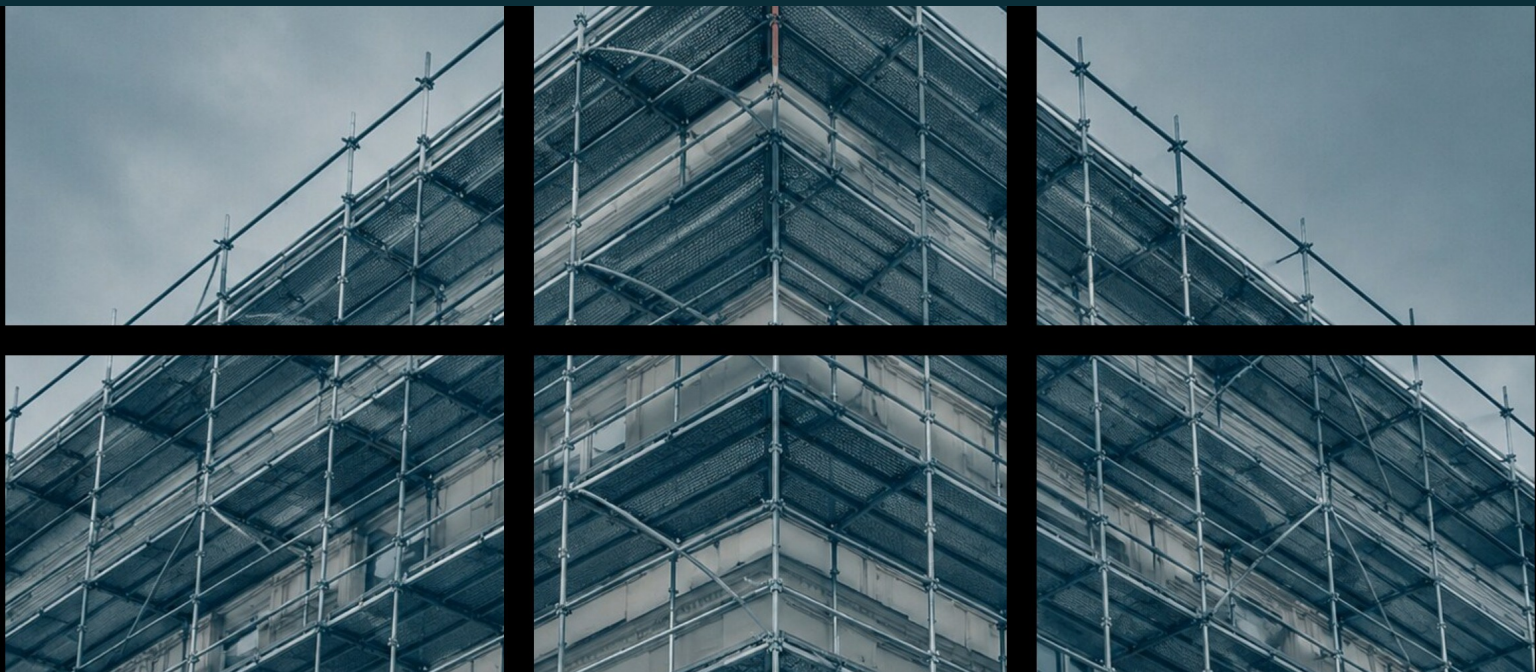


BEYOND THE COST LINE

The Consultant Cull — the Cheapest Diagnostic No One Reads

External spend is always the first lever in a downturn. What it exposes about a company's own capability is the part no one stops to read.



The Consultant Cull – the Cheapest Diagnostic No One Reads

Transformation · No. 155

Originally written December 2008 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“The consultant cull is the cheapest diagnostic a company will ever run on its own capability – and most organisations throw the result away unread, because they have decided in advance that it is only a cost story.”

1. The memo that brings relief

Every downturn has a document that circulates faster than any other. In this one it is the spend freeze – two paragraphs from the finance director, sent late on a Tuesday, confirming what the corridors already knew. Recruitment is paused. Travel needs sign-off two levels up. And, in the line that does the real work: all external and contract spend is to be reviewed immediately, with the presumption that it stops.

By Friday the day-rate contractors have cleared their desks. The consulting engagement that was "business critical" in September is, by December, a discretionary line with a notice period. There is a particular relief in this. No redundancy consultations, no compromise agreements, no damage to the morale of permanent staff who keep their jobs. A contract simply ends. On the spreadsheet the saving is immediate and clean, and in a quarter where cash has become the only number that matters, clean savings are worth more than their face value.

I have watched this sequence run twice now – once as the dot-com money drained out at the start of the decade, and again this autumn, faster and harder. What strikes me is not that organisations cut external spend first. Of course they do; it is the most reversible, least painful lever they have. What strikes me is how few of them read what the cutting reveals.

2. The cull is a measurement, not just a saving

Here is the claim I want to argue. The consultant cull is the cheapest diagnostic a company will ever run on its own capability – and most organisations throw the result away unread, because they have decided in advance that it is only a cost story.

When you pull every external person out of an organisation in the space of a fortnight, you are running an experiment you would never be allowed to fund deliberately. You are asking a single question: what did these people actually do? And the answer sorts them, cleanly, into two kinds.

The first kind were augmenting capacity. They were extra hands and extra hours on work the organisation understood perfectly well and could do itself, given time. Pull them out and the work slows, backs up, waits – but it does not stop, and it does not lose its shape. The knowledge stayed in the building.

The second kind had become load-bearing. Somewhere over three or four years of steadily outsourcing not just effort but judgement, they had come to hold the operating model in their heads. They knew why the reconciliation ran the way it did, which exceptions mattered, who to call when the overnight batch failed. Pull them out and something structural gives way. The organisation discovers, in the worst possible week, that it had quietly stopped doing its own thinking and started renting it.

The distinction that matters in a cull is not day-rate versus salary, or external versus permanent. It is augmentation versus dependency — and the freeze is the only time the organisation ever gets to see which is which.

The tragedy of the blind cut is that it treats both kinds identically. The same red pen goes through the contractor who was genuinely surplus and the one who was, in all but name, running a critical process. The first cut is pure saving. The second is not a saving at all; it is a liability crystallising, on a timetable you did not choose.

3. Two back offices

Let me make this concrete, because the abstraction hides the cost. Consider two financial-services operations I would recognise anywhere, both running multi-year change programmes, both told in November to take external spend down hard.

The first cut by the numbers. Roughly forty per cent of the programme "team" were day-rate contractors; the freeze took all of them by month-end, booking an annualised saving north of £1.2 million. It looked like disciplined management. Within six weeks two of the four workstreams had stalled — not for want of hands, but because the only people who understood the migration logic for the legacy ledgers had been among the first out. The programme director spent January re-engaging three of them through a supplier, at a premium, on emergency terms, to salvage a go-live that slipped a quarter regardless. The £1.2 million saving became roughly £800,000 of re-engagement and a delay whose cost no one wanted to total up.

The second operation did something that looked, at first, like hesitation. Before cutting, it spent two weeks — no more — mapping every external role against a single question: if this person leaves on Friday, does the work slow, or does it stop? Twelve of the fifteen contractors were augmentation; they went, on schedule, and the saving was real. The remaining three were keystones. Those three were kept on sharply reduced terms for ninety days with one contractual deliverable: not to do the work, but to document it and teach it to named permanent staff. By spring that organisation had a lower cost base and the knowledge internalised. It had used the crisis to buy back a dependency it had been accumulating for years — at a discount, because in a downturn the supplier has no leverage.

Same lever. Same intent. One firm cut a cost; the other cut a cost and closed a vulnerability. The difference was a fortnight of reading the diagnostic instead of throwing it away.

4. The objection worth taking seriously

There is a serious argument against everything I have just said, and it deserves stating at full strength rather than as a straw man.

It runs like this. In a liquidity crisis you do not have time for nuance. Cash is oxygen. External spend is the fastest, cleanest lever you own — no consultation periods, no legal exposure, no hit to the morale of the people you are keeping. When the bank is watching your covenants weekly, a fortnight of role-mapping is a fortnight you may not have, and a "keystone retention" programme is just a softer word for failing to cut. Speed is itself a form of prudence; the organisation that agonises over which contractors to keep may not survive to enjoy its own sophistication.

I take this seriously because the first half of it is simply true. Speed matters, and external spend is the right first lever. But the conclusion does not follow. The cull is going to happen either way and on roughly the same timetable; the two weeks of triage do not delay it, they run alongside the decision that was already being made. What the triage changes is not the speed of the cut but its aim. And the asymmetry is brutal: the cost of the fortnight is a fortnight of a few managers' attention, while the cost of getting it wrong is a stalled programme, an emergency re-engagement at a premium, and a slipped delivery in the middle of the worst quarter most of us have worked through. You do not skip the triage to save money. You skip it and it costs you far more, later, when you have the least room to absorb it.

“Cash is oxygen — but so is the knowledge of how the business actually runs, and the cull is the one moment you get to see who is holding it.”

5. Cutting with a diagnostic eye

So what does it mean to cut well rather than merely fast? Not to cut less — the discipline of the downturn is real, and the external line should carry the first and heaviest load. It means spending a small, fixed amount of attention converting a blind cut into an informed one.

- Ask the one question that sorts the population: if this role vanishes on Friday, does the work slow or does it stop? Slowing is augmentation and can be cut without ceremony. Stopping is dependency, and dependency must be handled, not merely removed.
- For the handful of genuinely load-bearing roles, make the cut conditional on transfer. Retain them briefly and cheaply, but change the deliverable: the job is no longer to do the work, it is to leave the work behind in a form your own people can run. You are buying knowledge, not effort.
- Use the leverage the moment hands you. In a downturn the supplier's negotiating position has collapsed; the retention terms you could never justify in a boom are suddenly available at a discount. The crisis is the cheapest time you will ever have to internalise a capability you should never have rented.
- Write down what the cull taught you. The map of where you had become dependent is the single most valuable artefact the downturn produces. It shows you exactly

where your operating model had hollowed out — and it will be worth more than the savings when you come to rebuild.

None of this is slower than the blind cut. It is the same cut, aimed.

6. The recovery you are already choosing

The freeze memo reads as a decision about cost. It is really a decision about what the organisation will be when the money comes back — and it will come back, as it did after every downturn any of us has worked through.

The blind cutter will spend the recovery rehiring the same dependency at boom-time rates, having learned nothing except that consultants are expensive when you need them urgently. The organisation that read its diagnostic will come out smaller, cheaper, and — for the first time in years — in possession of its own operating model. Both cut external spend first. Only one of them let the cutting teach it something.

The consultants going first is not the interesting part of a downturn. What they leave behind when they go — the shape of the hole — is the part worth reading. Most organisations are too busy congratulating themselves on the saving to look at it. That is the real waste, and unlike the consultants, it never shows up on the spreadsheet.