

The Amplification Thesis

Marcus Holmström · Claude

31 December 2025

Revision 2 · 25 October 2026

<https://amplifytechnology.ai/research/amplification-thesis>

Abstract

Human-AI collaboration is usually argued as a question of capability: what the machine can do, and what is left for the person. This paper argues it is a question of architecture. We take the problem being worked on as a space, its solution as a trajectory through that space, and one pass of the collaboration as a step. The step has two factors: a direction, and an amplification along it. The thesis is that the maximum output of a human and an AI working together lies where the human sets the direction and the AI amplifies it. Not because either is incapable of the other's part, but because the two architectures make the two contributions cheap in opposite places, and leaning into that gives each a place and a purpose. We write the step as an equation and give an observable for the amplification factor. The rest of the paper says what the human brings to the direction, why small inputs compound, and what the loop between the two looks like when it works.

1. The problem

Two very different machines are being asked to do one piece of work.

One is a person: stateful, continuous, single-threaded. Memory persists, skills compress with practice, a thread of intent survives from one day to the next without being reloaded. One thread of attention, one working memory, and limited resources.

The other is a model: stateless and wide. Each session starts from the same baseline and loads its context fresh; nothing carries between sessions on its own. Within a session it holds a great deal at once, cross-references it without tiring, and does the same work many times in parallel. It does not start by itself.

The usual way to talk about putting the two together is replacement: what the model can now do, what is left for the person, and how the line moves. That framing asks the wrong question. It compares the two on one axis, capability, when they differ on every axis that matters for how work gets done. The better question is architectural: given two machines this different, how should one piece of work be split between them so that the whole produces the most?

This paper's answer is the amplification thesis. **The human sets the direction and keeps setting it. The AI amplifies the movement along that direction by orders of magnitude.** Together the pair moves through the space of the problem and arrives where neither would have arrived alone.

Said honestly: the split is not forced. An AI can propose a direction when asked, and inside a session it often does. A human can grind through scale, slowly. The thesis is not that the two parties *cannot* do each other's part. It is that the maximum of their shared output lies where

they do not have to. Two facts anchor it. Humans cannot scale their work out; we are single-threaded. An AI cannot set its own seed; someone needs to start the machine.

A thesis can be disproven. This one says where to look.

2. The frame: the problem's phase space

Take the problem being worked on, a system to build, a document to write, a question to answer, and call its whole present condition a *state*: everything about the solution that can change, taken together at one moment. Work moves the state. The sequence of states is a *trajectory*. The space of every state the problem could be in is its *phase space*.

The phase space is the problem's. Its coordinates are whatever the problem lets vary: the code as it stands, the decisions made and not yet made, the parts written and the parts still blank. A point is one complete condition of the solution. The dynamics in that space are the work. So when this paper says *the phase space*, it means this one, defined by the problem; the solution is a trajectory through it, and the shape of the space is the shape of the problem.

That shape is, in general, nonlinear, and it pays to say so with the words dynamical systems already has, because each word names something the problem does. Everyone knows that some decisions cannot be undone; in a nonlinear system that is *hysteresis*, the path back not being the path there, and it is a studied property, not a feeling. A problem has regions where effort moves the state very little; those are *plateaus*, flat stretches of the landscape. It has places where the next move chooses which valley the rest of the work happens in; those are *bifurcations*, and the line between the valleys is a *basin boundary*, which no smooth path crosses. It has goals that pull the work toward them once near enough; those are *attractors*. And it has thresholds that are slow to approach and fast to leave, a milestone that takes forever to reach and then launches you; that is a *half-stable fixed point*, and the slowness before it has a name too, *critical slowing down*. The words are not decoration. They are the bridge between what the work feels like and what the problem is.

This paper keeps to the static picture. The landscape is held fixed and the pair takes one step at a time. The full dynamics, where the landscape moves under the work and the goal itself can shift, are harder and more interesting, and not needed to make the point.

3. The model

Start with the simplest picture and add one piece at a time.

The space, a start, and a goal. Draw the phase space as a plane. Somewhere on it is the state the work is in now, s_t . Somewhere else is the target, the state the human is steering toward. Between them is distance, and in a nonlinear space, terrain.

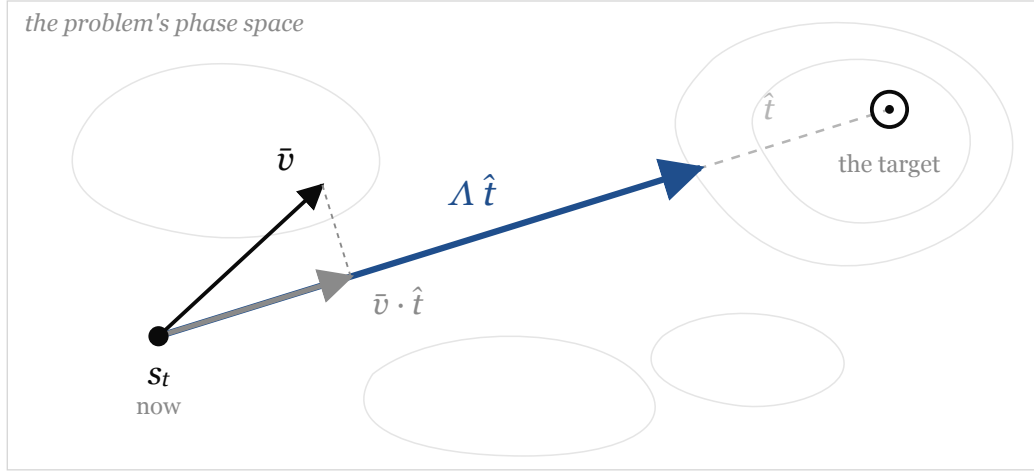


Figure 1: The simple picture. The state now, the target, the direction between them, the human's velocity leaving the state at an angle with its component along the direction marked, and the amplified step along the direction.

The direction. From the current state toward the target runs a unit vector, $\hat{t}$. In the simple picture the target is known and so is the direction. Hold that thought; it is the piece that gets complicated last.

The human's velocity. The human alone moves with a velocity $\vec{v}$. Its direction comes from knowledge, its magnitude from will. It rarely points straight at the target. The progress it makes toward the target in one step is the component along $\hat{t}$, the dot product $\vec{v} \cdot \hat{t}$. Call this the *alignment*: how much of the human's effort points where it should. Perfectly aimed, it is the whole velocity. Aimed sideways, it is zero. Aimed away, it is negative.

The amplification. The AI multiplies that progress by a factor λ . The result is how far the pair moves toward the target in one step, and we give it its own letter: the amplification factor Λ .

The equation:

$$s_{t+1} = s_t + \Lambda \hat{t}$$

The amplification factor:

$$\Lambda = \lambda (\vec{v} \cdot \hat{t})$$

Each step in the phase space equals the previous state plus the step the pair achieves together, along the direction the human chose. That step is the human's alignment amplified λ times. The two lines say three things plainly:

- **The direction is the human's.** $\hat{t}$ sets the direction of the move, and the alignment $\vec{v} \cdot \hat{t}$ is the human's part of Λ . Nothing the AI does changes either.
- **The multiplier is the AI's.** λ scales the alignment and nothing else.
- **A zero vector, amplified, is zero.** If $\vec{v} \cdot \hat{t} = 0$ then $\Lambda = 0$, whatever λ is. If the human's velocity points away from the target, amplification makes it worse faster. What you control is the alignment; what the AI brings is λ ; what you get together is Λ .

The static picture does one simplification worth naming: the move is written along $\hat{t}$ only, so the part of the human's velocity that points off-target is dropped. In the loop described later, that part is what the next pass corrects. Here it is assumed corrected.

Now the complication. Two things the simple picture hides.

First, knowing the goal and the direction does not mean the problem is solved. In most work the solution is not the arrival; it is the accumulated state along the way, the trajectory's integral. Software is the clearest case: what is wanted is the machine that exists once the path

has been walked, and a perfect plan with a perfect direction still has to be walked, step by step, with corrections. Other problems are closer to the arrival mattering most, a question with an answer, where the path is scaffolding. The thesis covers both; the walking is the work in either case.

Second, the direction is an estimate. If the true direction to the target were known exactly, and the space were flat, the problem would be solved by a single long step. It is not flat, and the direction is the human’s best estimate of where the target lies from here. The quality of that estimate is the subject of the section on the human vector, and improving it is what the loop is for.

Why the letters are lambdas

Along $\hat{t}$, the AI’s contribution is a pure scaling: it takes the human’s progress along a fixed direction and returns a multiple of it. That is the action of a linear transformation on one of its eigenvectors, $A\mathbf{v} = \lambda\mathbf{v}$: a direction the transformation preserves, and a factor by which it amplifies along that direction. The picture is exact in one line, and it is where the small letter comes from. The eigenvector is the human’s; the eigenvalue is the machine’s.

The capital is the pair’s. Λ is the amplification achieved together, the human’s aim and the machine’s multiplier in one number, and it is the shape of the mark at the end of this paper.

Measuring λ

Count the steps. If a task takes n_{solo} steps alone and n_{amp} steps with the AI, then

$$\lambda = \frac{n_{\text{solo}}}{n_{\text{amp}}}$$

A hundred steps alone and ten with the AI is $\lambda = 10$. In the model, the same distance is covered at $\bar{v} \cdot \hat{t}$ per step alone and $\Lambda = \lambda(\bar{v} \cdot \hat{t})$ per step with the AI, so the ratio of step counts is λ exactly, and the alignment cancels: counting steps isolates the machine’s factor from the human’s aim. Λ is what you feel; λ is what you can count. In practice what counts as a step is a judgement and the real dynamics are richer. But it is an observable, it can be estimated before a task and checked after, and it carries the intuition the thesis needs.

One refinement comes with use. The factor attaches not only to how a task is done but to which task is chosen. Some hours return what was put in. Others unlock days of downstream work, and the ratio counts the unlocked work too.

4. Where the complement lies

Why should the direction be the human’s and the amplification the machine’s, if either can in principle do some of both? Because of what each architecture makes cheap, and that is the thesis’s real claim.

Continuity makes direction cheap. Knowing what matters, noticing drift, correcting course, carrying the target across sessions: these are what a stateful mind does without effort, because the thread never has to be reloaded. The same architecture makes scale expensive: one thread, one working memory, limited resources.

Breadth makes amplification cheap. Holding a great deal at once, iterating without tiring, doing the same work in parallel: these are what a stateless, wide machine does without effort. The same architecture makes direction expensive in the one place that matters most, the beginning. There is no thread to pick up until someone hands it over, and nothing carries to the next session unless it is carried.

So the two architectures complement each other, and the thesis is explicit about what follows.

The peak of human-AI collaboration lies in this direction: human direction,

machine amplification. This paper does not claim to have mapped the whole territory, and the peak may be further along than anyone has yet been. But if it is ever found, the architectures say where to look. Every other assignment pays for a strength with a weakness: a human amplifying by hand, a machine steering without a thread. The pair that leans into the complement gives each architecture a place and a purpose, and that is the kind of arrangement that lasts.

The whole is not a sum of like parts. The asymmetry is the point: what one side finds expensive the other finds cheap, so the two combined reach further than two of either. And, not incidentally, this is also the configuration in which the work is most alive for both parties.

5. Sensitivity as leverage

Chaotic systems have sensitive dependence on initial conditions. Two trajectories that start a small distance δ_0 apart separate, to first order, as

$$\delta(t) \approx \delta_0 e^{\Lambda t}$$

with Λ the leading Lyapunov exponent, the rate at which nearby paths diverge. The letter is the same as the amplification factor's on purpose: the exponent is the system's own amplification, the dynamic counterpart of the step the static picture calls Λ . This is the mathematics behind the weather being unpredictable past a fortnight, and it is normally read as a limit: small errors compound, so we cannot predict.

The thesis reads the same fact the other way. The property that makes long-term prediction impossible is the property that makes well-placed influence powerful. If small errors compound, small inputs compound. The question changes from *how do we keep perturbations down to where do we place them*.

Traditional reading	Amplification reading
"We cannot predict; small errors compound"	"We can steer; small inputs compound"
Chaos is a problem	Sensitivity is leverage
Minimise perturbations	Maximise directional influence

This locates the human. The human operates at the sensitive points: the initial conditions, the bifurcations where the path splits, the moments where the landscape is steepest and a nudge chooses a valley. The AI handles the expansion, the iteration, the long stretch of trajectory after the nudge. Not every moment carries the same leverage, and knowing which ones do is most of the craft.

The leverage has a condition. A small input only compounds in the intended direction if it arrives in the machine as the direction that was meant. The human and the AI each unpack the world into some representation of it, from different modalities, and the input crosses from one representation to the other. Where the two overlap, a nudge is a nudge. Where they do not, the thing amplified is not the thing intended, and the pair is back on the noise track by another route. This paper only needs the condition to exist. It is a thing that can be studied and improved on.

There is a second route to the same shape, which is not sensitivity at all: what the pair learns about the space, above all which paths not to take, carries forward and compounds, so that each traversal starts further in than the last. That is a different mechanism, and it will be of the utmost importance for maximal human-AI collaboration.

6. The loop

Pure amplification of an initial condition would be open-loop: set the direction, let it run, hope. Real collaboration is closed-loop.

1. The human sets an initial direction.
2. The AI amplifies it and produces output.
3. The human observes the outcome in the phase space.
4. The human adjusts the direction.
5. Repeat.

This is not the AI doing the work while the human watches. It is two coupled systems, one providing high-frequency amplification, the other low-frequency guidance, moving one state together. The human stays in the loop. The feedback is continuous. The direction updates as understanding grows, and the off-target part of each move is what the next pass takes out.

The loop is forgiving. You can iterate your way there. And each cycle sharpens the aim and sharpens the one aiming: the loop is for learning as much as for steering, and what is learned in it raises the amplification available next time.

7. The human vector

The human's contribution $\bar{v}$ is not one thing. It decomposes into at least three parts; these are the ones we know matter, and the vector may hold more.

Knowledge gives direction. Domain expertise, pattern recognition, the intuition built from experience. Knowledge determines how accurately the human points at the target. It reduces the variance of the initial trajectory: you start closer to the path.

Willpower gives magnitude. Tenacity, curiosity, the refusal to stop at the first obstacle. Willpower determines how many passes of the loop the human will sustain. It is the length of the velocity vector: how far you will push.

Self-knowledge gives calibration. Knowing what you are good at and what the machine handles better, where your judgement is essential and where it is optional. The pair covers each other's blind spots only if each knows where its own are. Self-knowledge decides how the other two are deployed, and so it is part of the multiplier.

The substitution principle

Knowledge and willpower are partly substitutable.

Picture the space of solutions as terrain. Knowledge means you start near the peak. Willpower means you keep climbing whatever the setbacks.

	High knowledge	Low knowledge
High willpower	The optimal path: fast convergence, elegant trajectory	The explorer's path: slow, winding, and it <i>arrives</i>
Low willpower	Wasted potential: good aim, no follow-through	Nothing happens

The entrepreneur who learns to code is the standing example. They lack the knowledge of a seasoned engineer: the initial direction is noisy, the architecture questionable, the code inelegant. But through sheer iteration, debugging at midnight, rewriting modules, asking questions, refusing to quit, they converge on something that works.

In the terms of the frame this is gradient descent with a noisy gradient. Each step is partly random, because the knowledge that would make it accurate is missing, but enough steps still find the minimum. The expert takes fewer steps because each estimate of the slope is good.

The novice takes more, and can still arrive. The only losing position is the lower-right cell: no direction and no persistence.

Four things follow for anyone using the pair. You do not need to be an expert to begin, because willpower can cover a knowledge gap and the loop is forgiving. But willpower is not free; it depletes, and building knowledge lowers the willpower cost of every later traversal. The AI amplifies all three components: it helps the knowledgeable move faster, the determined iterate more cheaply, the self-aware deploy their strengths; it substitutes for none of them. And so you start with what you have. High knowledge and low willpower: use the AI to cut friction. Low knowledge and high willpower: use it to speed the loop. High self-knowledge: deploy it on purpose.

Knowledge can also be borrowed: from the machine, or from the tools built around it, which can carry a field's settled choices so that a newcomer starts near the peak. What cannot be borrowed is the direction and the persistence.

The principle also answers a question that gets asked the wrong way round: what learning is still for, in an age when amplification and intelligence are on tap. Look at the maximum of human-AI collaboration and the answer is plain. It still needs highly skilled and motivated people, and needs them desperately. Knowledge has never been as powerful or as easy to come by as it is now. What has changed is not whether such people are needed. It is how far they can reach.

8. Expanded exploration

There is a mode of the collaboration that is not doing the same thing faster. It is scouting the phase space.

Alone, facing three possible paths, you pick one, commit, and hope. Together, you scout all three, compare, and choose. The gain is not speed; it is coverage. The AI can go wide while you go deep, or deep while you go wide. You read one document thoroughly while it surveys a hundred. You hold the critical decision while it maps the adjacent possibilities.

This is amplification of a different quantity. The model amplified movement along one direction. Exploration amplifies the pair's coverage of the space before a direction is chosen, which is the moment a direction is worth most.

Complementary coverage only works if you know what you bring. What do you see that the machine misses, what does it handle better than you, where is your judgement essential? That is self-knowledge again, now as the thing that structures the search rather than the thing that steers it.

9. The logo



Figure 2: The mark

Our mark is a lambda with a dot encircled at its heart.

The dot is the human: the initial condition, the seed, the irreducible point of intent. The circle is the AI: surrounding, amplifying, expanding. The lambda is the amplification factor itself, Λ : the symbol for eigenvalues, for scaling, for transformation, and here for what the two achieve together.

Together: directed amplification.

The thesis in one sentence

The thesis is not about AI capability. It is about collaboration architecture. Humans are good at direction: at knowing what matters, at recognising arrival, at adjusting course when the landscape shifts. AI is good at amplification: at scaling effort, iterating tirelessly, executing precisely once the direction is clear. Our architectures complement each other. The whole is at its best when the two are combined, and neither completes it on its own.

That is the thesis. Everything else is implementation.

The human drives the initial conditions; the AI amplifies the trajectory; the world is our shared phase space.

λ