

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**



### **Full Episode Transcript**

**With Your Host**

**John E. Grant**

[The Agile Attorney](#) with John E. Grant

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

When your legal team is running at about 80% utilization, then any single piece of work is going to take about five times as long as it would if it were the only thing on your plate. When you get up to 95% utilization, that delay jumps to about 20 times. And I've mostly used that curve on this show as a warning. But what's interesting is the same thing runs in reverse. And in today's episode, I'm going to tell you how you can use this relationship between delivery time and utilization to your benefit.

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.

Hey, everyone, welcome back. So as we sort of slide into the groove of this fall season, hopefully you're feeling like you're returning to something of a normal workflow. There's a lot that can go on and it's interesting.

I'm going to take some of my cues for this episode from one of my coaching conversations with my clients, as I often do, and what this particular client was reporting, and they're not really the only one, is that even though they're back in what they think of as sort of the normal workflow of fall, right, summer's over, their kid's school has started, everything should be kind of settling down.

And yet somehow they're feeling even more overloaded than they were just a month or two ago. And I wanted to sort of break down for you why that is with this particular firm, maybe what are some things that might be causing a similar thing in your law practice? And then also talk about some solutions to it. So hopefully we can get to all of those things today.

And, you know, one of the things, and it's funny, I think that feeling that things were maybe moving a little bit more smoothly in this summer might have been a little bit of a false signal or, you know, who knows what it is.

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

We had the Labor Day holiday, obviously a few weeks ago, depending on when you're listening to this, and I was actually out of town this weekend or that weekend this year, but normally I kind of like staying in town on those weekends.

I live in Portland, Oregon, and because a lot of people leave town, what I find is that the city just feels easier, right? Even if there's people that are coming to town, they have a different pattern, a different thing they're trying to do, and so I just feel like I can get around and do things and kind of enjoy the city, enjoy what's going on out in the world a little bit more on those long weekends when maybe some of my neighbors and friends have vacated and are out doing different things. If you live in a really touristy town, probably you feel the opposite, right?

Probably you want to get the heck out because a lot of people are coming to you. Portland's big enough that it seems to absorb the tourists in one place, but for the residents, it allows us to do something a little bit different. And so I wonder if maybe that's a little bit what was happening in this person's law practice, right? Maybe some people on the team were out, maybe clients were doing their own thing.

And so I think maybe the firm or the practice could have just felt a little bit quieter just because some of the normal hustle and bustle wasn't there, even though maybe capacity was down because people are on vacation. And I happen to know this attorney also took some pretty good vacations this summer. So there's any number of things that could play into it. But the thing about this client in the recent call is they're reporting that all of a sudden, even though they feel like they're back in this place where things are normal, they're feeling overloaded and some specific language they use, right?

I feel like I'm even more of a bottleneck than I ever was, or maybe that the bottleneck feels more acute than it did before, feeling like there's a harder time keeping up with things. And then when we looked at their Kanban

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

board, we saw some interesting patterns, right? So number one, their total case count was up. And for this firm, I think it was a couple of things. Number one, again, over the summer when capacity was down, even if things felt lighter, they weren't closing as many matters, right? And so as I talk about all the time in this podcast, I like a law practice where the intake and the outflow are well aligned with each other, right?

This is the drum-buffer-rope stuff I talked about back in the Agile Attorney 101 series, and it's specifically that rope, right? We want the intake door tied to the outflow door, so we're not letting more work into our practice unless other work has already gone out. And I think when you're having holidays, summer seasons, whatever, at least for this particular firm, that didn't happen. So output or outflow, right? Completing matters had slowed down. And then the other piece of the summer season or the end of the summer season is, I think, that they were seeing an uptick in demand, right? So just like things slow down for you, in this case, things are probably slowing down for your clients or for this firm's clients.

And so when all of them were done with their summer breaks is when they started sort of contacting the firm to get some legal work done. Different practices have different kinds of cycles or different kinds of seasonality, so yours may not match that exact one, but I think they all, most of them at least, tend to have something of a seasonal nature to them. So even if it's not summer season, you might recognize that pattern. And so again, the upshot for this firm was their WIP was up, their work in process, right? The number of active matters that they're currently managing.

And as I've talked about a lot on this podcast, when WIP goes up, then the length of time or the amount of time it takes to deliver any one unit of work, right? In this case, a matter in the law practice that will lengthen. And there's a couple of things that I've talked about. And frankly, I'll admit, I've probably conflated these things a little bit in the past. So I'm going to try to pull them apart slightly. They're related. And so it's not always easy to do

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

that. But the first one is Little's Law, and this is basically the relationship between the total amount of work in your system and how long it's going to take you to deliver the work on average.

So if you're finishing 10 matters a month and you are carrying 100 matters, then on average, each matter is going to take about 10 months for you to deliver. If you're only carrying 50 matters and you can finish 10 matters a month, and you're going to deliver on average a matter in five months. This is kind of just how it works. And obviously, I know not all of you have the kinds of practices where you can just deliver a matter in a period of months. There's a lot of dependencies that come into play. So I'm not saying that it's a perfect thing for every single law practice, but the overall mechanism is still true, right? The more open work you have, the longer it's going to take you on average to deliver a unit of work.

And that's just sort of one of these facts of life that has turned out to be true. Now, the Kingman's formula is related to that. But instead of just looking at how much work you're carrying, it looks at how busy the system is, right? And in this case, the workers on the team and what I would say is how intensely the people on the team are being utilized. And let me clarify here, when I talk about utilization here, I'm not talking about billing utilization, right? I know that's a typical law firm metric. I'm talking about actual utilization, right? How much of their total capacity is taken up with demands from their work, from their job.

Now, there's another component to the Kingman's formula, and I won't dive too deep into it right now, although I think it's worth mentioning, which is the other thing that causes delay in addition to utilization, is the variability of the work, right? How unpredictable the work is.

And so basically, the way that the Kingman's formula works is the busier the system is, or the people on your team are, and the more variability you have in the type of work that you're doing, or the lack of predictability you have in the work that you're doing, then the longer anything is going to

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

take. And the thing about the Kingman's formula is that the relationship between these things is not linear, it's geometric.

And this is part of what I think is accounting for my clients kind of sudden feeling of overwhelm. So like I said, their WIP is up a little bit. They closed a few fewer matters than they normally do over the summer, they brought in a few extra matters than they would normally do relative to what they've been closing. So they're carrying more cases than they normally would, but the feeling of overwhelm, the feeling of busyness, the feeling of work getting stuck or stalled or that I'm the bottleneck has gone up quite a bit.

And the reason why has to do with this geometric nature of the Kingman's formula. And I'll share a graph of this in the show notes because it really does help to look at the picture. But describing it to you, the thing that is fascinating is up to about 50% utilization, you don't notice the delay hardly at all, right? The delivery time it takes for the average unit of work doubles when you get to 50% utilization. But that sort of movement from 1x delivery time to 2x delivery time doesn't really show up that viscerally, right? It's really easy to miss.

And then even as you move to about 70% utilization, it doesn't change that much. The delivery time goes to about three and change times the length of time it would take to deliver a piece of work if it were the only thing you're working on, right? That's what these multipliers are compared against. But what's interesting is once you get above about 70%, then the slope of the graph starts to change dramatically.

And I've talked about this before, but the thing to know is that at 100% utilization, what you have is an asymptote, right? So you can actually never quite get all the way to 100% utilization. Your delivery time will spike to infinity, right? You will have complete and total gridlock and you'll be unable to deliver anything at all.

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

So again, to kind of give you some audio markers or some mental markers for what it looks like as you sort of move to the higher levels of utilization on this graph, at 80% utilization, it takes about 5x the time that it would take to just do the work if it were the only thing you are working on, at 90% utilization, that moves to 10 times the length of time. And then at 95% utilization, it takes 20 times as long to deliver any one unit of work as it would have taken if it were the only thing that you were working on. Now again, like I said, that number is kind of based on consistent variability.

So Kingman's formula has this other component and the more variable the types of work that you do are those numbers may be even higher right the delay might be even more or if you do a very simple type of law then the delay might be a little bit less. But the shape of the curve is going to be the same either way.

And I think that explains what's going on in this particular client law firm of mine where probably they were operating somewhere above 80% capacity already and my guess is that with that just little bit of extra work, the tracking costs, the overhead, the switching costs, all the little things that add up to basically conspire to take up more and more of the individual worker's capacity and then for this attorney, their personal capacity started to sort of throw them into this overload place where they started to hit a steeper and steeper part of the curve and they're really feeling it and I don't know if they're all the way to 95% utilization, it's kind of impossible to know, and it doesn't matter so much.

But when you're feeling these sudden changes, my guess is it's a safe assumption that you're operating at one of those extreme higher ends of the utilization curve. Now the good news for this attorney and her team, and for you if this is something that you're feeling, is that the same thing works in reverse.

So as we looked at the number of cases on their Kanban board, part of what we did in this coaching call is start the exercise that I have talked

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

about before, which is close the closable. Going through and finding cases that can either be pushed all the way into the done column, at which point you don't have to worry about them anymore, you don't have any more overhead tracking, whatever, right? You can basically repurpose your team's capacity to work on the remaining cases and not those ones that are close to done.

Or in some situations you need to get to these fish-or-cut-bait moments with your certain matters, certain clients, where if the client is no longer a good fit, the matter is not a good fit, the clients not being responsive to your needs, is dragging their feet, whatever it happens to be, figuring out some way to to close those cases as well. And again, the good news under the Kingman formula is that you only need to close out or move about 5% of your total work out of your system, especially if you're up at those higher utilization ends, to be able to feel significant relief in terms of the stress, the delay, the bottlenecks, whatever that are going on in your practice.

For a really good illustration of this, I'm going to point you back to episode 21 of this podcast, where I talked about Laura's story and just a quick refresher, right? She was a small firm practice, but they had 65 total cases open in their system. They were doing family law at the time, still doing family law, although different things now. 12 of those were what they wound up designating as red file cases, and they basically counted those as double because those are the high complexity cases, so it was our way of figuring out the variability.

And that basically came to a weighted number of 77 open cases in this practice and that number on an absolute basis isn't useful but Laura was in this position a lot like my client from earlier this week where they just she was feeling overwhelmed right things were just moving slow and it was too much so again in that episode I asked Laura what her ideal number was and she said 40 and that's obviously more than 5% cuts, but over the

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

course of a few months, they really focused on that close the closable, the nearly done cases on the right side of their Kanban board.

They then went through the other client load and sent sort of these fish-or-cut-bait messages to non-responsive clients. Also, they got pickier about intake, right? So they slowed down, they did the intake pause, the whole thing where when your bathtub is overflowing, yes, you want to clear the drain, but you also need to turn off the tap. But over the course of about three months, they were able to go from that 77 weighted average down into the mid 40s, and they didn't quite get to 40. But Laura's stress and work-life balance changed remarkably as a result of that.

I talked about this in a slightly different way back in episode 105, where I was referring to a transactional law firm where we did some process improvement work to really try to smooth out the workflows and also setting something of a WIP limit, right, a total carrying capacity limit on the number of matters that they were going to hold open in this transactional workflow. And looking at the graphs here, we saw really clear trends that as we were able to close more cases and then only open cases that match the number of cases we were closing, their cycle time, right? The time it took to deliver a full matter dropped from 192 days at the 85th percentile down to 146 days.

So about a 25% improvement in speed just by being more intentional about their WIP and their throughput. So if you're in this position, right, if you're feeling like you are, for whatever reason, back into one of these places of overload, of overwhelm, of bottleneck, of delay, a few things that I think you can do almost immediately that is going to, again, not get you down maybe to where you would like it or a perfect situation, but at least help really turn down the temperature, right? Get that practice from the roiling boil down to the gentle simmer that I like to get people to.

One of those things is the intake pause. I know that it's harder to do in some practices. I know that it's a hard mindset to wrap your head around,

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

right? We all have this sort of scarcity mentality, or at least it's normal to have this scarcity mentality where we don't want to turn away work. But again, it's my experience across my clients that there are not that many practices that are suffering for want of work right now. If you have one of those, I'm not saying that you're doing it wrong. There's a lot of variability out there, but most of the people I'm working with basically have more demand than they can possibly handle, right? They are turning away cases and they probably should be turning away more.

A quick aside, when you talk about filtering for what cases you do bring in, don't bring in the complex ones, right? So again, variability is part of one of the factors in Kingman's equation. So the more you take on highly complex cases, the more it's going to tax your capacity, tax your utilization, and therefore slow down the delivery, not just of that case, but of all the cases in your workflow.

So trying to be really more selective, right, is one of the reasons I love to talk people out, talk law firms out of certain practice areas or talk them into dropping practice areas is because by doing that, we reduce some of the complexity, we reduce some of the variability. The other thing that you can do is that close the closable exercise, right? And so just as a quick refresher, there's a few different places where you can find success in closing the closable.

One of them is in cases that are already done or already close to done, not already done, obviously. But if they're close to done and they just need a few things to get them across that finish line, then I suggest you schedule the time, create the space, assemble the team so that you can get those cases all the way across the finish line. That's gonna help you free up capacity almost immediately. And it's gonna do the thing that you wanna do, which is deliver the work for the client, right? So hopefully that's a win-win situation.

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

One of the other filters you can use is for clients that aren't holding up their end of the deal. And of course this helps if you made it really clear up front what the deal is supposed to be. But either way, if you've got non-responsive clients or you've got clients that aren't being well engaged in the handling of their own matter, then either take steps to re-engage them or figure out if it's time to cut bait and maybe cut them loose, send them to another firm, another attorney, or some other way of solving their legal matter.

The third place where close the closable might come into play, or the third technique, I guess, for closing the closable, I talked about back in episode 97, which is that Marie Kondo test, right? Which is, does this matter still make sense to you? Does it spark joy? Does it fit your expertise? Does it fit your team? You don't have to store or work on cases that aren't right for you anymore. You obviously can't leave people in a bad situation, but you definitely are allowed to decommit if there's something that is no longer working for you for whatever reason.

The last thing I'll suggest, right, if you're doing that thing where you really do feel slammed or super overwhelmed, I would come up with a goal for how many cases you need to close. And I think the right number is probably in the neighborhood of about 5% of your total case count, right? If you can reduce your number of active cases by just 5%, you're going to feel a significant improvement, especially if you're at that running towards the red line part of the Kingman's formula graph.

All right. That is it for today. Hopefully you got something useful out of this episode. If you did, I would appreciate it if you would take a moment to rate or review the show in Apple Podcasts or Spotify or whatever tool you listen in. And if you don't already, be sure to follow or subscribe so you can stay up to date on all of my new episodes as they come up.

## **Ep #136: Law Firm Overload: Why Closing Just 5% of Your Cases Brings Big Relief**

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 time.