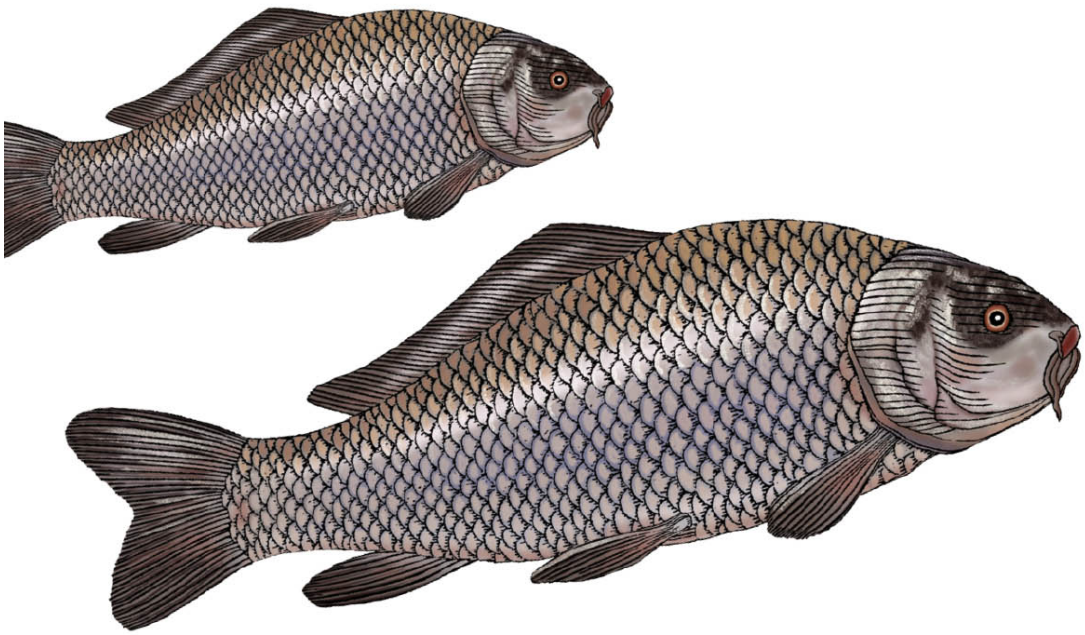
O'REILLY®

Vorwort von David J. Anderson

Kanban

kurz & gut

O'Reillys Taschenbibliothek



Susanne Bartel

Inhalt

KAPITEL 4

Limitiere die parallele Arbeit (das WIP)

Als Einstieg starten wir gleich mit einem kleinen Gedankenexperiment zur Motivation für diese Praktik. Anschließend finden Sie eine Zusammenfassung der Praktik mit wichtigen Steuerungselementen und der zugrunde liegenden Gesetzmäßigkeit namens Littles Gesetz. Auf dieser Basis schauen wir uns dann die Ursprünge, die Gestaltung und die Auswirkungen von Pull-Systemen an. Abschließend finden Sie Hinweise und Tipps, um WIP-Limits auch in Ihrer Umgebung erfolgreich anzuwenden.

Motivation für die Praktik

Lassen Sie uns mit einem Gedankenexperiment beginnen. Dieses sollte Ihnen einen guten Einstieg in die Thematik bieten.

Stellen Sie sich vor, Sie hätten einen Tank, in dem sich 100 gleichartige Bälle befinden und bearbeitet werden. Jede Stunde verlassen 10 Bälle unten den Tank (die Lieferrate). Die Mechanik im Tank ist so, dass die Reihenfolge der Bälle beibehalten wird, und wie gesagt: Alle Bälle sind gleich groß.

Gerade ist aus dem vollen Tank ein Ball unten herausgeploppt (Lieferung!), und Sie geben oben einen farbig markierten Ball als 100. Ball hinein (siehe Abbildung 4-1). Wie lange dauert es, bis der markierte Ball unten wieder aus dem Tank kommt?

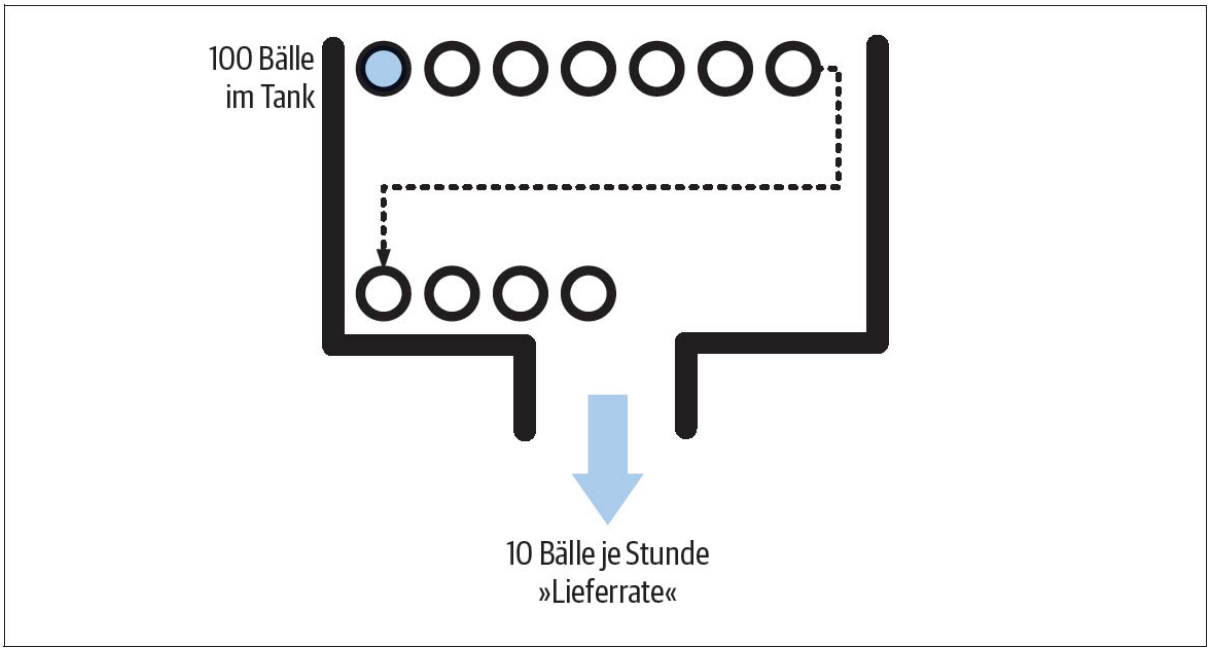


Abbildung 4-1: Bälle im Tank, Teil 1

[Denkpause. Bitte selbst überlegen, ehe Sie weiterlesen.]

Es dauert 10 Stunden, das ist recht naheliegend, oder? 100 Bälle sind im Tank, pro Stunde verlassen 10 Bälle den Tank, nach 10 Stunden verlässt der 100. Ball wieder den Tank.

Sie ahnen es sicher: Der Tank entspricht Ihrem Arbeitssystem, die Bälle Ihren begonnenen Aufgaben. Die Frage zu den Bällen im Tank beantworten die meisten Menschen nach kurzem Nachdenken korrekt. Interessanterweise fällt es jedoch vielen schwer, die gleichen Schlüsse in Bezug auf ihr eigenes Arbeitssystem zu ziehen.

Zurück zu unserem Tank: Denken wir nun weiter. Sie möchten, dass alle Bälle in der halben Zeit, das heißt innerhalb von fünf Stunden, durch den Tank kommen. Was müssen Sie dazu am Tank verändern?

[Hier haben Sie nochmals Zeit zum Nachdenken.]

Ich hoffe, Sie haben sich von mir aufs Glatteis führen lassen und »Den Ausgang verdoppeln!« gedacht. Das sähe dann aus wie in Abbildung 4-2.

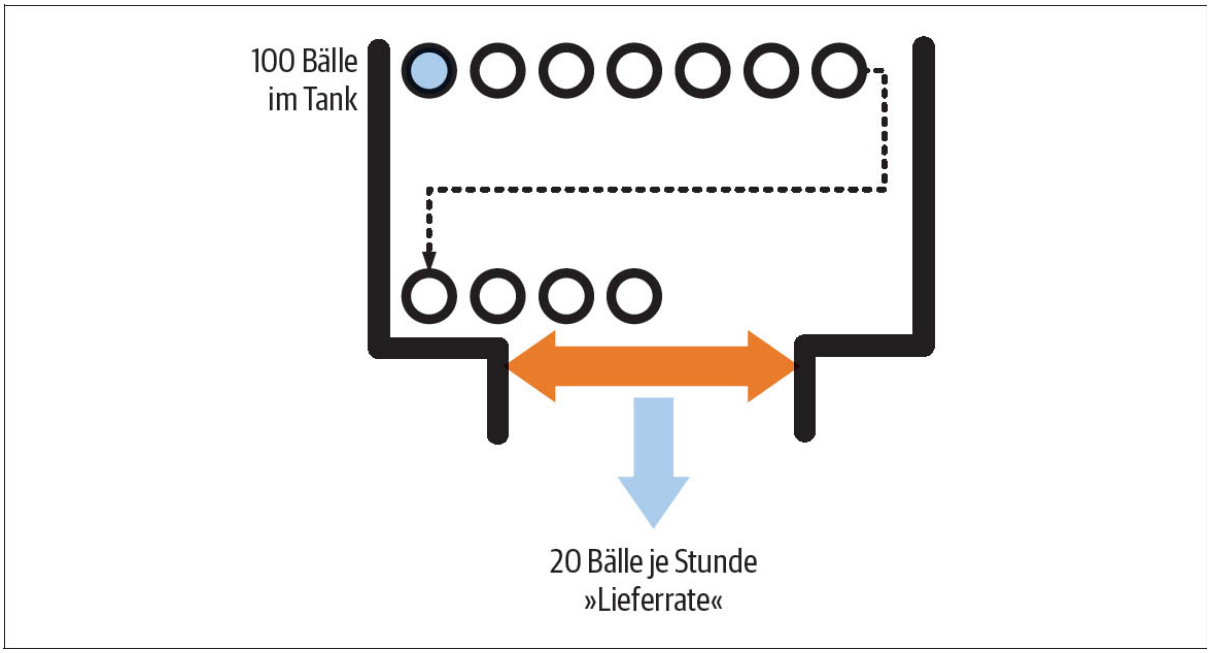


Abbildung 4-2: Bälle im Tank, Teil 2

Tatsächlich ist diese Antwort ein weitverbreiteter Managementreflex. »Wir brauchen mehr Leute!« oder »Wir brauchen noch ein Team!« Prinzipiell ist das in Grenzen natürlich richtig und würde helfen: Wenn wir den Tank so gestalten, dass pro Stunde nun 20 Bälle hindurchkommen, wird unser markierter Ball bereits nach 5 Stunden wieder herauskommen.

Wir haben allerdings mal so eben die Lieferrate verdoppeln müssen, um dies zu erreichen! Wie könnte man die Lieferrate verdoppeln? Für den Tank: bauliche Maßnahmen. Übertragen auf Wissensarbeitssysteme? Typischerweise mehr Leute, neue Teams, höherer Automatisierungsgrad, Prozessverbesserungen. Alles Dinge, die leider nicht über Nacht und nicht billig zu bekommen sind. Außerdem skalieren diese Systeme nicht linear. Sie können also nicht so einfach ein zweites Team danebensetzen und davon ausgehen, dass sich die Lieferrate mal eben verdoppelt. Ab einem gewissen Punkt kann die Lieferrate aufgrund von Abhängigkeiten und Kommunikations- sowie Koordinierungsaufwänden sogar wieder sinken.

Nun ja, zum Glück haben wir noch eine weitere Stellschraube, an der wir drehen können. Haben Sie eine Idee? Genau, die Größe unseres Tanks! Wir lassen nur noch 50 Bälle gleichzeitig hinein, das heißt, unser markierter Ball wird als 50. Ball schon nach 5 Stunden (bei einer Lieferrate von 10 Bällen pro Stunde wie zu Beginn unseres Gedankenexperiments) herauskommen.

Dieses Verhalten folgt den Gesetzmäßigkeiten von *Littles Gesetz*, die im Abschnitt »Littles Gesetz: vom Bälle-Tank zum Wissensarbeitssystem« auf Seite 86 näher erläutert werden. Hier vorab in der groben Form für dieses vereinfachte

Modell (Tank ist stets voll, alle Bälle sind gleich groß und behalten die Reihenfolge):

$$Laufzeit = \frac{Anzahl\ der\ Bälle\ im\ Tank}{Lieferrate}$$

Für den eben betrachteten Fall ergibt sich eine Situation wie die in Abbildung 4-3.

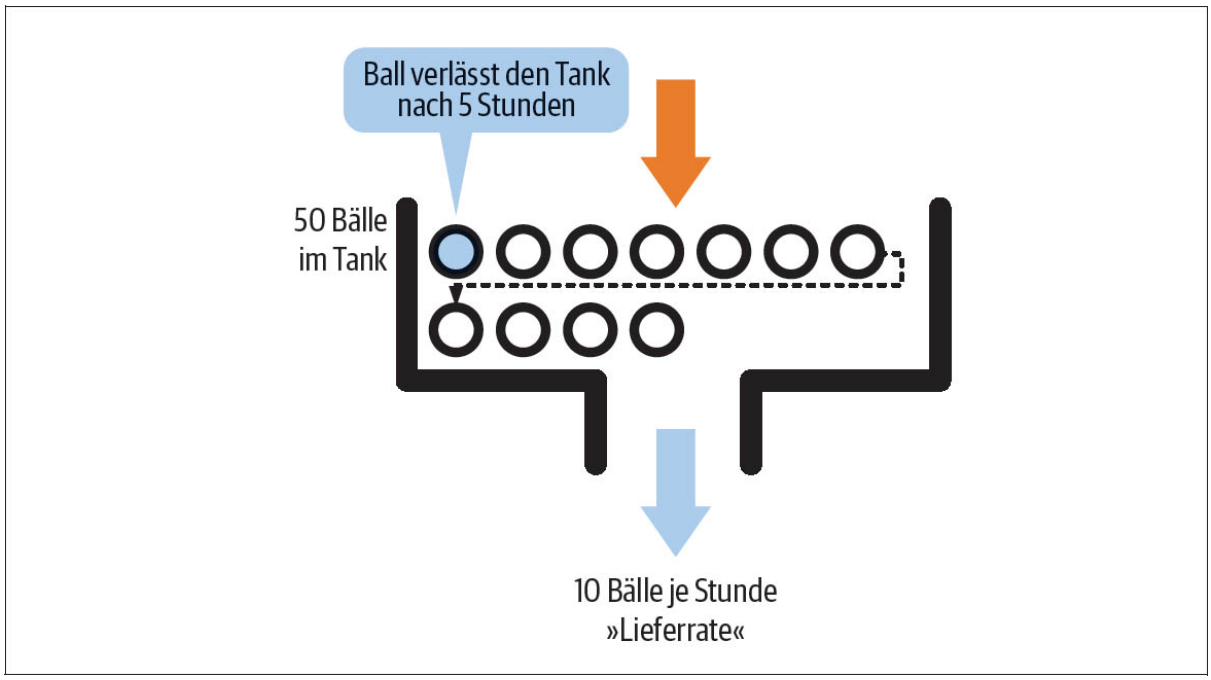


Abbildung 4-3: Bälle im verkleinerten Tank

Mittels der Formel ausgedrückt, ergibt sich also:

$$Laufzeit = \frac{50\ Bälle}{10\ Bälle\ je\ Stunde} = 5\ Stunden$$

An dieser Stelle haben vielleicht auch Sie ein »Aber das ist doch schummeln!« auf den Lippen. Na ja, es ist zumindest sehr schlau. Und es erzeugt sehr nützliche Effekte. Zudem ist es als rein organisatorische Maßnahme bzw. Vereinbarung deutlich günstiger als die weiter oben betrachtete Option.

Zu den Effekten: Schauen Sie sich doch »Ihre 100 Bälle« in Ihrem Tank an, also all das, was aktuell in Arbeit ist. Wenn Sie mal ordentlich auskehren, finden Sie wahrscheinlich jede Menge:

- Zombies – schon lange tot, aber geistern noch herum – sowie

- ewige Verlierer – werden immer übersteuert (»XYZ ist jetzt wichtiger«), dauern ewig, und einige werden schließlich zu Zombies.

Weniger Bälle gleichzeitig im Tank (oder in der Luft ;)) führen andererseits zu deutlich reduzierten Koordinierungsaufwänden. Dies werden Sie sehr schnell positiv wahrnehmen. Ganz nebenbei werden dadurch Kapazitäten freigesetzt. Außerdem werden unangenehme, aber für ein effektives Operieren notwendige Entscheidungen erzwungen, insbesondere das Selektieren und Sequenzieren am Eingang. Sieben Sie Zombies und ewige Verlierer frühzeitig aus, damit diese im Folgenden nicht Ihr System verstopfen.

Die von Ihnen bewusst gewählten Bälle (nämlich genau jene, die aktuell die höchste Wichtigkeit und Dringlichkeit haben), die es nun in den Tank schaffen, haben gute Chancen, fokussiert und mit minimalen Verzögerungen erledigt zu werden! Die Begrenzung parallel erlaubter Arbeiten im System hat außerdem den Effekt eines *Enabling Constraint* (einer befähigenden Einschränkung) und wirkt sich sehr positiv auf Zusammenarbeit aus.

Wenn Sie gerade mit Kanban starten, operieren auch Sie wahrscheinlich deutlich oberhalb Ihrer tatsächlich vorhandenen Kapazität. Dies erkennen Sie z.B. an:

- wechselnden Prioritäten und damit verbundenen
- häufigen Wechseln zwischen verschiedenen Arbeitsgegenständen,
- der Existenz von Zombies und ewigen Verlierern sowie einer
- der sehr langen Dauer, bis Arbeiten abgeschlossen werden.

Konsultieren Sie hierzu gern auch Ihr Bauchgefühl oder befragen Sie Kolleginnen und Kollegen.

Nach dieser Einführung wenden wir uns nun der darauf basierenden Kanban-Praktik zu.

Beschreibung der Praktik

Sie erfahren nun als Erstes, wie Sie die sogenannten *WIP-Limits* anwenden können, um die Größe Ihres »Arbeitstanks« zu begrenzen.

WIP und WIP-Limits

WIP ist eine Abkürzung für *Work In Progress* und auch im deutschen Sprachgebrauch üblich. WIP bezeichnet die Anzahl der parallel in Arbeit

befindlichen Dinge (z.B. in Arbeitsschritten, pro Person oder auch auf einem ganzen Board). Dies entspricht der Anzahl der Bälle im Tank.

Wie reduziert man nun praktisch die Bälle im Tank bzw. die Aufgaben im System? Dazu führt man sogenannte WIP-Limits (Work-in-Progress-Begrenzungen)[1] ein.

WIP-Limits

Ein WIP-Limit definiert die zu einem Zeitpunkt maximal erlaubte Anzahl an Karten in einem definierten Bereich des Kanban-Boards. Dies kann beispielsweise die Anzahl der Karten in einer oder in mehreren Spalten oder Zeilen sein oder auch die Anzahl der Karten pro Person.

In Kanban-Boards werden WIP-Limits typischerweise durch eine Zahl in einem Kreis oder am jeweiligen Bereich dargestellt.

Wie Sie bereits durch das Gedankenexperiment mit dem Tank erfahren haben, kann diese kleine, unscheinbar wirkende Zahl des WIP-Limits große Wirkung entfalten. Lassen Sie uns hier noch einmal das Board von 101 Brands anschauen und in den oberen Bereich hineinzoomen, zu sehen in Abbildung 4-4.

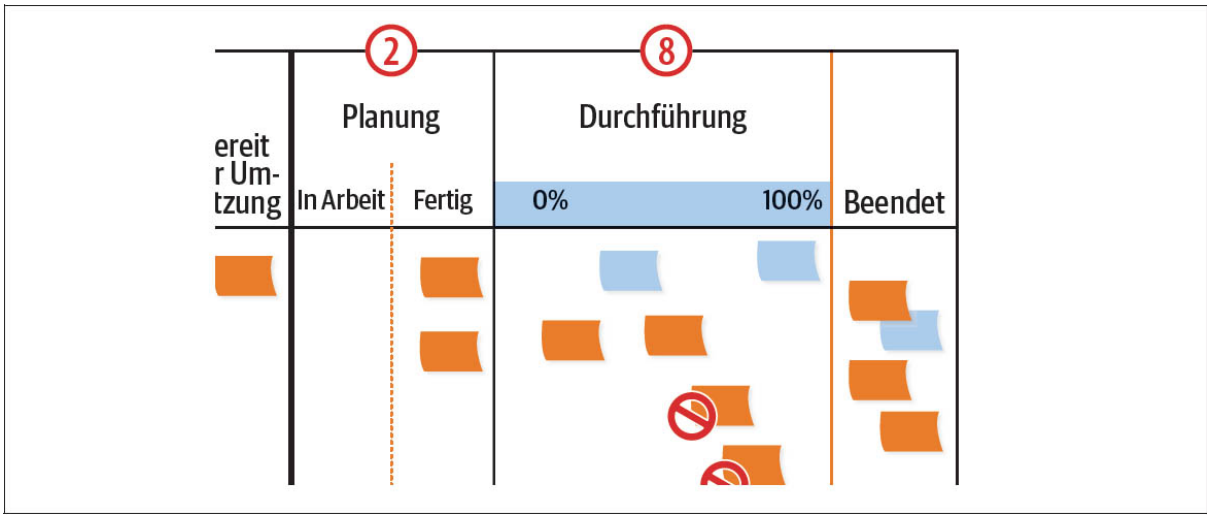


Abbildung 4-4: Beispiele für WIP-Limits, hineingezoomt

Die Zahlen 2 und 8 in den Kreisen über dem Board zeigen an, dass nur 2 bzw. 8 Projektkarten in der entsprechenden Spalte hängen dürfen. Die Darstellung in einem Kreis (und oftmals in roter Farbe) hat sich als Standard in haptischen Boards eingebürgert.

Lesen Sie den Kasten »WIP-Limits« oben gern noch einmal durch. Der Schlüssel hier ist »zu einem Zeitpunkt«. Dies bedeutet also nicht je Woche, Tag oder Monat. Sie können eine Spalte mit einem WIP-Limit von 3 haben, durch die in

einer Woche 15 Karten durchrauschen. Von denen sind jeweils aber nur 3 gleichzeitig in der Spalte. Wenn eine Karte den Arbeitsschritt verlässt, wird ein Platz (das virtuelle Pull-Signal, auch *kanban* genannt und hier üblicherweise klein geschrieben) frei, den die nächste Karte einnehmen kann. Es ist dabei unwichtig, welche der drei Karten weiterzieht und den Platz freimacht. Eventuell ist am Montag eine Karte in der Spalte gelandet, wurde dort z.B. durch fehlende Informationen blockiert (hält also einen Platz weiterhin besetzt), und durch die beiden anderen freien Plätze sind in der Zwischenzeit 10 andere Karten durch die Spalte gewandert. Andererseits kann es auch passieren, dass bei einem WIP-Limit von 3 dieselben drei Karten über Tage oder Wochen dort hängen.

Die Begrenzung der Arbeit, die in das System darf, ist ein wichtiger Schlüssel zur Reduzierung von Überlastung und den damit verbundenen Kontextwechseln und Umschaltverlusten. Ziel ist es, den Bedarf der Kunden und die Leistungsfähigkeit des Systems über die Zeit in Balance zu bringen.

Praktik »Limitiere die parallele Arbeit (das WIP)«

»Stop starting, start finishing!«[2] Zu viel gleichzeitig begonnene Arbeit führt zu negativen Effekten wie Umschaltverlusten und längeren Durchlaufzeiten.

Ziel der Praktik ist es daher, die begonnene Arbeit im System entsprechend der tatsächlich vorhandenen Kapazität zu reduzieren. Dazu werden WIP-Limits und das Pull-Prinzip eingesetzt, um ein Pull-System zu etablieren (siehe Abschnitt »Pull-Systeme« auf Seite 92).

Damit können Reaktions- und Lieferfähigkeiten verbessert werden. Durch den verbesserten Fokus und eine höhere Konsistenz im Prozess steigt die Qualität.

Die Wirksamkeit der getroffenen Vereinbarungen wird kontinuierlich beobachtet und an sich ändernde Gegebenheiten angepasst. Fertiggestellte Arbeit wird höher bewertet als angefangene. Dies stellt in der Organisation oft kulturell eine Änderung dar.

Littles Gesetz: vom Bälle-Tank zum Wissensarbeitssystem

Bevor wir uns Littles Gesetz, das wir oben bereits erwähnt haben, näher anschauen, lassen Sie uns kurz den Bälle-Tank aus dem einführenden Gedankenexperiment mit dem typischen Wissensarbeitssystem vergleichen. Auf einen Bälle-Tank übertragen, könnte sich unsere Wissensarbeit zum Beispiel so darstellen wie in Abbildung 4-5.

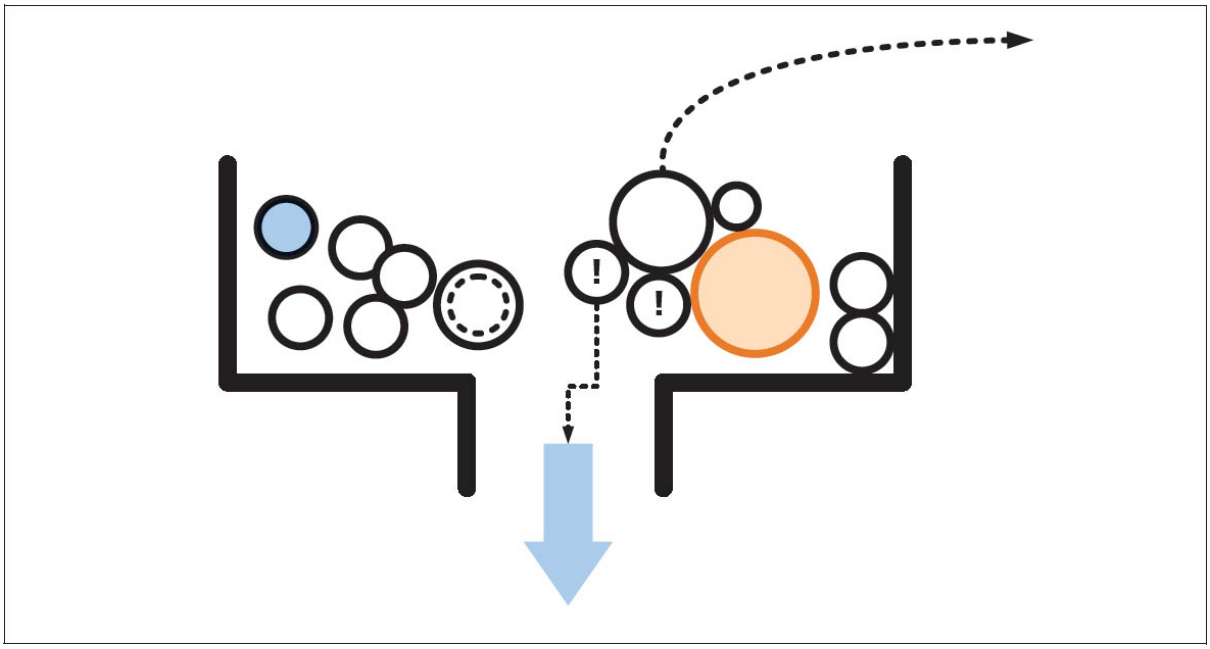


Abbildung 4-5: Wissensarbeit im Bälle-Tank

Die Bälle sind unterschiedlich groß. Im Tank herrscht Unordnung, Bälle überholen sich gegenseitig, einige werden zwischendurch rausgeworfen oder bleiben in einer Ecke hängen. Einige Bälle wachsen, andere schrumpfen. Die Anzahl der Bälle im Tank schwankt über die Zeit. Und manchmal werden sogar mehr Bälle hineingepresst, als eigentlich Platz ist. All dies sind Gründe, weshalb die oben genannte vereinfachte Formel für unsere Arbeitssysteme (Laufzeit = 50 Bälle / 10 Bälle je Stunde = 5 Stunden) so nicht anwendbar ist.

Littles Gesetz beschreibt das gleiche Verhalten, allerdings unter Verwendung von Durchschnittswerten und an bestimmte Voraussetzungen geknüpft, um auch für unsere dynamischen Wissensmanagementsysteme anwendbar zu sein.

Definition

Die Formel ist benannt nach Dr. John D. C. Little. Dieser formulierte und bewies 1961 diese Gesetzmäßigkeit in der Warteschlangentheorie.[3]

Littles Gesetz beschreibt in Produktionsprozessen jeder Art den Zusammenhang zwischen der Anzahl der Dienstleistungen, die einen Prozess durchlaufen sollen, und der Zeit, die zu ihrer Fertigstellung benötigt wird. Knapp umschrieben, besagt Littles Gesetz, dass sich der Arbeitsanfall zu stauen beginnt, wenn mehr Arbeitsaufgaben hinzukommen, als Arbeiten erledigt werden. In unsere heutige Begriffswelt übertragen, ergibt sich die im Kasten unten dargestellte Formel.

Littles Gesetz

Littles Gesetz bietet einen wichtigen Einblick in das Verhalten flussbasierter Systeme. Es beschreibt den Zusammenhang langfristiger Durchschnittswerte in einem stabilen System.

$$\text{Ø Laufzeit} = \frac{\text{Ø WIP}}{\text{Ø Lieferrate}}$$

Daraus folgt, dass wir die parallele Arbeit im System begrenzen müssen, um die Durchlaufzeit für Arbeitspakete zu optimieren. Damit liefert das Gesetz die Motivation für die Kanban-Praktik *Limitiere die parallele Arbeit (das WIP)*.

Die Gültigkeit von Littles Gesetz ist an gewisse Voraussetzungen geknüpft:

- Das WIP ist zu Beginn und am Ende der betrachteten Zeitperiode in etwa gleich.
- Das durchschnittliche WIP im System ist stabil, es gibt keine Trends (Entwicklung nach oben oder unten).
- Die Ankunftsrate entspricht der Abgangsrate.
- Alle Karten, die in das System kommen, werden schließlich das System verlassen, das heißt, es gibt keine Abbrüche.
- Das Durchschnittsalter der angefangenen Arbeit im System bleibt gleich.

Nicht zuletzt lohnt sich hier ein Gedanke zurück an Ihre Physiklehrerin. Natürlich müssen die verwendeten Einheiten konsistent sein! Achten Sie hierbei besonders auf die Laufzeit und die Angabe des Durchsatzes, gegebenenfalls müssen Sie normieren.

Sollten Sie nun gerade ein Fragezeichen auf der Stirn haben, hier ein Beispiel: Sie haben in den letzten 3 Wochen insgesamt 36 Features geliefert. In diesem Zeitraum befanden sich in Ihrem System durchschnittlich 6 Features. Sie möchten ermitteln, wie groß die durchschnittliche Laufzeit war. Schauen Sie gern noch einmal kurz in den Kasten oben zu Littles Gesetz:

$$\text{Ø Laufzeit} = \frac{\text{Ø WIP}}{\text{Ø Lieferrate}}$$

Für die Ermittlung der durchschnittlichen Laufzeit dieser Features (F) teilen Sie das WIP von 6 Features durch die Lieferrate von 36 Features in 3 Wochen. Daraus ergibt sich

$$\text{Laufzeit} = \frac{6\ F}{36\ F\ /\ 3\ Wochen} = \frac{6\ F}{12\ F\ /\ Woche} = \frac{6\ F}{12\ F\ /\ 7\ Tage} = \frac{6\ F}{1{,}7\ F\ /\ Tag} = 3{,}5\ \text{Tage}$$

Die durchschnittliche Laufzeit der Features war 3,5 Tage.

Da diese Laufzeiten aus Kundensicht wichtig sind und wir in 24/7-Zeiten leben, werden in der Regel Wochen mit 7 Tagen und Tage mit 24 Stunden angesetzt, also »tatsächliche« Zeit und keine Büroarbeitszeiten. Wenn dies in Ihrem Kontext unpassend ist, können Sie das natürlich anders vereinbaren. Falls Sie ein digitales Kanban-Tool einsetzen, wird dies all diese Fragen konsistent für Sie handhaben.

Achtung, diese mittels Littles Gesetz berechnete Durchschnittslaufzeit ist nicht geeignet, um Prognosen für die Laufzeit zukünftiger einzelner Arbeitspakete zu treffen. Mehr dazu finden Sie im Abschnitt »3. V: Kennzahlen für Vorhersagen (Prognosen)« auf Seite 168.

Praktisch auf ein Team angewendet, ergibt Littles Gesetz folgende Handlungsempfehlung: Wenn das Team bereits voll ausgelastet ist, sollte an weniger Arbeitspaketen gleichzeitig gearbeitet werden, damit Arbeiten zügiger abgeschlossen werden. Die Lieferrate wird dadurch nicht sinken, denn: Alles, was über die vorhandene Kapazität hinaus gestartet wird, wird sämtliche Arbeiten langsamer machen und *nicht* zu mehr abgeschlossener Arbeit in der gleichen Zeit führen! Wenn das Team hingegen freie Kapazitäten hat, kann es mehr gleichzeitig bearbeiten und abschließen, und die Arbeitspakete werden nicht länger dauern.

Schraube überdreht? Sollten Sie besorgt sein, dass Sie das WIP zu sehr senken, achten Sie auf diese Indikatoren:

- Es gibt vermehrt Leerlauf bei den Beteiligten.
- Es gibt übermäßig viele Übergaben bzw. Wechsel bei der Bearbeitung von Karten: Der durch die Kontextwechsel eintretende Kapazitätsverlust wird verschärft, wenn Teams oder Mitarbeitende überwiegend an Themen arbeiten, die außerhalb ihrer Kernkompetenz liegen und in denen sie nicht besonders effizient sind. Achtung, dies ist eine externe Form der Zusammenarbeit, die negativen Effekte treten oft deutlich später ein als erwartet. Daher überprüfen Sie am besten Ihre Kennzahlen.
- Messbarer Indikator: Die Lieferrate sinkt (vorausgesetzt, andere Faktoren sind stabil).

Zu Ihrer Beruhigung: Dies habe ich bisher in der Praxis noch nicht beobachten können. Einzige Ausnahme: bei Okaloa-Flowlab-Simulationen (siehe den Abschnitt »Okaloa Flowlab« auf Seite 214), wo ich von Zeit zu Zeit Teams anstachele, diesen »Kipppunkt« zu finden. Tatsächlich sind bei vielen Kanban-Systemen die gesetzten WIP-Limits immer noch zu hoch.

Weiter unten in diesem Kapitel im Abschnitt »WIP-Limits in der Praxis einsetzen« auf Seite 101 finden Sie Hilfestellungen dazu, wie Sie WIP-Limits einführen und

mit ihnen aktiv arbeiten können.

Littles Gesetz gehört auf jeden Fall auf die Liste der Themen für abendfüllende Gespräche in der Kanban-Community. Bei Interesse können Sie im Netz dazu weitere Artikel und Videos finden.

Nach diesem Ausflug in die Gesetzmäßigkeiten für das Verhalten flussbasierter Systeme wenden wir uns nun den Effekten des Einsatzes von WIP-Limits zu.

WIP-Limits machen Ungleichgewichte sichtbar

Tatsächlich kann wohl nicht oft genug darauf hingewiesen werden, dass WIP-Limits vorhandene Probleme aufzeigen und keine neuen Probleme (wohl aber Unbequemlichkeit) schaffen.

Ein Kernproblem ist typischerweise die Unausgewogenheit zwischen der Nachfrage (den angeforderten Leistungen) und der Leistungsfähigkeit bzw. den Lieferfähigkeiten. Dieses Ungleichgewicht bedeutet in den meisten Fällen, dass die Nachfrage die Leistungsfähigkeit übersteigt. In unbegrenzten Systemen ist dies jedoch oft unsichtbar. Die überschüssige Arbeit »versteckt« sich in überfüllten Arbeitssystemen (siehe Abbildung 4-6).

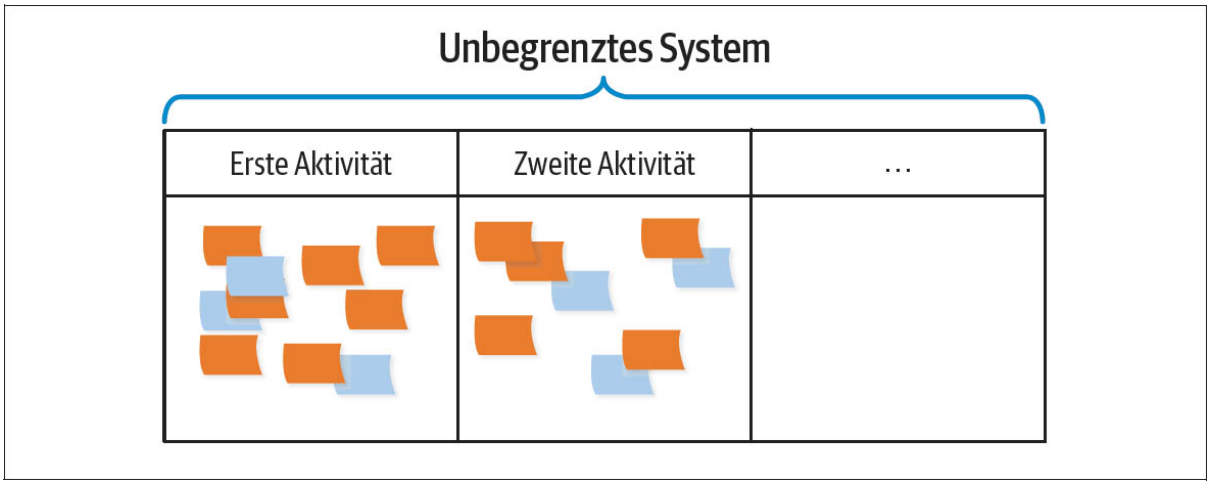


Abbildung 4-6: Ein unbegrenztes, überfülltes System

Werden Systeme auf die tatsächlich vorhandene Kapazität (d.h. die Fähigkeit, einen gewissen Mix an Arbeit gleichzeitig und in guter Qualität zu bearbeiten) begrenzt, verschwinden die überzähligen Anfragen natürlich nicht einfach. Diese sammeln sich – Sie ahnen es – *vor* dem System (siehe Abbildung 4-7). Falls Sie schon einmal eine Simulation wie Okaloa Flowlab oder auch FeatureBan durchgeführt haben, konnten Sie dies dort sehr plastisch erleben.

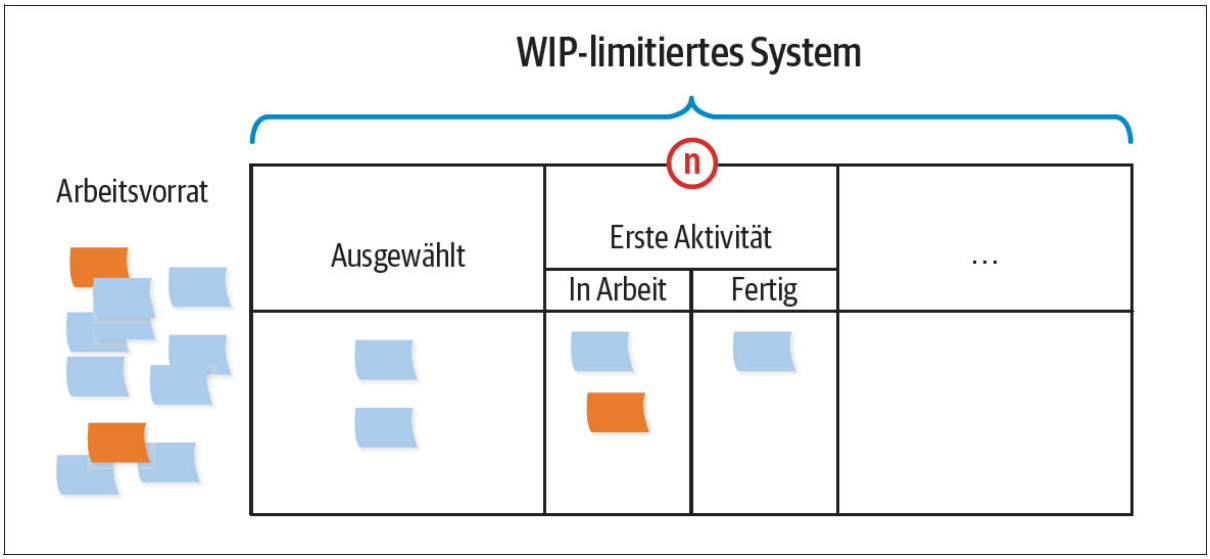


Abbildung 4-7: Arbeit staut sich vor dem WIP-limitierten System.

Dies ist zunächst unbequemer, führt aber zu deutlich klareren Erkenntnissen und zu mehr Steuerbarkeit. Wenn Sie beispielsweise beobachten, dass die Warteschlange vor Ihrem System anwächst, ist das eine Folge davon, dass Sie kontinuierlich mehr Anfragen bekommen, als Sie leisten können. Gleiches gilt auch für den umgekehrten Fall. Sie können also recht einfach überprüfen, ob Ihr System in Balance ist und ob ergriffene Maßnahmen Wirkung zeigen.

WIP-Limits sind schon die halbe Miete für Pull-Systeme. Diese schauen wir uns jetzt im nächsten Abschnitt an.

Pull-Systeme

Bevor wir uns *Pull* näher anschauen, lassen Sie uns kurz den gegenteiligen Ansatz, das *Push*, betrachten: Stellen Sie sich vor, Sie sitzen am Schreibtisch und sind in Ihre Arbeit vertieft. Eine Kollegin aus Ihrem Team oder einer anderen Abteilung kommt zu Ihnen, legt Ihnen eine Akte auf den Tisch mit dem Satz: »Ich benötige das bis Donnerstag!« und verlässt sofort den Raum. Vermutlich haben Sie eine ähnliche Situation schon erlebt. Sie bekamen Arbeit »aufgedrückt«, die Sie eventuell gar nicht mehr schaffen, und Sie hatten keine Gelegenheit, diese aktiv anzunehmen und die Bearbeitung zu dem genannten Termin zuzusagen.

Dies klingt etwas drastisch, ist aber in ähnlicher Form z.B. auch bei einem klassisch gesteuerten Projekt der Fall: Es gibt einen definierten, vorab bestimmten Zeitplan, und Arbeitspakete werden durch die Projektleitung den Beteiligten (die gegebenenfalls auch an anderen Projekten arbeiten oder durch operative Tätigkeiten beteiligt sind) zugewiesen. Ob es zu diesem Zeitpunkt tatsächlich Zeit und Kapazität dafür gibt, ist ungewiss.

Der Gegenentwurf dazu ist das Prinzip des *Pull*: Arbeit wird (weiter-)gezogen, wenn es freie Kapazität dafür gibt.

Wie praktisch alle Kanban-Konzepte kann auch das Pull-Prinzip auf verschiedensten Ebenen angewendet werden: von einem einzelnen Projekt bis zu neuen Projekten, die erst bei frei gewordenen Kapazitäten gestartet werden.

Wie könnte denn nun das *Push* der Akte in ein *Pull* gewandelt werden? Man könnte ein kleines Schild auf den Tisch stellen: »Neue Akten bitte hier rein (Posteingang). Wir geben innerhalb eines halben Arbeitstags Rückmeldung. Sehr dringende Angelegenheiten bitte mit einem roten Aufkleber versehen.«

Die neu eingetroffenen Akten werden dann zweimal täglich kurz gesichtet und entsprechend der jeweiligen Dringlichkeit und den vorhandenen Kapazitäten weiterbehandelt. Dies bezeichnet man als *verzögerte Zusage*, die im Kasten unten beschrieben wird.

Verzögerte Zusage

Die verzögerte Zusage (*Deferred Commitment*) bezeichnet die Trennung der Anforderung für eine Arbeitsleistung von der Zusage, eine Arbeit konkret zu verrichten. Dies bedeutet, dass angefragte Leistungen nicht direkt in Arbeit genommen werden.

Die verzögerte Zusage stellt eine wichtige Grundlage für Pull-Systeme dar.

Im nächsten Kasten finden Sie die Definition eines Pull-Systems, das dieses Prinzip anwendet.

Das Pull-System

Ein Pull-System ist ein System für die Steuerung von Arbeit. Darin wird Arbeit nur begonnen, wenn sowohl Nachfrage besteht als auch Kapazität für eine Lieferung verfügbar ist.

Kanban-Systeme verwenden WIP-Limitierungen, um die verfügbare Kapazität darzustellen und die Notwendigkeit für das Ziehen von Elementen zu signalisieren, sobald Kapazitäten zur Verfügung stehen. Diese Signalsetzung erfolgt mittels virtueller Pull-Signale.

Ein Pull-Signal entsteht, wenn das WIP-Limit nicht voll ausgeschöpft ist. Das heißt, es befinden sich in diesem Teil des Systems weniger Karten als möglich. Bewegt sich Arbeit auf dem Board nach rechts, wandern Pull-Signale nach links, also upstream, und signalisieren Kapazität, um Arbeit nach rechts weiterzuziehen.

Die Limitierung der Arbeitspakete im System erzeugt gemeinsam mit dem durch Vereinbarungen gestützten Pull-Prinzip einen kontinuierlichen Arbeitsfluss im System.

Oder wie ein Kursteilnehmer es ausdrückte: »Super, Kanban ist ja wie Rohr-frei!«

Eine weitverbreitete Fehlinterpretation ist folgende: »Pull« heißt, jede(r) nimmt sich die Karten, die ihm oder ihr am besten gefallen – also Rosinenpickerei. Diese Bedenken habe ich häufiger gehört. Dank der durch das Board erzeugten Sichtbarkeit und dem damit einhergehenden gewissen sozialen Druck ist dies in der Praxis jedoch üblicherweise kein Problem. Im Gegenteil – durch gemeinsam vereinbarte *Pull-Kriterien* wird hier konsistentes Verhalten gefordert und gefördert. Diese werden im Abschnitt »Pull-Kriterien« auf Seite 134 näher beschrieben.

An dieser Stelle spielt auch das in der Organisation mehr oder weniger stark ausgeprägte Vertrauen eine Rolle: Vertrauen Führungskräfte darauf, dass ihre Mitarbeitenden motiviert sind, das Beste zu leisten? Kanban hilft dabei, in eine positive Spirale des Vertrauensaufbaus zu gelangen: Zunächst wird Transparenz geschaffen, dann kontinuierlich (eventuell auch kleinere Aufgaben) und zuverlässig geliefert.

Leider werden in vielen Kanban-Systemen keine WIP-Limits eingesetzt. Damit kann die Kapazität nicht zuverlässig begrenzt und das Pull-Prinzip nicht umgesetzt werden. Das ist schade, denn zunächst einmal ist ein Pull-System die Basis für ein stabiles System und das Erreichen einer guten Lieferfähigkeit und Vorhersagbarkeit. Damit ist es auch die Grundlage für verschiedene wirklich coole weiterführende Techniken wie Kapazitätszuweisung, effiziente Regelung des Nachschubs und natürlich Prognosen.

Ich hoffe, dieses Buch hilft ein wenig, das zu ändern. In den nächsten Abschnitten finden Sie daher sehr praktische Überlegungen und Erfahrungswerte, um Sie auf den Weg zum »Pull« mit Ihrem System zu unterstützen.

Die Wurzeln von Pull-Systemen

Um zu verstehen, warum Pull-Systeme in einer bestimmten Art und Weise modelliert werden, hilft eine kurze Rückschau auf die Ursprünge der Methode. Das Pull-Prinzip beruht im Kern auf dem Pull in der *Lean Production* (schlanken Produktion).

Auf einer Produktionsstraße (wie zum Beispiel bei der Automontage) gibt es drei aufeinanderfolgende Arbeitsstationen A, B und C. Zwischen diesen wird Inventar gelagert, nennen wir sie »Puffer«. Das ist in Abbildung 4-8 dargestellt.

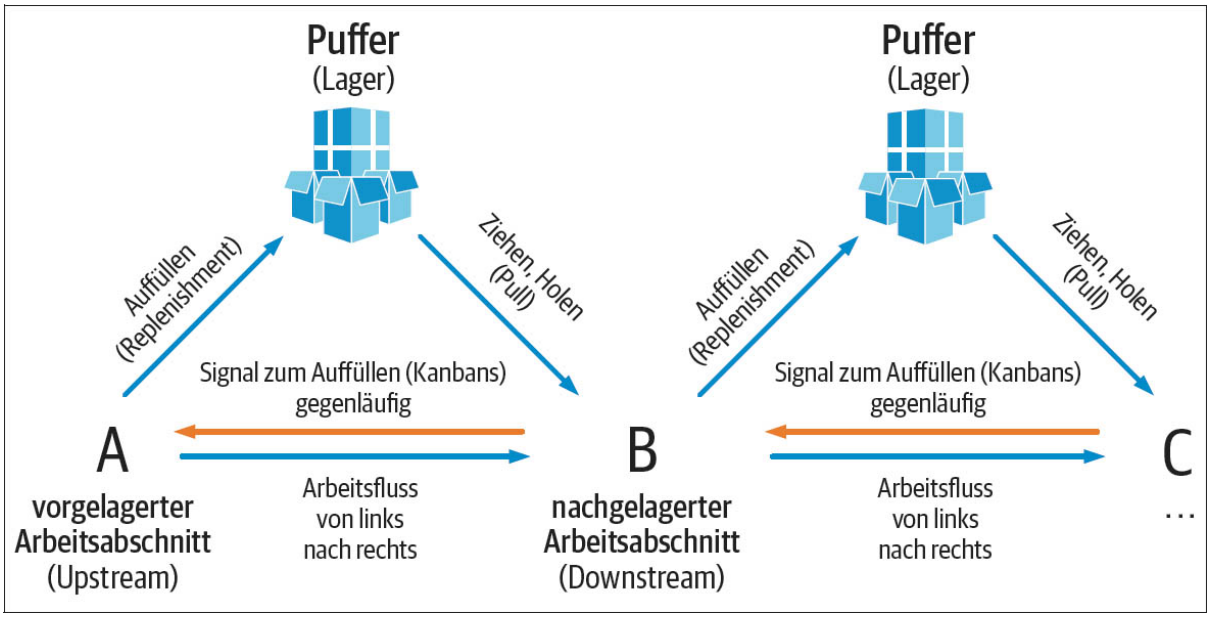


Abbildung 4-8: Nachschub und Pull-Signale in der Produktion

Damit B arbeiten kann, bedient sich B der Arbeitsgegenstände (*Work in Progress*) aus dem vorgelagerten Puffer. Falls A nicht ausreichend schnell für Nachschub sorgen kann, wird der Puffer leerlaufen, und B kann nicht weiterarbeiten, der Fluss wird stocken. Wenn andererseits B langsamer arbeitet als A, wird sich die angefangene Arbeit stauen.

Für einen optimalen Fluss sind daher klare Signale für optimale Pufferfüllstände (nicht zu viel und nicht zu wenig) erforderlich. In der Produktion werden dabei physische *kanban*-Karten als Pull-Signale benutzt. Hat eine Prozessstufe wieder Kapazitäten, werden Pull-Signale vom Prozessausgang flussaufwärts weitergeleitet, um benötigten Nachschub zu signalisieren, bis hin zum Start der Produktionsstrecke.

Hierzu ein einfaches Beispiel, das Ihnen eventuell sogar im Alltag weiterhelfen kann: Stellen Sie sich vor, Sie hätten ein Regalfach, in dem Sie Ihr Toilettenpapier lagern. Sie wohnen fußläufig zu einem Supermarkt, der Nachschub ist also mit Ausnahme von Sonntagen problemlos möglich. Sie kleben daher an die dritte Rolle von hinten eine Haftnotiz mit der Aufschrift »Klopapier kaufen!!«. Voilà, Ihr physisches Pull-Signal!

In unseren wissensbasierten Systemen nutzen wir virtuelle Pull-Signale. Der Effekt ist der gleiche: Wenn eine bestimmte Prozessstufe ausgelastet ist, werden vorgelagerte (Upstream-)Schritte gedrosselt. Auch hier kann in einem Extremfall der gesamte Prozess zum Stillstand kommen, wenn die Arbeit an einer Stelle stockt.

Im nächsten Abschnitt erfahren Sie, wie Pull-Systeme in der Wissensarbeit gestaltet werden.

Modellierung von Pull-Systemen in der Wissensarbeit

Mit diesem Hintergrund schauen wir uns nun an, wie das Prinzip auf die Wissensarbeit angewendet wird.

Das bereits im Rahmen der Praktik *Visualisiere* vorgestellte Beispiel stellt schon eine hervorragende Basis dar, denn hier wechseln sich Aktivitäts- und Pufferspalten ab. Wir müssen nur noch die Kapazitäten begrenzen. Dazu benutzen wir WIP-Limits, die oft als Zahlen in Kreisen über der Spalte bzw. den Spalten dargestellt werden. Dies sehen Sie in Abbildung 4-9.

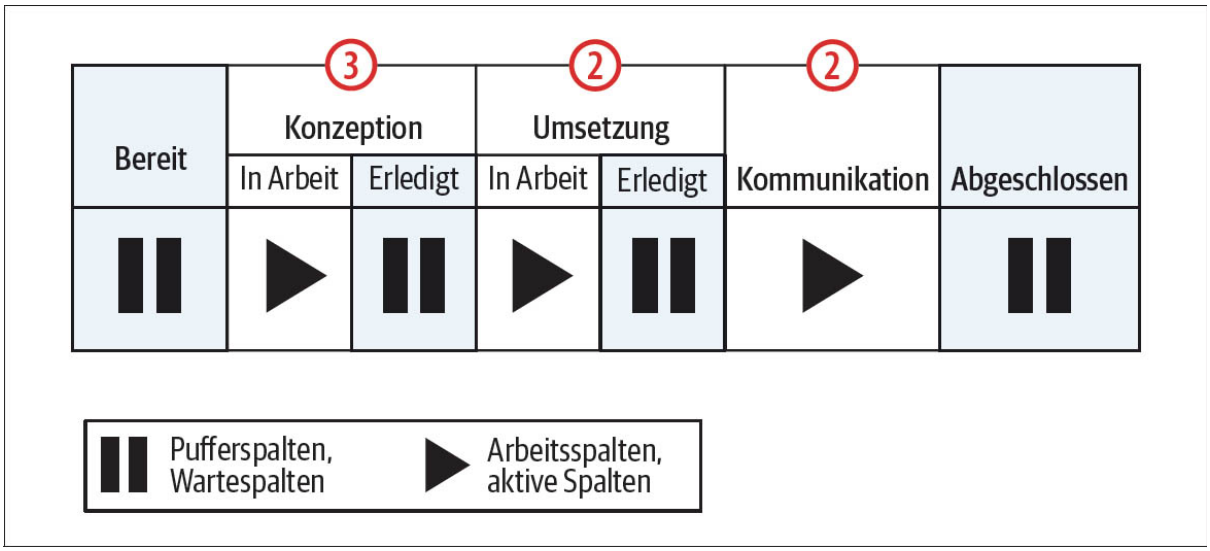


Abbildung 4-9: Pull-System mit Pufferspalten und WIP-Limits

Beachten Sie, dass analog zum Beispiel aus der Produktion die Limits typischerweise beide Spalten, die Aktivitätsspalte (*In Arbeit*) und den direkt nachgelagerten Puffer (*Erledigt*) überspannen. Pull-Signale ergeben sich aus der Differenz zwischen dem WIP-Limit und der im entsprechenden Bereich befindlichen Karten.

Nachfolgend einige beispielhafte Situationen, die die Wirkungsweise verdeutlichen sollten.

In Beispiel 1 (Abbildung 4-10) gibt es zwei freie »Plätze« auf dem Board. Die beiden Pull-Signale werden durch die gestrichelten Boxen angedeutet. Hier gibt es also diese Möglichkeiten:

- Karte A fertigstellen (Achtung, es muss immer möglich sein, fertige Karten in einen Puffer oder eine nicht limitierte Spalte zu bewegen)
- Pull-Signal: Karte B oder D in die Spalte *Kommunikation* ziehen
- Pull-Signal: die Konzeption für Karte F oder G beginnen

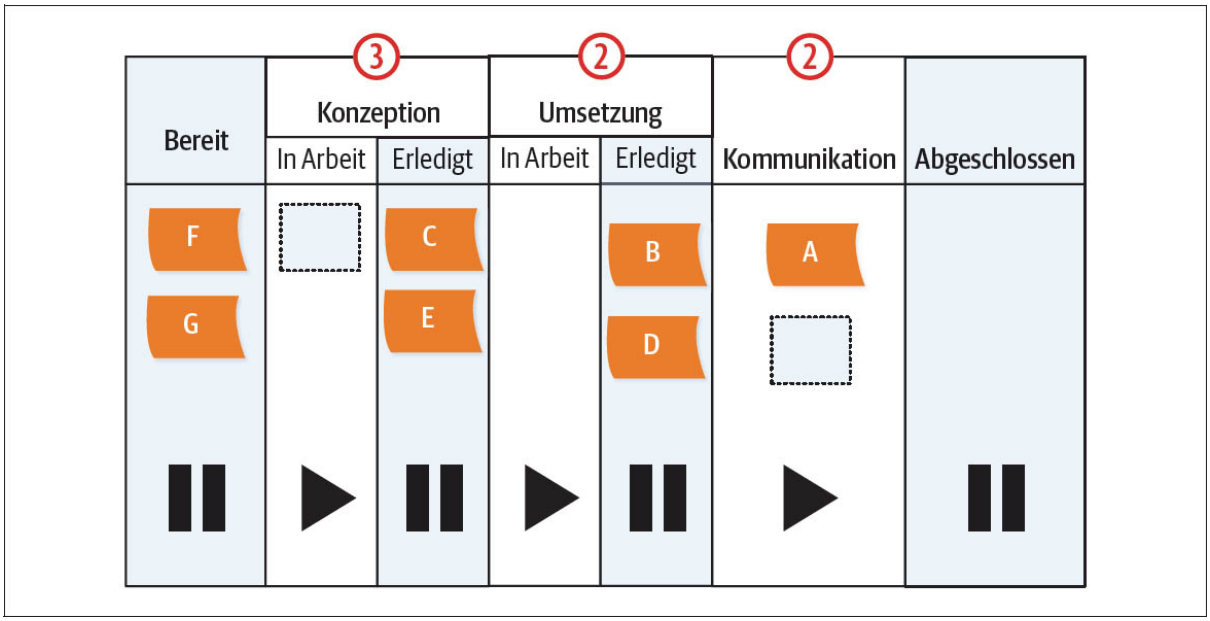


Abbildung 4-10: Beispiel 1: Zwei Pull-Signale

In Beispiel 2 (Abbildung 4-11) wurde Karte F in die Konzeption gezogen und ist erledigt. Für die Aktivitäten *Konzeption* und *Umsetzung* ist keine Kapazität mehr da. Obwohl alle Karten dort in *Erledigt* hängen, sind die WIP-Limits ausgeschöpft. Dies ist ein Signal dafür, dass Arbeit am Ende des Boards weitergezogen und abgeschlossen werden muss, um vorn Kapazität zu schaffen.

Hier ist nur noch möglich, Karte A in *Abgeschlossen* zu bewegen oder Karte B oder D in *Kommunikation* zu ziehen.

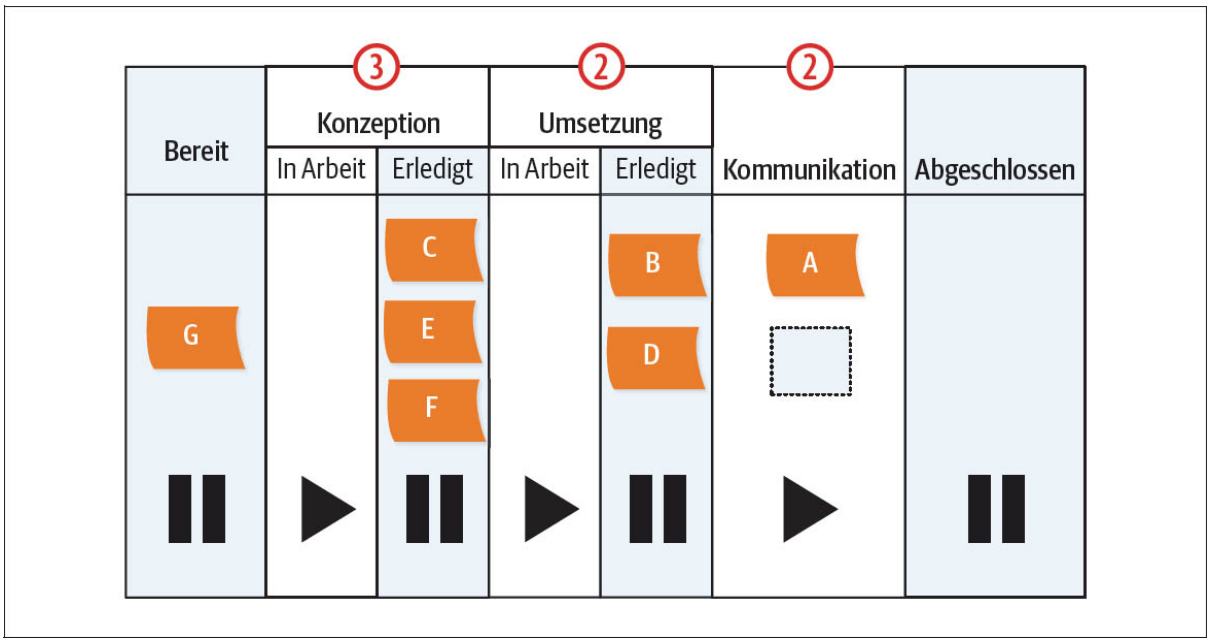


Abbildung 4-11: Beispiel 2: Ein Pull-Signal

Damit geschieht dann etwas Spannendes: Ein Pull-Signal wandert! Dies sehen Sie in Beispiel 3 (Abbildung 4-12): Karte D hat sich nach rechts in die Spalte *Kommunikation* bewegt. Dadurch wandert das Pull-Signal nach links und ermöglicht es nun, Karte C, E oder F in die Umsetzung zu ziehen.

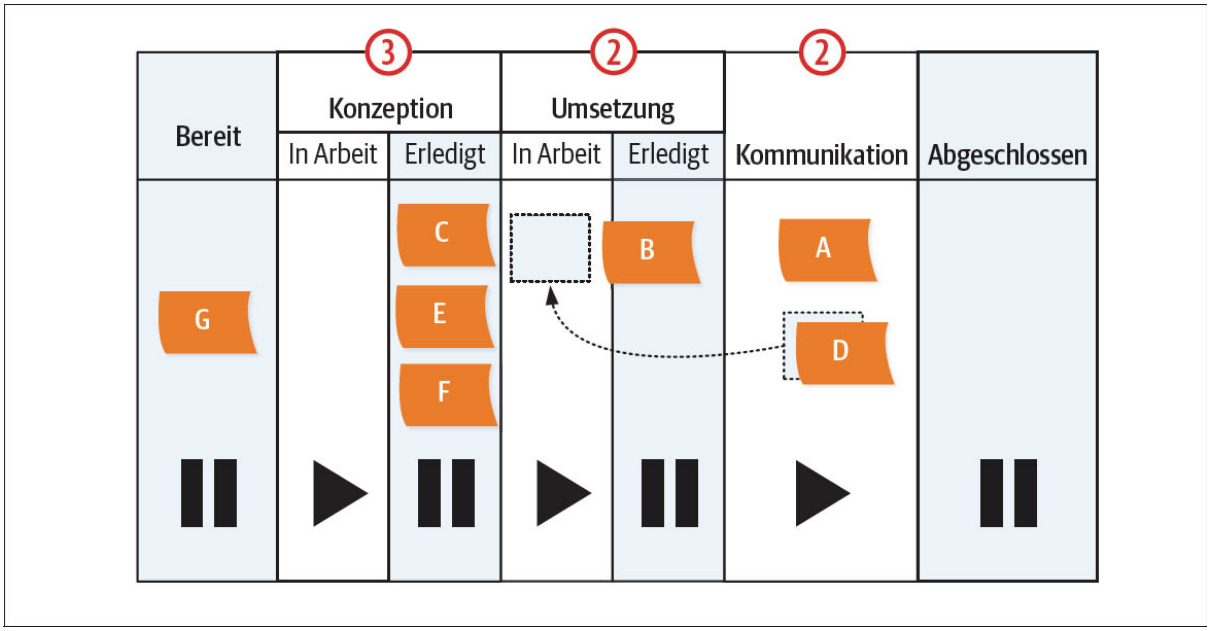


Abbildung 4-12: Beispiel 3: Das Pull-Signal wandert nach links.

Spulen wir ein wenig vor, könnte sich eine Situation wie die im Beispiel 4 (Abbildung 4-13) ergeben. Es hat sich Arbeit bewegt, und am Eingang des Pull-Systems (hier: *Konzeption* als erste WIP-limitierte Spalte) ist Platz entstanden, um neue Arbeit zu starten!

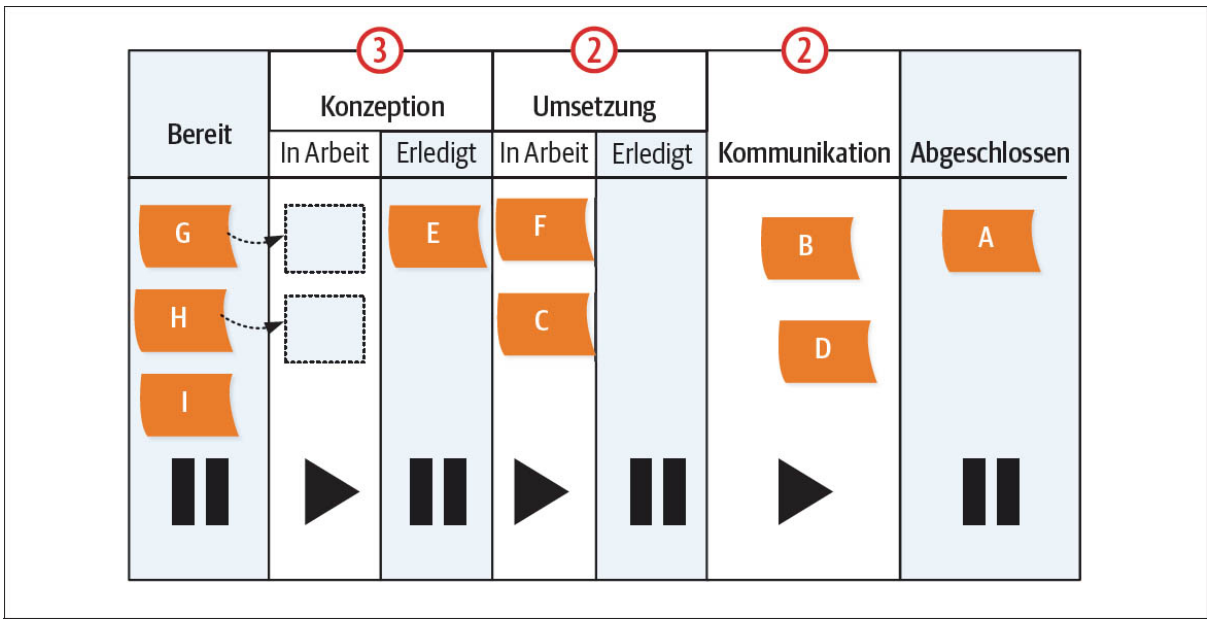


Abbildung 4-13: Beispiel 4: Platz am Systemeingang für neue Arbeit

Ausgezeichnet, in der Zwischenzeit gab es Nachschub in der *Bereit*-Spalte. Es stehen nun drei Karten zur Auswahl, um ins System gezogen zu werden. Für G und H beginnt die Konzeption.

Anhand dieser vier Beispiele konnten Sie die Funktionsweise von virtuellen Pull-Signalen erfahren, wie sie in Kanban-Systemen zur Wissensarbeit eingesetzt werden. Mehr zu den Möglichkeiten der praktischen Umsetzung von WIP-Limits erfahren Sie weiter unten in diesem Kapitel.

Das Pull-System in seiner Umgebung

In der Praxis trifft dieses Pull-System mit all seinen Vorzügen natürlich an seiner Systemgrenze auf die eher von *Push* geprägte Umgebung. Hier hilft die Umsetzung der *verzögerten Zusage*. Dafür wird als Puffer ein Arbeitsvorrat zwischengeschaltet. Für diesen gibt es auch andere Bezeichnungen, wie z.B. Backlog oder Ideenpool. Dieser Arbeitsvorrat fungiert als unbegrenzter »Posteingang«, in dem alle Anfragen der Kunden zunächst landen.

Danach greifen vereinbarte Regelungen und eine Eingangspufferspalte auf dem Board, um Arbeit gesteuert (und verzögert, aber wenn nötig sehr schnell) in das System zu ziehen. Die entsprechenden Praktiken werden in Abschnitt »Der Systemeingang« auf Seite 111 ausführlich beschrieben.

Ein Pull-System implementieren

Pull-Systeme sind ein schönes Beispiel dafür, wie verschiedene Kanban-Praktiken einander bedingen und gemeinsam wirken:

- Praktik *Visualisiere*: Das Board-Design muss durch Nutzung von Puffern so gestaltet sein, dass Pull technisch möglich ist (siehe Abschnitt »Modellierung von Pull-Systemen in der Wissensarbeit« auf Seite 96).
- Praktik *Limitiere die parallele Arbeit (das WIP)*: Kein Pull-System ohne WIP-Limits! Diese können auf Ebene von (Gruppen von) Spalten oder auch für das Gesamtsystem vereinbart werden.
- Praktik *Mache Regeln explizit*: Für Pull-Systeme braucht es eine Definition der Handhabung von WIP-Limits, die Vereinbarung von Pull-Kriterien sowie die Vereinbarung der Kriterien für »umsetzungsbereit«[4].
- Praktik *Verbessere gemeinsam, entwickle experimentell weiter*: WIP-Limits und sonstige Vereinbarungen (z.B. zu Pull-Kriterien) sollten ständig überprüft und angepasst werden. Hierzu bekommen Sie viele praktische Hinweise im folgenden Abschnitt.

Der effektive Einsatz von WIP-Limits ist für Pull-Systeme von zentraler Bedeutung. Für die Arbeit mit WIP-Limits geben die nächsten Abschnitte Tipps und Hinweise.

WIP-Limits in der Praxis einsetzen

Sie haben nun erfahren, wie wichtig die Praktik *Limitiere die parallele Arbeit (das WIP)* ist. Sie erinnern sich, dass die Limitierung angefangener Arbeit der Schlüssel zur fokussierten Arbeit mit weniger Umschaltverlusten ist. Daher sind die WIP-Limits ein wichtiges Element zum Steuern des Arbeitsflusses hin zur zuverlässigen, vorhersagbaren Lieferung an die Kunden. Dazu erfahren Sie mehr in Kapitel 5, *Manage den Arbeitsfluss*.

Trotzdem wird die Praktik der Begrenzung angefangener Arbeit oftmals nicht oder nur halbherzig als »Feigenblatt« angewendet. Und das aus gutem Grund: Durch die Begrenzung des Systems auf die tatsächlich vorhandenen Kapazitäten wird Stress erzeugt. Überlastung wird gnadenlos sichtbar gemacht, und Entscheidungen und auch Veränderungen sowie eine aktive Arbeit am System werden erzwungen.

Ich hoffe, die nachfolgenden Abschnitte helfen Ihnen, diese Praktik erfolgreich einzusetzen und bald die wohlverdienten Früchte zu ernten. Und vergessen Sie nicht: Alle Probleme, die möglicherweise in diesem Zuge hochkommen, gab es bereits vor Ihrer Kanban-Initiative. Kanban erzeugt sie nicht, sondern macht sie sichtbar und erzeugt Motivation für Veränderungen.

Das richtige WIP-Limit finden

Ehe wir uns der konkreten Dimensionierung der WIP-Limits widmen, hier nochmals eine kurze Übersicht über Arten von WIP-Limits.

Arten von WIP-Limits und deren Wirkung

Die wohl bekannteste und »klassische« Form ist die aus der Produktion abgeleitete Limitierung über Spalten (Sie erinnern sich, die Puffer zwischen den Stationen). Der nachfolgende Kasten gibt einen Überblick über gebräuchliche Arten von WIP-Limits.

Arten von WIP-Limits

WIP-Limits können auf verschiedenen Ebenen definiert werden:

- Limitierung je Person, z.B. »maximal drei Karten pro Person«.
- Limitierung in einer oder in mehreren Spalten. Dies können Puffer, Einzelschritte oder mehrere Aktivitäten im Prozess sein, z.B. »maximal drei Karten im Test«.
- Limitierung je Kategorie von Arbeit (z.B. in Zeilen oder über Kartenfarben unterschieden), z.B. »maximal drei Neuentwicklungen und maximal fünf Fehlerbehebungen gleichzeitig«.
- Limitierung des gesamten Kanban-Boards: »Wir wollen insgesamt maximal 15 Projekte in Arbeit haben.«

In einem Kanban-System können auch mehrere Arten von WIP-Limits eingesetzt werden und sich ergänzen.

Bei der Wahl der WIP-Limitierung spielen der Reifegrad der Organisation sowie die erhoffte Wirkung eine wesentliche Rolle.

Die eingesetzte Art der WIP-Limitierung wird verschiedene Auswirkungen haben: Limitieren Sie die Anzahl der Karten pro Person, werden Einzelne weniger überlastet sein, bekommen dadurch aber kaum Anreize zur Zusammenarbeit. Dennoch kann die Begrenzung der maximalen Anzahl der Karten pro Person ein wichtiger erster Schritt sein und schafft eine ruhigere Arbeitsumgebung, in der dann die Kraft und der Fokus entstehen können, um weitere Veränderungen anzugehen.

Etablieren Sie die Begrenzung auf Basis der Arbeitsschritte oder für das gesamte Board, werden die Beteiligten eher miteinander in Kommunikation treten. Die Vereinbarung von Begrenzungen angefangener Arbeit erschwert es, bei Auftreten von Hindernissen einfach neue Arbeit zu beginnen. Es entsteht schneller der Impuls, auf andere zuzugehen und Unterstützung für das Erledigen von Aufgaben oder die Mitarbeit an Arbeitspaketen anzubieten, weil es durch das WIP-Limit nicht möglich ist, einfach neue Arbeit zu beginnen.

Auch hier gilt: Haben Sie den Mut, ein wenig zu experimentieren, beobachten Sie die Wirkung und ändern oder ergänzen Sie gegebenenfalls die Art der angewendeten Limitierung.

Nun gilt es noch, eine wirksame Größe für die Limitierung zu finden. Gleich vorab: Bitte erwarten Sie keine einfache Formel!

Dimensionierung von WIP-Limits

In diesem Bereich gibt es zahlreiche und hartnäckige Legenden. Sollten Sie jemals den Ratschlag »Anzahl der Personen mal 2 + 1« hören, ignorieren Sie ihn höflich und suchen sich einen besseren Ratgeber. Natürlich kann das in

manchen Fällen das optimale Limit sein, in vielen Fällen würde eine solche Limitierung jedoch die Zusammenarbeit in einem Team effektiv unterbinden. Nehmen wir als Beispiel ein Softwareentwicklungsteam, in dem sich alle Entwickler gegenseitig unterstützen könnten. Wenn 6 Entwickler im Team sind und das WIP-Limit nach oben stehender Formel somit 13 wäre, käme sehr viel parallele Arbeit zustande, und die WIP-Limits gäben kaum Veranlassung dazu, zusammenzuarbeiten.

Bedenken Sie dies: WIP-Limits sind ein »Stressor«, der gewisse Verhaltensweisen erzwingt. Das ist das Wunderbare: Über vermeintlich einfache Regelungen erreichen Sie tatsächliche Verhaltensänderungen. Wie oft haben Sie schon »Wir sollten mal mehr zusammenarbeiten!« gehört oder selbst proklamiert? Mit WIP-Limits werden Sie wertvolle Impulse erhalten, dies tatsächlich zu tun!

Ein ausgeschöpftes WIP-Limit gibt das Signal, dass die Kapazität erreicht ist und damit keine neue Arbeit in den entsprechenden Bereich gezogen werden kann. Beispielsweise kann sich eine Person also keine weitere Karte ziehen, oder es kann keine neue Karte auf das Board kommen. Damit wird ein Impuls erzeugt, anders aktiv zu werden. Dies kann bedeuten, in einem anderen Bereich zu arbeiten, in dem es noch Kapazität gibt, oder auch andere zu unterstützen: »Gibt es eine Karte auf dem Board, an der ich mitarbeiten kann?« Entsprechende Eignung für die Arbeit und vorhandene Fähigkeiten natürlich vorausgesetzt. Im »schlimmsten« Fall gibt es freie Zeit, die Wissensarbeiter in der Regel gewinnbringend für Prozessverbesserungen, Reflexion, Weiterbildung etc. einsetzen können.

Hier nun die schlechte Nachricht, auf die ich Sie schon vorzubereiten versucht habe: Es gibt keine Formel zum Errechnen der »perfekten« WIP-Limits. Kanban ist eine empirische Methode und unterstützt Sie durch einen wissenschaftlichen Ansatz dabei, nach und nach die Anzahl der Arbeitspakete auf ein optimales Niveau zu bringen.

Sobald Sie eine Visualisierung etabliert haben, werden Sie schon einen ersten Eindruck davon bekommen, an welchen Stellen sich Arbeit staut. Diese rein visuelle Betrachtung wird Ihnen wertvolle Anhaltspunkte liefern.

Natürlich möchte ich Sie nun so nicht im Regen stehen lassen. Es gibt doch einige Überlegungen und Faustregeln, die Ihnen im ersten Schritt helfen sollten. Folgende Fragen können Sie beim Finden des optimalen WIP-Levels unterstützen: Wie viele Mitarbeitende bzw. Teams sind involviert? Wie hoch ist der Spezialisierungsgrad, das heißt, wie stark können die Beteiligten einander unterstützen? Wie stark bestehen externe Abhängigkeiten zum Beispiel von Zulieferungen aus anderen Bereichen?

Hier ein paar Daumenregeln: Je höher der Spezialisierungsgrad, desto höher werden Sie Ihren WIP-Level wählen müssen, damit auch bei stockendem Arbeitsfluss ausreichend Arbeitspakete zur Auswahl stehen. Ansonsten könnte dies die Auslastung zu negativ beeinflussen (längerfristiger Leerlauf).

Je stärker bereits eine Kultur der Zusammenarbeit etabliert ist und die Art der Arbeit dies zulässt, desto kleiner können Sie Ihre WIP-Limits wählen. Bedenken Sie, dass die Begrenzung der Aufgaben bei Menschen zunächst auch Ängste bzw. Druck hervorrufen kann. Daher kann es besser sein, zunächst mit einem etwas höheren WIP-Limit zu starten und gleichzeitig zu vereinbaren, dieses beispielsweise in zwei oder vier Wochen gemeinsam zu überprüfen, um dann zu entscheiden, ob es optimal war oder weiter gesenkt werden könnte.

Steuern Sie hier von Anfang an dem Schlendrian entgegen. Die WIP-Limits sind gemeinsam vereinbart, und alle Beteiligten sollen gemeinsam dafür Verantwortung tragen. Wenn das WIP-Limit ausnahmsweise gebrochen wird, sollte das besprochen werden. Kommt dies häufiger vor, ist das ein Signal für eine nötige Überprüfung und gegebenenfalls Anpassung. Verhindern Sie es, dass die WIP-Limits stillschweigend ignoriert werden. Wenn sich die Leute erst daran gewöhnt haben, dass das WIP-Limit (diese »Zahl über der Spalte«) ja gar nicht so wichtig ist, werden Sie ein wichtiges Instrument zur Steuerung frühzeitig verschleißen.

Und zu guter Letzt der Hinweis: Nach meiner Erfahrung liegt ein optimales WIP-Limit meistens deutlich unter dem, was die Beteiligten intuitiv annehmen. Typischerweise ist etwa ein Drittel davon ausreichend. Nutzen Sie geeignete Simulationen wie beispielsweise Okaloa Flowlab (siehe Abschnitt »Okaloa Flowlab« auf Seite 214), um in einem geschützten Rahmen neue Erfahrungen zu ermöglichen. Ihre Intuition muss hier erst trainiert werden.

Der Umgang mit WIP-Limits

So wichtig eine konsequente Beachtung der WIP-Limits ist: Eine Kultur des roboterhaften, unreflektierten Einhaltens der Begrenzungen gilt es zu vermeiden. Bedenken Sie, dass der Zweck der Limitierung ist, wichtige und notwendige Diskussionen zu führen. Ist eine Ausnahmesituation eingetreten, zum Beispiel durch gleichzeitige Krankheit mehrerer Personen oder durch eine ungewöhnlich hohe Anzahl blockierter oder ungewöhnlich langwieriger Aufgaben, sollten die Beteiligten dies besprechen und gegebenenfalls vereinbaren, das festgelegte WIP-Limit für einen begrenzten Zeitraum zu überschreiten.

Machen Sie es sich am besten zur Routine, die vereinbarten WIP-Limits in Ihrem System gemeinsam zu betrachten und zu hinterfragen, ob sie weiterhin optimal sind. Wenn sich beispielsweise der Mix Ihrer Arbeit ändert, es stärkere Änderungen bei den Beteiligten gibt oder aber sich Arbeitsweisen stark geändert haben, kann dies Änderungen an den WIP-Limits zur Folge haben.

»Dürfen wir bis zum Limit voll machen?« Interessanterweise bekomme ich diese Frage oft gestellt. Die Antwort ist: »Ja, bitte!« Es ist nicht sinnvoll, hier Reserven zu lassen. Dies erzeugt zusätzliche Überlegungen und Aufwände, die zu nichts führen. Also: Wenn eine 3 über einer Spalte steht, können Sie auch drei Karten hineinziehen. Sollten Sie alle das Gefühl haben, dass das zu viel wäre, senken Sie doch einfach das WIP-Limit! Haben Sie Mut, aktiv an und mit Ihrem System zu arbeiten.

Die Idee ist ja genau, den Fluss insgesamt durch die entwickelten Mechanismen (z.B. die WIP-Limits) und Vereinbarungen sehr effizient zu steuern, statt aufwendige Diskussionen zu den einzelnen Arbeitsaufträgen (also Einzelfallbehandlungen) zu führen. Sie werden staunen, wie viel Zeit und Energie hier freigesetzt werden können.

An dieser Stelle nähern wir uns automatisch der Kanban-Praktik *Manage den Arbeitsfluss*, die Thema des nächsten Kapitels ist. Beide Praktiken sind eng miteinander verknüpft.

Index

E

F

G

H

I

K

L

N

O

P

R

S

T

U

V

W

Z