

Nick Bostrom

Superintelligens

Vägar, faror, strategier

fri tanke

Innehåll

Sparvarna och ugglan – en oavslutad fabel	7
Förord av Anders Hansen	9
Författarens förord	13
1. Tidigare utveckling och nuvarande kapaciteter	17
2. Vägar till superintelligens	47
3. Former av superintelligens	92
4. En intelligensexplosions kinetik	108
5. Avgörande strategiskt övertag	132
6. Kognitiva superkrafter	150
7. Den superintelligenta viljan	171
8. Är standardutfallet undergång?	187
9. Kontrollproblemet	206
10. Orakel, andar, suveräner, verktyg	232
11. Multipolära scenarier	252
12. Att förvärva värden	292
13. Att välja kriterierna för att välja	327
14. Den strategiska bilden	356
15. Dags att mobilisera	397
Tack	405
Noter	407
Lista över figurer, tabeller och rutor	468
Litteratur	470
Register	509
Utförlig innehållsförteckning	516

KAPITEL I

Tidigare utveckling och nuvarande kapaciteter

Vi börjar med att blicka tillbaka. Historien, sedd i sin allra största skala, verkar uppvisa en följd av tydligt skilda tillväxtlägen, vart och ett mycket snabbare än det föregående. Mot bakgrund av detta mönster har det förmodats att ännu ett (ännu snabbare) tillväxtläge kunde vara möjligt. Men vi lägger ingen större vikt vid denna observation – detta är inte en bok om ”teknologisk acceleration”, ”exponentiell tillväxt” eller de olika idéer som ibland förs samman under rubriken ”singulariteten”. I nästa steg överblickar vi den artificiella intelligensens historia. Därefter går vi igenom vilka kapaciteter som idag uppnåtts på området. Slutligen kastar vi ett öga på några nyare expertenkäter och begrundar vår okunnighet när det gäller tidtabellen för framtida framsteg.

Tillväxtlägen och ”big history”

För bara några miljoner år sedan klättrade våra förfäder fortfarande omkring i de afrikanska trädtopparna. Ur såväl geologiskt som evolutio- närt tidsperspektiv var utvecklingen av *Homo sapiens* från vår senaste gemensamma förfader med de stora aporna en mycket snabb process.

Vi utvecklade upprätt hållning, motsättbara tummar och – framför allt – vissa relativt små förändringar i hjärnstorlek och neurologisk organisation som ledde till ett stort språng i kognitiv förmåga. Till följd av detta kan människor tänka abstrakt, kommunicera komplexa tankar och genom kulturen ackumulera information från generation till generation långt bättre än någon annan art på planeten.

Dessa förmågor tillät människor att utveckla allt effektivare produktionsteknologier, vilka gjorde det möjligt för våra förfäder att migrera långt bort från regnskogen och savannen. Särskilt efter jordbrukets införande och spridning ökade befolkningstätheten, tillsammans med den mänskliga befolkningens totala storlek. Fler människor innebar fler idéer; större täthet innebar att idéer kunde spridas lättare och att en del individer kunde ägna sig åt att utveckla specialiserade färdigheter. Denna utveckling ökade *tillväxttakten* i fråga om ekonomisk produktivitet och teknologisk kapacitet. Senare utveckling, relaterad till den industriella revolutionen, medförde ett andra, jämförbart omslag i tillväxttakt.

Sådana förändringar i tillväxttakt har viktiga konsekvenser. För några hundra tusen år sedan, i den tidiga mänskliga (eller hominida) förhistorien, var tillväxten så långsam att det tog omkring en miljon år för människans produktionskapacitet att öka tillräckligt för att livnära ytterligare en miljon individer på existensminimum. Omkring 5000 f.Kr., efter den agrikulturna revolutionen, hade tillväxttakten ökat till en punkt där motsvarande tillväxt tog bara två århundraden. Idag, efter den industriella revolutionen, växer världsekonomin med i genomsnitt lika mycket var nittionde minut.¹

Redan den nuvarande tillväxttakten kommer att producera imponerande resultat om den upprätthålls under någorlunda lång tid. Om världsekonomin fortsätter att växa i samma takt som den har gjort de senaste femtio åren, kommer världen år 2050 att vara omkring 4,8 gånger rikare än idag, och år 2100 omkring 34 gånger rikare.²

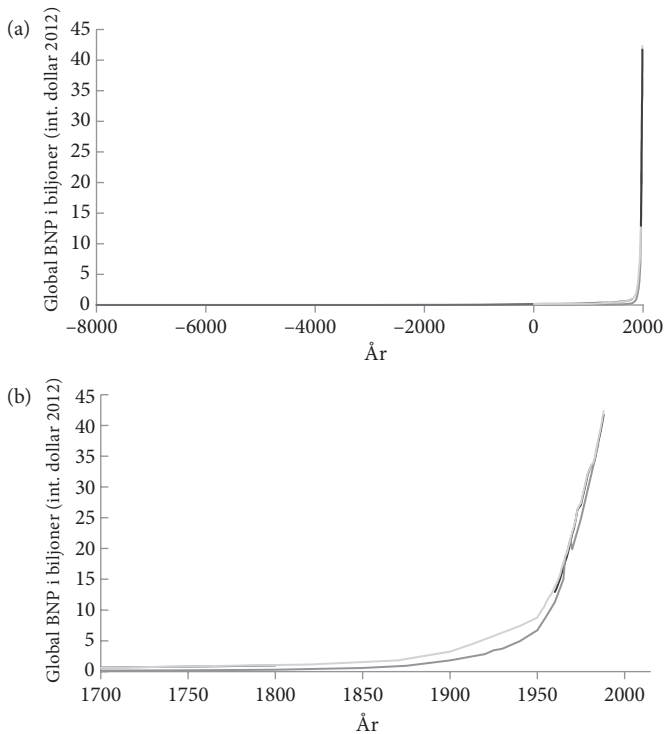
Men tanken på en fortsatt stadig exponentiell tillväxt bleknar i jämförelse med vad som skulle hända om världen genomgick ett nytt

omslag i *tillväxttakt*, jämförbart i storlek med de som den agriskulturella och den industriella revolutionen innebar. Ekonomen Robin Hanson uppskattar, baserat på historiska data om ekonomi och befolkning, att en typisk fördubblingstid för världsekonomin under pleistocens jägar- och samlarkulturer var 224 000 år; för jordbrukssamhället 909 år; och för industrisamhället 6,3 år.³ (I Hansons modell är den nuvarande epoken en kombination av jordbruks- och industrisamhällets tillväxtlägen – världsekonomin som helhet växer ännu inte med en fördubblingstakt på 6,3 år.) Om ännu en sådan övergång till ett annat tillväxtläge skulle inträffa, och om den var av liknande storleksordning som de två tidigare, skulle det leda till ett nytt tillväxtläge där världsekonomin fördubblas i storlek ungefär varannan vecka.

En sådan tillväxttakt framstår som rena fantasin ur dagens perspektiv. Observatörer i äldre epoker kan ha funnit det lika befängt att anta att världsekonomin en dag skulle fördubblas flera gånger inom loppet av en människas livstid. Men detta är det extraordinära tillstånd som vi numera betraktar som ordinärt.

Idén om en kommande teknologisk singularitet har med tiden blivit allmänt bekant, med början i Vernor Vinges inflytelserika essä och vidare genom texter av Ray Kurzweil och andra.⁴ Men termen ”singularitet” har använts på ett oklart sätt, i många olika betydelser, och dragit på sig en ohelig (men samtidigt nästan millenaristisk) aura av konnotationer till teknologiska utopier.⁵ Eftersom de flesta av dessa betydelser och konnotationer är irrelevanta för vårt resonemang, kan vi vinna i klarhet genom att lägga bort ordet ”singularitet” till förmån för en mer precis terminologi.

Den till singulariteten relaterade idé som intresserar oss här är möjligheten av en *intelligensexpllosion*, och särskilt utsikten till maskinell superintelligens. Tillväxtdiagram som de i Figur 1 övertygar kanske vissa om att en ny drastisk förändring i tillväxtläge är på gång, jämförbar med den agriskulturella eller industriella revolutionen. Kanske finner samma personer sedan det svårt att föreställa sig ett scenario där världsekonomin fördubblas på bara några veckor, om det inte samtidigt



Figur 1. Världens BNP i ett långt historiskt perspektiv. När världens ekonomiska historia ritas upp i en linjär skala liknar den en platt linje som smyger sig fram längs x-axeln tills den plötsligt skjuter iväg rakt upp. (a) Även när vi zoomar in på de senaste 10 000 åren förblir mönstret i huvudsak en enda vinkel på 90 grader. (b) Först under de senaste 100 åren eller däromkring lyfter kurvan märkbart över nollnivån. (De olika linjerna i diagrammet motsvarar olika uppsättningar data som ger något olika uppskattningar.⁶)

innebär att medvetanden skapas som är mycket snabbare och effektivare än den välbekanta biologiska sorten. Men för att ta möjligheten av en revolution i maskinintelligens på allvar behöver man inte förlita sig på statistiska kurvor eller extrapoleringar utifrån tidigare ekonomisk tillväxt. Som vi ska se finns det starkare skäl att vara på sin vakt.