

Dagen värd för Sommar i P1 är filosofiprofessor Nick Boström. I dagens program berättar han om den slingriga väg fram till att han blivit en av världens främsta tänkare kring artificiell intelligens och hur teknisk och medicinsk utveckling kommer att påverka mänskligheten. Han har bland annat skrivit boken Superintelligens som översatts till ett trettiotal språk och som starkt påverka den internationella debatten om AI. Dagen program är inspelat i Oxford där Nick Boström bor och där han också byggt upp ett stort forskningsinstitut.

Den bra grejen med en fågleskit är att den är liten, så tänker en filosof. Andra människor som fått en fågelskit på sig tänker ”Det var inte bra”. Jag får en fågelskit på mig och jag tänker ”Det kan ju vara ett lyckotecken. Jag fick en fågelskit på mig man alla de andra fågelskitarna missade ju mig. DET är lycka”. Jag heter Nick Boström och är en filosof. Välkommen till mitt Sommar i P1 härifrån Oxford där jag bor.

*Musik: Signatur Sommar Sommar Sommar med Radiosymfonikerna*

Jag växte upp i Helsingborg, precis intill stranden vid Öresund. Det var en trygg och lycklig barndom men det var en mask i äpplet. En stor, äcklig, rälig mask. Skolan! Jag tyckte jag kunde redan läsa, skriva, räkna långt innan jag började skolan och att vad man ville lära mig där var namnet på en massa älvar. Var det verkligen värt att investera hälften av sin barndom för att lära sig namnen på några älvar och län? Det betvivlade jag. Min strategi i skolan blev att göra så lite som möjligt. När det närmade sig slutet av mellanstadiet och vi skulle upp i 7:e klass och högstadiet fick alla träffa läraren, alla var och en för sig. Tanken var nog att vi skulle få lite uppmuntran och kanske några goda råd och visdomsord inför resten av våra liv. Vi satt där i rummet och då sa läraren till mig ”Du kommer att misslyckas”.

Jag fortsatte att vantrivas även när jag kom upp i högstadiet efter sommarlovet. Men något år senare hände något märkligt. Det var en dag, jag var väl i 8:e klass eller kanske i 9:an, jag gick omkring för mig själv och hade inget speciellt att göra och jag fick för mig att gå till Helsingborgs stadsbibliotek. Varför minns jag inte. I min familj fanns inga akademiker och jag förknippade allt som hade med böcker att göra med skolan, som jag ju avskydde. Hur som helst, jag gick runt där i salarna och plockade lite i hyllorna. Började bläddra i en bok som visade sig vara en antologi med diverse tyska filosofer från 1800-talet. Jag tror ett av namnen var Schopenhauer och Nietzsche var väl kanske med också. I vilket fall som helst lånade jag den och några andra böcker och tog dem med mig hem. Jag tog boken med de tyska 1800-tals filosoferna till mormor och morfar som bodde nära oss. De hade en trädgård och där satte jag mig under ett äppelträd och började läsa. Den värld som öppnade sig där under äppelträdet likande ingenting som jag hade erfarit förut. Jag läste mer och mer i boken och det var som en stor dörr öppnade sig till en värld av idéer och tankar som jag dittills varit helt omedveten om i hela mitt liv.

Under den tiden som följe hittade jag i snabb följd in i filosofi, vetenskap, konst och litteratur. Jag upptäckte att när jag läste texter som betraktades som svåra så verkade de ofta helt klara och begripliga för mig. Jag började läsa dem på originalspråken, vilket var ännu bättre. Jag var i 15-årsåldern men det kändes som om jag fram till den här punkten hade sovit mig genom mitt liv och först nu hade öppnat ögonen. Jag drabbades av en väldigt stark känsla av att jag hade slösat bort min tid fram till det här uppvaknandet och att jag nu inte ville förlora en enda

minut till. Det här blev början på en intellektuell upptäcktsresa. Jag kastade mig med all min energi i mitt eget utbildningsprojekt och försökte bemästra en massa olika områden, allt från att läsa vetenskapliga artiklar till att skriva dikter och skådespel och framförallt så började att tänka. ”Okej, så det finns alltså ett universum som börjat existera och här befinner vi oss då på planeten jorden i Skåne, av alla platser, och vi ser att det finns minst en miljard, miljard andra solsystem där ute i rymden, många mycket äldre än vårt solsystem och den moderna tidsåldern har just börjat efter att vi har varit jägare och samlare ett par hundra tusen år. Och så sitter den en gosse här som heter Nicklas Boström under ett äppleträd och han har ett kranium och i det kraniet finns en hjärna som väger 1,5 kg och allt det som jag ser och hör och tänker är elektriska processer inuti den hjärnan som kommer att dö om några årtionden, eller så. Hm, hm, hm...”. Det hade föds en känsla i mig att tiden ständigt rinner som sand mellan våra fingrar och försvinner. Vi har ett liv. En tidsbegränsad möjlighet att det livet till något som räknas och kanske försöka fatta vad det hela handlar om.

*Musik: Sonata För Violin Solo Nr 2 A-Moll Bwv 1003 med Johann Sebastian Bach*

En av alla grejerna som jag höll på med under den tiden var att måla tavlor. Jag byggde en ateljé inte långt ifrån äppleträdet, i källaren i mormor och morfars hus. Närmare bestämt i tvättstugan. Det blev fler och fler tavlor därnere, oljemålningar och teckningar. Jag vet inte hur många gånger min mormor fick se sina nytvättade lakan hänga där med en skvätt målarfärg, men aldrig att hon klagade. I närheten av deras hus ligger Pålsjöskog som vätte ner mot stranden vid Öresund. Jag fick en idé om att ha en utställning där om våren när löven började spricka ut och marken täcktes av vitsippor. Jag cyklade runt om i stan och satte upp affischer om den här stora konstutställningen som skulle hållas. Jag lånade en trädgårdskärra och en skottkärra av morfar och med en i var hand så rullade ett urval av mina tavlor ut till en glänta i skogen. Jag hängde mina tavlor direkt på träden. Utställningen skulle hålla på i fyra dagar. Dag 1, ösregn! Hjulen på trädgårdskärran och skottkärran gräver ner sig i leran. Dag 2, ösregn! Dag 3, ösregn. Dag 4, bara vanligt regn. Nästa år gjorde jag om det med en ny utställning. Det blev regn igen. Tavlornas träramar vred sig sneda i blötan som de vore plågade själar. Trots allt hade jag lyckats sälja några tavlor. Helsingborgs dagblad kom förbi och gjorde en notis men kort därefter slutade jag att måla. Det fanns annat att lägga den tiden och energi på. Sådant som jag tyckte var viktigare. Om man tar i objektivt beaktande vad världen har mest behov av så är det väl inte så sannolikt att det korrekta svaret är fler oljemålningar.

*Musik: An Die Laute med Franz Schubert*

Under den här tiden brukade jag sätta mig på en cykel efter skolan och köra ut i skogen. Där läste jag mina böcker, skrev dikter och aforismer och skådespel och tänkte på filosofi. Och där upplevde jag höga nivåer av lycka. Ofta blev jag som genomborrad av naturens skönhet. Jag kunde sitta där med blicken fäst och bara stirra på en scen eller en sak som fångade min uppmärksamhet i tio, tjugo eller trettio minuter. Det kunde hända på andra platser också. Ibland var det inte ett fysiskt objekt eller scen utan en filosofisk kontemplation. Det var som om medvetandet fylldes med en idé i den utsträckningen att mitt egna jag tillfälligtvis försvann. En ren observation, en ren kontemplation av en sak i sig, utan hänvisning till dess nytta eller konventionella betydelse. Nu skulle man kanske kunna tro att detta språng in i världen av idéer och litteratur skulle ha gjort att jag blev mer intresserad av skolan men det

blev faktiskt tvärtom. Om skolan tidigare varit tråkig så fann jag den nu som ett helt outhärdligt slöseri med tid. Jag kände att jag kunde utvecklas snabbare på egen hand hemma eller i skogen, än i skolan bakom min bänk. Detta blev ett allt större problem för mig. I gymnasiet hittade jag lösningen på detta problem. Jag fann en väg som annars mest användes av elever som varit på ett utbytes-år. Om man hade lärt sig ett ämne kunde man då examineras i det genom särskilda prov och sedan inte behöva gå till lektionerna. Det visade sig att det inte fanns någon regel att göra detta inom alla ämnen. Jag kunde knappt tro att det var sant. Alla runt omkring tyckte däremot att det var en dålig idé och avrådde mig men jag bestämde mig för att stanna hemma från skolan och sålunda lyckades jag tenta av naturvetarlinjens sista år, ämne efter ämne. Det tog tio veckor vilket jag tyckte bekräftade att skolan verkligen hade varit ett stort slöseri med tid.

Men nu, efter att jag hade blivit fri från gymnasiet dök det upp ett nytt problem. Kallelsen till mönstring inför värnplikten. Det sista jag ville efter att just ha frigjort mig från skolsystemet var att hamna i en annan statlig institution. Jag minns hur jag satt där i väntrummet på mönstringscentret och läste Mclellan Dremel Heart and the PDP research Groups banbrytande Parallel Distributed Processing Volume 2, Explorations In The Microstructures of Cognitive, Psychological and Biological Models. Bredvid mig satt en kille som lekte med sin tändare så att han nära på tände eld på skjortan på en av de andra inkallade. Så, jag gjorde de olika testen av kondition, armstyrka, benstyrka, handgreppsstyrka, syn, hörsel och intelligens. Till slut var det dags att ha ett samtal med den tjänsteman som skulle bedöma hur jag psykiskt skulle fungera i det militära. Han frågade vad jag hade för intressen? ”Jag gillar att sitta ute i skogen på en sten och skriva dikter.” Han nickade och frågade ”Men om det skulle bli krig, vad skulle du göra då?” ”Jaa, jag kan ju kunna skriva dikter om hur hemskt kriget är”, svarade jag. ”Ja, det måste ju någon också göra”, svarade mannen. Och så fick jag min frisedel.

#### *Musik: Kantat Nr 82 med Johann Sebastian Bach*

När friheten till slut var min hade jag i bilden av att det fanns två vägar att gå. Den ena var att bosätta mig i ett skjul i en skog någonstans och fortsätta med det som jag redan höll på med. Som tur var valde jag inte den vägen. Den andra vägen var universitetet. Jag var tveksam efter de tröga åren i skolan men jag bestämde mig i alla fall för att pröva. Det visade sig att universitetsvärlden var något helt annorlunda än skolvärlden så jag dök in i den, med hull och hår. Efter en sväng i Lund flyttade jag Göteborg. Jag studerade matematik, filosofi, matematisk logik och artificiell intelligens samtidigt som jag fortsatte läsa en hel del på egen hand. Känslan att jag inte hade någon tid att förlora bar jag fortfarande med mig. Jag har alltid varit nattuggla och behöver mycket sömn så jag sprang till föreläsningarna på morgonen samtidigt om jag tryckte i mig en ostmacka som jag hade förberett kvällen innan, se ingen tid skulle gå till spillo. Vid den här perioden hade jag ibland en hink kallt vatten i mitt rum. När jag blev trött och okoncentrerad stoppade jag ner huvudet i hinken för att försöka pigga upp mig. Det funkade inte särskilt bra. Hur som helst, jag var rastlös i Göteborg och efter ett år eller så blev jag rekommenderad av en bekant att flytta till Umeå. Det universitetet skulle vara bra. Jag började också studera psykologi där bland annat. Efter två, tre månader blev jag kallad till ett möte på rektorsexpeditionen på psykologinstitutionen. En hög tjänsteman där sa att han upptäckt att jag vid sidan av psykologin samtidigt bedrev tre andra heltidsstudier. Fysik, matematik och filosofi. Han hade missat att jag även studerade logik på distans. Han sa att jag var tvungen att hoppa av alla andra kurser annars skulle han slänga ut mig från

psykologinstitutionen på grund av att det var psykologiskt omöjligt att studera så mycket på en gång. Jag försökte säga ”Du vet, så här långt har det funkat bra”. Jag stod där med intygen som visade sig att jag hade goda betyg i kurserna. Vore det inte bättre att försöka entusiasmera studenter som inte hade tillräcklig motivation än att köra ut de som verkligen hade en hunger att lära sig? ”Nee”. Så jag blev utslängd. Kanske den första som blivit utslängd från ett utbildningsetablissemang för att ha pluggat för mycket.

I vilket fall som helst så gjorde det inte så mycket. Det var på höstterminen och jag hade märka att varje vecka i Umeå var kallare och mörkare än den föregående och jag kunde se var det skulle barka. Jag var glad att återvända till Göteborg. Under hela den här perioden som varade en 6-7 år hade jag en känsla i mig av djup, djup ensamhet och isolering. I hela mitt liv hade jag aldrig träffat en enda person som delade mina intressen. Jag kände mig som en främmande art som på nåt sätt hade råkat landa på den här planeten. Trots denna känsla av isolering, eller kanske på grund av den, så bar jag också på en känsla av djupt medlidande för allt det som människor och för den del djur måste uthärda i sina liv. Allas vår utsatthet gentemot de mörka krafterna i universum och tillvaron, mot vilka vi än i dag mestadels är försvarslösa, hur mycket modern teknik vi än har. De här tankarna kanske kunde spåras tillbaka till allt jag hade läst av de tyska filosoferna och den svenska diktaren och biskopen Esias Tegnér. Faktum var att jag hade läst så mycket Tegnér att andra svenska klassiker att jag började prata gammal svensk grammatik med verb i pluralformer. ”Grodorna äro lustiga att se, ty de hava inga svansar”. Att tala på det sättet måste nog betraktas som en av de minst goda idéerna jag har fått. Även om jag själv tyckte att det lät väldigt fint på den tiden. Det bidrog i alla fall knappast till att minska min isolering. Märkligt nog var det aldrig någon som kommenterade hur jag talade. Folk antog väl att jag pratade med någon extra knäpp skånsk dialekt. Och sedan dess, även om jag nu har slutat tala med gammal svensk dialekt, så har min svenska försämrats. Jag har ju bott utomland så länge nu att jag nästan förlorat förmågan att tala modersmålet, så det är därför det låter lite konstigt. Så jag får be er ha lite patience (engelska-tålamod) med detta.

*Musik: Se´N Har Jag Er Frågat mera med Jean Sibelius*

Efter ett år eller så flyttade jag till Stockholm där jag studerade teoretisk fysik och filosofi. Vid den här tiden höll jag på att genomgå en ny transformation. En nästa lika djupgående som den som jag hade i mina tonår när jag upptäckte idéernas värld. Man kanske kan beskriva den som en övergång från en romantisk till en modern epok. Jag tappade intresset för 1800-tals filosoferna och för att läsa klassiker och för den forna tidens hjältemod. Jag blev mer intresserad av analytisk filosofi och modern vetenskap och potentialen i teknologi. Men man kan säga att i klivet från klassiskt till nutid så missade jag målet och flög över det och blev mer intresserad av framtiden istället. Det blev ett enormt skifte i mitt fokus. Jag kände mig färdig med att skriva dikter och sådana saker. Den sista dikten jag skrev hette Rekviem och den är faktiskt ett försök att beskriva den här tankemässiga omvandlingen. Det handlar om en kavallerigeneral som har försovit sig och som vaknar upp och ser sitt läger övergivet. Den tar oss på en komprimerad resa från en gammal heroisk era till en helt modern, eller till och med postmodern era. Jag ska läsa den för er nu. Om ni behöver att gå till tvättstugan eller fixa något annat så är detta ett riktigt bra tillfälle.

När tidens tupp för tredje gång gal, ur hjälteslummern väcks vår general. Han står upp. Det är försent. Lägret ligger övergivet. Hans trupp har dragit av färde. Det är kallt. Gräset hans stövlar träder på är stelt av den första frosten. De djupt insjunkna ögonen vittnar om utståendets lidande. Anletsdragen som åren utmejslat bär prägel av ansvarsfullhet. Han vandrar stilla runt bland tälten med manteln över axlarna, tigande. Det är ingen där. Han verkar överlägga något. Sen går han raskt fram till sitt riddjur, sätter foten i stigbygeln och svingar sig upp. Han smackar åt hästen och rider iväg utan ord. Sporra ditt riddjur. Kanske det inte är försent ännu att samla styrkorna, att i spetsen för en här segra för humaniteten. (smack, smack) Hoppa över röset, plaskar genom man, rider under lundarnas bokar och ut över fälten där morgondimman ligger drömmande kvar. Och förbi kärret där ljudsalvorna från en hackspetts knackande hamrar bort över ytplanet. Hans vilja tvingar fram oanade krafter ur hästens väsen. Han tar sig upp på åsen. Piskrappen går som elektricitet genom nävarna. Får den svettglänsande muskulaturen att uträtta arbete och sporrarnas blodade järngaddar tillhandahåller en smärta som stegras till det outhärdliga om steget mattas det minsta. Och fradga bildas från djurets gipa och dess ögon är vilda, näsborrarna är vilda. Då hörs i den galna galoppen ett dån, större än åska. Hjälten griper hårt om värjskaftet och vänder trotsigt blicken uppåt. Makten musik och himlens trafik, ur skyn skjuter ett jaktplan. Men bakom gudens styrpanel sitter en människoapa. Fåglarna flyr i samtliga väderstreck som om denna örnen jagade dem. Ljusstrålar som reflekteras av föremålet bryts samman av linsen och fokuseras på näthinnan där fotoreceptorer stimuleras att avge aktionspotential som fortplantas utefter nervus opticus och projiceras på synbarken som sitter här och framför den sitter själen upp och ner vänd i tv-fåtöljen och glori och käkar popcorn. Wow! Vad det är spännande! Emellertid avlägsnar sig det snabba flygplanet snart. För krigsman står vid sin fåle och följer med blicken, pricken, över horisonten som så småningom försvinner helt i det blå. Reformatorn står nersänkt i renen och passerar ovationer som han själv har installerat. Ångan flåsar ur deras munnar, som ur två föråldrade lokomotiv. Stannar tiden om vi står alldeles stilla? Perplex är den vise, se det blir hans häst som skjuter fram propositionen så omdefiniera våra ideal. Hä! Även du min Brunte, även jag, är alltså min art dömd till undergång som din till att äta hö. Slutligen ger sig dock saken av sig själv, falla ner från himmelen som en rostad duva på tallriken. Sticka vapnet i skidan, bege sig hem, äta frukost, läsa tidningen. Brigadgeneralen som städ sig gott mot faran utan att blinka, blundar inte nu för att hans yrkesverksamma liv är över. Han kan under alla omständigheter påräkna pension från staten. Kanske är det bäst så? Han till och med visslar en melodi. Då kommer det fram en råtta, vit med röda ögon och säger "Vad är Mozarts musik mot stimulering av nucleus accumbens?". Befälet sparkar undan råttan. Han överlämnar kommandot åt högre existensformer. Optimerade samhällssystem, artificiella nevrala nätverk och medicinerade övermänniskor. Luxiterna... (säger något på Latin)

*Musik: Acroyear 2 med Autechre*

Ibland när man är utomhus då känner man en droppe landa på hjässan. I alla fall om man är lite flintskallig som jag blev på gamla åren. Då hoppas man att det börjar regna för alternativet, om det inte var en regndroppe, är ju fågelskit. I januari 1996 hade jag kommit till London. Först var det tänka att jag skulle bara vara där på en utbytestermi på Kings collage där jag studerade Competition Neuroscience men när det närmade sig slutet kände jag att jag ville stanna i den här stimulerande miljön och fortsätta med min doktorsutbildning där. Frågan var i vad? Jag var intresserad av teoretisk fysik och av hjärnvetenskap men jag tänkte att om

jag gjorde en doktorsavhandling i ett av de ämnena skulle jag till slut veta väldigt mycket av något väldigt avgränsat, om toppkvarken eller någon region i hypo-campus. Istället bestämde jag mig för att göra min doktorsutbildning i vetenskapsfilosofi och sannolikhetsteorins grundvalar. På så sätt skulle jag få mer av en överblick över vetenskapens stora byggnation och jag skulle kunna plocka det mest intressanta delar från olika områden. Vid den här tiden hade jag blivit lite mer social, en smula mer mogen. Jag studerade och jobbade fortfarande väldigt hårt men jag sprang inte längre till förlängarna med en macka i munnen.

London, i mitten av 90-talet var som sagt en spännande plats. Det var åren under den korta perioden av Cool Britannia med britpop och en blomstrande club-scen för housemusik och ny konst, och allt vad det var. I princip visste folk två saker om Sverige. ABBA och Ingmar Bergman. På varenda fest i mitten 90-talet spelades alltid Dancing Queen och kanske svävade Bergmans ande över den turistbok om Sverige som jag sprang över vid ett tillfälle. Bland fraser som "Vad är klockan?" och "Hur kommer man till järnvägsstationen?" fanns där också frasen "Tänk på döden!" som en av de nödvändiga kunskaperna för att kunna turista i Sverige. Vid den här tiden blev jag också mer involverad i aktiviteter utanför schemat, i dramakurser och debatt-tävlingar. Jag gjorde också stan-up comedy. En scen och genre som var högst livaktig på Londons pubar och teatrar. En sak som jag gillade med stand-up var ärligheten och direktheten i feedbacken från publiken. I stand-up är det ju helt enkelt så att om publiken skrattar så är du bra. om publiken inte skrattar är det skit. Ingen tvetydighet. Det är en stark kontrast mot hur det är som filosof där man kan hålla på i årtal och kanske i ett helt liv utan att få ett besked om man är bra eller inte.

*Musik: What's He Building med Tom Waits, Ralph Carney, M Mark Reitman, Jeff Sloan*

Klivet in i Londonvärlden sammanföll också för mig med klivet ut på Internet. Jag är nämligen så fruktansvärt gammal att jag växte upp utan Internet. Men när jag anlände till London år 1996 kom jag äntligen ut på nätet och för mig betydde det mycket. På nätet upptäckte jag att det fanns andra människor i världen som var intresserad av de frågor jag hade funderat på under många år. Jag hade länge haft en bild av att människans förutsättningar en dag i grunden skulle förändras som en konsekvens av utvecklingen inom artificiell intelligens. Eller att biomedicinska metoder skulle kunna höja människans kognitiva kapacitet eller kunna manipulera mänskliga hjärnor för att ändra sinnesstämning. Men jag hade aldrig träffat någon annan som varit alls intresserad av sådana här frågor. På nätet hittade jag snabbt en community som ibland kallad sig tranhumanister eller extropianer, som också förde diskussioner kring de här frågorna. Att jag lyckades ramla in i det här var faktiskt resultatet av arbetet av en annan svensk som heter Anders Sandberg och som idag är en forskningskollega på mitt institut. Anders hade då en stor webbsida med en massa länkar som jag hittade och började följa och blev snabbt involverad i diskussionsgrupper och mejlinglistor. Det var en ganska unik intellektuell miljö på den tiden. Det var visserligen högljutt och mycket trams och galna idéer men det var också en del rätt intressanta idéer. Det folks om deltog var ofta utanför mittfåran, men det fanns en del akademiker och en del som jobbade som mjukvaroringenjörer. Men också människor som levde alla möjliga olika sorters liv och som ägnade sig åt att tänka hur AI, nanoteknologi, rymdteknik, virtual reality, humans enhancement medicin, life extension och så vidare kunde komma på att påverka den mänskliga tillvaron i framtiden. Vid den här tiden var det i praktiken det enda forum där sådana diskussioner hölls. Men min bild var att de här informella konversationerna på Internet

var visserligen betydligt bättre än ingenting men också betydligt sämre än vad det borde vara. Det fanns två stora begränsningar som jag såg det. För det första var det framförallt ett tidsfördriv för amatörer som diskuterade de här viktiga frågorna på sin fritid. Ur mitt perspektiv förtjänade de här frågorna samma seriösa och systematiska tålamod och uppmärksamhet som per automatik ges av forskare kring alla sorters mindre viktiga områden. För några år sedan genomförde vi på mitt institut en studie som visade att gjordes tio gånger mer forskning om dyngbaggar än om existentiella risker och mänsklighetens framtid. Så det behöver ändras. Hela samtalet behövde lyftas upp till en nivå för seriös akademisk forskning.

Det andra problemet med den här tidiga varianten av transhumanismen var att den allt för ofta förföll till att bli en slags enfaldig hejarklack för teknikutvecklingen. De gjorde det slappa antagandet att teknik av sig själv helt säkert kommer vara en kraft för det goda i världen. I mina ögon framstod det som ett begränsat perspektiv. Min hållning var ju att jaa, det är rimligt att teknologin vid någon punkt i framtiden kommer att fundamentalt förändra mänsklighetens natur och förutsättningar, även om det råder osäkerhet när i tiden detta kommer att ske. Och jaa, det finns också en enorm uppsida och potential att skapa någon som är långt mer underbart och humant än den värld vi ser runt oss idag. Inte bara genom att styra saker i omvärlden utan också genom att förändra och uppgradera människans natur. Men på samma gång fanns det ingen garanti för att resultaten av att öka våra tekniska styrkor skulle bli bra. Det vore kanske lika sannolikt att de nya förmågorna skulle komma användas för destruktiva ändamål och orsaka katastrofer med effekter som är oåterkalleliga för människan och planeten. I min forskning har jag kallat det för existentiella risker. Arbetet borde ligga på att förstå de här olika sätten teknologikutvecklingen kan förändra världen och hitta de punkter där man med lite god vilja kan öka sannolikheten att utvecklingen blir gynnsam för människan.

För att göra det möjligt att tänka på det här andra sättet grundade jag The World Humanist Association tillsammans med en bekant, David Pierce, kring 1997. Det var en internationell gräsrotsorganisation för att underlätta samtal om de här sakerna och för att försöka göra det mer akademiskt respekterat. WHA växte snabbt till flera tusen medlemmar och organiserade konferenser och hade en vetenskaplig tidskrift. Och jag tror den spelade en betydelsefull roll under några år för det var en forum där de här diskussionerna kunde utvecklas. Men själva har jag sen länge lämnat transhumanismen. Redan i början av 2000-talet hade vi lyckats få de här frågorna att flyta in i den akademiska världen. Förbättring av människan diskuterades som en del av bioetik till exempel och en annan ännu mer intressant gren ledde till arbetet som nu har fortsatt av till exempel The Future of Humanity Institute som jag ska prata mer om straxt.

### *Musik: Simulation Argument med Pasztor*

Efter mitt doktorsarbete i London flyttade jag till USA och till Yale University. Under min tid på Yale publicerade jag en vetenskaplig artikel som introducerade simuleringsargumentet. Jag ska inte ge mig in i något slag föredra här utan läser bara sammanfattningen av artikeln. Titeln var *Lever vi i en datorsimulering?*

*Den här artikeln argumenterar för att minst ett av de här tre antaganden är sanna.*

- 1. Människoarten kommer mest sannolikt döta ut innan ett postmänskligt tillstånd har uppnått*
- 2. En postmänsklig civilisation kommer sannolikt inte att bedriva ett betydande antal simuleringar av den mänskliga utvecklingens historia eller varianter av det*

3. Vi lever med all sannolikhet i en datorsimulering. Av det följer att tron att det finns en chans att vi kommer bli post-humans som kan skapa simuleringar av våra förfäder är falsk, om vi inte redan lever i en simulering. Ett antal konsekvenser diskuteras också här.

Simuleringsargumentet fick en stor påverkan när det först publicerades i Philosophy Quarterly och har fortsatt ha en stor påverkan ända sedan dess. Det verkar som om uppmärksamheten kring den går i vågor när någon ny community får syn på den för första gången.

Simuleringsargumentet har även dykt upp i populärkultur. Tesslas grundare Elon Musk twittrar återkommande om det och det finns ett flertal filmer och noveller, musik och en teaterproduktion på Broadway som har direkt inspirerats av det. Just nu är det tydligen någon som håller på och skriva ihop en balett om den här akademiska artikeln i volym 53, nummer 211 i Philosophy Quarterly 2003.

*Musik: Simulation Argument med Pasztor*

2003 kom jag tillbaka till Europa, till Oxford University. Efter ett par år fick jag möjligheten att grunda The Future of Humanity Institute. Institutet kom ursprungligen till som en konsekvens en stor donation till universitetet från filantropen Dr James Martin. Han ville stötta tvärvetenskapligt arbete med inriktning på mänsklighetens utmaningar under 2000-talet. FHI startades som ett av tio institut i den här satsningen inom universitetet i Oxford. Ja namnet, The Future of Humanity Institute, låter ganska bombastiskt men det är faktiskt också ett ganska relevant namn. Det beskriver de frågor vi jobbar med. Den stora bilden, de stora frågorna som rör mänskliga civilisationen och den mänskliga artens överlevnad i en framtid där teknologin i grunden kan förändra alla parametrar och förutsättningar för det liv vi lever. På institutet försöker vi forma koncept och skapa intellektuella verktyg som hjälper oss att tänka systematiskt kring de här frågorna och de stora linjerna är; Vilka möjligheter som finns för den som vill försöka göra världen bättre? Hur kan man koppla långsiktiga resultat för mänskligheten till konkreta handlingar och projekt idag? Namnet är också tillräckligt brett för att ge oss ett stor manöverutrymme.

En av de saker som har uppstått ur arbetet på institutet är boken *Superintelligens – vägar, faror, strategier* som jag gav ut 2014 och den kom ut på svenska 2017. Boken tog mig 6 år av intensivt arbete att skriva. Inledningsvis var det faktiskt tänkt att det skulle bli en bok om existentiella risker i allmänhet och att ett kapitel bara skulle handla om artificiell intelligens. Men när jag började skriva på det kapitlet så började det växa och växa och växa. Det var som om det var en bok som försökte få sig själv skriven, men den handlade om AI, artificiell intelligens i framtiden. Så jag bytte fokus och skrev hela boken om det. Märkligt nog, trots att boken var skriven med ett helt akademiskt perspektiv med ett ton fotnötter och en rätt tung prosa så blev den ändå end New York Times bestseller. Det handlar bland annat om vad som händer om och när vi människor lyckas skapa generell artificiell intelligens. Alltså inte bara AI som är bättre än människor på vissa speciella saker utan som överträffar den mänskliga hjärnan i allmän problemlösningsförmåga och inlärningsförmåga. Med andra ord vad som händer när vi får superintelligens. Med superintelligensens framväxt står vi inför ett betydande problem. Hur kan man programmera superintelligenta intellekt på ett sådant sätt att de gör de vi vill att de ska göra, att de är säkra och att de har bra effekt för mänskligheten?

När boken gavs ut var detta ett väldigt förbisett område. Men under de senaste åren, kanske delvis som en konsekvens av boken så har det förändrats. Numera finns det ett aktivt

forskningsfält som försöker utveckla metoder för att man ska kunna kontrollera AI och som då ska kunna fortsätta att fungera även om artificiell intelligens i framtiden uppnår övermänniskliga nivåer av tankeförmåga. På FHI har vi en grupp som forskar om detta just idag och som samarbetar nära med The Deep Mind och andra ledande aktörer inom AI området.

*Musik: River med Joni Mitchell*

Sedan starten har FHI växt. Vi närmar oss nu hundra medarbetare. När man går igenom FHI idag så ser det vid första anblick väldigt mycket ut som ett vanligt akademiskt forskningscenter; kontor, mötesrum, Whiteboards med matematiska symboler. Men om du hänger kvar en stund så tror jag du märker en särskild stämning, en påtaglig känsla av syfte. Oavsett tid på dygnet eller vilken veckodag det är så är det nästan alltid någon där och praktiskt taget varenda en som jobbar där tror jag inte går till jobbet i första hand för att få en lön, utan för att de känner att vad de gör verkligen har betydelse. Det är en väldigt speciell miljö.

Vi har ett antal olika forskningsgrupper. En fokuserar på artificiell intelligens och att utarbeta kontrollmetoder som kan funka även om det blir superintelligens i framtiden. En annan grupp, statsvetare och sådär, studerar hur samhället skulle kunna fungera om har AI och som kan göra alla möjliga grejer som människorna gör. Sedan har vi en grupp som arbetar med bioteknologi och biosäkerhet, makrostrategin är ett fortsatt stort område. Jag själv för närvarande håller på att tänka på den moraliska statusen av digitala intellekt. Om man tänker att vi utvecklar artificiell intelligens kanske i nära framtiden som har motsvarande kapacitet som djur, som en mus eller en katt, där vi tror djur har en viss moralisk status. Eh, man kan inte göra...man kan inte sparka sin hund. Så hur ska vi då tänka på det om våra dataprogram kanske får medvetande eller förmåga att känna smärta och vad innebär det moraliskt sätt?

The Future of Humanity Institute är bara en liten del av ett större globalt nätverk som vuxit fram ur det vi och andra har gjort. En annan (?) i detta nätverk är The Future of Life Institute som skapats av en annan svensk, Max Tegmark i Bostonregionen. Det har faktiskt varit överraskande många svenskar involverade i allt det här. Det fanns en tid tidigt i FHI:s historia, vi var fortfarande väldigt små enbart en handfull anställda och av en händelse var det då tre av oss som var svenskar. Min stora rädsla vid den tiden var att när vi utlyst ett nytt forskningsjobb skulle det bli så att den mest kvalificerade personen var en annan svensk, för då skulle jag ju ha tvingats anställa den personen och till omvärlden skulle de ju då ses som oerhört nepotistiskt eftersom det vore statistiskt osannolikt att av en slump skulle bli så många från mitt hemland. Som tur är har proportionerna sedan dess jämnats ut.

Många av de som är inblandade i det här är i sig själva märkvärdiga personer, till exempel en kille som var här och hjälpte FHI för något år sedan var så hängiven till målet att göra så mycket nytta till världen som möjligt att han inte bara hade skänkt alla sina pengar till välgörande ändamål utan också donerat sin ena njure. En annan helt underbar person var en student som försökte donera så mycket som möjligt av sin inkomst till välgörenhet. Så för att spara ihop mer pengar bodde han i ett tält på universitetsområdet. Han har sedan dess haft en kometkarriär inom den amerikanska försvarsbyråkratin där han så småningom hoppas kunna knuffa saker i en mer humanistisk och mänsklig riktning. Intellectuellt sett så är FHI också en speciell plats. Forskarna kommer från en mängd olika områden. Vi har ett flertal matematiker, datorforskare, AI-forskare samt några statsvetare, ingenjörer och filosofer. Flera

av dem är mångkunniga och fogar samman flera olika områden. För mig är det ett stort privilegium att komma till jobbet dag efter dag.

När jag var yngre brukade jag ofta läsa medan jag spatserade. En av mina forskare berättade för mig att när han var liten så brukade han läsa en bok medan han simmade. Jag känner mig inte längre som en utomjordning som har kraschlandat på den här planeten. Eller i vilket fall så verkar det som det var några andra utomjordingar som också överlevde kraschen.

*Musik: All is Full Of Love med Björk Gudmundsdottir*

Det börjar närma sig slutet av programmet och detta är väl en bra plats att rikta några tack till viktiga personer. Jag skulle vilja tacka mina föräldrar för att de hela tiden hundra procentigt stöttat mig under den här knasiga resan. Jag skulle vilja tacka alla kollegor som jag lärt mig så mycket ifrån och min fru för allt det som hon har gjort, och vår lilla kille som är en sån liten underbar varelse. Ja, om man tänker på det skulle man kunna utvidga det här tackandet till hundratals tidigare generationer och till hundratals miljoner kvinnor och män vars hårda arbete, mod, självupppoffring och kreativa sätt att lösa problem har byggt vår moderna civilisation vars välstånd och frihet saknar motstycke. Vi har kommit så långt och vi behöver kanske bara bära stafettpinnen en liten bit till innan vi lämnar över den till artificiell intelligens eller en uppgraderad version av mänskligheten. Vi behöver verkligen all hjälp vi kan få.

Jag ska avsluta det här programmet med att utsätta er för ytterligare en dikt. Den här tar oss tillbaka till den tid när jag var student i Stockholm. Det var en vinter under en kall natt. Jag hade suttit på Fysikum och analyserat experimentresultat till sent och jag tog tunnelbanan hem till mitt studentrum uti Solna. Jag minns att jag promenerade från stationen genom en skogsdunge. Det snöade. Det var så tyst, snön dämpade alla ljud och dessa stora snöflingor föll tungt och rofyllt ner från natthimlen. Och ännu en gång blev jag genomborrad och stod där i natten, ensam i snön och bara tog in skönheten och storslagenheten i det. Och när jag till slut fortsatt vidare, som i en trans, så skrev jag det här

*Stilla faller snön, på gatorna och torgen och på den frusna sjön och på glädjen och på sorgen. Singlar sakta ner på svalnade förälskelser. Läger sig på de eldiga lockar jag till slut på det flammande hjärtat, det vita flockar. Och varje fotspår suddas ut och alla ojämnheter täckas över. Vi hör vid våra dagars slut hur dödens vaggsång barnet sover. Men vaken stund och se flingorna falla som människosjälar ur kosmos det kalla. Det snöar på järnek, lav och en. Snart snöar det även på din gravsten.*

Jag heter Nick Boström. Tack för att ni lyssnat.