Ep #106: Breaking the Cycle of Urgency in Law Firms [Agile Lawyering Part 6]

Full Episode Transcript

With Your Host

John E. Grant

There are a lot of law practices out there that think of themselves as running on priorities. But in reality, they run on urgency, and that is not a good thing. Of course, some things in practice will inevitably become urgent. And when they do, it can be all hands on deck to solve that particular problem. And of course, we naturally want to reward the effort when the problem gets solved. But if you aren't careful, then that kind of system quietly trains people to manufacture urgency just to get their work done.

Fires will multiply, both real and imagined. Stress increases, and the truly important work only winds up getting attention when it's already close to late. Today's episode is about breaking that cycle, not with willpower or more rules, but by establishing a simple, predictable operational rhythm that changes how work gets prioritized and what behaviors get rewarded.

You're listening to *The Agile Attorney Podcast*, powered by GreenLine. I'm John Grant, and it is my mission to help legal professionals of all kinds build practices that are profitable, sustainable, and scalable for themselves and the communities they serve. Ready to become a more Agile Attorney? Let's go.

A quick note. The concepts I'm discussing to talk about today should be useful to you no matter what kind of law practice you're part of or what tools you use. If you'd like, stay tuned at the very end where I will briefly discuss how my software tool GreenLine helps reinforce the practices from today's episode.

Welcome back, everyone. We're past the halfway point in this Agile Attorney 101 series. And if you've gotten this far, then hopefully you're starting to feel the value of these foundational concepts. We've covered a lot of ground, focusing on finding smooth and predictable flow in your law practice, making work visible, understanding bottlenecks, setting quality standards, and managing capacity. And I hope you're starting to see how these concepts not only make the workflow better, they're also will make everything easier on you and your team members too.

Of course, these are all key concepts from Capital A Agile along with Lean and other systems thinking approaches to work. But as we cross this midpoint, I want to briefly remind you about my emphasis in episode 101 about lowercase a agility: the core strength, the fluidity, focus, and adaptability of the regular things we see in the world as agile.

Now, last week in episode 105, my focus was on capacity management. And I called for you to look at the open work inside your law practice and close the closable. And I tied that back to the Marie Kondo approach to tidying up your physical space. We also covered WIP limits as a quality policy that protects your finite capacity and prevents you from getting overloaded again in the future.

Then toward the end, I introduced the drum buffer rope framework. And I explained the rope part, which is tying your entrance door to your exit door, so new work only comes in when old work goes out. That's kind of like that nightclub capacity policy. And I talked about buffers, the queue management piece, like the various approaches restaurants might use when they're at capacity and you've got to wait for a table.

But I saved the drum for today because it really deserves its own focus. And I ended by asking you to imagine that you're running or working in a law practice where things very rarely rise to the level of being truly urgent, where your work life is just pretty normal, but you're still doing the work and earning the money you need to support all the important things in your life. I want you to conjure up that image again. How chill must that sound?

But I bet you're a little skeptical that could actually happen, maybe even a little worried about what it might look like in practice. There might even be some part of you thinking that a practice like that sounds a little boring. And

I get the concern, but I promise you, if you get this transition right, you will not be bored.

One of the crazy things about this profession is how dependent many people become on this cycle of fear and reward, carrots and sticks. And one of the reasons I hear so often from people that I always feel like I'm fighting fires or I feel like I'm on the hamster wheel is that we are. We're indoctrinated to it. The whole profession seems to have these reward systems in place that are tied to our lizard brains, and that constantly keeps us on edge.

So there are a lot of things that are important about establishing the drum or setting the cadence of how work gets addressed in your law practice. But I think the most important one is that we can use the cadences I'm about to teach you to slowly but surely change the feedback loops and ultimately change what behaviors get rewarded for everyone involved in delivering the legal work in your practice.

Now, before I get to the cadence, I want to go back for a minute and talk about what happens when teams aren't intentional about their capacity and how that influences the human level reward systems in your practice. When a law practice becomes urgency coded, when diving catches and individual heroics get celebrated, or when all hands on deck surges combine to deliver those very real results, then some natural human behaviors start to emerge that are understandable, but are ultimately destructive to team dynamics and to system wide flow.

And consciously or not, people start to think, "Well, that was cool, but I guess if I want my thing to get done by the team, then I'd better make it seem urgent too." I talked about this a little last week in my discussion of urgency camouflage. And sometimes that's when people yell and scream and pound the desk to give the false sense of urgency. And sometimes that's when people smuggle their work to the front of the queue using back channels or personal favors.

But if those techniques are allowed to work, then it becomes this self-perpetuating, vicious cycle where people feel like they have to yell a little louder or play increasingly absurd political games just to get otherwise routine work to the front of the line. And because of all that queue jumping, the everyday important things don't get worked on until they also become urgent. And so again, everyone feels like they're fighting fires all the time because they are.

And WIP limits create the conditions to break that cycle, but we also have to be very intentional about creating those different types of reward systems. And that's a different pattern of feedback loop that incentivize getting things done proactively before they get urgent instead of saving the day on something that never should have become urgent to begin with.

Now, one of the things about WIP limits is that they inevitably create queues. Once we have capacity for new work underneath our WIP limit then, the question becomes, how do we decide what to work on next? How do we prioritize our work?

And this is one of the rare places in the series where I'm just going to come out and tell you exactly what I think you should do. Adopt a first in, first out default prioritization policy. It's the same one they're using at that nightclub, the restaurant, the grocery store, the bank. And they do it because it's simple, it's fair, and it's effective. Everybody understands how they work.

Now, once you have that default policy, of course you're going to have exceptions. Sometimes things catch fire, maybe a client gets in an accident or gets arrested, property gets stolen. Sometimes work shows up with a short deadline. But the key thing is that they are exceptions. They're handled in the context of that baseline expectation that work enters your finite capacity on a first in, first out basis.

And I want to give you a sort of a new framework for how to manage and think about those expectations. And this is based on the notion of cost of delay. It's an interesting mental trick because instead of asking why should this work come next, which can create a lot of that politicking and sort of advocating for my particular pet project, instead we ask, what would be the consequences if we don't do this work right now? That's nice because it forces a more objective, dispassionate analysis that is less likely to get gamed and politicked.

Now, within that cost of delay framing, there are four main archetypes of work. And in Agile, we call them classes of service. The first is known as standard work, and that should make up the vast majority of the work in your practice. If we chart cost of delay on a graph where the x-axis is time and the y-axis is severity, then standard work shows a steady line moving up and to the right.

The cost of delaying a thing to tomorrow isn't that much higher than doing it today. But if you put off work for too long, that slow and steady growth becomes a problem. And first in, first out prevents that, especially when combined with WIP limits, so you finish old work before you're committing to something new.

The second class of service is what we call expedite, and that covers the emergency work. If you think of that cost of delay graph, then the expedite line has a way steeper slope than the standard work one does. I think of it as where something has caught fire and that fire is spreading. The severity on the y-axis is accelerating really quickly over a short span of time. And obviously, these jump to the front of the line. In fact, they might even displace existing work that's already in progress if the emergency is severe enough.

The third is what we call fixed date work. And that's when a deadline is approaching. And the graph for this one is interesting because the cost of delay before the deadline is actually relatively flat. But once that deadline passes, the consequences are immediate and severe. The line spikes vertically all at once. And then oddly enough, it tends to flatten out again because once the damage is done, it's done, and all you can do is try to recover.

What's interesting though is in a healthy system working within its capacity, the fixed date work item types don't necessarily need to jump the queue. You can usually address them with the same FIFO policy that you use for standard work. You only need to move it to the front of the line when the deadline is near and the work isn't on track to finish before that deadline arrives.

Now, this leads me to an important aside about terminology around dates. And I want to strongly urge you and your team to draw a clear distinction between a deadline, which is that immediate and severe consequence thing, the thing that leads to that fixed date work item type, and other types of due dates, like target dates. The target dates are useful in terms of milestoning and making sure work is on track for completion, but missing a target date or a due date just doesn't have that same consequence as blowing a deadline.

And this is going to take a little practice because colloquially, we use these terms interchangeably. But I'm urging you, think about being more specific and urge your team to be more specific. Here's how I like to do it. For internal milestones or progress states, I like to use target date, partly because it's less likely to get confused with the word deadline.

If you are going to use due date, which is that more confusing term, I would save it for communications with clients. Giving them due dates is a really effective strategy. And save the word deadline, specifically preserve it for the types of work that has that severe and immediate consequence if we don't hit that date.

And I'll give you a quick example about how those terms might work in combination. Let's say you're working on an M&A deal. The date that the deal is supposed to go through, that's probably a deadline. If you need some information from a client so that you can actually do some due diligence work, give them a due date. Now, if you want your associate to review that due diligence material and send you a memo by next Thursday, that's a target date.

Okay, back to classes of service because there's one more, and that fourth class of service is called intangible work. And this is work that rarely becomes urgent right up until it does. And here's the example I like to use. If you're a homeowner, you probably know this, right? If you've got a 12-year roof on your house, you don't necessarily need to replace it right at year 12. You can probably stretch it, 13, 15, maybe longer.

But if a big storm blows in and your roof is beyond its useful life and it springs a leak, then you're going to have a lot more work to do than just repairing the roof. You've basically created an emergency situation because we put off something longer than we should have.

In a law practice, this comes up with things like policy updates or template updates, taking CLE so you stay on top of the law. And I think the best way to deal with them is the same way that you would do it with oil changes in your car, and that's just to put them on a schedule and stick to that schedule. If you don't, then these low risk items have a way of converting themselves into real emergencies.

So we've got the WIP limits in place. We've got these definitions of ready and done that are improving our consistency, preventing rework. And then we've got now our default first in, first out prioritization approach with the logical exceptions as you need to make them. The next question is how will we keep our promise-keeping machine running at a reasonable speed and on a predictable timeline? And that's finally where the drum comes in, right? And the drum is about establishing a regular, predictable cadence of team interactions that keep the system healthy and keep work moving.

Now, when done well, there's really only three of them: the weekly planning meeting, the daily stand-up, and periodic retrospectives. And it's important to note that these all fall under the umbrella of a very specific meeting type called cadence meetings. And again, their purpose is around alignment and brief information sharing. It's around coordinating interdependencies, and ultimately making sure that your matters are moving with some momentum.

And the key to successful cadence meetings is that they should be short, they should be frequent, they should be scheduled, and they should be mandatory. More importantly, they're feedback loops. They're where the system checks itself, and we start recognizing those different behaviors that we want to reward, like progress over effort, transparency over opacity, and reliability over heroics.

Now, there are two important things that people should not be doing in cadence meetings. One of them is reporting on their own activity as opposed to reporting on the deliverables progress. When I'm running a cadence meeting, I try to banish the words I worked on from the conversation. I don't so much care about effort, I care about progress. And it often takes a while for teams to get the hang of this because they're so used to being measured by effort-based metrics. But to badly paraphrase Yoda, I'd say the work is either done or not done. There is no worked on.

The other thing that cadence meetings should not be allowed to do is devolve into a working meeting. And these meetings will inevitably daylight things that need to happen in order to move the work closer to done. That's their whole point.

But the actual work needs to happen elsewhere, maybe a co-working session with a couple of team members after the stand-up or the other cadence meeting, or more than likely someone just going back to their desk and doing their job. But as soon as you allow actual work to start happening in the status meetings, I guarantee you will lose control of the whole thing and nobody's going to come out happy.

With that said, let me talk about the first of the cadences, the weekly planning meeting. And as the name implies, this is a weekly check-in that is going to help us set the tone for the week ahead. Although we actually start it not by looking forward, we start by looking backward. And I like to open a weekly planning meeting with celebration and reflection. What did we actually deliver in the last week? Acknowledging those wins is really the first step in rewiring the reward pathways away from the heroic saves and towards the competent accomplishment we're trying to achieve.

Then of course, when you do that, you want to also look at things that we thought we'd finish but we didn't. And this isn't to assign blame, but to close a feedback loop. Did we bite off more than we could chew? Did we underestimate how long something would take? Was there hidden work that popped up or did something catch fire that took our attention?

And these are the things that come up in a regular basis, but they rarely get explicitly acknowledged. And we need to acknowledge them. It isn't so much a problem that they happen, but it becomes a problem if they keep happening.

And this is the thing where we can recalibrate that optimism bias I've talked about before against our actual delivery. And what's great is that if you've committed to closing five matters in a week but you've only finished three, that's useful information. It tells you something that will help everyone make more credible commitments the next time around. That reflection will generally be pretty short. It usually takes about five minutes, maybe as long as 15.

Then you start to look forward, and you want to do it by going to the far right side of your Kanban board. And that's where you'll find the work that is closest to done. Remember from my close the closable talk last week? That's probably also the oldest work in your system under your first in, first out policy. And so there's lots of good reasons that you want to start and read your Kanban board from right to left.

And once you see that nearly done work, I want you to start making some commitments. What matters will we as a team fully and finally close this week? Then, what in progress work can we confidently move downstream to its next phase? We'll also daylight anything with one of those true deadlines that's approaching that seems at risk.

And that's important, but remember, in a healthy system, deadline driven work shouldn't be at risk because it entered with enough lead time. If you're constantly seeing deadlines at risk, that's actually a signal that you're still over your comfortable carrying capacity.

Then finally, as you're planning your week, you look at new work, but remember, you only get to commit to new work if you've created capacity underneath your WIP limit. As you'll recall from the rope discussion last week, the entrance door only opens when something else has exited. Now, one important thing about the weekly planning meeting, everyone involved in delivery should be there. This isn't a manager's meeting, this is a team meeting. So not just lawyers, but paralegals, legal assistants, anyone who participates in the actual movement of the legal work.

You'll also want to make sure that you've made a record of what commitments you've made, partly because you want to celebrate them when next week comes around, and partly because you want to be able to close that feedback loop if for some reason you don't get to all of the commitments that you've made this week. If you're using Kanban board software, hopefully there's a way that you can do this inside of the tool itself.

So the second cadence meeting is the stand-up, although some teams call it a huddle or maybe a pulse check. This is basically the same structure as the weekly planning meeting, it's just zoomed in to a shorter time scale. Now, I like to run them daily, but I've seen teams that do them three times a week and that can work too.

Here's the thing with the stand-up. In my book, a well-run stand-up meeting should take no longer than two minutes times the number of people attending the meeting. And that means everyone needs to show up prepared to report on their progress since the last stand-up and also already have a sense of what they plan to make progress on before the next one.

And the way that this works is we just go around the room and each person will take their turn to state three specific things. Number one, since our last stand-up, here's the progress I actually made on the matters I committed to last time. Number two, between now and our next stand-up, these are the matters I plan to make progress on and the specific progress I'm going to help them make.

And then number three, if they have worries, then they can say, "I'm concerned about these roadblocks or these dependencies that might prevent me from making the progress that I need to make." And I'll warn you again, this is the meeting where people are really going to want to use that pesky term "I worked on" instead of reporting actual progress.

But what I want to hear is specific things, like, "Yesterday I got the Johnson complaint package drafted and ready for review, and it's cued up for Sarah to check it over," or "Yesterday I drafted the pleadings for the Johnson complaint, and today I'll finish the ancillary documents and get the whole package over to Sarah for review." Both of those are fine because they talk about specific advances, not just about the effort that people made.

And the reason why I'm so insistent on this is that phrases like "I'm working on X" or "I'm busy with Y" have this tendency to become productivity theater. And I actually think of them as warning signs that people aren't actually holding to their priorities or aren't willing to hold themselves accountable or be held accountable for actual progress.

Now, one other thing to look out for, if people keep hitting unexpected work, right? They didn't get what they needed to because of unplanned client follow-ups or emergencies that came up, time tracking down documents. That's a signal of failure demand that I've talked about before. And if those same patterns keep showing up, it tells you that there's an upstream process that needs attention or maybe that you're still operating over capacity because work keeps colliding with itself.

Now here too, there sometimes is a tendency for people to want to dive in and just start working, and I want you to prevent that too. So if someone surfaces a need, like this needs a strategy session or I'm blocked until Malcolm finishes his piece, acknowledge it, but then schedule time to address it and move on. Don't solve it in the stand-up. The stand-up is about scanning and surfacing work items, not workshopping them.

Now, hopefully you can see that those two core cadence meetings, the weekly planning and the daily stand-up, they work in parallel, right? The weekly planning looks back to last week and forward to the next one. Daily stand-ups look back at yesterday and forward to today. They've got the same structure but a different time horizon. They're both focused on deliverables, and they're both oriented towards surfacing issues and making commitments.

And again, both of these are really important for that mindset change because this is where “we should” becomes a pattern of “we do.” This is where the work that never really feels urgent finally gets scheduled attention. And this is where that pattern of consistent and predictable progress helps tone down those urgency signals, real and faked, that ramp up all the stress levels.

Now, the third cadence I'm going to touch on briefly, and that is the periodic retrospective. Unlike weekly planning and stand-up meetings, this one is less about the work itself and more about the team's tools, processes, and skills they need to actually do the work well. And I've done a few episodes on retrospectives already, so I'm not going to deep dive here.

But I will say they are borderline magical for capturing team concerns and aligning around opportunities for improvement. My biggest lament is that teams don't hold retrospectives nearly often enough. In my opinion, they should happen at least quarterly, and once every month or two is great.

So hopefully that gives you a good idea about how the right cadences can help you organize work and align around progress. But I also want to make sure you see how they start to provide that new reward structure that favors predictability and consistency instead of behaviors that might seem sexier, but are ultimately problematic. But I want to warn you, if you have people on your team who already rely on those problematic behaviors, they might resist these changes. And I want you to be careful. Do not listen to this podcast and then come and hand down a new rule from on high.

There's a great line from systems thinking expert Peter Senge, and he says it's not so much that people don't like change, but they definitely don't like being changed. So to help you avoid resistance to that change, it might be helpful for me to name some archetypes of working people that you might run into.

The first is the hero. This is someone whose identity is built around being the emergency responder. In an urgency coded practice, that gets celebrated with gratitude and status and visibility. But when your law practice becomes calm and predictable, that can threaten their sense of purpose.

I liken it a little bit to Loki from Norse mythology and obviously the Marvel movies. When things are going well, he creates chaos because that's where Loki finds meaning. And your team member might unconsciously do the same thing. They might create urgency because that's when they feel valuable, or they might subtly undermine these new systems that are creating stability that threatens their identity.

The second archetype you might run into, I think of as the operator, the back channel fixer who facilitates the queue jumping through personal relationships. And in return, they get special treatment, insider status, or sometimes personal favors. And I think of this a little bit like Wormtongue from Lord of the Rings, if you know that story. And he derives his power from being that special advisor with access and influence.

If you're more of a Harry Potter fan, Peter Pettigrew kind of falls into this category as well. But either way, when FIFO and transparency remove the need for back channels, these operators lose their special status and they might quietly work to recreate the conditions where they're the ones greasing the skids. This archetype is a little more insidious than the hero because it's harder to spot, but I want you to look out for it.

The third archetype is someone that I refer to as hiding under the busy cloak. These are folks who have cultivated an identity as kind of indispensable across multiple areas while maintaining this like funny opacity around what it is that they actually do. And they use this defense that I'm too busy, and it's almost unfalsifiable, which is the point. If they're late with X, they say, “I was doing Y.” Didn't finish Y, then they say, "Well, I was tied up with Z."

In an urgency driven system, then being, quote, too busy becomes sort of this proof of value, and it also protection from accountability. But when cadence meetings require accountability around actual progress towards deliverables, and when your prioritization policies make your priorities transparent, the busy cloak gets pulled back and that can leave some people feeling naked without it.

The thing I want you to keep in mind with all of these archetypes is that what we're asking for is nothing short of an identity shift, and that's one of the hardest things humans can do. So you have to be really intentional about creating and really celebrating those new reward structures so people have something to move toward instead of worrying about what they're losing from this change. I need you to really celebrate delivery and completion. And I want you to be careful to treat shortcomings not so much as personal failings, but as learning opportunities.

And importantly, I want you to recognize contributions to improving the system and make it clear that is important work too. The value that we're doing, it comes from reliability and not reactivity. And even with all that, you might fail with some people. And in my improvement work with firms, it isn't unusual for someone to leave as part of these process improvement efforts because they want to find a firm where they can maintain their comfortable identity.

And I get that that's hard, especially if they're skilled or if they've been with you for a long time. But remember that building a systems-dependent practice means reducing reliance on individual effort. We want systems over heroes, and that makes you way more resilient and more sustainable. Frankly, it's better to lose someone who needs the chaos than it is to preserve chaos in order to keep them.

So let me bring all this together. With WIP limits protecting your capacity, FIFO as your default prioritization, and cadence meetings driving that drum beat of progress, you're really building towards something that is fundamentally different than how most law practices operate.

And what you're building is a workplace where calm competence replaces urgency, where true deadlines get hit safely with time to spare, and where work gets delivered in a predictable order in a consistent way, and even where that intangible sort of maintenance work gets scheduled and dealt with before it becomes an emergency.

And as I just said, more importantly, you're building a place where the rewards are changing, where we're really favoring that steady progress over the drama of heroic saves. We want transparency over opacity, reliability over reactivity, because all of these things are better for system performance. Not only that, they're better for client outcomes and client experience. And ultimately, they're better for everyone's well-being, including your own.

Think of it this way, if it starts to get boring, you're no longer firefighting, you're conducting. And that's the chill practice I promised you at the beginning. And I guarantee you, it won't be boring. If anything, it'll be liberating. You'll get your excitement from finding new and profitable ways to apply this superior approach.

So next week, I'm going to build upon the cadences I've talked about today and the more subjective feedback loops that they provide. And that's a great place to start. But next week, I want to introduce a series of objective measures that you can also use to tighten and reinforce the behaviors we want to see. So stay tuned for that.

Alright, a lot of the things I discussed today are actually big reasons why my partners and I created GreenLine to begin with. We've built GreenLine to help you focus on just in time task assignments and to give you visibility into your team's existing commitments.

Because one of the easiest ways to overwhelm a team, despite having good cadences, is when too many tasks get assigned to people too early, too broadly, or too optimistically. And if you're using any sort of task automation in your existing law practice management tool, you probably know what I'm talking about.

Tasks pile up on people long before those tasks are actually ripe for work, and usually with no regard for whether the people actually have the capacity to work on them. And then guess what? Everyone suddenly looks really busy even though nothing is moving. And that's exactly the opposite of what WIP limits and FIFO are meant to protect.

With GreenLine, we're really intentional about supporting pull and not push-based systems. Work gets accepted or assigned only when someone has actual capacity to take it on, not weeks in advance as kind of wishful thinking. And we have specific tools that let you see at a glance what's actually on each person's plate right now, not as a performance scoreboard, but as a capacity mirror.

And those things help make your stand-ups run more smoothly, they make your planning meetings more honest, and they help prevent that quiet slide back into the world of overload and urgency.

And GreenLine doesn't create the discipline, the cadences do, but it's designed to reinforce the behaviors that you're trying to build instead of undermining them. To see how GreenLine supports these and other Agile lawyering practices, head on over to greenline.legal to learn more. And be sure to hit that book a demo button and we'll be happy to give you the full tour.

Alright, that's it for today. If you found today's episode helpful, please spread the word by sharing it with a friend or a colleague. And of course, to keep up with this 101 series, hit the follow or subscribe button in your favorite podcast app.

If you have thoughts or questions about today's episode or topics you'd like to hear me discuss, please don't hesitate to reach out to me at john.grant@greenline.legal.

As always, this podcast gets production support from the fantastic team at Digital Freedom Productions, and our theme music is “Hello” by Lunareh. Thanks for listening, and I will catch you again next week.