

Innehåll

1. SLAVERI OCH AUTONOMI 7

Autonomi och ansvar 19

Syntetiska känslor 22

Relationen mellan människa och robot 29

21 nödvändiga gram för autonomi? 33

Problemet med andra medvetanden 41

Slaveri, autonomi och ansvar 47

2. KONSTGJORD MORAL 51

Metaetik: finns det en objektiv moralisk sanning? 61

Normativa moraliska teorier 68

Olika idéer om hur moral kan byggas in i robotar 84

3. STOPPA MÖRDARROBOTARNA! 91

»Det kommer alltid att finnas en människa i beslutsloopen« 96

Hackning och störningar 98

UAV:er – drönare 99

Argument kopplade till jus ad bellum 103

Argument kopplade till jus in bello 106

Den moraliska grunden för krigets lagar 109

Varför skulle en robot döda? Rättfärdigandet att döda i krig 116

4. ROBOTAR, POLITIK OCH MENINGEN MED LIVET 123

Områden där robotar är överlägsna 131

Områden där människan är överlägsen 136

Prognosen 140

Kan vi stoppa utvecklingen? 154

5. SCIENCE FICTION, SINGULARITET OCH COPETECH 157

NOTER 171

1. SLAVERI OCH AUTONOMI

NOTHING IS STRANGER *to man than his own image*. Ingenting är så märkligt för människan som att se sin egen avbild. Orden kommer från den tjeckiske författaren Karel Čapek i den text från 1921 där termen *robot* användes för första gången. Det tjeckiska ordet *robot* betyder i vid mening slavarbetare, och det är vad robotarna har varit för oss, där de år efter år har utfört tunga och monotona arbetsuppgifter inom industrin. Tysta, medgörliga slavar, som arbetar dygnet runt utan att ställa krav. Men är det en bild av oss själva vi ser, när vi tittar på dessa konstgjorda slavar? Ser vi en variant av oss själva, eller ser vi något väsensskilt? Väcker synen av den tysta, ickeorganiska slaven några känslor inom oss, och påverkas i så fall dessa känslor av robotens utseende?

Slaveri tycks vara intimt sammankopplat med människan och följer henne genom historien. Det finns skriftliga belägg för slaveri i några av våra äldsta bevarade skrifter. Slaveri nämns i mer än 4000 år gamla kilskriftstavlor från Mesopotamien, samt i Gamla testamentet. Historiskt sett har majoriteten av alla kulturer tillämpat slaveri.

Slaveriet speglar kanske de allra mörkaste sidorna i den mänskliga naturen. Förslavandet av andra människor visar på vad som verkar vara en stark och inneboende önskan, eller åtminstone för många, att dominera och styra, inte bara över materiella ting utan även djur och människor, parad med förmågan att avhumanisera och tänka på »den andre« som någonting annat än just en mänsklig. »Människan har utvecklats«, kan man invända. »Slaveriet hör

oupplysta tider till, när man trodde att världen var platt och att Tor slog med sin hammare så det åskade.« Men slaveri har funnits – och finns – ända in i modern tid. Det avskaffades officiellt i USA under 1800-talet, och i vissa delar av Afrika och Asien först på 1900-talet. I en del fundamentalistiska muslimska länder finns det fortfarande kvar. Ett exempel är Mauretanien, där det fortfarande är möjligt att födas in i slaveri. Dagens slaveri finns dock närmare än vad vi vill tro. Skrämmande exempel på detta är människohandel – trafficking – inom prostitution och barnarbete.

Slaveriets kanske främsta drivkraft är emellertid inte önskan att dominera för dominerandets skull, utan att låta någon annan utföra de arbetsuppgifter som är farliga, tunga, enahanda eller på andra sätt oattraktiva för de flesta människor. Och det är där robotarna kommer in. En grundsten i Immanuel Kants filosofi – som vi kommer att se lite närmare på längre fram i denna bok – är att man aldrig får använda en annan människa som *medel*, bara som *mål*. Det betyder att man inte ska utnyttja andra för sina egna syften; man måste alltid se mänskligheten i den andre, se sig själv i sina medmänniskor. »Humanitetsformeln«, som Kant kallade den. Med robotar som slavar – det vill säga *artificiella* slavar – är det ingen risk att vi utnyttjar våra medmänniskor. Eller är det?

Slaveriets motpol är frihet och självständighet. Den förslavade önskar ständigt – åtminstone på något plan – att bli fri. Fri att få bestämma över sig själv och sitt liv, i stort som smått. Åtminstone den slav som inte blivit fullständigt nedbruten eller hjärntvättad, men ett sådant tillstånd är inget skäl för att andra inte ska träda in och hjälpa honom ur hans belägenhet. Försöken att förhindra slaveri är också mycket gamla. Det äldsta belägget för förbud mot slaveri finns troligen i den persiske kungen Kyros II:s dekret till babylonierna. Och i Sverige förbjöds träldomen – åtminstone delvis – i 1337 års Skara stadga. Frihet – i betydelsen grundläggande

personlig frihet – kan nog sägas vara en av de allra mest grundläggande rättigheterna.

Men robotarna, våra konstgjorda slavar, drömmer inte om frihet och självständighet – de drömmer inte om något alls. Det är därför de är så exemplariska slavar. De kan varken lida eller ha någon egen vilja, och ingen moralisk skugga kan därmed falla över »slavägaren« eftersom en robot inte kan berövas något, åtminstone inte i mänsklig mening. När nu mänskligheten går framåt och blir mer moralisk, mer civiliserad – studier visar till exempel att mänskligheten blir mindre våldsam, vilket skulle kunna vara ett tecken på att vi faktiskt blir alltmer civiliserade¹ – allt fler får det bättre ekonomiskt, och mänskliga rättigheter sprids över världen, så kommer ingen människa mer att förslavas. Vi har ju robotarna, som aldrig kommer att klaga eller kräva frihet och självständighet i något som helst avseende. Man kan dock fråga sig om detta verkligen stämmer.

Det paradoxala är att när vi nu har möjlighet att ha konstgjorda, artificiella slavar, så går den tekniska utvecklingen mot att ge dem ökad självständighet – ökad frihet, skulle man kunna säga. Denna forskning pågår framför allt inom det militära och handlar om att robotarna där ska kunna fatta vissa beslut på egen hand. Just det – inom det militära, där robotarna faktiskt i många fall har möjlighet att *döda* oss.

När man talar om robotars självständighet används ofta termen *autonomi*. Termen används även om människor, och då handlar det vanligtvis om att vara fri – att vara fri från yttre tvång, men också att ha en fri vilja. Det är inte helt lätt att slå fast vad det innebär att ha en fri vilja, även om alla har en vardaglig förståelse av vad det innebär. Det handlar om att inte vara styrd av någon annan, att fatta beslut på egen hand. Det kan också handla om att man ska – eller bör, för att verkligen anses vara autonom – se igenom propaganda, stå emot försäljningsknep och diverse mer eller mindre

subtila övertalningsförsök. Det kan även handla om att ha förmåga att lägga sina mest grundläggande mänskliga drifter och behov åt sidan, och ståndaktigt välja äpplet istället för chokladkakan när man går på diet. När man talar om autonomi i sådana sammanhang och fördjupar sig i dess betydelse glider man snart in i filosofiska diskussioner kring fri vilja och vilka kriterier som ska vara uppfyllda för att man verkligen ska vara fri och till exempel ska kunna hållas ansvarig för ens handlingar, moraliskt och juridiskt. Här går uppfattningarna hos de akademiska filosoferna isär, vilket vi ska undersöka längre fram i denna bok.

När man däremot talar om robotar och autonomi, så verkar autonomi vara något som kommer i grader. Det vill säga, det är inte antingen eller, svart eller vitt, huruvida någon – eller något – är autonom. Man talar ju till exempel om autonoma gräsklippare, dammsugare och bilar. Det är värt att nämna att Volvo ligger i framkant när det gäller autonoma bilar, som bland annat ska klara att köra i bilköer i upp till 50 kilometer i timmen. Volvo har som mål att ha 100 självkörande bilar i Göteborg 2017.² Även Google har tillverkat en självkörande bil utan ratt eller pedaler för gas och broms.³

Allt sköts med hjälp av sensorer, aktuatorer och mjukvara i bilen. Målet är att få ner antalet trafikolyckor genom trötta och stressade förare och potentiella rattfyllerister, som då inte längre behöver köra, utan bara kan sätta sig i den autonoma bilen. Regeringen i Storbritannien har meddelat att förarlösa bilar kommer att tillåtas köra på allmänna vägar i landet från och med 2015, och tekniktester kommer att börja göras i början av året.⁴ Nissans självkörande bil var den första autonoma bil som körde på allmän väg i Japan, vilket skedde i slutet av 2013.⁵ Flera stora biltillverkare håller på att utveckla helt eller delvis autonoma bilar och i USA har flera självkörande bilar redan godkänts.



Google testkör en Toyota Prius som modifierats för att kunna köras utan förare.

Angående de här bilarna har vi kvar en grundläggande betydelse av termen autonomi – att maskinen har ett visst mått av självständighet – men man inser samtidigt att det är en väldigt vid term som kan innefatta mycket. Även om bilen klarar av att navigera i bilköer och kan köra undan och stanna om något händer, kommer det troligen att dröja innan man kan skicka iväg den för att hämta barnen i skolan, medan man själv är hemma och gör i ordning middagen.

Idag bestäms robotarnas autonomi framför allt av vilken relation de har till människan. Man talar till exempel om att en robot är *fjärrstyrd*, *halvautonom* eller *fullt autonom*. Alan Winfield – professor, robotforskare och författare till boken *Kort om robotik* – skriver att när robotforskare talar om autonoma robotar menar de vanligtvis robotar som bestämmer vad de ska göra närmast helt utan mänsklig inblandning eller styrning.⁶ Han menar att vi måste vara försiktiga här eftersom de inte talar om »verklig«, »mänsklig« autonomi och att fullt autonom snarare ska förstås som *kontrollautonomi*, vilket

betyder att roboten kan utföra sin uppgift eller sitt uppdrag utan mänsklig inblandning, men att detta uppdrag fortfarande är programmerat eller övergripande styrt av en människa.

Men en fullt autonom robot är, precis som Winfield antyder, inte detsamma som en fullt autonom människa – det ställs olika krav på vad som menas med att vara »fullt autonom« när det gäller robotar och när det gäller människor. För robotars del handlar autonomi alltså i själva verket om var människan befinner sig i förhållande till den mer eller mindre autonoma roboten. En fjärrstyrd robot styrs helt och hållet av människan, ungefär som man styr en radiostyrd leksaksbil eller helikopter. De obemannade flygande farkoster som används av militären – UAV:er (*unmanned aerial vehicles*) eller drönare som de oftast kallas i vardagligt tal – är fjärrstyrda, men har idag viss autonomi när det gäller exempelvis navigering, och kan därför kallas halvautonoma. Autonomi blir alltså en bedömningsfråga och definitionen därmed flytande. En fullt autonom robot kan sköta sig själv »helt och hållet«, och fatta »egna« beslut utifrån vad den har för uppdrag. Men den kan ju inte sköta sig själv »helt och hållet« i den meningen att den en vacker dag kan bestämma sig för att lämna sin militära bana och ägna sig åt trädgårdsarbete istället. Den sortens fri vilja skulle vi ju kräva att en människa har, för att anse att hon är fullt autonom (och, som vi snart ska se, är det inte säkert att ens det skulle vara tillräckligt).

Graden av autonomi hos en robot kan också avgöras av var människan befinner sig i något som man kallar *beslutsloopen*. Det vill säga var människan befinner sig när roboten fattar beslut. Om människan befinner sig *inuti* beslutsloopen betyder detta i militära sammanhang att ett mänskligt befäl har fullständig kontroll över robotens »beslut« och handlingar, exempelvis vilka mål roboten kan välja att beskjuta. Roboten har en högre grad av autonomi om

människan är *ovanför* beslutsloopen. Här har roboten förmågan att välja mål och skjuta utan mänsklig inblandning, men människan har möjligheten att övervaka och är hela tiden redo att gå in och styra och ändra. Sedan har vi det tillstånd då människan befinner sig *utanför* beslutsloopen. Roboten kan då klara av att göra allt helt utan en människas inblandning – det som Winfield benämner *kontrollautonomi*. Det är denna tredje kategori man menar när man talar om fullt autonoma robotar, åtminstone idag. Men, som vi kommer att se längre fram, finns det även filosofiska diskussioner när det gäller olika grader av »full autonomi« hos robotar. Det här med var människan befinner sig i beslutsloopen och hur det påverkar henne är dock inte så enkelt som det kanske verkar vid en första anblick.



Det var en vacker morgon den 3 juli 1988, och kriget mellan Iran och Irak led mot sitt slut. Passagerarna på Iran Air Flight 655 med destination Dubai väntade på att planet skulle lyfta från Bandar Abbas flygfält i södra Iran. Det skulle komma att lyfta först klockan 10:17 – hela 27 minuter efter tidtabell – vilket var mycket med tanke på att den totala flygtiden över Persiska viken bara var cirka 30 minuter. Vissa menar att förseningen kan ha påverkat den tragedi som några minuter senare var ett faktum.

Samtidigt befann sig den amerikanska kryssaren USS *Vincennes* i Hormuzsundet i Persiska viken, som Flight 655 skulle flyga över. *Vincennes* var utrustad med det autonoma stridssystemet AEGIS.

AEGIS – som också är namnet på den grekiska guden Zeus sköld – hade fyra olika inställningar med olika grader av autonomi. Den första inställningen med lägst autonomi var den *halvautomatiska*, i vilken en människa fanns i systemet för att avgöra när det



Iranska barn kastar blommor i havet på 24-årsdagen efter nedskjutningen av Iran Air flight 655, då 290 personer dog. 66 av dessa var barn.

skulle skjutas och på vad. Den andra nivån var *automatisk special*, där en människa bestämde prioriteringarna, som till exempel att skjuta bombplan före jaktplan, men överlät till datorn att bestämma hur detta skulle skötas. Den tredje nivån var *automatisk*, i vilken information gick till den person som hade befälet, så att det fanns en överblick och möjlighet att styra, men datorn kunde sköta sig själv. Den fjärde nivån var *casualty*, eller vad man skulle kunna kalla »olycksläget« eller »krisläget«, där datorn hade möjlighet att på egen hand »bestämma« hur skeppet skulle försvaras, och alltså hade tillstånd att göra det den ansåg bäst för att skydda skeppet. Människan kunde dock gå in och styra datorn i alla dessa fyra nivåer av autonomi, även den sistnämnda.

AEGIS noterade passagerarplanet Flight 655, en airbus, som höll sin kurs och hastighet, och som skickade en signal som talade om att det var civilt. AEGIS registrerade dock planet som ett

iranskt stridsplan av typen F-14 via en ikon på skärmen. En fiende, med andra ord. Trots att data berättade för besättningen att planet inte var ett stridsplan, litade den mer på datorns varning. AEGIS var då inställd på den lägsta nivån av autonomi – halvautomatisk – men inte en enda av de 18 soldaterna och officerarna gick emot datorn. Klockan 10:24 skickade kapten Rogers upp två stycken SM-2BR antifygmissiler. Passagerarflyget var i luften i bara sju minuter innan det träffades, föll mot vattnet och sjönk.

Planets svarta låda återfanns aldrig. *Vincennes* svarta låda fanns dock och kunde ge detaljerad information om skeendet.



Att lita mer på datorn än på sitt eget mänskliga omdöme kallas på engelska för *automation bias*. Det är tendensen att lita på ett autonomt system trots att man har belägg för att systemet har fel eller är opålitligt i en speciell situation. Ett senare, liknande exempel är när U.S. Patriot-missiler under Irakinvasionen 2003 »råkade« skjuta ner två allierade flygplan, som systemet hade klassificerat som irakiska raketer. Det fanns endast några sekunder för att fatta beslutet, så människorna litade på maskinen. Deras roll i beslutsloopen bestod av att de hade veto, det vill säga att de kunde stoppa det datorn gjorde, men detta veto var de inte villiga att använda mot datorns snabbare och – enligt vad de bedömde – bättre omdöme.

Forskare på amerikanska DARPA (*Defence Advanced Research Project Agency*) menar att människan håller på att bli den svagaste länken i kedjan – att den här beslutsloopen man talar om kommer att bli kortare och snävare rent tidsmässigt, så att det till slut blir svårt att över huvud taget få in människor i den, eftersom vi helt enkelt är för långsamma.⁷ Ray Kurzweil, amerikansk uppfinnare

inom områden som maskinläsning och talförståelse, och författare till bland annat *The Singularity is Near*, är känd för *the law of accelerating returns*, det vill säga påståendet att den tekniska utvecklingen är exponentiell. Med detta menas att utvecklingen inte kommer att ske i en jämn, gradvis takt, utan att den blir snabbare och snabbare ju mer avancerad tekniken blir. Kurzweil hävdar att försäkringen om att det alltid kommer att finnas en människa i beslutsloopen bara är en teoretisk beskrivning. Han menar att människan kanske *tror* att hon har kontroll, men att det i själva verket är så att hon bara har kontroll på vissa nivåer.

En annan faktor som gör att människan stegvis flyttas ut ur loopen är vår svårighet att hantera flera system samtidigt. Ekonomiska faktorer kan kräva att varje operatör kontrollerar – eller snarare övervakar – flera robotar på samma gång istället för bara en enda. Eller, ännu bättre ur ett ekonomiskt perspektiv, att en operatör kontrollerar hela svärmar av robotar. Problemet är att redan när man låter en människa övervaka så få som två drönare istället för en, minskar människans prestationsnivå med ungefär femtio procent. En studie från Nato visade att målet där en operatör kontrollerar flera drönare är väldigt ambitiöst och svårt att uppnå.

Även om kommunikationskedjan inte är bruten kan det ändå uppstå situationer där det inte finns tid för en människa att hinna reagera. Den bästa mänskliga stridspiloten behöver åtminstone 0,3 sekunder för att reagera på enkel stimulans och dubbelt så lång tid för att välja mellan flera olika alternativ. En robotpilot behöver mindre än en miljondels sekund. Det finns också robotinställningar som direkt beskjuter den som skjuter – något som skulle ta flera sekunder för en människa och eventuellt skulle fienden då hinna komma undan. Att låta robotar få lov att skjuta tillbaka om de beskjuts skapar undantag till regeln att det alltid ska finnas en människa i beslutsloopen. Många menar att framtidens vapen

kommer att vara för snabba, för små eller för många, och därmed helt enkelt skapa en miljö som är för komplex för människor att hantera.

Varför går då utvecklingen mot ökad autonomi för robotar? Varför vill vi ge våra konstgjorda slavar – som ju i många fall tycks bli både snabbare och starkare än människan – ökad självständighet och frihet?

När en människa styr exempelvis en drönare går en signal från människan till roboten – en signal som kan avbrytas eller störas. Det är då bra om roboten har en viss autonomi. Många menar dessutom, som redan nämnts, att en operatör bör kunna hantera två eller flera robotar samtidigt av ekonomiska skäl, och eftersom den mänskliga förmågan brister just när det gäller simultankapacitet, är det bra om robotarna har viss autonomi och kan fatta åtminstone enkla beslut på egen hand.

En annan, minst sagt viktig faktor när det gäller robotar i krig är fienden. Om en robot inte kommer att skjuta förrän en människa godkänner det, så behöver fienden bara bryta kommunikationen dem emellan. Militärer menar att även om man inte vill ta bort människan ur loopen, så måste det finnas en reservplan när kommunikationen är bruten. Särskilt om alternativet är att roboten ska återvända till basen och riskerar att visa fienden vägen.

Som redan nämnts är risken för störningar ett allvarligt problem och en av drivkrafterna bakom ökad autonomi. Ett slagfält fullt av elektromagnetiska vågor som styr maskiner och robotar innebär förstås en säkerhetsrisk. Att exempelvis störa en GPS är inte svårt – sådana apparater kan vem som helst köpa på nätet.

Men även vid ökad autonomi finns risken att robotarna blir hackade av motståndarsidan, så att fienden får kontroll över en robot och kan få den att göra vad de vill. I en U.S. Army-artikel skriver Ralph Peters hur framtida krig skulle kunna inkludera

elektroniska »övertygningslag« där de två sidornas system skulle försöka övertyga det andra systemet att göra som de själva vill – att få motståndarens robot att köra utför en klippa, eller att koda om sig själv så de slåss för motståndarsidan istället.⁸

Det påstås faktiskt att den amerikanska militären är extra öppen för hackerattacker, eftersom nittiofem procent av kommunikationen sker över kommersiella nät, inklusive satellitsystem. Det var den kinesiska militärakademien som först noterade problematiken med att amerikanerna använde sig av informationens »motorvägar«.

Militära system måste förstås försöka hålla oönskade gäster borta, och den amerikanska militärens interna nätverk, SIPRNet (*Secret Internet Protocol Router Network*), ska vara helt omöjligt att bryta sig in i, men ändå har det vid upprepade tillfällen hänt att ett internetvirus också har spridits på SIPRNet.⁹

För att undvika åtminstone vissa problem med hacking kan man tänka sig att robotarna görs så autonoma att de blir stängda system som inte kommunicerar med varandra via ett nät, utan att de kommunicerar med varandra som människor gör – genom kommunikationsradio och kroppsspråk, just för att undvika att fienden fångar en robot som i sin tur kan påverka resten av den egna armén genom virus eller trojaner. Här ligger en del av drivkraften för att göra militära robotar mer autonoma. Problemet med att göra robotarna helt säkra för hackare är dock att möjligheten att styra och kontrollera dem samtidigt minskar drastiskt.

Amerikanska National Science Foundation menar att en sorts nätverksbaserad telepati kan vara möjlig inom två årtionden. Men ett problem är att även om man kan koppla upp sig till en drönare och styra den med tanken, så är drönaren mycket mer uthållig och fokuserad än en människa, och vi behöver dessutom en dator för att styra drönaren. Då är vi tillbaka till resonemanget om människan i beslutsloopen. Och om våra hjärnor skulle vara kopplade

till maskiner betyder det också att de kan kopplas till varandra – och därmed bli tillgängliga för någon annan, utomstående. Peters menar att det helt enkelt inte kommer att gå att göra något som det är omöjligt att hacka sig in i.

Och det vore ju mardrömmen – att ha en armé bestående av robotar som vänds mot en själv. Som vi strax ska se verkar människor ha ganska lätt för att faktiskt *bry sig om* robotar. Det motsatta, det vill säga robotar som bryr sig om människor, är emellertid inte fallet – åtminstone inte inom den närmaste framtiden. Däremot kan vi människor få för oss att våra robotar bryr sig om oss på något sätt – en sorts projicering av mänskliga känslor på maskinen. Förmodligen kommer detta robotbeteende, där roboten verkar bry sig om oss – vilket i sin tur gör oss bättre till mods och får oss att lita på roboten – att utvecklas tillsammans med autonomi. Det är en säkerhetsåtgärd och skulle då handla om en sorts lojalitetsalgoritm som gör att hur hackad en robot än är, så kommer den aldrig att vända sig mot sin ägare.

Det här med komplexitet är ett kapitel för sig. En robot som fattar egna beslut är förstås väldigt komplex. Men ju mer avancerad en robot blir, desto större är risken för fel i antingen hårdvara eller mjukvara.

Autonomi och ansvar

De mer spektakulära filmrobotarna kittlar vår fantasi i dystopiska framtidsscenarioer, där de får självmedvetande, slår sig fria och hotar att förgöra oss. Vi skakar av oss oron i samma stund som vi lämnat biosalongen, men på senare år har det blivit tydligt att robotarna inte bara är på frammarsch – de blir dessutom alltmer avancerade. Man kan fråga sig vad detta kommer att innebära. Det kanske inte

är så troligt att de blir som i filmens värld, men vad som kan hända är att robotarnas ökade komplexitet gör det mer problematiskt att avgöra var ansvaret ligger om något går snett när en robot gör ett »misstag«.

Anledningen till att vi talar om autonomi hos människor är ofta för att avgöra om vi kan hålla dem moraliskt eller juridiskt ansvariga för deras handlingar. Om någon agerar på ett visst sätt när hon är under tvång, så håller vi henne normalt inte juridiskt ansvarig, men det beror förstås på vilken typ av tvång det rör sig om. Om någon hallucinerar håller vi henne eventuellt inte ansvarig – men även här måste det ofta avgöras från fall till fall. Den som ser tomtar och troll hamnar troligen inom den psykiatriska vården snarare än i fängelset, men den som medvetet berusar sig hålls oftast ansvarig. Vad det handlar om är delvis att vi kräver att den som agerar verkligen har förstått vad som pågår och sin egen roll i skeendet – att personen är vid sina sinnens fulla bruk – för att hon ska kunna ställas till svars. I den så kallade Rödebytragedin år 2007, där en pappa försvarade sin familj och sköt en ung person till döds, bedömdes pappan vara tillfälligt sinnesförvirrad. Den norske massmördaren Anders Behring Breivik ansågs i första instans inte vara ansvarig – man menade att det måste föreligga någon sorts sinnessjukdom. Detta visar att det här med autonomi och ansvar inte är någon lätt fråga, vare sig i de filosofiska seminarierummen, i domstolarna eller samhället i stort. Och då kommer vi in på frågan om det verkligen skulle vara möjligt att hålla en robot ansvarig för sina handlingar – hur avancerad den än vore. Sådana frågeställningar tar oss in i spekulativa områden inom filosofin, och många är benägna att säga att det här med robotar och ansvar i princip är en ickefråga. De som anser detta menar att eftersom en robot alltid är programmerad av en människa, så kan den aldrig hållas ansvarig för sina handlingar.