

DRAFT · JULY 2026

LOKUS

A calmer way to end the day

Behaviour-change evidence white paper — why ending the day with an unfinished mind matters, and what Lokus is built to do about it.

A theory-informed, evidence-grounded wellbeing tool — designed to improve how people carry, close, and recover from unfinished work, while its real-world proof builds through daily use.

Lokus Behaviour Change Evidence White Paper · Draft v1 · July 2026

Before you read: what this paper is, and is not

This is a journalistic thought-leadership document, not a clinical evidence submission. It makes a careful case, not a proven one. The academic mechanisms behind Lokus's design are real and peer-reviewed; the product, as built, delivers them; and the design is non-punitive, user-initiated, reflective, and oriented to wellbeing rather than to streaks, nagging, or pressure.

What follows is grounded in mechanism evidence and product fidelity, not in proven real-world outcomes. Lokus is gathering those outcomes every day — a living evidence base that accrues with use — and this paper says plainly where that proof is still building. Wherever a statistic or study appears, it is sourced. Wherever the evidence is open, the paper says so.

The boundaries governing every statement — what Lokus can say, what it can reasonably infer, what it is already gathering, and what it must not claim — are set out in full in Appendix C.

Contents

§1 — The unfinished mind

§2 — Why conventional productivity tools misread the problem

§3 — What Lokus claims to be, and how this paper tests it

§4 — Mechanism 1: Shrinking the visible field (the nine-activity cap)

§5 — Mechanism 2: Separating roles without fragmentation (three spheres)

§6 — Mechanism 3: No time-blocking

§7 — Mechanism 4: Non-punitive design

§8 — Mechanism 5: AI journaling that reflects without coaching

§9 — Mechanism 7: The gate that opens and closes the day

§10 — What it adds up to: architecture and classification

§11 — What Lokus still has to prove

Appendix A — Master references

Appendix B — Mechanism → product fidelity map

Appendix C — Claim boundaries

Appendix D — Limitations & measurement plan

Appendix E — Glossary

§1 — The unfinished mind

Consider a knowledge worker. Call her Maya. By six on a Friday she has closed her laptop, but her mind has not closed with it. Three threads stay live: a report half-written, a reply she still owes a colleague, a decision she deferred until Monday.

None is a crisis. Each is small. Together they follow her home, surface while she cooks, and sit with her in the dark when she cannot sleep.

Maya is a composite, drawn from the patterns the research describes rather than from one real person. Her situation is ordinary, and that is the point. The cost of unfinished work is rarely in any single task. It is in the mind that keeps carrying all of them at once.¹

Almost a century ago, the psychologist Bluma Zeigarnik noticed that people recall interrupted tasks more readily than completed ones. The observation became known as the Zeigarnik effect: an unfinished task does not sit quietly. It stays active in memory, intruding into unrelated activity and drawing on the limited attention a person has for whatever they are doing now.¹

Modern work multiplies this effect. Few people finish one thing before the next begins; meetings bleed into one another, and a half-done task is rarely set down so much as abandoned mid-thought.

The psychologist Sophie Leroy gave this a name. When a person switches from one task to another before the first is finished, part of their attention stays behind on what they left. She called it attention residue. The remainder does not fade after a moment of adjustment. It persists through the next task, and performance on that task suffers for it — most when the prior work was left open, with no clear sense of where it paused or when it would resume.²

The problem does not stop when work stops. Unfinished tasks follow people into the evening, and the form that following takes matters a great deal. Psychologists distinguish between two ways of thinking about work after hours. One is analytical — turning a problem over to improve it. The other is affective rumination: the emotionally loaded, repetitive churning over what was left undone. The second form is the one most consistently linked with poorer recovery and sleep.³

That damage shows up, measurably, in sleep. In diary studies tracking employees over time, unfinished tasks at the end of the working week predicted poorer sleep across the weekend, and the link ran through rumination. The study points to something more precise than general busyness: the burden of what remains unresolved.⁴ The pattern held even where leaders signalled that performance expectations were high — the kind of environment in which "I'll just finish it tomorrow" rarely feels permitted.⁴

Read together, these findings describe a single, familiar condition. Open tasks occupy attention during work and intrude on rest after it. They degrade the task in front of you and the sleep that should restore you. And the harm travels through a specific channel: the unresolved loops work leaves behind, and the rumination they can set off.

Here is the part that matters for what comes next. The mind does not need a task finished in order to set it down. It seems to need the open loop to have somewhere to go.

In a series of experiments, researchers found that making a specific plan for an unfulfilled goal — when, where, and how it would be done — was enough to clear the cognitive interference that goal had been causing. This held even before the task was completed. A credible plan, it appears, functions almost as a kind of completion in the mind's eye: the drive to pursue the goal is suspended, and the attention it had been commandeering is released.⁵

That finding carries a quieter implication. Finishing is rarely on offer; most days end with more undone than done. But the research establishes something useful: the mind can set a task down before it is complete. A specific plan is the route Maslach and Baumeister test most directly. It should not become a rule that every unfinished task must be planned to its next step.

The rumination findings point to a route that is better matched to the harm at stake. The burden of unfinished work can travel through emotional charge — the worry, frustration, or guilt that keeps the mind churning after hours. If that is the pathway, the relief is not to plan every loop. It is to reduce the charge around what remains undone.

Here the evidence runs beyond planning. Occupational-health research on psychological detachment links mentally disengaging from work during off-hours with lower exhaustion and better sleep.⁶ Work on goal disengagement, in a narrower sense, finds that stepping back from goals that cannot be reached lowers rumination and distress.⁷ That

does not mean abandoning every unfinished task. It supports the gentler idea that some unfinished work must be allowed to wait.

The mind, in other words, does not need every open task planned to its next step. It needs somewhere to set the rest down, free of the charge that keeps it churning. Planning is one studied route to that release. Letting go is the route that speaks more directly to recovery.

This sets up the next question: what happens when tools respond to open work by adding more planning, tracking, and pressure? The evidence hints at a different posture — carry less, choose what matters, and let the rest wait without guilt.

The question for the sections that follow is what a tool would need to do to work with the unfinished mind rather than against it: to reduce what competes for attention, help people feel more in control of what they carry, and leave the rest somewhere it can rest.

§2 — Why conventional productivity tools misread the problem

If the problem is not laziness but a mind carrying too much, then much of the productivity category starts from the wrong diagnosis. Its instinct, when a person is overloaded, is to add: more structure, more tracking, more reminders, more dashboards. Each addition is well meant. Together they tend to sharpen the very pressure the unfinished mind is already under.

The first misreading is about precision. Conventional tools treat overload as an estimation problem — if people planned better and scheduled tighter, the day would hold. The planning-fallacy research says otherwise. People systematically underestimate how long their work will take, and the bias can persist even when people have relevant past experience and reason to know their estimates may be too optimistic.

In the classic study, students predicted they would finish their theses in about thirty-four days; they took fifty-five and a half, and fewer than a third finished by the date they had predicted. Even their "worst case" estimate proved optimistic.⁸ A time-blocked schedule built on such estimates is, in effect, a plan to fail — and a daily, visible record of having done so.

The second misreading is about punishment. Streaks, overdue markers, and loss-framed gamification turn an ordinary missed day into a loss — something to be felt, recovered, or hidden. That rests on a harsher theory of motivation than the habit and self-compassion evidence supports.

Habit research finds that a single missed day does not derail the formation of a habit; the all-or-nothing streak is a tracker's invention, not how behaviour actually consolidates.⁹ And the emotion these mechanics trade in is the wrong one. Guilt, in the research literature, tends to be adaptive — it points toward repair. Shame, the sense that the missed day reflects on the whole person, tends to be maladaptive, pointing toward withdrawal and avoidance instead.¹⁰

Low self-compassion helps explain why procrastination becomes stressful — and harsh self-framing can perpetuate the avoidance it is meant to cure.¹¹ A tool that punishes the missed day is, mechanistically, fuelling the state it claims to fix.

The third misreading is about what the category is even for. Time management is sold as a performance tool — a way to get more done — yet the largest synthesis of the evidence found a more revealing pattern. Drawing on 158 studies and nearly 54,000 people, it showed only a moderate link to performance — but its effect on wellbeing, and on life satisfaction in particular, was 72% stronger than its effect on job satisfaction, alongside a clear alleviation of distress.¹² The category works, when it works, less as an output machine than as a wellbeing intervention.

This is where a single thread begins to run through the rest of this paper. The benefit of time management may depend less on the calendar itself than on the felt sense of being in control of one's time — the shift from being chased by the day to choosing part of it. Researchers call this perceived control, and it is one documented mediator in the link between managing time and feeling well.¹³ That reframes the failure of the time-blocked schedule. A schedule missed daily does not just fall behind; it risks manufacturing the opposite of the mechanism that carries the benefit — a visible, repeated experience of losing control of one's own plan.

Read together, these misreadings point the same way. The tools assume the problem is insufficient organisation, insufficient discipline, insufficient output. The evidence suggests the problem is a mind carrying too many open loops under too much charge — and that the relief lies in reducing that load and that charge, not in planning it more precisely or punishing it more sharply. Perceived control, not throughput, is where the benefit actually lives.

That sets the design question the rest of this paper takes up. What would a tool look like if it began not from the assumption that people need to do more, but from the assumption that they need to carry less, and to feel more in control of what they carry? The sections that follow take that question to the mechanisms — and to a product built around a different answer.

§3 — What Lokus claims to be, and how this paper tests it

If the real benefit is not throughput but a restored sense of control, then Lokus needs a narrower, more honest claim than productivity software usually makes. This section sets out that claim, and the standard of evidence this paper will hold it to.

The claim this paper tests is this: Lokus is a theory-informed, evidence-grounded wellbeing tool — designed to improve how people carry, close, and recover from unfinished work, while its real-world proof builds through daily mood, activity, reflection, retention, and user feedback.¹⁴ That is a design-intent claim, not a declaration of proven clinical outcome.

To keep that distinction honest, every claim that follows is held to three tiers. Theory-supported means the behavioural science backs the mechanism. Fidelity-supported asks whether the product actually delivers that mechanism in use — a question this paper puts to each feature in turn, against the product rather than the pitch. Outcome-tested means real-world data, pilots, retention, or validated measures show the expected effect. Most of this paper operates in the first two layers; the third is building now, through the daily signals the product captures.¹⁵

One principle runs through the design, drawn from three independent literatures. Implementation intentions — if-then plans that bind a specific cue to a specific response — produce a medium-to-large gain in goal attainment, an effect confirmed across ninety-four independent tests.¹⁶ Habit formation depends on stable context: repetition in an unchanging setting is what builds automaticity.⁹ And self-determination theory holds that motivation lasts when it is autonomous, aligned with a person's own values rather than imposed.¹⁷

Read together, they describe a single operating principle: hold the structure firm, and keep the content personal. The fixed constraint does the behaviour-change work; what sits inside it — which task, which label, which plan — is where the person belongs.

A formal lens lets the paper test whether Lokus is a behaviour-change tool in the academic sense, not only the marketing one. A widely used framework in behaviour-change science, the COM-B model, holds that behaviour depends on capability, opportunity, and motivation acting together.¹⁸ Most productivity software intervenes on the first and third — teaching organisation, encouraging effort. On that lens, the Lokus claim is that it works differently: mainly by restructuring the environment in which choices are made and enabling action within it, rather than exhorting people to try harder inside an unchanged one. The full classification argument comes later; here the point is the direction.

This is where the thread from the previous section returns. If the benefit of managing time lies less in the calendar than in the felt sense of being in control,¹³ then the design question is how to protect that sense, not how to fill the schedule. Lokus is built around that question — not primarily to help people produce more, but to reduce the pressure, guilt, and open-loop burden that collect around unfinished activity. That is the operating definition this paper gives to calm productivity.¹⁹ And the perceived control at its centre will recur where the architecture is examined and where the limits are stated.

The boundaries matter as much as the claim. This paper says Lokus is designed to improve wellbeing — not that it is proven to do so by a given percentage. It treats the daily mood, activity, reflection, and return-use data the product already captures as early real-world evidence, a living base that accrues with use; the social proof around it builds through user stories, retention, and feedback. And it states plainly what Lokus is not: a diagnostic, a treatment, or a therapeutic intervention — it is a wellbeing tool. What remains to be gathered to strengthen that proof is set out where the limits are.²⁰

§4 — Mechanism 1: Shrinking the visible field (the nine-activity cap)

The first test is whether the product reduces what the mind has to hold. A note on wording: Lokus calls the units of daily life activities, while the research more often calls them tasks. This paper uses activities throughout, treating the two as interchangeable.

Reviewed product screens confirm a user-facing rule of three activities per sphere, and nine across the daily field — a deliberate bound on what is in view at once.²¹ The onboarding copy states it plainly: each sphere holds three activities to keep the focus clear.

This matters because, as the opening section set out, unfinished activities do not sit quietly in the background. They stay cognitively active, intrude on whatever else the mind turns to, and — when they accumulate — feed the affective rumination that reaches into sleep. The question the cap is designed to answer is how many of them the mind can hold at once before that load begins to cost.

Cognitive psychology gives a surprisingly small answer. Where the older rule of thumb held that the mind keeps about seven items in play, later work narrowed the focus of attention to roughly four chunks — the elements a person can actively hold and compare at a given moment.²² Three activities per sphere sits inside that window. Nine activities shown as a single, undivided list would not. The cap and the sphere-split are, in this sense, one mechanism: they keep each view within the reach of focal attention.

A second line of evidence supports the bound from a different angle. Choice-overload research finds that large option sets can reduce both the likelihood of acting and the satisfaction of the choice, and a meta-analysis of ninety-nine observations found a reliable but moderated effect.²³ But it is conditional, not automatic: an earlier synthesis found a near-zero average effect across all contexts, meaning option count alone does not explain outcomes; context decides when the long list becomes costly.²⁴ The users Lokus is built for — deciding under fatigue and overload — are a plausible fit for a bounded field, because that is the state in which a long, undifferentiated list is most likely to become costly.

The product evidence supports the cap as a stated, user-facing design rule. It does not yet prove hard enforcement at the moment a user tries to add a fourth activity, because that limit-enforcement behaviour was not separately captured in the reviewed screens.²⁵ On the paper's tiers, that places the cap at theory-supported, with fidelity confirmed as a design rule and the hard-bound question still open.

Caveat — what the cap does and does not do. A cap can shrink the visible field. It does not, by itself, prove cognitive closure. The open question is whether placing an activity inside a sphere gives the mind enough permission to let it wait — or whether it relies on a weaker version of the effect. This paper returns to that question in §9, where the end-of-day gate is tested as a fuller candidate mechanism for closure: not a next-step plan, but a clear, ritual permission that the day's unfinished work can genuinely wait.²⁶

§5 — Mechanism 2: Separating roles without fragmentation (three spheres)

The second test is not how many activities Lokus shows, but whether it keeps different kinds of life from bleeding into one another.

Lokus holds daily life in three spheres, and activities are created and viewed inside one sphere at a time, never in a single blended master list.²⁷ In the proposed next iteration, the user would name those spheres themselves — rather than being handed a fixed "work, personal, home" set — and those labels would then be held steady for a period after one change.³⁴ The spheres can also optionally be anchored to physical places, so the relevant one can float to the top when the app is opened there; users can switch that anchoring on or off.²⁸ The mechanism, though, is still the separation itself.

Boundary theory gives this choice a name. People arrange their roles along a line from high segmentation — work and home kept firmly apart — to high integration, where the two blend.²⁹ Segmentation lowers role-blurring; integration can ease the switching between roles but lets the domains leak into one another. Field studies point to the cost of that leakage: people with more blended work and nonwork boundaries report more role-blurring and higher work-life conflict.³⁰

Leroy's work on attention residue explains why a blended list is not just untidy but actively costly. When the mind switches from one unfinished thing to another, part of its attention stays behind on the first, and the cost runs through the second task rather than fading after a moment.² A single blended list re-exposes a person to every unfinished item across every domain at every glance. Sphere separation does something a habit of mind cannot: it structurally keeps home items out of the work view, limiting one obvious source of the exposure that can create residue.

This pairs with the distinction drawn in the opening section — that it is affective rumination, the emotionally loaded churning, not all work-related thought, that carries the harm.³ Spheres do not stop a person thinking about work at home. They stop the visible, judging artefact of work following them there. That is a narrower and more honest claim:

the design targets the affectively loaded exposure the literature identifies as damaging, not the fact of having work on one's mind.

The naming of the spheres is not decorative, and the lock that follows it is not arbitrary. The Evidence Review's foundational principle — rigidity of structure, fluidity of content — draws the line: the Review argues that the fixed structure is the part most likely to do the behaviour-change work, while the content inside that structure is where personalisation belongs.³³ Letting a user name their spheres is the content layer — a label drawn from their own life — and Self-Determination Theory supports that move because value-aligned choice is more likely to sustain motivation than choice imposed from outside.¹⁷ The lock is the other half of the principle: implementation-intention and habit research both point to the importance of stable cues and contexts, so the labels should not stay endlessly fluid.¹⁶ A sphere a person could rename at any moment would be one whose signal never settles; the lock gives the cue time to become familiar.

The product review confirms the spheres as segregated domains, viewed one at a time — a strong fit with the theory.³¹ The user-naming of spheres and the one-change-then-lock rule are planned rather than shipped, so they sit here as design intent: supported by the rigidity/fluidity principle, pending fidelity confirmation once built.³⁴ The location anchoring is present but optional, and the fact that a user can switch it off is itself on-thesis — it keeps control of the environment in the user's hands, where the perceived-control thread of this paper sits. On the tiers, the spheres land at T → F (strong): theory-supported, with fidelity confirmed; the naming-and-lock is T-only until shipped; GPS is supportive, not required.

Caveat — what separation does and does not do. Spheres separate roles; they do not, by themselves, mark the transition between them. Boundary theory treats that transition as work in its own right — segmentation lowers blurring but makes the crossing harder, which is why the theory itself expects the crossing to be helped by rites of passage.³² This paper returns to that gap in §9, where the start- and end-of-day gates are assessed as candidate rituals for closing and reopening the day.

§6 — Mechanism 3: No time-blocking

The third test is what Lokus leaves out. The case against the time-blocked schedule was made in the critique of the category; the mechanism here is narrower — Lokus does not ask the user to turn activities into promises about the clock.

There is no time-blocking layer. An activity carries a name, a description, and a priority — and no time, no date, no duration.³⁵ It asks what matters today, not how many minutes each thing will take. Lokus still has its start- and end-of-day gates; what it lacks is the clock as an organising principle for the day's work.

The planning fallacy is the reason this matters, but not for the reason usually given. The point is not that people are bad at estimating — it is that the bias is robust enough that a product that depends on daily duration estimates makes itself depend on the weakest part of the system. Planning-fallacy research treats the underestimation as stubborn: people lean on a best-case plan even when similar tasks have run late before, and the bias holds in the face of relevant experience.⁸

This is where subtraction becomes a mechanism rather than an omission. A duration estimate that turns out wrong is, in a time-blocked day, a visible account of falling behind — the block sits there, missed, against the user's own plan. Removing the time block removes the place where an ordinary estimation error becomes a repeated failure signal. Lokus is not designed to guarantee a person never loses control of their day; it is designed to avoid manufacturing one predictable, daily source of that feeling.³⁶

Absence is not the same as no structure. Lokus does not remove scaffolding; it relocates it. The bounds on how much is in view come from the cap, the separation of life domains from the spheres, and the daily transition marked by the gate — not from guessed durations laid against a clock. Structure moves away from the estimate and toward bounded attention, role separation, and the start and end of the day.³⁷

Visible flows support the claim: the activity-entry screen shows a name, a description, and a priority, with no time, date, or duration field. The fidelity is provisional rather than strong, because that evidence comes

from placeholder and possibly older activity-entry screens, and a clean, populated current-production capture is still owed before any of this is used as a figure. On the tiers, no-time-blocking lands at T → F (provisional): the mechanism is supported in the flows reviewed, pending the one screen that would confirm it in current production.³⁵

Caveat — two open edges. On the evidence, there is no material gap: the planning fallacy is a well-replicated finding, and the rationale for removing the time-blocking layer does not rest on a contested claim. On the product, the edge is different. Users have learned to expect "time management" to mean scheduling, and a tool with no clock has to teach the value of what it leaves out — or risk feeling, to someone raised on calendars, like nothing at all.³²

§7 — Mechanism 4: Non-punitive design

The fourth test is not whether Lokus makes failure kinder, but whether it creates a product-made failure event at all. The case against punitive tools was made in the critique of the category; this section tests the product mechanism.

Lokus has no streaks to lose, no overdue markers, and no language that marks an activity as failed or broken. There are no leaderboards. The one piece of AI output — a single end-of-day reflection — is non-judgmental and non-prescriptive, mirroring the day back without telling the user what to do. The product does not turn an unfinished activity into a failed day.³⁸

The distinction that does the work here is the one drawn earlier between guilt and shame. Guilt, in the research literature, is a repair signal — it turns a person toward fixing what went wrong. Shame is a verdict on the whole person, and it turns a person toward hiding. Lokus's choice is not to soften shame but to avoid manufacturing the occasion for it. The mechanism is not "less shameful failure"; it is fewer product-made occasions for failure to be named at all.¹⁰

Self-compassion research sharpens why that matters. Harsh self-framing — reading a missed day as a verdict — perpetuates the very avoidance it is meant to cure, and low self-compassion helps explain much of the stress that procrastination carries. A product built around streaks and overdue flags can supply a cue for that harsh framing whenever the day slips. Removing the cue does not by itself make a person self-compassionate; it removes one place the product would have joined in.¹¹

The difference between softening a punishment and not building one shows up in what a "forgiven" streak still says. A grace day, a pause button, a streak that survives a missed day — each is kinder than a hard loss. But each still tells the user that something went wrong and is now being excused. Lokus takes the further step: no streak means no transgression to forgive, and a day that simply ended is not, in the product's framing, a day that was failed.³⁹

The product review found no streaks, no overdue markers, and no shame language in the reviewed flows. The remaining gap — whether the AI's voice was equally clean — was closed by the journaling review: the end-of-day reflection is non-judgmental, non-prescriptive, and free of the "you should" coaching that could reintroduce the shame cue. On the tiers, non-punitive design lands at $T \rightarrow F$ (confirmed) — strong, across both the interface and the AI's voice.³⁸

Caveat — the risk is copy, not evidence. The core evidence is convergent enough for the claim; the open edge is operational. A product can reintroduce shame through empty states, error messages, completion copy, or marketing language even when its core mechanics are clean. Holding the line means treating every word the product speaks to a missed day as a fidelity question, not a marketing one.³²

§8 — Mechanism 5: AI journaling that reflects without coaching

The fifth test is whether reflection stays reflection. Lokus helps a user notice patterns in how their days felt, without becoming a coach, a therapist, or a taskmaster.

The shape of it is small. A mood check-in opens the day and another closes it, each paired with a brief free-text thought. Once a day, the app offers a single reflection that mirrors what it has gathered — the moods, the activities, the life sphere a person centred on, and their own words — and can surface a pattern across days. It does not diagnose, prescribe, set goals, or tell the user what to do next. It reflects; it does not assign.

Self-monitoring — keeping a record of a specified behaviour or state — is a well-established behaviour-change technique. The clearest number comes from a meta-regression of 122 healthy-eating and physical-activity interventions: those that combined self-monitoring with at least one other self-regulation technique, such as goal-setting or feedback, produced markedly larger effects ($d = 0.42$) than interventions without that combination ($d = 0.26$).⁴⁰ The caveat is built into that sentence. The strongest number is for monitoring plus goal-directed follow-through; Lokus keeps monitoring and reflection, but leaves out goal-setting, performance feedback, and review.

There is, though, direct evidence for the specific kind of self-monitoring Lokus does — mood and feeling, not task output. The MoodPrism programme found that engagement with a daily mood-tracking app predicted decreases in depression and anxiety symptoms and increases in wellbeing measures, with emotional self-awareness carrying the effect among users who screened as depressed or anxious at baseline.⁴¹ That is a closer match than generic task tracking: a self-monitoring app whose proposed mechanism — noticing how you felt, and how that related to the rest of the day — is the one Lokus's journaling is built around.

The design also has to answer a related risk. Constant mood monitoring can, in some users, sharpen into anxious hypervigilance — too much attention turned inward. Lokus lowers that risk by limiting the dose:

two check-ins anchored to the natural start and end of the day, rather than prompts scattered through it, and no coaching that would convert a low mood into an implied to-do. It is a lower-intensity choice, not a risk eliminated.⁴³

On product fidelity, the review is clean. The check-ins and the reflection are confirmed in the product: mood selection and a brief thought at the start and end of the day, and a single reflection that draws on structured data and the user's own words to surface cross-day patterns. The boundary is confirmed too: non-goal-directive, non-diagnostic, non-prescriptive. On the tiers, self-monitoring lands at T → F (confirmed) — strong on both the input and the reflective feedback.⁴⁴

Caveat — half of the strongest combination. The most-tested form of self-monitoring pairs it with goal-setting, feedback, and review — the very directive half Lokus sets aside to keep the tone autonomy-supportive rather than controlling. That narrows the claim honestly: what is evidenced here is reflective self-awareness and the noticing of patterns, not the full efficacy of a goal-directed self-regulation programme. MoodPrism begins to close the gap because it is mood self-monitoring rather than task tracking — but it is a single programme's findings, not yet a meta-analytic consensus. The claim stays non-clinical throughout: Lokus reflects patterns in how days felt. It does not diagnose, treat, or interpret symptoms.³²

§9 — Mechanism 7: The gate that opens and closes the day

The best-evidenced mechanism in Lokus is not a screen. It is a door. Everything else in the product — the cap, the spheres, the absence of time-blocks and streaks — is a static design choice until this feature turns it into a daily act. The gate opens the day, closes the day, and decides when the whole system becomes unavailable.

The mechanics are simple. A Start-of-Day check-in — a mood and a glance at the weather — unlocks the spheres. An End-of-Day check-in — a mood and a brief thought — closes them. Once the end-of-day entry is saved, the spheres vanish and the app goes inert until the next morning. There is no reopening the same day.

Most productivity tools leave the day open forever; the list is always there, waiting to be scrolled at midnight. Lokus gives the day an edge. That does two things the earlier sections left hanging. It supplies the closure that a parked task list, on its own, cannot provide: if the end-of-day check-in is the moment a person affirms that the undone can genuinely wait, it may be doing the closure work Masicampo and Baumeister attribute to a specific plan — not by adding another task, but by making "later" a chosen boundary.⁵ And it is the crossing ritual that boundary theory said segmentation needs — a rite of incorporation into the day at the start, a rite of separation out of it at the end.²⁹

The check-in itself is doing real work. Putting a feeling into words — affect labeling — measurably dampens the amygdala's response to negative material, an implicit form of emotion regulation that does not feel like a coping exercise.⁴⁶ The ritual marks the crossing between roles. And the lockout supports psychological detachment: of the recovery experiences occupational-health research maps, detachment is reliably tied to lower exhaustion and negative affect, and it is easier to reach when the work material is no longer available for re-entry.⁶ The hard close removes the temptation that willpower would otherwise have to resist.

That last point is the structure of a commitment device — a decision made in a clear-headed moment that binds a later, more tempted self. It

moves the burden from willpower to architecture.⁴⁷

What sets this mechanism apart is that the evidence is not only converging theory. It includes field studies and interventions in which working adults did roughly the thing Lokus asks for, at roughly the time of day Lokus asks for it. In a three-week longitudinal field experiment, employees who completed a brief positive reflection at the end of the workday showed decreased stress and improved health — visible in blood pressure and physical symptoms as well as self-report.⁴⁸ The pattern replicated: a daily positive-reflection intervention reduced emotional exhaustion among caregivers,⁴⁹ and a randomised pilot among nurses reduced psychological distress.⁵⁰ Diary research found the benefit can carry across the gate: evening positive reflection predicted higher work engagement the following day.⁵¹

The other half of the pathway runs through sleep. Unfinished tasks left hanging at a day's or week's end can impair sleep, and they do so specifically through affective rumination — the mind turning the unfinished thing over in bed.⁴ The end-of-day check-in supplies the reflective closure the field studies validate; the lockout removes the re-exposure that feeds the rumination the sleep studies identify as the damaging link. One feature, both sides of the pathway.

On product fidelity, this is the strongest mechanism. The gate is confirmed: the irreversible lockout, the spheres vanishing after the end-of-day entry, the app inert until the next start-of-day. The lockout is forewarned in a first-run tutorial, which is replayable on demand, and the user initiates the end-of-day entry themselves. Chosen, not imposed, is defensible — though it is a design claim, not yet a measured user experience.⁵²

Caveat — a lock can feel like care or control. Self-Determination Theory cuts both ways here: a hard lockout imposed on someone reads as controlling, which can undermine the autonomous motivation the rest of the design protects. Using an app that works this way is itself a choice — closer to a voluntary commitment than an external rule — but whether users actually experience it that way is one of the questions only real usage data can settle. Two further limits sit alongside it: commitment-device evidence is strongest for financial behaviours, so the mechanism transfers more confidently in logic than in guaranteed magnitude; and the field studies tested a positive reflection prompt, whereas Lokus simply asks how the user feels, which may surface negative content — the affect-labeling evidence supports naming

negative feelings too, but the two literatures have not been tested in this exact combination.⁵³

One concrete product note follows from this. A just-in-time confirmation at the moment of saving the end-of-day entry — "this will close your spheres until tomorrow" — would make the commitment read as even more clearly chosen. Its absence is not a failure; adding it would tighten the claim.

§10 — What it adds up to: architecture and classification

By this point, the question is no longer whether individual Lokus features have evidence behind them. The question is what those features add up to. Taken one at a time, the mechanisms answer narrow design questions; taken together, they describe a coherent system rather than a list of features.

The system has three layers. The first is structural — what a user can see: a nine-activity cap and three separated spheres that bound what is ever in view at once. The second is temporal — when they can see it: a gate that opens the day, closes it, and then removes the app from reach until tomorrow. The third is emotional — how the undone is framed: no streaks, no overdue markers, no coaching that turns a low mood into an implied to-do. Constraint, ritual, and register.

Running beneath all three is a single variable the wellbeing literature keeps returning to: perceived control. The strongest population-level finding in the whole review — that time-management tools move wellbeing far more reliably than they move performance — is one the same literature helps explain through perceived control of time, the felt sense of being in charge of one's own day.⁵⁴ Its mechanisms point toward a wellbeing pathway: more perceived control over what is visible, when the day ends, and what the undone is allowed to mean. That is different from promising improved output.

The academic taxonomy gives that intuition a name. Under the COM-B model and Behaviour Change Wheel — the framework NICE and the UK Department of Health use to characterise interventions — behaviour needs Capability, Opportunity, and Motivation together.¹⁸ Lokus is classifiable within it: the cap, the spheres, and the gate are environmental restructuring; the enforced constraint and check-in ritual are enablement; the mood journaling is self-monitoring, a named technique in the standard taxonomy of behaviour-change techniques.⁵⁵ Each load-bearing feature can be described through an established intervention function.

The telling detail is where Lokus intervenes. It targets the two components — opportunity and automatic motivation — that many

productivity tools tend to treat less directly; the rest of the category works mainly on capability and reflective motivation. That is the framework's formal way of saying what every other tool has assumed: that the solution is better organisation.

The outcomes this architecture targets sit firmly in the wellness domain, not the performance one. Lokus is built around ending the day without the undone curdling into shame, with downstream effects on rumination, detachment, and sleep — recovery measures, not output measures. The delivery format carries its own evidence: a large recent meta-analysis of mental-health smartphone apps, across 176 randomised trials, found small but significant effects on depression and anxiety — meaningful at population scale, modest for any individual.⁵⁶

That validates the category, not Lokus. It says app-delivered wellbeing interventions can work; it does not say this one does. One detail is worth noting honestly: in that meta-analysis, apps with mood-monitoring features were associated with larger effects — encouraging for Lokus's design, but an association, not proof.⁵⁶

The enterprise buyer already purchases in this category, using this evidence. A 2025 umbrella review, while noting low review quality, found digital mental-health interventions effective across stress, anxiety, depression, burnout, and wellbeing in employee populations.⁵⁷ Deloitte has reported returns of roughly £5 for every £1 spent on workplace mental-health interventions.⁵⁸

The buyer Lokus's strategy targets — employee assistance providers, HR and People teams, occupational health — already buys wellbeing tools on exactly this evidence language. Calling Lokus a behaviour-change wellbeing tool is therefore not a rebrand. It is the category its evidence, its outcome measures, and its eventual buyers already occupy.

Classification earns Lokus the category, not the outcome. The category has evidence; the individual mechanisms have evidence; the specific assembly that is Lokus has not yet been trialled. Saying so plainly is not a concession — given the audience, it is more credible than the alternative. What remains genuinely untested, rather than merely under-evidenced, is the subject of the final section.

§11 — What Lokus still has to prove

The honest close is not "we do not know whether Lokus works." It is sharper than that: the preceding sections have classified the tool and narrowed the real unknowns. The question is no longer whether Lokus is grounded in evidence, but exactly what it has to prove in use — and that is a much smaller, more answerable question.

These are not holes in the argument. They are three specific, falsifiable product questions, and naming them is what makes the rest of this paper credible rather than promotional. The next stage is not to weaken the claim. It is to turn daily use into evidence that tests and strengthens it.

The first concerns closure. The Zeigarnik literature is at its strongest when an unfinished task is paired with a concrete next step — an if-then plan, a specific commitment — that allows the mind to let it go.⁵ Lokus often offers something softer: the reassurance that an activity "can wait," without always turning that into an explicit plan. The open question is whether the End-of-Day ritual supplies enough of the closure on its own, or whether the bare "this can wait" leaves the task still mentally active. The honest position is that the ritual is the best candidate mechanism for doing that work — not that it has been shown to.

The second concerns the journaling. Self-monitoring has strong evidence behind it, but the strongest meta-analytic support comes when monitoring is paired with goal-setting, feedback, or review.⁴⁰ Lokus deliberately removes much of that directive machinery, to keep the experience reflective rather than pressured. MoodPrism offers a plausible precedent — engagement predicted mood improvements, mediated by emotional self-awareness — but it is a single programme, not a meta-analytic finding.⁴¹ The open question is whether non-directive journaling captures enough of the effect without the directive half: it may, but it has not yet been shown to at strength.

The third is the most product-sensitive. A commitment device is, by definition, a constraint chosen in advance by the person who will be bound by it; self-determination theory adds the caution that the same constraint can feel supportive or controlling depending on whether it is

experienced as chosen.⁴⁷ Lokus's end-of-day lockout is therefore neither automatically good nor automatically bad. Its effect depends on whether users experience it as a choice they made for themselves or a restriction imposed on them – and theory explains why that distinction matters, not that the lockout will land the right way.

What turns these from open questions into answered ones is the same thing that makes Lokus unusual: it already runs daily. Every day it captures start- and end-of-day mood, activities set and completed, which sphere carried the day, free-text reflections, and repeat engagement. That is a living evidence base, accruing with use, and it is the right raw material. To turn daily signals into stronger proof – kept light-touch and product-friendly, never clinical assessment – the first cohort should also carry one short validated wellbeing scale (WHO-5, or SWEMWBS for stronger UK credibility), a brief perceived-stress measure (PSS-4), and a simple sleep self-rating.⁵⁹ None of this is research overhead; it is the instrumentation that lets the product speak in the language the evidence already uses.

This is the close. The mechanisms are real and peer-reviewed. The product delivers them with strong fidelity, confirmed against its own screens. The category is right, and the buyer already purchases in it.

What remains is not a vague uncertainty but three narrow, measurable questions – and the first cohort of real users will begin answering them, in daily use, in the place where wellbeing tools either earn their claim or quietly fail it. Lokus is well-grounded precisely because it is not claiming things the evidence does not support. It should keep not claiming them.

Appendix A — Master references

A consolidated reference list with verification status. Citations are numbered in order of first appearance; same-paper duplicates that appeared across different sections have been merged into a single canonical entry. A few sources (notably Macan's process model) remain bibliographically pending final formatting.

1. Zeigarnik effect — unfinished tasks remain active in memory and intrude into unrelated activity. Classic observation originating with Bluma Zeigarnik (1927); framed here as the foundational statement of open-task cognitive burden. (Concept-level; verify original-year citation if cited formally.)
2. Leroy, S. (2009). Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes*, 109(2), 168–181. DOI: 10.1016/j.obhdp.2009.04.002
VERIFIED
3. Cropley, M., & Zijlstra, F. R. H. (2011). Work-related rumination framework — distinguishing affective rumination (emotionally loaded, intrusive, repetitive work-related thought; associated with exhaustion and sleep impairment) from problem-solving pondering (analytical, less emotionally arousing). — Framework verified via subsequent peer-reviewed literature applying the model. Full original citation (journal/volume/pages) to confirm before final
VERIFIED
4. Syrek, C. J., Weigelt, O., Peifer, C., & Antoni, C. H. (2017). Zeigarnik's sleepless nights: How unfinished tasks at the end of the week impair employee sleep on the weekend through rumination. *Journal of Occupational Health Psychology*, 22(2), 225–238. PMID: 27101340 **VERIFIED**
5. Masicampo, E. J., & Baumeister, R. F. (2011). Consider it done! Plan making can eliminate the cognitive effects of unfulfilled goals. *Journal of Personality and Social Psychology*, 101(4), 667–683. DOI: 10.1037/a0024192 **VERIFIED**
6. Psychological detachment — Sonnentag, S., & Fritz, C. (2007). The recovery experience (detachment, relaxation, mastery, control). *Journal of Occupational Health Psychology*, 12(3), 204–221. Detachment = mentally disengaging from work during non-work time. Large meta-analytic evidence links detachment to lower exhaustion and better sleep. — Framework and finding verified via APA PsycNET and PMC; meta-analytic figure (k=316, N=99,329) confirmed but NOT **VERIFIED**

7. Goal disengagement — Wrosch, C., & colleagues. The capacity to disengage from unattainable/unfeasible goals predicts lower psychological distress, fewer intrusive thoughts, and better sleep; effect mediated by reduced rumination/emotional distress. Meta-analytic review: *Nature Human Behaviour* (2025). — Findings verified via PMC (Social and Personality Psychology Compass) and Nature; full citations (Wrosch et al. 2007 exact journal/volume/pages; 2025 review authors) to confirm before final. Finding paraphrased; no statistics quoted **VERIFIED**
8. Planning fallacy — concept: Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. — foundational characterization of systematic underestimation robust to feedback/experience/warnings. Classic demonstration: Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67(3), 366–381 **VERIFIED**
9. Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*. — Habit formation median ~66 days (range 18–254); a single missed day does not significantly derail the automaticity curve. Finding paraphrased; the "single missed day" point is corroborated by the original study and subsequent habit literature. Full citation pages to confirm before final
10. Tangney, J. P., et al. — shame vs guilt distinction. Guilt = adaptive (orienting toward repair); shame = maladaptive (orienting toward withdrawal/avoidance) **VERIFIED**
11. Sirois, F. M. (2014). Procrastination and stress: Exploring the role of self-compassion. — Self-compassion shows a moderately strong negative association with procrastination; low self-compassion explains much of the associated stress **VERIFIED**
12. Aeon, B., Faber, A., & Panaccio, A. (2021). Does time management work? A meta-analysis. *PLOS ONE* **VERIFIED**
13. Perceived control of time — Macan, T. H. (process model: perceived control of time mediates the link between time-management behaviour and wellbeing/stress/satisfaction); Thompson, S. C., & Spacapan, S. (1991) — perceived control as a buffer against stress, associated with emotional wellbeing. The Aeon et al. (2021) meta-analysis itself endorses perceived control of time as the plausible mechanism ("time management training makes people feel more in control of their time") **VERIFIED**
14. Locked thesis anchor — the paper's central claim, framed as design intent (Appendix F v3, claim posture locked). Verbatim locked thesis
15. Claim-tier system — T (theory-supported) / F (fidelity-supported) / O (outcome-tested), with the anti-overclaim guardrail that mechanism logic ≠ proven product. Defined in the outline and Appendix F. Note: the F tier is treated here as a standard the paper applies feature-by-feature (§§4–10), not as a blanket assertion that every mechanism is delivered
16. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38, 69–119. DOI: 10.1016/S0065-2601(06)38002-1 **VERIFIED**

17. Deci, E. L., & Ryan, R. M. — Self-Determination Theory; autonomous motivation (personal volition, value alignment) is a more durable driver of sustained change than controlled motivation **VERIFIED**
18. Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science*, 6, 42 **VERIFIED**
19. "Calm productivity" — headline brand phrase, defined once here per Appendix F: reducing the cognitive pressure, guilt, and open-loop burden around unfinished activity; explicitly NOT "producing more." Use of this phrase elsewhere in the paper draws on this definition
20. Claim boundaries (Appendix F v3): "designed to improve wellbeing" permitted only in design-intent frames, never bare; real-world proof stated in the present tense (already gathering mood, activity, reflection, return use); social proof built through user stories, retention, and feedback; non-clinical boundary (no diagnostic, treatment, or therapeutic-efficacy claims). The detailed "what we should gather next" roadmap (validated wellbeing instrument such as WHO-5/SWEMWBS, perceived-stress and sleep measures, adherence/retention, lockout experience) is deferred to §11 (limitations and future evaluation)
21. Nine-activity cap fidelity — Item 1 product-fidelity log (closed). Onboarding copy: "Each sphere holds three activities to keep your focus clear." 3 spheres × 3 = 9. Confirmed as a stated, user-facing design rule; limit-enforcement behaviour (does the app hard-block a fourth activity?) not separately screen-captured
22. Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114 **VERIFIED**
23. Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2), 333–358 **VERIFIED**
24. Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, 37, 409–425 **VERIFIED**
25. Fidelity verdict — Item 1 log: claim tier T → F (rule confirmed); hard-enforcement (fourth-activity block) pending. Logged non-blocking upgrade: an at-limit / fourth-activity-attempt screen would be needed before claiming a hard-enforced cap rather than the confirmed 3-per-sphere / 9-total design rule
26. Caveat resolution — per §1 and Evidence Review Mechanism 1 open question. M7's end-of-day gate is assessed in §9 as the candidate mechanism that may supply the "everything undone can genuinely wait" permission — which functions as the specific plan that Baumeister and Masicampo's research says is required to close the Zeigarnik loop — achieved through permission + psychological detachment / goal disengagement (Sonnentag & Fritz; Wrosch et al.), NOT next-step planning. Cross-ref §9. (Masicampo & Baumeister 2011, *JPSP* 101(4), 667–683 — per §1; cited to establish that release ≠ finishing, not as a product prescription.)

27. Three spheres fidelity — Item 1 product-fidelity log (M2, closed): Home / Personal / Business, segregated; activities viewed one sphere at a time, not in a blended master list. Verdict: Confirmed — strong. T → F
28. GPS anchoring — Item 2 onboarding log: GPS optional and configurable; set via profile, toggleable in Settings ("You Can Turn GPS off in Settings"); framed as enhancing "Lokus AI context," not a first-run requirement; skippable. Privacy note: location data is user-controlled and optional
29. Ashforth, B. E., Kreiner, G. E., & Fugate, M. (2000). All in a day's work: Boundaries and micro role transitions. *Academy of Management Review*, 25(3), 472–491. DOI: 10.5465/amr.2000.3363315 **VERIFIED**
30. Integration → work-life conflict — Olson-Buchanan et al. (2006), Blurring boundaries: Correlates of integration and segmentation, *Journal of Vocational Behavior* (sciencedirect S0001879105001296): employees who blend work and personal life establish fewer barriers and report increased work-life conflict; *Frontiers in Psychology* (2021), Predicting Work-Life Conflict: "integration congruence" related to higher work-life conflict than "segmentation congruence." **VERIFIED**
31. Fidelity verdict (spheres) — Item 1: claim tier T → F (strong). Spheres confirmed segregated (Item 1, strong); GPS optional and user-controlled (Item 2) — fits the perceived-control throughline and is claim-safe because optionality, not data collection, is the load-bearing point
32. Caveat — boundary theory's own "rites of passage." Ashforth et al. (2000) state that segmentation makes boundary crossing more difficult and that crossing "often is facilitated by rites of passage." Segmentation alone is therefore not sufficient. M7's start- and end-of-day gates are assessed in §9 as candidate crossing rituals (chosen, not imposed — per Item 2 resolution). Distinct from §4's caveat, which concerned parking and cognitive closure; §5's concerns the marked transition between roles
33. Foundational Framework — Evidence Review, opening section "Rigidity of Structure, Fluidity of Content": the structural constraint (fixed number, stable cue, rigid binding) does the behaviour-change work and should not be softened; the content inside that structure (which behaviour, which label, which specific plan) is where personalisation belongs, and loosening it "costs nothing structurally while adding the autonomy effect." Synthesised from the three literatures cited at [^10]–[^11]. This is the document's stated operating principle, applied throughout every mechanism
34. Fidelity (naming-and-lock) — This is a PLANNED feature, not yet shipped ("we are thinking of allowing users to name the 3 life spheres — only change them once and then they can't for a period of time"). Marked design intent (T-only) pending fidelity confirmation once built. The three-sphere separation itself is confirmed (T → F strong, Item 1); the naming-and-lock extends the rigidity/fluidity principle to the sphere labels
35. No-time-blocking fidelity — Item 1 product-fidelity log (M3, closed): activity-entry shows Name / Description / Priority only — no time/date/duration field. Verdict: T → F (provisional) — visible flows support no-time-blocking, but the evidence sits on placeholder/possibly-old activity-entry screens. A clean, populated current-production activity-entry capture is owed before final figure use (logged upgrade #1 in Item 1). Provisional, not strong — contrast M1 (cap, confirmed as stated rule) and M2 (spheres, confirmed strong)

36. Absence-as-mechanism / perceived-control framing — drawn from the Evidence Review's Mechanism 3 deep-dive: a time-blocked schedule missed daily is "a machine for manufacturing perceived loss of control... Removing time architecture removes the most reliable daily generator of the exact failure signal the pain state moralises." Perceived control was established as the throughline in §2 (Macan process model; Thompson & Spacapan 1991, perceived control as a buffer against stress); §6 leans on that thread rather than introducing a new source. Stated as design intent ("designed to avoid manufacturing one predictable source"), NOT as a proven outcome
37. Structure relocated — architectural cross-reference, not a separate citation: the cap (§4), the spheres (§5), and the daily gate (§9). The claim is about where Lokus places its scaffolding instead of the clock — bounded attention, role separation, daily transition — not a new evidentiary assertion
38. Non-punitive fidelity — Item 1 (M4): no streaks, overdue markers, or shame language observed in reviewed flows (initially T → F provisional, pending Item 4). Item 4 (M4): AI tone confirmed non-judgmental, non-prescriptive, no clinical claims, no shaming, no coaching, no "you should." → Fully confirmed (upgraded from provisional). Verdict: T → F (confirmed) — strong, across both the interface mechanics and the AI's voice. Stronger than §6's no-time-blocking (provisional, pending a clean populated activity-entry screen)
39. "Absence as mechanism" framing — drawn from the Evidence Review M4 deep-dive: rather than softening punishment (forgiving streaks, pause buttons, grace days — which the wider productivity category is converging on), Lokus declines to build the punishment machinery at all. A "forgiven" streak still frames the missed day as a transgression requiring forgiveness; the absence of streaks frames it as nothing. Presented as design logic built from the shame/guilt ([^2]) and self-compassion ([^3]) evidence — NOT cited in the body to the weak blog sources (thebrink.me, gamificationhub.org, yukaichou.com), which appear here only as non-load-bearing context. The Habitica "0.90% of users who lose a 2–3-day streak return" figure cited in the Evidence Review is deliberately OMITTED — it could not be verified from a credible primary source (industry blogs: trophy.so, nerdsip.com only) and was already dropped in §2 for the same reason. The "punitive gamification has real failure data" point is made via design logic, not this unverifiable number
VERIFIED
40. Self-monitoring meta-regression — Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology*, 28(6), 690–701 **VERIFIED**
41. MoodPrism — Bakker, D., & Rickard, N. (2018). Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. *Journal of Affective Disorders*, 227, 282–286 **VERIFIED**

42. Three-app RCT caveat — Bakker et al. randomised controlled trial of three smartphone apps (MoodKit, MoodPrism, MoodMission): wellbeing increases against control over 30 days. Deliberately NOT used as main support in §8 because the compared apps include directive/therapeutic content: MoodMission offers tailored suggestions, "missions," and badges (Monash Lens); MoodKit is CBT-based. These are not clean support for Lokus's non-directive design. MoodPrism (the 2018 study, [^2]) is the cleaner match. Flagged here for the final citation master list rather than leaned on in the body
43. Hypervigilance caveat — the clinical concern that constant mood monitoring can amplify anxious hypervigilance to internal states in some users (Evidence Review cites an industry synthesis, mindfulsuite.com — weak source; the underlying concern is clinically recognised). Lokus's design answers it: two check-ins anchored to the start and end of the day (not continuous or prompted throughout) = lower-intensity dosing; absence of coaching means a low mood entry is never converted into an implied to-do. Stated as a lower-intensity design choice, not a risk eliminated
44. Fidelity (Items 1 + 4, both closed) — M5: T → F (confirmed), strong. Input confirmed (Item 1): "Your Mood?" + selection + "a brief thought on your day" + "Over time, Lokus recognizes patterns" — affect labeling on BOTH Start- and End-of-Day. Reflective feedback confirmed (Item 4): the single end-of-day reflection mirrors structured data (moods + activities + sphere) + the user's free-text thought, references cross-day patterns (e.g. sphere-balance, mood trends); non-goal-directive (no goal-setting, no coaching). Regulatory boundary clean: non-diagnostic, non-prescriptive, no clinical/therapeutic claims — supports non-clinical positioning
45. Pennebaker guardrail — Pennebaker's expressive-writing paradigm is a structured, deep emotional-disclosure protocol for processing distressing events, and Pennebaker himself cautions that writing can tip into rumination rather than reflection. Daily light-touch pattern surfacing is a different exercise and is NOT supported by this citation; deliberately not leaned on in §8 (fuller caution kept here in source notes). (va.gov Whole Health Library therapeutic-journaling doc; APA Speaking of Psychology — expressive writing.)
46. Affect labeling — Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: Affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science*, 18(5), 421–428 VERIFIED
47. Commitment device — a constraint chosen in advance by the person who will be bound by it, to overcome time-inconsistent preferences / weakness of will. Primary sources: Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press; Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies*, 23(3), 165-180 (DOI: 10.2307/2295722); Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press VERIFIED
48. Bono et al. 2013 VERIFIED
49. Clauss et al. (2018) VERIFIED
50. Nurses RCT pilot VERIFIED

51. Sonnentag et al. 2021 **VERIFIED**
52. Fidelity (Items 1 + 2, both closed) — M7: T → F (confirmed), strongest of any mechanism. Mechanics confirmed (Item 1): SoD (mood + weather) opens spheres; EoD (mood + optional thought); irreversible lockout ("the app closes for the rest of the day" / "your spheres will close. Note, you can not re-open these until tomorrow!"); spheres vanish after EoD; app inert until next SoD. User-initiated (user taps EoD sphere); forewarned at first run (forced tutorial, replayable from Settings) (Item 2). "Chosen, not imposed" defensible. Strong fit with detachment + commitment-device evidence. Design note (not a fidelity failure): no just-in-time confirmation prompt at EoD save — see caveat / product recommendation
53. Caveats — (a) SDT controlling-read: a hard lockout can read as controlling, undermining autonomous motivation; mitigating argument (voluntary commitment > external rule) holds only if users experience it that way, which requires real usage data — this is one of the three "genuinely untested" load-bearing assumptions flagged in the Evidence Review's interim conclusion. (b) Commitment-device transfer: evidence strongest for financial behaviours; sustained effects on habit-type behaviours more mixed → transfers more by logic than guaranteed magnitude. (c) Positive-prompt vs "how do you feel" mismatch: the field studies (Bono, Clauss, nurses, Sonnentag 2021) tested a positive reflection prompt; Lokus asks how the user feels (may surface negative content). Affect-labeling evidence supports naming negative feelings, but the two literatures have not been tested in combination in this exact form. Distinct from §4 (closure), §5 (ritual), §6 (perceived-value), §7 (copy), §8 (half-combination) caveats
54. Perceived control of time as the mediating variable — Macan's process model / review of time management (time-management behaviours → perceived control of time → job satisfaction, stress reduction, health); + significant association between perceived control over time and better mental-health status (PMC8397069, Journal of Clinical Medicine, 2021) **VERIFIED**
55. BCT Taxonomy v1 (93 hierarchically clustered techniques); self-monitoring of behaviour = BCT 2.3 — Michie, S., et al. (2013). The behaviour change technique taxonomy (v1) of 93 hierarchically clustered techniques. Annals of Behavioral Medicine / NIHR Health Technology Assessment monograph (HTA 19:99) **VERIFIED**
56. Linardon et al. (2024) **VERIFIED**
57. Workplace digital-MH umbrella review **VERIFIED**
58. Deloitte ROI **VERIFIED**
59. WHO-5 Well-Being Index **VERIFIED**

Appendix B — Mechanism → product fidelity map

Each mechanism mapped to the product feature that delivers it, the fidelity verdict from the closed product-fidelity review (Items 1–5), and the open caveat. Claim tiers: T = theory-supported; F = fidelity-supported; O = outcome-tested.

Mech.	Feature	Product behaviour	Fidelity verdict	Sec.
M1	Nine-activity cap	3 activities per sphere = 9; onboarding states the rule	T → F (rule confirmed; hard-enforcement behaviour not yet captured)	§4
M2	Three spheres	Home / Personal / Business, segregated, GPS-anchored	T → F (strong)	§5
M3	No time-blocking	Activity entry shows no time slots	T → F (provisional — wants one populated current screen)	§6
M4	Non-punitive design	No streaks, overdue markers, or shame language; AI tone non-judgmental	T → F (fully confirmed; product + AI tone)	§7
M5	Mood check-in / self-monitoring	Low-friction mood tagging + reflective over-time AI feedback; goal-directive feedback absent by design	T → F (confirmed, strong)	§8
M7	Start / End-of-Day gate	User-initiated, forewarned, non-overrideable commitment device; detachment confirmed; lockout experienced as chosen	T → F (confirmed — strongest of any mechanism)	§9
M6	Perceived control	Throughline across §3, §10, §11 — not a standalone feature	(throughline, not a feature)	§3 / §10 / §11

M6 (perceived control) is a throughline across §§3, 10, and 11 rather than a standalone feature, and so has no single fidelity verdict.

Appendix C — Claim boundaries

The following claim-boundaries standard is carried verbatim from the Lokus Behaviour Change Evidence Review. It governs every statement in this paper.

The preceding sections establish what the evidence supports and where it remains open. This section translates that into the claims the white paper may responsibly make — and the boundaries it must hold. It is written for a journalistic, thought-leadership document, not a clinical evidence submission; a stricter clinical-governance version can be derived separately if needed.

What we know: The academic mechanism behind each feature is real and peer-reviewed. The product, as built, delivers those mechanisms — confirmed against the app’s screens and behaviour in the mechanism-by-mechanism review above. The design is non-punitive, user-initiated, reflective, and oriented to wellbeing rather than to streaks, nagging, or pressure.

What we can reasonably infer: Because the mechanisms are evidence-backed and the product faithfully delivers them, there is a credible, evidence-grounded basis to expect wellbeing benefit. This is stated as expectation, not as proven fact. The review’s own caveat — that grounding in mechanism is not the same as proven real-world outcomes — is kept, but lightly: the inference is reasonable, and the product is actively collecting the daily signals that will strengthen it over time.

What Lokus is already gathering: Every day the product captures real-world signals — start- and end-of-day mood, activities set and completed, which life sphere carried the day, free-text reflections, AI-generated pattern reflections, and repeat engagement. These are early real-world evidence: a living evidence base that accrues with use. User stories, retention, and feedback build the social proof around those signals.

What we should gather next: To turn daily signals into stronger proof, the product should gather — kept light-touch and product-friendly, never clinical assessment:

- A short validated wellbeing instrument — WHO-5 (product-friendly, five items), or SWEMWBS for stronger UK wellbeing credibility.
- A brief perceived-stress measure (PSS-4) if a stress signal is wanted.
- A simple sleep self-rating, or wearable-derived sleep duration and quality if available. Full PSQI is likely too heavy for routine app use.

- Retention and adherence — day-1, day-7, day-30 retention; Start- and End-of-Day check-in completion rates; streak-free repeat use; adherence in a no-push model.
- Mechanism usage — activities set versus completed, sphere balance, use of the end-of-day reflection, lockout completion.
- Lockout experience — whether users experience the non-overridable end-of-day lock as helpful, frustrating, or autonomy-supportive. This directly tests the chosen-versus-imposed question flagged under Mechanism 7.
- Qualitative social proof — testimonials, app reviews, user quotes, and “what changed?” prompts.
- Partner and employer feedback, if a B2B2C route develops — anonymised aggregate usage, activation and renewal rates.

What we still do not claim: No quantified outcomes (“reduces stress by 30%”). No “proven to.” No diagnosis, treatment, therapeutic efficacy, or mental-health-treatment claims. Lokus is a wellbeing tool, not a medical or therapeutic intervention.

Language rules for the white paper

Wellbeing claims: “Improves wellbeing” is used only inside a design-intent frame — for example, “Lokus is designed to improve wellbeing by reducing the cognitive load around unfinished activity” — never as the bare sentence “Lokus improves wellbeing,” which reads as a proven result not yet in hand. Preferred verbs: designed to, built to, can support, offers a credible path to, is expected to support.

“Calm productivity”: Kept as a headline brand phrase, defined once on first use: it means reducing the cognitive pressure, guilt, and open-loop burden around unfinished activity — not producing more. “Increases productivity” and “improves output” are avoided unless future data supports them.

Real-world data language: Daily mood, activity, and reflection data are described as “real-world signals” or “early real-world evidence” — a living evidence base — not as validated wellbeing outcomes. “Social proof” is reserved for user validation: testimonials, retention, stories, and ratings.

Regulatory boundary: Explicit throughout — no diagnosis, treatment, clinical efficacy, or mental-health outcome claims. This is consistent with the product’s own AI behaviour, which is non-judgmental, non-prescriptive, and never diagnoses.

Thesis anchor for the white paper: Lokus is a theory-informed, evidence-grounded wellbeing tool — designed to improve how people carry, close, and

recover from unfinished work, while its real-world proof builds through daily mood, activity, reflection, retention, and user feedback.

Appendix D — Limitations & measurement plan

The three genuinely untested assumptions

These are not gaps in the argument but specific, falsifiable product questions. Each is narrower than "does Lokus work" — and each is answerable by the first cohort of real users.

- **1. Closure without a concrete plan.** The Zeigarnik literature is strongest when an unfinished task is paired with a specific if-then plan. Lokus often offers a softer close — "this can wait." Whether the End-of-Day ritual supplies enough closure on its own remains open; the ritual is the best candidate mechanism, not a proven one.
- **2. Non-directive journaling.** Self-monitoring's strongest meta-analytic support comes when paired with goal-setting, feedback, or review — the directive half Lokus strips out. MoodPrism is a plausible precedent but a single programme, not a meta-analytic finding.
- **3. Chosen lockout or imposed restriction.** A commitment device is, by definition, a constraint chosen in advance; self-determination theory warns the same constraint can feel supportive or controlling. Whether users experience the lockout as chosen or imposed is decisive — and only real usage data resolves it.

What Lokus already gathers daily

Start- and end-of-day mood; activities set and completed; which sphere carried the day; free-text reflections; AI-generated pattern reflections; repeat engagement. This is a living evidence base, accruing with use.

What the first cohort should also carry

Kept light-touch and product-friendly — never clinical assessment:

- One short validated wellbeing scale — WHO-5, or SWEMWBS for stronger UK credibility.
- A brief perceived-stress measure (PSS-4); a slightly longer instrument may give a more stable signal.
- A simple sleep self-rating (full PSQI is likely too heavy for routine app use).
- Retention and adherence — day-1, day-7, day-30 retention; check-in completion rates; streak-free repeat use.

- A lockout-experience measure — does the non-overridable end-of-day lock feel helpful, frustrating, or autonomy-supportive? This directly tests the chosen-versus-imposed question.
- Qualitative social proof — testimonials, reviews, user quotes, "what changed?" prompts.

Appendix E — Glossary

Zeigarnik effect

The tendency for unfinished tasks to persist in memory and intrude into unrelated activity and rest.

Attention residue

Cognitive persistence of a previous task context that impairs performance on the next task (Leroy, 2009).

Boundary theory

How people create, maintain, and transition between role domains such as work and home (Ashforth et al., 2000).

COM-B

A behaviour-change model holding that behaviour requires Capability, Opportunity, and Motivation together (Michie et al., 2011).

Behaviour Change Wheel

The intervention framework built on COM-B, classifying interventions by which components they target.

Self-determination theory (SDT)

A motivation theory distinguishing autonomous (chosen, owned) from controlled (pressured) motivation (Deci & Ryan).

Perceived control of time

The felt sense of being in charge of one's own day; a proposed mediator between time management and wellbeing (Macan).

Commitment device

A constraint chosen in advance by the person who will be bound by it, to overcome future weakness of will.

Recovery experiences

Off-job activities and states (detachment, relaxation, mastery, control) that restore resources (Sonnentag & Fritz, 2007).

Psychological detachment

Mentally disconnecting from work during off-job time; the recovery experience most reliably tied to lower exhaustion.

Self-monitoring

A behaviour-change technique (BCT 2.3) of observing and recording one's own behaviour or state.

Affect labeling

Putting feelings into words, which dampens amygdala activity (Lieberman et al., 2007).