

THE CREATIVE OPERATING SYSTEM

This document defines how we plan, produce, measure and learn from paid creative. It applies to every account and every strategist, and we follow it as written.

Two rules govern everything that follows. We write the question before we make the asset, and we write every question against exactly one level of the creative variable ladder. Together these rules make results comparable across assets, months and brands, which is what allows the system to accumulate knowledge rather than output.

THE THREE TESTING BUCKETS

Each bucket produces the input the next bucket requires

LEVELS 2-6 · BROAD

Moonshot

We test broadly and at low cost across subavatars, angles and formats on which we hold no data. We fund moonshots to generate metastudy signal, and we judge them on what they teach us. We do not judge them on return.

PRODUCES

Metastudy signal

LEVELS 3-6

Net New

We build new angles, values and formats on patterns the metastudies already support. Net New produces our winners and establishes the performance ceilings we then scale against.

PRODUCES

Winners

LEVELS 5-10

Iteration

We move one variable on a proven asset and hold every other variable static. Iteration compounds a winner into scale, and it is the only bucket that produces unicorns.

PRODUCES

Unicorns



We run all three buckets every month because each one consumes what the previous one produces. A portfolio built only on Iteration exhausts its supply of winners and stops finding new ceilings. A portfolio built only on Moonshots generates signal it never converts into scale.

The mix between the three is not a matter of preference. Rolling reach and creative insight maturity determine it, and the monthly agenda records that determination explicitly. Section 01 sets out how we make that decision.

THE OPERATING RHYTHM

We run the same sequence every month, and every stage feeds the next one

WHEN	WHAT WE RUN	WHY WE RUN IT THEN
First week of month	Monthly agenda	We set the batching strategy and the portfolio layout for the month, working from the retro's shortlist.
Every week	Creative batch launch	We ship each batch with its control and with one variable isolated per test.
Monday	Analytics report, then creative report	We publish the same view every week so that we read change directly instead of reconstructing it.
Real time	Winner and unicorn alerts	Both tiers require action in the same week: we scale the creative and open the iteration set.
Launch + 2 weeks	Creative debrief	Two weeks clears the launch window and still falls inside the period in which the team remembers its intent.
Last week of month	Monthly retro	We close the month, amend the metastudies, restate maturity and produce the next shortlist.
Continuous	Creative metastudies	We amend them as debriefs land and review them formally at every retro.

THE CREATIVE VARIABLE LADDER

We write every hypothesis and every insight against one level of the ladder

The ladder lists everything we can change about an ad, ordered from the broadest change to the most granular. It is not a sequence we work through in order. It is the vocabulary that makes separate tests comparable to one another.

Every pitch names the single level it moves. Every debrief files its result against that level. Every metastudy entry sits under the level it describes. When a strategist cannot name the level a concept tests, the concept does not enter the pipeline.

Higher levels carry more variance than lower levels. A test at level 2 or level 3 therefore requires more spend than a test at level 9 or level 10 before its confidence interval narrows enough to support a decision. We budget the measurement alongside the production, and we reject any pitch that tests a high level on a budget that cannot read the result.

LEVEL	WHAT IT CONTROLS		BUCKET
STRATEGIC — CHANGING THESE PRODUCES A DIFFERENT AD			
1	Portfolio	The composition of the slate itself. We test this month over month at the agenda level rather than asset by asset.	AGENDA LEVEL
2	Subavatar	The specific segment inside the avatar that the ad addresses.	MOONSHOT / NET NEW
3	Angle	The argument the ad makes about the product.	NET NEW
4	Value	The payoff the ad promises, and the person to whom that payoff matters.	NET NEW
EXECUTIONAL — CHANGING THESE PRESERVES THE UNDERLYING BET			
5	Hook Attributes · Emotion · Open loop	The opening beat. A hook change on a proven angle is Iteration. A hook built for a new subavatar is Net New.	ITERATION / NET NEW
6	Format	The container: static, short edit-driven video, long form, or creator-led.	NET NEW / ITERATION
7	Copy Structure	The order in which the ad sequences its argument.	ITERATION
8	Attributes	The specific claims, features and proof points the ad carries.	ITERATION
9	Psychological Principle	The persuasion mechanic the ad relies on.	ITERATION
10	Elements	The smallest movable parts: captions, music, pacing, thumbnail and call to action.	ITERATION

CREATIVE INSIGHT MATURITY

Maturity is the sum of our retros, metastudies and debriefs, and we measure it directly

We measure maturity on two axes. **Coverage** counts the ladder levels on which we hold a statistically significant read. **Tightness** measures how narrow the confidence intervals behind those reads are.

Maturity determines the number of ads we launch, the mix of buckets we run, and the ladder levels we are permitted to test in a given month. Maturity is also the Confidence axis in ICE. We score confidence from maturity and from nothing else.

Stated this way, maturity tells us what to build. A brand with tight intervals at levels 5 through 10 and no data at levels 2 through 4 executes well and does not know where its next ceiling sits, so that brand runs Net New and Moonshot. A brand with the opposite profile knows its angles and is not compounding them, so that brand runs Iteration.

ICE PRIORITIZATION

We score every candidate on headroom, maturity and resource draw

IMPACT — HEADROOM. We read impact from the marginal impact to rolling reach. When each increment of spend moves rolling reach very little, the current strategy retains headroom, and scaling it delivers high impact. As rolling reach approaches its cap, impact moves away from scaling the current strategy and toward diversifying it.

CONFIDENCE — MATURITY. We score confidence from coverage and interval tightness at the ladder level under test. Higher maturity produces higher confidence. We do not score confidence from experience or conviction.

EASE — RESOURCE DRAW. We score ease inversely to the strategy and production resource a concept consumes. Short edit-driven video and statics score highest. New-concept shoots score lowest.

Ease and headroom move in opposite directions, and that relationship drives the portfolio. **When we hold headroom we can afford low-ease concepts. When rolling reach caps we cannot, and we need volume and variety instead.** We score ICE to force this argument before production rather than after it.

PORTFOLIO LAYOUT AND MONTHLY PACING

Rolling reach sets the state, and the state sets the mix, the volume and the ease ceiling

STATE A

Reach not capped

Heavy Iteration · lower volume · lower ease · levels 5–10

We hold headroom and we know what works. We spend the month on craft rather than count, and we compound proven winners toward unicorn scale. This state requires mature insights.

STATE B

Reach capping

Iteration and Net New · higher volume · higher ease · levels 3–6

The current strategy has reached its audience. We raise ease and diversify into new angles and formats at a controlled rate while we continue to harvest the winners we already hold.

STATE C

Reach capping hard

Net New and Moonshot · highest volume · maximum ease only · levels 2–6

We produce only very easy formats, which means short edit-driven video and statics, and we diversify into moonshots. We are not scaling in this state. We are searching for the metastudy that unlocks the next winner.

The monthly agenda records this decision explicitly in the first week of the month. The strategist states the reach state, the bucket mix, the ladder levels in play, the ad volume and the week-by-week batch plan. No line enters the agenda without a hypothesis and a ladder level attached to it.

THE THREE-STAGE PIPELINE

Every asset clears three gates, and nothing advances until it clears its gate

STAGE ONE

Pitch

The strategist writes the concept and its hypothesis against one ladder level in one bucket. The pitch states what we believe, states why we believe it, and states the result that would disprove it.

GATE

Does this test a level on which we do not already hold a tight read?

STAGE TWO

Script

The strategist converts the concept into a shootable and editable document, names the isolated variable explicitly, and identifies the control the asset will run against.

GATE

Does this still move one variable and only one variable?

STAGE THREE

Produce

Production builds the asset to spec and tags it with its bucket, its ladder level and its control, so that the debrief can locate it later.

GATE

Does the finished asset match the approved script, and can we read it against its control?

We enforce the gates because drift causes most of our creative failures. A strong pitch becomes an average ad when nobody verifies at the script gate that the isolated variable survived the rewrite.

WEEKLY BATCHING

We batch every week, and batch discipline carries our testing integrity

- We batch creative every week against the plan set in the monthly agenda. The agenda decides what a batch contains. The batch itself decides whether we can read the result at all.

1

Move one variable per test. One ladder level changes and every other level holds static.

2

Hold everything else static. This includes talent, set, edit template, pacing, music, copy structure, call to action, placement and landing page. When two variables move we cannot attribute the result, and the asset becomes spend rather than evidence.

3

Batch by production cost and isolate by variable. Assets that share talent, set or edit template shoot together, and inside the batch each asset differs from its control by exactly one variable.

4

Launch every batch with its control. Without a control we cannot read the batch, and the debrief carries no information.

5

State the confounders you cannot eliminate in the pitch, not in the debrief. A strategist who cannot name them in advance will construct them after the result arrives.

Batching creates pressure toward cheap production, which pulls the portfolio down the ladder and toward iteration. The ease ceiling set by the reach state counteracts that pressure. When batching overrides the ease ceiling month after month, we are setting the portfolio after production instead of before it.

WINNER AND UNICORN THRESHOLDS

We define both tiers by percentile against the account’s own creative history

We set thresholds by percentile rather than against a fixed target so that the bar rises as the account improves. We rank creatives by spend and require that the creative held efficiency against the account’s goal metric, because a creative accumulates spend only while it continues to clear that metric.

BELOW THRESHOLD

Live

The creative has launched and is either still accumulating spend or is readable and sitting below the tenth percentile. It runs to its scheduled debrief and triggers nothing.

TOP 10TH PERCENTILE

Winner

The creative has cleared the tenth percentile and held efficiency against the goal metric. Net New produces most of our winners, and a winner tells us we have validated an angle or a value.

ON ALERT

Confirm significance, scale the creative, and open an iteration set at levels 5–10 in the same week.

TOP 1ST PERCENTILE

Unicorn

The creative has cleared the first percentile and sustains efficiency at scale. Our largest unicorns carry more than \$1M in spend on a single ad. Iteration on an existing winner produces almost every one of them.

ON ALERT

Open a dedicated iteration lane and amend the metastudy in the same week.

Net New produces winners and iteration produces unicorns. We do not launch unicorns. We build them from winners, one isolated variable at a time. A winner alert that we close without opening an iteration set is a failure of the system, because the opportunity to compound was available and we did not take it.

THE EVIDENCE STANDARD

Every hypothesis, report, debrief and metastudy entry carries a level, a significance and an interval

Significance tells us whether the result reflects a real difference. The confidence interval tells us the size of that difference and the precision of our estimate. We treat the interval as the more important of the two, because the interval determines whether we scale.

We re-test a significant result that carries a wide interval rather than scaling it. We act on a narrow interval around a small lift ahead of a wide interval around a large one. We do not enter a claim into a metastudy without its interval attached.

WEEKLY REPORTING AND DEBRIEFS

We publish two documents every Monday, and one debrief for every asset we launch

Weekly Analytics Report MONDAY

We publish the numbers in the same format every week. The report covers account-level and creative-level performance: current spend, current efficiency, decay, the creatives that are live but not yet readable, and the position of rolling reach against its cap. The report describes results and does not judge them.

Weekly Creative Report MONDAY, AFTER

The strategist writes the interpretation from the analytics report and organizes it by ladder level rather than by asset. For each level the report states what moved, the significance, the interval and the implication, and every conclusion names a bucket and a required action: iterate, extend, kill or hold. We keep these as two separate documents so that the strategist separates measurement from interpretation.

Creative Debrief LAUNCH + 2 WEEKS

We write a debrief for every launched asset two weeks after launch. Two weeks removes the distortion of the launch window and still falls inside the period in which the team remembers its intent. The debrief answers three questions in order: what did we predict, what happened and with what significance and interval, and what do we now believe. Only the third answer carries forward.

ROLLING CREATIVE METASTUDIES

The metastudy holds what we know about a brand’s creative, level by level

A metastudy answers one question: what do we know about this brand’s creative at each level of the ladder. We build it by aggregating debriefs and identifying the results that repeat. We do not publish it as a report. We amend it continuously and index it by the ladder so that a strategist can query any level directly.

Each level records four things: the claims that have held, the claims we have disproven, the confidence interval behind each claim, and the open questions. The open questions feed next month’s Net New and Moonshot slate directly.

A single debrief carries no weight on its own. It acquires weight when it narrows or breaks a pattern already recorded in the metastudy. Moonshots exist to supply this document and we fund them for that purpose alone.

We file every debrief against the ladder level it tested rather than the asset that tested it. That filing convention is what allows us to aggregate debriefs at all. Without a maintained metastudy we cannot score the Confidence axis in ICE, and ICE stops functioning.

MONTHLY RETRO

We run the retro against the agenda we set, not against our impression of the month

- 1

Review against plan. State what we planned to batch, what we batched, and where the mix or the volume broke.
- 2

Review against hypothesis. Work level by level through the debriefs and state what reached significance, how wide the intervals are, and which buckets paid.
- 3

Amend the metastudies. Record what we add, what we revise and what we retire, then restate insight maturity.
- 4

Set up next month. State the reach state, the resulting portfolio layout, and the shortlist that carries into the agenda.

The retro must produce two outputs: **a restated insight maturity and a shortlist for next month’s agenda.** The retro feeds the agenda directly, and without both outputs the agenda has nothing to work from.

DIAGNOSTICS

How to tell the system is failing

Each symptom below has one specific cause. We check this list at every retro.

An asset moves two variables.	We cannot attribute the result. That asset spent budget and produced no evidence.
A winner alert closes without an iteration set opening behind it.	We are finding winners and not building unicorns. We are skipping the step that compounds.
We judge a moonshot on its return.	We are removing the only bucket that finds new patterns.
We run heavy Iteration while rolling reach is capping.	We are compounding into a ceiling. We ignored State B or State C.
A hypothesis or an insight carries no ladder level.	We cannot aggregate the result. The month produces output and no knowledge.
We test a high ladder level on a low budget.	The interval never narrows, so maturity does not move and confidence never rises.
We score confidence high while maturity is low.	ICE is justifying the work we had already decided to make.
Batching determines the slate.	We are deciding the portfolio after production instead of before it.

Every one of these failures removes the same thing: our ability to attribute a result to a cause. We accept lower volume, slower diversification and more expensive reads in order to protect attribution, because attribution is what turns a month of spend into an asset we still hold next quarter.