

The plates they threw away

What a forgotten stomach bacterium, a temperature that waited 24 years for its question and a laser nobody could build from the paper teach us about the results science loses

Stromal Research

Science keeps its results remarkably well. What it keeps badly is the reason a result means what it means. Four episodes from twentieth-century research — in microbiology, astronomy and laser physics — show the same pattern: the observation survived, and the context needed to read it did not. That gap is widest for the most common outcome at the bench, the experiment that was neither a success nor a failure. This essay follows the stories to the problem, explains why lab notebooks are structurally blind to it, and ends with a three-line habit any lab can adopt this week.

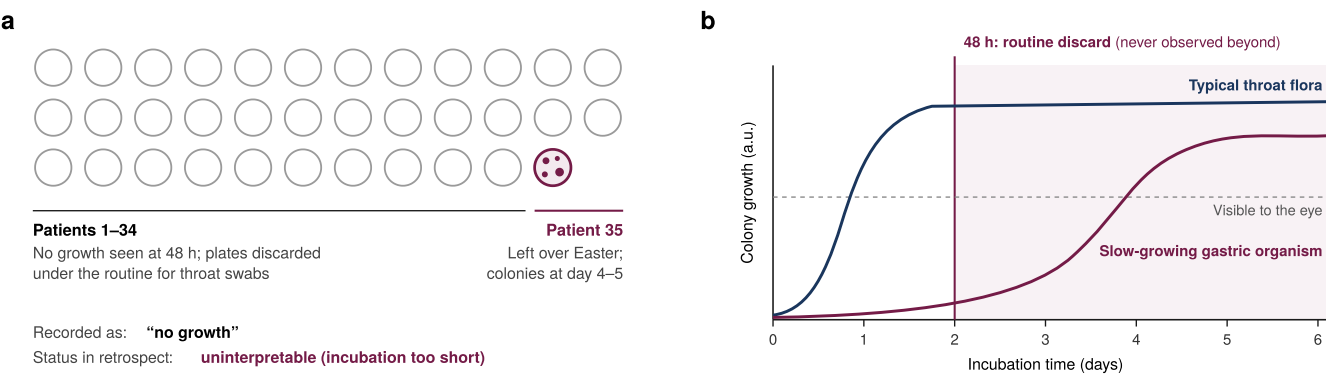


Fig. 1 | The 48-hour rule. **a**, Early attempts by Barry Marshall and Robin Warren to culture the organism now known as *Helicobacter pylori* repeatedly yielded no growth. A plate from patient 35, left in the incubator over the Easter holiday of 1982, produced colonies after 4–5 days. The laboratory had been discarding plates at 48 h^{1,2}. **b**, Schematic (not measured data) of why: a cut-off that suits fast-growing organisms ends observation before a slow-growing one becomes visible. The first 34 results were not negative. They were uninterpretable, and nothing on record said so.

In brief

- Labs rarely lose their data. They lose the reason a result means what it means.
- That reason matters most for *inconclusive* results, which are neither successes nor failures.
- Notebooks record experiments when they happen; the judgement about them usually comes later, somewhere else.
- A shared status label plus one sentence of reason, written at the moment of judgement, closes much of the gap.

Easter, 1982

In the spring of 1982, a young doctor in Perth named Barry Marshall had a problem that was not going away. He and the pathologist Robin Warren had been staring at something in stomach biopsies that, according to medical consensus, should not exist: small, curved bacteria living in

the stomach lining, often exactly where patients had gastritis and ulcers. Warren had been seeing them since 1979. Everyone knew that bacteria could not survive stomach acid, so to prove the organisms were real, Marshall needed to grow them in a dish. He could not. Biopsy after biopsy went to the hospital’s microbiology laboratory, and plate after plate came back with nothing. Thirty-four patients; thirty-four failures. Then came Easter. A biopsy taken on the Thursday before the long weekend went into the incubator like all the others. But the laboratory was swamped by an outbreak of drug-resistant *Staphylococcus*, and nobody got to Marshall’s plates for four or five days. When someone finally looked, there were colonies. Patient 35 was the first successful culture of what we now call *Helicobacter pylori*¹. What Marshall learned next is the part of the story that matters here. The laboratory had not been failing to grow his bacterium. It had been throwing it away. If nothing

interesting was on a plate after 48 hours, the plate was discarded¹. That was the routine for throat swabs, where 48 hours is plenty. Nobody had written down that the rule might not suit a slow-growing organism from the stomach, because nobody had thought to ask (Fig. 1).

So thirty-four “negative” results were not negative at all. They were inconclusive — and the reason they were inconclusive lived nowhere: not in a notebook, not in a protocol, only in a habit.

The negative that lasted thirty years

Marshall and Warren were not the first to see spiral bacteria in human stomachs. They were not even close.

In 1940, Stone Freedberg and Louis Barron at Harvard reported spiral organisms in roughly 40% of stomachs removed during surgery for ulcers or cancer². At the presentation, people in the audience mentioned ulcers clearing up in syphilis patients treated with bismuth and other heavy metals — a clue that points straight at an infection.

Then, in 1954, one of America’s leading gastro-enterologists, Eddie Palmer, published a study of more than a thousand stomach biopsies. He found no bacteria at all, and concluded that the organisms others had seen were contamination that appeared after death².

The field accepted it. Palmer had the larger sample, the bigger name and a clean answer. The question was considered closed for most of the next three decades, while millions of ulcer patients were treated for stress and excess acid.

No one is entirely sure why Palmer saw nothing. What is clear is that his result should have been carried forward as “we did not find them, using this method” and was instead remembered as “they are not there”. That one difference cost a generation of patients. Marshall and Warren received the Nobel Prize in 2005.

A number without a question

In 1941, a Canadian astronomer named Andrew McKellar was studying faint absorption lines from cyanogen molecules drifting between the stars. From the way the molecules were

excited, he worked out a temperature for their surroundings: about 2.3 degrees above absolute zero³. He published it as a curiosity of interstellar chemistry, of uncertain wider significance, and the number sat in the literature.

Twenty-four years later, two radio engineers at Bell Labs, Arno Penzias and Robert Wilson, were fighting a faint, persistent hiss in a giant horn antenna in New Jersey. It came from every direction, day and night. They checked their equipment. They evicted a pair of pigeons nesting in the horn and scrubbed out the droppings. The noise stayed, equivalent to about 3.5 degrees above absolute zero⁴.

What they had found, once physicists at nearby Princeton explained it, was the afterglow of the Big Bang: the cosmic microwave background. It won Penzias and Wilson the 1978 Nobel Prize in Physics⁵.

McKellar had measured the same background two decades earlier. What he did not have — and what nobody reading his paper had — was the question his number answered. A measurement without its question is just a number, and numbers are easy to walk past.

The laser nobody could build from the paper

In the early 1970s, the sociologist Harry Collins went looking for how scientific knowledge actually moves. He followed laboratories in Britain and North America that were trying to build a TEA laser, a carbon-dioxide laser whose design had been published in detail⁶.

On paper, anyone with the article and a workshop should have been able to build one. In practice, Collins found that the people who got a working laser had all spent time with someone who already had one. The people with something on the bench that looked like a laser but would not fire had not.

The written account left out the things that made the difference: which details mattered, what to ignore, what a real failure looked like compared with a fixable glitch. That knowledge passed from person to person, in rooms, and nowhere else.



Fig. 2 | Results that existed before their meaning did. In each episode the observation was recorded and survived. What was missing was the context needed to read it: the question it answered, the limits of the method, the judgement of the people who did the work. In two cases that context was recovered — once by a new theory, once by an accidental holiday^{1–6}.

What the stories share

We usually tell these stories as tales of genius and luck: the Easter weekend, the pigeons, the prize. Read them another

way and they are about something much more ordinary.

In every case, the result existed. The plates were incubated. Palmer’s biopsies were examined. McKellar’s temperature was published. The laser design was in print.

What was missing was the information that told you what the result *meant*.

Science is good at recording two kinds of outcome. When something works, it goes into the figure, the paper and the grant renewal. When something clearly fails, it at least gets mentioned in lab meeting. But much of what happens at the bench is a third kind: **the inconclusive result**. The signal is there, but the control drifted. The effect showed in two replicates and vanished in the third. The cells came from the wrong batch. Nobody will call it a success, and nobody will call it a failure.

An inconclusive result is the one kind of result that means nothing without its reason. Strip away “we don’t trust this because…” and the data quietly turn into a weak negative. That is how 34 discarded plates became “it doesn’t grow”, and how a thousand biopsies became “it isn’t there”.

The result existed. What it meant did not.

Why the reason disappears

The obvious objection is that this is what lab notebooks are for. The trouble is timing. A notebook entry is written when the experiment is done. The judgement that makes a result inconclusive usually comes later, and somewhere else.

Here is how it typically goes in a modern lab. On Tuesday, you run the assay and record what you did and what the plate reader said. On Thursday, at lab meeting, someone notices that the positive control is lower than usual. The PI asks whether the incubator was serviced that week. It was. The room decides, out loud, not to count the run: “Let’s repeat it once the control’s sorted” (Fig. 3).

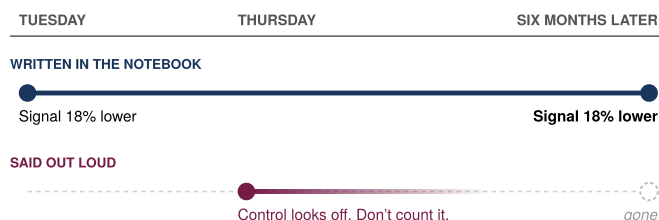


Fig. 3 | The number gets written down. The verdict doesn't. The result goes into the notebook on Tuesday. The decision that it doesn't count is made out loud on Thursday and never written down. Six months on, only the number is left, and it looks like the experiment simply didn't work. Illustrative example.

That decision — the most important fact about the experiment — was made in conversation. It lives in five people’s memories and perhaps a message thread. It almost never makes it back into Tuesday’s entry, because by Thursday, Tuesday feels finished. Electronic notebooks have not changed this, because it was never a paper-versus-software problem. Notebooks are organized around when work is done; interpretation happens at a different time, in a different place.

The clock is running

Once the reason is separated from the result, time does the rest. Three findings give a sense of how fast context decays, even for published work — the best-kept kind.

When Timothy Vines and colleagues tried to retrieve the data behind 516 studies published between 2 and 22 years earlier, the odds that a dataset still existed fell by about 17% for every year since publication; the odds of even reaching the authors by e-mail fell by about 7% a year⁷. When the Reproducibility Project: Cancer Biology set out to repeat 193 experiments from 53 high-profile papers, not one was described well enough to repeat without asking the original authors for help; in the end, 50 were completed⁸. And in a 2016 *Nature* survey of 1,576 researchers, more than half said they had failed to reproduce one of their *own* experiments⁹. That figure has many causes, but it shows that being the person who did the work does not guarantee you can recover it.

In 1979, the psychologist Robert Rosenthal named the field-wide version of this the file-drawer problem: positive results get published, null results go into the cabinet¹⁰. Inconclusive results sit at the very bottom of that drawer. They rarely leave the lab, and inside the lab they rarely outlast the people who ran them.

The cost comes in two forms. The first is the **silent repeat**: a new student, or the same student a year later, runs the experiment again because nobody can say what happened last time — same reagents, same instrument time, same weeks, nothing new learned. The second is worse: the **false negative**. The inconclusive result is remembered as a failure — “we tried that; it didn’t work” — and a promising line of work ends, not because anyone decided it should, but because the reason the first attempt did not count was forgotten. Palmer’s study did this to an entire field. Labs do it to themselves every week.

What changes when stopped work gets a name

The usual advice is to document more. It rarely survives the first deadline. A more interesting answer comes from Arcadia Science, a research organization that decided to publish the projects it stopped working on¹¹.

It did not call them failures. It called them “iced” and gave them a public home called the Icebox. By late 2023, 11 of Arcadia’s 34 publications were iced projects. In an internal survey, two-thirds of the scientists who responded had led or contributed to iced work, and in most cases the decision to stop had come from the scientists themselves. The stopped work drew thousands of readers, and two outside groups asked for materials from it.

The lesson is not that every lab should publish its abandoned projects. It is that **a name changed behaviour**. Once “stopped, and here’s why” was a legitimate outcome with its own label, people started writing down why things stopped. The reasoning that usually evaporates had somewhere to go. Inconclusive results need the same thing: a name, and a reason attached to it.

The three-line habit

Here is something any lab can start this week, with whatever notebook it already uses. Whenever a result is not clearly a success or a failure, write three lines and attach them to the

experiment itself — not to the meeting notes.

Status. One word from a short list that the whole lab shares: *supported* (holds as far as we can tell), *unreliable* (something compromised the run), *stopped* (we chose not to continue) or *blocked* (cannot proceed until something changes). **Because.** One sentence on why. **Would resolve it.** What would turn this into a real answer.

Now imagine the Perth laboratory had done this. Each discarded plate would have carried a note like the one in Fig. 4a — and 34 such notes would have pointed straight at the answer an accidental holiday eventually delivered.

a

CLOSEOUT · PERTH CULTURES, 1982 (IN HINDSIGHT)

SUPPORTED **UNRELIABLE** STOPPED BLOCKED

BECAUSE Plates discarded at 48 h; a slow-growing organism may not be visible by then.

WOULD RESOLVE IT Hold plates for ≥5 days before calling no growth.

b

CLOSEOUT · THE TUESDAY ASSAY (ILLUSTRATIVE)

SUPPORTED **UNRELIABLE** STOPPED BLOCKED

BECAUSE Positive control drifted after incubator service; run cannot support the inference.

WOULD RESOLVE IT Repeat with a fresh control; hold treatment conditions fixed.

Fig. 4 | Three lines, written when the judgement is made. **a**, The note the Perth cultures could have carried, written in hindsight. **b**, The same habit applied to the Tuesday assay. Each records what the result can be used for, why, and what would settle it.

Two details make the habit work. First, **write it when the judgement is made, not when the experiment is run.** Usually that means straight after lab meeting, by whoever presented the data. That is the moment a notebook misses, and it is the moment that matters. Second, **keep the status list short and the same for everyone.** Arcadia's "iced" worked because everyone knew what it meant. If "unreliable" means the same thing in every notebook, a newcomer can pull up every unreliable run in a project and know exactly which results not to trust, and why.

A year from now, someone in your lab will ask the question every lab eventually asks: *have we tried this before?* Without the three lines, the answer is a shrug, a dig through the shared drive and probably a repeat. With them, it is: *yes, once, but we marked it unreliable because the control drifted after the incubator service, and the plan was to repeat with a fresh control. That repeat never happened.* That is not a dead end. It is the next experiment, already designed.

Why this matters more now

AI systems can now read every notebook entry, file and figure a lab has ever produced. That makes the missing sentence more consequential, not less. Shown a plot with an 18% drop, a model will happily propose a biological explanation. It cannot know the lab later decided the run was compromised unless that decision was written down next to the plot. Models will keep improving; they still cannot

recover a judgement that was never recorded. The record has to carry the memory, so that people and machines can reason over it.

The oldest unsolved problem in the lab

None of this is new. In 1661, Robert Boyle — one of the founders of experimental science — published essays on what he called the "unsuccessfulness" of experiments: why trials fail, why they disagree, and why experimenters should not be discouraged by it¹². Three and a half centuries later, we have journals, databases, electronic notebooks and machines that can read millions of papers. We still do not have a reliable way to keep the one sentence that explains why a result did not count.

The fix does not start with technology. It starts with treating "we're not sure, and here's why" as a real scientific outcome, worth writing down at the moment someone decides it.

Somewhere in your lab is a stack of plates that someone threw away at 48 hours. Start by writing down why.

References

- Marshall, B. J. *Helicobacter* connections. Nobel lecture, 8 December 2005. <https://www.nobelprize.org/prizes/medicine/2005/marshall/lecture/>
- Marshall, B. One hundred years of discovery and rediscovery of *Helicobacter pylori* and its association with peptic ulcer disease. In *Helicobacter pylori: Physiology and Genetics* (eds Mobley, H. L. T., Mendz, G. L. & Hazell, S. L.) (ASM Press, 2001).
- McKellar, A. Molecular lines from the lowest states of diatomic molecules composed of atoms probably present in interstellar space. *Publ. Dominion Astrophys. Obs.* **7**, 251 (1941).
- Penzias, A. A. & Wilson, R. W. A measurement of excess antenna temperature at 4080 Mc/s. *Astrophys. J.* **142**, 419–421 (1965).
- Wilson, R. W. The cosmic microwave background radiation. Nobel lecture, 8 December 1978.
- Collins, H. M. The TEA set: tacit knowledge and scientific networks. *Sci. Stud.* **4**, 165–185 (1974).
- Vines, T. H. et al. The availability of research data declines rapidly with article age. *Curr. Biol.* **24**, 94–97 (2014).
- Errington, T. M. et al. Challenges for assessing replicability in preclinical cancer biology. *eLife* **10**, e67995 (2021).
- Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
- Rosenthal, R. The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 638–641 (1979).
- Arcadia Science. Icebox is a science-sharing strategy designed to encourage risk-taking. <https://doi.org/10.57844/arcadia-8c5c-e47c> (2023).
- Boyle, R. *Certain Physiological Essays* (London, 1661).

About this essay

The Working Record is a series from Stromal on how scientific work is recorded, interpreted, forgotten and carried forward. Stromal builds tools for keeping a lab's working record connected; the habit described here needs none of them and works on paper, in a spreadsheet or in any notebook.

A note on sources

Historical details are drawn from first-hand accounts and the works cited. Why Palmer's 1954 study found no bacteria remains uncertain. Figures 1b, 3 and 4b are illustrative, not measured data.

Correspondence

hello@mylabos.net

© Stromal 2026