ACHILLES TRADE SYSTEM · EDITION SIX

The Achilles Trading System

A, B, Classic, Bar and SMC — every feature, what it means, and how to use it.

Achilles A, SMC, B and Bar running together on a live chart.

A
PRICE PANE

B
LOWER PANE

Classic
LOWER PANE · ORIGINAL

Bar
CONTEXT STRIPS

SMC
STRUCTURE

Welcome

What the Achilles Trading System is, and what it is for

The Achilles Trading System is a family of five indicators built to work as one. They were made because most trading tools share the same problem: they tell you what to do, and they never tell you how often that worked. These do the opposite. They are built to **describe the market accurately** — where the important prices are, how stretched things have become, whether real money is moving — and then hand the decision back to you, with an honest note about how much each reading is actually worth.

The five pieces cover different jobs and are designed to be read together:

- **Achilles A** puts the reference prices on your chart — the bands, the key averages, the day's volume-weighted average price — and keeps a running confluence score in the corner.
- **Achilles B** is the momentum pane: one clear wave, a handful of readings, and a compact card that tells you the state of everything at a glance.
- **Achilles B Classic** is the same engine in the denser, more familiar layout, for traders who already think in that visual language.
- **Achilles Bar** adds context strips — how volatile, how heavy, how bought or sold this session is.
- **Achilles SMC** does the structural work: it finds the zones, the untaken highs and lows, and the support and resistance that everything else is measured against.

> IT WORKS ON ANY MARKET TRADINGVIEW CARRIES
>
> Nothing here is tuned to a single coin or a single exchange. The calculations use nothing but open, high, low, close and volume, so the suite runs on **crypto, stocks, forex, futures, indices and commodities** — anything you can pull up a chart of. It works on any timeframe from seconds to months.
>
> ⚠️ One honest caveat: the volume-based features — the whale flags, the volume strip, the volume profile — need a feed that actually reports volume. Most do. A few spot-forex and index feeds do not, and on those the volume readings will simply sit flat. Everything else still works.

A word on how to read this book. Every feature carries a small label saying whether it ships switched on, whether it is still experimental, and — where we have tested it — what the test found. **Some of those labels say the feature does not predict anything.** Those are the honest ones, and they are there on purpose: a tool that describes the market well is genuinely useful, and knowing which of your tools are description and which are prediction is most of what separates a consistent trader from a hopeful one.

Start here

Three things to know before you turn anything on

> THE ONE RULE
>
> **Structure first, indicators second.** Decide where the important prices are — the highs, the lows, the level you actually care about — and only then look at what these tools say is happening there. None of them is a system on its own. Every one of them is a way of describing a level you have already chosen.

Everything switched on at once. A real chart running Achilles A and Achilles SMC with almost every option enabled — shaded zones, session bands, diamonds, crosses, a phase label and two cards, all at the same time. Every one of them is working correctly, and together they are very hard to read. You will not start here: the setup this book recommends turns most of it off, and you add things back one at a time as you find a use for them.

1 · The level — you choose it → **2 · The reading** — the indicators go here → **3 · The decision** — yours again

The suite is deliberately honest about what it knows. Most indicators tell you they predict things. These have been measured, and most of them *don't* — they describe. That is still useful, and it is a great deal more useful than a signal you trust for no reason. Every feature in this book carries a label:

LABEL	WHAT IT MEANS
ON BY DEFAULT	You will see it the moment you add the indicator.
OFF BY DEFAULT	Present, but switched off so your first chart is readable. Turn it on when you want it.
EXPERIMENTAL	Built and working, but not validated. Treat as a map, not a signal.
MEASURED NULL	We tested it. It does not predict direction. It is drawn because it describes something, not because it wins.
MEASURED	A number we have actually checked, with the sample size stated.

The five indicators, in one table

INDICATOR	WHERE IT LIVES	WHAT IT IS FOR
Achilles A	On the price chart	Bands, key averages, and a confluence score
Achilles B	Own pane	The momentum wave — current design
Achilles B Classic	Own pane	The original wave, for those who prefer it
Achilles Bar	Own pane	Four context strips, made to sit under Achilles B
Achilles SMC	On the price chart	Structure, blocks, levels and sessions

DON'T RUN EVERYTHING AT ONCE

Achilles B and Achilles B Classic draw the same wave in two designs. **Pick one.** Running both is not more information, it is the same information twice — and two panes of it will crowd the chart you are actually trying to read.

Achilles A

Sits on the price chart · price levels and confluence

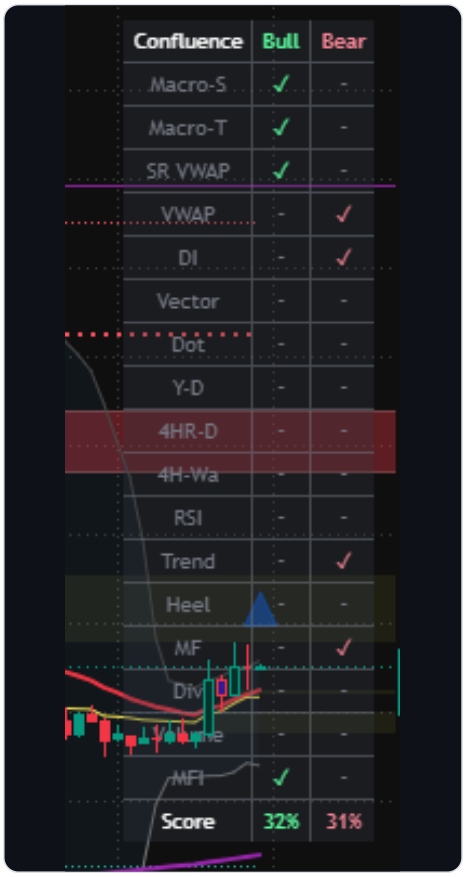
WHAT IT IS

The context layer. It draws a small number of well-known reference prices directly on your candles, and keeps a scorecard in the corner that counts how many separate things currently lean bullish versus bearish.

WHAT IT DRAWS

FEATURE	WHAT IT IS	STATUS
Achilles Bands	A middle average with a band either side, set a fixed distance out by how volatile the market has been. Price spends most of its time between them, so the edges mark an unusually stretched price — not a reversal. Presets for scalping and day trading change how quickly they react.	ON
The Achilles Line	Two moving averages in one line. It is drawn at the **midpoint of the 9 and 21 EMA** and coloured by their relationship — green while the 9 is above the 21, red while it is below. One line instead of a crossing pair.	ON NULL
200 EMA	The long-term average almost every trader watches. It matters partly because so many people are looking at it.	ON
SR VWAP	The volume-weighted average price since midnight UTC — the day's true "average cost". Above it leans bullish, below it leans bearish.	ON MEASURED
Whale detectors	Marks candles that traded on unusually heavy volume.	ON EXPERIMENTAL
Confluence Scorecard	Fifteen separate checks, weighted and totalled into a Bull % and a Bear %.	ON MEASURED

The Confluence Scorecard — the main event



Fifteen checks, two columns, one score. Each row is an independent question — is the session VWAP bullish, is momentum positive, is the trend up — and each carries a weight earned by how it performed in testing. The bottom row totals them into a Bull % and a Bear %.

This is the part of Achilles A worth learning properly, because it is the one reading in the whole suite with evidence behind it. **Read the two percentages, not the ticks.**

- **Both low and close together** — as in the 32% / 31% above — means the checks disagree with each other. That is a market with no consensus, and it is the most common state.
- **One side clearly ahead** means many unrelated measurements are pointing the same way at once. That agreement is the signal, not any individual row.
- **A single tick means nothing.** The card works because fifteen things agree, and any one of them alone is noise.

The weights are not opinions. The daily session VWAP carries **14** because it held up in testing; money flow carries **5** because it did not. That is the difference between a scorecard and a wishlist.

The markers on your candles

Achilles A prints small shapes above the candles. Each one means something specific, and there are only five to learn:

Bullish flag
Heaviest-volume candle, closing up

Bearish flag
Heaviest-volume candle, closing down

Blue triangle
Momentum turning up from a low

Yellow & red crosses
Momentum stretched — a caution mark

Blood diamond
The strongest of the caution marks

All of them
Notes, not instructions — see below

THE FLAGS ARE THE ONES TO ACTUALLY WATCH

The **bullish and bearish flags** are the Whale Detector, and they are deliberately rare: they print only on the *heaviest* tier of volume, not merely above-average. An earlier version flagged the lighter tier too and printed so often that the mark stopped meaning anything.

A flag tells you a lot of money changed hands on that candle and which way it closed. It does **not** tell you the move will continue. Treat it as "something happened here worth remembering", and use the level it happened at.

How to use it

Read the SR VWAP first. It answers one question cleanly — is price above or below today's average cost? Of the eleven components we measured individually, it was one of only three that held up on both halves of the test, and it carries the heaviest weight on the scorecard.

Then read the scorecard total, never a single row. The card works through *breadth* — the fact that many unrelated things agree. One tick on one row means nothing at all.

THE RESULT WE WITHDREW

The scorecard once measured **+2.7 percentage points** on data it had never seen, and for a while we believed it. **It was wrong, and the error was ours.** A signal that goes both ways was being scored against an always-long benchmark, which quietly hands a short-biased signal free points from its mix rather than its timing. Corrected against a like-for-like control the edge fell to **+0.31 and +0.37 points** across two holdout sets — statistically indistinguishable from zero.

So the card is **not** a proven edge, and we are not going to tell you it is. What it still does well is compress fifteen separate checks into two numbers, which is a genuinely useful summary of whether the market currently agrees with itself.

Achilles B

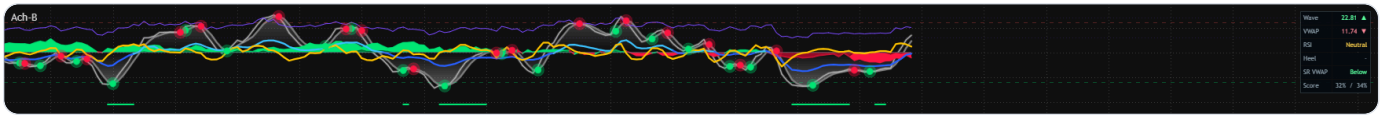
Own pane · the next-generation wave design

WHAT IT IS

The next-generation wave design. Same engine as the original, rebuilt from the display up: one clear wave with a soft glow instead of a crowded field of overlapping lines, a compact readout card that states each reading in plain words rather than numbers you have to interpret, and colours chosen by measuring their contrast on a dark screen rather than by eye. It is the modern version of the layout, and it is where you should start if you are starting fresh.

WHAT IT DRAWS

FEATURE	WHAT IT IS	STATUS
The Wave	The main momentum line. Above zero is bullish pressure, below is bearish. Its *turns* matter more than its level.	ON
Trigger line	A faster line that crosses the wave. The cross is where the dots come from.	ON
Buy / Sell dots	Mark a wave cross. A "live" dot appears on the still-forming bar in a second colour — it can vanish before the bar closes.	ON NULL
Money Flow	Whether buying or selling has been dominant.	ON NULL
Momentum Line	Trend strength and direction, from the DI system.	ON
RSI	How stretched price is. Banded into three states on the readout.	ON MEASURED
Achilles Heel	Marks when the DI and VWAP readings disagree unusually strongly — a stretch reading, not a signal.	ON EXPERIMENTAL
The readout card	Five lines: Wave, VWAP, RSI, Heel, SR VWAP — plus the Confluence score.	ON

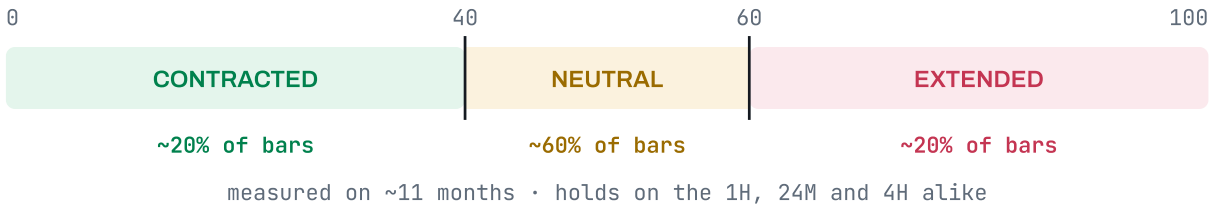


Achilles B, as it actually looks. The wave with its glow, buy and sell dots on the crossings, and the readout card on the right — the six lines that tell you the state of everything without reading the lines themselves.

How to read the card

The card is the fastest way in. Each line is one plain-English state rather than a number you have to interpret:

LINE	READS	MEANING
Wave	+12.4 ▲	Momentum and whether it is rising
VWAP	-3.1 ▼	The wave's own VWAP, and its direction
RSI	Extended / Neutral / Contracted	How stretched price is — see below
Heel	Exhaustion / Pressure / -	Whether the stretch reading is firing
SR VWAP	Below / Above	Where the day's average price sits relative to price. **Below** means price is holding above it — bullish.
Score	32% / 31%	Confluence Bull % and Bear %. Dims when they are within 5 points, because that is a tie.



THE RSI BANDS ARE CALIBRATED, NOT GUESSED

Extended is 60 and above. Contracted is 40 and under. Everything between is Neutral. Those cut-offs were measured on roughly eleven months of data: they put about 20% of bars in each outer band and 60% in the middle, and they do it consistently on the 1-hour, 24-minute and 4-hour charts alike. So "Extended" genuinely means the top fifth of readings — not "above average".

WHAT THE DOTS ARE NOT

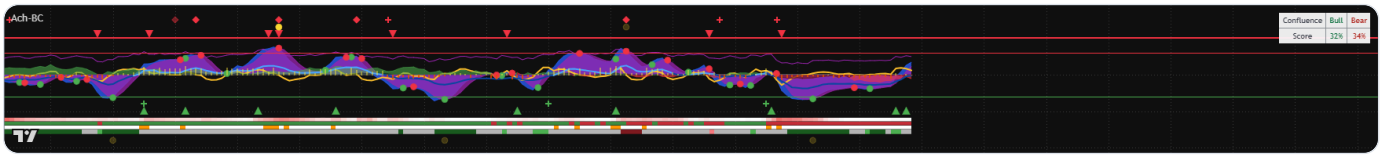
The buy and sell dots mark a momentum cross and nothing more. Measured across **22,500 instances they were right 50.0% of the time** — exactly chance — and they fire a median of **19 bars before** the actual turn. Use them to notice that momentum has shifted. Do not use them as an entry.

Achilles B Classic

Own pane · the original design

WHAT IT IS

The wave as it was originally built — everything Achilles B has, plus the extra layers that were later moved out or retired. Kept for people who learned on it and prefer the denser display.



Achilles B Classic. If you have used a classic momentum oscillator before, this will look immediately familiar — the filled wave, the dots on the crossings, the yellow money-flow line and the coloured strips underneath. That familiarity is the entire reason Classic still ships: if you already read charts this way, there is no reason to make you relearn it.

Classic adds, on top of everything in Achilles B:

EXTRA FEATURE	WHAT IT IS	STATUS
Odysseus Wave	A second, slower wave shown as a histogram behind the main one.	ON EXPERIMENTAL
Divergences	Marks where price and the wave disagree.	ON NULL
The four bars	Money Flow, Volume, Volatility and Heel strips along the bottom. *These are now their own indicator — see Achilles Bar.*	ON
Market Maker arrows	Marks heavy-volume climax candles.	ON EXPERIMENTAL
EMA cross	The 9/21 crossover.	ON NULL

WHICH ONE SHOULD YOU RUN?

Achilles B if you want a clean pane and a readout you can take in at a glance. **Classic** if you want everything on screen and are comfortable with a busy chart. They compute the same wave from the same formula — the difference is what is drawn, not what is calculated.

Two of Classic's extras — divergences and the EMA cross — have both been measured and neither predicts direction. They are on in Classic because Classic is the "everything" build.

Achilles Bar

Own pane · made to sit with Achilles B

WHAT IT IS

The four context strips from the bottom of Classic, lifted into their own pane so they can be used alongside the cleaner Achilles B. Each strip is a coloured band that answers one question about *what kind of market this is right now*.

STRIP	WHAT IT TELLS YOU	STATUS
Volatility	How active the market is against its own recent range. Dark when quiet, bright when moving.	ON
Money Flow	Green while buying pressure dominates, red while selling does.	ON NULL
Volume	Brightens when volume runs above its 30-bar average, and again when the 4-hour is also spiking.	ON
Achilles Heel	How stretched the momentum spread is.	OFF EXPERIMENTAL

Achilles Bar. Each strip is one question answered along the whole chart — how volatile, bought or sold, and how heavy the volume. Nothing here points a direction; they describe the session you are trading in.

The Heel strip ships **off** here for a specific reason: Achilles B already shows the Heel. The same reading appearing twice on one screen looks like two pieces of evidence when it is one. Turn it on if you are running Bar without Achilles B.

> **READ THEM AS WEATHER, NOT AS SIGNALS**
>
> None of the four strips tells you which way to trade. They tell you what sort of session you are in — quiet or active, bought or sold, ordinary or unusual. That is genuinely useful when deciding *whether* to trade at all, and useless as a reason to pick a direction.

Achilles SMC

Sits on the price chart · structure, zones and levels

WHAT IT IS

The biggest tool in the suite and the one that does the most drawing. It marks market structure, zones where price previously moved away sharply, and the handful of prices that matter each day. It is also the one you should switch *off* most of.

FEATURE	WHAT IT IS	STATUS
Achilles Blocks	Zones where price moved away sharply, marked as areas to watch on a return.	ON EXPERIMENTAL
Achilles Magnets	Untaken highs and lows — places where resting orders are likely to sit.	ON NULL
Volume Profile	The price where the most volume traded, plus levels that were never revisited.	ON EXPERIMENTAL
SS/R levels	Swing support and resistance from confirmed pivots.	ON
Market structure	Higher highs, lower lows, and the external bias card.	ON
Daily & prior levels	Today's high and low, yesterday's, and the overnight range.	ON
Vector candles	Colours candles by how heavy their volume was.	ON
Swing Failure finder	Marks where price swept a level and closed back. Also marks *candidate* levels where one could happen.	OFF EXPERIMENTAL
Fib retracement	Retracement levels of a swing, automatic or hand-placed.	OFF NULL
Market phases	Labels the chart with accumulation, markup, distribution and markdown — the four stages a market is usually described as moving through.	ON EXPERIMENTAL
Market sessions	Shades the Tokyo, London and New York hours.	OFF

An Achilles Block, and why it matters. Price traded quietly inside the shaded zone, then left it in a single decisive candle. That zone is where the orders that caused the move were sitting — and the horizontal lines carry it forward, so when price returns weeks later you are looking at a level with a reason behind it rather than a line you drew by hand.

What SMC gives you that a blank chart does not

Most traders mark structure by hand: draw a line at the high, another at the low, shade a zone where price took off, redraw it all when the timeframe changes. SMC does that work continuously and consistently, and it does it the same way every time — which is the part hand-drawing cannot promise.

WHAT IT FINDS	WHY IT HELPS
Achilles Blocks	The zones price left in a hurry. When price comes back to one, you are trading against a level where something already happened — not a guess. Blocks are tracked from the moment they form, marked when price engages them, and retired when they break.
Support & resistance that updates itself	SS/R draws from confirmed swing pivots, so lines appear only once a swing is genuinely complete. They extend forward, retire when price closes through them, and can be locked to a higher timeframe so a 5-minute chart still shows the hourly levels everyone else is watching.
Market structure, marked for you	Higher highs and lower lows identified continuously, with an external-structure card that states the current bias in one word. No more squinting at whether that was really a higher low.
Achilles Magnets	The highs and lows that were never taken — where resting orders still sit. It scores and ranks them, and shows the distance to each, so you can see at a glance what is above you and what is below.
Volume Profile	The price where the most business was actually done, plus the levels price has never returned to. The profile timeframe is fixed, so the levels do not shift under you when you change charts.
Daily and session levels	Yesterday's high and low, today's, the overnight range and the trading sessions — the reference prices most of the market is already reacting to.

WHY THIS IS THE FOUNDATION, NOT A DECORATION

Every other tool in this suite answers the question *"what is happening here?"* — and none of them can answer it until you have decided what *here* means. **SMC is what produces that shortlist of prices.** Blocks and magnets give you a handful of levels with a reason attached; the wave, the score and the strips then tell you what is going on when price arrives at one.

That is the whole method, and it is why this section sits at the structural end of the book rather than the decorative one.

How to use it

Blocks and magnets are the two to learn first. A block tells you where price previously left in a hurry. A magnet tells you where unfilled orders probably rest. Between them you get a short list of prices worth caring about — which is exactly the "structure first" the rest of the suite is built on.

THE CANDIDATE MARKS ARE THE NOISIEST THING HERE

The Swing Failure finder can mark every level where a sweep *might* happen. On a busy chart that is dozens of marks that are not signals — they are only places to watch. The switch is called **Show candidate levels**, and turning it off is the single quickest way to quieten the chart.

A KNOWN ROUGH EDGE

SMC currently carries around 275 settings, which is more than anyone needs and more than is comfortable to navigate. Trimming that down is active work. If a setting is not described in this book, you almost certainly do not need to touch it.

What ships on, what ships off

The exact settings a new install starts with

Add all five indicators to a fresh chart and this is what you get. The principle behind the list is simple: **a first chart should be readable, not complete.** Everything switched off is still there, one tick away, waiting for you to have a reason to want it.

Achilles A — everything on

All five features run by default: the **Achilles Bands**, **The Achilles Line**, the **200 EMA**, the **SR VWAP**, the **Whale Detector** flags and the **Confluence Scorecard**. Nothing here is noisy enough to need hiding, and the scorecard is the reading we most want you to see.

Achilles B — everything on

The wave and its trigger, the buy and sell dots including the live one on the forming bar, money flow and its pressure line, the Momentum Line, RSI, the Achilles Heel, and the full readout card with the SR VWAP row and the Confluence score. The glow and the gradient wash are on too — they are what make one wave readable where four lines would not be.

Achilles B Classic — everything on except one

FEATURE	STATE	WHY
Odysseus Wave, divergences, EMA crosses, Market Maker arrows, all four bottom bars	ON	Classic is the “everything” build. That is the point of it.
Confluence Scorecard	OFF	Achilles A already draws it. Two identical cards on one screen is clutter, not confirmation.

Achilles Bar — three of four strips

STRIP	STATE	WHY
Volatility, Money Flow, Volume	ON	Three different questions, three different answers.
Achilles Heel	OFF	Achilles B already shows the Heel. The same reading twice on one screen looks like two pieces of evidence when it is one.

Achilles SMC — the one with real choices made

This is where the defaults do the most work, because SMC can draw far more than any chart can carry.

ON BY DEFAULT	OFF BY DEFAULT
Achilles Blocks	Fib retracement
Achilles Magnets, plus the nearest untaken level each side	Swing Failure finder
SS/R levels — *prices shown, lines hidden*	Candidate SFPs
Macro External Structure and its card	Session boxes
Vector candles and vector zones	SS/R level lines
Volume POC, yesterday's POC and naked POCs	Liquidity section on the structure card
Daily open, day high/low, PDH/PDL, Psy-Hi/Lo	Colouring of non-vector candles
Market sessions — *labels only, no boxes*	
Market phases	

TWO CHOICES WORTH EXPLAINING

SS/R shows you the prices but not the lines. You get the numbers on the axis where the levels are, without eight horizontal lines across your candles. Turn the lines on when you are working a specific level.

Sessions show labels but not boxes. You can see where Tokyo, London and New York begin without three shaded rectangles sitting behind your price action.

Both are the same idea: keep the information, drop the ink.

WHY SOME UNVALIDATED THINGS STILL SHIP SWITCHED ON

Three of the defaults — **Achilles Blocks**, **Achilles Magnets** and the **Volume Profile** — carry an EXPERIMENTAL or MEASURED NULL label in this book, and are still on by default. That is deliberate, and it is worth being clear about.

They are on because they are **descriptive tools that do a job no test is required to justify**: they tell you where price previously moved away sharply, where untaken orders sit, and where the most business was done. Those are facts about the chart. What has not been shown is that any of them *predicts* what happens next — and that is what the labels are telling you.

Use them to build your shortlist of prices. Do not use them as a reason to enter. That distinction is the whole point of labelling them at all.

THE TWO BIGGEST THINGS ARE OFF

The **Fib retracement** and the **Swing Failure finder** ship switched off. They are the two that draw the most, and both are experimental. Turn them on one at a time, and give each a few sessions before adding the next — that is the only way to tell whether a tool is helping you or just filling the screen.

How this compares to other momentum indicators

An honest answer, including where we are the same

If you have used a momentum-style oscillator before, the wave in Achilles B will look familiar — a line that rises and falls around a centre, with marks where it turns. That is deliberate. Momentum is a well-understood idea and there is no advantage in reinventing how it is drawn, so the pitch here is not that we invented a better wave.

> THE ACTUAL DIFFERENCE
>
> **We measured ours, and we tell you what we found — including the parts that do not work.** Every claim in this book carries a sample size. We have not seen another indicator in this family publish a single one.

We ran the comparison ourselves

The obvious question about any of these tools is whether the money-flow reading actually predicts anything. So we tested it — ours *and* theirs, on the same bars:

MONEY FLOW	FORWARD EDGE	SAMPLE
Achilles	`+0.6 / +0.9pp — zero`	`453,176 bars, holdout-confirmed`
Momentum-style equivalent	`zero forward information`	`measured head-to-head`

Neither one predicts direction. That is not the answer we hoped for, and it is the answer. Our test ran across three coins and four timeframes with a proper holdout, and an early result that looked promising on one coin over two months *reversed sign* once the full sample was in — which is exactly why single-chart impressions are worth so little.

So money flow is on the chart because it describes buying and selling pressure, and it carries a weight of **5 out of 100** in our Confluence Scorecard. That number is not a guess. It is what the measurement earned it.

What you get here that we have not seen elsewhere

FEATURE	WHY IT MATTERS
Published test results	Every reading in this book is labelled with what we found and how much data we found it on. The last section is nothing but results, including the unflattering ones.
Weights earned by measurement	The scorecard weights each component by how it performed, not by how important it feels. Money flow gets 5; the session VWAP gets 14.
Thresholds calibrated to the market	Our RSI bands are 40 and 60, taken from eleven months of this instrument's own distribution — not the textbook 70/30, which on this market would call a third of all bars "extended".
Candle-type-proof readings	Every calculation samples the real price feed. Switch to Heikin Ashi and the numbers stay honest — on many scripts they quietly change, because Heikin Ashi candles are averages rather than real prices.
We withdrew our own headline result	Our scorecard measured +2.7 percentage points on unseen data. We found the control was wrong, corrected it, and the edge disappeared. We published the retraction. **We make no performance claim at all** — and we would rather tell you that than sell you a number we could not defend.

WHY WE ARE TELLING YOU THIS

An indicator that promises an edge and never shows a test is asking you to take its word. The most useful thing we can give you is the opposite: a clear list of what describes the market well, and a short list of what actually predicted anything. You will trade better knowing which is which.

What we actually tested

Every result, in one table

Most indicator documentation has no section like this. This one exists because the alternative — asking you to trust a tool because it looks like it works — is how people lose money slowly.

The headline: nothing in this suite has been shown to predict direction. Not one feature, including the one we built the scorecard around. What follows is the evidence, and then why these tools are still worth having on your chart.

How to read the results

- **Measured null** — the test could have found a useful effect and did not.
- **Could not tell** — the effect, if any, was smaller than the test could resolve. Not the same as "no effect".
- **Withdrawn** — a result we published and later found to be an artefact of how it was measured.
- **Description** — verified to draw what it claims to draw; never tested as a forecast.

Samples count **non-overlapping events**, not bars. An effect must exceed **2.8 times its own standard error** to count. For scale: **the largest effect this project has ever measured is +0.024R** — about 2.5% of one unit of risk. Treat any larger claim, from anyone, with suspicion.

The results

WHAT WAS TESTED	RESULT	SAMPLE	VERDICT
Confluence Scorecard	+2.7pp → +0.31 / +0.37pp	2 holdouts	WITHDRAWN
SR VWAP — price above/below session VWAP	51.1% / 54.3%	2 halves	WEAK
Support & resistance levels — three independent routes	-0.033R to -0.036R	18,243	MEASURED NULL
Order blocks — does zone quality matter?	0 of 12 cells clear the floor	40,000/symbol	MEASURED NULL
Order blocks — the earlier "60% vs 33%" claim	nothing above ~6pp survives	40,000/symbol	REFUTED
RSI bands	no effect above ~0.4 ATR	~8,000 bars	MEASURED NULL
Divergences	-0.065R	32	MEASURED NULL
Swing failures (SFPs)	could not tell	0.041R floor	COULD NOT TELL
SFP volume filter	could not tell	—	COULD NOT TELL
Market structure & phase labels — accuracy check	5 of 5 phases matched, in order	live chart	DESCRIPTION
Market phases — four scoring variants tested	zero shipped; the volume term is barely about volume	—	MEASURED NULL
Buy / sell dots	50.0% — exactly chance	22,500	MEASURED NULL
Dot timing	median 19 bars early	22,500	MEASURED NULL
9 / 21 EMA cross	50.2% · 50.3% · 50.7%	ETH · BTC · SOL	MEASURED NULL
Our money flow	+0.6 / +0.9pp	453,176	MEASURED NULL
A leading momentum indicator's money flow	zero forward information	head-to-head	MEASURED NULL
Liquidity magnets — do bigger pools pull harder?	~99% swept either way	34,951	MEASURED NULL
Fading a sweep	48–49% regardless of size	34,951	MEASURED NULL
Fib 0.618 vs its neighbours	-0.14pp	—	MEASURED NULL
Volume ladder — higher-timeframe confirmation	-0.855 ATR, MDE 2.449	237 events	COULD NOT TELL
Control — every bar, no condition	-0.032 ATR	2,494	BASELINE

Achilles Bands vs Keltner channels	`indistinguishable`	—	DESCRIPTION

Read the control row. Across 2,494 samples a bar's own direction continues essentially never — the average outcome is zero. That is the bar every result above has to clear, and none of them do.

The one we got wrong

The Confluence Scorecard measured **+2.7 percentage points** on unseen data, and for a while we believed it. The control was wrong: a signal that trades both ways was being scored against an always-long benchmark on a sample that drifted down, so it collected points for its *direction mix* rather than its timing. Scored against its own mix, the edge fell to **+0.31 and +0.37 points** across two holdouts — zero.

We found it, we corrected it, and we withdrew our only positive result. That is what the process is for.

> **OUR NULLS MEAN LESS THAN WE WOULD LIKE, TOO**
>
> A null only counts if the test could have found something. The tightest significance floor anywhere in this work is **0.0676R** — nearly three times the largest effect we have ever measured. **An edge of realistic size would slip through most of these tests unnoticed.** So nothing here is described as disproven, and "could not tell" is kept separate from "measured null" throughout.

Two things that are arithmetic, not prediction

- **The fee floor is 0.20% of price.** Below that stop distance, costs eat more than 15% of your risk — about $4.80 at a $2,400 price, and short-timeframe stops routinely fall under it.
- **A tighter stop costs more, not less.** Halving your stop distance doubles what fees take as a share of your risk. This is the most commonly inverted idea in retail trading.

Neither predicts anything. Both change which trades are worth taking, which in practice matters more than most signals would even if they worked.

What this means for you

Every indicator on a chart is doing one of two jobs: **describing** what has happened, or **predicting** what will. Description is reliable — where price left in a hurry, where orders rest, how stretched things are, whether this session is busy. Prediction is where nearly every claim in this industry, including ours, falls apart under measurement.

These tools are good at description. Use them to build a short list of prices worth caring about, and to know what kind of market you are in when price arrives at one. Then make the decision yourself, with a stop wide enough that fees do not eat it.

THE ONE THING TO TAKE AWAY

Knowing which of your tools describe and which claim to predict is worth more than any indicator. Most traders never find out. You now have it written down for every feature on your chart, with the sample sizes attached.

APPENDIX

Every test, in full

The table you have just read is the summary. What follows is the complete record behind it — every test this system has been put through, what each one asked, how it was measured, and what it returned.

It is written for a trader rather than a statistician. Each test leads with its question in plain words, then how it was tested in a sentence, then the number.

Every result is drawn on the same scale: **how big the finding was next to the smallest finding that test could have detected**. That last quantity is called the floor, and it depends on how much data there was. A dot inside the shaded band means the test could not tell.

77	tests, across eleven families of tools
33	produced a result and a floor in the same units, so they can be plotted
29	of those 33 fall inside their own floor — too small for the test to see

Everything we tested, and what it actually showed

77 tests, in plain English. Most indicator documentation has no section like this. This one exists because the alternative — asking you to trust a tool because it looks like it works — is how people lose money slowly.

77	11	33	88%
tests run	families of tools	gave a measurable number	of those, too small to see

> **The headline, up front.** Nothing in this suite has been shown to predict direction. Not one feature — including the one we built a scorecard around, published, and then withdrew. What follows is the evidence, and then why these tools are still worth having on a chart.

How every test came out

14 no effect found · **23** claim did not hold · **14** we withdrew this · **8** could not tell · **7** describes, cannot predict · **7** works as claimed · **4** made it worse

How to read every chart here

No test can spot an effect of any size. Below a certain size, a result is indistinguishable from luck — and that size depends on how much data the test had. We call it the **floor**.

So rather than show you results in four different units, every chart here shows one thing: **how big the result was next to the smallest result that test could have spotted.** The shaded band is the floor. A dot inside it means the test could not tell. A dot outside it means the test found something real.

Every measurable test, on one scale

Sorted smallest to largest. The shaded band is each test's own floor.

Price levels and order blocks

Support lines, order blocks, breakers, untouched volume levels and Fibonacci retracements all make the same promise: that certain prices are special, and price behaves differently when it arrives there. Each one was tested against a fake line or zone placed the

same distance from price, so the only difference was whether the level was real. None of them beat the fake. The levels are a useful map of where to pay attention; not one of them has been shown to tell you which way price goes next.

Support and resistance levels

NO EFFECT FOUND

Does price respect a real swing high or low any more than it respects a random line the same distance away?

How it was tested. Every swing level was scored forward against a fake line pushed the same distance to either side, with that distance then mathematically extrapolated down to zero.

RESULT

`-0.03 ± 0.29pp on 1-hour, and -0.23 ± 0.21pp once the distance is extrapolated to zero; 24-minute agrees at -0.35 ± 0.35pp.`

← too small to see | big enough to trust → 0.04× the floor

Real levels scored 47.6 / 49.2 / 47.5% against fake lines at 47.6 / 48.5 / 47.0%. Three independent runs put the average outcome at -0.033R to -0.036R. Both check groups agree, four calibration checks pass, and the profile is flat rather than growing with distance. An earlier +1.5pp reading was withdrawn once the fake line was fixed to differ in one respect only.

HOW MUCH DATA

`19,688 separate 1-hour trades across eight check symbols, plus a separate 24-minute run`

SMALLEST EFFECT IT COULD SPOT

`0.0115 to 0.0162R on 1-hour, about 0.6 to 0.8 percentage points; 0.0194R on 24-minute - the best-resolved test in the project`

A real swing level is worth nothing over an arbitrary line the same distance from price, so use levels to decide where to pay attention, never as a reason to expect a bounce.

Order block retests

COULD NOT TELL

When price comes back into an order block, does it turn there more often than at any similar zone?

How it was tested. Every armed retest was scored against a matched fake zone - first shifted sideways, then, after that comparison proved rigged, matched instead on how deep the pullback was - and twenty combinations of the block settings were swept alongside.

RESULT

Baseline blocks scored 50.6 / 51.2 / 48.4%, giving +1.9 / -1.0 / -2.5pp against the fake zone.

← too small to see | big enough to trust →

0.80× the floor

The sideways-shifted comparison read +4.6 / +4.8pp and turned out to be an artefact: pullback depth on its own is worth 20pp, the fake running 41.2% in the shallowest tenth to 61.1% in the deepest while the real block sat flat near 55%. Matched properly on depth and timing: +1.6 / +1.2pp on 3-minute and -0.8 / -0.6pp on 1-hour, with a separate control saying the real block is 3.3pp worse than a matching non-swing zone. Not one of the 20 setting combinations cleared +2pp on both check symbols, and 0 of 12 quality groups cleared the smallest effect the test could spot. An archived 50.0% at 1,867 events was withdrawn as an overlapping-trades artefact, and the older "60% versus 33%" claim left nothing above about 6pp on retest.

HOW MUCH DATA

3 symbols on 1-hour and 3-minute, 40,000 bars per symbol; 20 declared setting combinations; 12 quality groups

SMALLEST EFFECT IT COULD SPOT

0.040R, about 2 percentage points - still roughly 1.7 times too coarse to settle it

A block retest looks like a coin flip and the test was not sharp enough to see anything smaller, so treat any future block claim as false until it has been matched on how deep the pullback was.

Twenty-five order block filters

CLAIM DID NOT HOLD

Can any filter sort the order blocks that work from the ones that do not?

How it was tested. About 25 filter variants were tried across two halves of recent data on one symbol, and the two survivors were then re-run on a genuinely different symbol they had never touched.

RESULT

```
In the build data both winners held in both halves: a minimum block height of 0.75 ATR
gave 53.3% against a 49.3% baseline, and a same-bar overshoot rule gave 62.1% at 6 ATR
and 68.0% at 7
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

ATR, rising cleanly with its own threshold. On the fresh symbol the baseline was 50.0%, minimum height 50.4%, overshoot-6 45.3% and overshoot-7 42.3%. The overshoot rule did not merely fade, it inverted. Separately, ranking blocks by score instead of filtering them gave -0.01 / -0.48 / +0.40pp, and a model using all nine candle features captured only 1.6 to 2.9% of the room available.

HOW MUCH DATA

`1,101 events in the build data, 1,117 on the fresh symbol`

A filter that survives both halves of one coin over one period, rises neatly with its own setting and comes with a good story can still be pure noise - it only counts once it survives a coin it has never seen.

Breaker retests and the volume filter

CLAIM DID NOT HOLD

When a broken block gets retested by a big-volume candle, does that retest actually work?

How it was tested. Every breaker retest was rebuilt offline across nine coins and three timeframes, then big-volume returns were compared with quiet ones four separate ways, including pairs taken inside the same zone.

RESULT

2x volume versus under-2x: -3.83pp on 1-hour, +4.31pp on 24-minute, -3.07pp on 3-minute. The shipped gate: -3.75 / +5.99 / -2.49pp. Top third against bottom third: -4.84 / +3.57 / -2.89pp.

← too small to see | big enough to trust →

0.91× the floor

Paired inside the same zone: -7.89 / -6.31 / -5.27pp. Not one of the 12 cells cleared its own floor, and the deepest samples lean the wrong way. On a separate run the filter read +1.15pp on the coin it was built on and inverted to -2.26pp and -0.91pp on two unseen coins. The original claim was 60% versus 33% - a 27-point gap.

HOW MUCH DATA

2,369 resolved retests on 1-hour, 1,153 on 24-minute, 11,473 on 3-minute; nine coins; 25,345 blocks became 7,162 breakers

SMALLEST EFFECT IT COULD SPOT

0.0676R, about 3.4 percentage points - 2.8 times the project's best-ever finding

The 27-point gap is not in the data, so treat a breaker as a place on the chart rather than a reason to take the trade - though waiting for the return does cut your risk from about 3 ATR to about 1.73 ATR.

Fibonacci retracement levels

NO EFFECT FOUND

Is there anything special about the 0.618, 0.5 or 0.786 retracement?

How it was tested. The same leg, same bar and same direction were scored with only the ratio changed, measured from the close of the touch bar and compared against a 62-rung ladder running from 0.30 to 0.95.

RESULT

`0.618 scored -0.14pp against the mean of the rungs either side, 0.500 read -0.08pp and 0.786 read -0.01pp.`

← too small to see | big enough to trust → 0.44× the floor

Against all 62 rungs the three fib ratios ranked 8th / 19th / 34th of 62 on 3-minute and 9th / 43rd / 49th on 1-hour. The whole triple against the same triple nudged 0.01 to 0.06 came to -0.000pp. The only real pattern is depth: +1.7 to +1.8pp at a retracement of 0.44 to 0.53, sliding to -0.77pp at 0.95 - a smooth curve the fib rungs sit on unremarkably. Hand-drawn fibs scored -1.60 and -0.23 against their fake levels, both worse than the fake.

HOW MUCH DATA

`nine coins on 3-minute and 1-hour, both check groups`

SMALLEST EFFECT IT COULD SPOT

`0.32pp (0.0064R) on 3-minute; 0.009 to 0.015R on 1-hour - three to five times finer than the project's best-ever positive result`

Price retraces but the golden ratio adds nothing, and if you ever backtest a fib with a resting order sitting at the level you will get a fake winner, because the fill itself only picks the moves that carried on.

Naked point of control as a magnet

WE WITHDREW THIS

Does price get pulled back to an untouched high-volume price more than to any other price?

How it was tested. Each untouched point of control was tracked over a 240-bar horizon against a fake level whose distance was drawn from a different event on the same side, so the two sat the same distance from price on average.

RESULT

`The first attempt read +17.3 / +15.2 / +18.1pp and was wrong, because its fake level was built from the day's open and sat a different distance from price.`

← too small to see | big enough to trust → 0.05× the floor

Corrected, the naked point of control was reached 88.4 / 88.1 / 87.5% of the time against the fake at 88.2 / 88.1 / 87.8% - a difference of +0.2 / +0.0 / -0.3pp, with the mean distances matching to two decimal places.

HOW MUCH DATA

`3 coins on 1-hour, 240-bar horizon of about ten days`

SMALLEST EFFECT IT COULD SPOT

`about 0.12R, roughly 6 percentage points - and it was the wrong measure anyway`

Price reaches a naked point of control 88% of the time inside ten days, and it reaches any old price at that distance 88% of the time too, so it is a price, not a magnet.

Order block zone width

DESCRIBES, CANNOT PREDICT

Does a wider order block make the trade harder to win?

How it was tested. Straight arithmetic - your risk is the zone width plus a 1 ATR buffer, and risk is what R is measured against - checked against how far price actually travelled after each retest.

RESULT

2R was reached in 9 of 22 cases at a 0.45 ATR zone (41%), 7 of 22 at 0.73 ATR (32%), 6 of 22 at 1.00 ATR (27%), 2 of 22 at 1.50 ATR (9%) and 1 of 22 at 2.01 ATR (5%).

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The median move after a retest is 2.34 ATR from the zone edge, 75th percentile 3.92 and 90th percentile 4.56. On 3 ATR of risk the median move is 0.78R - it does not even reach 1R. Drawing the zone to the wicks rather than the bodies changed the outcome in only 2 of 15 paired setups.

HOW MUCH DATA

22 retests; observed widths median 0.62 ATR, 90th percentile 1.16 ATR, maximum 1.60 ATR

SMALLEST EFFECT IT COULD SPOT

not applicable - this is arithmetic, not a measurement

A fat zone demands a bigger move for the same reward and no sample size can argue with that, so past roughly 1 ATR of width the maths stops closing whatever the zone is called - worry about width, not about wick versus body.

Momentum tools

Momentum tools — RSI bands, the 9/21 moving-average cross, divergences, and the buy/sell dots — all claim to tell you when a move is running out or a new one is starting. Tested across tens of thousands of signals on nine coins, every one of them landed on a coin flip, and the yellow dot typically prints about 19 bars before the turn it is supposed to

mark. The only reading that moved off zero moved the wrong way: taking raw wave crossovers loses a small, consistent amount on every trade.

The yellow buy and sell dots

CLAIM DID NOT HOLD

When the yellow dot prints, does price go my way?

How it was tested. Every dot was scored on whether price travelled one average bar's range in its favour before travelling the same distance against it, on three coins and five chart speeds, and was also compared against simply noticing the wave was already deep.

RESULT

50.0% — exactly chance, and flat at every speed: 3-minute 49.43%, 15-minute 50.10%, 30-minute 49.61%, 1-hour 49.38%, 4-hour 51.99%.

← too small to see | big enough to trust → 0.00× the floor

Median 19 bars before the real turn, with price running a median 3.0 ATR against you first. As a top/bottom finder a random bar marks a pivot 2.9% of the time, the dot 6.4%, 'the wave is simply deep' 9.8%, and 'deep but no dot' 10.5% — the dot loses to plain depth in 29 of 30 cells.

HOW MUCH DATA

22,656 dots on ETH, BTC and SOL across five timeframes; 22,500 scored for direction; 18,000 oversold episodes

SMALLEST EFFECT IT COULD SPOT

0.021R on the race; per-cell error bars of 0.72 to 2.35 percentage points

The dot only tells you momentum has crossed; it is a coin flip that usually fires around 19 bars too early, so never treat it as an entry.

Every attempt to fix the dot

CLAIM DID NOT HOLD

Is there a filter, a slower chart, or a waiting rule that turns the dot into something worth trading?

How it was tested. About 1,100 rule variants across six angles were screened, the seven that passed were handed to an independent re-tester who ran the control the first screen had skipped, and separately every fast-chart-to-slow-chart pairing, every waiting length from 1 to 21 bars, and resting limit orders were tested against matched controls.

RESULT

All seven died. The closest survivor was +0.38pp against a ±0.90pp error bar, and it was the best of 94 tried.

← too small to see | big enough to trust → 0.42× the floor

On data it had never seen, the volume filter reversed: the quietest 2% of dots won 56.46% against the loudest 2% at 52.64%. Waiting failed at every length from 1 to 21 bars; limit orders lifted the win rate to 51-53%, but the same order placed from a random bar did 50.5-52.1%. Lift found on the tuning coin carried over to fresh coins at a slope of 0.20 — about 80% of it was noise.

HOW MUCH DATA

~1,100 variants screened and 7 re-tested, over 22,656 dots; 9 fast-to-slow chart pairs x 16 agreement rules across 174 comparison cells

SMALLEST EFFECT IT COULD SPOT

±0.90 percentage points on the best surviving filter

Stop hunting for the setting that fixes the dot — filtering it, confirming it on a slower chart, and waiting for a better price have all been tried and none of them work.

The raw wave crossover as an entry

MADE IT WORSE

If I simply buy when the fast wave crosses up and sell when it crosses down, what does that cost me?

How it was tested. Every crossover was scored offline on nine major coins against a control matched on six things, so the comparison bars were as alike as possible apart from the crossover itself.

RESULT

-0.0240R per trade, 3.07 standard errors from zero, against a smallest spottable effect of 0.0219R. The signal lands inside a matching ribbon only 27.3% of the time.

← too small to see | big enough to trust → 1.10× the floor

HOW MUCH DATA

Nine major coins; the signal fires about 13 times a day on the 15-minute chart

SMALLEST EFFECT IT COULD SPOT

0.0219R

This one is not a nothing — trading raw wave crossovers loses a small, steady amount every time, before you even pay fees.

The 9/21 EMA cross

NO EFFECT FOUND

Does the 9 crossing the 21 tell me which way price is going?

How it was tested. The forward outcome after every cross was measured on nine coins and three chart speeds, then 25 other fast/slow pairs chosen in advance were scored on the same machinery.

RESULT

50.2%, 50.3% and 50.7% win rate on ETH, BTC and SOL. On the hour -0.04 to -0.09pp against a floor of about 0.024R;

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

on the 3-minute -0.55pp, negative on 9 of 9 coins, against a floor of about 0.012R. Across the 25-pair grid the entire spread is smaller than one cell's own measuring error, and 9/21 ranks 10th, 6th, 20th and 21st of 25. The earlier evidence for 9/21 was withdrawn — its best figure, 50.24%, sits inside its own 0.44 unit of noise.

HOW MUCH DATA

Nine coins on 1-hour, 24-minute and 3-minute, including 313,133 events on the 3-minute; ETH, BTC and SOL re-checked separately

SMALLEST EFFECT IT COULD SPOT

about 0.024R on the hour and 0.012R on the 3-minute; 0.37 to 0.94 percentage points per archived cell

The cross is a coin flip on every coin tested and 9/21 is in no way a special pair, so keep it as a picture of trend, never as a reason to click.

Waiting for a cross before entering

CLAIM DID NOT HOLD

At a level I already like, am I better off waiting for the moving averages to cross before I get in?

How it was tested. The same level, stop and target were traded with only the trigger changed, comparing entering straight away against ten different crossover pairs.

RESULT

Entering immediately: 53% win, +0.282R. Fast pairs (2/6, 3/9, 2/8, 3/8, 3/6, 3/7, 3/10): 36-46% and +0.05 to +0.13R. 4/8: 36% and +0.008R.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Slow pairs including 9/21: 27-42% and -0.11 to -0.28R. Every crossover lost to entering immediately, and the ranking tracks how slow the pair is.

HOW MUCH DATA

39 pivots on the 3-minute over 365 bars; 38-39 trades per condition

SMALLEST EFFECT IT COULD SPOT

not stated; the spread among neighbouring fast pairs is itself the noise band, and each cell holds only 38-39 trades

Your stop does not move while you wait, so every bar you spend waiting for confirmation buys a worse price for the same risk — thin evidence at 38-39 trades a cell, but it all points one way.

Divergences

NO EFFECT FOUND

When momentum makes a higher low while price makes a lower low, does the turn actually come?

How it was tested. Five ways of spotting a divergence were measured on one coin and re-checked on two coins that had not been used to pick them, then sliced by trend, depth and confirmation, and finally used as a filter on dot entries against the control of taking every dot.

RESULT

On data it had never seen the five detection rules scored 50.31%, 49.81%, 50.18%, 50.40% and 50.37%.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Slices: against the trend 50.3/50.6%, with the trend 52.7% where it was fitted but 49.9% on the check, deep turn 50.1/50.7%, confirmed 50.6/50.5%. Used as a filter on dots it made things worse — every dot with no filter 40% and +0.002R, classic divergence 38% and -0.065R, trend-aligned hidden divergence 34% and -0.237R.

HOW MUCH DATA

Cells from 480 to 7,862 trades on ETH, BTC and SOL from 15-minute to 4-hour; 820-910 divergences per coin; the filter arm had 32 classic and 29-37 hidden against a control of 269

Divergences describe momentum fading, not a turn arriving — and filtering other signals with them made those signals worse, not better.

RSI bands

COULD NOT TELL

Does an overbought or oversold RSI reading tell me where price goes next?

How it was tested. The bands were first set at 40 and 60, where the top and bottom fifth of readings actually fall on this market over about eleven months rather than the textbook 70/30, and then the move 12 bars later was measured from each band in units of a typical bar's range, counting each move only once.

RESULT

1-hour: contracted +0.079 ATR, neutral -0.104 ATR, extended +0.158 ATR; 24-minute neutral +0.185 ATR. Nothing bigger than about 0.4 ATR anywhere.

← too small to see | big enough to trust →

0.49× the floor

On calibration, median RSI is 49.7 (1H), 50.3 (24M) and 49.7 (4H), and cut points of 40 and 60 give a 21.2 / 58.7 / 20.1 split on the hour, where the old 55 line had been calling a third of all bars 'extended'.

HOW MUCH DATA

About 8,000 bars on 1-hour and 24-minute; 124 contracted, 338 neutral and 203 extended hourly readings, plus 314 neutral on the 24-minute

SMALLEST EFFECT IT COULD SPOT

0.375 to 0.797 ATR depending on the band

Read RSI as how far a move has already stretched, not as where it goes next — and keep the lines at 40 and 60, because at 55 a third of all bars looked 'extended'.

Volume and money flow

Volume tools promise to tell you whether a move has real money behind it: a heavy bar is supposed to confirm a break, and a money-flow reading is supposed to show buying and selling pressure. Testing found no version of that idea that predicts direction — our own money flow, a leading paid indicator's money flow, and the volume filters tried on sweeps all came out at zero or could not be resolved at all. Worse, the flagship rule of the whole

framework, the volume tier ladder, turns out to need a number that does not exist yet at the moment you would trade it.

The volume tier ladder

COULD NOT TELL

When a big-volume bar fires on the fast chart and the hour above it is also big, does price run further?

How it was tested. Every 24-minute bar trading at least twice its 30-bar average volume was followed for the next 24 bars, and the ones whose last completed hourly bar was also 2x were compared against the rest, using events that shared no bars and rules written down before any number was looked at.

RESULT

Confirmed 17 events at -1.033 ATR vs partial 220 events at -0.178 ATR, a gap of 0.855 ATR, against a smallest detectable effect of 2.449 ATR.

← too small to see | big enough to trust → 0.35× the floor

The peeking version (82 vs 155 events) gave 0.497 ATR against a floor of 3.022 ATR.
Control of every bar: -0.032 ATR.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
8,000 x 24M bars and 8,000 x 1H bars of ETHUSDC.P; 237 qualifying events; 2,494-bar control	2.449 ATR (peeking version 3.022 ATR)

The most trusted rule in the framework has never been shown to work, and the usable version fired 17 times in 133 days, so settling it on one coin would take decades.

The confirming hour arrives too late

CLAIM DID NOT HOLD

The ladder says the hour above must also be twice normal volume, so can you actually read that number when the signal fires?

How it was tested. Worked out how the bars nest inside each other, then measured what share of the confirming hour's volume was the signal bar's own volume.

RESULT

No. The 1-hour bar contains the 24-minute bar that fires, so at the moment of entry that hour is still forming.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

On the two events measured the signal bar was 53% and 54% of the whole hour's volume.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
arithmetic on bar nesting, plus the two largest events of 2026-08-28	not applicable — arithmetic, not a measurement

Over half of the 'confirmation' is the same trades counted twice, so the ladder has never been tested the way it is actually used, and the unbroken 5-for-5 live record is anecdote against 237 measured events that show nothing.

How rare is high volume, really

DESCRIBES, CANNOT PREDICT

Is a bar at twice its average volume actually a rare event worth reacting to?

How it was tested. Divided every bar's volume by its own 20-bar average and counted how often each level came up, with no other conditions attached.

RESULT

`1.0x on 36.19% of bars, 1.3x on 22.40%, 1.5x on 16.71%, 1.75x on 11.81%, 2.0x on 8.49% (1 bar in 12), 2.5x on 4.68% (1 in 21), 3.0x on 2.66%, 4.0x on 1.01%.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
every bar on 9 crypto majors, 1-hour, after a 300-bar warm-up	no floor applies — this is a count, not a test

Twice-average volume turns up roughly once every twelve bars, so a big volume number by itself does not make a setup rare or special — whatever rarity a setup has comes from where the bar sits, not how loud it is.

Demanding big volume on the sweep

CLAIM DID NOT HOLD

Does only taking sweeps that arrive with heavy volume make them better trades?

How it was tested. Scored swing-failure entries with and without a rule that the sweep candle trade at least twice normal volume, then stacked that filter on top of the existing entry rule, with a live forward log checked the same way.

RESULT

`Filtered +0.121R against +0.086R unfiltered — a gap of 0.035R against a smallest detectable effect of 0.041R.`

← too small to see | big enough to trust → 0.85× the floor

Stacked on the entry rule it made things worse: +0.131R against +0.227R for the entry rule alone. Median volume on sweeps was already 2.07x normal.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
the offline swing-failure sample, plus 21 live sweeps logged forward	0.041R

A sweep is a heavy-volume event by definition, so calling one unconfirmed for low volume is not supported — the best sweep in the live log (6.17 ATR) fired on 0.41x volume, the quietest of all 21.

Our money flow as a signal

WE WITHDREW THIS

Does our money-flow reading tell you anything about where price goes next?

How it was tested. Bars were sorted into fifths by their money-flow reading across three coins and four timeframes, the top-minus-bottom gap was measured, and the whole thing was rechecked on coins the test had never seen.

RESULT

The original -6.5pp contrarian lead reversed to +0.62pp on the fitting coin and +0.86pp on the unseen coins — the opposite sign, and both effectively zero.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
453,176 resolved bars (the original claim came from 6,344)	not stated; the run-to-run wobble was 0.20-0.29pp per cell

The eye-catching lead was a two-month accident on one coin; money flow has no shown ability to predict direction, and agreement across horizons on the same bars proves nothing — only a fresh coin does.

A leading paid indicator's money flow

NO EFFECT FOUND

Does the well-known paid indicator's money flow do any better than ours?

How it was tested. Pulled 6,344 bars of the paid indicator's own plotted output straight off the chart, sorted each reading into fifths and measured the top-minus-bottom gap at three forward horizons.

RESULT

Money flow -0.66 / +0.13 / +0.30pp across the three horizons against a 1.44pp per-cell floor. Its RSI -1.25 / -1.63 / -1.65pp, its Stochastic RSI -2.99 / -1.71 / -1.65pp.

← too small to see | big enough to trust →

0.46× the floor

Its wave matched ours at a correlation of 1.0000.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
6,344 bars of 15-minute ETHUSDC.P	1.44pp per cell

The paid tool has no mechanical edge over ours, and its popular free clone reads a different quantity entirely (-6.35 against -1.11 on the same bar), so never quote the clone as if it were the paid number.

What the money flow cloud measures

CLAIM DID NOT HOLD

Is the green-and-red money flow cloud really weighted by volume, and does the darker inner band add anything?

How it was tested. Read the shipped formula line by line, checked a copy of it against the live chart bar for bar, then sorted every bar into blank, an exact copy of the outer cloud, or something new.

RESULT

```
The outer cloud has no volume term at all — it is candle body divided by candle range,
averaged over 60 bars.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The inner pressure band is blank on 36.8% of 1-hour bars and an exact copy on 19.2%, leaving 44.0% that says anything new (15-minute: 36.9 / 17.3 / 45.8). The code matches its own formula to 5.2e-14.

HOW MUCH DATA

`298,200 bars across 3 coins and 2 timeframes, plus 201 bars checked against the code`

The name is inherited, not earned: the cloud is a slow read of candle bodies, and on about 55% of bars the volume band is blank or just repeating the cloud — with blank meaning both 'neutral volume' and 'volume disagrees'.

Liquidity and sweeps

The idea is that price runs to where stop orders are resting, takes them out, then reverses — so a wick that pokes through a level and closes back inside should be a reversal worth trading, and a fat pool of orders should pull price toward it. Every tradeable version of that failed testing. The two biggest tests failed the same way: entering after a sweep loses about 3.6% of the money you put at risk on every trade, and it does that while winning 69% of the time. The one genuinely useful thing that survived is a measurement rather than a

signal — sweeps only poke about a third of an ATR past the level, so a stop about 1 ATR beyond the wick survives roughly nine out of ten of them.

Entering after the sweep

MADE IT WORSE

If I wait for price to poke through a level and close back inside, then enter with my stop tucked behind the wick, does it make money?

How it was tested. Every sweep was traded exactly as specified — enter at the close of the sweep bar, stop 1 ATR beyond the wick, risk capped at 2 ATR, target the nearest untaken level, real exchange fees charged — and scored against entering at a random bar with the identical stop and target.

RESULT

`Sweep entry -0.0362R per trade at a 68.7% win rate, against -0.0325R for the random entry with identical geometry, so the sweep is worse than random.`

← too small to see | big enough to trust → 3.62× the floor · beyond the scale

Moving the target does not rescue it: nearest level -0.0362R, 1.0R target -0.0725R, 1.5R -0.0807R, 2.0R -0.0856R, 3.0R -0.0857R, and the sweep is worse than random at every one. Average win 0.155R against a full 1R loss.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
18,243 trades on six fresh symbols, 1-hour	0.010R — the sharpest test in the whole project

You can win 69% of the time and still lose money, because the wins are small and the losses are full size — and no choice of target fixes it.

The swing failure pattern itself

MADE IT WORSE

Does a wick that takes out a high or low and closes back inside tell you which way price goes next?

How it was tested. Fading the sweep was scored as a race from the sweep bar's close — does price travel 1 ATR the reversal way before it travels 1 ATR against — on nine symbols, with the long/short mix matched, checked in all four calendar quarters and on data it had never seen.

RESULT `49.6% on the data used to build it and 49.3% on data it had never seen.`

← too small to see | big enough to trust → 1.50× the floor

Re-run against a matched comparison group it comes out at -1.64pp, which is -0.033R in the reversal direction, and that repeats on both check groups, in all four quarters, and at every sweep size from 0.5 to 2.0. Every slice leaves it there: wick size 49.1-50.3%, swing age 45.3-49.8%, direction 48.8-50.1%.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
31,710 sweeps, nine symbols, four timeframes	0.022R

Traded mechanically as a reversal the pattern is not even a coin flip — it leans slightly against you, and the lean is smaller than what the fees cost, so it is information about location, not money.

Does a better-looking sweep trade better?

CLAIM DID NOT HOLD

Is a sweep with a long wick and a decisive close back inside a better trade than a scrappy one?

How it was tested. 73 sweeps were split into thirds by how deeply price closed back inside, by wick length and by swing size, with the grading rule written down in advance, and each third was compared against simply taking every sweep.

RESULT

`Baseline of all sweeps: 42% win, +0.071R. The 'strong reclaim' third — the deep, decisive close back inside — scored 24% win and -0.292R. The weak reclaim third scored 44% and +0.144R.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Strong reclaim was negative on all three timeframes: -0.02 / -0.30 / -0.84R. Graded by swing size instead: strong -0.095R, weak -0.179R, both worse than doing nothing.

HOW MUCH DATA

`73 sweeps across 30-minute, 1-hour and 4-hour`

The most impressive-looking sweep is the worse trade, because a big reclaim means the move already happened and you are entering late.

Big volume on the sweep candle

CLAIM DID NOT HOLD

Does insisting on heavy volume in the sweep candle pick out the better sweeps?

How it was tested. Sweeps were scored with and without a rule demanding at least twice normal volume on the sweep candle, that filter was then stacked on top of the entry rule, and the same idea was recorded live on every sweep as it formed.

RESULT

Filtered scored +0.121R against +0.086R unfiltered — a gap of 0.035R, smaller than the 0.041R the test could see.

← too small to see | big enough to trust →

0.85× the floor

Stacked on the entry rule it made things worse: +0.131R against +0.227R for the entry rule alone. Median volume on a sweep is already 2.07x normal, so the filter screens out almost nothing. In the live log the best sweep recorded (best move 6.17 ATR) fired on 0.41x volume, the lowest reading of any logged sweep. The opposite claim — that heavy-volume sweeps are worse, at 47.5% and 48.5% — did not reproduce when re-tested properly.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
the 73-sweep offline set, plus 21 independent sweeps logged live	0.041R

A sweep is a high-volume event by definition, so calling one 'unconfirmed' because volume looks low has nothing behind it.

Do bigger liquidity pools pull harder?

CLAIM DID NOT HOLD

The indicator scores every untaken high and low by how much volume sits behind it — does a big score mean price gets there faster, or that the level holds when price arrives?

How it was tested. Every scored level was tracked for how long price took to reach it, how often price closed straight through it within 20 bars, and how a fade of the sweep did, with each level's distance at birth held constant so a nearer level cannot look magnetic for the wrong reason.

RESULT

It runs backwards. From 1-2 ATR away, median bars to first touch by score band was 9 / 10 / 13 / 19; from 2-4 ATR, 22 / 24 / 30 / 42;

← too small to see | big enough to trust → 0.78× the floor

from 4-plus ATR, 43 / 60 / 52 / 77 — a high-scoring level takes about twice as long to be reached, and the check symbols agree almost exactly. Once touched, price closed through within 20 bars 86.2 / 84.7 / 83.4 / 84.8% of the time, flat at every score, and fat levels produced smaller pullbacks first, not bigger (2.39 / 2.18 / 2.06 / 1.83 ATR). About 99% of levels get swept eventually whatever their size, and fading the sweep wins 48-49% across every pool size.

HOW MUCH DATA

34,951 levels price later reached; 43,076 levels scored; about 27,000 sweeps across three symbols and four timeframes

SMALLEST EFFECT IT COULD SPOT

1.16 percentage points on the fade comparison

A big score is not a stronger level, it is a slower target — use it to see where price might eventually go, never to decide whether a level will hold.

The untouched volume shelf as a magnet

WE WITHDREW THIS

Does price get dragged back to an untouched high-volume price more than to any other price the same distance away?

How it was tested. How often price reached the untouched shelf was compared against how often it reached a fake target drawn the same distance away on the same side, after a first attempt that used a fake target sitting at a different distance had to be thrown out.

RESULT **The first reading of +17.3 / +15.2 / +18.1pp was wrong.**

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Corrected: the shelf was reached 88.4 / 88.1 / 87.5% of the time against 88.2 / 88.1 / 87.8% for the matched fake target — a difference of +0.2 / +0.0 / -0.3pp, with the two average distances matching to two decimal places.

HOW MUCH DATA

3 symbols on 1-hour, 240-bar window (about ten days)

Price reaches an untouched volume shelf 88% of the time within ten days — and it reaches any price at that distance 88% of the time. It is a price, not a magnet.

How far a sweep runs past a level

DESCRIBES, CANNOT PREDICT

How much room does my stop need to survive a stop hunt?

How it was tested. For every sweep, the distance the wick travelled past the level before price turned back was measured as a multiple of that chart's average range, on four timeframes and re-run on a second independent sample.

RESULT

`Median depth 0.26 / 0.38 / 0.30 / 0.29 times ATR. 90th percentile 0.93 / 0.95 / 0.90 / 0.71. Largest ever seen 1.65 / 1.33 / 1.22 / 0.89.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The old claim that 1.5x ATR clears every sweep is dead: the largest observed sweep grew every time a sample was added, 0.89 to 1.22 to 1.33 to 1.65, while the median and 90th percentile barely moved. In dollars the same medians read $2.98 on 30-minute but $9.13 on 4-hour, which is why the number must be kept in ATR.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
`85 sweeps across 15-minute, 30-minute, 1-hour and 4-hour, spans of 3.1 to 83.2 days`	`not applicable — a direct measurement`

Stop hunts are shallow: about 1 ATR beyond the swept wick clears roughly nine in ten of them, so surviving one is cheap — but this says how far price pokes, never which way it goes next.

Market structure and phases

This family covers the tool that splits the chart into trends and ranges, then guesses whether a range is accumulation (buyers loading up, break likely up) or distribution (sellers unloading, break likely down). Testing found it is an honest description of what price has already done — it never rewrites history to look right — but it does not forecast. The label is correct about half the time whatever settings you use, its volume ingredient turned out to

be roughly 71% a disguised count of up-closes, and the one reading that looked like it called breakouts was simply watching the break happen.

How often the phase label is right

NO EFFECT FOUND

When the chart says a range is accumulation or distribution, how often does it call the eventual break correctly?

How it was tested. Scored each range's first label against the side price actually broke out of, then repeated the scoring at every conviction level from 1 to 7 to see if being pickier helps.

RESULT

First-call accuracy 49.41% / 51.93% / 49.93%; never leaves 48-53% at any conviction level from 1 to 7. Coverage 100%, and the label appears a median of 1 bar into the range.

← too small to see | big enough to trust →

0.41× the floor

HOW MUCH DATA

4,894 range phases across nine major coins

SMALLEST EFFECT IT COULD SPOT

4.72pp on the sealed-range version, 2.52pp on the faster version

It is right about half the time no matter how picky you are, so read it as a description of where price has been, not as a call on where it goes next.

Does the accumulation lean call the break

CLAIM DID NOT HOLD

Does the accumulation-versus-distribution reading tell you which way a range will break before it breaks?

How it was tested. Used only ranges that actually broke, froze the reading at the last bar before the break and then further and further back in time, and scored each freeze against how often that side broke anyway.

RESULT

Looked like +5.5 / +9.3 / +10.5pp. Frozen at the last bar it reads +23.4 / +27.5 / +31.0pp; three bars back +9.7 / +11.9 / +17.4pp; ten bars back -0.6 / +7.3 / +6.2pp;

← too small to see | big enough to trust →

0.12× the floor

twenty bars back -3.3 / +0.2 / +0.8pp. A plain count of up bars against down bars scores the same (78.3 vs 77.7).

HOW MUCH DATA

about 500 ranges per coin on three coins

SMALLEST EFFECT IT COULD SPOT

0.130R, and 0.16-0.20R for the rows the verdict rested on (roughly 6.5pp)

The reading was watching the break form, not forecasting it — any number that shrinks the earlier you take the snapshot was never a forecast.

Is there any volume in the volume term

NO EFFECT FOUND

Does the volume part of the phase engine actually know anything about volume, or is it just counting up-closes?

How it was tested. Replaced volume with a flat 1.0 on every bar and counted how often the vote changed, then ran a version with volume shuffled against price as a fake check, keeping the phases, ranges and every gate identical.

RESULT

With volume replaced by 1.0 the engine casts the SAME vote on 70.9% of ranges (67.0-74.8% across the nine coins). Real minus fake volume: +2.40 / +0.22 / +1.55pp.

← too small to see | big enough to trust → 0.98× the floor

HOW MUCH DATA

nine major coins, 40,000 1-hour bars each (2022-01 to 2026-08); 4,894 range phases, 4,773 resolved, 167,623 bar readings

SMALLEST EFFECT IT COULD SPOT

2.46-5.45pp (about 0.050-0.110R)

About 71% of the volume term is a disguised up-close count, and taking volume out altogether leaves how often the label is right unchanged (+0.13 / +0.06pp).

The volume term's hidden short lean

NO EFFECT FOUND

The volume part of the engine votes short more often than long by accident – does correcting that make it better at calling direction?

How it was tested. Measured the built-in lean, re-centred it, and re-ran four scoring methods across three splits of the data, two settings and two read points, with a deliberately fake arm run alongside as a check.

RESULT

The lean is real: the shipped version votes 9.6pp net short (12.3pp on the older version). Re-centring removes it, moving the vote to +0.3 to +1.7pp.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Direction-calling ability does not move — every cell sits inside its own floor, and the fake arm threw up numbers the same size as the real ones.

HOW MUCH DATA

four scoring methods x three data splits x two settings x two read points, plus a fake arm; repeated on 24-minute bars

SMALLEST EFFECT IT COULD SPOT

not stated as one number; every cell was inside its own

Changing how often a tool leans long or short moves the scoreboard without changing whether it knows anything — the third time in this project a result turned out to be that in disguise.

Is the upthrust term's sign backwards

CLAIM DID NOT HOLD

A failed push above the top of a range was flagged as being counted the wrong way round – does that hold up?

How it was tested. Reproduced the original number exactly, re-scored the term against a benchmark that leans the same way it does, compared it against the mirror-image spring term, and tried flipping the sign.

RESULT

Reproduces exactly (43.25 / 46.44 / 48.09 against the flagged 43.4 / 46.6 / 48.7). Scored fairly it reads -1.60 / -1.36 / -2.02pp. Upthrust minus spring is +0.31pp plus or minus 2.05.

← too small to see | big enough to trust →

0.47× the floor

Flipping the sign gives +0.18pp and destroys 86% of the distribution calls (2,354 down to 330).

HOW MUCH DATA

nine coins, three data splits; the original flag came from about 412 pooled ranges on three coins

SMALLEST EFFECT IT COULD SPOT

4.3-5.0pp; 4.33pp for the flip

No code bug here — the alarm came from comparing the term against the wrong benchmark, and flipping it would silence most of the labels for nothing in return.

Adding textbook Wyckoff shapes to the label

MADE IT WORSE

If proper Wyckoff ideas are added – effort versus result, climax, dry-up, shakeout – does the phase label get better?

How it was tested. Built four textbook constructs with thresholds fixed in advance and no tuning, then measured both how quickly the label arrives and how often it is right.

RESULT

All four make the call LATER by 0.71 to 0.98 bars, and none is more accurate (could not tell at 3.2-3.8pp).

← too small to see | big enough to trust →

2.51× the floor

Deleting the volume term and adding nothing does the same damage (-2.99 / -3.15) as the constructs (-2.72 / -4.82).

HOW MUCH DATA

same nine coins, 40,000 bars each, 4,894 range phases

SMALLEST EFFECT IT COULD SPOT

0.11-0.39 bars on speed; 3.3-5.0pp (0.067-0.100R) on accuracy

The textbook shapes make the label arrive later without making it any more accurate — the delay is big enough to measure, the improvement is not.

Does the phase engine repaint history

WORKS AS CLAIMED

Does the phase drawing only name a range after price has genuinely left it, or does it quietly rewrite the past so it looks like it knew?

How it was tested. Replayed the rules in Python from the chart's own bars and compared them against what the indicator actually drew, across the whole timeframe stack, plus a side-by-side check of the offline copy against the live chart.

RESULT

554 of 554 sealed ranges were named only after a real close outside a bracket that was already on screen.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Price alignment error 0.0000 on every timeframe, and phases tile the chart with no gaps or overlaps. The offline copy matched the live logic on 40,000 of 40,000 bars per coin with 0 mismatches, and 5 of 5 phase names matched the live chart in order.

HOW MUCH DATA

about 39,000 bars (4H 3,114 / 1H 12,453 / 24M 10,515 / 3M 13,087), plus 40,000 bars each on three coins for the copy check

SMALLEST EFFECT IT COULD SPOT

not stated — a correctness check, not an edge test

What you saw live is what you see later, which is more than most phase indicators can say — but drawing honestly is not the same as predicting anything.

Volatility bands

These are the bands drawn around a moving average — the Achilles Bands on the chart, plus the standard Bollinger and Keltner versions of the same idea. The folklore is that a close outside the edge is a snap-back trade, and that the bands pinching together warns of a move coming. Across roughly 24,000 hourly trades on nine coins and five calendar years, none of that survived: fading the band does lose consistently, but the band edge is not

what causes it, the leftover is smaller than the round-trip fee, and switching band type, changing the settings or waiting for a squeeze all measured as nothing.

Fading a close outside the band

CLAIM DID NOT HOLD

When price closes outside the band, should I bet on a snap back?

How it was tested. Every hourly close outside the band was traded in the fade direction, each trade counted once, against a control matched on the long/short mix and against a scrambled version of the tape, with all 88 cells written down before any number was computed.

RESULT `Fading loses 1.33pp, on 9 of 9 symbols and all 5 calendar years.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

But the band threshold itself adds only +0.72pp ± 0.38 over the same rule on ordinary bars, and only +0.13pp on the second check group. The response is not orderly: the moderate bucket reads higher than the more extreme one. It is +1.59pp before Oct 2025 and +0.19pp after. 24-minute reads +0.29pp and 3-minute reverses sign. Gross +0.027R against 0.081R of round-trip fees.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
23,886 trades, nine symbols, five calendar years, 1-hour	not stated as one number

Do not fade a close outside the band on the hourly — but the work is being done by the 20-bar average, not the band, and what is left costs three times more to trade than it pays.

Bollinger versus Keltner, head to head

NO EFFECT FOUND

Would switching to the other common band type be better?

How it was tested. Two teams ran it blind to each other, each matching how often the two band types fire so both were scored on the same number of trades, with no width setting ever tuned against results.

RESULT

`Bollinger +1.333pp, Keltner +0.911pp — a gap of +0.422pp, which is +0.0084R, with a range of -0.008R to +0.027R, against a 1.27pp floor.`

← too small to see | big enough to trust → 0.33× the floor

The fitting symbol read -1.74pp, the opposite sign, favouring Keltner. 14 of 15 cells sat below the floor and none cleared on both check groups with the same sign.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
23,886 Bollinger trades and 18,714 Keltner trades on 1-hour; 48 cells declared for one design, 18 for the other	1.27pp; the finest either design could resolve was 0.025R on 1-hour and 0.0145R on 3-minute

Keep the band type you already have — not because it won, but because the gap between them is three times smaller than the smallest effect the test could spot.

Middle line or band width

NO EFFECT FOUND

If the two band types differ at all, is it the centre average or the way the width is measured?

How it was tested. One change at a time — swap only the centre average and hold the width measure fixed, then swap only the width measure and hold the centre average fixed.

RESULT

Changing only the centre average gives +0.057pp ± 0.262 against a 0.73pp floor — the tightest nothing in the whole study.

← too small to see | big enough to trust → 0.42× the floor

Changing only the width measure gives +0.579pp ± 0.498 against a 1.39pp floor, and the fitting symbol flips sign at -2.22pp.

HOW MUCH DATA

1-hour, nine symbols; trade counts matched to within 0.6% on the centre-line swap

SMALLEST EFFECT IT COULD SPOT

0.73pp for the centre average, 1.39pp for the width measure

The centre average is not where anything lives, so arguing over a 20-bar simple versus a 21 exponential is arguing over nothing.

The band settings sweep

NO EFFECT FOUND

Would a different length or width setting work better than the shipped one?

How it was tested. All 20 combinations of two lengths by two width multipliers by five average types, every cell written down before running.

RESULT 19 of 20 combinations favour price continuing rather than snapping back.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The shipped 20-bar / 2.0-width / simple-average setting sits mid-pack, and the single cell that leaned the other way on the fitting symbol reversed on both check groups.

HOW MUCH DATA

20 declared cells, nine symbols, 1-hour

Leave the length, the multiplier and the average type exactly where they are — the whole grid is one flat field.

The squeeze

CLAIM DID NOT HOLD

When the bands pinch together, does a real move follow in the direction they point?

How it was tested. Five separate rule families were fitted on one coin and then scored by different people on two coins the fitter never saw, after which the squeeze machinery was ripped out and replaced with "every 16 bars, just bet the direction rule".

RESULT

54.9% right where it was fitted, 50.28% on the two untouched coins. Zero of ten unseen cells beat the 52.5% bar set in advance, and two came out backwards.

← too small to see | big enough to trust →

0.11× the floor

Stripping the squeeze out entirely and using a plain timer scored 50.30% against the full rule's 51.90%. The detector fires on 8.6-9.1% of bars in all nine cells, near-identical across three coins and three timeframes.

HOW MUCH DATA

2,534 fitted events and 5,075 scored on unseen data; 19 agents

SMALLEST EFFECT IT COULD SPOT

52.5% was the bar set in advance — 2.5pp above a coin flip

The squeeze is an alarm clock, not a signal — a timer that ignores the bands completely does the same job.

Never fade the band in a trend

WE WITHDREW THIS

Is the standing rule "only fade the band in a range, never in a trend" actually supported?

How it was tested. The rule was traced back to the study it came from and re-run against the bands that are actually on the chart.

RESULT

The rule was measured on a different band type — a 20-bar exponential average with a 2x average-range width — not the bands on the chart, and it does not carry over.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Its own numbers were inconsistent anyway: top of band while trending read +0.44 on 30-minute and +1.46 on 1-hour, while ranging read -1.20 and -0.19, and the plain top-of-band reading flipped sign across timeframes at +0.05, +1.18 and -0.63 average-range units.

HOW MUCH DATA

the original study's key ranging buckets held 13 and 19 observations, with samples reused, no control and no second coin

SMALLEST EFFECT IT COULD SPOT

not stated; ranging buckets of 13 and 19

This rule was being issued live every hour for weeks and was measured on the wrong indicator — it is gone, and nothing replaced it.

The biggest number favouring Keltner

CLAIM DID NOT HOLD

One reading in the band study was enormous. Was it real?

How it was tested. The single largest cell — the most extreme 1% of 24-minute bars — was checked against the matching 1-hour cell, split by era, and cross-checked against 3-minute data covering the same days.

RESULT

```
-10.16pp favouring Keltner on 24-minute, against +3.54pp favouring Bollinger in the
matching 1-hour cell — the exact opposite.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

It collapses on an era split, -15.7 in the first half and -3.9 in the second, and is invisible in 3-minute data covering the same days. The second design found the mirror image, its two clearing cells pointing in opposite directions.

HOW MUCH DATA

```
15 declared readings; 3-minute cross-check covering the same 312 days
```

When you look at fifteen readings, the biggest one is usually the noise — which is why the bar gets set before looking, not after.

The Confluence Scorecard

The Confluence Scorecard is a 15-part checklist built into the indicator. Every bar it adds up bull and bear votes, weighting each part somewhere between 2 and 14 points out of 100, and prints a score meant to tell you which way the market is leaning. It was once the project's only winning result — a 2.7-point edge on fresh data — until the comparison

behind that number was found to be broken; corrected, the edge is about a third of a point, the weights turn out to carry no information, and the fifteen parts behave like four or five.

The 2.7-point edge, withdrawn

WE WITHDREW THIS

Does the 15-part score tell you which way price is about to go?

How it was tested. The card was re-scored against a like-for-like benchmark that holds its own long/short mix fixed, because the original had compared a signal that trades both ways against simply buying and holding on a sample that drifted down.

RESULT

Originally +2.7pp (+1.49pp and +1.76pp on the two check groups). Corrected: +0.31pp and +0.37pp, pooled +0.34pp. Individual symbols run from -0.80pp to +1.66pp.

← too small to see | big enough to trust → 0.17× the floor

HOW MUCH DATA

32,809 trades across eight check symbols, 1-hour bars, 4.5 years

SMALLEST EFFECT IT COULD SPOT

2 percentage points, the size fixed in advance as worth having

Never quote the 2.7 figure again; the card's real reading is a third of a point, which is nothing you can trade.

Best in ranges and downtrends, withdrawn

WE WITHDREW THIS

Is the card worth more when the market is falling or going sideways?

How it was tested. One trade at a time, split by market phase, scored first against buy-and-hold and then against a benchmark that keeps the card's own long/short mix fixed.

RESULT Originally +0.47pp in uptrends, +2.04pp in ranges, +2.66pp in downtrends.

← too small to see | big enough to trust → 0.41× the floor

Corrected, downtrends turn negative on both check groups at -0.24pp and -0.81pp, ranges fall to +0.99pp and +0.04pp, and uptrends read -0.08pp then +2.41pp.

HOW MUCH DATA

eight check symbols, 1-hour bars, 4.5 years

SMALLEST EFFECT IT COULD SPOT

2 percentage points, set before the test

The phases where the card looked best are exactly the phases where it leans short — it was being paid for the lean, not for the timing.

The card with a real stop and target

NO EFFECT FOUND

Once you trade the card's calls with a real stop, a real target and fees, does it make money?

How it was tested. The calls were run through the trade harness at a 1 ATR stop for 1R and a 3 ATR stop for 3R, counted first on every qualifying bar and then one trade at a time.

RESULT

Counting every bar: -0.0233R at 1 ATR/1R and +0.0652R at 3 ATR/3R. One trade at a time: -0.0341R and +0.0019R, break-even.

← too small to see | big enough to trust →

0.02× the floor

The overlapping count inflated the result thirtyfold and made its error bars five times too small.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
103,305 overlapping bars against 3,702 one-trade-at-a-time trades	0.080R

It breaks even before fees are even in the picture, and any bar-by-bar result should be treated as inflated until it is re-run one trade at a time.

Do the weights do anything

NO EFFECT FOUND

The card weights its parts from 2 to 14 points. Does the weighting change the answer?

How it was tested. The same card re-scored with all fifteen parts weighted equally, picking the same number of signals as the shipped version.

RESULT

Shipped weights give +0.31pp and +0.37pp; equal weights give +0.09pp and +0.56pp. The gap between the two is 0.22pp one way on the first check group and 0.19pp the other way on the second.

← too small to see | big enough to trust →

0.11× the floor

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
the same 32,809-trade universe, top 30% of readings	2 percentage points

Stop tuning the weights — both versions read zero, so there is nothing to tune toward.

Drop one of the fifteen parts

NO EFFECT FOUND

Is one bad part dragging the other fourteen down?

How it was tested. The card was re-summed fifteen times, each time with a different part removed; a removal had to beat the full card by 2 points on both check groups to count.

RESULT

Nothing qualifies and nothing comes close. The best candidate, dropping Macro-S (weight 8), gives +0.22pp and +0.93pp.

← too small to see | big enough to trust → 0.47× the floor

Removing the two heaviest parts, SR VWAP and Trend, hurts on both groups: -0.49/-0.24 and -0.10/-0.32.

HOW MUCH DATA

the same 32,809-trade universe, both check groups

SMALLEST EFFECT IT COULD SPOT

2 percentage points on both check groups

There is no part to blame and no part to cut — the card is not one broken row away from working.

Fifteen parts, how many real opinions

DESCRIBES, CANNOT PREDICT

A 15-part score sounds like fifteen opinions agreeing. How many is it really?

How it was tested. A version of the indicator that publishes every part separately was dumped off the live chart, and how closely the parts move together was measured.

RESULT

The card behaves like 3.6 independent votes on 15-minute and 4.7 on 1-hour, against a weight-implied maximum of 6.8 and 7.2 — 53% and 66% of the breadth it claims.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The three heaviest parts hold 40 of the 100 points and move together at +0.52 to +0.73 (DI against Trend +0.73 and +0.61). The part holding the fourth-heaviest weight, 7 points, votes on only 2.4-2.6% of bars and contributes 0.2 of 100. Two parts agreed on 100% and 97.3% of bars, which is a fixed offset rather than a vote.

HOW MUCH DATA

703 bars each of 15-minute and 1-hour

A high score is not fifteen things agreeing; it is four or five things said several ways, so read it as less confirmation than it looks like.

Does the offline copy match the live card

WORKS AS CLAIMED

Is the copy used for all this testing really the same card that is on the screen?

How it was tested. All fifteen parts were rebuilt in Python by three modules working in parallel and compared bar by bar against dumps of the live indicator's own per-part output.

RESULT

`All fifteen parts reproduce bar for bar. The assembled net score is 99.90% exact, the seven misses all being a file's final still-forming bar.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

A real bug was found in the trend-strength code along the way, affecting 11 bars of 13,398 on coins with coarse price steps.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
`13,398 scored bars across nine symbols and two timeframes; 6,745 bars for the assembled score`	`not stated - this is a correctness check`

Everything else measured here is about the real card, not a rough imitation of it.

Costs and trade geometry

This group is about the money a trade loses before you are right or wrong about direction: exchange fees, and how the width of your stop and the size of your target change what you keep. It is the one corner of the project where something clearly survived testing, but what survived is a cost saving, not a way to predict direction. A round trip costs roughly 0.03 to 0.05R, which is larger than the biggest effect this project has ever measured, and the

popular fixes for it — a tighter stop, a fixed dollar volatility cutoff, banking half early — all made things worse or were withdrawn.

Tighter stops cost more in fees

WORKS AS CLAIMED

If I pull my stop in closer, do I pay less in fees?

How it was tested. Every entry was traded both long and short with the same stop and target and the two averaged, so market drift cancels out exactly, then the whole thing was repeated on a shuffled price tape where no trend effect can exist.

RESULT

At a 1R target the cost runs -0.0958R, -0.0480R, -0.0320R, -0.0240R and -0.0160R as the stop widens from 0.5 to 1.0, 1.5, 2.0 and 3.0 ATR - almost exactly one divided by the stop distance.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Widening from 1 ATR to 3 ATR saves +0.0333R per trade on the real tape and +0.0332R on the shuffled tape, the same number. The harness matched the fee-only line to 4 decimal places at every stop width.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
nine symbols on 1-hour bars; the original grid used seven symbols and 3,000 entries each	not applicable - this is arithmetic, matched to 4 decimal places

A 0.5 ATR stop costs six times what a 3 ATR stop costs for the identical dollars at risk, because halving the stop doubles your position size and fees are charged on the position, not on the risk.

What one round trip costs you

DESCRIBES, CANNOT PREDICT

Before I am right or wrong about anything, what does a single trade cost me?

How it was tested. Measured straight off the random-entry comparison arm, which loses money for one reason only: fees.

RESULT About 0.03 to 0.05R per trade.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

HOW MUCH DATA

18,243 trades at a 2 ATR risk cap on the cheapest fee tier

Any idea has to clear 0.03 to 0.05R before it is worth taking, and the largest effect this project has ever measured is smaller than that.

Fee share of risk on each chart

DESCRIBES, CANNOT PREDICT

On each of my charts, how much of the money I am risking goes to the exchange?

How it was tested. Worked the fee out as a share of risk - fee rate times price, divided by twice the ATR, since the method risks about two ATR - then ran it at each timeframe's current ATR and checked the formula reproduced every row of the original fee table.

RESULT

```
At the 0.060% losing-trade round trip: 4H (ATR 40.99) 1.84%; 1H (ATR 24.37) 3.10%; 24M (ATR 17.24) 4.38%; 3M (ATR 5.41) 13.97%. A winning trade at 0.030% halves every figure.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

At a 0.12% round trip the same formula gives ATR 25 -> 5%, ATR 12 -> 9-10%, ATR 7.50 -> 15%, ATR 5 -> 23%, ATR 3 -> 38%.

HOW MUCH DATA

```
4 timeframes at one price snapshot (ETH 2,519.40); formula cross-checked against all 5 rows of the source fee table
```

SMALLEST EFFECT IT COULD SPOT

```
not applicable - this is arithmetic, not a measurement
```

Fees are a rounding error on the 4-hour and nearly a seventh of your risk on the 3-minute, so the quieter the chart the bigger the exchange's cut of the same trade.

The do-not-bother floor is a percentage

WE WITHDREW THIS

How quiet does the market have to get before a trade is not worth taking?

How it was tested. Three fixed dollar thresholds had been written down at different times and each treated as a constant; re-deriving the formula showed price sits on top of the fraction, so the figure has to move with the price of ETH.

RESULT

```
$7.50, $5.12 and $3.76 turned out to be the same quantity measured at three different prices.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The floor is 0.20% of price: $3.76 at ETH 1,880 but $5.04 at ETH 2,519.40, so quoting $3.76 today understates it by about a third. The 15% cut-off the floor is solved from was inherited convention and was never tested, and the fee gate built on it was about three times stricter than a limit-order trader actually faces (23% assumed, versus about 15% limit-in/market-out and about 8% limit both sides at ATR 4.86). It was deleted as a rule on 2026-08-06.

HOW MUCH DATA

```
3 previously recorded dollar figures, re-derived at current price
```

SMALLEST EFFECT IT COULD SPOT

```
not applicable - this is arithmetic, not a measurement
```

Never skip a setup because of a fixed dollar volatility figure: the real threshold is 0.20% of price and moves as price moves, and the cut-off behind it was a convention nobody ever measured.

Do wide targets collect extra?

WE WITHDREW THIS

Do moves that get going keep running, so that a bigger target picks up more than a coin flip would pay?

How it was tested. The same both-ways grid, compared against what a coin-flip market would pay after fees, then re-priced once trades sharing the same market move stopped being counted as separate ones, and re-checked on two separate groups of symbols.

RESULT

Originally +0.0238R at a 3 ATR stop and 3R target, against a floor of 0.0676R - the smallest effect that test could spot.

← too small to see | big enough to trust →

0.46× the floor

Re-priced, the range runs about -0.014R to +0.064R and includes zero. The estimate later rises to +0.0312R with 8 of 9 symbols leaning positive, but it fails on both checking groups, it lives disproportionately in one calendar year (2025), and the only slice sharp enough to see the target effect reads negative. That cell's actual net expectancy is +0.0051R.

HOW MUCH DATA

344,763 trades across nine symbols on the dense re-run; 3,702 trades sharing no bars in the earlier version

SMALLEST EFFECT IT COULD SPOT

0.0676R originally; 0.0535R on 1-hour; 0.0308R on the 3-minute cut

Keep wide stops for the fee saving, not because trends were shown to carry - the carrying-on half of the story is the part that died.

Midpoint stop versus far-edge stop

CLAIM DID NOT HOLD

For an order block trade, should my stop sit in the middle of the zone or past its far edge?

How it was tested. Identical entry and identical target on every trade with only the stop moved, then the answer was required to hold for longs and shorts and for the first and second half of the sample.

RESULT

```
Midpoint stop: 29% win, -0.091R (n=62). Stop 1 ATR beyond the far edge: 68% win,
+0.044R (n=59).
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

In 22 of 62 cases the midpoint stop was taken out by a move the wider stop would have survived to reach target, and zero cases ran the other way. Sweeping the multiple, the sign only holds steady across every split from 1.5x ATR (76% win, +0.087R) through 1.75x (78%, +0.081R) to 2.0x (80%, +0.077R); 1.0x (+0.044R) and 1.25x (+0.053R) both flip sign.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
65 zones, 62 of them retested, over 360 hourly bars across 15 days	not stated; 62 trades over 15 days on one timeframe

Put the stop past the far edge of the zone rather than at its midpoint - the 22-to-0 count never inverted - but 15 days of data cannot pin down the exact multiple beyond a rough 1.5x to 2x ATR band.

Taking half off at 0.5R

CLAIM DID NOT HOLD

Does banking half the position early and pulling the stop to breakeven make me more money?

How it was tested. The same trades run four different ways, with the result then broken down into win rate, average win and average loss.

RESULT

```
Plain: 36.5% win, +0.035R, average win +1.94R, average loss -1.06R. All out at 1R:
52.3% win, -0.015R.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Half off at 0.5R plus a breakeven stop: 67.0% win, -0.032R, with the average win collapsing to +0.47R while the average loss does not move at all. Reproduced from win rate times average win and loss to four decimal places.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
a 360,000-bar panel across nine symbols	not applicable - this is a breakdown of the same trades

Taking half off nearly doubles your win rate and turns a small profit into a loss - it relabels losers as winners instead of making the trades better.

Calls Made In Advance

The idea is simple: write the trade down before it happens - the trigger price, the target, the give-up level, the exchange and a deadline - then score it either way once the clock runs out. Six such calls were made between 25 and 29 August, plus three zones drawn by hand with the give-up rule fixed at the moment of drawing. Four of the six calls passed, but two of those wins were handed over by the shape of the trade rather than the idea, the three hand-drawn zones went nought for three, and the one flattering forward record from earlier collapsed once it turned out no pass rule had ever been set before the outcome.

Six calls made in advance

COULD NOT TELL

If I write the trade down before it happens - trigger, target, give-up level, deadline - how often does the call come good?

How it was tested. Six calls were fixed in writing with a trigger price, a target, a give-up level, the exchange and a deadline, then scored on 24-minute closing prices, counting only bars strictly after the trigger bar.

RESULT

```
4 PASS, 1 FAIL, 1 EXPIRED. Two of the four passes (TEST 2 and TEST 4) were nearly free on their own stated geometry, so on evidence rather than tally the run produced exactly 1
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

properly-shaped pass out of 6.

HOW MUCH DATA

```
6 pre-registered calls, 2026-08-25 to 2026-08-29, ETHUSDC.P on Hyperliquid
```

SMALLEST EFFECT IT COULD SPOT

```
not stated as an effect size; the log's own target for a meaningful record is roughly 90 calls
```

Four out of six reads like a system that works; the honest count is one clean result out of six, which tells you nothing yet.

The one properly built call

COULD NOT TELL

With the target and the give-up level set at sensible distances apart, does the break actually pay?

How it was tested. Named 2026-08-26 with the geometry stated first - trigger a 24-minute close above 2514.90, target 2534.00 (+18.00), give-up 2505.50 (-10.50), 1 to 1.71, break-even 36.8%, deadline 2026-08-28 23:00 UTC - and checked that the target had not already been touched when the trigger fired.

RESULT

```
PASS - fired at close 2515.60, target 2534.00 hit at 08:24 on a high of 2545.30, worth
+18.40. The give-up was never approached (lowest low after the trigger 2516.10).
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The bar that did it ran 29.20 points on 3.23x average volume; the trigger bar itself ran 1.77x.

HOW MUCH DATA

```
1 call; fired 2026-08-27 08:00, target hit 24 minutes later
```

The one call out of six that was built properly, and it is still a single trade decided by one 29-point bar.

Two passes that were nearly free

COULD NOT TELL

When a call passes, was there actually anything at risk?

How it was tested. Two of the six passing calls were re-measured afterwards: how far the target already sat when the trigger fired, and what the risk-to-reward really was at the price the trigger actually filled at.

RESULT

```
TEST 2 risked 41.60 points to make 10.40 - 1 to 0.25, needing an 80.0% win rate just
to break even - and its target was already only 4.20 points away when the trigger
fired.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

TEST 4 declared 1 to 1.18 and a 45.9% break-even, but the trigger bar's own low of 2483.40 was already 4.90 points through the target, and the close-basis fill landed 14 points past the named level, making the real geometry 1 to 0.38 and the real break-even 72.6%.

HOW MUCH DATA

```
2 of the 6 pre-registered calls (TEST 2 and TEST 4)
```

Two of the four wins were handed over by the shape of the trade, not the idea - measure the risk-to-reward at the fill, not at the level you named.

Both range edges named a day ahead

COULD NOT TELL

If you name both edges of the range a day in advance, does price actually close through either one?

How it was tested. Named 2026-08-25 23:20 UTC - up side a 24-minute close above 2469.00 (target 2483.50), down side a close below 2437.80 (target 2419.70), deadline 2026-08-26 15:20 UTC, closing prices only, so a wick through the level does not count.

RESULT

```
EXPIRED, neither side fired. Best close up 2468.30 - short by 0.70. Best close down 2439.60 - short by 1.80.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Both edges were wicked through (2474.40 and 2431.60) across a 42.80-point range and neither closed through.

HOW MUCH DATA

```
39 closed 24-minute bars inside the 40-bar window
```

Whether you judge a level on closes or on wicks changed the answer more than the market did - on wicks this same call would have fired both ways and been counted twice.

The same call, two exchanges

COULD NOT TELL

Would my call have scored the same if I had used a different exchange's prices?

How it was tested. The first named-level call - a 24-minute close below 2477.60, target 2453.40, give-up 2488.20, 40-bar limit, all fixed in public at 09:00 UTC on 2026-08-25 - was scored on Hyperliquid as written, then re-scored on Binance prices over the identical window.

RESULT

```
PASS on Hyperliquid - trigger filled 09:36 at 2473.80, target 2453.40 reached 13:12, lowest price after the trigger 2437.50, highest 2485.60, so the 2488.20 give-up was never reached.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

FAIL on Binance - it printed 2489.00 three minutes after the trigger, clearing the give-up by 0.80. A 3.40-point gap between venues flipped the answer.

HOW MUCH DATA

```
1 call scored twice; resolved 11 bars into a 40-bar window
```

SMALLEST EFFECT IT COULD SPOT

```
not stated as a number; the log said roughly 90 such calls would be needed to show a large effect
```

Name the exchange before the call - a few points of difference between venues can turn the same trade from a win into a loss.

Zones drawn by hand, rule fixed first

CLAIM DID NOT HOLD

When I draw a support or resistance zone myself, does it hold?

How it was tested. Three order blocks drawn on the live chart, each with its give-up rule - a body close outside the zone - stated at the moment of drawing so it could not be moved afterwards, then left alone and scored when price arrived.

RESULT

```
3 drawn, 3 failed, 0 held. The 2497.70-2506.90 zone broke on its first touch (body close 2493.50, 4.20 below the base).
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The 2496.20-2503.30 zone broke within two cycles (close 2508.60, 5.30 above the top). The 2483.50-2495.80 zone survived four tests including an 18-point wick and a close 1.20 above its base, then died 96 minutes later on a close of 2471.20.

HOW MUCH DATA

`3 zones drawn, 2026-08-27 to 2026-08-29`

On the only three zones scored under a rule fixed before the fact, hand-drawn levels went nought for three - and the cleanest-looking one broke fastest.

The forward record that was withdrawn

WE WITHDREW THIS

The level-plus-volume-plus-structure checklist had a 6-out-of-8 record on live breaks - was that real?

How it was tested. The checklist was graded on live breaks over roughly 20 monitoring cycles with a running record of correct and incorrect calls, then the record itself was audited for whether any pass rule had been written down before the outcomes existed.

RESULT

```
The stated record was 6 correct declines, 1 correct positive, 1 failed positive - WITHDRAWN as unverified on 2026-08-05.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

No success rule was ever set beforehand, so the time horizon and the threshold for each 'correct decline' were chosen once the outcome was already known.

HOW MUCH DATA

`8 graded instances`

If you decide after the move what counts as being right, you will be right most of the time - which is exactly why the trigger, target and deadline have to be written down first.

Where we got it wrong

This project keeps a list of its own mistakes. Every entry below was once written down as a finding and then failed a check run by the same people who produced it — a signal scored

against the wrong yardstick, a confirmation that could not be read until after the trade was gone, a run of sweeps counted three times over, a record graded once the outcome was already known. The first entry matters most: the only positive result the project ever had turned out to be an accident of the comparison, and it was caught days before it went into print.

The +2.7pp edge that was not there

WE WITHDREW THIS

Was the one number in this project that said something actually real?

How it was tested. The card leans short and the test window drifted downwards, so scoring it against simply buying and holding paid it for its mix of longs and shorts rather than its timing; it was re-scored against a comparison that holds that mix fixed and tests only the timing.

RESULT

Originally reported as +2.7pp. Corrected: +0.31pp on the first check group and +0.37pp on the second, pooled +0.34pp, against a bar of +2pp set before the test.

← too small to see | big enough to trust → 0.17× the floor

Individual symbols run from -0.80pp to +1.66pp. The phase pattern that made it look strongest in ranges and downtrends was the mistake itself: corrected, downtrends read -0.24pp and -0.81pp.

HOW MUCH DATA

32,809 trades across eight check symbols, 1-hour bars, 4.5 years; the offline copy matches the live card on 13,398 bars

SMALLEST EFFECT IT COULD SPOT

+2pp, set before the test

Never quote the +2.7pp figure again — a signal that trades both ways has to be scored against something that also trades both ways, or it collects free points for the market's drift.

Our tests were blinder than we said

WE WITHDREW THIS

When a test here says it found nothing, how small a thing could it really have spotted?

How it was tested. The original check wobbled a fake reaction by only 0.03 ATR inside a 0.10 ATR counting band, so every fake landed dead centre and looked easy to detect; it was redone using the real height of an actual order block.

RESULT

A reaction inside a median-height block can sit 0.40 ATR away from the quoted line, 13 times the wobble originally used.

← too small to see | big enough to trust →

0.13× the floor

Recalibrated, the smallest detectable effect moves from 1.03% to 4.02% — from about 1 reaction in 100 to about 1 in 25, roughly 0.18R per trade. Fees are 0.035R and the largest effect the project ever measured was +0.024R.

HOW MUCH DATA

22,687 blocks, median height 0.800 ATR, 90th percentile 1.121 ATR

SMALLEST EFFECT IT COULD SPOT

about 1 reaction in 25, or 0.18R per trade

Read every 'nothing found' here as 'no large effect' — these tests could not have seen an edge the size of the biggest one this project ever measured.

A confirmation you cannot read in time

WE WITHDREW THIS

The volume rule says the hour above must also be heavy – can you actually know that at the moment you would press the button?

How it was tested. Plain arithmetic on how bars nest inside each other, plus measuring how much of the confirming hour's volume was the signal bar's own volume.

RESULT

No. A 1-hour bar contains the 24-minute bar that fires the signal, so when the signal closes its parent hour is still forming.

← too small to see | big enough to trust →

0.35× the floor

On the two big events the signal bar was 51,621 of 97,579 (53%) and 59,690 of 110,159 (54%) of the whole hour's volume — over half the confirmation is the same trades counted twice. Tested using only what was visible at the time, confirmed minus partial is -0.855 ATR against a detection floor of 2.449 ATR, and the usable version fires 17 times in 133 days.

HOW MUCH DATA

Arithmetic plus the two largest live events of 2026-08-28; the measured version ran on 8,000 bars and 237 separate events

SMALLEST EFFECT IT COULD SPOT

2.449 ATR

You cannot read a confirming hourly volume figure until the hour has closed, and by then the entry is long gone — the most trusted rule in the framework had never been tested the way it is used.

The resting order that invented an edge

CLAIM DID NOT HOLD

If I test a fib the way I would actually trade it – an order sitting at the level waiting to be filled – does it work?

How it was tested. The identical test was re-run with entry taken at the level, as a resting order would fill, instead of at the close of the bar that touched it.

RESULT

`+2.5pp at the 0.500 level, +7.6pp at 0.786 and +15.5pp at 0.95 — bigger the deeper the level, with confidence readings up to 25.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

It is entirely construction: a bar that reaches a level closes on the favourable side of it 57-58% of the time, and at deep levels the fact that the move is still going guarantees it. This would have been the eleventh false positive.

HOW MUCH DATA

`The same nine-symbol panel`

Any backtest that fills a resting order at a level is quietly picking the moves that carried on — the fill does the selecting for you, so test at the close of the touch bar instead.

A fake level placed at the wrong distance

WE WITHDREW THIS

Does price get pulled back to an untouched volume shelf more than to any other price?

How it was tested. The first fake comparison level was built from the day's open and sat a different distance from price, so it was harder to reach for reasons that had nothing to do with being a real level; it was rebuilt at a matching distance on the same side.

RESULT `The first attempt read +17.3 / +15.2 / +18.1pp and was wrong.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Corrected, the untouched point of control was reached 88.4 / 88.1 / 87.5% of the time against 88.2 / 88.1 / 87.8% for the matched fake — a difference of +0.2 / +0.0 / -0.3pp, with average distances matching to two decimal places.

HOW MUCH DATA

`3 symbols on 1-hour bars, 240-bar horizon (about ten days)`

SMALLEST EFFECT IT COULD SPOT

`not stated in the same units`

Price reaches an untouched volume shelf 88% of the time within ten days, and it reaches any other price at that same distance just as often — it is a price, not a magnet.

Thirty logged sweeps that were really sixteen

DESCRIBES, CANNOT PREDICT

How many real, separate swing-failure observations does the forward log actually hold?

How it was tested. Every row of the log was checked for exact duplicates and for repeated pokes at the same price shelf inside a single session.

RESULT

```
3 exact duplicates — one entry appears three times with an identical timestamp,
another twice — and 14 rows collapse into their first because they are repeat sweeps
of the same shelf in one
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

session. Nine separate shelves were each swept 2-3 times. The 30 raw rows become 16 independent observations against a stated target of about 30; the count has since grown to 21.

HOW MUCH DATA

```
Every row logged 2026-08-26 to 2026-08-28,
about 30 raw rows
```

SMALLEST EFFECT IT COULD SPOT

```
not applicable - a count
```

Poking the same shelf three times in one session is one observation, not three — count shelves, not rows, or the sample inflates and manufactures a result out of nothing.

A record graded after the outcome

WE WITHDREW THIS

Did the level-plus-volume-plus-structure checklist really call breaks correctly six times out of eight?

How it was tested. The claimed live record was re-examined for whether a rule saying what counts as a correct call had been written down before the calls were graded.

RESULT `The stated record was 6 correct declines, 1 correct call and 1 failed call.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Withdrawn as unverified on 2026-08-05: no success rule existed in advance, so both the time horizon and the threshold were chosen once the outcome was already known.

HOW MUCH DATA

```
8 graded instances over roughly 20 monitoring cycles
```

Write down what counts as a win before the move happens — grade afterwards and you will be right almost every time, on paper only.

Does the indicator do what it says

Every tool on the chart makes a promise in its name and its manual: this cloud is a money-flow reading, this label marks a market phase, this dot sits on a real swing high. These tests set aside whether any of it predicts price and ask a much narrower question — does the code actually compute the thing written on the label? Mostly yes: the core features rebuild exactly, and the phase labels never rewrite history to look clever. But three features were

not what their names said, including a "money flow" cloud with no volume in it at all and an eight-line ribbon that turned out to be one line.

Is anything in the indicator broken?

WORKS AS CLAIMED

Does every reading on my chart actually compute what its label says it computes?

How it was tested. Four features were rebuilt in plain code from their written formulas and compared value-by-value against the live chart, with the checking rules fixed in writing before any test ran.

RESULT

```
All four pass. The money-flow replica matches the live plot to within 3.6e-14, and the
pressure band to 5.2e-14.
```

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Divergence markers: 1,004 comparisons, zero false fires and zero misses, all 5 events matched exactly. The scorecard code is byte-identical in both indicators, with 17 bull and 17 bear parts all correctly paired and none left unpaired. Two things do mislead, though: on the hourly the money-flow pressure band is flat on 36.8% of bars and simply retraces the outer cloud on another 19.2%, so it says something new on only 44.0%; and the divergence row stays lit for 10 bars after one event, so the hidden card counts each diamond 9.86 bars for every one you can see. The Dot row lights on 19.89% of all bars (19.76-20.06% across six series).

HOW MUCH DATA

```
298,200 bars across 3 symbols and 2 timeframes for the money-flow check; 1,004 bar-by-bar comparisons for the divergence markers; 15,576 characters of shared scorecard code compared byte for byte
```

SMALLEST EFFECT IT COULD SPOT

```
not applicable - correctness check, no edge was measured
```

Nothing is broken, but the hidden scorecard and the chart in front of you do not always tell the same story, and the pressure band says nothing new on more than half of all bars.

Do the phase labels rewrite history?

WORKS AS CLAIMED

Does the market-phase label only name a range after price has genuinely left it, or does it quietly redraw the past so it looks like it knew?

How it was tested. The phase rules were replayed in plain code from the chart's own bars and compared against what the indicator actually drew, across the whole timeframe stack.

RESULT

554 of 554 finished ranges were named only after a real close outside a bracket that was already on screen.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Price alignment error 0.0000 on every timeframe, and the phases tile the chart with no gaps and no overlaps. Switching to smoothed candles changes nothing (identical to 1e-9), and the labels read the same on the 4H, 24M and 3M charts when anchored to the 1-hour (identical to 1e-8). Against the live chart the phase names matched 5 of 5, in order.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
about 39,000 bars - 4H 3,114 / 1H 12,453 / 24M 10,515 / 3M 13,087	not applicable - correctness check; this checks the names and their order, not the exact bar each phase begins on

What you saw live is what you see when you scroll back later, which is more than most phase tools can say - though drawing it honestly is not the same as it being worth trading.

Does the offline copy match the live card?

WORKS AS CLAIMED

When the scorecard is tested over years of history, is it the same card that is actually on my screen?

How it was tested. All 15 parts of the card were rebuilt in plain code and compared bar by bar against 18 dumps of the real card's own output pulled off the live chart.

RESULT

All 15 parts reproduce the live card bar for bar. The assembled score is 99.90% exact over 6,745 bars, and the 7 misses are each file's final still-forming bar.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

A real bug surfaced on the way: rounding dust of about 1e-17 made two mathematically equal one-tick moves compare as unequal, flipping the trend-direction reading on 11 of 13,398 bars (0.15%), with zero effect on BTC, ETH or BNB.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
13,398 scored bars across nine symbols and two timeframes; 6,745 bars for the assembled score	not applicable - correctness check, no edge was measured

The live chart only holds about 1,000 bars while the questions needed 40,000, so this copy is what made every later scorecard test possible - and it can be trusted.

Do the levels sit where they claim?

WORKS AS CLAIMED

Is every liquidity level drawn on my chart really on a swing high or low, or are some of them just lines?

How it was tested. Every drawn level was checked against the actual candle data over the full history, with the swing count per 1,000 bars compared across five timeframes and the score checked against the running median volume.

RESULT

`3,809 of 3,809 sit on a real candle extreme, and 3,809 of 3,809 are the extreme over the full 15-bar window, with zero swings skipped.`

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

Swings per 1,000 bars at the same setting: 3m 96.6, 15m 94.8, 30m 94.8, 1h 95.2, 4h 92.4. A score of 5 lands on exactly 1.00x the median volume, each step up about 1.23x. Two faults were found and fixed: 27.5% of levels used to pin at the top score of 10 (16.9% after), and brand-new levels were hidden 56% of the time against 34% for levels over 100 bars old.

HOW MUCH DATA

`3,809 levels on 1-hour ETH; density on 3m, 15m, 30m, 1h and 4h; 43,076 levels for the score calibration; 12.8 million level-bars for the age check`

SMALLEST EFFECT IT COULD SPOT

`not applicable - correctness check, no edge was measured`

The lines are drawn exactly where they claim to be and one setting works on every timeframe - but whether a bigger score means a stronger level is a separate question, and no edge was found there.

Does the money flow cloud use volume?

CLAIM DID NOT HOLD

The green and red cloud is called money flow - is it measuring money, or volume, at all?

How it was tested. The shipped formula was read straight out of the code and the darker inner pressure band was traced through its arithmetic, then checked against live readings.

RESULT

There is no volume term in the cloud at all. It is (close minus open) divided by (high minus low), averaged over 60 bars - a slow reading of candle body direction.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

The inner pressure band is forced to sit between zero and that volume-free cloud, so it can never exceed it or take the opposite sign; when volume actually opposes the cloud, pressure draws a flat zero, which looks identical to neutral volume. The same pass found that the plot labelled VWAP inside the oscillator panel is simply the fast wave minus the slow wave, residual exactly 0.0 - true of the paid script and of ours. Both the guidebook and the caption were corrected.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
source code check plus live 15-minute readings and the audit's bar-by-bar count over 298,200 bars	not applicable - correctness check, no edge was measured

A row you may have been reading as a volume signal is a price-shape average carrying an inherited name, and it can never warn you that volume disagrees because it never looks at volume.

Is an eight-line ribbon really eight readings?

CLAIM DID NOT HOLD

A ribbon of eight moving averages looks like a lot of information - is it?

How it was tested. The ribbon's direction and its width were compared against just the 20 and 55 moving averages on every bar where the ribbon was stacked in order.

RESULT

Ribbon direction matches the sign of the 20 minus the 55 on 100.0000% of bars (275,265 of 275,265), and ribbon width matches the gap between those same two lines on 100.0000%.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
275,265 bars - the 76.6% of bars on which the ribbon is stacked	not applicable - correctness check, no edge was measured

The six middle lines cannot carry anything the outer two do not, so it is one reading dressed up as eight.

Fake prices on smoothed candles

CLAIM DID NOT HOLD

If I switch my chart to smoothed candles, are my indicators reading the real price or the smoothed one?

How it was tested. Every price read in the three scripts was routed through the real candles, then every numeric reading on smoothed candles was compared against standard candles on the same bar, matched by timestamp.

RESULT

Before the fix, the same 1-hour bar read wave 17.88 against 11.94, and money flow 41.62 against 16.18. After the fix, 54 of 54 readings are identical, with zero differences.

This test's result and its floor were measured in different units, so the two cannot be placed on the same scale.

HOW MUCH DATA	SMALLEST EFFECT IT COULD SPOT
54 readings across the three scripts (12 + 13 + 29)	not applicable - correctness check, no edge was measured

The indicators were quietly reading prices that never traded - it is fixed now, but any reading taken off a smoothed chart before the fix described something that did not happen.

Samples count separate, non-overlapping events rather than bars, so one large move cannot be counted many times over. Every figure here is reproducible from the logs in this project.

Nothing in this document is advice, and none of it is a reason to take a trade.