



UPPSALA  
UNIVERSITET

# Kunskap och Artificiell Intelligens

En undersökning av kunskapsbegreppet och dess relation till  
artificiell intelligens.

*Kalle Kindvall*

Ämne: Filosofi

Nivå: C

Poäng: 15 hp

Ventilerad: VT 2017

Handledare: Lars-Göran Johansson

Examinator: Pauliina Remes

Filosofiska institutionen

<b>1. Introduktion</b>	<b>1</b>
1.1. Bakgrund	1
1.2. Syfte	2
1.3. Frågeställning	2
1.4 Metod	2
<b>2 Kunskap</b>	<b>3</b>
2.1 Kunskap Standardanalysen	3
2.2 Det största problemet med definitionen av kunskap: Gettierproblemet	4
2.3 Internalism och Externalism	5
2.4 Reliabilism	8
2.5 Williams kontextualism	8
2.6 Information istället för Kunskap	10
2.7 Kunskapens värde	11
2.8 Skepticism	13
2.9 Kunskapsteroiernas krav sammanfattade i relation till artificiell intelligens	15
<b>3 Artificiell Intelligens</b>	<b>17</b>
3.1 Artificiell Intelligens Historia	17
3.2 Artificiell Intelligens Bedrifter	18
3.3 Människors förhållningssätt till maskiner	20
3.4 Intelligens	21
3.5 Är Artificiell Intelligens också intelligens?	22
3.6 Testa Artificiell Intelligens	23
3.7 Framtidens Artificiella Intelligens	24
<b>4.Diskussion</b>	<b>26</b>
4.1 Kunskap som endast en Artificiella Intelligens kan besitta	26
4.2 Kan en Artificiell Intelligens besitta kunskap?	28
4.3 Slutsats	33
<b>Referenser</b>	<b>34</b>

# 1. Introduktion

## 1.1. Bakgrund

*Artificiell intelligens* är ett begrepp som inte är självklart i sin innebörd. Ofta syftar det på ett program i en datorer, med mer eller mindre smarta algoritmer. Som beskrivs mer i detalj i avsnittet om den artificiell intelligensens historia så har inställningen till vad som bör räknas som artificiell intelligens ändrats under åren. Nivån för hur avancerat det ska vara för att inte bara kallas för en algoritm utan faktiskt något intelligent, har ständigt flyttats fram. Allt från miniräknare till superdatorer används som hjälpmedel vid forskning och i sökande efter ny kunskap. Ett sammanhang där datorn går från att vara ett redskap till att kanske bli något mer är något som kallas *Eureka Machine*, som tagits fram av forskare på Cornell University. Egentligen är det inte en maskin utan snarare ett datorprogram som körs på flera sammankopplade datorer. Den är designad för att hitta underliggande samband i stora mängder av data. Som test så har maskinen fått analysera data om hastighet och massa hos ett fallande äpple samt från experiment med pendlar som sätts i svängning med mera. Utifrån dessa har den sedan tagit fram beskrivande samband och ett av dessa var  $F=m*a$  det vill säga Newtons andra rörelselag.  $F$  i sammanhanget används som en okänd variabel som visar på relationen mellan massa och acceleration och i detta fallet så har datorn brutit ut den, vilket ger det en form vi är välbekanta med.

Efter att ha visat sig fungera på data som vi redan förstår så fick den sedan analysera data om celler. Sambandet som den beskrivs testades sedan experimentellt och visade sig stämma men forskarna förstår inte själva formeln på grund av systemets komplexitet. Istället för Newtons samband med sina tre termer så innehöll detta samband över tusen termer och detta gör det såklart svårt för människor att få en överblick (Sample, 2009). Om vi tar detta längre så går det att föreställa sig problem där det finns för mycket information som behöver hanteras för att en människa ska hinna ta till sig allt under en livstid.

Vart tar det här oss? Först verkar det som att det kan finnas saker som är bortom mänsklig förståelse men som att dessa saker fortfarande kan hanteras med hjälp av datorer och artificiell intelligens. Men kan den artificiella intelligensen faktiskt besitta kunskap och vad kan vi säga om den samlade mänskliga kunskapen i fall där vi inte förstår det som behandlas?

## 1.2. Syfte

Syftet med denna text är att utreda relationen mellan kunskap och artificiell intelligens. Texten kommer visa på hur artificiell intelligens påverkar eller hänger ihop med olika definitioner av kunskap och vilka krav som ställs på artificiell intelligens för att den ska kunna sägas besitta kunskap.

## 1.3. Frågeställning

- Kan en artificiell intelligens besitta kunskap?
- Vilka olika definitioner av kunskap finns det och hur kan dessa olika definitioner förstås eller relateras till artificiell intelligens?
  - Vilka krav sätter de olika definitionerna av kunskap på artificiell intelligens förmåga att ha kunskap?
  - Vilka problem med de olika definitionerna av kunskap kan artificiell intelligens belysa?
- Vad är artificiell intelligens och vad är skillnaden på intelligens och artificiell intelligens?

## 1.4 Begrepp och tillvägagångssätt

Den här uppsatsen börjar med att beskriva olika definitioner av kunskap. Den utgår ifrån en kunskapsfilosofiskt perspektiv. Många av dessa tar avstamp i tidigare definitioner och dessa finns därför med som en bakgrund, och för att ge en bättre förståelse för vad exakt definitionerna syftar till. Genom att se vilka olika potentiella sätt som kunskap kan ses på så skapas en grund för att förstå vilka egenskaper hos artificiell intelligens som kan vara relevanta.

Begreppet artificiell intelligens är genom denna uppsats relativt löst definierat som någon form av, mänskligt skapat, smart system. Detta kan syfta både till en avancerad algoritm i en dator, så väl som en robot eller maskin med smart beteende, eller många datorer som tillsammans bidrar till ett smart system. Efter avsnittet om kunskap kommer en fördjupning i vad artificiell intelligens är, vad det kan vara och vad artificiell intelligens åstadkommit fram till idag. Slutligen så knyts dessa två delar ihop i diskussionen.

## 2 Kunskap

### 2.1 Kunskap Standardanalysen

Utgångspunkten för att definiera kunskap i denna text är den så kallade standardanalysen om kunskap. Den definierar kunskap med orden sann, berättigad trosföreställning. Formaliserat så säger standardanalysen att en person P har kunskap om att A föreligger om följande gäller.

- 1) A är sann
- 2) P håller A för sann
- 3) P har goda skäl för detta

Alla dessa tre ord kan sedan brytas ner och analyseras noggrant för ökad förståelse av kunskapsbegreppet och också påvisa olika problem med sättet att se på kunskap. Detta leder till olika mer eller mindre skilda definitioner av kunskap och flera av dessa kommer presenteras nedan för att sedan kopplas till artificiell intelligens. Innan dessa skilda definitioner tas upp nedan så tittar vi först lite närmre på de tre orden.

Till att börja med säger definitionen alltså att det måste vara sant, det vill säga att vi kan inte ha kunskap om något som är falskt. Jag kan inte ha kunskap om att ett dygn är 10 timmar. Att det måste vara sant kan verka självklar men detta är något som i vissa fall gör det svårt att säga att någon kan besitta kunskap alls.

Ordet trosföreställning innebär att för att besitta kunskap måste man tro på det. Om något förklaras, bevitnas eller om en person tänker ut något som är sant och berättigat så är det endast kunskap om personen också tror att det är så.

Tillsist kommer då berättigandet av trosföreställningen. Att trosföreställningen är berättigad innebär att den som besitter kunskapen har goda skäl till att tro det den gör. Vad som räknas som goda skäl är en sak som de olika inriktningarna inom kunskapsteori är oense om (Johansson et al., 2014). Denna oenighet är en stor del av vad som utreds närmare i de följande avsnitten av denna rapport.

Med standardanalysen så finns nu en utgångspunkt för funderingar kring vad kunskap är. Vissa inriktningar fokuserar på att göra mer exakta definitioner kring de tre orden och andra, som i nästa avsnitt, ifrågasätter upplägget från grunden.

## 2.2 Det största problemet med definitionen av kunskap:

### Gettierproblemet

Det var Edmund Gettier som först introducerade det här problemet, som visar på en brist i den grundläggande logiken bakom definitionen enligt standardanalysen. Standardanalysen säger ju som beskrivet ovan att kunskap är sann berättigad trosföreställning. Gettierproblemet visar att även om en person tror att något är sant, det är sant och personen har goda skäl till att tro det så vill vi inte alltid att detta ska kallas kunskap. Det finns flera varianter på Gettierproblemet som på olika sätt genererar scenarion som uppfyller alla officiella krav på kunskap men som vi ändå inte är benägna att kalla för kunskap (Johansson et al., 2014).

Till en början presenterar jag här en egenformulerad variant på Gettiers ursprungsexempel. En person ser att någon rigger en hästkapplöpning så att en häst som kallas Fishy Business garanterat kommer vinna. Personen ser också att Fishy Business har nummer 9 på ryggen. Hen har alltså god anledning att tro att Fishy Business kommer vinna loppet och att den hästen har nummer 9. Av detta drar hen slutsatsen att för att vinna ska man satsa på häst nummer 9. Det som sedan händer i detta exempel är att Fishy Business blir påkommen och diskvalificerad innan loppet börjar och att en annan häst får hoppa in och ta både dess plats och dess nummer. Denna nya häst råkar sen vinna loppet. Personen hade alltså rätt i sin logiskt dragna

slutsats om vilket nummer som skulle vinna och den underliggande premissen var baserades på goda skäl eftersom personen klart och tydligt sätt både riggandet av loppet och hästens nummer. Men det var ju ändå inte riktigt rätt och det är detta som är Gettierproblemet.

Enligt standardanalysen så hade personen sann kunskap om vilken häst som skulle vinna. Detta trots att den premiss som det fanns goda skäl att tro på visade sig bli falsk. Det är detta som gör att vi inte vill behöva kalla detta för kunskap då det endast blev sant av en tillfällighet. Om den nya hästen som tog över nummer 9 inte av en tillfällighet varit den som vunnit så hade det ju blivit falskt.

Det finns en annan variant på Gettierproblemet med delvis annorlunda struktur men i grunden så är poängen densamma. Standardanalysen är bristfällig. En lösning skulle vara att helt enkelt lägga in ett till krav till standardanalysen som säger att för att det ska vara kunskap får inte heller premissen vara falsk. Detta täcker dock inte in den andra sorters Gettierproblem.

Ett exempel på detta är en fåraherde som blickar ut över ett fält och ser vad som ser ut att vara ett får och drar därför slutsatsen att det finns ett får på detta fält. I detta fall så visar det sig dock att fåret som fåraherden såg egentligen var en ulv i fårakläder, bokstavligen talat. Premissen om att fåraherden såg ett får är hittills falsk och ger en falsk slutsats, det vill säga inga problem än. Om det däremot visar sig att vad fåraherden trodde var en grå sten egentligen visar sig vara ett smutsigt får så blir både premissen och slutsatsen sanna. Fåraherden såg ett får, drog slutsatsen om att det fanns ett får på fältet och hade rätt, men det fåret som fåraherden trodde sig se var ju inte ett får (Johansson et al., 2014).

Det finns olika försök till lösningar på detta problem och vissa menar att det inte går att lösa problemet som sådant utan att definitionen för kunskap behöver förändras radikalt, vilket beskrivs mer nedan i avsnittet om information. Ett annat sätt är att försöka ta sig runt problemet är att omvärdera vad som räknas som goda skäl, vilket görs i Reliabilismen som beskrivs nedan. Reliabilismen bygger på ett externalistiskt synsätt så innan vi går vidare måste vi först skilja på detta och internalism.

## 2.3 Internalism och Externalism

Detta är två olika varandra uteslutande inställningar till kunskap. De är generellt överens om den grundläggande definitionen om kunskap som sann berättigad trosuppfattning. Där de skiljer är var de tycker att berättigandet kommer ifrån. För internalister så handlar det om att personen som besitter kunskap själv kan förstå och redogöra för varför det den tror är kunskap faktiskt är det. Om jag påstår mig ha kunskap om något är det rimligt att även säga att jag förväntas ha starka skäl för min trosuppfattning. Detta kallas att den som påstår sig ha kunskap har ett epistemiskt ansvar. Den som säger sig ha kunskap måste kunna lägga fram goda skäl för sin trosuppfattning. Om de goda skälen inte kan läggas fram så finns inte heller någon anledning att tro att personen med den påstådda kunskapen faktiskt har en intern förklaring. Det är i denna interna förklaring som internalister lägger vikten. För att förklaringen ska vara godkänd måste internalisten ha tillgång i sitt medvetande till allt det som behövs för rättfärdigandet av kunskapen. Utan denna tillgång eller medvetenhet om skälen till sin tro går det ju inte heller att med säkerhet att säga sig ha skäl.

Det epistemiska ansvaret har konsekvensen att om någon presenterar bättre anledningar för en motstridig kunskapsposition så följer det att internalisten tvingas anpassa sin trosuppfattning. Förutsatt såklart att det inte finns ytterligare anledningar att inte acceptera de nya anledningarna. Att internalisten hade kunskap om något, men nu har kunskap om något annat anses acceptabelt inom internalismen och är en del av det epistemiska ansvaret. Förutsatt att kunskapen baserats på de bästa anledningarna som fanns tillgängliga för internalisten samt att dessa anledningar kan räknas som goda.

Nu kommer ett exempel för att sammanfatta detta. När jag för en timme sedan kollade ut genom fönstret och såg jag vattendroppar falla mot marken och drar därför slutsatsen att det regnade. Att jag såg vattendropparna falla är en anledning till min slutsats. Att jag minns korrekt är ett annat antagande som jag håller för sant. Sedan måste jag även tro att det är sant att vattendropparna faktiskt var regndroppar. Det viktiga för att jag med ett internalistiskt synsätt ska kunna kalla



detta för kunskap är att jag vet om alla dessa anledningar och håller dem för sanna. Det förutsätter även såklart att min anledning av att jag såg vattendropparna falla kan ses som goda skäl för min slutsats. Vart exakt den gränsen dras specificeras inte av ett internalistiskt synsätt, men det finns åtminstone vissa skäl som räknas som tillräckligt goda. När någon sedan berättar för mig att de stått uppe på taket med en trädgårdsslang och vattnat träden utanför mitt fönster så har jag nu alltså fått tillgång till bättre skäl. Mitt epistemiska ansvar tvingar mig nu att omvärdera min slutsats och nu har jag kunskap om att det inte var regn.

En konsekvens av internalism är något som kallas mentalism. Med mentalism menas att eftersom berättigandet kommer från medvetandet, och vilken information som finns tillgänglig internt, så skulle två likadana medvetanden ha samma berättigande för samma trosuppfattningar. Vi kan tänka oss att det hypotetiskt finns två likadana medvetanden som råkar befinna sig i två olika världar. I den ena världen så råkar de berättigade trosuppfattningarna vara sanna medan i den andra världen så är samma, lika berättigade, trosuppfattningarna istället falska. Detta betyder att internalism innebär att berättigade trosuppfattningar inte nödvändigtvis är mer troliga att vara sanna. Om vi vill försöka hitta någon garanti för att en trosuppfattning är mer trolig att vara sann när den är berättigad så behövs ett externalistiskt synsätt.

Externalismen vill alltså att när en trosuppfattning är berättigad så är detta också en indikation på att den är sann. Internalismen och mentalismen som följer av den kan inte garantera en sådan koppling mellan sanning och berättigande. För att externalismen ska försöka göra det så väljer den istället att säga att rättfärdigandet av kunskap måste hänga ihop med yttvärlden. Detta är ett av de steg som externalister tar för att inte drabbas av problemen med mentalism. Det andra innebär ett antagande om att det måste finnas en koppling mellan yttvärlden och personen som besitter kunskap. Externalism kräver dock inte att den specifika kopplingen är känd av personen som besitter kunskap. Tillgängligheten i medvetandet till anledningarna som motiverar kunskap, som internalister kräver, är inte nödvändiga i externalism. En anledning till varför det skulle vara bättre med ett berättigande av sann trosuppfattning är att det i ett externalistiskt synsätt leder till fler sanningar än oberättigade trosuppfattningar. Reliabilismen som beskrivs nedan, säger till exempel att en tillförlitlig process är det som bör ligga till grunden för kunskap. Om

kunskap då alltså tas fram med den och detta leder till fler sanningar, så är detta ett faktum oavsett om den som besitter kunskapen känner till processen. Generellt så menar externalister att det går att ha kunskap, det vill säga sann berättigad trosuppfattning utan att ha kunskap om att det är kunskap. Detta går dock ingen internalist med på och detta är en av de stora skillnaderna mellan externalism och internalism(Johansson et al., 2014).

## 2.4 Reliabilism

Reliabilism är en inriktning inom kunskapsteori och en utveckling av naturalism. Kortfattat kan det sägas att naturalism kopplar ihop kunskap med vetenskapliga resultat. Det finns olika grader av naturalism som gör denna koppling olika tvingande. Att säga att vetenskapliga resultat kan vara relevanta för kunskap är ett steg mot naturalism. Att säga att kunskap endast bör vara svar på vetenskapliga frågor är ytterligare ett steg mot att koppla ihop vetenskap och kunskap. Dessa frågor måste nämligen besvaras med vetenskapliga resultat och alltså kan kunskap endast vara sådant som bevisats med vetenskap, enligt det synsättet.(Johansson et al., 2014)

Reliabilism fokus ligger på tolkningen av berättigad eller goda skäl. För att något ska klassas som goda skäl så måste de ha producerats med en tillförlitlig kunskapsprocess. För att en kunskapsprocess ska räknas som tillförlitlig så ska den producera sanna trosföreställningar för det mesta. Det är alltså hur bra processen fungerar som avgör om den är bra och inte något speciellt i hur den ser ut. Exakt hur ofta den måste ge korrekta resultat är oklart, men kanske behövs det inte definieras utan man kan utifrån sammanhanget avgöra om det är tillräckligt hög kvalitet. Att som naturalismen gör och kopplar ihop vetenskap och kunskap kan vara ett sätt att tolka reliabilismens krav på en tillförlitlig process. Den vetenskapliga metoden blir alltså grunden för hur en bra process skall se ut(Johansson et al., 2014).

## 2.5 Williams kontextualism

Denna inriktning är Michael Williams förslag på kunskapsteori och bygger i grunden på att kunskap och frågan om den är berättigad måste utvärderas i en kontext. Det finns flera andra inriktningar som också kan kallas kontextualism och som tar upp att kontexten är relevant. Williams inriktning har sina rötter både i fundamentalismen samt koherentismen. Kontexten påverkar alltså här vad som är kunskap och inte och det kan delas upp i fem, av Williams, definierade anledningar.

Den första är *semantiska villkor* och syftar till att i vissa sammanhang så finns det vissa påståenden som inte meningsfullt kan ifrågasättas. De måste helt enkelt accepteras vara sanna för att öppna upp möjligheten att komma vidare. Som exempel citeras ett, inom sammanhanget välkänt, bevis från Moore. Ursprunget är att om diskussionen förs på ett sådant sätt att inte ens det mest grundläggande kan tas för givet så kommer inga slutsatser heller någonsin att kunna dras och det finns ingen poäng med att diskutera. I exemplet tycker Moore att en sådan minsta gräns för vad vi i sammanhanget måste ta för sant är att han har två händer. Han bevisar detta enkelt genom att sträcka fram sina båda händer. I sammanhanget att jag påstår mig hålla en penna i min hand så kan det ju såklart ifrågasättas om detta ens är möjligt och om jag ens har händer. I detta sammanhang så är det alltså ett semantiskt villkor för påståendet att jag faktiskt har händer och om man inte tror på detta kan jag enkelt bevisa det på samma sätt som Moore (Johansson et al., 2014).

Den andra sortens anledning som Williams tar upp kallar han för *metodologiska nödvändigheter*. Den påminner om den första anledningen i det att utan vi utan den inte kan komma vidare till meningsfulla frågeställningar. Om vi inte accepterar att det finns djur så kan vi inte heller ställa frågor om deras beteenden.

Nummer tre på listan är *dialektiska faktorer* och är den som tydligast kopplas till ordet kontextualism. I ett visst sammanhang så är det endast vissa argument och relaterade invändningar som räknas. Om en forskare presenterar en ny teori uppbackad med data som stödjer teorin så kan det inte ifrågasättas med argument från någon annan av stilen "Jag personligen tycker inte att det borde vara så." utan

diskursen kräver motsägande data eller argument för varför presenterat data eller slutsatserna av det är felaktiga.

Potentiellt den viktigaste faktorn som presenteras, i relation till datorers förmåga att besitta kunskap, är *ekonomiska faktorer*. Ekonomiskt i sammanhanget syftar till att kostnaderna för att utgå ifrån att något är sant eller falskt respektive, vägs mot varandra. Som ett exempel så är det väldigt farligt att ifrågasätta om gravitationen verkligen fungerar och baserat på detta ett steg ut för ett stup medan kostnaderna för att lita på att detta är kunskap inte är lika höga.

Den sista aspekten kallas situationsfaktorer och med det menas att kunskapen är kunskap endast i denna situation men att om omständigheterna skulle ändras så förväntas vi ompröva vår trosuppfattning (Johansson et al., 2014).

## 2.6 Information istället för Kunskap

Vi har nu sett ett antal olika vinklar på hur kunskap kan definieras, vad som kan räknas som goda skäl samt hur dessa kan ifrågasättas. Ett annat alternativ som kommer presenteras i detta avsnitt är att helt enkelt använda ordet *Information* i stället. Varför detta är rimligt samt vilka konsekvenser det kan få kommer tas upp.

Det finns flera olika sätt att använda ordet information. Vår internetuppkoppling kan överföra en viss mängd information varje sekund. Jag kan ha information om att föreläsningen börjar kl 15:15. Ditt DNA har information till dina celler om hur kroppen ska byggas upp. Den andra av dessa användningar, det vill säga att jag har information *om* något kallas för semantisk information. Semantisk information måste vara sann, för om det inte är sant så är det inte heller information. Begreppet falsk information kan användas i det sammanhanget men med semantisk information så menas helt enkelt sann information. En liknelse kan göras med falska pengar, som precis som falsk information inte har ett värde och falska pengar är inte pengar vilket betyder att falsk information inte heller bör räknas som information.

Personligen så tycker jag inte riktigt att denna liknelse räcker som förklaring till varför falsk information inte har ett värde. Om jag är informerad om att det brinner i byggnaden och därför väljer att fly men det visade sig att det var en giftig gas som släppts lös istället så var likväl informationen om branden värdefull för mig. Falska

pengar kan också ha ett värde, om den som får betalt i dem inte märker att de är falska. Problemet är analogt med att tillskriva kunskap värde över sann trosföreställning, som diskuteras i nästa avsnitt och som inte är helt enkelt. Tills vidare så används helt enkelt orden information och semantisk information med utgångspunkten att de syftar till sann information.

En stor skillnad mellan information och kunskap är att besittaren av informationen inte måste vara medveten. En dator kan ha utan problem eller ifrågasättande innehålla information. Men eftersom information inte är kunskap så hjälper detta inte för att reda ut huruvida datorer kan besitta kunskap. Det finns dock anledning att fundera på om den grundläggande definitionen av kunskap bör ses över med information som utgångspunkt.

Gettierproblemet är enligt Floridi olösbart och personligen håller jag med (Johansson et al., 2014). Han säger sig visa detta genom ett problem han kallar koordinationsproblemet. Om olika agenter har olika information och bara en har full information så måste informationen föras vidare, det vill säga koordineras till de andra agenterna men det går inte att garantera att informationsöverföringen sker perfekt. Som exempel kan vi tänka tillbaka på Gettierproblemet med får som beskrevs innan. Säg att det fanns en till person på ängen som såg både fåret som fåraherden trodde var en sten och att det andra fåret faktiskt var en varg som klätt ut sig. Då skulle denna persons kunskap om att det fanns ett får på ängen uppfylla det vi förväntar oss av kunskap och om den extra informationen som denna person besitter vidarebefordras till fåraherden så skulle båda ha äkta kunskap. Men i varje liknande scenario så går det såklart inte att garantera att denna andra person kan vidarebefordra all information. Det går därför inte att garantera undvika Gettierproblemet. Floridis lösning är en definition av kunskap som innebär att *information t är kunskap för en agent A* om följande tre villkor gäller.

- 1) Det finns en *Redogörelse* för hur *t* blev information
- 2) *A* är informerad om *t*
- 3) *A* besitter informationen att *Redogörelsen* finns

Redogörelsen är en term i sammanhanget, speciellt definierad, och innebär en beskrivning av kunskapskällan samt olika noder med kunskap och hur dessa relaterar till varandra. Redogörelsen blir en garanti för att kunskapen är pålitlig (Johansson et al., 2014). Eftersom agenten i Floridis definition inte behöver vara en människa så är detta sätt att se på information och kunskap kanske kompatibelt med att datorer kan besitta kunskap. Mer om detta senare.

## 2.7 Kunskapens värde

Att ha kunskap är självklart värdefullt då det kan vara instrumentellt för att uppnå diverse mål. I vissa fall kan dock detta uppnås utan att det man tror sig ha kunskap om är berättigat, det vill säga man har endast en sann trosföreställning. Att tro rätt om något kan i praktiken ge samma resultat som verklig kunskap. Detta skapar ett problem i sökandet efter kunskapens värde. En anledning till att säga att kunskap är värd mer än sann trosföreställning är att berättigandet ger någon form av garanti för att den sanna trosföreställningen faktiskt är sann och inte falsk. Falsa trosföreställningar kan orsaka problem. Om jag precis som i exemplet om falsk information befinner mig i ett hus som påstås brinna och för att rädda mig hoppar ut genom fönstret och bryter ett ben. Detta räknar jag som en vinst framför att dö tills det visar sig att det inte alls brann. Det finns därför ett värde i att inte bara ha sanna trosföreställningar om vissa saker utan att även undvika de falska.

Ett annat sätt att ge värde till kunskap är att se till den som besitter kunskapen och säga att det är dennas rationalitet som ledde fram till den sanna trosföreställningen. Detta kallas dygdteori eftersom det är den som besitter kunskapens dygdighet som ger kunskapen ett extra värde. Ordet dygd kommer från det engelska ordet *virtue* och i detta sammanhang syftar det till en mer allmän god egenskap än som det ofta används i det svenska språket. Helt enkelt, om vi litar på att personen som har kunskapen är bra eller duktig på det berörda ämnet så litar vi också på att personens kunskap inom detta ämne har ett värde.

Ytterligare ett sätt att se på det är att kunskap faktiskt ändå är värdefullare oavsett om den sanna trosföreställningen är lika effektiv. Detta är en åsikt som reflekteras i sammanhang där kunskap eftersträvas utan att ett tydligt eller kanske

ens avlägset användningsområde som till exempel inom ren matematik(Johansson et al., 2014).

Dock tycker jag att detta värde på något sätt är kopplat till en idé om att även om vi inte vet vad nyttan är än så har vi sett tillräckligt många gånger att oväntad kunskap senare kommer till nytta. Men även utan detta så finns det också något som jag inte kan definiera mer än som en känsla av att även om det inte kommer till nytta så är kunskap ändå värdefull. I nästa avsnitt så tas konceptet skepticism upp som ifrågasätter om det är möjligt att besitta kunskap. Ofta motiveras det med olika hypotetiska, potentiellt möjliga, beskrivningar av världen där sammanträffanden leder till en illusion av kunskap. En anledning till att jag och många med mig känner att kunskap har ett värde är kanske helt enkelt att dessa hypotetiska beskrivningar, samt de hypotetiska scenarion som tas upp i detta avsnitt om hur falska trosuppfattningar kan vara av värde, helt enkelt *känns* för osannolika.

Några som antagligen inte alls håller med om denna känsla, eller åtminstone skulle argumentera för att den är felaktig är däremot skeptikerna som tas upp i nästa avsnitt.

## 2.8 Skepticism

Att det ens är möjligt för människor att besitta kunskap är något som i många sammanhang, inte minst på andra ställen i denna text, tas för givet. Förespråkare av skepticism håller dock inte med om att det är så enkelt. Det finns två grundinställningar som båda är former av skepticism. Den första inställningen är att det omöjligt att ha kunskap. Den andra säger att eftersom vi inte kan veta säkert om det är möjligt eller inte så bör vi inte uttala oss alls om det går eller inte går att besitta kunskap. Det finns sedan flera mer eller mindre radikala tillämpningar av dessa grundinställningar som leder till olika inriktningar inom skepticismen.

En sådan inriktning kallas Pyrrhonsk skepticism och den säger att vi inte kan veta något säkert. Till att börja med så bör vi inte lita på våra sinnesintryck. Hur vi upplever en situation påverkas av många olika faktorer som vårt humör, vår kultur och vårt tillfälliga eller permanenta fysiska tillstånd. Taklampan i sovrummet är för ljus på morgonen men för mörk när jag ska läsa och samtidigt kan den då vara för

ljus för min sambo som försöker sova. Den kanske också är för högljudd om man fokuserar på det lilla surrande ljud den ger ifrån sig. Det finns alltså motstridande påståenden om lampan både från olika personer och från olika tider eller sinnesintryck hos samma person och vi bör därför inte säga att det finns en sanning om lampan. Genom att ta detta till sin spets så blir slutsatsen alltså att vi inte kan veta något alls med säkerhet.

Ett motargument mot att ta skepticism så här långt är om man kopplar en persons kunskap till dess handlande. Att jag inte hoppar ut från ett högt stup beror (förutom bristen på höga stup i min närhet) på att jag har kunskap eller åtminstone en övertygad trosföreställning om att detta kan leda till min död. En skeptiker har svårt att formulera ett argument mot att ta steget ut för stupet, utan att gå ifrån den påstådda övertygelsen om att vi inte kan eller bör säga oss veta något. De kan såklart svara att de inte har någon övertygelse men inte heller någon anledning att ta det steget. Även om Pyrrhonsk skepticism inte är så stor idag så är ändå lärdomen att allt kan ifrågasättas och att vi måste formulera goda anledningar till vad som ska räknas som kunskap. Men också att vi, för att vi inte alla ska hoppa till vår död, måste ha någon form av begränsning i vår skepticism.

Humes använder skepticismen för att förbättra kunskapen istället för att förkasta den helt. Han säger att vi bör försöka formulera så bra teorier som möjligt och sen använda skepticism för att ifrågasätta dem. Genom ifrågasättandet så filas teorierna så att de blir ännu vassare. Att använda en så kallad lokal form av skepticism, som alltså inte ifrågasätter kunskap som sådant utan istället ifrågasätter vissa specifika kunskapsanspråk. Ett exempel på ett sådan ifrågasättande är resonemang som bygger på induktion. Eftersom solen har gått upp varje morgon så kommer den även gå upp imorgon. Dock så finns möjligheten, om än minimalt sannolik, att någon fritt flygande himlakropp kommer och rubbar ordningen i vårt solsystem vilket leder till att solen inte alls går upp. Eftersom detta inte går att utesluta helt kan vi inte heller garantera att det inte händer. Hume säger att om vi ser till hur vi faktiskt gör så kommer vi fram till att vi ändå inte kastar bort kunskap bara för att den är baserad på induktiva resonemang. Vi är medvetna om att de är provisoriska men vi använder dem ändå som kunskap. Jorden existerade igår och i förrgår och eftersom jag induktivt resonerar mig fram till att den kommer göra det



imorgon också så ställer jag väckarklockan för att inte missa min föreläsning dagen efter.

Kanske det viktigaste området för lokal skepticism relaterat till artificiell intelligens är det som beror andras medvetande. Nicholas Malebranche formulerade en skillnad mellan det egna och andras medvetande. Jag vet att jag tänker eftersom jag tänker att jag tänker. När det gäller andra så är det inte lika enkelt utan då måste man betrakta vad det påstår eller deras handlande. Ett argument för likställande är analogiargumentet. Jag tänker och är en människa alltså tänker även andra människor eftersom de är som mig. Analogier är dock inte det samma som bevis och detta sätt att motivera det har ifrågasatts. Att använda induktion för att sluta oss till andra människors medvetande är inte heller helt garanterat eftersom induktiva resonemang inte ens helt kan garantera att solen går upp.

Det verkar alltså inte gå att garantera att andra människor faktiskt har ett medvetande. Men kanske kan vi ändå gå vidare. Ett sätt att resonera kallas för abduktion. Det innebär att man använder slutledning för att nå den bästa förklaringen och att det helt enkelt är den som används till vidare. I detta sammanhang innebär det alltså att den rimligaste och alltså bästa förklaringen till andra människors beteende är att dom också har ett medvetande som ens eget (Johansson et al., 2014).

Detta avsnitt om skepticism är det sista som beskriver inriktningar inom kunskapsteori. I nästa avsnitt kommer en sammanfattning av dessa och för att koppla ihop det till fortsättningen av denna rapport så kommer de även kortfattat att relateras till artificiell intelligens.

## 2.9 Kunskapsteoriernas krav sammanfattade i relation till artificiell intelligens

Relationen mellan kunskapsbegreppet och artificiell intelligens skiljer sig åt beroende på vilken av definitionerna och inriktningarna ovan som används som utgångspunkt. Här kommer jag försöka sammanfatta det viktigaste i varje definition i förhållandet till artificiell intelligens. Detta för att göra det tydligare vilka aspekter av artificiell intelligens förmågor som är viktiga i följande avsnitt och vad vi behöver reda ut.

Kunskap ska ju enligt standardanalysen vara sann. Denna del av problematiken förändras generellt inte när vi istället utvärderar maskiners potentiella kunskap jämfört med människor. Är det sant så är det sant. Att informationen finns där i sig räcker dock inte utan den måste kunna beskrivas som ett påstående för att kunna vara sant eller falskt.

Vad som krävs för att säga att en agent faktiskt har en trosföreställning gällande något blir mer problematiskt när det är en artificiell intelligens som är agenten, istället för en människa. Generellt så accepterar vi i kunskapssammanhang att människor som säger sig tro det också gör så. Undantaget är när vi vill bygga vår egen kunskap baserat på vittnen eftersom vi då för rättfärdigandet av vår egen kunskap inte gärna kan basera den på en lögn. Det första alternativet är att artificiell intelligens måste kunna visas besitta en förmåga att faktiskt ha en trosföreställning. Alternativ nummer två är att se på artificiell intelligens som påstår sig ha en trosuppfattning som en potentiell lögnare. Detta är dock ett stort problem för även om det är sant och berättigat så vill vi inte kalla det kunskap hos agenten om denne inte kan sägas tro på det. Alternativet att helt enkelt lita på den potentiella lögnaren finns ju såklart också.

Det underliggande antagandet som görs är dock att andra människor har ett medvetande vilket möjliggör trosuppfattningar. Om det går att säga att något utan medvetande kan ha en trosuppfattning är tveksamt. Om en artificiell intelligens potentiellt kan sägas ha ett medvetande är alltså relevant att undersöka närmare.

Reliabilismen fokuserar på processen så om artificiell intelligens ska kunna sägas besitta kunskap så är det viktigaste här att processen är pålitlig. Vad kan vi då säga om processen som en artificiell intelligens utför för att tillskansa sig kunskap. För att säga att artificiell intelligens har kunskap enligt reliabilismens synsätt så är det viktigaste att helt enkelt visa att processen är pålitlig.

Att se till kunskapens värde ger generellt ett synsätt som inte borde skilja mycket om det är en maskin eller människa som besitter kunskapen. Detta eftersom värdet kunskapen bidrar med är densamma. Dydgteorin passar dock inte ihop med detta då man förutsätter en rationalitet hos den som besitter kunskapen och att det är detta som ger den dess värde. För att vi ska kunna säga att maskiner har kunskap

enligt detta synsätt måste vi alltså undersöka maskiners förmåga att resonera rationellt.

När det gäller datorer och information är det som behöver redas ut helt enkelt skillnaden på att ha information och att vara informerad. Om man kan säga att en artificiell intelligens kan vara informerad så kan den enligt Floridis definition också besitta kunskap. Olika ordval i liknande definitioner ger olika slutsatser. Om betydelsen av att informera används som att man tränar eller instruerar någon så blir slutsatsen att vi kräver av maskiner att dom kan lära sig. En annan definition säger att det är att överföra kunskap vilket alltså inte hjälper i detta sammanhang utan helt enkelt leder till att vi måste kunna likställa maskiner med människor till en viss grad. Om man använder ordet som att helt enkelt överföra information så behöver vi inte mer utreda datorers förmåga gällande detta utan kan direkt säga att de kan besitta kunskap. Att välja den definitionen skulle nog av många anses som otillräckligt i sammanhanget och de övriga aspekterna kommer därför undersökas vidare.

Skepticismen visar att det inte är en självklarhet att ens människor kan besitta kunskap. Med den utgångspunkten hjälper det inte ens att lyckas visa att artificiell intelligens är likvärdig med mänsklig intelligens. När de mindre radikala formerna av skepticism betraktas så kanske flera av de argumenten som används för att ändå godta att människor kan ha intelligens också kan appliceras på maskiner. Detta kräver lite olika saker. Om analogiargumentet ska kunna användas så måste maskinerna vara väldigt lika oss. Abduktion kanske kan leda till att vi bör betrakta maskiner som intelligenta och som besittare av kunskap om de visar upp färdigheter av sådan nivå att detta blir den rimligaste förklaringen. Att visa på när artificiell intelligens presterat som bäst blir alltså nödvändigt för att kunna utvärdera detta noggrannare

### 3 Artificiell Intelligens

### 3.1 Artificiell Intelligens Historia

Här kommer jag beskriva hur artificiell intelligens började och hur det har utvecklats. Med hjälp av historien samt hur samtalet kring artificiell intelligens har sett ut under olika tider vill jag dels sätta dagens bedrifter i ett sammanhang. Det har varierat mycket mellan väldigt optimistisk till väldigt pessimistisk syn på hur snabbt man tror att artificiell intelligens kommer att förbättras och vad den kommer klara av. Med medvetenhet om historien kan vi bättre förstå utvecklingen i framtiden.

I slutet av 1950-talet så fanns det väldigt stora förväntningar på det nya fenomenet artificiell intelligens. Det gjordes också en del väldigt lovande framsteg. 1957 utvecklades ett program som kallades för General Problem Solver. Denna lyckades lösa problem inom matematiken och ta fram teorem som matematiker tidigare inte kunnat göra. Tillsammans med andra lovande program så skapade detta ett väldigt stort hopp på vad artificiell intelligens skulle kunna göra inom den närmaste framtiden.

Det visade sig att mycket av förhoppningarna var lite för optimistiska och efter väldigt mycket investeringar under 1970-talet så minskade investeringarna under 1980. Detta har ibland kallats för AI-vintern. Men efter vintern kommer vår och datorerna som användes blev fortfarande snabbare. Artificiell intelligens som område fortsatte att utvecklas och nya nivåer uppnås (Kurzweil, 2005).

En historiskt viktig och välkänd händelse är när IBM's dator Deep Blue den 11 maj 1997 blev den första datorn som besegrade den bästa mänskliga schackspelaren (ibm.com, 2016). Även om det var ett stort steg och något som många inte trodde skulle vara möjligt förrän långt senare så är man ändå inte villig att kalla Deep Blue för intelligent. En anledning som läggs fram till varför detta inte är intelligens är att den endast testar flera tusen positioner utan att egentligen förstå vad den gör. En annan anledning att inte vilja kalla den för intelligent är att detta är det enda den kan och den är inte intelligens när det gäller något annat. Det skulle kunna kallas intelligens inom ett snävt fält (Nyu.edu, 2015).

## 3.2 Artificiell Intelligens Bedrifter

Här presenteras en lista på olika saker som artificiell intelligens har lyckats göra. Detta för att kunna utreda om några av dessa kan beskrivas som att den artificiella intelligensen besitter kunskap.

Det första exemplet är ett artificiell intelligenssystem utvecklat av forskare vid Massachusetts Institute of Technology(MIT) som kallas ConceptNet 4. Andra forskare från University of Illinois satte detta system på prov och gav det standardtest för att utvärdera barns intelligens. Resultatet blev att ConceptNet 4 fick ungefär lika många poäng som en fyraåring. Den var dock väldigt ojämn, bra på vissa delar av testet som vokabulär och att känna igen likheter men var sämre på förståelsefrågor som att besvara "varför"-frågor (Reilly, 2013; ScienceDaily, 2015).

AlphaGo har, vid tiden av denna rapports skrivande(våren 2016), uppnått en av de potentiellt största bedrifterna i artificiell intelligens historia. AlphaGo är ett program utvecklat av Google för att spela det kinesiska spelet Go. Reglerna för Go är väldigt enkla. Spelplanen är ett rutnät och varje runda så placerar de två spelarna ut varsin sten med svart respektive vit färg. Målet är att ringa in den andre spelarens stenar. Trots sin enkelhet är Go väldigt komplext då brädet är väldigt stort och besluten som måste fattas om vart det bästa stället att placera nästa sten baseras på väldigt många faktorer. Brädet är 19x19 rutor stort, att jämföra med schacks 8x8 stora spelplan. Detta gör att det finns fler möjliga drag än det finns atomer i hela universum. Även om det fanns lika många parallella universum som det finns atomer i universum så räcker inte mängden atomer i alla dem universumen till. Detta illustrerar omöjligheten i att i en dator representera alla möjliga drag. Med bakgrund av detta så har Go ansetts vara väldigt beroende av mänsklig intuition och att möjligheten för en dator att vinna i detta därför legat minst 10 år in i framtiden. I mars 2016 så ändrades allt detta då AlphaGo besegrade Lee Sedol, en Go-spelare som vunnit världstiteln 18 gånger. Matchen spelades bäst av 5 och AlphaGo tog hem 4 vinster. Tidigare samma år besegrades även den europeiska mästaren Fan Hui. Det som gjorde att AlphaGo lyckades var att den inte försökte beräkna alla drag utan istället arbetade mer likt så som mänsklig intuition fungerar. Efter att ha tränats med

väldigt många tidigare matcher i Go så utvecklade den själv ett sätt att bedöma vad som är ett önskvärt tillstånd på brädet, och utifrån det kan den sedan göra sina drag för att sträva efter ett så fördelaktigt bräde som möjligt. Att den har möjlighet att lära sig på detta sätt är för att den använder så kallade artificiella neurala nätverk som är ett avancerat algoritmiskt system som försöker imitera hur den mänskliga hjärnan hanterar information (Nature.com, 2016)(Deepmind.com, 2016)(Official Google Blog, 2016).

Det finns även saker som datorer vida överträffar människor på. När det gäller beräkningar och minne så är människan chanslös i jämförelse. Ett exempel på när detta använts för att ta fram kunskap är bevis inom matematik. Detta har gjorts flera gånger, och det största är på 200 terabyte. Det vill säga 200 000 000 000 000 bytes av information, alldeles för stort för att någon människa ska kunna gå igenom det och verifiera det. Beviset är för ett problem som kallas *Boolean Pythagorean triples* (Lamb, 2016).

### 3.3 Människors förhållningssätt till maskiner

Det finns även saker som maskiner kan göra som kanske kan tyckas väldigt orelaterat till kunskap. Anledning till att jag tar upp det är för att visa på *hur vi förhåller oss* till maskiner, datorer och artificiell intelligens. Vårt förhållningssätt till saker är ibland relevant för hur vi bör definiera dessa och det är därför jag tar upp det här. Att säga att någon besitter kunskap enligt standarddefinitionen innebär ett erkännande av att denna person har en trosuppfattning och alltså medvetande nog att tro. Frågan som detta avsnitt ska undersöka är alltså om vi förhåller oss till maskiner som om de har ett medvetande.

Flera olika studier har visat på att människor känner empati för robotar. Ett test gjordes med foton på händer. Vissa händer var människohänder och andra tillhörde en robot. I olika versioner av dessa så fanns det knivar som antingen var säkert positionerade eller såg ut att vara redo att skada handen. Försökspersoner visades dessa bilder och deras reaktioner mättes med elektroder på deras huvud. Undersökningen kom fram till att de bilderna där handen såg ut att vara i fara gav samma reaktion oavsett om det var en mänsklig hand eller inte. Försökspersonerna

reagerade marginellt mer på människohanden. När båda händerna visades i säkra positioner så reagerade försökspersonerna oberört oavsett (Suzuki et al., 2015).

Detta visar att människor instinktivt reagerar väldigt lika på att se en människa och en livlös robot potentiellt komma till skada. Denna reaktion har ingen nödvändig koppling till vad vi faktiskt tycker om dessa olika scenarion. En annan studie gjorde ett liknande test där försökspersoner visades olika korta filmklipp. Vissa med robotar och vissa med människor och varianter där dessa behandlades både bra och dåligt. Resultatet var samma som ovan, det vill säga att människor som behandlas dåligt aktiverar hjärnan mer men att samma sorts aktivering i hjärnan sker när det gäller robotar. Försökspersonerna fick även själva beskriva hur de kände inför vad de såg och beskrivningarna matchade det som uppmätts (Sciencedirect.com, 2016).

Testen visar hur vi känner inför det, och detta påverkar såklart också hur vi beter oss. En annan studie visar tydligt på detta. Barn ombads att hålla i en docka, en hamster och en Furby, upp och ner så länge de ville. När de höll i dockan var det att de blev trötta i armen som gjorde att de inte ville hålla längre. Hamstern kravlade och verkade inte uppskatta det vilket ledde till att barnen nästan omedelbart slutade. Furbyn, som är en hårig liten leksak som kan ge ifrån sig ljud, var programmerad att gråta när den hölls upp och ner vilket fick barnen att även där välja att släppa ner den mycket snabbare än dockan (Fisher, 2016).

Inte bara barn utan även vuxna känner såhär. När robotföretaget Boston Dynamics släppte en video som visade på dess senaste robots stabilitet var det många som var upprörda över att de gjorde detta genom att sparka roboten (Phoebe Parke, 2016).

### 3.4 Intelligens

Om vi undersöker ordet intelligens och dess definition kanske vi kan avgöra om en artificiell intelligens bör räknas som något eget eller om det kan tillskrivas samma egenskaper. Vi människor har intelligens och denna intelligens möjliggör besittande av kunskap. Men när vi talar om intelligens så kan vi mena många olika saker. Så kanske är definitionen inte tillräckligt precis för att hjälpa oss att förstå artificiell

intelligens förmåga till kunskap. Det som presenteras nedan är alltså inte ett svar på vad intelligens är utan bara en sammanfattning av olika .

IQ är ett vanligt mått på en logisk form av intelligens men det finns även många andra former av intelligens. Lärande, kommunikation, emotionell kunskap, minne och kreativitet är alla aspekter av tänkande som kräver former av intelligens (Detterman, 2005). Vad intelligens exakt innebär är dock inte självklart eller fullt definierat.

I Oxfords Compact Dictionary(Waite and Hawker, 2009) med flera (The American Heritage dictionary of the English language, 2000) så har dom definierat det som "Förmågan att ta till sig och tillämpa kunskap och färdigheter." vilket leder oss till ett cirkelresonemang i detta sammanhang.

Detta är dock inte helt fruktlöst då vi omvänt kan använda kunskapsbegreppet för att uttala oss om maskiner bör räknas som intelligenta eller inte. Om vi kommer fram till att de kan besitta kunskap så har vi alltså också kommit fram till att de är intelligenta, förutsatt att vi accepterar den definitionen av intelligens.

För att gräva lite djupare så kan man jämföra med andra försök att definiera intelligens. World Book Encyclopedia säger att det är förmågan att anpassa sig till sin omgivning(The World Book encyclopedia, 2006). En lite längre definition kommer från en psykolog vid namn Anastasi (Anastasi, 1992) som säger "Intelligens är inte en enda enad förmåga utan istället en sammansättning av flera funktioner. Termen syftar till en kombination av förmågor som är nödvändiga för överlevnad och avancemang inom ett visst sammanhang.". Att vara snabb är ju en förmåga som haren har och som hjälper den överleva men är detta också då intelligens? Andra definitioner hanterar detta genom att specificera att det är en kognitiv förmåga(Saklofske, wishes and Sattler, 2002).

Generellt är dessa olika synsätt ganska överens om vad intelligens är, även om det använder olika ord. Att kunna tänka ut hur ett problem skulle kunna lösas kräver intelligens. Med detta kan vi nu lättare utvärdera artificiell intelligens.



### 3.5 Är Artificiell Intelligens också intelligens?

Eftersom Deep Blue lyckas anpassa sig och lösa problemet av vilket som är det bästa schackdraget i en ny situation så kan den kanske då sägas ha någon form av intelligens. I fältet som forskar om artificiell intelligens brukar man skilja på snäv intelligens och generell intelligens. Enligt den definitionen så har Deep Blue mycket riktigt intelligens, men detta är av den snäva varianten. För att kunna likställa eller säga att en dator har likvärdiga kognitiva förmågor som en människa bör den dock uppnå generell artificiell intelligens. För att den skulle räknas som en generell artificiell intelligens skulle det krävas att den kunde lösa även många andra problem.

Alltså bara för att artificiella intelligenser uppnår vissa definitionen av intelligens så kan vi inte sluta oss till att den är av tillräckligt hög nivå för att också tillskriva en maskin förmågan att besitta kunskap. Detta kan vara problematiskt eftersom vi inte endast kan reda ut maskiners förmåga till intelligens utan även måste undersöka dess nivå. Omvänt kan vi säga att maskiner åtminstone kan uppvisa viss form av intelligens och om de uppvisar tillräcklig förmåga för att tillskrivas kunskap så behöver inte existensen av deras intelligens också ifrågasättas. I enlighet med Oxfords Dictionarys definition ovan så är en maskin som besitter kunskap implicit intelligent. Detta kan verka som en återvändsgränd, men det fyller ändå en viktig funktion. Kanske kan det användas åt andra hållet. Kunskapsbegreppet kan hjälpa oss att avgöra den svåra frågan om när en maskin ska räknas som intelligent.

### 3.6 Testa Artificiell Intelligens

Turing menade att frågan om maskiner kan tänka är för svår att svara på direkt. Som den här uppsatsen visar så kan kunskap ses på olika sätt och detsamma gäller för intelligens. Att ha en enhetlig definition av vad tänka är och som alla är överens om är kanske en omöjlig uppgift. Även ordet maskin behöver definieras för att kunna svara på den frågan. Det kan synas lättare, men kommer en definition av maskiner att inkludera eller exkludera att människor ses som maskiner? Turings

föreslagna alternativ till att försöka definiera vad vi menar med "kan maskiner tänka?" kallar han för "The Imitation Game"(Turing, 1950).

Genom att skapa ett test som kräver intelligens kan vi säga att om en dator klarar av det så är den intelligent. Testerna gör det lättare att försöka komma åt problemet utan att exakt påstå sig förstå eller kunna analysera intelligent tänkande. Grundidén är att om vi inte kan skilja på vad en dator eller robot gör jämfört med en människa när de utför något som kräver intelligens så bör vi räkna dem som intelligenta.

Det finns flera varianter av testet men huvudprincipen som Turing föreslog är att en utvärderare kommunicerar med antingen en dator eller en människa, via textmeddelanden och utan att se den andra. Personen som utvärderar ska sedan avgöra om den kommunicerade med en dator eller en människa Om datorn lyckas lura tillräckligt många att tro att de kommunicerar med en människa så anses den ha klarat testet (Turing, 1950)..

En förutsättning är att utvärderaren ställer bra frågor, vilket såklart är problematiskt. Om en maskin klarar av att lura en dålig utvärderare så säger det inte så mycket att maskinen klarar av testet. Att börja fråga hur den andre mår kan vara en bra utgångspunkt. Den frågan är dock lätt för en maskin att låtsas förstå och svara på. Så för att verkligen testa den behöver många olika sorters frågor ställas, men även kanske mer avancerade mentala utmaningar. Det finns dock en möjlighet menar Turing för personen som utvärderar maskinen att introducera nästan alla former av mänskligt tänkande som utmaningar, och det är detta som krävs för att utvärderingen ska bli riktigt bra. Som exempel tar han upp att man kan fråga om personen spelar schack, följt av att beskriva ett bräde och fråga om bästa draget. Värt att påpeka är att detta exempel gjordes 1950, det vill säga långt innan datorn Deep Blue vann mot Kasparov 1997. Detta är delvis ett exempel på hur gränsen för vad vi kräver av datorerna ständigt flyttas när de når nya bedrifter, dock kan man argumentera att det är stor skillnad på en dator som är skapad för att spela schack och en dator som är skapad för att imitera en människa och som bland många andra saker kan spela schack. Det vill säga skillnaden mellan specifik och generell intelligens.

En annan kritik som tas upp är att testet är orättvist mot maskinerna. Om man istället vände på det och en människa skulle försöka imitera en maskin så skulle det vara lätt att identifiera människan för att den inte klarar av att räkna ut ett väldigt stort matematikproblem på extremt kort tid. Detta är en intressant poäng och en definition av intelligens som inkluderar och erkänner maskiners specifika egenskaper som intelligenta kanske ger en bättre förståelse. Dock så är detta inte ett problem om maskinen klarar av testet då dess intelligens då fått ett erkännande (Turing, 1950).

### 3.7 Framtidens Artificiella Intelligens

Det finns många olika teorier, idéer och åsikter om hur artificiell intelligens kan utvecklas i framtiden. Syftet med att ta upp det är att kunna ställa frågan på sin spets och ta de exempel som kommer upp gällande dagens artificiell intelligens ett steg längre. I grunden är frågan inte beroende av att teknologin faktiskt når dit, även om det så klart gör den mycket mer intressant att potentialen finns.

En möjlighet är att i framtiden så kommer det finnas datorer som är mer intelligenta än människor. Det vill säga som kan överträffa oss på alla aspekter som vi tycker räknas som intelligens. Att försöka hitta en anledning till att dessa ändå inte kan besitta kunskap är svårt om man samtidigt vill hålla kvar vid att människor har möjlighet att besitta kunskap.

Singulariteten är något som ofta kommer upp i sammanhang om framtida artificiell intelligens. Idén bakom detta är en dator som är programmerad för att programmera sig själv smartare. Det är alltså en artificiell intelligens som vars uppgift är att förbättra sig själv. Samtidigt som den gör sig själv bättre så ökar den också sin egen potential för att fortsätta göra sig själv bättre och snabbare. Detta leder till en exponentiell utveckling och tillslut kommer en punkt när denna utveckling sker så snabbt och artificiella intelligensen är så avancerad att det i princip är oändligt. Det är detta som kallas för singulariteten. Detta är såklart endast en idé om vad som skulle kunna hända. Den har dock stöd från flera håll som en trolig teori. Ray Kurzweil är en väl respekterad forskare och i sin bok "The Singularity Is Near" lägger han fram många argument som stödjer teorin om singularitet som en trolig framtid (Kurzweil, 2005).

Här kommer en sammanfattning av några få av argumenten. Generellt så är vi människor vana att tänka i linjära termer vilket gör att när vi tänker kring framtiden så tänker vi oss en linjär utveckling. När en exponentiell utveckling studeras under en kort tid kan den verka linjär, men över lite större tidsskala så blir skillnaden enorm. Ett exempel på en relevant exponentiell utveckling är datorernas beräkningsförmåga. Kurzweil tar även upp priset på teknologiska komponenter samt vår förmåga att skanna och på så sätt kartlägga den mänskliga hjärnan som andra områden som förbättras exponentiellt Skillnaden hur datorer och mänskliga hjärnor funkar, till exempel binärt kontra analogt, säger han inte är ett problem då en tillräckligt snabb dator kan simulera hjärnan. Ytterligare en sak som tas upp som kan hjälpa till att möjliggöra denna utveckling är att datorer kan ladda ned och dela information med varandra i hastigheter långt bortom vad människor har möjlighet till. När de börjar bli intelligenta nog att utnyttja detta till fullo så kan det därför gå väldigt snabbt att bygga vidare på det (Kurzweil, 2005).

Många av dessa saker behöver ju såklart inte garanterat leda till singularitet och det finns flera anledningar att ifrågasätta den slutsatsen. Kanske kommer inte den exponentiella utvecklingen kunna fortsätta på grund av något hinder vi idag inte känner till (Kurzweil, 2005). Att det inom den närmsta framtiden kommer leda till mer avancerad artificiell intelligens är däremot inte särskilt kontroversiellt För diskussionen nedan kan vi därför föreställa oss datorer med förmågor från allt mellan dagens artificiella intelligens ända upp till förmågan hos en dator som uppnått singularitet. Med förbehållningen att det kanske är mindre troligt att någonsin realiserar ju närmare singulariteten vi närmar oss i nivå.

## 4.Diskussion

Här knyter jag ihop artificiell intelligens med kunskap. Jag diskuterar kraven på besittaren av kunskap, som ställs av de olika inriktningarna, och jämför detta med vad olika nivåer av artificiell intelligens kan sägas uppnå. Både det som finns idag utvärderas men även hypotetisk artificiell intelligens. De framtida artificiella intelligenserna kan i sin tur kan delas upp utifrån att antingen klara vissa specifika test för artificiell intelligens eller helt enkelt förutsägelser om framtida enormt

avancerade artificiella intelligenser. Även om dessa artificiella intelligenser är hypotetiska så fyller de ändå ett syfte för att undersöka våra gränser för när vi är villiga att tillskriva en dator förmågan att besitta kunskap.

#### 4.1 Kunskap som endast en Artificiella Intelligens kan besitta

En fundering som jag hade när jag började denna uppsats handlar om kunskap som endast en maskin besitter. Denna frågeställning är lite avskild från de andra och i följande stycke kommer jag försöka reda ut den, innan jag går vidare till att undersöka datorers förmåga att över huvud taget besitta kunskap. Kunskapsteori utgår oftast utifrån syftet att hitta ett synsätt om hur vi ser på, eller bör se på kunskap som vi själva tror oss ha. Ramverk, teorier om eller beskrivningar av vad som skiljer våra mer allmänna trosföreställningar från kunskap. Dessa teorier kan även ofta tillämpas för att avgöra om någon annan besitter kunskap eller inte. Om de goda skälen som denna andra person presenterar för oss accepteras och vi går med på att det är kunskap, så blir detta också kunskap för oss. Reliabilister skulle kanske vilja se eller själva utföra experimentet, om ett sådan låg till grund för kunskapen, innan de var beredda att acceptera de goda skälen eller benägna att själva tillskriva sig den kunskapen. Men när de goda skälen är accepterade så är det generellt avgjort.

Detta förutsätter dock att vi, som utvärderare själva kan förstå kunskapen eller resonemangen, och de goda skälen bakom. I Williams kontextualism så används orden dialektiska faktorer som syftar till att i ett visst sammanhang så är de endast en viss sorts ifrågasättande som räknas som relevanta. När det kommer till att avgöra om en expert på gravitationsvågor faktiskt besitter kunskap om ett specifikt påstående inom sitt område så är jag inte tillräckligt insatt för att meningsfullt ifrågasätta det. Jag kan inte veta om det är kunskap. Det finns däremot andra som är insatta och kan göra det, och jag kanske också skulle kunna det om jag valde att sätta mig in i ämnet.

Vad händer om det inte finns någon som kan göra det avgörandet? För den som besitter kunskapen kan kunskapsteorin hjälpa till med svar. Om en dator kan besitta kunskap så innebär detta att den också kommer besitta kunskap som

människor inte kan besitta. Som vi sett i beskrivningen av AI, så kan datorer göra saker som människor inte kan göra. Hantera problem som kräver alldeles för många termer för människor att hantera för att nämna ett exempel. Alltså, om de kan ha kunskap så kommer de kunna ha det om saker som vi människor inte kommer kunna förstå. Av det jag läst om kunskapsteori så verkar denna aspekt vara något som kan utvecklas och diskuteras mer. Möjligheten med avancerad artificiell intelligens som besitter kunskap, belyser att vi kanske ibland tagit vissa saker för givna. Vem ska vara ifrågasättaren i ett sådant sammanhang? En annan dator kanske. Om en internalist ber datorn förklara vilka interna motiv som ligger till grunden för kunskapen och datorn sedan på grund av premissernas storlek måste rada upp dem i flera hundra år så kan vi ju inte veta om det var kunskap eller inte. Samma sak gäller om reliabilismen följs och processen ska undersökas men denna process på något sätt är bortom mänsklig förmåga.

Människor kommer inte kunna avgöra om det faktiskt är kunskap eller inte då vi inte kan utvärdera de goda skälen, men är detta en anledning att inte acceptera det som kunskap? Kanske behövs ett förtydligande ord *övermänsklig* kunskap eller något i den stilen, för att referera till detta, där dom goda skälen finns men är bortom mänsklig förståelse. Vi säger ju generellt att experter, som håller på med sådant som vi inte är insatta nog i för att förstå, ändå har kunskap. Där sker ett antagande om att det finns någon som granskar och accepterar de goda skälen, eller åtminstone att det finns någon som skulle kunna förstå sammanhanget nog för att avgöra om det är rimliga skäl eller inte. Kanske bör vi ha samma inställning till dessa potentiella framtida datorer.

Floridis idé om information som ett alternativt sätt att definiera kunskap kan dock erbjuda ett svar på funderingen om kunskap som bara datorer kan utvärdera. Kraven på informationen i sammanhanget är att det ska finnas en redogörelse för hur den blev just information. För att en person ska kunna sägas ha kunskap måste denna känna till informationen och att redogörelsen finns. Det sägs dock inte att den måste veta exakt hur denna redogörelse ser ut. Detta är nyckeln. Om en dator tar fram ny kunskap, men metoderna är bortom vår förståelse, så finns ändå en redogörelse. Vi kommer inte kunna utvärdera om det garanterat är kunskap, men till

skillnad från andra synsätt på kunskap så möjliggör Floridis informationsbaserade beskrivning av kunskap ett sätt för oss att ta till oss kunskapen som vår egen.

## 4.2 Kan en Artificiell Intelligens besitta kunskap?

Går det att säga att en artificiell intelligens kan besitta kunskap? Det går alltid att säga, men kommer någon hålla med? Som avsnittet om kunskap visar så finns det många olika åsikter som säger många olika saker. Dessa är dessutom oftast specifikt utformade med människor i åtanke. Generellt är de flesta överens om att besittande av kunskap är möjligt eller att vi åtminstone bör använda ordet på ett sådant sätt att något vi människor tror, och som uppfyller de andra kraven på kunskap, faktiskt kan räknas dit. De inriktningar som inte går med på detta, framförallt tänker jag på en radikal form av skepticism, är knappast mindre strikta mot datorers förmåga till kunskap. Om vi dock föreställer oss att en superintelligent dator långt bortom vår förståelse skapas någon gång i framtiden så kanske det finns en möjlighet.

Denna tankelek är inte endast hypotetisk utan som vi såg i avsnittet ovan något som förutspås av flera experter. Säg att denna dator är så smart att den kan veta Allt. Då kan den till och med veta vilken inställning till kunskap som är den absolut sanna. Det vill säga om det går att ha kunskap, om det inte går att ha kunskap eller om det att det inte går att veta vilken av dessa två som är rätt. Detta tvingar dock slutsatsen att det går att ha kunskap, för om datorn som vet allt vet med säkerhet att det inte går att ha kunskap, så har den ju kunskap om det och därför kan det inte vara sant. Det vill säga en självmotsägelse. En dator som vet allt vet garanterat allt. Inte så intressant kanske då de flesta säkert ifrågasätter att det någonsin kommer existera en sådan dator men inte heller helt möjligt att utesluta. Hur påverkar det oss dock? I slutändan, eftersom vi inte är den datorn så vet vi ju inte om den faktiskt vet allt eller bara säger sig veta allt.

Detta tar oss till internalism kontra externalism. Kanske kan en internalist acceptera att en dator har kunskap. När den säger sig ha kunskap om något så har den tagit på sig ett epistemiskt ansvar. För att uppfylla sitt ansvar kräver vi att den

har en intern förklaringsmodell baserad på olika antaganden och dess samband. Alla dessa antaganden ska kunna presenteras som påståenden. Om datorn alltså även kan presentera dessa och de accepteras som giltiga av den som lyssnar så bör den också tänka att datorn har kunskap. Det är dock ingen garanti att detta är kunskap, eller att internalisten som undersöker datorns kunskapsförmåga är nöjd. Internalisten kräver ju att om vi vet något så vet vi att vi vet det. Med andra människor så görs ofta många antaganden om att deras medvetande existerar. Som en förklaring mot skeptikerna såg vi ett antal olika motiveringar till detta antagande. En av motiveringarna var att eftersom jag är människa och tänker är det också rimligt att anta att andra människor också tänker. Internalisten kanske inte är villig att sträcka detta till att inkludera datorer och skulle därför behöva hitta något annat sätt att försäkra sig om datorns medvetenhet. Abduktion som var ett annat sätt att motivera antagandet av andras medvetande som användes mot skeptiker kanske kan appliceras på datorer. Om de betar sig på ett sådant sätt att det bästa antagandet vi kan göra är att de har ett medvetande så får vi också anta det. Detta kräver kanske att internalisten förutom frågor om datorns motiv till sin kunskap även behöver ställa fler frågor för att bekräfta medvetandet.

Så hur ska vi då veta om datorn har kunskap eller bara säger så. Hur kan den övertyga oss? Turingtestet kanske är ett sätt. Om en artificiell intelligens klarar av det testet så kan vi inte skilja på en människa och en dator. Det finns som diskuterat många problem med detta. Framförallt kanske att frågorna är tillräckligt utmanande. Om vi föreställer oss en dator som pålitligt och regelbundet klarar testet även med de bästa frågorna. Vi pratar nu om mycket mindre avancerade datorer än de som närmade sig singulariteten, men fortfarande en bättre dator än det som finns idag. Har vi i detta sammanhang en dator som kan besitta kunskap?

Vissa skulle nog kunna hålla med om det. Att göra antagandet om dess medvetenhet skulle vara lättare än innan den klarar testet. På samma sätt som vi faktiskt behandlar gråtande robotar mer som människor än dockor så skulle antagligen en så avancerad artificiell intelligens som säger sig besitta kunskap behandlas på liknande sätt som en människas kunskap. Om en sådan dator sa att en komet kommer träffa jorden om 500 år och att detta riskerar att orsaka massdöd



så skulle vi antagligen undersöka lösningar på problemet på samma sätt som om en forskare kom på ett sådant påstående. Det i sig skulle nog de flesta hålla med om. Vad är skillnaden på den artificiella intelligensen och en miniräknare där en människa skrivit in värdena och fått fram resultatet. Vi skulle ju inte säga att det är miniräknaren som har kunskapen.

Att det är en i sammanhanget om kunskap, relevant skillnad mellan miniräknaren och den artificiella intelligensen, är nog inte lika många överens om. Båda behandlar något sant och berättigat. Kanske ligger en relevant skillnad i graden av berättigande. Miniräknaren är byggd så att den ska följa en pålitlig process vilket reliabilister skulle kunna räkna som ett berättigande, även om detta är ett gränsfall. Den artificiella intelligensen har möjlighet att förklara relevansen av siffrorna. Detta kan tillfredsställa naturalister som kan få bekräftat att processen som den artificiella intelligensen använt är pålitlig. I miniräknarens sammanhang så är det forskaren som är insatt i den relevanta processen. Internalisterna kan inte heller ställa frågor om vilka motiv miniräknaren har så som de kan göra med den artificiella intelligensen.

Den kanske viktigaste skillnaden är dock trosföreställningen. Detta saknas hos miniräknaren, men finns det hos den artificiella intelligensen? Detta blir alltid kärnan och att få ett enda korrekt svar är kanske omöjligt. Avsnittet ovan om intelligens var en förhoppning om att ge någon form av svar här. Som nu blir tydligt så hjälper det oss dock inte.

Men kanske behöver vi inte heller svara på den frågan på samma sätt som vi inte behöver veta helt säkert att andra människor har ett medvetande. Det verkar så till tillräckligt stor grad och vi behandlar andra människor som om dom har det. Så kanske kan vi säga att även om vi inte vet helt säkert så är det detta som gäller. Andra människor har ett medvetande. Jag tror att samma resonemang bör appliceras på avancerad artificiell intelligens. Vi kanske inte kan veta helt säkert att den är medveten, men det kommer vara vårt bästa antagande med en tillräckligt avancerad dator. Inte bara vårt bästa antagande utan också det vi faktiskt kommer att göra, vilket är väldigt relevant.

Williams kontextualism kan appliceras här på frågan om maskiner kan ha en trosuppfattning eller inte. De ekonomiska faktorerna kanske blir för stora med att

ifrågasätta kunskapen om maskiners trosuppfattning. Alternativt om vi väljer att ta till oss det för att det är sant och berättigat oavsett om det faktiskt är maskinens trosuppfattning. När kunskapens värde diskuteras så skiljs det på sann berättigad trosuppfattning och sann trosuppfattning. Jag har dock inte hittat mycket diskussion inom kunskapsteori om skillnaden mellan att det är en trosuppfattning eller inte. Kanske är detta ett sådant fall där artificiell intelligens kan belysa aspekter inom kunskapsteori som kan utforskas vidare.

Vad är ens något som är sant och berättigat men inte en trosuppfattning? Om vi ser till människor så kommer ett enkelt svar. Det är sant och berättigat men vi tror inte på det. Låt säga att jag inte tror på en klimatförändring orsakad av människan trots att jag har sett bevisen för att människans tillförsel av koldioxid till atmosfären påverkar detta och så vidare. Då är ju detta inte kunskap, men i det här sammanhanget så skulle jag ha gjort ett aktivt val att inte välja det som min trosuppfattning. Antingen genom att tro på motsatsen att det inte finns någon sådan klimatförändring orsakad av människor eller att helt enkelt säga att jag inte tror varken det ena eller det andra. Man kanske kan tänka sig att jag besitter någon form av kombination av trosuppfattningar i frågan men ett helt fjärde alternativ verkar inte finnas. Inte när det kommer till människor, men kanske för datorer.

Kan vi använda detta för att resonera kring datorns trosuppfattning? Om vi utgår ifrån att datorn inte ljuger kanske vi helt enkelt kan acceptera att det den presenterar som sin position i en fråga representerar dess trosuppfattning. Beroende på våra krav på berättigande så kan vi ge kunskap till ganska många datorer med denna ståndpunkt. Kanske för många. Vi är inte beredda att tillskriva en dator kunskap bara för att den säger sig tro något som är sant och som vi vet är berättigat.

Jag tror att det värde vi tillskriver kunskap spelar en stor roll i varför vi inte vill säga att alla datorer som säger saker som för oss är kunskap också besitter kunskapen. Om jag har kunskap om Pythagoras sats,  $a^2+b^2=c^2$ , så betyder det att jag inte bara känner till och förstår sambandet den beskriver utan att jag också vet vad a,b och c är samt hur jag mäter dem i en triangel. Jag kan använda det och dra nytta av kunskapen. Om en dator av idag säger sig ha kunskap om Pythagoras sats men inte har förmågan att sätta det i ett sammanhang och faktiskt dra nytta av

kunskapen har inte värdet av kunskapen och därför är det inte riktigt samma sak som kunskap.

Detta tror jag är kärnan för många och när datorer uppvisar en tillräckligt hög förståelse av sammanhanget så tror jag de flesta av oss kommer åtminstone behandla det de påstår som kunskap och kanske också tillskriva dem besittande av kunskap. Vissa datorer idag kan ju dock analysera en bild, förstå vad som är sidorna och sen räkna ut hypotenusan med hjälp av formeln. Kanske har den då, inom ett litet sammanhang visat tillräckligt med förståelse för sammanhanget. Innan Deep Blue kom så var vi redo att kalla en dator som klarar av schack för intelligent, men nu när den finns så har vi höjt ribban. Det kan vara så att kraven på datorer som säger sig besitta kunskap också kommer fortsätta höjas så att framtida datorer som vi idag skulle tillskriva kunskap inte godkänns.

### 4.3 Slutsats

Så vad har vi kommit fram till? Från allvetande dator till de vi lever med varje dag. Vissa kommer aldrig vara tillfredsställda med datorers försök att bevisa att de besitter kunskap. Samtidigt så är vissa skeptiker heller aldrig redo att gå med på att människor, eller att någonting alls för den delen, kan besitta kunskap. Eller att vi inte kan veta om så är fallet eller inte. De flesta skulle nog påstå att det dagens datorer sysslar med inte är kunskap. Kanske är detta just för att de finns. Exakt vilken motivering vi använder för att diskvalificera dem beror lite på vår inställning till hur kunskap bör definieras. Internalister kan bli övertygade när datorn kan antas ha ett medvetande. Vart exakt den gränsen går är dock oklart och vissa skulle inte ens bli helt övertygade om detta av en allvetande dator. Naturalister köper nog redan idag datorns goda skäl i vissa sammanhang. Men antingen krävs en större förståelse av sammanhanget från datorns sida eller något som visar att det faktiskt är en trosföreställning och inte bara text på en skärm för att det ska räknas som kunskap.

Slutsatsen är helt enkelt att det inte finns en tydlig gräns eller något som kan övertyga alla om en dators förmåga att besitta kunskap. Att säga att dagens datorer besitter kunskap är det nog dock väldigt få som vill göra, men kanske bör vi tänka

om och säga att de besitter en begränsad förmåga till kunskap. Frågan är hur långt in i framtiden de flesta av oss helt enkelt likställer deras kunskap med vår. I slutändan tycker jag det verkar bli en godtycklig gränsdragning och vi kan alla dra den som vi känner passar bäst. Min personliga rekommendation är dock att Floridi's informationsbegrepp används, dels på grund av Gettierproblemet, men även för att den är mer kompatibel med artificiell intelligens som besittare av information.

## Referenser

ANASTASI, A. (1992). What Counselors Should Know About the Use and Interpretation of Psychological Tests. *Journal of Counseling & Development*, 70(5), pp.610-615.

Deepmind.com. (2016). *AlphaGo | Google DeepMind*. [online] Available at: <https://deepmind.com/alpha-go> [Accessed 5 Sep. 2016].

DEETTERMAN, D. (2005). Robert J. Sternberg (Editor) 2004. International Handbook of Intelligence. Cambridge, UK: Cambridge University Press. ISBN 0-521-80815-4 (Hbk), 0-521-00402-0 (Pbk). Pp. xi + 496.#Detterman#2005#107#108#D. *Intelligence*, 33(1), pp.107-108.

Fisher, R. (2016). *Is it OK to torture or murder a robot?*. [online] Bbc.com. Available at: <http://www.bbc.com/future/story/20131127-would-you-murder-a-robot> [Accessed 5 Sep. 2016].

Johansson, L., Ekenberg, T., Masterton, G. and Remes, P. (2014). *Kunskapsteori*. Lund: Studentlitteratur AB

Kurzweil, R. (2005). *The singularity is near*. New York: Viking.

Lamb, E. (2016). Two-hundred-terabyte maths proof is largest ever. *Nature News*, [online] 534(7605), p.17. Available at: <http://www.nature.com/news/two-hundred-terabyte-maths-proof-is-largest-ever-1.19990> [Accessed 5 Sep. 2016].

Nature.com. (2016). [online] Available at: <http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html> [Accessed 5 Sep. 2016].

Nyu.edu. (2016). *"How Intelligent is Deep Blue?" by Drew McDermott*. [online] Available at: <http://www.nyu.edu/gsas/dept/philo/courses/mindsandmachines/Papers/mcdermott.html> [Accessed 5 Sep. 2016].

- Official Google Blog. (2016). *What we learned in Seoul with AlphaGo*. [online] Available at: <https://googleblog.blogspot.se/2016/03/what-we-learned-in-seoul-with-alphago.html> [Accessed 5 Sep. 2016].
- Phoebe Parke, f. (2016). *Is it cruel to kick a robot dog?*. [online] CNN. Available at: <http://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/> [Accessed 5 Sep. 2016].
- Saklofske, D., wishes, B. and Sattler, J. (2002). Corrections to: Assessment of Children: Cognitive Applications (4TH Edition) Jerome M. Sattler. *Canadian Journal of School Psychology*, 17(1), pp.115-116.
- Sample, I. (2009). *'Eureka machine' works out laws of nature*. [online] the Guardian. Available at: <http://www.theguardian.com/science/2009/apr/02/eureka-laws-nature-artificial-intelligence-ai> [Accessed 5 Sep. 2016].
- Scencedirect.com. (2016). *Investigations on empathy towards humans and robots using fMRI*. [online] Available at: <http://www.sciencedirect.com/science/article/pii/S0747563214000090> [Accessed 5 Sep. 2016].
- Suzuki, Y., Galli, L., Ikeda, A., Itakura, S. and Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, [online] 5, p.15924. Available at: <http://www.nature.com/articles/srep15924> [Accessed 5 Sep. 2016].
- The American Heritage dictionary of the English language. (2000). Boston: Houghton Mifflin.
- The World Book encyclopedia. (2006). Chicago: World Book.
- TURING, A. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236), pp.433-460.
- Waite, M. and Hawker, S. (2009). *Compact Oxford dictionary and thesaurus*. Oxford: Oxford University Press.

Www-03.ibm.com. (2016). *IBM100 - Deep Blue*. [online] Available at:

<http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/> [Accessed 5 Sep. 2016].