Four Points, Elaborated

Lesson 21 appendix · the four claims from *The Leaderboard Nobody Can Score*, each taken apart · corpus-sourced where sourceable, argument flagged where not

1 · Why K/D does not predict victory

The ratio is a per-engagement measure. A war is not decided per engagement. It is decided by whether you can keep presenting engagements until the other side cannot. Those are different quantities, and conflating them is the whole error.

A good ratio is not evidence of skill. It is evidence of what you were short of.

Gabriel states the Mongol constraint without noticing it is a constraint: "a well-led and well-trained army, although smaller, will almost always defeat a larger army that is poorly led." Read the first clause. **Mongol armies were small, and no amount of generalship changes a birth rate.** The exchange rate was not a Mongol preference. It was the only variable available to them, so they built an entire doctrine — reconnaissance, the *nerge*, march divided and fight united, the deliberate gap — around squeezing it.

The Soviet Union had the opposite endowment and behaved accordingly. A continent of conscripts and an industrial base relocated east. When your scarce resource is time and your abundant resource is men, you spend men to buy time, and the ratio is simply not the thing you are optimising.

> **The testable form of this.** The importance of exchange rate varies *inversely* with the depth of your reserves. If replacement is fast, ratio is a vanity metric. If replacement is impossible, ratio is the only metric. Subotai and Zhukov are not two answers to one question — they are correct answers to two different questions.

And the Mongol campaign in Europe was not ended by attrition at all. The armies withdrew from Hungary in May 1242 after the death of Ögödei required the princes to return for the succession. Gabriel's own counterfactual is explicit — had the Great Khan not died, there was nothing between them and the Atlantic; the Grand Master of the Templars wrote to Louis IX that "no army of any significance stood between the Mongols and France," and Gabriel's verdict on the letter is "This was no exaggeration."

> *They had the best ratio in recorded history and it did not decide the campaign. A political clock did.*

2 · Why Stalin's number is team damage and not a ratio

A ratio requires two disjoint populations. K from set A, D from set B. Subotai had that: the dead at Liegnitz were Europeans, the dead Mongols were Mongols, and the two sets never intersect.

Stalin did not have that, and the consequence is arithmetic, not rhetorical. When the numerator is drawn from the same population as the denominator, every kill *increments the denominator*. The quantity being computed is not efficiency. It is net loss, with a sign error.

> **Formally: if both terms are drawn from one population N, the ratio is maximised by inaction.** Any metric with that property is not measuring performance. It is measuring restraint — and it will score a regime that does nothing above a regime that does anything, which is why no one keeps this statistic and why Stalin's figure is not comparable to Subotai's in any direction.

The purge of the Red Army officer corps is the clean demonstration. In 1937–38 Stalin destroyed a large part of his own command structure. **Enemy cost: zero. Own capability cost: severe.** It is the

only move on this entire board that is pure denominator — and the Soviet catastrophes of summer 1941 are, in part, the invoice.

And the system did not stop at its victims. Applebaum, at the moment the Terror begins: the NKVD "controlled the fate of more than a million prisoners. But the relative calm was not to last. Abruptly, the system was about to turn itself inside out, in a revolution that would **destroy masters and slaves alike**." Even the operator class was inside the denominator. Yagoda ran the secret police and was consumed by it; the terror is named after Yezhov, who followed him and followed him.

> **The transferable version, which has nothing to do with war.** Whenever a system's success metric is computed over a population that includes the system's own inputs, the metric will reward self-harm. A team measured on tickets closed that also files the tickets. A model evaluated on a benchmark its own outputs were used to build. **Check for set intersection before you trust a ratio.**

3 · Why nine sacks of ears beat 2026

The ear-count was a good instrument for three specific reasons, and none of them is that the Mongols were honest people.

One · a physical token with a uniqueness constraint. One right ear per body. You cannot inflate the count without producing the objects, and you cannot double-count a corpse, because it only has one right ear. The measurement is forgery-resistant by construction.

Two · the collector had no stake in the number. Special squads walked the field after the fighting. They were not the units whose performance the number reflected.

Three, and decisively · the cost of a wrong number fell on the person receiving it. The count went to a commander who had to decide what to attack next, with how many men. Overstate the enemy dead and he under-commits and loses a battle. **The consumer of the metric was the party punished by its error.**

> **The rule this yields: trust a number in proportion to how badly the person receiving it is hurt by it being wrong.**
> That single test explains why internal management figures are usually better than published ones, and it is why the Mongol count is more usable than a 2026 press briefing.

Modern casualty figures invert all three properties. No physical token. The compiler is a belligerent or is funded by one. And the consumer — a public that cannot verify and will not reconvene — bears no cost for believing a wrong figure, so nothing in the loop corrects it.

Gabriel supplies the symmetric medieval case, which proves the mechanism rather than the era. Western commentators "exaggerate[d] the size of Mongol armies" in order "to explain their defeats," because being outnumbered is a better excuse than being outgeneralled. **Same defect, opposite direction, same cause: the number was being produced for an audience instead of for a decision.**

4 · Why 2,749,163 is the least trustworthy number on the page

Precision and accuracy are different claims, and significant figures are a statement about method. Seven of them asserts that the counting procedure could resolve individual human beings. It could not.

Where the digits come from is the problem. They come from a bureaucracy that was itself measured on its paperwork — the same apparatus that bid arrest quotas upward to show enthusiasm. Such a record is *internally consistent and externally incomplete*, which is the exact combination that produces false confidence: every column adds up, and the columns are the wrong columns.

Applebaum says so twice, in one paragraph. The figure is "incomplete, and does not cover all categories of prisoner in every year." And: "Even if it were complete, however, this figure still would not reflect all of the victims of the Stalinist judicial system." She cites it "reluctantly" — and the reluctance is well-founded, because a precise number travels and its caveats do not.

Her real objection is prior to the arithmetic. "'How many died?' is in fact an imprecise question… those who ask such a question should first consider **what it is that they really want to know.**" Camp deaths only? Exile villages? Executions outside the camps? The famine? **Each boundary gives a different answer, and the answers differ by more than the precision of any of them.** A number without a stated boundary is not a measurement, it is a vibe with decimals.

Compare the two numbers on this board directly.
65 battles — rounded, from a hostile archive, with an undefined category ("pitched battle"). Its roundness advertises its error bar.
2,749,163 — exact to the person, from an interested archive, with an undeclared boundary. Its last six digits conceal its error bar.

The rounder number is the more honest one. Reporting more precision than your method supports is a false claim that contains no false statement — which is why it survives fact-checking.

What the four points share

All four are failures of the same kind: a number that is well-formed but not well-posed.

The ratio is well-formed and does not measure what wins. Stalin's figure is well-formed and the sets intersect. The modern casualty count is well-formed and nobody in the loop pays for its error. The Gulag total is well-formed to seven digits and its boundary was never stated.

> *In each case the arithmetic is fine. The question was broken before the arithmetic started — which is why Applebaum's instinct to interrogate the question, rather than improve the estimate, is the only move that works.*