Stojąc na ramionach olbrzymów

Idee organizacji produkcji a ich praktyczne realizacje

Przykład firmy Hitachi Tool Engineering

© Eliyahu M. Goldratt, 2008

Tłumaczył: Wojciech Śledziński

Wprowadzenie

Nietrudno wywieść popularność produkcji odchudzonej (Lean production) z sukcesu Toyoty. Sukces Toyoty jest niezaprzeczalny. Toyota obecnie wytwarza tyle samo samochodów, co tradycyjny lider – GM (General Motors), a przy tym osiąga zyski. W ciągu ostatnich pięciu lat zysk netto Toyoty był o 70% wyższy od średniej w branży, podczas gdy GM ponosi straty. Ten sukces firmy można całkowicie przypisać systemowi produkcyjnemu Toyoty (TPS)[11]. Takie jest przynajmniej przekonanie kierownictwa firmy[22] – deklarowanym wyzwaniem numer jeden Toyoty jest przekazanie TPS-u jako kodu genetycznego firmy następnemu pokoleniu.

Biorąc pod uwagę fakt, że Toyota jest okrętem flagowym japońskiego przemysłu, powinniśmy spodziewać się szerokiego stosowania produkcji odchudzonej w Japonii. O dziwo, wcale tak nie jest. Jest rzeczą powszechnie wiadomą w Japonii, że mniej niż 20% firm produkcyjnych wprowadziło w życie system produkcji odchudzonej. Jak to możliwe?

Jest tak nie dlatego, że nie próbowano zastosować tego systemu. Wiele firm w Japonii włożyło mnóstwo wysiłku w próby wprowadzenia produkcji odchudzonej, ale skończyły się one porażką. Jedną z takich firm jest Hitachi Tool Engineering. Niemożności wprowadzenia u nich systemu produkcji odchudzonej nie da się wyjaśnić zaniechaniem poważnych wysiłków w tym kierunku. Firma ta wielokrotnie próbowała wprowadzić produkcję odchudzoną, lecz pogarszające się wyniki produkcyjne zmuszały ją do powrotu do bardziej tradycyjnych sposobów zarządzania produkcją.

Faktu, że większość przemysłu japońskiego nie zastosowała produkcji odchudzonej, nie można też przypisywać brakowi wystarczającej wiedzy. Toyota nad miarę szczodrze dzieliła się swoją wiedzą. Firma ta udostępniła szerokiej publiczności całość wiedzy o swoim systemie produkcyjnym (TPS), posuwając się nawet do zapraszania swych bezpośrednich konkurentów do odwiedzania swoich fabryk. Firma Hitachi, podobnie jak tyle innych, korzystała z dostępnej wiedzy i nie wstydziła się zatrudniać do pomocy najlepszych dostępnych ekspertów.

Istnieje wyjaśnienie niepowodzenia owych firm we wprowadzaniu odchudzania, wyjaśnienie, które jest oczywiste dla każdego obiektywnego obserwatora firmy takiej jak Hitachi Tool Engineering. Owo niepowodzenie wynika z podstawowej różnicy środowisk produkcyjnych. Kiedy Taiichi Ohno rozwijał TPS, nie czynił tego abstrakcyjnie. Tworzył ten system dla swojej firmy. Nic zatem dziwnego, że ów potężny system praktyczny, jaki Ohno stworzył, może nie działać w zasadniczo odmiennych środowiskach wytwórczych. Nie oznacza to jednak, że praca Ohno nie może być nader cenna dla innych środowisk. Geniusz Ohno objawi się nam w całej pełni, gdy uświadomimy sobie, że i on sam wcześniej znalazł się w dokładnie takiej samej sytuacji. W owym czasie systemem wytwórczym, który zrewolucjonizował produkcję, była metoda linii potokowej, opracowana przez Henry Forda. Metoda Forda była już stosowana nie tylko w nieomal całym montażu pojazdów, lecz również w wielu bardzo różnych branżach, jak wytwarzanie napojów i amunicji. W owym czasie panowała już też zgoda co do tego, że linie potokowe można i należy wprowadzać wyłącznie w środowiskach, w których wymagana wielkość produkcji uzasadnia przeznaczenie wyposażenia do wytwarzania jednego wyrobu. Gdy tylko wytwarzane ilości nie były dostatecznie duże, nikt nie rozważał możliwości zastosowania linii – nikt oprócz Ohno.

Ohno pojął, że idee, leżące u podstaw systemu Forda są ogólne. Ich realizacja praktyczna jest ograniczona do określonego typu środowisk, niemniej same idee są uniwersalne. Ohno miał jasną

---

[1] http://moneycentral.msn.com/investor/invsub/results/compare.asp?Page=ProfitMargins&Symbol=TM

[2] System produkcyjny Toyoty (TPS) stał się powszechnie znany w świecie najpierw pod nazwą produkcji "dokładnie na czas" (Just-In-Time, czyli JIT), a później - produkcji odchudzonej (Lean). Sama Toyota twierdzi, że produkcja odchudzona nie oddaje w pełni ducha TPS z powodu zniekształceń w trakcie informowania o tym systemie i wprowadzania go do praktyki.

wizję idei jako punktu wyjścia, siłę geniuszu, aby zaprojektować realizację praktyczną odpowiednią dla środowiska Toyoty, w którym nie jest możliwe przeznaczenie wyposażenia do produkcji jakiejś jednej części składowej, oraz nieustępliwość w pokonywaniu olbrzymich przeszkód na drodze wprowadzania takiego systemu organizacyjnego do praktyki. Wynikiem tego wszystkiego jest TPS.

Zamiast więc powstrzymywać się od stosowania poprawnych idei lub, co gorsza, próbować na siłę ich praktycznej realizacji w środowiskach w sposób oczywisty zbyt odmiennych, powinniśmy raczej pójść śladami Ohno.

W tym artykule przedstawimy:

- podstawowe idee łańcuchów dostaw – idee, na których opiera się system produkcji odchudzonej;
- ogólne zastosowanie tych idei, z którego można skorzystać w znacznie szerszym zakresie środowisk wytwórczych, oraz
- imponujące wyniki, jakie osiągnęła firma Hitachi Tool Engineering, dzięki owej szerszej realizacji praktycznej.

Perspektywa historyczna

Przemysł wytwórczy został ukształtowany przez dwóch wielkich myślicieli, Henry Forda i Taiichi Ohno. Ford zrewolucjonizował produkcję masową przez wprowadzenie linii potokowych. Ohno przeniósł idee Forda na kolejny poziom w postaci TPSu, systemu, który zmusił cały przemysł do zmiany swego pojmowania zapasów, nie jako elementu majątku, lecz składnika pasywów.

Punktem wyjścia dla Forda było uznanie, że kluczem do efektywnej produkcji jest skupienie się na poprawie całkowitego przepływu wyrobów przez ciąg operacji. Jego wysiłki w celu poprawy przepływu były tak udane, że do roku 1926 całkowity czas realizacji od wydobycia rudy do przekazania kompletnego, składającego się z ponad 5000 części, samochodu na platformę kolejową gotową do wysyłki, wynosił 81 godzin![3] Osiemdziesiąt lat później, żaden producent samochodów na świecie nie jest w stanie osiągnąć tak krótkiego czasu realizacji, czy choćby zbliżyć się do takiego wyniku.

Przepływ oznacza, że zapasy w trakcie operacji przesuwają się. Gdy zapasy się nie przemieszczają, to się akumulują. Gromadzące się zapasy zapasów zajmują określone miejsce (fizyczną przestrzeń). A zatem, intuicyjnym sposobem na osiągniecie lepszego przepływu jest ograniczenie miejsca, w którym zapas ma prawo się gromadzić. Aby uzyskać lepszy przepływ, Ford ograniczył przestrzeń przeznaczoną na produkcję w toku między każdą parą gniazd wytwórczych. Oto istota linii potokowych, co może potwierdzić fakt, że pierwsze linie potokowe nie posiadały żadnych mechanizmów, takich jak transportery, do przemieszczania zapasu z jednego gniazda wytwórczego do następnego. Śmiałość metody Forda ujawni się, gdy zdamy sobie sprawę z bezpośredniej konsekwencji ograniczenia przestrzeni, tej mianowicie, że gdy miejsce przeznaczone na zapas jest zapełnione, pracownicy, zasilający to miejsce, muszą przerwać pracę. Dlatego więc, aby osiągnąć równomierny przepływ, Ford musiał położyć kres stosowaniu lokalnych mierników efektywności. Innymi słowy, linie potokowe stanowią wyzwanie dla obiegowej wiedzy, dla poglądu, że, aby być wydajnym, każdy pracownik i każde gniazdo wytwórcze musi pracować przez 100% czasu.

Ktoś mógłby pomyśleć, że uniemożliwienie zasobom ciągłej pracy obniży przepustowość (produkcję) danej operacji. Taki niepożądany efekt mógłby mieć miejsce, gdyby Ford zadowolił się li tylko ograniczeniem przestrzeni składowej. Jednak z ograniczenia akumulacji zapasów wynika i inny efekt. Pozwala ono w sposób bardzo naoczny wykrywać prawdziwe problemy, zagrażające przepływowi – gdy jedno gniazdo wytwórcze w linii przestaje produkować i trwa to dłużej niż krótką chwilę, wkrótce staje cała linia. Ford wykorzystał te wynikową dobrą widzialność, aby lepiej wyrównać przepływ, poprzez zajęcie się owymi wyraźnymi przestojami i ich eliminację. Ostatecznym rezultatem likwidacji lokalnych mierników efektywności[4] i zrównoważenia przepływu jest istotne zwiększenie przepustowości. Henry Ford osiągnął najwyższą przepustowość na jednego pracownika

[3] Ford, Henry, *Today and Tomorrow* , Productivity Press, 1988 (originally published in 1926).

[4] Równoważenie przepływu nie jest równoznaczne z równoważeniem zdolności wytwórczych, czyli dążeniem do sytuacji, gdy zdolność wytwórcza każdego gniazda jest dostosowana do jego obciążenia. Jest to powszechnie popełniany błąd przy równoważeniu linii potokowych.

pośród wszystkich producentów samochodów w jego czasach.

Podsumowując, linie potokowe Forda opierają się na tych oto czterech ideach:

1. Poprawa przepływu, czyli, co równoważne, czasu realizacji, jest głównym celem działalności operacyjnej.
2. Ten zasadniczy cel trzeba przełożyć na praktyczny mechanizm, który będzie wskazywał operatorom, kiedy nie powinni produkować, a więc będzie zapobiegał nadmiernej produkcji.
3. Trzeba skończyć z praktyką mierzenia lokalnej efektywności.
4. Trzeba ustanowić praktykę skupiania się na równomierności przepływu.

Podobnie jak i u Forda, głównym celem Ohno była poprawa przepływu – skrócenie czasu realizacji, jak wskazuje na to jego odpowiedź na pytanie co robi Toyota:
Wszystko, co robimy, to obserwowanie odcinka czasu od chwili, gdy klient składa nam zamówienie do momentu, gdy odbieramy gotówkę. I skracamy ów odcinek czasu… [5]

Ohno, gdy zaczął stosować tę drugą ideę, stanął przed przeszkodą niemalże nie do pokonania. Kiedy popyt na pojedynczy produkt jest wysoki, wydzielenie linii do wytwarzania poszczególnych części składowych, jak to uczynił Ford, jest uzasadnione. Jednakże, w owym czasie w Japonii popyt na rynku był na małe ilości różnorodnych samochodów. Z tego względu Ohno nie mógł wydzielić wyspecjalizowanych linii w Toyocie. Jak już mówiliśmy, wszystkie inne branże, które znalazły się w takiej sytuacji, po prostu nie brały pod uwagę stosowania linii. Ohno jednak rozważał pomysł zastosowania linii, w których wyposażenie nie byłoby wyspecjalizowane, a więc takich, w których każde gniazdo wytwórcze produkowałoby różnorodne części składowe. Problem polegał na tym, że w takim przypadku zastosowanie mechanizmu ograniczonej przestrzeni doprowadziłoby do zatorów – nie wszystkie detale byłyby dostępne na stanowisku montażu, a więc montaż nie mógłby pracować, gdyby cała wydzielona przestrzeń składowa była już zapełniona, bo nie mogłyby wtedy pracować linie zasilające. Ohno pisze, że wpadł na rozwiązanie, gdy usłyszał o supermarketach (było to na długo przedtem, nim faktycznie zobaczył supermarket podczas swojej wizyty w Stanach w 1956 r.). Zrozumiał, że zarówno supermarkety, jak i linie zasilające w Toyocie muszą radzić sobie z dużą różnorodnością wyrobów. Produkty w supermarketach nie tarasują półek i regałów, lecz zamiast tego większość towarów trzyma się w magazynie na zapleczu. W samym magazynie każdy produkt ma przydzielone ograniczone miejsce na regałach. Dopiero, gdy klient weźmie towar z półki w sklepie, uruchamiana jest procedura uzupełnienia z magazynu na zapleczu, aby zapełnić przydzielone temu produktowi miejsce na półce sklepowej. Ohno zaplanował mechanizm, który umożliwiłby mu wskazywanie operatorom Toyoty, kiedy nie należy produkować. Zamiast zastosowania jednego ograniczonego miejsca między gniazdami wytwórczymi dla ograniczenia gromadzenia produkcji w toku, musiał on ograniczyć dozwoloną ilość każdego detalu. Opierając się na tym wymaganiu, Ohno zaprojektował system Kanban.

Kanban jest opisany w licznych artykułach i książkach. W tym artykule opiszemy tylko jego istotę, aby pokazać, jak słuszne były podstawowe idee Ohno. Między każdą parą gniazd i dla każdego detalu oddzielnie[66] gromadzenie się zapasu jest ograniczone przez ustawienie określonej liczby pojemników i ustalenie liczby sztuk w pojemniku. Pojemniki te, podobnie jak każdy pojemnik w jakiejkolwiek fabryce, zawierają też odpowiednie dokumenty. Taki pojedynczy dokument, zazwyczaj karta (czyli po japońsku kanban), określająca jedynie nazwę detalu i liczbę sztuk w pojemniku, jest jednak traktowany niekonwencjonalnie. Gdy kolejne gniazdo robocze pobiera pojemnik z częściami do dalszej obróbki, owa karta nie przemieszcza się się wraz z pojemnikiem, ale jest zwracana do poprzedzającego gniazda roboczego. Stanowi to sygnał dla tamtego gniazda, że pojemnik został pobrany i ustalony zapas detalu nie jest pełny. Tylko w takim przypadku poprzedzające gniazdo robocze ma prawo produkować (jeden pojemnik detali określonych na karcie). W istocie system Kanban wskazuje każdemu gniazdu roboczemu, kiedy i co produkować, ale, co ważniejsze, również kiedy nie należy produkować. Nie ma karty – nie pracujemy. System Kanban jest praktycznym mechanizmem, który wskazuje operatorom, kiedy nie należy produkować, przez co zapobiega nadmiernej produkcji. Ohno z powodzeniem rozwinął idee Forda, zmieniając podstawę tego mechanizmu z ograniczonego miejsca na ograniczony zapas.

[5] Ohno, Taiichi, *Toyota Production System* , Productivity *Press*, Inc. 1988, str. ix (w Przedmowie Wydawcy). Warto też odnotować, że zarówno w tej, jak i w innych swoich książkach, Ohno w pełni uznaje, że podstawowe idee są zasługą Forda.

[6] Dla zmniejszenia liczby miejsc przeznaczonych do składowania pojemników, Ohno szeroko stosował U-kształtne komórki produkcyjne, zamiast gniazd roboczych wyposażonych w jeden typ maszyn.

Przestrzeganie zasady równomiernego przepływu wymaga skończenia z praktyką stosowania lokalnych mierników wydajności. Ohno podejmował ten temat nieustannie w swych książkach, podkreślając, że nie ma sensu zachęcanie ludzi do produkowania, gdy określone produkty nie są potrzebne w bardzo krótkim czasie. Ten nacisk był prawdopodobnie powodem tego, że poza Toyotą TPS pierwotnie stał się znany jako system produkcji dokładnie na czas (Just-in-time production).[7]

Jak tylko system Kanban, wskazujący operatorom, kiedy nie należy produkować, wprowadzono do hal fabrycznych, natychmiastowa redukcja przerobu wymagała olbrzymiego wysiłku, aby zrównoważyć przepływ. Wyzwanie, przed jakim stanął Ohno, było o niebo większe od tego, z czym musiał się zmierzyć Ford. Aby uświadomić sobie, jak wielkie było owo wyzwanie, wystarczy zwrócić uwagę na jeden tylko z wielu aspektów tego zadania. W odróżnieniu od środowiska, w którym stosuje się linie dedykowane poszczególnym wyrobom, system Ohno zmuszał gniazdo wytwórcze do częstego przestawiania się z produkcji jednego detalu na inny. Dla większości gniazd wytwórczych każde takie przestawienie wymaga czasu na wykonanie potrzebnego przezbrojenia. Ponieważ pojemniki z założenia "zamawiały" stosunkowo niewielką liczbę detali, wyznaczone przez nie partie produkcyjne niejednokrotnie były śmiesznie małe w stosunku do wymaganego przezbrojenia. Początkowo dla wielu gniazd wytwórczych czas potrzebny na przezbrojenie był dłuższy, niż czas produkcji, co powodowało znaczący spadek przerobu. Nic dziwnego, że Ohno napotkał niesłychany sprzeciw, do tego stopnia, że, jak sam pisał, od późnych lat czterdziestych do wczesnych sześćdziesiątych ubiegłego wieku o jego systemie mówiono "ten obrzydliwy system Ohno". Ohno, tak jak i jego zwierzchnicy, przejawiał oczywiście niezwykłą determinację[8] i dalekowzroczność, aby kontynuować wprowadzanie systemu, który dla każdego, kto spoglądał nań z perspektywy lokalnej, a takie spojrzenie muszą mieć w większości ludzie pracujący bezpośrednio na produkcji, po prostu nie miał sensu.

Ohno musiał utorować nową drogę, aby pokonać przeszkodę w postaci czasów przezbrojeń. W tamtych czasach i aż do chwili, gdy TPS zdobył ogólnoświatowy rozgłos, tradycyjnym sposobem radzenia sobie z przezbrojeniami było zwiększanie rozmiarów partii produkcyjnej. Popularnym terminem, na którego temat napisano tysiące artykułów, była "optymalna wielkość partii". Ohno zignorował całą tę wiedzę,[9] bowiem ustępstwo na rzecz stosowania "ekonomicznie uzasadnionych" rozmiarów partii skazałoby na niepowodzenie dążenie do skrócenia okresów realizacji. Skłonny był zatem raczej z uporem twierdzić, że niezbędne czasy przezbrojeń nie zostały wykute na kamiennych tablicach i że sposób postępowania można modyfikować tak, by te czasy drastycznie skrócić. Kierował wysiłki na opracowanie i wprowadzenie w życie metod redukcji czasów przezbrojeń. Wysiłki te w końcu doprowadziły do skrócenia wszystkich przezbrojeń w Toyocie do tego stopnia, że wynoszą one najwyżej kilka minut. Nie budzi więc zdziwienia fakt, że obecnie Lean kojarzy się silnie[10] z małymi rozmiarami partii produkcyjnych i z metodami redukcji przezbrojeń. Jednakże potrzeba równoważenia przepływu wymagała znacznie więcej, niż tylko pokonania przeszkody wynikającej z przezbrojeń. Fakt, że większość gniazd wytwórczych nie była dedykowana do obróbki jakiejś jednej części składowej, czynił prawie niemożliwym ustalenie przez bezpośrednią obserwację prawdziwych problemów, zakłócających przepływ. Ohno był w pełni świadom tego, że istnieje zbyt wiele rzeczy, które można poprawić, i że bez skupienia wysiłków ku poprawie procesu na rzeczach najistotniejszych, równoważenie przepływu będzie trwało zbyt długo. System Kanban dostarczył mu rozwiązanie tego dylematu. Stosowana w Lean analogia potoku i skalistego dna pomaga zrozumieć, jak osiąga się równoważenie przepływu. Poziom wody w potoku odpowiada poziomowi zapasów, podczas gdy skały, pokrywające dno potoku – to problemy, zakłócające przepływ. Na dnie potoku jest wiele skał i ich usunięcie wymaga czasu i wysiłku. Powstaje pytanie, usunięcie których skał jest istotne. Odpowiedzi udziela obniżenie lustra wody – te skały, które wynurzą się znad wody, należy usunąć.

W początkach stosowania systemu Kanban, aby uzyskać rozsądny przerób, Ohno musiał zaczynać z wieloma pojemnikami, z których każdy zawierał niemałą ilość określonej części składowej. Stopniowo Ohno zmniejszał liczbę pojemników, a następnie i liczbę detali w każdym pojemniku. Jeśli przepływ nie uległ zauważalnemu zakłóceniu, wówczas kontynuowano zmniejszanie liczby

[7] Niemniej jednak w literaturze poświęconej systemowi Lean nie podkreśla się wyraźnie faktu, że TPS wymaga zniesienia stosowania lokalnych mierników efektywności.

[8] Ohno, Taiichi and Setsuo, Mito, *Just-In-Time For Today and Tomorrow* , Productivity Press, 1988.

[9] Pierwszy artykuł opublikował Ford W. Harris w czasopiśmie *Factory, The Magazine of Management,* Tom 10, Nr 2, Luty 1913, str. 135-136, 152. Od tamtego czasu artykuły na ten temat są ogłaszane nieomal co miesiąc.

[10] Przykładowo, czas zmiany matryc w fabrykach Toyoty uległ skróceniu z dwóch do trzech godzin w latach 1940 do mniej niż jednej godziny i zaledwie 15 minut w latach 1950, aby dojść do 3 minut w latach 1960 (Ohno pisał o tym w swej książce, *Toyota Production System*).

pojemników i liczby części w każdym z nich. Gdy przepływ ulegał zakłóceniu, stosowano metodę "Pięciokrotnego dlaczego?" dla precyzyjnego ustalenia podstawowej przyczyny owego zakłócenia. Tak zdefiniowany problem należało rozwiązać zanim można było dalej redukować liczbę i zawartość pojemników. Wymagało to czasu, ale ostatecznym rezultatem takiego postępowania była niesłychana poprawa wydajności. Trzeba zauważyć, że mimo iż w ostatnim dwudziestoleciu każda inna firma samochodowa wprowadziła tę lub inną wersję systemu Toyoty, osiągając przez to spore korzyści, niemniej nadal żaden inny producent samochodów nie dorównuje wydajności Toyoty. Pokazuje to, jak ważny jest poprawny wybór postępowania, która skupia wysiłki lokalnych udoskonaleń na rozwiązywaniu najważniejszych problemów. Niestety, w innych firmach wysiłki na rzecz poprawy są źle ukierunkowane, albowiem ich wiodącym motywem jest osiągnięcie obniżki kosztów, zamiast całkowitej koncentracji na poprawie przepływu. Ohno wkładał tak wiele wysiłku w skracanie przezbrojeń nie po to, aby uzyskać jakąś oszczędność kosztów. Gdyby jego celem była obniżka kosztów, nie "marnowałby" zaoszczędzonego czasu na dalsze zmniejszanie rozmiarów partii i w konsekwencji wykonywanie znacznie częstszych przezbrojeń. Ohno starał się zmniejszać liczbę części wybrakowanych nie w celu zmniejszenia jakichś (niewielkich zresztą) kosztów, lecz czynił to, aby wyeliminować istotne zakłócenia przepływu, które powstają w wyniku wykonania braku. Ohno nie próbował nawet wymuszać lepszych cen na dostawcach Toyoty, czy też obcinać płace w Toyocie, a są to przecież dwie największe pozycje kosztów. Zamiast tego wkładał on całą energię w poprawę przepływu.

Czynnikiem zamazującym ten obraz jest to, że owa koncentracja wysiłków na poprawie przepływu i ignorowanie rozważań, dotyczących lokalnych kosztów, w ostatecznym rezultacie prowadzi do znacznie niższego kosztu jednostkowego. Dokładnie tak samo, jak końcowym wynikiem zarzucenia mierników efektywności lokalnej jest znacznie wyższa wydajność siły roboczej. Jeśli wydaje się to dziwne, to jest tak dlatego, że kadra kierownicza jeszcze nie przyswoiła sobie pojęciowej różnicy między zarządzaniem działalnością operacyjną, skoncentrowanym na poprawie przerobu, zamiast skupiania się na obniżce kosztów. Jedną z konsekwencji koncentrowania uwagi na obniżce kosztów jest to, że niemal wszystkie inicjatywy, zmierzające do ustanowienia procesu permanentnego doskonalenia, szybko dochodzą do punktu malejących efektów i w rezultacie wiele z nich przekształca się w gołosłowne deklaracje. Jest to jednak kwestia zbyt rozległa i zbyt ważna, by ją pobieżnie potraktować w tym artykule.

Podsumowując, i Ford i Ohno realizowali cztery idee, o których od tej chwili będziemy mówili, jako o koncepcjach łańcucha dostaw:

1. Głównym celem działalności operacyjnej jest poprawa przepływu, czyli, co równoważne, skrócenie czasu realizacji.
2. Ten zasadniczy cel trzeba przełożyć na praktyczny mechanizm, który będzie wskazywał operatorom, kiedy nie powinni produkować, a zatem będzie zapobiegał nadmiernej produkcji. Ford użył do tego miejsca, Ohno - zapasu.
3. Trzeba skończyć z praktyką mierzenia lokalnej efektywności.
4. Trzeba ustanowić praktykę skupiania się na równomierności przepływu. Ford zastosował bezpośrednią obserwację. Ohno użył stopniowego zmniejszania liczby pojemników, a następnie stopniowego zmniejszania liczby części w każdym pojemniku.

Granice TPS

Podejście Ohno do rozwoju systemu Lean ilustruje ważną ideę, a mianowicie to, że istnieje różnica między praktyczną realizacją, a podstawową koncepcją, na której owa realizacja się zasadza. Idee podstawowe są ogólne, praktyczna realizacja jest zaś przeniesieniem tych idei na grunt specyficznego środowiska. Jak już widzieliśmy, owo przeniesienie jest niebanalne i wymaga wielu składników rozwiązania. Musimy pamiętać, że każda praktyczna realizacja przyjmuje jakieś założenia, niekiedy w sposób niejawny, dotyczące danego środowiska. Nie powinniśmy oczekiwać, że konkretna praktyczna realizacja zadziała w środowisku, w którym jej założenia nie są spełnione. Możemy zaoszczędzić sobie wiele wysiłków i rozczarowań, jeśli zadbamy o to, by wyraźnie zwerbalizować te założenia. Najbardziej wymagającym założeniem, jakie czyni TPS w odniesieniu do środowiska produkcyjnego, jest postulat, że środowisko to jest stabilne. A postuluje on stabilność w trzech różnych ujęciach.

Pierwsze ujęcie odsłania się, gdy zwrócimy uwagę na to, że nawet wtedy, gdy wybrano właściwe środowisko i najlepsi specjaliści nadzorują wprowadzanie systemu w życie, ustanowienie systemu Lean wymaga sporo czasu. Liker w swej książce *The Toyota Way* wskazuje, że wdrożenia systemu Lean, prowadzone przez organizację Toyota Supplier Support Center (TSSC, to organizacja, którą Toyota powołała do uczenia zasad TPS firm amerykańskich), potrzebowały co najmniej od sześciu do dziewięciu miesięcy na jedną linię produkcyjną.[11]

Nie jest to zaskakujące dla kogoś, kto ma świadomość, jak wiele czynników, zakłócających przepływ, występuje w każdym niemal środowisku produkcyjnym, i jak wrażliwy jest system Kanban w czasie, gdy zaczyna on dochodzić do docelowego niskiego poziomu zapasów. Ponieważ system Kanban wymaga czasu na wprowadzenie w życie, jego założeniem jest, że środowisko jest stosunkowo stabilne, czyli że procesy wytwórcze i same produkty nie zmieniają się w sposób istotny przez dłuższy czas. Toyota cieszy się stosunkowo stabilnym środowiskiem. Przemysł samochodowy dopuszcza zmiany tylko raz w roku (zmiana rocznika modelu) i zazwyczaj z roku na rok ogromna większość części składowych pozostaje ta sama. W wielu innych branżach sprawa wygląda zgoła inaczej. Na przykład, w głównych sektorach przemysłu elektronicznego okres życia większości wyrobów jest krótszy niż sześć miesięcy. Do pewnego stopnia niestabilność wyrobów i procesów ma miejsce w większości innych branż. Przykładowo, firma Hitachi Tool Engineering wytwarza narzędzia skrawające – względnie stabilny rodzaj wyrobów, jednak zacięta konkurencja zmusza tę firmę do wprowadzania nowych narzędzi skrawających, wymagających nowej technologii, co sześć miesięcy. Wprowadzanie systemu Lean w takim środowisku jest pracą syzyfową.

Drugi aspekt stabilności, wymaganej przez TPS, to stabilność w czasie popytu na każdy wyrób. Przypuśćmy, że czas realizacji pewnego wyrobu wynosi dwa tygodnie, ale popyt na ten wyrób jest sporadyczny, powiedzmy, że średnio na wyrób ten przychodzi tylko jedno zamówienie na kwartał. W obecnej chwili wyrób ten ma swój udział w produkcji w toku tylko przez dwa tygodnie w kwartale, a przez resztę czasu nie pojawia się on na produkcji. Inaczej będzie wyglądała sytuacja w warunkach funkcjonowania systemu Lean, który wymaga stałego utrzymywania pojemników dla każdego wyrobu pomiędzy każdą parą sąsiednich gniazd wytwórczych.

Firma Hitachi Tool Engineering wytwarza ponad dwadzieścia tysięcy różnych wyrobów – pozycji magazynowych. Popyt na większość tych pozycji jest sporadyczny. Konieczność stałego utrzymywania zapasu każdej pozycji magazynowej między każdą parą sąsiednich gniazd roboczych doprowadziłaby, w przypadku Hitachi, do znacznie większego zapasu produkcji w toku, niż to ma miejsce obecnie. Jest to w sposób oczywisty środowisko, nie nadające się do zastosowania rozwiązania Ohno.

Jednak najtrudniejszym aspektem stabilności, wymaganej przez TPS jest stabilność ogólnego obciążenia różnych rodzajów zasobów przez zamówienia odbiorców. Przypuśćmy, że zamówienia nie są jednolite, jak to ma miejsce w większości firm. Jest bardzo prawdopodobne, że obciążenie w tym tygodniu jakiegoś konkretnego gniazda wytwórczego okaże się znacznie niższe od jego zdolności, podczas gdy w następnym tygodniu przekroczy ono tę zdolność. W tym bardzo powszechnym przypadku, system Kanban, czyli system zapobiegający wytwarzaniu z wyprzedzeniem, spowoduje niedotrzymanie terminów dostawy w drugim tygodniu. Zamówienia w Toyocie są stosunkowo stabilne, niemniej jednak Toyota musiała ustalić tryb przyjmowania zamówień (a tym samym, przyrzekania dostaw), który ograniczałby zmiany wytwarzanego asortymentu z miesiąca na miesiąc. Większość firm nie jest w stanie wymusić na swoich klientach takich korzystnych warunków zamawiania.

Warto odnotować, że wymagana stabilność wykracza poza możliwą do osiągnięcia przez produkcję zdolność poprawy funkcjonowania. Wszystkie trzy aspekty stabilności dotyczą sposobu, w jaki firma projektuje i sprzedaje swoje wyroby, a nie sposobu ich wytwarzania. Niestety większość firm cierpi z powodu przynajmniej jednej z form niestabilności, o ile nie wszystkich trzech naraz.

Powyższe nie oznacza, że w środowiskach, dla których założenia systemu Lean nie są spełnione, w ogóle nie można stosować jakichś elementów tego systemu (np. komórki produkcyjne ustawione w kształt podkowy mogą być bardzo użyteczne w wielu środowiskach, a metody redukcji przezbrojeń można stosować w niemalże każdym środowisku). Nie znaczy to bynajmniej, że w takich

[11] Liker, Jeffrey K., *The Toyota Way* , McGraw-Hill, 2004. Wyd. polskie *Droga Toyoty*, MT Biznes, 2005, tłum. Grzegorz Łuczkiewicz.

środowiskach nie powinniśmy oczekiwać uzyskania tej samej skali wyników, jaką osiąga Toyota, a więc wyników, które sprawiają, że firma ta zajmuje dzisiaj taką pozycję, jaką zajmuje. Niemniej stosowania pewnych specyficznych metod systemu Lean, czy osiągania satysfakcjonujących wyników dzięki jakimś programom oszczędnościowym, nie należy utożsamiać z wprowadzeniem do praktyki systemu Lean.

Znaczenie przepływu w środowiskach stosunkowo niestabilnych

Ford i Ohno otworzyli nam oczy na fakt, że lepszy przepływ, czyli skrócenie czasów realizacji, prowadzi do znacznie wydajniejszej działalności operacyjnej. Udowodnili to w przypadku środowisk stabilnych. Jaki jest jednak wpływ ulepszonego przepływu na środowiska stosunkowo niestabilne?

Pierwsza forma niestabilności, to niestabilność wywołana przez krótki czas życia wyrobów. Gdy okres życia wyrobu jest krótki, nadmierna produkcja może stać się zapasem przestarzałym i przez to bezwartościowym. Ponadto, jako że okres życia wyrobu jest krótki, długie czasy wytwarzania mogą doprowadzić do utraty popytu rynkowego. Przypuśćmy, dla przykładu, że okres życia jakiegoś wyrobu wynosi około sześciu miesięcy, a czas realizacji tego wyrobu wynosi dwa miesiące. Długi czas realizacji powoduje utratę sprzedaży nie dlatego, że nie ma popytu, ale dlatego, że przez znaczną część okresu rynkowej żywotności wyrobu produkcja nie potrafi zaspokoić tego popytu.

Drugi rodzaj niestabilności to niestabilność popytu na poszczególne produkty w czasie. Powszechną praktyką na ograniczanie kłopotów w środowiskach, w których występuje wiele pozycji magazynowych, podlegających sporadycznemu popytowi, jest dążenie do zaspokajania owego popytu z zapasów magazynowych. Wadą takiej praktyki jest wysoki poziom zapasu wyrobów gotowych o niezwykle długim okresie przechowywania, w połączeniu z wysokim poziomem braków. System wytwórczy, który potrafi zorganizować działalność w hali produkcyjnej w sposób umożliwiający uzyskanie znacznie lepszego przepływu, ma radykalny wpływ na takie środowiska.

Środowiska, odczuwające skutki trzeciego przejawu niestabilności – niestabilności łącznego obciążenia mocy – mogą najbardziej skorzystać z istotnie lepszego przepływu. Rezultaty przeciążenia różnych zasobów powodują, że takie firmy zazwyczaj mają słabe wyniki w zakresie terminowości dostaw (< 90%) i w rezultacie są one skłonne zwiększać swoje moce. Doświadczenie pokazuje, że z chwilą, gdy takim firmom udaje się radykalnie poprawić przepływ, nie tylko terminowość ich dostaw dochodzi do do poziomu powyżej 95%, ale równocześnie ujawnia się nadmiar mocy nawet do 50%.[12]

Ohno udowodnił, że idee wprowadzone przez Forda nie są ograniczone do masowej produkcji wyrobu jednego rodzaju. I aczkolwiek przeszkody przy stosowaniu tych idei w mniej restrykcyjnym środowisku wydawały się nie do pokonania, analiza geniuszu i uporczywości Ohno pokazała nam nie tylko, że jest to wykonalne, lecz również, jak tego dokonać. Obecnie zdajemy sobie sprawę, że:

- TPS jest ograniczony do środowisk relatywnie stabilnych;
- większość środowisk boryka się z niestabilnością, oraz że
- środowiska relatywnie niestabilne mają znacznie więcej do wygrania dzięki poprawie przepływu, niż środowiska stabilne.

Czy teraz, mając świadomość tych faktów, nie powinniśmy pójść śladami Taiichi Ohno? Czyż nie powinniśmy cofnąć się do idei łańcucha dostaw, aby wywieść z nich skuteczną praktyczną realizację, dostosowaną do relatywnie niestabilnych środowisk?

Zastosowanie idei łańcucha dostaw, oparte na limicie czasu

Najbardziej intuicyjną podstawą mechanizmu ograniczającego nadmierną produkcję jest nie

[12] Mabin, Victoria J. and Balderstone, Steven J., *The World of the Theory of Constraints* , CRC Press LLC, 2000. W międzynarodowej literaturze na temat teorii ograniczeń przeanalizowano przeciętne uzyskane wyniki: 70% skrócenie czasów realizacji, 44% poprawa terminowości dostaw i 76% wzrost dochodów/ przerobu/zysku.

miejsce, czy zapas, lecz czas. Jeśli zamierzamy zapobiec wytwarzaniu przed czasem, nie powinniśmy przed czasem wydawać materiałów. Zastosowanie czasu, jako podstawy, jest nie tylko bardziej intuicyjne, a wskutek tego, łatwiejsze do zaakceptowania przez zatrudnionych bezpośrednio na produkcji, ale ma też tę zaletę, że jest odpowiednie dla niestabilnych środowisk, albowiem jest znacznie mniej wrażliwe na zakłócenia przepływu.

Odporność mechanizmu opartego na czasie wynika stąd, że ogranicza on wprost całkowitą ilość pracy w systemie, a nie tylko poprzez ograniczenie ilości pracy między każdą parą sąsiednich gniazd roboczych. Na liniach potokowych lub w systemach opartych na kartach (Kanban) zapas dozwolony między gniazdami roboczymi jest ograniczony do absolutnego minimum (co zazwyczaj odpowiada ilości pracy znacznie poniżej jednej godziny). Z tego względu, gdy jakieś gniazdo robocze ma przestój dłuższy od krótkiej chwili, następne gniazda robocze prawie natychmiast zostają pozbawione pracy, a praca gniazd poprzedzających zostaje "zablokowana". Jeśli w przypadku któregokolwiek z gniazd roboczych, łączny czas trwania przerw w dostawach materiałów do pracy i blokad, jest dłuższy niż nadwyżka mocy przetwórczych tego gniazda, przerób firmy ulega redukcji. Czułość linii potokowych i systemów, opartych na kartach, wynika z faktu, że zakłócenie powstałe w jednym z gniazd zużywa również zdolności zarówno gniazd następujących, jak i poprzedzających. Zjawisko takie prawie nie występuje w systemach, opartych na limicie czasowym, bowiem praca, z chwilą, gdy została zwolniona do wykonania, nie jest sztucznie wstrzymywana.

Trudność w stosowaniu systemu opartego na limicie czasu polega na tym, że dla każdego zlecenia trzeba wstrzymywać wydanie odpowiednich materiałów tak, aby nastąpiło ono dopiero w stosownym czasie przed ustalonym terminem realizacji tego zlecenia. Ale jak ten stosowny czas uruchomienia zlecenia wyliczyć? Gdy na scenie produkcyjnej pojawiły się komputery, czyli we wczesnych latach 60., wydawało się, że wreszcie otrzymaliśmy właściwe narzędzie do przetwarzania olbrzymiej ilości szczegółów i wykonywania rachunków niezbędnych do obliczenia stosownych czasów wyprzedzenia dla każdego materiału i zlecenia. W ciągu dziesięciu lat w licznych firmach na całym świecie opracowano mnóstwo programów komputerowych przeznaczonych wyłącznie do realizacji tego zadania. Niestety, spodziewane rezultaty w postaci poprawionego przepływu i zmniejszonego zapasu produkcji w toku nie urzeczywistniły się.

Problem w tym, że czas potrzebny na przekształcenie materiału w końcowy wyrób, gotowy, aby dostarczyć go klientowi, zależy bardziej od czasu, jaki musi on spędzić w kolejkach w oczekiwaniu na zasób, zajęty przetwarzaniem innego zlecenia, lub przed stanowiskiem montażu w oczekiwaniu na nadejście jakiejś innej części), niż od samego czasu właściwego przetwarzania. Powszechnie wiadomo, że niemalże w każdej produkcji przemysłowej – za wyjątkiem linii procesowych i firm, stosujących system Kanban – czas, w którym partia detali jest fizycznie obrabiana, stanowi zaledwie około 10% całkowitego czasu realizacji zlecenia. W rezultacie decyzja o tym, kiedy wydać materiał, określa, gdzie powstaną i jak długie będą kolejki, co z kolei określa, ile czasu zajmie wykonanie zlecenia, a to określa, kiedy wydać materiał. Mamy więc do czynienia z problemem jajka i kury. W latach siedemdziesiątych zasugerowano, by problem ten rozwiązać przez iteracyjne powtarzanie procedury (planowanie potrzeb materiałowych – MRP – z pętlą sprzężenia zwrotnego), czyli wykonanie przebiegu programu komputerowego, sprawdzenie obecności planowanych przeciążeń różnych zasobów (długości kolejek), modyfikację wymaganych terminów realizacji zleceń tak, aby wyeliminować przeciążenia, i kontynuowanie tego procesu, aż nie usuniemy wszystkich znaczących przeciążeń. Niewiele czasu było potrzeba, aby doświadczenie pokazało, że proces ten nie jest zbieżny - bez względu na liczbę iteracji, przeciążenia przemieszczają się po prostu z jednego rodzaju zasobów na inny.

W rezultacie, już w latach siedemdziesiątych, systemy komputerowe stosowano nie do wyznaczania dokładnych harmonogramów zwalniania materiałów do produkcji, lecz raczej do dostarczania lepszej informacji o ilościach (i terminach) zamawiania materiałów u dostawców. Oficjalna nazwa tych systemów – planowanie potrzeb materiałowych (MRP) - miała odzwierciedlać ich główne zastosowanie. [13]

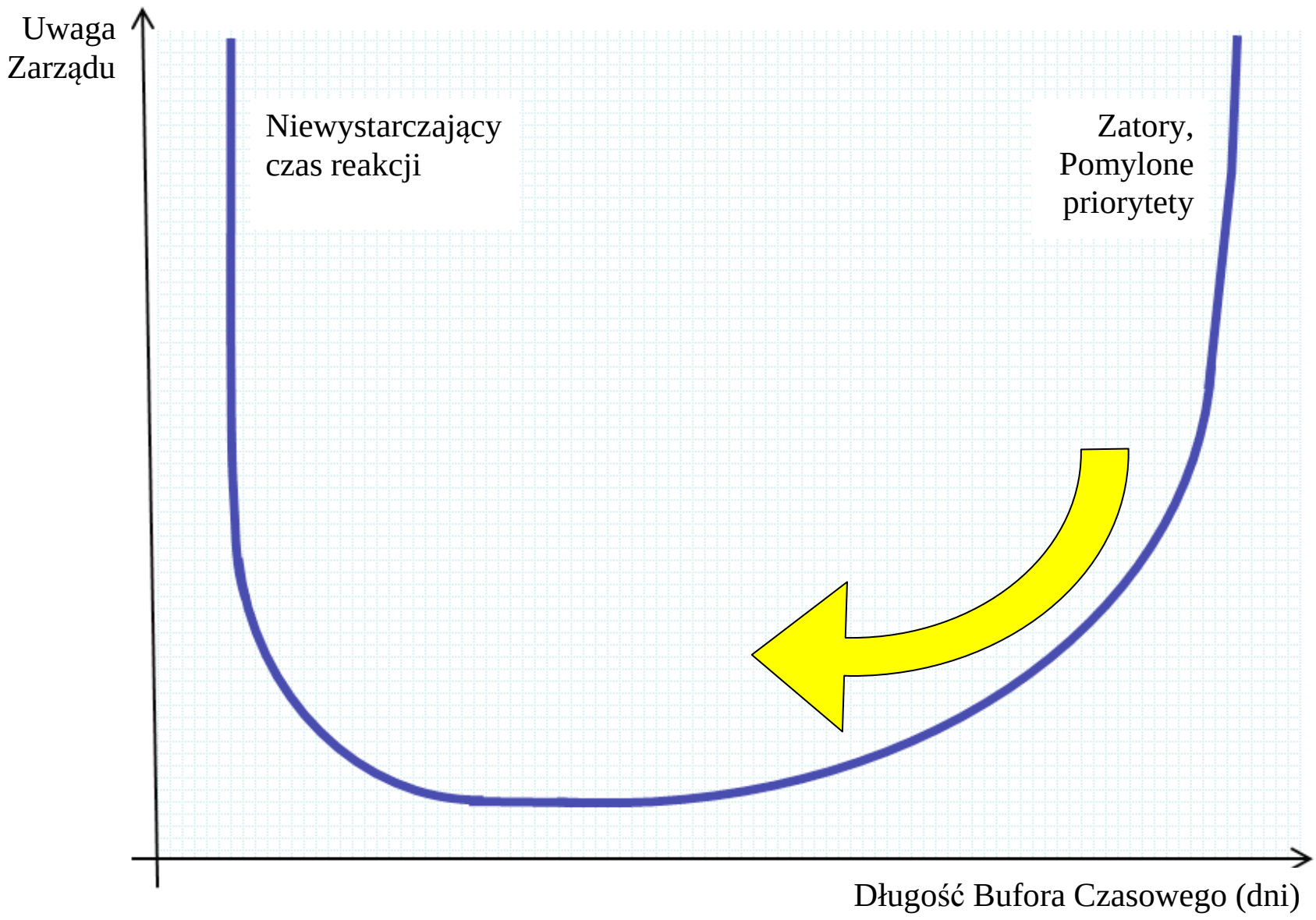
Wracając do podstaw, czyli kierując się ideą łańcucha dostaw, celem jest poprawa przepływu, a zatem skrócenie czasu realizacji. Przyjęcie czasu, a nie miejsca, czy zapasu, za podstawę mechanizmu, wskazującego operatorom, kiedy nie należy produkować, zobowiązuje nas do tego,

[13] Orlicky, Joseph, *Material Requirements Planning* , McGraw-Hill Book Company, 1975, wyd. polskie, *Planowanie potrzeb materiałowych,* PWE, 1981.

byśmy starali się wydawać odpowiedni materiał w stosownie krótkim czasie przed wymaganym terminem realizacji zlecenia, a więc dokładnie wtedy, gdy jest on potrzebny (just-in-time). Ale, co właściwie rozumiemy przez "dokładnie wtedy, gdy potrzeba" („just-in-time”)? Mimo iż termin „dokładnie na czas” jest podstawowym pojęciem w systemie Lean, stosowany jest on raczej w znaczeniu przenośnym, niż ściśle wartościowym. W systemie Lean, przez wytwarzanie dokładnie na czas, nie rozumiemy, rzecz jasna, że dana konkretna część, którą właśnie wykonaliśmy, musi się znaleźć na stanowisku ekspedycji, gotowa do wysyłki, w następnej sekundzie…, czy też minucie…, czy nawet godzinie. W rzeczywistości, w najlepszym nawet systemie Kanban przypuszczalnie owa część nie będzie zaraz przetwarzana przez następujące gniazdo robocze, co można wyprowadzić z faktu, że rutynowo pełne pojemniki oczekują między gniazdami roboczymi. A zatem, jaki odstęp czasu będziemy uważali za spełniający warunek „dokładnie na czas”? Zapytajmy jeszcze wyraźniej: jeśli zamierzamy ograniczyć nadmierną produkcję przez limitowanie wydań materiałów, to z jakim wyprzedzeniem w stosunku do wymaganego terminu realizacji zlecenia powinniśmy zwolnić materiał dla tego zlecenia?

Jednym ze sposobów uzyskania rozsądnej odpowiedzi jest zbadanie, jaki wpływ ma wybór owego wyprzedzenia na to, ile uwagi musi poświęcać kierownictwo produkcji, aby dotrzymać wszystkich przyrzeczonych terminów realizacji zleceń. Przypuśćmy, że wydajemy materiał z wyprzedzeniem, równym dokładnie czasowi, jaki jest faktycznie potrzebny na wykonanie danego zlecenia. Taki wybór będzie wymagał wiele uwagi kierownictwa, aby ściśle obserwować poszczególne operacje, bowiem jakiekolwiek opóźnienie na jakiejś operacji, czy nawet opóźnienie w przemieszczaniu części między operacjami, spowoduje przekroczenie terminu realizacji. Ponadto konieczne będzie sporządzenie precyzyjnego harmonogramu dla zapewnienia, że nie utworzą się żadne kolejki, gdyż każda kolejka powoduje opóźnienie obróbki części w niej oczekujących. Taki wybór jest oczywiście niepraktyczny, nawet nieskończenie wiele uwagi kierownictwa nie wystarczy, aby dotrzymać wszystkich zadeklarowanych terminów realizacji. Musimy wybrać jakiś dłuższy okres wyprzedzenia, okres, który będzie zawierał zabezpieczenie, pochłaniające opóźnienia. Potrzeba uwzględnienia zabezpieczenia uzasadnia nazwanie okresu między zwolnieniem materiału, a terminem realizacji zlecenia „buforem czasu”.

Wybór dłuższych buforów czasu wydłuża czas realizacji i zwiększa produkcję w toku, ale ponieważ dłuższe bufory czasu oznaczają większe zabezpieczenie, można się spodziewać, że przy znacznie mniejszej ilości uwagi ze strony kierownictwa, większy procent zleceń będzie wykonywany w deklarowanych terminach lub przed nimi. Stwierdzenie to jest słuszne dla stosunkowo krótkich buforów czasu, natomiast gdy bufory czasu są znaczne, zaczyna występować inne nieprzyjemne zjawisko. Musimy pamiętać, że im dłuższy jest wybrany bufor czasu, tym wcześniej materiał zostaje wydany na produkcję, co oznacza, że więcej zleceń znajduje się równocześnie w realizacji. Gdy w hali produkcyjnej jest zbyt wiele zleceń, zaczynają pojawiać się zatory. Im więcej zatorów, tym więcej uwagi musi poświęcać kierownictwo na porządkowanie priorytetów zleceń. Ilość niezbędnej uwagi ze strony kierownictwa, jako funkcja długości wybranego bufora czasu, została przedstawiona schematycznie na rys. 1



Rysunek 1

Zakłady produkcyjne, które zastosowały system Forda lub Ohno mają przeciętny czas realizacji zaledwie nieco dłuższy od faktycznego czasu operacji, a kierownictwo nie musi prawie wcale angażować swej uwagi, aby podpowiadać pracownikom produkcyjnym, jakim zleceniami mają się w danej chwili zajmować. Zakłady te zdecydowanie znajdują się w lewej części niskiego, płaskiego odcinka wykresu. W którym, jednak, miejscu na wykresie znajduje się ogromna większość zakładów, stosujących bardziej konwencjonalne praktyki zarządzania produkcją?

Jak mówiliśmy, w fabrykach konwencjonalnych partie części zaledwie przez 10% czasu realizacji podlegają faktycznej obróbce. Przez mniej więcej 90% czasu partie te albo oczekują w kolejce do zasobu, albo czekają na jakiś inny rodzaj części przed montażem. Od Forda, a w jeszcze większym stopniu od Ohno, nauczyliśmy się, że nie powinniśmy przyjmować rozmiarów partii jako z góry określonych, że partia ekonomicznie uzasadniona jest nieekonomiczna i raczej powinniśmy dążyć do osiągnięcia ciągłego przepływu (przepływu jednej sztuki). Mając takie przekonanie, łatwo zrozumieć, że gdy przetwarzana jest partia części, to za wyjątkiem procesów takich, jak mieszanie czy suszenie, faktycznie obrabiana jest tylko jedna sztuka, podczas gdy pozostałe sztuki z danej partii czekają na swoją kolej. Oznacza to, że w zakładach konwencjonalnych, które stosują rozmiary partii większe niż 10 sztuk w partii, a jest to przypadek większości środowisk produkcyjnych, faktyczny czas obróbki stanowi mniej niż 1% całkowitego czasu realizacji. Występuje też inne zjawisko, typowe dla tych firm, a mianowicie, jakikolwiek byłby formalny system ustalania priorytetów, o ile w ogóle jakiś formalny system funkcjonuje, system faktyczny jest taki: “pilne”, “bardzo pilne” i “odłóż wszystko inne i rób to natychmiast”. Takie firmy znajdują się najwyraźniej wysoko na stromym odcinku w prawej części wykresu zależności absorbowanej uwagi kierownictwa od bufora czasu (rysunek 1).

Położenie na stromiźnie po prawej stronie oznacza sytuację przegraną z każdego punktu widzenia, bowiem czasy realizacji są bardzo długie (w stosunku do faktycznych czasów obróbki), zapasy są wysokie i w wielu przypadkach firma cierpi na kiepską terminowość dostaw (<90%), mimo absorbowania znacznej uwagi kierownictwa. Wiedząc o tym, że wybór przez kierownictwo krótszego bufora czasu (a więc przesunięcie się na długi, płaski obszar wykresu) bardzo znacząco poprawiłby tę sytuację, jak to możliwe, że ogromna większość konwencjonalnie zarządzanych firm znajduje się na takiej ogólnie przegranej pozycji?

Odpowiedzi udzielili Ford i Ohno. Poprzez swoje działania udowodnili oni w sposób zdecydowany, że, wbrew powszechnemu przekonaniu, dążenie do stałej aktywizacji wszystkich zasobów przez cały

dostępny czas, nie jest receptą na wydajną działalność operacyjną. Wręcz przeciwnie, prawda jest całkiem inna: aby osiągnąć wydajną działalność operacyjną, trzeba położyć kres stosowaniu lokalnych mierników efektywności. Tradycyjne firmy próbują jednak osiągnąć pełną aktywizację zasobów. Gdy zasoby w górze strumienia nie są wąskimi gardłami, a tak się sprawy mają w większości środowisk, zawsze będą one co jakiś czas bez pracy. Dla zapobieżenia temu, wydaje się na produkcję materiał, niezbędny do wykonania zleceń, przyjętych z bardziej odległym terminem realizacji, czy nawet zleceń przewidywanych. Nieuniknioną konsekwencją takiego postępowania są dłuższe kolejki. Dłuższe kolejki powodują, że niektóre zlecenia nie są realizowane na czas, co z kolei interpretuje się, jako argument za jeszcze wcześniejszym zwalnianiem materiału na produkcję. A także jako oznakę tego, że nie posiadamy wystarczających zdolności wytwórczych. Nietrudno wyobrazić sobie, jak takie siły wypychają firmy w górę stromizny wykresu.

Dobrym punktem wyjścia ku poprawie przepływu będzie taki wybór bufora czasu, aby był on równy połowie aktualnego czasu realizacji. Taki wybór zapewni, że firma znajdzie się gdzieś na płaskim odcinku wykresu. Nie ma sensu tracić czas na próby odnalezienia lub obliczenia punktu optymalnego. Natychmiastowe korzyści są zbyt znaczne, aby je odraczać, a kolejne wysiłki na rzecz bardziej równomiernego przepływu zmodyfikują sam wykres.

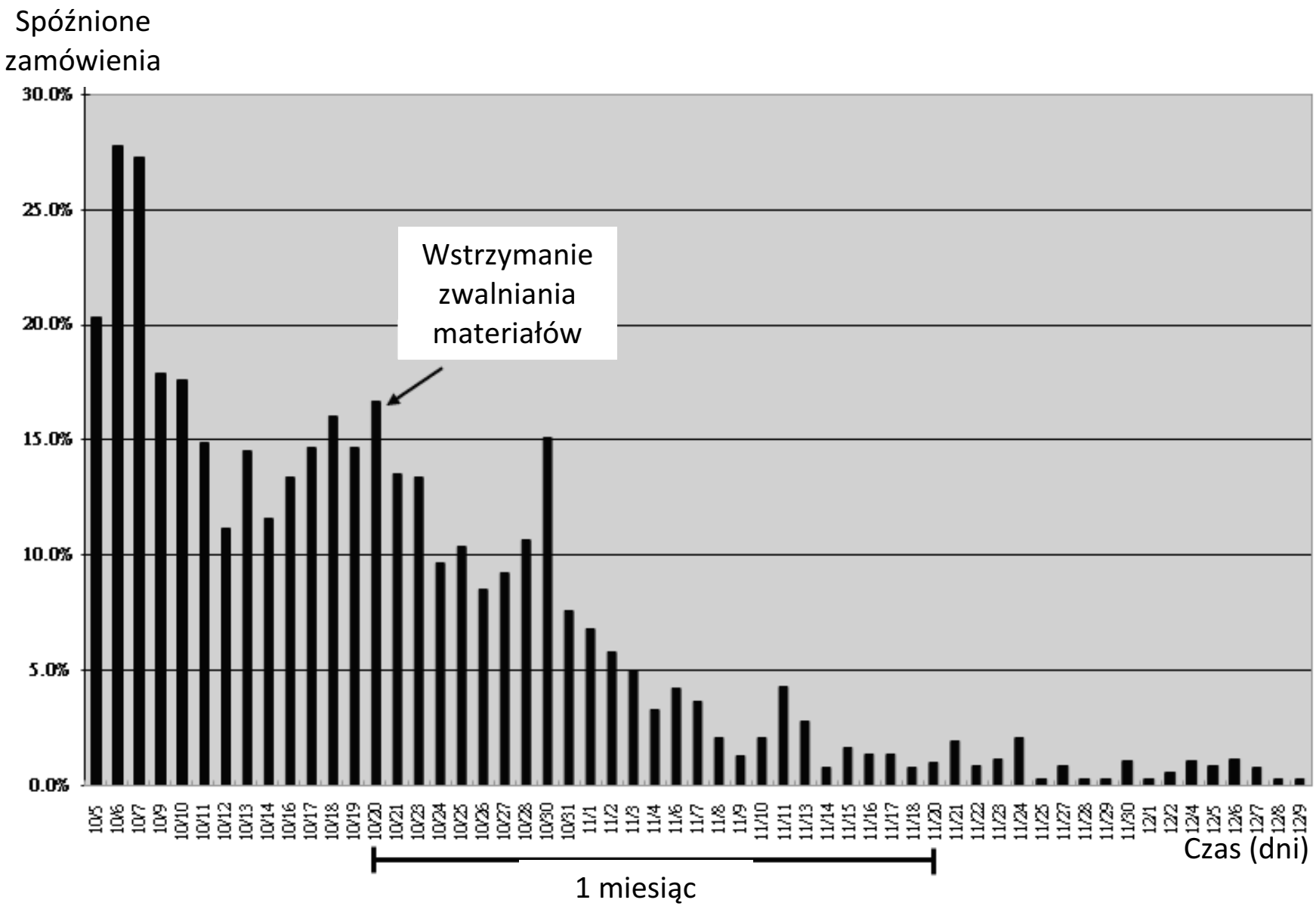
Ograniczenie wydań materiałów tak, aby wyprzedzały one terminy realizacji poszczególnych zleceń dokładnie o długość bufora czasu - równego, przypomnijmy, połowie aktualnego czasu realizacji, znacząco poprawi wskaźnik terminowości, skróci czas realizacji do połowy obecnego, a tym samym, wskutek likwidacji nadmiernych zapasów, zmniejszy produkcję w toku o więcej niż połowę obecnego poziomu.

Nie można jednak oczekiwać, że sama tylko ta zmiana podniesie wskaźnik terminowości dużo ponad 95%. Po prostu nadal na produkcji znajduje się wiele zleceń, występują kolejki przed zasobami, a pozostawienie kolejności, w jakiej zlecenia są przetwarzane, grze przypadku spowoduje, że wiele z nich zostanie ukończonych z opóźnieniem. Potrzebny jest system określania priorytetów. Sama potrzeba takiego systemu nie powinna jednak otwierać furtki dla poszukiwania wyrafinowanych algorytmów ustalania priorytetów. Liczba napływających zleceń stale się zmienia, zakres prac do wykonania różni się w zależności od zlecenia, długość kolejki zmienia się nieustannie i nie zapominajmy wreszcie, że nadal zdarzają się zakłócenia. Krótko mówiąc, jest to środowisko o wysokiej zmienności. Nauka, jaką Shewhart przekazał produkcji z doświadczeń fizyki, a Deming rozpowszechnił w całym świecie, mówi, że próby postępowania z dokładnością większą niż poziom szumu (w naszym przypadku oznacza to próbę stosowania wyrafinowanych algorytmów, uwzględniających każdy możliwy parametr w środowisku o wysokiej zmienności) nie poprawiają, lecz pogarszają stan rzeczy. Wynikiem będzie najpewniej nie poprawa, a pogorszenie wskaźnika terminowości.

Prosty system ustalania priorytetów otrzymamy, gdy tylko uświadomimy sobie, że bufor czasu, równy połowie obecnego czasu realizacji, jest wciąż znacznie dłuższy od faktycznego czasu obróbki, a ponieważ radykalnie redukuje on zatory, więc bez żadnej interwencji wiele zleceń zostanie ukończonych w ciągu zaledwie jednej trzeciej bufora czasu, a większość z nich - w ciągu pierwszych dwóch trzecich bufora. Opierając się na świadomości tych faktów, priorytety przypisuje się zleceniom, poprzez „zarządzanie buforami". Dla każdej partii śledzimy czas, jaki upłynął od chwili wydania materiału do jej produkcji. Jeśli upłynęło czasu mniej, niż jedna trzecia bufora, kolor priorytetu jest zielony, jeśli ponad jedną trzecią, ale mniej od dwóch trzecich - kolor jest żółty, jeśli ponad dwie trzecie - kolor jest czerwony, wreszcie jeśli termin realizacji zlecenia już minął- kolor jest czarny. Zlecenia czarne mają wyższy priorytet od czerwonych i tak dalej. Jeśli dwie partie mają ten sam kolor, próba decydowania o tym, którą z nich przetwarzać najpierw jest doskonałym przykładem dążenia do dokładności, przewyższającej poziom szumu.

Wprowadzenie takiego systemu na produkcję jest stosunkowo łatwe. Na pierwszym etapie nie ma żadnej potrzeby dokonywania jakichkolwiek zmian fizycznych, wystarczy tylko wstrzymywać wydawanie materiałów wcześniej, niż o połowę historycznie ustalonego czasu realizacji przed odpowiednim terminem dostawy, i polecić produkcji, by przestrzegała systemu priorytetów, oznaczonych kolorami. Rezultaty są imponujące, zwłaszcza w porównaniu z włożonym wysiłkiem. Aby uzyskać pierwsze wrażenie o tym, jakie wyniki i jak szybko osiąga się już na tym pierwszym etapie, rysunek 2 przedstawia faktyczne udziały procentowe opóźnionych zleceń w zatrudniającej

2000 osób fabryce produkującej tysiące różnych rodzajów metalowych przyborów kuchennych.



Rysunek 2

Rzecz jasna, trzeba zrezygnować ze stosowania lokalnych mierników wydajności, bowiem w przeciwnym razie nadal będzie wywierany nacisk na zbyt wczesne zwalnianie materiałów do produkcji. Doświadczenie pokazuje, że szybkość, z jaką wszyscy pracownicy produkcji dostrzegają dodatni wpływ wprowadzonego rozwiązania, czyni tę zmianę niemal pozbawioną oporu.

Niemniej jednak, w większości środowisk wciąż jeszcze zdarzają się zlecenia, opóźnione w stosunku do zadeklarowanych terminów realizacji, i wciąż jest tam ogromny potencjał dla doskonalenia. Trzeba urzeczywistnić również i czwartą ideę - ustanowienia procesu koncentracji wysiłków na równoważeniu przepływu.

Pierwszy krok w kierunku równoważenia przepływu jest stosunkowo łatwy. Wstrzymywanie zwalniania materiałów ujawnia obfitą nadwyżkę zdolności wytwórczych, która wcześniej była niewidoczna. Jest jednak prawdopodobne, że niektóre gniazda robocze mają mniejszą nadwyżkę zdolności od pozostałych. Takie gniazda są oznakowane, gdyż gromadzi się przed nimi kolejka oczekujących zapasów. Fakt rezygnacji z lokalnych mierników wydajności pomaga w określeniu prostych działań, niezbędnych do zwiększenia zdolności wytwórczych tych gniazd. Działań tak prostych, jak zapewnienie, że gniazdo robocze o ograniczonej zdolności nie będzie stało bezczynnie podczas przerwy na posiłek, czy podczas zmiany obsługi, odciążenie go przez przekazanie części zadań gniazdom mniej wydajnym, ale posiadającym dostatecznie dużą nadwyżkę mocy itd. [14]

Ponieważ powyższe działania zwiększają praktycznie zdolności gniazd roboczych, które są źródłem kolejek, same kolejki ulegają skróceniu i mniejsza liczba zleceń dochodzi do stanu, w którym otrzymują czerwony priorytet. Oznacza to, że bufor czasu staje się niepotrzebnie długi. Skuteczna reguła regulowania bufora czasu bez podejmowania ryzyka pogorszenia wysokiego wskaźnika terminowości dostaw mówi, aby skracać bufor czasu wtedy, gdy liczba zleceń czerwonych jest mniejsza niż 5% ogólnej liczby zleceń zwolnionych do produkcji, a wydłużać go, gdy udział zleceń czerwonych przekracza 10%.

Firma, która postępuje zgodnie z powyższymi wskazówkami, osiągnie w ciągu paru miesięcy bardzo wysokie wskaźniki terminowości, znacznie krótsze czasy realizacji i ujawni sporą nadwyżkę mocy. Właśnie wtedy zaczyna się prawdziwe wyzwanie. W przeszłości, niejednokrotnie i nazbyt często

[14] Goldratt, Eliyahu M. and Cox, Jeff, *The Goal: A Process of Ongoing Improvement* , North River Press, 1984, wyd. polskie, *Cel. Doskonałość w produkcji*, Werbel, 2000, tłum. Małgorzata Duraj-Czyżewska.

odpowiedzią wyższego kierownictwa na w pełni ujawnioną nadwyżkę mocy było dopasowanie zdolności wytwórczych do „właściwych rozmiarów” i uzyskanie oszczędności przez obniżkę kosztów. Jest to poważny błąd. Owa nadwyżka mocy – to pracownicy, ci sami, którzy dopiero co pomogli firmie poprawić wyniki i w prostej konsekwencji zostają za to „wynagrodzeni” zwolnieniem ich samych, albo ich przyjaciół, z pracy. We wszystkich przypadkach, w których takie działania podjęto, nieunikniony gniewny sprzeciw szybko pogorszył wyniki fabryki do poziomu niższego niż w punkcie wyjścia. Miejmy nadzieję, że takie postępowanie zarządów mamy już za sobą.

Bardziej sensowne podejście do owej ujawnionej nadwyżki zdolności wytwórczych, polega na tym, aby obrócić ją w dodatkową korzyść, poprzez zachęcanie sprzedawców, aby zwiększali sprzedaż dzięki poprawionym wynikom działalności firmy. Wzrastająca sprzedaż może z łatwością spowodować pojawienie się rzeczywistego wąskiego gardła. Ignorowanie ograniczonych zdolności wąskiego gardła przy podejmowaniu zobowiązań co do terminów dostaw nowych zamówień pogorszy wskaźnik terminowości i sprzedaż rozczarowanym klientom zacznie gwałtownie spadać. Sprawą zasadniczą jest wzmocnienie powiązań między działem sprzedaży i produkcją - oto prawdziwe wyzwanie. Trzeba ustanowić system, który zapewni, że za każdym razem zobowiązanie co do terminu dostawy jest podejmowane wyłącznie na podstawie jeszcze nie przydzielonej zdolności wytwórczej wąskiego gardła.

Wąskie gardło staje się werblem, wybijającym rytm akceptacji zamówień, bufor czasu przekształca obiecane terminy realizacji w daty zwolnień zleceń na produkcję, a działanie wstrzymujące wydania materiałów tworzy linę, wiążącą zamówienia ze zleceniami przekazywanymi na produkcję. Oto powód, dla którego owo oparte na kontroli czasu zastosowanie teorii ograniczeń stało się znane jako system "werbla-bufora-liny" (po angielsku „the Drum-Buffer-Rope system”, w skrócie DBR). W chwili obecnej szeroko rozpowszechnione są działania doświadczalne w kierunku doskonalenia tego procesu tak, aby poprawiać działalność operacyjną poprzez rejestrację i analizę przyczyn pojawiania się zleceń czerwonych.

Przykład Hitachi

Firma Hitachi Tool Engineering Ltd., której roczne obroty wynoszą 24 miliardy jenów, projektuje i produkuje ponad 20 000 rozmaitych narzędzi skrawających. Popyt na większość wyrobów jest sporadyczny, a klienci tej branży zmuszają ich do wypuszczania nowych typów narzędzi co sześć miesięcy. Gdy na rynku pojawiają się nowe typy wyrobów, wcześniejsze typy narzędzi stają się przestarzałe. Nic dziwnego, że ich wysiłki w kierunku zastosowania systemu Lean kończyły się niepowodzeniem.[15]

Firma Hitachi Tool Engineering Ltd. rozpoczęła wprowadzanie DBR w jednej ze swych czterech fabryk w Japonii w 2000 roku. Skokowa poprawa wskaźnika terminowości (z 40% do 85%), związana ze zmniejszeniem produkcji w toku i skróceniem czasów realizacji o połowę, której towarzyszyło zwiększenie o 20 % ilości wysyłanych wyrobów przy tym samym zatrudnieniu, zachęciła ich do rozszerzenia wdrożenia. Do 2003 roku wprowadzili oni DBR we wszystkich czterech fabrykach. [16]

Radykalne skrócenie czasu realizacji i znacznie poprawiona wrażliwość na potrzeby klientów umożliwiły obniżenie zapasów w łańcuchu dostaw – u dystrybutorów – z 8 do 2.4 miesięcy wartości obrotów. Obniżka zapasów poprawiła w sposób radykalny stopę zwrotu zainwestowanego kapitału u dystrybutorów, uwolniła ich gotówkę i wzmocniła ich stosunki z firmą Hitachi. Nic dziwnego, że dystrybutorzy rozszerzyli oferowany przez nich asortyment narzędzi Hitachi, co doprowadziło do 20% wzrostu sprzedaży (na stabilnym rynku).

Prawdziwy wpływ zastosowanego rozwiązania poznamy, gdy dokonamy oceny rentowności tej firmy w świetle faktu, że w okresie od 2002 do 2007 roku cena surowców (metali) wzrosła znacznie bardziej od ceny sprzedaży narzędzi skrawających. W takich warunkach zyski firmy powinny stopnieć. Natomiast faktycznie roczny zysk netto przed opodatkowaniem firmy Hitachi Tool Engineering Ltd. wzrósł z 1,1 mld. jenów w roku budżetowym, kończącym się w marcu 2002, do 5,3

[15] Umble, M., Umble E., and Murakami, S., “Implementing theory of constraints in a traditional Japanese manufacturing environment: the case of Hitachi Tool Engineering,” *International Journal of Production Research* , Vol. 44, No. 10, 15, Maj 2006, str. 1863 – 1880.

[16] Ibid.

mld. jenów w roku, kończącym się w marcu 2007, co stanowi pięciokrotny wzrost zysku netto w ciągu pięciu lat. Stopa zysku firmy Hitachi Tool Engineering Ltd wzrosła z 7.2% w 2002 r. do 21.9% w 2007 r., co jest najwyższym wskaźnikiem zanotowanym kiedykolwiek w tego typu przemyśle.[17]

Granice DBR

Jak już uprzednio zwracaliśmy uwagę, jakakolwiek praktyczna realizacja przyjmuje pewne założenia, niekiedy niejawne, co do środowiska i nie należy oczekiwać, że dana realizacja zadziała w środowiskach, co do których jej założenia są niesłuszne. Założenia, jakie czyni system DBR są oczywiste. Przyjmuje on mianowicie, że faktyczny czas przetwarzania (obróbki) jest bardzo krótki (<10%) w porównaniu z aktualnym czasem realizacji zlecenia. Założenie to jest słuszne dla wielu, jeśli nie dla większości, typowych środowisk produkcyjnych. Niemniej, jest ono zdecydowanie niesłuszne dla bardzo szerokiego spektrum środowisk, które tradycyjnie określa się mianem środowisk projektowych, czyli środowisk zorientowanych na realizację przedsięwzięć.

W środowiskach projektowych czas przetwarzania jest stosunkowo długi, a niecierpliwe oczekiwanie ze strony klientów szybkiego zakończenia przedsięwzięć zmusza wykonawców do przyrzekania terminów realizacji, które są zaledwie dwu, lub, rzadziej, trzykrotnie dłuższe niż czas przetwarzania. Nie dziw, że osiągnięcia w tych środowiskach są mierne, do tego stopnia, że nikt nie oczekuje zakończenia przedsięwzięcia w terminie, przy nieprzekroczonym budżecie i w pełnym przewidzianym dla niego zakresie. Niemniej jednak, fakt ten nie powinien odwracać naszej uwagi od wniosku, że DBR jest nieodpowiednim mechanizmem dla środowisk projektowych, ponieważ założenia DBR nie są tu spełnione. Potrzebna jest inna realizacja praktyczna, która wprost zajmie się kwestią stosunkowo długiego czasu przetwarzania.[18]

[17] A GUIDE TO MAKING EVER FLOURISHING COMPANY - PRODUCTION, DISTRIBUTION, MARKETING AND SALES. Chukei Publishing, 2008. Satoru Murakami, Jun Takahashi, Shotarou Kobayashi, str. 196, str. 207

[18] Goldratt, Eliyahu M., *Critical Chain* , North River Press, 1996. Wyd. polskie: Łańcuch krytyczny, Werbel, 2000, tłum. Ewa Kawczyńska.