see the groups. Early, and in every new city, the mechanism cannot see them and should do nothing rather than guess. The earlier paper made this point about the bridging signal generally. It applies with particular force here because the conversation rooms and the perspective mechanics of chapter eleven depend on it, and they will be thin before they are useful. There is a complication Cityverse should own. Cityverse is also working against the groups it is trying to measure, since the Super Thread exists to dissolve the silos that produce them. The design is shrinking the problem while it measures it, which is the right thing to do and makes a clean early reading impossible.

The medal may be decoration. The claim of chapter eight is that a reward aimed at the act brings a person back to the act. That is a hypothesis. It is possible that the people who earn medals are the people who would have attended anyway, and that the medal records participation without producing any. If so, the medal is a ledger, and the design should know which.

Standing will be used as a proxy. Cityverse keeps conduct private and lets no amount of money buy the top. It cannot control what the city does with a public standing once one exists. Employers will look at it. Candidates will campaign on it. Organizations will invite the Elders first. A retiree with free evenings will climb faster than a parent working two jobs, and the breadth gate softens that without erasing it, because time is a form of privilege as surely as money is. A public standing becomes a reputation through social use whether or not the platform intends one. The design's answer is to keep the number off the front of the page, to lead with a name that signals belonging rather than rank, and to watch the correlation between standing and wealth as closely as it watches the correlation between money and reach. That is a mitigation, and the paper claims nothing more for it.

The tap is a hypothesis. Chapter eleven rests on the claim that people who will never type will tap, and that the resulting count says something true about where the city stands. Both halves are untested. A low-friction signal can broaden the sample and thin the deliberation at the same time, and a room full of taps is a picture of who tapped, which is closer to the city than a comment thread and still short of a poll. Cityverse shows the count only after a person has chosen and labels it for what it is. Until the fourth prediction below has been tested, the count is a reading of the room and should be presented as one.

This paper describes one city. Everything in Part Two is written against St. Petersburg, where a newsroom already holds the city's attention and the design attaches to existing gravity. A city starting without that has a harder behavioral problem, because several of these mechanisms, the real cue, the echo, the conversation room, presuppose that things are already happening on the platform for a person to return to. The earlier paper's sequence, buy the one link and give away the rest, is the answer. It is not a fast one.

The objection, met a second time

Part One answered the manipulation charge in principle: there is no neutral platform, so the question is how the machinery is aimed and who keeps the value. Part Two has now described the machinery, and the charge can be put more sharply. A reader who has followed the mechanics can say: you have built a habit loop and named its parts, and a habit loop is a habit loop whatever it is pointed at.

The answer is yes, and here is what was pointed where. The cue answers to a real event or it is not sent. The surprise is confined to places where it is a gift and kept away from the act the person came to do. The reward is for having done something, in the stands or in a room the person built, and it points at the two hardest things, bridging and knowing. The stage in the clubhouse is reached by an act, and everything said on it is free to everyone in the room. The record is of the person's own participation and cannot be revoked. The controls that matter are in the person's hands and the disapproval of strangers cannot reach them as a number. The Flow cannot see the money. Every one of those is a design choice the incumbents could have made and did not, and every one of them costs something on the dashboard the incumbents optimize. That is the pattern, and it has a cause. A habit loop aimed at what a person does looks like this, and it looks like this because the value is being left with the person. The reader is invited to name the mechanism in Part Two that fails the test of chapter five. Cityverse's claim is that there is not one.

What would prove this wrong

A behavioral design should say what evidence would count against it. This one makes four predictions that can be checked, and they are stated now so that the test cannot be redrawn to fit whatever happens.

The first concerns what people do. If the design is aimed at doing, its members should do more in their city than comparable residents who are not members: attend more, volunteer more, show up more. They should also build more on the platform than a passive audience would, the quizzes and rooms and calendars that did not exist before. And within the membership, the people who earn medals for attending should return to real events at a higher rate than comparable members who do not. If the rates are the same, the medal is a ledger, and the loop is aimed at something other than what this paper claims.

The second concerns satisfaction over time. Cityverse's other headline metric is whether members, asked afterward, say their time was worth it. On a platform aimed at watching that number should fall with tenure, because the ratio degrades the experience even as it lengthens the session. Here it should rise, or at minimum hold, as a person's record and standing accumulate. If members who have used the design longest rate it lowest, the loop is aimed at hours whatever the paper says.

The third concerns the shape of a session. A platform aimed at watching produces a characteristic distribution of session lengths, with a long tail of sessions that run far past the point of any stated intention, the signature of the scroll. This design should not show that tail. If it does, the echo is not doing what chapter seven claims, and the ratio has found its way back in.

The fourth concerns who participates. Chapter four claims the door and the ranking drive out a specific set of people, the careful, the expert, the conflict-averse, and chapter eleven claims the room is rebuilt to bring them back. That is measurable. The share of members who contribute, and the retention over time of the groups that harassment most reliably drives off other platforms, should be higher here than on open platforms in the same city. If participation here is the same small, self-selected sliver, the room has not changed and the professor is still home.

None of these can be tested before the design has run long enough to matter. They are stated so that when it has, the standard is already on the record.

The argument in ten sentences

Each sentence is defended in the chapter it names. A reader who remembers only these has the design.

1. The same machinery builds a habit and an addiction. Aimed at what a person watches it extracts; aimed at what a person does, in the city or on the platform, it serves. (*Chapter five*)
1. A reward is a jackpot only against the misses around it. Make every post a hit and the hit becomes the floor. (*Chapter two*)
1. No platform competing for the evening hour holds it without engineering it. The only choice left is whether to say so. (*Chapter three*)

1. A cue is only as good as the story it opens, and a platform's stories are made of what it can produce. The incumbents produce no content of their own, so the fight is the one serial they can make. Cityverse produces the news, the tools, and the rooms, so its cues open things that actually happened. (*Chapters one, six and ten*)
1. The right to speak and the fact of speech are different things. Free food feeds the fastest; open access feeds the loudest. (*Chapter four*)
1. Each kind of content arrives in the form a person already knows, and what they have seen travels with them. Nothing is missed and nothing repeats. (*Chapter seven*)
1. A like counts agreement and inflates until it counts for nothing. The reward stays; it moves to the harder things: bridging a disagreement, and knowing what you are talking about. (*Chapter eight*)
1. The record a person builds and leaves in place is what makes returning worth it and leaving cost something. (*Chapter nine*)
1. Autonomy, competence, and belonging are the rewards that do not wear out, and a city can supply all three, in its own rooms and on the platform. (*Chapter ten*)
1. The quiet majority will tap where they will not type. Build the tap, hold the tally until they have chosen, and the room fills with people who never spoke before. (*Chapter eleven*)

Sources

Figures and findings in this paper are drawn from the following. Where a source is vendor or company data rather than peer-reviewed research, or where a finding is contested, that is noted. Company documents describing the design itself are not cited; the design is described as it is.

THE REWARD MECHANISM

Schultz, Dayan and Montague, "A Neural Substrate of Prediction and Reward," *Science* 275(5306), 1997. Dopamine neurons encode reward prediction error, firing to unexpected reward and shifting to the predictive cue. The basis for chapter two's account of surprise as a relative quantity.

Anselme, "Dopamine, Motivation, and the Evolutionary Significance of Gambling-like Behaviour," *Behavioural Brain Research* 256, 2013, and Anselme and Güntürkün, "How Foraging Works: Uncertainty Magnifies Food-Seeking Motivation," *Behavioral and Brain Sciences* 42, 2019. Uncertainty of reward increases seeking in animals foraging under scarcity; the basis for chapter two's account of the hunt as the origin of the mechanism.

Ferster and Skinner, *Schedules of Reinforcement*, 1957. Variable-ratio schedules produce the most persistent and extinction-resistant behavior.

Brickman and Campbell, "Hedonic Relativism and Planning the Good Society," in *Adaptation-Level Theory*, 1971. The adaptation of subjective wellbeing to changed circumstances; the origin of the hedonic treadmill.

Brady, McLoughlin, Doan and Crockett, "How Social Learning Amplifies Moral Outrage Expression in Online Social Networks," *Science Advances* 7(33), 2021. Outrage is reinforcement-learned through social reward; the reward signal, more than the belief, escalates it.

WHAT ENGAGEMENT RANKING SELECTS FOR

Milli, Carroll, Wang, Pandey, Zhao and Dragan, "Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media," *PNAS Nexus* 4(3), 2025.

Engagement ranking amplified out-group animosity and anger; users rated the amplified content lower on stated preference than what they would have chosen.

Rathje, Van Bavel and van der Linden, "Out-group Animosity Drives Engagement on Social Media," *PNAS* 118(26), 2021. Out-group language the strongest single measured amplifier of resharing.

Anderson, Brossard, Scheufele, Xenos and Ladwig, "The 'Nasty Effect,'" *Journal of Computer-Mediated Communication* 19(3), 2014. Uncivil comments polarize readers' perception of the underlying content, independent of its merits.

Facebook's "Meaningful Social Interactions" ranking change, 2018, as documented in internal materials reported by journalists in 2021. Not peer-reviewed. The weighting of reactions and comments above passive consumption, and the subsequent amplification of divisive content, is chapter one's example of a platform optimizing interaction and rewarding outrage.

THE UNFINISHED STORY

Zeigarnik, "Das Behalten erledigter und unerledigter Handlungen," *Psychologische Forschung* 9, 1927, and Ovsiankina, "Die Wiederaufnahme unterbrochener Handlungen," *Psychologische Forschung* 11, 1928. Interrupted tasks are better remembered and are spontaneously resumed. The classic Zeigarnik memory effect has had mixed replication; the resumption tendency is more robust. Cited for the mechanism only.

Green and Brock, "The Role of Transportation in the Persuasiveness of Public Narratives," *Journal of Personality and Social Psychology* 79(5), 2000. Absorption into a narrative and its hold on attention and belief. The basis for chapter one's account of the serial.

WHO SPEAKS AND WHO STAYS HOME

Noelle-Neumann, "The Spiral of Silence: A Theory of Public Opinion," *Journal of Communication* 24(2), 1974.

Hampton, Rainie, Lu, Dwyer, Shin and Purcell, "Social Media and the 'Spiral of Silence,'" Pew Research Center, 2014. Eighty-six percent of respondents willing to discuss a contested issue in person; forty-two percent willing to post it. Correlational.

Coe, Kenski and Rains, "Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments," *Journal of Communication* 64(4), 2014. Frequent commenters are more civil than infrequent ones; the design's reason for never penalizing volume.

St. Pete Catalyst comment moderation, 2018 to the present. First-hand observation by the newsroom's editors over that period. The basis for chapter six's account of the room itself as a cue.

INTRINSIC MOTIVATION

Deci and Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior*, 1985, and Ryan and Deci, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being," *American Psychologist* 55(1), 2000. Autonomy, competence and relatedness as basic psychological needs; the undermining of intrinsic motivation by contingent extrinsic reward.

INTERVENTIONS BEFORE AND AT THE MOMENT OF POSTING

Hertie School Data Science Lab, field experiment comparing pre- and post-publication moderation in comment sections, 2025. Pre-moderation reduced toxicity by roughly a quarter with no loss of participation. Working paper; not yet peer-reviewed.

Nextdoor "Kindness Reminder," company transparency reporting, 2019 and 2024; OpenWeb and Jigsaw, real-time feedback case study, 2020; Twitter "Read Before You Retweet," company data, 2020. Vendor data throughout. Cited for direction and rough magnitude only: roughly a third of people shown a reconsider prompt edit or withhold, and a read-first prompt increased opens by about forty percent.

Jhaver, Bruckman and Gilbert, "Does Transparency in Moderation Really Matter?" *PACM HCI* 3(CSCW), 2019. Giving a reason for a removal reduces the odds of future removals, across roughly thirty-two million posts.

Katsaros, Tyler, Kim and Meares, "Procedural Justice and Self-Governance on Twitter," *Journal of Online Trust and Safety* 1(3), 2022. Enforcement experienced as procedurally fair reduces reoffending.

BRIDGING

Ovadya and Thorburn, "Bridging-Based Ranking," Belfer Center, 2023. Ranking by approval across disagreeing groups rather than by engagement.

Wojcik et al., "Birdwatch: Crowd Wisdom and Bridging Algorithms Can Inform Understanding and Reduce the Spread of Misinformation," 2022. Preprint. The cross-group agreement gate as implemented in Community Notes.

Renault, Restrepo-Amariles, Pröllochs et al., "Community Notes Reduce Engagement With and Diffusion of False Information Online," *PNAS*, 2025. Causal estimates of reduced resharing following bridged notes.

CLASSIFIER BIAS

Sap, Card, Gabriel, Choi and Smith, "The Risk of Racial Bias in Hate Speech Detection," ACL 2019. Toxicity classifiers flag African-American English up to about twice as often as equivalent standard text; dialect-priming of annotators reduces the disparity. The basis for chapter twelve's limit on the classifier.

HABIT DESIGN AS PRACTICED

Eyal, *Hooked: How to Build Habit-Forming Products*, 2014. The trigger, action, variable reward, investment loop, described from the builder's side. Cited as the incumbents' account of their own method.

COMPANION DOCUMENT

Hamilton, *Engineering the Commons: A Design for Rebinding American Cities*, Cityverse, 2026. Canonical for the core design this paper views through a behavioral lens: gravity, chains, the bought link, the civility floor, and the consent gate.

Joe Hamilton joe@cityverse.life