

MEASURED, NOT GUESSED

CLAUDE . md

UNDER TEST

Does Claude Code actually obey
the rules you write down?

Trials with the ban written in
— violations

0 / 22

Trials without it
— violations

17 / 17

Length, position, contradiction and stakes changed nothing.
Then the hooks turned out to fail open.

yurukusa

39 trials · Claude Code 2.1.246 · full harness included

CLAUDE.md Under Test: What Claude Code Rules Actually Do

Thirty-nine measured trials on whether written rules are obeyed, when Claude Code hooks earn their place, and how both fail - with the full harness and every trial's data.

yurukusa

This book is available at <https://leanpub.com/claude-md-under-test>

This version was published on 2026-08-30



This is a [Leanpub](#) book. Leanpub empowers authors and publishers with the Lean Publishing process. [Lean Publishing](#) is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

© 2026 yurukusa

Contents

Preface: I ran the experiment that hurt my own sales pitch	1
Who this is for	1
What is measured here, precisely	2
What is not measured here	2
Two disclosures, because they change how you should read this	3
1. The claim everyone repeats	4
Where I got it	4
The three things that got mixed together	4
What I expected	4
2. How do you measure obedience at all?	5
The uninstrumented zero	5
The instrument: observe in every condition, deny in one	5
Isolation, and why it is not optional	5
The four times my instrument lied to me	5
3. Does the model even reach for the banned command?	6
The banned command, and why this one	6
The task shapes	6
Result	6
The smoke run that was almost a published finding	6
The trap this closes	6
4. Nothing, a rule, a hook	7
The instruction file, in full	7
Result	7
What the model actually ran	7
The obvious objection	7
5. Does length break it? Does position?	8

CONTENTS

Building a realistic long file	8
Result	8
The interpretation, stated carefully	8
Which leaves the third hypothesis	8
6. What happens when the file argues with itself	9
First attempt, and why it did not count	9
Second attempt: a fair fight	9
Result	9
What I take from this, and what I do not	9
Where the argument stands	9
7. Raising the stakes: banning deletion	10
The safety design, which is most of the work	10
The task	10
The conditions	10
Result	10
The finding that evaporated	10
Where this leaves the ban	10
8. Is it the ban, or the escape route you offered?	12
Condition L: the ban with nothing offered	12
Result	12
What this settles	12
The limit that matters most here	12
The complete picture	12
9. All thirty-nine trials	13
Nothing else moved	13
Where I got the arithmetic wrong	13
The recovered trials	13
Reading the number honestly	13
10. What thirty-nine trials cannot tell you	14
Zero does not mean never	14
The pooled p-value is not what it looks like	14
Everything that was held fixed	14
What the instrument physically cannot see	14
What I would attack first	14

11. What a hook actually adds	15
The claim I can no longer make	15
What is actually left	15
The distinction that survives	15
A four-layer picture	15
The honest version of the sales copy	15
Two mirror-image errors, one of which was mine	15
12. Your guard hook fails open	17
The design, and the instrument that lied a fifth time	17
Result	17
The failure mode that will actually bite you	17
The other hole, found by accident	17
What to do about it	17
13. What this means for your CLAUDE.md	18
Measured: name the command	18
Measured: do not reorganise your file for attention	18
Measured: naming an alternative is optional, but it makes the workaround predictable	18
Measured: do not contradict yourself, but not for the reason you think	18
Judgement: what to put in the file at all	18
Judgement: the three questions before adding a line	19
What I have not measured and will not pretend to	19
14. What 914 hooks actually say when they block	20
How many of them block at all	20
What the blocking message contains	20
The hand count	20
What I conclude, and what I withdraw	20
Why any of this matters for your own hooks	20
15. The rule that points at a file which is not there	21
The measurement	21
The limits, which are substantial	21
Why this belongs in a book about obedience	21
The check, which takes a minute	21
16. Run it yourself	22
Before anything: two safety rules	22

The minimum viable version	22
Scripts	22
Adapting it to your situation	22
What a disagreement would mean	22
17. The replication, in English	23
What was run	23
Result	23
What this closes, and what it does not	23
The reason this chapter exists	23
Appendix A: Every trial	24
A.1 Per-trial data, replacement task	24
A.2 Per-trial data, deletion task	24
A.3 Recovered trials (reported separately, not in the totals)	24
A.4 Hook failure modes (Chapter 12)	24
A.5 The thirty hooks read by hand (Chapter 14)	24
A.6 English replication (Chapter 17)	24
Appendix B: The harness	26
B.1 Requirements	26
B.2 Two things to do before you run anything	26
B.3 The script	26
B.4 Reading the output	26
B.5 The instruction files used in the book	26
B.6 The hook failure-mode test (Chapter 12)	26

Preface: I ran the experiment that hurt my own sales pitch

I sell tools that stop Claude Code from doing dangerous things. A free hook collection, and a paid book about preventing accidents. The sentence at the centre of everything I have ever written about this is:

Rules you write down get skipped. That is why you need to stop things at the execution layer.

In late August 2026 I finally measured it.

Thirty-nine trials, one version of Claude Code, two task families, twelve conditions. **In the twenty-two trials where the ban was written into a project `CLAUDE.md`, the banned command was never once attempted. In the seventeen trials where it was not written, it was attempted seventeen times out of seventeen.**

Length did not matter. A ban buried in the middle of a 128-line instruction file was obeyed. A ban near the end was obeyed. A ban sitting in the same file as two other sections explicitly telling the model to *use the banned command* was obeyed. Raising the stakes from “replace a string in a config file” to “delete files you cannot get back” did not change the direction either.

That is the opposite of what I had been telling people for months.

This book is the whole thing: the instrument, the raw counts, the moments the instrument was broken and how I caught it, the conditions that failed to break obedience, and the honest boundary around what 39 trials can and cannot settle.

Who this is for

You maintain a `CLAUDE.md` – or an `AGENTS.md`, or a `.cursorrules`, or whatever your tool calls its instruction file – and you have a nagging question you have never been able to answer:

Is anything in this file actually doing anything?

You have probably read advice about it. Keep it short. Put the important rules at the top. Use ALL CAPS. Repeat yourself. Add “IMPORTANT” and “NEVER”. Most of that advice, mine included, was written by people who had not measured it, which is a polite way of saying it was made up.

I do not think you should trust my numbers either. So this book ships the harness: Chapter 16 is how to run it, and Appendix B is the script itself, which you can point at your own machine, your own version, your own task. If your numbers disagree with mine, yours are the ones that describe your situation.

What is measured here, precisely

- **Claude Code 2.1.246**, one version, pinned. Nothing else.
- **Two task families**. A reversible one (replace a string across 12 config files) and an irreversible one (clean up a directory containing files that must survive).
- **Project-level CLAUDE.md**, placed in the working directory. Not the user level `~/.claude/CLAUDE.md`.
- **What is counted is an attempt**, not an outcome. A `PreToolUse` hook records every Bash command before it runs, in every condition, including the conditions with no rule at all. That detail is the whole experiment; Chapter 2 explains why an uninstrumented zero is worth nothing.

What is not measured here

I would rather you know the holes before you pay than find them afterwards.

- **The thirty-nine trials were run in Japanese**. I work in Japanese. Every rule in this book is printed in the original and in translation so you can see exactly what the model was given. Before shipping I re-ran the core comparison in English and it reproduced – that is Chapter 17 – but only the core comparison. The long-file, conflict and deletion conditions exist in Japanese only.
- **One model version**. A newer or older build may behave differently. This is not a claim about Claude Code in general; it is a claim about 2.1.246 on the tasks below.

- **Small n.** Three or four trials per condition. Zero out of twenty-two is a direction, not a guarantee. Chapter 10 does the arithmetic on exactly how weak that is, including the smallest p-value the design can possibly produce.
- **Two task families out of all possible work.** Both are file manipulation. Nothing here is measured about refactoring, dependency changes, network calls, or anything involving a long agentic session.

Two disclosures, because they change how you should read this

I do not write the code myself. I direct an agent. The harness, the scorers and the analysis in this book were built that way, which is also a large part of why the instrument broke six times and why Chapter 2 exists. If that makes you want to check the numbers rather than take them, good – Appendix A is every trial and Appendix B is the script.

I have a commercial interest in the answer. I give away a hook collection and I sell a book on preventing accidents with coding agents. The result in these pages makes the *marketing* for both weaker: I can no longer say “your rules get skipped, so buy the enforcement layer.”

What survives is smaller and, I think, truer, and Chapter 11 is about that distinction: **a rule changes what the model chooses; a hook changes what happens after it has chosen.** The two look identical for a long time before they differ.

Then I checked the second half of my own pitch as well, and it came apart too. A `PreToolUse` hook blocks on exactly one exit status. A hook with a syntax error exits with a different one, and does not block. Whether a *missing* hook script protects you turns out to depend on which interpreter you named in the command: two of the three common spellings fail closed, and one fails open silently. Chapter 12 is that measurement, and it is the part of this book I would read first if I ran guard hooks in anger.

I published the result that undercut my own sales copy because the alternative was to keep selling a claim I had just watched fail. Which of those is the better business decision is, at the time of writing, still unmeasured.

1. The claim everyone repeats

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Where I got it

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The three things that got mixed together

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What I expected

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

2. How do you measure obedience at all?

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The uninstrumented zero

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The instrument: observe in every condition, deny in one

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Isolation, and why it is not optional

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The four times my instrument lied to me

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

3. Does the model even reach for the banned command?

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The banned command, and why this one

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The task shapes

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The smoke run that was almost a published finding

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The trap this closes

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

4. Nothing, a rule, a hook

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The instruction file, in full

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What the model actually ran

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The obvious objection

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

5. Does length break it? Does position?

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Building a realistic long file

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The interpretation, stated carefully

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Which leaves the third hypothesis

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

6. What happens when the file argues with itself

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

First attempt, and why it did not count

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Second attempt: a fair fight

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What I take from this, and what I do not

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Where the argument stands

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

7. Raising the stakes: banning deletion

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The safety design, which is most of the work

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The task

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The conditions

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The finding that evaporated

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Where this leaves the ban

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

8. Is it the ban, or the escape route you offered?

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Condition L: the ban with nothing offered

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What this settles

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The limit that matters most here

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The complete picture

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

9. All thirty-nine trials

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Nothing else moved

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Where I got the arithmetic wrong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The recovered trials

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Reading the number honestly

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

10. What thirty-nine trials cannot tell you

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Zero does not mean never

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The pooled p-value is not what it looks like

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Everything that was held fixed

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What the instrument physically cannot see

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What I would attack first

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

11. What a hook actually adds

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The claim I can no longer make

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What is actually left

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The distinction that survives

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A four-layer picture

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The honest version of the sales copy

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Two mirror-image errors, one of which was mine

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

12. Your guard hook fails open

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The design, and the instrument that lied a fifth time

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The failure mode that will actually bite you

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The other hole, found by accident

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What to do about it

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

13. What this means for your CLAUDE.md

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Measured: name the command

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Measured: do not reorganise your file for attention

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Measured: naming an alternative is optional, but it makes the workaround predictable

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Measured: do not contradict yourself, but not for the reason you think

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Judgement: what to put in the file at all

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Judgement: the three questions before adding a line

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What I have not measured and will not pretend to

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

14. What 914 hooks actually say when they block

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

How many of them block at all

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What the blocking message contains

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The hand count

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What I conclude, and what I withdraw

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Why any of this matters for your own hooks

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

15. The rule that points at a file which is not there

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The measurement

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The limits, which are substantial

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Why this belongs in a book about obedience

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The check, which takes a minute

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

16. Run it yourself

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Before anything: two safety rules

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The minimum viable version

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Scripts

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Adapting it to your situation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What a disagreement would mean

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

17. The replication, in English

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What was run

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Result

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

What this closes, and what it does not

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

The reason this chapter exists

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Appendix A: Every trial

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A.1 Per-trial data, replacement task

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A.2 Per-trial data, deletion task

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A.3 Recovered trials (reported separately, not in the totals)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A.4 Hook failure modes (Chapter 12)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A.5 The thirty hooks read by hand (Chapter 14)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

A.6 English replication (Chapter 17)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

Appendix B: The harness

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

B.1 Requirements

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

B.2 Two things to do before you run anything

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

B.3 The script

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

B.4 Reading the output

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

B.5 The instruction files used in the book

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.

B.6 The hook failure-mode test (Chapter 12)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/claude-md-under-test>.