De Syntax-Semantiekredenering van Searle

Searles kritiek op Harde KI

Doctoraalscriptie Cognitieve Kunstmatige Intelligentie
Eline Spauwen
0133914

Eerste begeleider: Prof. J-J. Ch. Meyer
Tweede begeleider: Dr. M. Lievers
Utrecht, mei 2007

Abstract

John Searle heeft op meerdere manieren kritiek geuit op de meest ‘sterke’ vorm van KI: de Harde KI. De meest helder opgestelde redenering die hij beschrijft bestaat hoofdzakelijk uit drie aannames (axioma’s) en een hoofdconclusie, die stelt dat digitale computers geen minds kunnen bezitten of genereren. In mijn literatuuronderzoek heb ik bekeken hoe deze redenering, die in 'de consensus' (onder grote groepen wetenschappers) als ‘ongeldig’ of ‘irrelevant’ wordt beschouwd zowel in de praktijk van de KI als door velen in de filosofie van de KI, opgebouwd en onderbouwd is. De hoofdvraag van het onderzoek luidt: *‘Waarom is de syntax-semantiekredenering van Searle een probleem voor de Harde KI?’* Om de stelling *dat* het een probleem is te bekijken, heb ik vooral Searles positie en zijn dieperliggende aannames over minds en computers bekeken; waarom ziet Searle een probleem voor systemen in de Harde KI en hoe is dit probleem op een vergelijking met minds gebaseerd? Daarnaast heb ik kritieken gezocht die expliciet tegen aannames van de hoofdredenering ingaan, vooral wat Searles opvattingen over computers betreft, om deze vervolgens in debat met Searle te bekijken. Als mogelijk antwoord op de vraag *waarom* Searles kritiek (nog steeds) een probleem vormt, heb ik uit de analyse van Searles positie en de directe kritieken op zijn aannames, een aantal factoren die impliciet (op een niveau van andere, fundamentelere aannames) de redenering beïnvloeden, expliciet gemaakt. Deze (drie) kwesties bieden opheldering over de bijzondere (en onafgeronde) aard van het debat, omdat ze ofwel naar algemenere problematiek of naar nog te verkrijgen verscherpte inzichten verwijzen. Met deze bespreking van (het debat rond) Searles kritiek op KI heb ik kunnen concluderen dat zijn kritiek, in de vorm van zijn axiomatische redenering, op dit moment nog niet als ongeldig is te bestempelen. Daarnaast heb ik enkele redenen voor de complexiteit van het ontkrachten van zijn argumentatie expliciet kunnen maken.

Voorwoord

Voor u ligt mijn afstudeerscriptie voor de studie Cognitieve Kunstmatige Intelligentie. Deze scriptie gaat over een vraag en een persoon, die mij tijdens mijn studie voortdurend hebben geïntrigeerd: John Searle. Ik ben zelf overtuigd dat Searle een hele sterke redenering heeft neergezet, en dat hij 'gelijk' heeft; dit wil ik niet onder stoelen of banken schuiven. Toch bestaat er, ook onder CKI-studenten, de consensus dat Searle een 'vervelende' filosoof is waar we ons niet zoveel van zouden aan hoeven trekken[1]. Dit vreemde verschil in opvattingen, die mijns inziens gebaseerd zijn op het niet gronding willen bekijken van Searles ideeën, vormt mijn persoonlijke motivatie om juist het debat rond Searle als afstudeeronderwerp te kiezen.

De manier waarop ik het debat heb bekeken is dan ook echt mijn manier, dit is een van de redenen waarom ik blij ben dat ik dit onderzoek gedaan heb. Mijn eerste begeleider, Prof. John-Jules Meijer, wil ik bedanken voor zijn vertrouwen in 'mijn manier', omdat dit mijn vertrouwen in mijn manier ook heeft aangewakkerd. Voor iemand die in de daadwerkelijke praktijk van de KI werkt, is dit onderwerp waarschijnlijk lastiger dan voor 'pure filosofen' die niet in de praktijk van de KI bezig zijn, zoals Searle (en ikzelf, moet ik toegeven). Een van de redenen dat ik meer in de filosofie dan in de praktijk bezig ben geweest is waarschijnlijk mijn overtuiging dat filosofen zoals Searle goede argumenten tegen het werken in de praktijk hebben. Mijn tweede begeleider, Dr. Menno Lievers, is juist de filosofische criticus geweest die me heeft aangescherpt om precies datgene wat ik nu eigenlijk bedoelde (mijn manier) te kunnen verwoorden en uitwerken.

'Helaas' ben ik als student CKI geen 'echte' filosoof, zoals Menno Lievers, en ook geen 'echte' wetenschapper die bezig is in de praktijk, zoals John-Jules Meijer. Natuurlijk zit een 'echte' CKI-student altijd ergens tussen disciplines in. Ik ben er met dit onderzoek mijns inziens in geslaagd om 'acte de présence' te geven van de typische eigenschappen die een CKI-student dient te hebben (voldoende diepgang *met* behoud van breedte, het zogenaamde 'multidisciplinair karakter van CKI'). Natuurlijk is mijn dank ook groot aan degenen (CKI-studenten en anderen) die alle vormen van tekst die ik heb geproduceerd hebben willen bekijken en bekritiseren en mij met hun hulp hebben aangescherpt.

[1] Dit heb ik ervaren in de vele malen dat ik studenten, die het artikel en de redenering voor het eerst zagen, mocht uitleggen en duidelijk maken hoe het in elkaar stak, en hun reacties heb kunnen zien, in mijn student-assistentschappen voor filosofie van de Cognitiewetenschappen / filosofie van de Geest. Had ik toen maar de kennis en inzichten over de redenering gehad die ik nu heb!

Inhoud

0. Inleiding

Het onderzoek waarvan deze scriptie verslag doet, is gericht op een redenering waarmee John Searle een bepaalde variant van de Kunstmatige Intelligentie (KI) wil bekritiseren. Hij kaartte de problematiek die hij ziet voor deze variant, de Harde KI, in 1980 voor het eerst middels een intuïtief plausibel gedachte-experiment aan; de (axiomatische) redenering die hierbij hoorde heeft hij enkele jaren later expliciet uiteengezet. Het probleem dat Searle beschrijft is in mijn ogen bijzonder essentieel voor de fundamentele opvattingen in de KI. Mijn onderzoeksvraag draait dan ook om deze redenering en het debat dat eromheen ontstaan is. Mijn hoofdvraag komt voort uit de voortdurende discussie over Searles artikel uit 1980: *Minds, Brains and Programs*. Hierin heeft Searle kritiek geuit op de Harde KI, door het uiteenzetten van een gedachte-experiment waarin hij zichzelf probeerde voor te stellen op welke manier een computer zou kunnen begrijpen. Zijn conclusie luidt, dat er geen enkele vorm van begrip, enigszins gelijkend op menselijk begrip, te verwachten is in welk soort digitale computer dan ook, wanneer die computer dezelfde basis heeft als de huidige digitale computers. De mogelijke vernietigende invloed van deze conclusie op de Harde KI is hierin hopelijk al te herkennen.

Helaas (voor de voortgang van de wetenschap) is zijn redenering niet zodanig algemeen geaccepteerd of afgewezen, dat het nu, in 2007, duidelijk is of hij wel of geen gelijk heeft. Het debat is nog steeds aan de gang; er is zelfs in 2002 een boek verschenen dat geheel aan zijn gedachte-experiment en redenering gewijd is. Met zijn gedachte-experiment en redenering heeft Searle zelfs een van de invloedrijkste artikelen voor de filosofie van mind (geest) of filosofie van cognitiewetenschappen, filosofie van de KI (voor zover deze ‘filosofie’ een eigen status heeft), en ook taalfilosofie van de afgelopen eeuw aan filosofen en anderen gepresenteerd. De invloed van het artikel wordt in ieder geval wel algemeen geaccepteerd (en beklaagd). En, zoals dat filosofen betaamt, hebben zij zich er en masse voor en (vooral) tegen uitgesproken. De consensus binnen de praktijk van de (Harde) KI bestaat dat Searle ongelijk heeft; de genadeslag voor zijn argument is echter nog door niemand gegeven. Deze persoon zou daarmee zeker een beroemdheid of beruchtheid gelijk aan die van Searle verwerven. De complexe, onafgesloten status van het debat is het onderwerp en het interessegebied van mijn onderzoek. Hoewel het niet ‘strookt’ met een algemene consensus die bestaat in de praktijkwereld van de KI, acht ik Searles opvattingen

intuïtief zeer plausibel. Het is dan ook een doel van dit onderzoek om te bekijken in hoeverre zijn opvattingen standhouden ten opzichte van opvattingen vanuit verschillende invalshoeken van de KI (de filosofische, praktische en cognitiewetenschappelijke benaderingen). Het multidisciplinaire karakter van CKI speelt hierbij een belangrijke rol.

Omdat Searle een filosofische invalshoek heeft ten opzichte van de KI, is deze invalshoek voor dit onderzoek de meest relevante. Omdat Searles positie centraal staat in de onderzoeksvraag, is een filosofische beschouwing van zijn standpunten en de kritieken erop, en daarmee relevante conclusies over het debat rond zijn standpunten, het gewenste eindresultaat van dit onderzoek. De vraag die dit literatuuronderzoek aanstuurt luidt dan ook:

Waarom is de syntax-semantiekredenering van Searle een probleem voor de Harde KI?

Voor een analyse van bepaalde aspecten van het debat, en vooral voor een studie van Searles aannames en de reacties hierop, heb ik hierbij de volgende deelvragen gevormd:

-Wat is Harde KI? Wat is de manier waarop Searle de KI typeert en de status van dit paradigma? Wat is de claim van Harde KI in de huidige vorm?

-Wat is de *syntax-semantiekredenering* (SSR) van Searle? Wat zijn de aannames van Searle en hoe zijn deze te verdedigen en te bekritiseren?

-Heeft de redenering nog steeds effect, staat hij nog overeind? Wat zijn de redenen voor de bijzondere status van de redenering?

Ik denk dat deze vragen over Searles redenering een bijdrage kunnen leveren aan het debat, omdat deze analyse inzicht kan bieden in de complexe status ervan.

De methode die ik in dit onderzoek heb toegepast, is het bestuderen van literatuur; in hoge mate van Searle zelf omdat zijn invalshoek centraal staat, en natuurlijk ook van tegenstanders die de redenering aanvallen. De impliciete aanname van de hoofdvraag is *dat* de redenering daadwerkelijk een probleem is. Mijn analyse is gespitst op het kritisch bekijken in hoeverre Searles redenering standhoudt in het debat. Hiervoor heb ik gezocht naar tegenstanders die in mijn ogen en wellicht ook die van Searle op de juiste manier tegen zijn aannames ingaan. De juiste kritieken zijn degene die de aannames van zijn redenering letterlijk ontkennen. Dit type kritiek is het meest nuttig voor een fundamentele bespreking van de aannames. Omdat een rechtstreekse ontkenning van de redenering nog niet algemeen geaccepteerd is, is de

vraag die ik telkens bij de kritieken in de juiste vorm gesteld heb, welke andere aannames mogelijk zorgen voor het meningsverschil tussen Searle en de tegenstander. Mijn hypothese hierbij is, dat deze onderliggende aannames zo fundamenteel en nog ‘onuitgewerkt’ (in de wetenschap) of ‘überhaupt niet uit te werken’ (gebaseerd op kwesties van overtuiging die buiten de wetenschap vallen en dus eeuwig zullen blijven bestaan), dat deze doorwerken in het debat over de redenering van Searle die een conclusie erover voorlopig of überhaupt onmogelijk maakt. Vooral kwesties die neerkomen op een ‘welles’-‘nietes’-discussie, kunnen wezenlijk 'onmogelijk af te handelen’ lijken te zijn. Als het debat rond Searle gebaseerd is op kwesties die in ieder geval op dit moment en wellicht in de verre toekomst ook niet uitgewerkt kunnen worden, kan dit een verklaring zijn voor de onafgeronde status van *dit* debat. Dit is een hele sterke claim, en daarom ook bijzonder interessant. Ik hoop daarom met mijn onderzoek te kunnen suggereren van welke kwesties in de wetenschap het debat *ook* afhangt, om de vraag ‘*Waarom* is het probleem als zodanig’ te kunnen beantwoorden.

In hoofdstuk 1 bespreek ik de vragen rondom Harde KI als paradigma van de KI, en de status ervan. In hoofdstuk 2 bespreek ik vooral de expliciet uitgeschreven vorm van de redenering van Searle en de relatie met het gedachte-experiment, om inzicht te krijgen in waartegen de redenering gericht is en wat de relevante vormen van kritiek kunnen zijn. Tevens verantwoord ik hierin mijn keuze voor de naam ‘syntax-semantiekredenering’. In hoofdstuk 3 bespreek ik de achterliggende opvattingen van Searle bij zijn aanname over ‘minds’ en semantiek bij mensen, het essentiële verschil in opvatting met Daniel Dennett hierover, en twee opvattingen van andere auteurs die Searles positie kunnen ondersteunen. In hoofdstuk 4 bespreek ik de uitleg van Searle over syntax en semantiek en het toepassen van deze terminologie op de (Harde) KI en de problematiek die Searle voor de Harde KI beschrijft; de (relevante) kritiek hierop van William Rapaport en John Haugeland bespreek ik in hoofdstuk 5. Daarbij probeer ik te analyseren op welke (andere) aannames de kritieken van deze auteurs berusten. In hoofdstuk 6 bespreek ik wat de knelpunten van Searles redenering zijn en waarom deze wel of niet een probleem vormen voor Searle zelf; ik bekijk waar de bewijslast uiteindelijk komt te liggen. De algemene conclusies over de status en kracht van de redenering trek ik in hoofdstuk 7, de conclusie, waarin ik ook enkele suggesties tot verder onderzoek doe.

1. Wat is Harde KI?

De door Searle 'aangevallen' opvattingen binnen de KI zijn die van de Harde KI. Om de status van de redenering van Searle te bekijken, zal in dit eerste hoofdstuk de vraag 'Wat is Harde KI?' centraal staan: hoe kunnen we de Harde KI beknopt beschrijven in de meeste algemene vorm, en wat is de status van de Harde KI zoals Searle deze heeft 'gedoopt'? Hoe kunnen we de opvattingen die het meest worden aangevallen door Searle ('de Harde KI zoals hij het noemt') typeren?

1.1. Harde en Zwakke KI

In zijn oorspronkelijke Chinese Kamerartikel heeft John Searle als eerste de termen 'Strong AI' en 'Weak AI' gebruikt om een onderscheid binnen de KI aan te geven. (Minds, Brains, and Programs, 1980). 'Harde KI'[2] is het beste te beschrijven als een vertrouwen in de haalbaarheid van een doel, en dus het proberen te werken richting dat doel. Harde KI, zoals door Searle beschreven, beschrijft de visie dat wanneer een computer op de juiste manier geprogrammeerd wordt, we kunnen zeggen dat het een 'mind'[3] *is.* Van dergelijke systemen kunnen we zeggen dat ze *begrijpen* en andere cognitieve toestanden hebben (Searle, 1980). Ware intelligentie kan op via de juiste programmatuur manier bereikt worden. De gebruikte programma's vormen zo een volledige uitleg van psychologische processen. Deze visie is dus niet alleen voor 'pure' informatici relevant: de (cognitieve) psychologie kan van dergelijke systemen een inzicht krijgen in hoe in de *mens* deze cognitieve, psychologische processen plaatshebben.

Aanhangers van de 'Zwakke KI' (Weak AI) hebben niet de doelstelling om ware minds te creëren. Zij zien KI als een methode of richting om nuttige instrumenten te maken die vergevorderde (cognitieve) functies automatiseren, waarvan gedacht wordt dat alleen mensen het voor deze taken benodigde intelligentieniveau kunnen bereiken. In deze visie kan dus bijzonder veel gemaakt en geprobeerd worden zonder dat daarop fundamentele kritiek kan worden geleverd; aanhangers maken geen 'gevaarlijke' claims, of ontkennen zelfs dat kunstmatige

[2] Strong AI wordt hier vertaald met de gangbare doch niet zo geweldige vertaling 'Harde AI'. 'Sterke AI' is een meer letterlijke vertaling, beide dekken helaas niet de lading van 'strong'.
[3] Ik gebruik hier het Engelse 'mind', omdat hier mijns inziens geen adequate vertaling in het Nederlands voor is. Het woord 'geest' is een veelgebruikte vertaling, maar die dekt in mijn ogen een geheel andere lading en heeft een wezenlijk andere connotatie dan 'mind' in het Engels. In de rest van deze scriptie zal ik dan ook consistent dit Engelse woord in Nederlandse zinnen gebruiken.

intelligentie mogelijk is. Een betere manier om deze stroming te beschrijven is met de naam ‘cognitieve simulatie’ (Preston, 2002, p. 14 voetnoot).

Deze tweedeling door Searle, van oorsprong een filosoof, is door sommigen als ongeldig bestempeld: Harde KI zou geen valide categorie denkers bestempelen, niemand zou ‘toegeven’ deze opvattingen te hebben. Het is echter na verschijning van ‘Minds, Brains and Programs’ in 1980 duidelijk geworden dat Harde KI geen ‘straw man’[4] (een door Yorick Wills gebruikte term, 1982, bron: (Can Chinese Rooms Think? (Map 4))) is, vooral door het bestuderen van de uitspraken en stellingen van de pioniers van de kunstmatige intelligentie (Preston, 2002, p. 15) en (Can Chinese Rooms Think? (Map 4)).

1.2. Harde KI en stromingen binnen KI en Cognitiewetenschap

Het is belangrijk om te weten uit welke hoofdstromingen en methodes de (Harde) KI bestond toen Searle deze classificatie maakte, om duidelijk te krijgen tegen welke aannames hij zijn redenering richtte. In de volgende paragrafen worden deze paradigma’s en verschillende opvattingen kort beschreven. Preston (2002) geeft in zijn introductie tot het huidige debat over de Chinese Kamer van Searle een uitgebreide beschrijving van de basis van Kunstmatige Intelligentie, die Searle opsplitste in Harde en Zwakke KI. In deze paragraaf zal ik hieraan refereren en sommige delen van zijn introductie parafraseren.

1.2.1. Klassieke KI

De tijd tussen het benoemen van het onderzoeksgebied tot “Artificial Intelligence” door McCarthy in 1956 en het verschijnen van Searles artikel bestrijkt de hoogtijdagen van wat gezien kan worden als Klassieke KI, ook wel beschreven als symbolistische KI, ‘GOFAI’ (Good Old-Fashioned AI), ‘Representations and Rules theory’, of ‘Symbol-System AI’ (Cunningham, 2000, p. 191). De pioniers van de KI waren ook bekend met een niet-symbolistische, *connectionistische* manier, maar de symbolistische aanpak is als eerste door velen toegepast. Het is dus is niet zo dat klassieke, symbolistische KI verder terug gaat dan het connectionisme (Preston, 2002,

[4] Een ‘straw man’ redenering (stropopredenering) is een vorm van drogreden, een logische redeneerfout die gebaseerd is op een verkeerde weergave van de opvatting van een tegenstander. Het toeschrijven van een verkeerde representatie van de redenering van een tegenstander wordt in een stropopredenering gebruikt om de redenering te ontkrachten. Omdat de positie van de tegenstander echter onjuist wordt weergegeven, is een stropopredenering dus een misleidende en onvruchtbare strategie van ontkrachten. Bron: (Straw man) en (Stropopredenering), Wikipedia.

pp. 11-13), maar de paradigmaverschuiving van klassieke naar niet-klassieke, connectionistische KI, heeft pas na enkele decennia plaatsgevonden.

De klassieke KI maakt gebruik van 'representaties en regels': klassieke KI-theorieën werden en worden voornamelijk uitgevoerd in systemen die uit deze twee hoofdcomponenten bestaan: representaties (symbolen), en regels om de symbolen te bewerken. In principe zijn de symbolen van deze representaties niet gedefinieerd door datgene waaraan ze refereren: er bestaan geen intrinsieke connecties tussen symbool en referent. De interpretatie van de symbolen is vrij (Cunningham, 2000, p. 192).

Allen Newell en Herbert Simon behoren tot de pioniers van de KI. Hun '*Symbol System Hypothesis*', een belangrijk basisbegrip voor hoe fysieke systemen moeten kunnen redeneren, luidt: redeneren is symboolmanipulatie (Poole, Mackworth, & Goebel, 1998, p. 4). Via de Church-Turing thesis, die stelt dat elke symboolmanipulatie op een Turing Machine[5] kan worden uitgevoerd, kan worden aangenomen dat redeneren in een symbolisch, computationeel systeem gebaseerd op de Turing Machine, mogelijk is (1998, p. 4). Gemakshalve is te zeggen dat de huidige traditionele computersystemen gebaseerd zijn op de ideeën van Alan Turing; Turing is de bedenker van een abstract 'apparaat' (de Turing Machine).

Poole e.a. plaatsen wel een kanttekening bij de Symbol System Hypothesis en de notie van computatie:

> 'This hypothesis doesn't imply that every detail of computation can be interpreted symbolically. Nor does it imply that every machine instruction in a computer or the function of every neuron can be interpreted symbolically. What is does mean is that there is a level of abstraction in which you can interpret reasoning as symbol manipulation, and that this level can explain an agent's actions in terms of its inputs' (1998, p. 5).

Dit abstractere niveau waarop redeneren als symboolmanipulatie te interpreteren is, is dus het niveau waarop 'agents' (intelligente actoren) geprogrammeerd en geïmplementeerd worden.

1.2.2. Computationalisme en functionalisme

De concreetste[6] aannames waartegen Searle argumenteert, zijn de opvattingen van het functionalisme en het computationalisme over minds en computers:

---

[5] De detaillistische beschrijving van een Turing Machine laat ik hier achterwege, omdat ik deze bekend veronderstel; de relevante eigenschappen van een Turing Machine zullen, wanneer nodig, aan bod komen. Zie voor een uitgebreidere beschrijving Copeland (1993, pp. 134-135).

[6] Concreet staat hier voor 'operationeel bruikbaar binnen de KI': opvattingen die filosofisch onderbouwd zijn maar ook beschrijvingen geven van hoe de KI hier in de praktijk mee aan de slag kan.

> ‘One could summarize this view - I call it ‘strong artificial intelligence, or ‘strong AI’ – by saying that the mind is to the brain, as the program is to the computer hardware’ (Searle, 1984, p. 28).

Deze computeranalogie voor de werking van de hersenen is kenmerkend voor het (Turing Machine) functionalisme. De uitgangspositie van het **functionalisme** is de opvatting dat *mentaliteit* gebaseerd is op functionele processen, en niet op fysische processen. De functionele toestand van een systeem bepaalt de mentale toestand; hoe deze functionele toestand fysisch is gerealiseerd is hiervoor niet relevant. Deze opvatting maakt het mogelijk om niet alleen de mens, maar ook de digitale computer, gebaseerd op het principe van de Turing Machine, te bekijken als een ‘ding’ dat mentale toestanden kan hebben, middels een functionele beschrijving (Preston, 2002); mentale toestanden kunnen zo multipel realiseerbaar zijn (hier komen we in hoofdstuk drie op terug).

Het aan functionalisme gerelateerde **computationalisme** is een opvatting binnen de Cognitiewetenschap (‘Cognitive Science’), met Jerry Fodor als een van de pioniers[7]. Cognitiewetenschap, en daarmee ook de Kunstmatige Intelligentie, houdt zich bezig met de studie van intelligent, coherent en rationeel denken, volgens Fodor (Preston, 2002). Fodor gebruikt het model van Turing, om zo de essentie van denken en cognitie te definiëren als computationele transformaties van mentale representaties. Deze representaties zijn symbolen met **semantische** en **syntactische** kenmerken. De transformaties verlopen op grond van de syntactische (**vorm**-) kenmerken van mentale representaties, de semantische (**inhoudelijke**) kenmerken zijn irrelevant voor het transformatieproces. De semantische kenmerken worden in de transformatieprocessen wel in stand gehouden, omdat de processen verlopen volgens strikt logische processen. Deze processen zijn bewijstheoretisch onderbouwd; de waarheidswaardes (waarin de ‘inhoud’ van de mentale representaties ligt) worden met behoud van logische consistentie in de transformaties bijgehouden (Preston, 2002):

> ‘The basic idea in cognitive science is the idea of proof theory, that is, that you can *simulate* semantic relations – in particular, semantic relations among thoughts – by syntactical processes. That is what Turing suggested, and that is what we have all been doing in one or the other area of mental processing’ (Fodor, 1995, p. 88)(mijn nadruk).

Let wel: Fodor ondersteunt het doel van KI niet: hij ziet geen heil in het maken van machines en het simuleren van cognitie. Hij wil de menselijke cognitie onderzoeken

[7] Hilary Putnam wordt, naast Fodor, bestempeld als ‘bron’ van het ontstaan van een empirische computationele theorie van mind (Block, 1994, p. 323), omdat hij (ook) de vergelijking tussen mentale toestanden en computationele toestanden van een computer maakte (Lycan, 1994, p. 318).

via computationele modellen van mind, niet namaken. Hij ziet het nut van de pogingen, maar vindt niet dat op die manier cognitie echt onderzocht kan worden (Fodor, 1995, p. 87). Hoewel het functionalisme dat gebruik maakt van Fodors ideeën als theoretische achtergrond gebruikt wordt in Harde KI, is Fodor niet als Harde KI-wetenschapper te zien.

1.3. Huidige status van Harde KI

De redenering tegen Harde KI is uiteraard geen relevante redering (meer), als Harde KI als zodanig beschreven geen gangbare werkwijze of doelstelling meer is. Hoewel er ook wordt gesuggereerd dat Harde KI nooit een valide 'stroming' heeft aangeduid (een 'straw man' is), is het, er vanuit gaande dat dit wel ooit het geval was, een vraag of er nu nog over Harde KI gesproken kan worden, en of de huidige opvattingen afgezwakte of juist sterkere claims inhouden dan twintig jaar geleden.

1.3.1. Tegenwoordig: Wel of geen Harde KI?

Als Searles redenering echt zo goed is, is er dan vijfentwintig jaar na dato nog wel sprake van mensen die Harde KI verdedigen en in praktijk proberen te brengen? Het zou een logisch gevolg van de conclusies van Searle en de voortdurende discussie over de Chinese Kamer kunnen zijn, dat vele wetenschappers hun doelstelling op zijn minst bijgesteld hebben. Searle zegt in een interview dan ook, dat hij wetenschappers ervan verdenkt hier niet open voor uit te komen:

> 'I think, in fact, that today very few people defend Strong Artificial Intelligence. Of course, they do not say that they have changed their mind, but they have. I do not hear as many extreme versions of strong Artificial Intelligence as I used to' (1995, p. 205).

Dat Harde KI een gangbare opvatting is gebleven, is onder andere te zien in de volgende opvatting van Poole e.a.: 'The central scientific goal of computational intelligence is to understand the principles that make intelligent behavior possible, in natural or artificial systems. The main hypothesis is that reasoning is computation. The central engineering goal is to specify methods for the design of useful, intelligent artifacts' (1998, p. 1). Poole e.a. noemen het onderzoeksgebied zelfs Computationele Intelligentie, omdat 'kunstmatig' een bron van verwarring zou vormen.

Poole e.a. (1998, p. 3) maken wel onderscheid tussen Computationele Intelligentie en andere Cognitiewetenschappelijke disciplines; in de Computationele Intelligentie worden hypotheses over de aard van intelligentie *getest* door machines te maken die intelligent zijn en meer kunnen dan slechts mensen nabootsen. Het is de

bedoeling dat deze machines de mens zullen gaan evenaren in intelligent gedrag. Ook zij zeggen letterlijk dat ze redeneren als computatie zien; dit is dus dezelfde sterke aanname als die van de klassieke, Harde KI. Zie voor andere huidige voorstanders van Harde KI bijvoorbeeld ook: (Russell & Norvig, 1995), Stan Franklins '*Artificial Minds*' (1995); in zijn voorwoord beschrijft Franklin dat hij ons wilt rondleiden in het 'nieuwe paradigma van mechanisms of mind' (p. ix), en werken van Aaron Sloman en John McCarthy.

1.3.2. Benoemen van 'het project'

De verwarrende invloed van de termen 'kunstmatig' en 'intelligentie' wordt niet alleen door Poole e.a. erkend, zie bijvoorbeeld ook (Cunningham, 2000, p. 191). Ook Selmer Bringsjord geeft aan dat de hoofdterm 'kunstmatige intelligentie' en de verschillende kampen binnen KI het niet gemakkelijk maken om een algemene uitspraak te doen over wat KI precies is. Hij stelt voor om een neologisme te vormen om te refereren aan deze 'inter- and multidisciplinary collection of camps and fields' (Bringsjord, 1992, p. 5). Hij kiest voor 'Cognitive Engineering'. Deze term is, toegegeven door Bringsjord, niet geheel neutraal, maar dekt wel goed de lading. Hij noemt het Cognitive Engineering project het project om een persoon te bouwen: het Person Building Project (PBP). Dit is *geen idealisatie* van bestaande doelen, zo hebben bijvoorbeeld Charniak en McDermott letterlijk beschreven: '"The ultimate goal of AI research (which we are very far from achieving) is to build a person, or more humbly, an animal' [Charniak & McDermott (1985), p.7]"[8] ' (Bringsjord, 1992, p. 7).

John Pollock heeft deze woordkeus al eerder gebruikt:

> 'My general purpose in this book is to defend the conception of man as an intelligent machine. Specifically, I will argue that mental states are physical states and persons are physical objects' (1989, p. 1).

In het voorwoord van ditzelfde boek schrijft hij bovendien: 'This book is a prolegomenon to the enterprise of building a person' (p. viiii). Hij doet in dit boek een voorstel voor een daadwerkelijk Person Building Project: OSCAR.

Ook Preston geeft aan dat de aanpak van de kunstmatige intelligentie is veranderd door Searles redenering:

---

[8] Verwijzing Bringsjord: E.Charniak & D. McDermott, 1985. *Introduction to Artificial Intelligence*. (Reading, MA: Addison-Wesley).

> ‘It *[het Chinese Kamer gedachte-experiment]* can also be used to raise large methodological questions about how cognitive science should be done (computationalism versus ‘cognitive neuroscience’, versus **some more person-centered alternative**?), [..]’ (2002, p. 47)(mijn nadruk).

Het centraal stellen van ‘persoon’ als onderzoeksthema en als datgene dat gemaakt zou moeten worden, is een nieuwe conceptuele aanpak. Deze aanpak is echter niet drastisch anders dan stellen dat een ‘mind’ gemaakt moet worden, of ‘intelligentie’ bereikt moet worden, als ‘een mind hebben’ en de ‘persoonstatus’ als onderling afhankelijk worden gezien.

Het ‘Person Building Project’ als naam biedt duidelijkheid over het doel, het willen ‘bouwen’ van een persoon, en daarnaast het een project is met de aanname dat het mogelijk om het doel te bereiken. Het is niet zo dat het een direct doel is, maar wel een ultiem doel:

> ‘But the fact is, when ambitious members of the field get together to talk about ultimate goals, personhood invariably pops up as the prime candidate’ (Bringsjord, 1992, p. 7).

De belangrijkste aanname van het PBP is: ‘Persons are automata’ (Bringsjord, 1992, p. 8). ‘Automaton’ is een abstracte notie: het is mogelijk om hier ‘digitale computer’, of ‘universele Turing Machine’, of iets dergelijks in te vullen. Hierin is het abstracte aspect van het computationalisme en het functionalisme terug te vinden; de functie (de computatie) staat los van de implementatie (het principe van multipele realiseerbaarheid). Deze ‘nieuwe’ naam is dus een bijgewerkte en duidelijkere benoeming van het ultieme onderzoeksgebied van Harde KI.

1.4. Simulatie versus duplicatie.

Het verschil tussen simulatie en duplicatie van processen en fenomenen (zoals die van minds), is een belangrijk verschil voor Harde KI. Voor nu moet het duidelijk zijn waarom en hoe deze twee begrippen van elkaar verschillen. Niet zelden blijkt een verschil in opvattingen te zijn gebaseerd op de opvatting over de relatie tussen deze twee. Bij het simuleren van een proces wordt het proces nagebootst, en niet gedupliceerd. ‘Simulatie’ houdt dus in dat er iets werkelijks of intrinsieks zou kunnen missen. Bij het dupliceren van een proces kan er geen twijfel zijn over de echtheid of mate van gelijkheid aan het originele proces.

Harde AI moet duplicatie nastreven. Doet men dat niet, dan is het de vraag of systemen die slechts minds simuleren, serieus moeten worden genomen. Is er wel sprake van ‘echte’ inhoud, van een ‘persoon’? Searle gebruikt een analogie van

computerprogramma's, die regenstormen simuleren, om dit aannemelijk te maken (1980, p. 249). Het is volgens hem niet zo dat door gebruik van een dergelijk programma er daadwerkelijk ergens regen uit de lucht komt vallen (met alle gevolgen van dien).

Een andere belangrijke opvatting die vaak centraal staat is dat minds niet kunnen bestaan bij de gratie van interpretatie *alleen*. Dit is niet algemeen geaccepteerd, Daniel Dennett deelt deze opvatting niet; zijn positie komt in hoofdstuk drie uitgebreider aan bod. Een simulatie kan wellicht als mind geïnterpreteerd worden, maar dit zegt niets over de intrinsiek inhoudelijke status van de simulatie. In de verschillende citaten die tot nu toe zijn gebruikt is vaak al aan de woordkeuze te zien dat er over simulatie en vaak interpretatie gesproken wordt, alsof deze garant staan voor een intrinsiek inhoudelijke status. Dit is volgens Searle geen correcte manier om over computers te oordelen:

> 'Such intentionality as computers appear to have is solely in the minds of those who program them and those who use them, those who send in the input and those who interpret the output' (1980, p. 248).

Hier komt al een stuk van de redenering van Searle aan het licht, waarmee hij de relevantie van 'intrinsieke kenmerken' in tegenstelling tot 'kenmerken onder interpretatie' wil aangeven. Deze redenering komt in hoofdstuk vier uitgebreid aan bod.

1.5. Conclusie

De basis van de Harde KI waartegen Searle redeneert, is het computationalisme; dit paradigma vormt het meest een concreet theoretisch kader voor de praktijk van de Harde KI. Bij het computationalisme in relatie tot minds, hoort de filosofische stroming van het computationeel functionalisme. De aannames van deze stromingen waren en zijn nog steeds opvattingen die gebruikt worden om in de Harde KI minds te maken. Het staat niet vast dat er sprake is van een drogredenering van Searles kant als hij het over zijn representatie van Harde KI heeft, en daar zijn redenering tegen richt: Harde KI zoals hij deze beschrijft is geen straw man. Er zijn voldoende wetenschappers die de doelen van de Harde KI op dezelfde manier beschrijven. Een (mogelijk) neologisme voor het (ultieme) doel van de Harde KI is het 'Person Building Project'. Juist onder deze naam is het onderzoeksgebied van de Harde KI gevoelig voor de argumentatie van Searle, die ingaat tegen de abstractere opvatting

dat minds of personen (niet synoniem maar wel gerelateerde) met functioneel en computationeel gespecificeerde systemen gemaakt kunnen worden. In de Harde KI dient te worden gestreefd naar duplicatie van het fenomeen mind; duplicatie (en niet slechts simulatie) is deel van het 'ultieme' doel. Slechts simulatie is geen garantie voor het maken van daadwerkelijk als persoon te beschouwen systemen: simulatie is geen voldoende voorwaarde voor duplicatie. Het nog steeds bestaande en concreet te beschrijven 'doelwit' van Searles redenering is dus Harde KI, en zouden we ook kunnen zien als het de (ultieme) onderneming om een persoon te bouwen (dupliceren). Nu we weten welke stromingen Searle wil bekritiseren, kunnen we bekijken en begrijpen wat die kritiek van Searle is.

2. Wat is de syntax-semantiekredenering?

Nu we weten waar ‘Harde KI’, waartegen Searles redenering gericht is, voor staat, is het mogelijk en relevant om eerst te bekijken wat de vorm en inhoud van de redenering precies zijn. Welke (axiomatische) vorm kunnen we het best bekijken en gebruiken als heldere en te weerleggen redenering? Welke rol is hierbij voor Searles gedachte-experiment weggelegd? Welke aannames doet Searle om tot zijn conclusie te komen, en hoe zijn deze het beste weer te geven? Welk type kritiek sluit aan op die aannames, en welk type in ieder geval niet? Daarnaast is het de vraag wat er binnen en wat er buiten de reikwijdte van de redenering valt, en hoe het in welke termen de Harde KI door Searle beschreven wordt. Dit zijn de vraagstukken die in de afzonderlijke paragrafen van dit hoofdstuk aan bod komen.

2.1. De logische opbouw van de redenering

Searles redenering tegen de mogelijkheid van Harde KI wordt door hem zelf op verschillende manieren uiteengezet. Om de redenering te kunnen bekijken en bekritiseren op het juiste niveau van argumentatie, is het nodig de structuur van de redenering uit te schrijven. Wat zijn de premissen of aannames (axioma’s) van de redenering, en wat is de (hoofd)conclusie? Als deze expliciet worden gemaakt, is het ook duidelijk welke kritiek expliciet tegen de redenering ingaat, en welke kritiek dat niet doet. De ‘juiste’ vorm van kritiek kan dan getypeerd worden. In het gedachte-experiment beschrijft Searle een scenario, welk punt wil hij hiermee aantonen? Allereerst wil ik duidelijkheid verschaffen over Searles gebruik van de Chinese Kamer als ‘redenering’ en daarna bekijken welke interpretatie of uitwerking ervan het ondersteunendst is voor Searles positie.

2.1.1. Chinese Kamer gedachte-experiment

Hieronder volgt de verhalende opzet van het scenario van het Chinese Kamer gedachte-experiment van Searle.

> ‘Suppose that I’m locked in a room and given a large batch of Chinese writing. Suppose furthermore (as is indeed the case) that I know no Chinese, either written or spoken, and that I’m not even confident that I could recognize Chinese writing as Chinese writing distinct from, say, Japanese writing or meaningless squiggles. To me, Chinese writing is just so many meaningless squiggles. Now suppose further that after this first batch of Chinese writing I am given a second batch of Chinese script together with a set of rules for correlating the second batch with the first batch. The rules are in English, and I understand these rules as well as any other native speaker of English. They enable me to correlate one set of formal symbols with another set of formal

> symbols, and all that "formal" means here is that I can identify the symbols entirely by their shapes. Now suppose also that I am given a third batch of Chinese symbols together with some instructions, again in English, that enable me to correlate elements of this third batch with the first two batches, and these rules instruct me how to give back certain Chinese symbols with certain sorts of shapes in response to certain sorts of shapes given me in the third batch. Unknown to me, the people who are giving me all of these symbols call the first batch a "script," they call the second batch a "story," and they call the third batch "questions." Furthermore, they call the symbols I give them back in response to the third batch "answers to the questions," and the set of rules in English that they gave me, they call the "program." Now just to complicate the story a little, imagine that these people also give me stories in English, which I understand, and they then ask me questions in English about these stories, and I give them back answers in English. Suppose also that after a while I get so good at following the instructions for manipulating the Chinese symbols and the programmers get so good at writing the programs that from the external point of view—that is, from tile point of view of somebody outside the room in which I am locked—my answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at my answers can tell that I don't speak a word of Chinese. Let us also suppose that my answers to the English questions are, as they no doubt would be, indistinguishable from those of other native English speakers, for the simple reason that I am a native English speaker. From the external point of view—from the point of view of someone reading my "answers"—the answers to the Chinese questions and the English questions are equally good. But in the Chinese case, unlike the English case, I produce the answers by manipulating uninterpreted formal symbols. As far as the Chinese is concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purposes of the Chinese, I am simply an instantiation of the computer program.
> Now the claims made by strong AI are that the programmed computer understands the stories and that the program in some sense explains human understanding. But we are now in a position to examine these claims in light of our thought experiment.' (1980, pp. 236-237).

Dit is de verhalende opzet van Searles eerste omschrijving van zijn kritiek op Harde KI, het Chinese Kamer gedachte-experiment. Wat is nu de door Searle bedoelde interpretatie van dit gedachte-experiment?

2.1.2. De interpretatie van de Chinese Kamer redenering.

In het oorspronkelijke Chinese Kamerartikel van Searle wordt het gedachte-experiment van de Chinese Kamer uiteengezet om een stelling over computers aannemelijk te maken. Het gedachte-experiment is echter sterk intuïtief: Driessen beschrijft dit als een gedachte-experiment waarin de oppervlakteredenering, dat wil zeggen de redenering in het verhaal dat het experiment beschrijft, een geheel andere structuur heeft dan de dieperliggende redenering (2002, p. 39). Deze dieperliggende redenering is erg diep, eigenlijk te diep, verstopt in de verhalende opzet. De *conclusie* van de oppervlakteredenering is een *aanname* in de dieperliggende redenering. Dit maakt de verhalende opzet tot slecht of geen bewijs voor de dieperliggende

redenering (Driessen, 2002). Dit is overigens geen 'gebrek' van de opzet van het gedachte-experiment van Searle; hij claimt niet dat het gedachte-experiment bedoeld is om bewijs te leveren. Wat de bedoeling ervan is, is het aannemelijk maken, illustreren (Searle, 1984, p. 31) dan wel demonstreren van een aanname:

> 'Now let me add the point that the Chinese room demonstrated. Having the symbols by themselves--just having the syntax--is not sufficient for having the semantics' (1990, p. 21).

Deze zin gaat vooraf aan het expliciet stellen van zijn een van de aannames, die hij voor de dieperliggende redenering gebruikt. Het gedachte-experiment is dus een intuïtief appel op de waarheid van een aanname van de dieperliggende hoofdredenering, zie hiervoor ook Driessen (2002).

Omdat het gedachte-experiment dus niet de hoofdredenering in zijn geheel ondersteunt, wil ik de verhalende opzet niet benadrukken door deze uitgebreid te bespreken, maar juist die onderliggende redenering. Gelukkig is deze redenering door Searle meerdere keren ook afzonderlijk van het Chinese Kamer Gedachte-experiment uiteengezet. Deze redenering is de redenering die ik aanduid met '**de syntax-semantiekredenering'** (hiervoor zal ik de afkorting 'SSR' gebruiken). De verwarring met de 'redenering' van het gedachte-experiment is de reden dat ik het liever niet aanduid met de 'Chinese Kamer'-redenering, zoals gebruikelijker is. Deze aanduiding is niet specifiek genoeg.

Jack Copeland schrijft, onder andere in '*The Curious case of the Chinese Room*' (1993, pp. 125-130) en '*The Chinese Room from a Logical Point of View*' (2002), over de 'logica' van de Chinese Kamer. Hij heeft het echter ook slechts over het gedachte-experiment (de logica van de redenering in de oppervlaktestructuur) en stelt dat Searle dit scenario als bewijs gebruikt voor zijn redenering:

> 'Moreover, even if some version of the argument were sound, the argument could not possibly establish – as Searle claims it does – 'his key thesis that whatever is 'purely formal' or 'syntactical' is neither constitutive of nor sufficient for mind' (Copeland, 2002, p. 109).

Het gedachte-experiment is echter geen bewijs voor een stelling, maar een intuïtieve demonstratie van een van de stellingen van de hoofdredenering. Copeland geeft aan dat Searles 'redenering' in het scenario ongeldig is, omdat hij een onverantwoorde generalisatie vanuit één geval naar alle gevallen of vanuit een deel naar een geheel maakt (2002, p. 110). De manier waarop Copeland het gedachte-experiment beschrijft, is zodanig dat hij een geldig argument ertegen kan maken; hij kan aantonen

dat de ‘redenering’ in de verhalende opzet niet valide is. Deze (mis)representatie van het doel van de verhalende opzet doet vermoeden dat Copeland een ‘straw position’ aan dit doel toeschrijft. Searle heeft namelijk niet beweerd dat deze verhalende opzet een dergelijke redenering bevat: hij wil met het scenario een aanname van de onderliggende redenering ‘demonstreren’. De strategie van Copeland, die de algemene conclusies die Searle trekt ontkent omdat de logische structuur van het gedachte-experiment wankel is, is dus niet bruikbaar als tegenargument tegen de dieperliggende redenering. Tegen die redenering gaat hij niet in, niet op een manier die het tegendeel betoogt (Preston, 2002, p. 29). Hij beklaagt slechts de oppervlaktestructuur van de Chinese Kamer redenering. Searles argumentatie is het sterkst als Searle, sprekend over de ‘Chinese Kamer redenering’, verwijst naar de onderliggende SSR en deze verdedigt. De relevante kritiek op Searle is dan ook kritiek die deze redenering aanvalt, en hierbij het gedachte-experiment niet centraal stelt.

2.1.3. De dieperliggende redenering

Searle geeft in ‘*Minds, Brains and Science*’ (1984, p. 39), een serie lezingen, voor het eerst (Preston, 2002, p. 28) zijn dieperliggende redenering voor de syntax-semantiekredenering kort en helder weer. Ook in het artikel ‘*Is the brain’s mind a computer program?*’ (1990) schrijft Searle de redenering uit. In beide publicaties trekt hij twee conclusies uit vier axioma’s, al verschilt de volgorde van het opstellen van de redenering in de twee publicaties. De nummering hieronder is zoals Searle deze in 1990 gebruikte. Ter introductie van de aannames heb ik de uitleg van de aannames uit zowel de lezingen als het artikel in letterlijke vertalingen gebruikt, en deze zelf niet van extra conclusies of commentaar voorzien. Dit is dus allemaal Searles eigen uitleg, door mij vertaald.

> **Aanname 1) Computerprogramma’s zijn geheel formeel (syntactisch) gedefinieerd.**
>
> Informatieprocessen in een digitale computer bestaan uit het manipuleren van gecodeerde symbolen volgens een voorgedefinieerde set regels. Deze regels vormen de essentie van het computerprogramma. De symbolen en programma’s zijn abstracte noties; ze hebben geen essentiële fysieke eigenschappen en zijn multipel realiseerbaar. De symbolen worden gemanipuleerd zonder referentie aan enige betekenis; de maker of gebruiker kan elke gewenste betekenis aan de

symbolen geven. In deze uitleg heeft het programma syntax maar geen semantiek (1984, p. 39), (1990, p. 21).

Aanname 2) Menselijke 'minds' hebben mentale inhoud (semantiek).

Dit is (voor Searle) een duidelijk feit. Fenomenen als gedachten, waarnemingen, begrip, enzovoorts, hebben een mentale inhoud (1984, p. 39), (1990, p. 21).

Aanname 3) Syntax op zichzelf is noch constitutief noch voldoende voor semantiek.

Er bestaat een onderscheid tussen formele elementen zonder intrinsieke betekenis of inhoud, en fenomenen die wel intrinsieke inhoud hebben. Dit is een conceptuele waarheid (1984, p. 39), (1990, p. 21).

Conclusie 1) ***(uit aanname 1 t/m 3)*** **Een computerprogramma is op zichzelf niet voldoende om een systeem een mind te geven.**

Programma's zijn geen minds en zijn niet voldoende om minds te creëren. Dit is een erg krachtige conclusie, omdat dit betekent dat het project van het proberen te creëren van minds door slechts het ontwerpen van programma's vanaf het begin gedoemd is. Het is een manier om te zeggen dat Harde KI 'false' is (1984, p. 39), (1990, p. 21).

Aanname 4) Hersenen veroorzaken minds.

Alle mentale fenomenen die wij beschouwen als constitutief voor minds worden geheel veroorzaakt door neurofysiologische processen in het brein (1984, p. 39), (1990, p. 23).

Conclusies 2 en 3) ***(uit aanname 4 en conclusie 1)*** **Om een mind te kunnen veroorzaken, moet een systeem dezelfde causale krachten / invloeden hebben als de hersenen.** *en* **Een artefact dat mentale fenomenen veroorzaakt, moet in staat zijn om de specifieke causale krachten van het brein te dupliceren, dit kan een artefact niet door slechts een formeel programma te doorlopen** (1984, p. 40), (1990, p. 23)**.**

De hoofdredenering bestaat uit de aannames 1 tot en met 3 (A1 t/m A3), en conclusie 1 (C1). Aanname 4 (A4) en conclusie 2 (C2) met 3 (C3) benadrukken het bijzondere van het brein en de noodzaak van deze causale krachten. Searle draait de

nummering van C2 en C3 in het latere artikel (Sear1e, 1990) om, ten opzichte van de eerdere publicatie (Searle, 1984). C3 heeft geen nieuwe aannames nodig. Deze twee conclusies vormen de (extra) opvatting van Searle dat een artefact alleen mentale fenomenen kan hebben als het de causale krachten van het brein kan evenaren (hoe het ook plaatsvindt in de hersenen, dit kan niet worden volbracht door 'slechts' een computerprogramma te doorlopen), en zijn dus een uitwerking van de redenering richting zijn ideeën over causale krachten in het brein.

Preston (2002, p. 28) geeft een 'samenvatting' van de vele versies van de dieperliggende redenering die Searle uiteen heeft gezet. Hij noemt dit, in navolging van Searle en Larry Hauser, het "Brutally Simple Argument":

A1) Programma's zijn puur formeel (syntactisch).

A2) Minds hebben semantiek, mentale (semantische) inhouden.

A3) Syntax alleen is niet gelijk aan of voldoende voor semantische inhoud.

C1) Programma's op zichzelf zijn niet constitutief of voldoende voor minds.

Deze simpele vorm beperkt zich tot de relevante stellingen en de eerste en belangrijkste conclusie. De andere conclusies van Searle kunnen worden gezien als verklaring of verdere uitwerking van de eerste. Searles A4, die stelt dat hersenen minds kunnen veroorzaken, zal ik als subaanname van A2 bespreken.

2.2. Logische aanvallen op de redenering

Aanvallen op Searles Chinese Kamer 'Redenering' ('Chinese Room Argument' in het Engels, ook veel aangeduid met 'CRA') zijn er meer dan voldoende, maar de vraag bij deze aanvallen is welke strategie ze volgen. Betwijfelen ze de aannames of de conclusie van de redenering, of hebben ze andere redenen om de hele redenering als ongeldig te zien? En nemen ze het 'Chinese Room Argument' *of* het "Brutally Simple Argument" als uitgangspunt? Vaak wordt slechts de verhalende opzet betwijfeld en veranderd of als onzinnig bestempeld. Zoals we hebben gezien handhaaft Copeland een strategie waarin hij de 'logische' opzet van het gedachte-experiment bekritiseert, en hiermee wil aantonen dat elke onderliggende redenering daarmee ook kracht verliest. Copeland voert echter geen argumenten tegen de zojuist uitgeschreven onderliggende redenering aan. Searle heeft zelf toegegeven dat het gedachte-experiment geen bewijs levert voor de onderliggende redenering. Hij ziet ook dat weinig van zijn *critics* ooit tegen de *'sheer logical structure'* van de redenering zijn ingegaan, en niet aangeven welke redeneerstappen er bevochten worden, in citaat van

Preston (2002, p. 28). Searle constateert zelf dat hij nog geen kritieken heeft gezien die argumenten bevatten die tegen die van hem zelf ingaan:

> ‘But, once again, why? Why can't I in the Chinese room also have a semantics? Because all I have is a program and a bunch of symbols, and programs are defined syntactically in terms of the manipulation of the symbols. The Chinese room shows what we should have known all along: syntax by itself is not sufficient for semantics. (**Does anyone actually deny this point, I mean straight out?** Is anyone actually willing to say, straight out, that they think that syntax, in the sense of formal symbols, is really the same as semantic content, in the sense of meanings, thought contents, understanding, etc.?)’ (The Failures of Computationalism: I, 2001) (mijn nadruk).

In de volgende hoofdstukken beschrijf ik uitgebreid in hoeverre Searles aannames geldig zijn (volgens zijn beschrijving) en hoe hun geldigheid ervoor kan zorgen dat zijn redenering nog steeds overeind staat. Als niemand er op de juiste manier tegenin gaat, is het nog niet op de juiste manier omvergeworpen. Om dit te kunnen bekijken, bekijk ik in de volgende hoofdstukken per aanname wat Searles onderbouwing voor en eventuele latere opmerkingen over de aannames zijn en hoe en door wie hier *direct* tegenin gegaan wordt.

2.3. Wat is het doelwit van de SSR, en wat valt buiten de reikwijdte?

De syntax-semantiekredenering is gericht tegen het computationeel functionalisme van de Harde AI. De impact van de SSR en de CRA is niet gering geweest, het bereik (datgene dat en degenen die erdoor aangevallen worden) is in principe en feitelijk ook erg groot. Harde KI is een serieuze tak van de wetenschap, er zijn veel mensen die zich ermee bezig houden en er wordt veel financiering in gestoken. Searle (2004a, p. 63) verklaart de impact door zijn aanpak: vorige kritieken op Harde KI bestonden vooral uit “Een computer kan nooit A”, waarop men in de Harde KI een programma maakte dat “A” mogelijk maakte. Searle heeft op het hart van de Harde KI gericht, door aan te geven dat computatie en symboolmanipulatie in computers geen mind kunnen genereren. De Harde KI is hiermee ‘gedoemd’. Het connectionisme heeft een andere aanpak in hoe informatie wordt gecodeerd en gebruikt, en het is niet direct uit de SSR af te leiden dat het ook op het connectionisme van toepassing is. Hiervoor is een extra redenering nodig. De uitgebreide behandeling van deze redenering valt helaas buiten mijn onderzoek. Heel beknopt, luidt de redenering dat alle processen of systemen die op een digitale computer *geprogrammeerd* kunnen worden nooit voldoende zijn om de computer semantiek te geven. Het feit dat hersenprocessen gesimuleerd kunnen worden in neurale netwerken, bewijst niet dat deze neurale

netwerken dezelfde causale krachten hebben als het brein (of enige andere kracht) (Searle, 1980, p. 244).

Searle bekritiseert dus niet diegenen die zich niet uitlaten over of zich niet begeven op het vlak van 'hebben van een mind' of 'persoon zijn'. In Zwakke KI is geen sprake van het willen creëren van personen. Ook in de Cognitiewetenschap, waarin de computer alleen als instrument en testmiddel voor modellen gebruikt wordt, hoeft men zich in principe niet aangevallen te voelen. Alleen als deze psychologen de (volgens Searle aan de hersenen eigen) causale krachten in de mens 'vergeten', en overtuigd zijn dat een computersimulatie van cognitie laat zien hoe cognitieve processen in menselijke hersenen daadwerkelijk plaatshebben, moeten ook zij zich aangesproken voelen.

Voor de filosofie van mind is de redenering in zoverre van belang, dat de puur functionele (computermodel) beschrijvingen van de geest, niet voldoende zijn om de essentie van een mind te beschrijven. Bringsjords hoofdstelling luidt: 'robots will largely *do* what we do, but won't *be* one of us' (Bringsjord, 1992, p. 43). Een mind hebben, een persoon zijn, is waar het om gaat, en dit is niet te bereiken via Harde KI.

De reikwijdte van de syntax-semantiekredenering hangt sterk af van waar 'syntax' en 'semantiek' voor staan in de systemen die bekritiseerd worden. Het domein van de redenering, de objecten waarover de aannames gedaan worden, zijn abstract gedefinieerd als 'digitale computers' en gebaseerd op de werking van Turing Machines. Digitale computers en vooral de hardware worden constant verbeterd, wat vooral inhoudt dat ze steeds snellere processoren en grotere geheugens hebben. Voor Searles oorspronkelijke redenering is het in principe irrelevant hoe snel of geavanceerd deze systemen worden of kunnen zijn. 'Syntax' en 'semantiek' blijven naar hetzelfde type proces of toestand in de systemen verwijzen. Het is dus erg belangrijk wat Searle bedoelt met 'syntax', en nog belangrijker wat hij bedoelt met 'semantiek', en waarom semantiek zo belangrijk is. Dat deze 'leentermen' uit de linguïstiek en de logica gebruikt worden is niet vreemd. Het gebruik van de formele aspecten van taal om de manier waarop de processen in een computer (en in de hersenen) te typeren is de klassieke KI en het computationeel functionalistische beeld van mind niet onbekend. Symboolsystemen zijn talig gestructureerd; de symbolen zijn basiseenheden met betekenis (als woorden), en de regels die ze manipuleren zijn als een grammatica. Deze analogie wordt ondersteund door het idee dat mentale toestanden en 'denken' sterk verbonden zijn met taal. Vooral Fodor, die de Language

of Thought Hypothesis verdedigt, wil aantonen dat gedachten de logische en structurele regels van taal volgen. Uitgaand van Fodor, is het voor de praktijk van de KI is het logisch om in een Symbol System deze analogie als houvast te zien. (Cunningham, 2000, p. 193). Zoals gezegd, is Fodor zelf geen aanhanger van KI (maar ook geen vervent tegenstander, de positie is wat ingewikkelder dan 'voor of tegen'). Zijn ideeën zijn wel zeer bruikbaar voor de KI.

In hoeverre is Searles gebruik van de termen syntax en semantiek slechts een analogie voor, en niet een daadwerkelijke beschrijving van de ontologie van het domein van de Harde KI? Zijn het 'slechts' leentermen? Hoe letterlijk moet je de terminologie nemen? In eerste instantie lijkt het erop dat we dit niet te letterlijk moeten opvatten: 'In the linguistic jargon, they *[the formal symbol manipulations]* have only a syntax but no semantics' (Searle, 1980, p. 248). Searle gebruikt hier het jargon om een 'stap' of niveauverschil aan te geven dat juist het probleem voor Harde KI goed aangeeft. In de producten van de KI die onder vuur liggen, gaat het dus niet alleen om taal, om bijvoorbeeld het kunnen spreken van een taal zonder de betekenis ervan te kennen (zoals Searle in de Chinese Kamer). Het gaat om alle algemene cognitieve informatieprocessen, die puur syntactisch van aard zijn. De inhoud van deze processen is niet te vatten door gebruik van de syntactische structuur. Deze manier van gebruikmaken van de terminologie maakt de eigenschappen van syntax en semantiek *in de taalkunde en logica* minder relevant voor deze discussie. Het gebruik van de terminologie door Searle zal hopelijk verduidelijkt worden in de uitleg van zijn aannames, waardoor we zullen kunnen concluderen hoe we dit gebruik kunnen typeren.

2.4. Conclusie

Het Chinese Kamergedachte-experiment van Searle bespreekt zijn kritiek op de Harde KI. Het gedachte-experiment heeft echter een andere 'redenering' in de oppervlaktestructuur dan op dieperliggend niveau. De dieperliggende redenering van de Chinese Kamer redenering is de syntax-semantiekredenering, en niet de verhalende oppervlakte structuur van de Chinese Kamer. De dieperliggende redenering is de (enige) echte redenering van Searle tegen Harde KI; de oppervlaktestructuur beschrijft geen echte redenering. Argumenten die tegen de aannames van de SSR ingaan, zijn de enige bruikbare argumenten, maar niet gemakkelijk te vinden; Searle vraagt zich af of iemand zijn kernaanname(s) aanvalt! In de SSR wordt de terminologie uit de

taalkunde en logica gebruikt om over het domein van het computationalisme conclusies te trekken. De systemen (en hiermee ook de makers ervan) in Harde KI in de vorm van het computationalisme is het doelwit van de redenering.

3. Minds hebben semantiek

Aanname 2 van Searle luidt: minds hebben semantiek. De mentale fenomenen die wij associëren met een mind, hebben inhoud, gaan ergens over. Deze stelling is voor Searle een vaststaand feit:

> 'And that, I take it, is just an obvious fact about how our minds work. My thoughts, and beliefs, and desires are about something, or they refer to something, or they concern states of affairs in the world; and they do that because their content directs them at these states of affairs in the world' (1984, p. 39).

Alhoewel Searle deze opvatting als 'overduidelijk feit' ziet, is het voor dit onderzoek belangrijk *hoe* hij dit als gegeven ziet. Searle vat het op als een 'obvious fact', alsof er geen argumentatie aan vooraf gaat. Toch is het relevant voor dit debat om ons af te vragen, wat voor Searle de redenen zijn om deze stelling als 'obvious fact' te zien. Hierin ligt een groot deel van de reden waarom het voor Harde KI en computationalisme een onhaalbaar doel is (volgens de filosoof Searle) om een mind te creëren. In dit hoofdstuk komen dan ook de volgende vragen aan de orde: "Hoe komen minds aan semantiek?", "Wat wordt bedoeld met semantiek?" en "Waarom is semantiek relevant?". Hiermee hoop ik Searles positie te kunnen verduidelijken en uit te kunnen leggen op welke expliciet te stellen wijze deze intuïtief plausibel is. Daar waar zijn positie mogelijk een hiaat heeft, gebruik ik opvattingen van Haugeland en Rychlak om suggesties te doen om het standpunt aan te vullen.

3.1. Searles positie - Hoe komen minds aan semantiek?

Het antwoord op de vraag van deze paragraaf is in subaanname 2a te vinden: hersenen veroorzaken minds. In deze paragraaf geef ik een uitleg van Searles positie in de filosofie van mind, wat betreft zijn opvatting dat minds semantiek hebben. Wat is volgens Searle de 'bron' van deze eigenschap? In de korte beschrijvingen bij de aanname in hoofdstuk twee zagen we al de term 'causale krachten' genoemd worden. Voor Searle staan in het veroorzaken van minds causale krachten centraal:

> '[..] I am a certain sort of organism with a certain biological (i.e. chemical and physical) structure, and this structure, under certain conditions, is causally capable of producing perception, action, understanding, learning, and other intentional phenomena. And part of the point of the present argument is that only something that had those causal powers could have that intentionality' (1980, p. 247).

De term 'causale krachten' zoals die hier geïntroduceerd wordt, is er een die geen voor de hand liggende interpretatie heeft, en is moeilijk uit te leggen in andere (reducerende of vertalende) terminologie. Voor Searle is het 'sophisticated common

sense' (Biological Naturalism, 2004b) dat het brein in ieder geval 'bezitter' van deze krachten is. Wat zijn de achterliggende aannames over minds bij mensen die deze uitspraak kunnen ondersteunen?

3.1.1. Biologisch naturalisme: minds bij mensen

Zijn aanpak over de werking van de mind en mentale toestanden in het algemeen (bij mensen), noemt Searle het *biologisch naturalisme* (2004b, p. 1). Searle vindt dat een 'frisse' (onbeladen door filosofische traditie) opvatting in de huidige wetenschap past en dat dit de manier zou moeten zijn om tegen het mind-body ('geest-lichaam') probleem aan te kijken:

> 'Well, here is what I came up with; and if you could just forget about Descartes, dualism, materialism, and other famous disasters I think you would come up with something very similar' (2004b, p. 1).

Hij doet een appel op het 'intuïtieve' gezond verstand van tegenwoordig om in te zien dat *mentale toestanden* niet buiten het wetenschappelijke wereldbeeld hoeven te vallen. De terminologie legt hij als volgt uit: 'biologisch' omdat het de nadruk legt op het biologische niveau als het juiste niveau om een uitleg te geven van mentale verschijnselen (2004b, pp. 6-7) en 'naturalisme' omdat mentale verschijnselen deel zijn van de natuurlijke wereld. Het inpassen van mentale verschijnselen in de natuurwetenschappen vereist daarom geen extra 'naturalisatie' (in de vorm van reductie, herdefiniëring of eliminatie), want deze verschijnselen zijn al deel van onze natuur (2004b, p. 7), en daarmee, idealiter, ook van de natuurwetenschappen.

Een belangrijke notie voor het biologische naturalisme is de 'ontologische subjectiviteit' van mentale toestanden, en deze kan het beste uitgelegd worden in Searles eigen woorden:

> 'The objective-subjective distinction is ambiguous and we need to disambiguate it before we go any further. First, there is an epistemic sense of the objective-subjective distinction. The claim that Rembrandt was born in 1606 is a matter of objective fact. The claim that Rembrandt was a better painter than Rubens is a matter of subjective opinion. Objectivity and subjectivity in this epistemic sense are features of claims. But in addition to the epistemic sense there is an ontological sense of the distinction. Most things, such as mountains, molecules and tectonic plates exist apart from any experiencing subject. They have an objective or third person ontology. Some things, such as pains and tickles and itches, only exist when experienced by a human or animal subject, and they have a subjective or first person ontology. Consciousness *[en andere*

> *mentale toestanden*[9]*]* is ontologically subjective in the sense that it only exists when experienced by a human or animal subject. It is important to emphasize that you can have epistemically objective knowledge of a domain that is ontologically subjective. It is for this reason that an epistemically objective science of ontologically subjective consciousness is possible' (2004b, pp. 2-3).

Dus, alhoewel het domein ontologisch subjectief is, is het wel mogelijk om kennis van dit domein te vergaren die epistemologisch objectief (niet afhankelijk van opinie van degene die de kennis heeft) is. Met deze redenering geeft Searle een verantwoording van waarom het eerste persoonsperspectief een speciale status in de wetenschap heeft: het is wel een subjectief geconstitueerd (epistemologisch subjectief toegankelijk) perspectief, maar dat houdt niet in dat de kennis erover ook epistemologisch subjectief is: objectieve kennis over een ontologisch subjectief domein is mogelijk, volgens Searle. De (mogelijke) kritiek dat een eerste persoonsperspectief op geen enkele manier tot objectieve kennis kan leiden, wordt hiermee onderschept.

Een andere tegenstelling die Searle wil wegnemen, is de schijnbare tegenstelling dat mentale verschijnselen vanwege hun zogenaamde mysterieuze status niet binnen de natuurwetenschappen passen. Mentale verschijnselen zijn eigenschappen van een hoger niveau, en toch objectieve natuurverschijnselen:

> 'The consciousness that is caused by brain processes is not an extra substance or entity. It is just a higher-level feature of the whole system. [...] There is nothing metaphysical about water being wet, just as there is nothing metaphysical about consciousness' (Searle, 1995, p. 211).

Het niveauverschil tussen mentale verschijnselen en hun causale basis is geen reden om mentale verschijnselen als 'mysterieus' te beschouwen. Het beschouwen van mentale verschijnselen (waaronder bewustzijn) als niet-mysterieus is lastig in de bestaande traditie van de filosofie van mind, maar niet onmogelijk:

> 'Consciousness is a natural biological phenomenon that does not fit comfortably into either of the traditional categories of mental and physical. It is caused by lower-level micro processes in the brain and it is a feature of the brain at the higher macro levels' (Searle, 1997, p. xiv).

---

[9] De substitutie 'en andere mentale toestanden' is verantwoord door de volgende uitspraak: 'Biological Naturalism is a theory of mental states in general but as this book* is about consciousness I will present it here as a theory of consciousness' (Searle, Biological Naturalism, 2004b, p. 1). *'This book' verwijst naar mijn beste weten niet naar een boek, dit document online is de enige verschijning van deze tekst die ik heb gevonden.

Hoe de juiste ‘causale krachten’ voor mentale fenomenen tot stand komen, is een empirisch-theoretische vraag voor de neurobiologie: ‘The problem is to figure out how the system, the brain, works to produce consciousness; and that is an empirical-theoretical issue for neurobiology’ (Searle, 1995, p. 213). Dat de neurobiologie hiervoor de aangewezen wetenschap is, is fundamenteel voor de ideeën van het biologisch materialisme.

De neurobiologie moet uitwijzen hoe **processen op macro- en microniveau** in de hersenen samenhangen en hoe de hersenen mentaliteit veroorzaken (Searle, 1984, pp. 20-21). Op het macroniveau bestaan er bijvoorbeeld mentale fenomenen, die door elementen op het neuronale microniveau ‘veroorzaakt’ en ‘gerealiseerd’ worden. Dit gebeurt tegelijkertijd; mentale fenomenen komen niet ‘na’ de hersenprocessen:

> ‘Just as the liquidity of the water is caused by the behaviour of elements at the micro-level, and yet at the same time it is a feature realised in the system of micro-elements, so in exactly that sense of ‘caused by’ and ‘realised in’ mental phenomena are caused by processes going on in the brain at the neuronal or modular level, and at the same time they are realised in the very system that consists of neurons. Just as we need the micro/macro distinction for any physical system, so for the same reasons we need the micro/macro distinction for the brain’ (1984, p. 22).

Deze vorm van veroorzaking bestaat dus uit het tegelijkertijd voorkomen van oorzaak en ‘gevolg’ op de verschillende niveaus van een fysisch systeem. Alle vormen van macroprocessen (op verschillende abstractieniveaus) zijn gerealiseerd in microprocessen op het laagste niveau. Mentale fenomenen zijn op deze manier als macroprocessen met microprocessen als uiteindelijke basis te zien.

Een kanttekening die te plaatsen is bij deze aanduiding van het macro/micro onderscheid van processen, is dat er naast een niveauverschil tussen macroprocessen (zoals mentale processen met semantiek) en microprocessen (in zekere zin is de ‘syntax’ van mentale toestanden een aanduiding voor ‘microprocessen’ op lager niveau) voor mentale toestanden ook een niveauverschil tussen syntax en fysica is voor hersentoestanden (1994b, p. 547), (1992, p. 207) en (1995, p. 210). Dit niveauverschil vormt óók een barrière voor de beschrijving van minds; het geeft het probleem aan waarmee neurobiologie te kampen heeft; wat is de relatie of de brug tussen die macro- en micro-eigenschappen? Het begrip ‘syntax’ zoals in de SSR gebruikt wordt, geeft niet perse het laagste niveau van microprocessen aan. Syntax is volgens Searle een *waarnemerrelatieve* notie: syntax is niet intrinsiek aanwezig in fysische processen. (1992, p. 207) (en bovengenoemde bronnen). Deze opvatting is

belangrijk voor de Searles interpretatie van ‘computatie’, hieraan wordt in hoofdstuk vier meer aandacht besteed.

Deze mogelijke barrière tussen niveauverschillen is voor Searles biologisch naturalisme echter geen probleem. Als syntax niet intrinsiek is aan fysische processen, is een syntactische beschrijving ervan niet essentieel voor het bestaan van het proces. Het is afhankelijk van de waarnemer welke syntactische interpretatie er aan een proces wordt gegeven. Deze opvatting van Searle geeft een fundamentele reden weer voor de opvatting dat de fysische beschrijving (van de processen op microniveau) de enige essentiële beschrijving is; het fysische is het enige niveau waarop de processen plaatsvinden, niet afhankelijk van een waarnemer (intrinsiek). Dit betekent ook dat de syntactische beschrijving niet het niveau is in termen waarvan de essentie van de processen in de hersenen te definiëren is, zie ook Dennetts volgende parafrase van Searle;

> ‘*The crucial powers of brains have nothing to do with the programs they might be said to be running*, so “giving something the right program” could not be a way of giving it a mind’ (1987, p. 326)(mijn nadruk).

Een zeker niveau van beschrijven van macroprocessen is dus het niveau waarop de processen in termen van syntax beschreven kunnen worden, maar deze syntactische beschrijving van de werking van de hersenen is dus nooit voldoende voor de semantische beschrijving. De syntactische beschrijving van processen in de hersenen beschrijven niet de meest fundamentele microprocessen in de hersenen. Beschrijvingen van de hersenen op niveau van macroprocessen zijn niet intrinsiek of essentieel voor de beschrijving op niveau van microprocessen. Het is eerder andersom: het niveau van de microprocessen is essentieel en maakt een juiste beschrijving van de macroprocessen mogelijk. De microprocessen dienen bestudeerd te worden in de neurobiologie om de juiste bijbehorende macroprocessen te ontdekken.

In het geval van mentale toestanden is het hoogste niveau van beschrijving de mentale toestand zelf en niet het syntactische of fysische proces waaraan het identiek is. Juist dit niveau is *ontologisch subjectief* en dus niet naar een *objectieve ontologie* te reduceren. In het geval van mentale toestanden en het eerste persoonsperspectief dat voor de ontologie van mentale toestanden essentieel is, lijdt een *causale* **reductie** van mentale toestanden niet tot een *ontologische* **reductie** (naar een derde

persoonsontologie) door middel van herdefiniëring, omdat de herdefiniëring het belang van het hebben van het concept ‘mentale toestanden’ teniet zou doen (2004b, p. 11). In een reductie op ontologisch niveau (naar derde persoonsontologie) zou het essentiële eerste persoonsperspectief verdwijnen: dit is ongewenst. Dit betekent niet dat er geen causale reductie van mentale toestanden (een beschrijving op het niveau van microprocessen uit de neurobiologie, bijvoorbeeld) mogelijk is. Er bestaat dus geen contradictie in Searles opvattingen over reductie: hij is **geen reductionist** wat het ontologische niveau van mentale toestanden betreft (2004b, p. 11).

3.1.2. Minds in andere systemen dan de hersenen

Het biologisch naturalisme (en de Chinese Kamerredenering dan wel de syntax-semantiekredenering) sluiten niet uit dat minds in andere systemen dan het brein kunnen bestaan. Deze opvatting wordt echter wel vaak als bijkomende conclusie van Searles positie gezien:

> ‘Another misunderstanding of the Chinese Room Argument is to suppose that I am arguing that as a matter of logic, as an a priori necessity, only brains can have consciousness and intentionality. *But I make no such claim. The point is that we know in fact that brains do it causally. And from this it follows as a logical consequence that any other system that does it causally, i.e. that produces consciousness and intentionality, must have causal powers to do it at least equal to those of human and animal brains. But it does not follow that other systems have to have neurons to do it.* (Compare: airplanes do not have to have feathers in order to fly, but they do have to share with birds the causal powers to overcome the force of gravity in the earth’s atmosphere.). The question of which systems are causally capable of producing consciousness and intentionality is an empirical factual issue, not to be settled by a priori theorizing’ (1994b, p. 547) (mijn nadruk).

De vraag of andere systemen dan hersenen minds kunnen veroorzaken, is een open empirische vraag. Een systeem dat of machine die daartoe in staat is zou echter volgens Searle van een bijzonder soort moeten zijn (met de juiste causale krachten), zoals de hersenen:

> ‘“Could a machine think?” My own view is that only a machine could think, and indeed only very special kinds of machines, namely brains and machines that had the same causal powers as brains’ (1980, p. 251).

Searle claimt over Harde KI en computationalisme wel dat deze methodes *a priori* niet de juiste causale krachten kunnen genereren, omdat ze op het verkeerde niveau de causale krachten proberen te dupliceren:

> ‘No one would suppose that we could produce milk and sugar by running a computer simulation of the formal sequences in lactation and photosynthesis, but where the mind is concerned many people are willing to believe in such a miracle because of a deep and abiding dualism: the mind they suppose is a matter of formal processes and is

> independent of quite specific material causes in the way that milk and sugar are not' (1980, p. 251).

Voor dupliceren is slechts simuleren niet voldoende. Waarom het gebruikte niveau zo fundamenteel verkeerd is, wordt besproken in hoofdstuk vier, waar de aanname dat computers 'slechts' syntax en geen semantiek hebben, wordt bekeken.

Het cruciale 'ingrediënt' voor *welk systeem dan ook* om mentale fenomenen te kunnen veroorzaken is dus het hebben van de juiste **causale krachten**. Hoe dit gebeurt, en hoe dit kan worden beschreven, zijn vragen voor neurobiologie en een is kwestie van interpretatie van de essentiële kenmerken van hersenprocessen. Het is inderdaad jammer, dat Searle zelf geen uitleg of verklaring van mentale fenomenen kan geven om zijn positie te ondersteunen (Hauser, 2002, p. 134 en p.141). Dit is echter een logische consequentie van het verschuiven van de bewijslast naar de neurobiologie enerzijds, en anderzijds het als een 'obvious fact' zien dat dit verschil werkelijk bestaat en dus überhaupt onderzocht dient te worden in de neurobiologie.

3.2. Causale krachten: perspectieven – Dennett en Searle

Als we voor deze redenering willen bekijken waarom de syntax-semantiekredenering een probleem is voor Harde KI, is het om verder te gaan met de opvattingen van Searle nodig dat we aannemen *dat* de hersenen de juiste causale krachten, zoals die hierboven besproken zijn, en nu bekijken wat die causale krachten bewerkstelligen en waarom ze van belang zijn. Over wat de essentie van deze causale krachten precies is, is (vooral) Daniel Dennett het niet met Searle eens; wat is de basis van deze discussie?

Dennett en Searle zijn beiden onder de indruk van de causale krachten van het menselijk brein, maar ze zijn het niet eens over welke causale krachten ertoe doen en waarom (Dennett, 1987, p. 325). Dennett is ervan overtuigd dat de juiste programmering 'iets' wel een mind kan geven, maar dat het empirisch gezien onwaarschijnlijk is dat het juiste type programma gedraaid kan worden op iets anders dan organische, menselijke breinen (1987, p. 326).

> 'So the causal powers required to control the swift, intelligent, intentional activity exhibited by normal human being can be achieved only in a massive parallel processor – such as a human brain. (Note that I have not attempted an a priori proof of this; I am content to settle for scientific likelihood)' (Dennett, 1987, p. 327).

Het kan zo blijken te zijn dat het enige systeem dat een dergelijke snelheid kan genereren een organisch brein is (Dennett, 1987, p. 328). Hiermee geeft Dennett dus wel aan dat er *iets* is wat gegenereerd wordt! Maar hij ziet de essentie van het streven om dit 'iets' in de wetenschap te proberen te passen dan wel te verklaren niet in, omdat alles via gedrag te verklaren zou zijn:

> 'Behavior, in this bland sense, includes all intersubjectively observable internal processes and events (such as the behavior of your gut or your RNA). No one complains that models in science only account for the "behavior" of hurricanes or gall bladders or solar systems. What else is there about these phenomena for science to account for? This is what makes the causal powers Searle imagines so mysterious: they have, by his own admission, no telltale effect on behavior (internal or external) – unlike the powers I take so seriously: the powers to guide a body through life, seeing, hearing, acting, talking, deciding, investigating, and so on' (1987, p. 334).

Dennett heeft niet dezelfde opvatting of intuïties als Searle over wat 'causale krachten hebben' inhoudt, en hoe we dit moeten benaderen. Haugeland poogt het punt waarop Searle en Dennett van elkaar beginnen te verschillen te beschrijven: hij ziet hun verschillende typeringen van welke samenstelling van en welke interacties tussen de delen van hersenen (of een systeem in het algemeen) als hun fundamentele verschil in opvattingen. Daaruit komen ook andere opvattingen over welke systemen wel en welke niet de juiste samenstelling en interacties hebben voort, en daarmee de verschillende opvattingen over hoe we dit al dan niet kunnen herkennen:

> 'Where they diverge, then, is in the way they characterize that "suitable arrangement and interaction" of matter, and hence in their accounts of which systems might have it, and how we can tell' (Haugeland, 1998, p. 291), '[..] and, moreover, put this way, I think Searle is closer to the truth' (p. 294).

Zoals Haugeland hier toegeeft, is hij (ook) geneigd om het meer met Searle dan met Dennett eens te zijn. Volgens Dennett is de juiste benadering het derde persoonsperspectief ('gedrag omvat alle intersubjectief waarneembare interne processen'), volgens Searle het eerste persoonsperspectief, zoals eerder aangegeven. Deze botsende opvattingen lijken uit te komen op een 'welles'-'nietes' dialoog.

Mijn persoonlijke uitgangspunt is dat Searles redenering intuïtief plausibel is, en naar de redenen hiervoor ben ik in dit onderzoek (onder andere) op zoek. Ik denk dat een van die redenen dat de redenering intuïtief plausibel is, ligt in het feit dat ik (blijkbaar) ook van mening ben dat het eerste persoonsperspectief relevant is voor de wetenschap en vooral voor dit debat. Als in dit debat het eerste persoonsperspectief wordt genegeerd, dan lijkt het ultieme doel van de Harde KI te verdwijnen of 'devalueren': om een systeem de status van persoon te laten bereiken, is het niveau

van het hebben van mentale toestanden van groot belang. Semantiek hebben op de manier die Searle probeert te beschrijven (zie 3.3) is hiervoor essentieel. In de Harde KI kan men deze relevantie haast niet ontkennen. Wat ik hopelijk tot nu toe, via Searle, duidelijk en intuïtief plausibel heb gemaakt, is dat feiten die wij verwerven via het ontologisch subjectieve eerste persoonsperspectief net zo goed wetenschappelijke feiten zijn als bijvoorbeeld het bestaan van zwaartekracht of fotonen:

> 'the fact that it reminds us [..] that actual semantics in a human mind, its actual mental contents, are ontologically subjective. So the facts are accessible only from the first-person point of view. Or rather, I should say that the facts are accessible from the first-person point of view in a way that they are not accessible from the third person' (Searle, 1995, p. 209).

De 'welles'-kant van dit debat, die een uitgangspunt is voor een verdere bespreking van de positie van Searle, is dus dat het eerste persoonsperspectief óók essentieel is voor de wetenschap. Hiermee is het derde persoonsperspectief in zekere zin irrelevant voor de ontologie van de mind:

> 'First, I have tried to argue that as far as the ontology of the mind is concerned, behavior is simply irrelevant. Of course in real life our behavior is crucial to our very existence, but when we are examining the existence of our mental states as mental states, the correlated behavior is neither necessary nor sufficient for their existence' (1992, p. 77).

Alleen het bestuderen van *gedrag*, dat vanuit het derde persoonsperspectief te observeren is (de 'andere opvatting' zoals van Dennett), is in deze uitleg van Searle voldoende noch noodzakelijk voor het bestuderen van minds. Als een systeem wordt gemaakt zonder de intentie bij dit systeem een mind gelijkend op minds bij mensen te veroorzaken, is dit een andere (gedevalueerde) vorm van KI. Minds maken zonder de relevantie van het eerste persoonsperspectief te erkennen is eigenlijk niet daadwerkelijk minds proberen te maken.

3.3. Wat wordt bedoeld met semantiek?

'Semantiek', zoals door Searle gebruikt, is op zijn minst (het meest 'los' opgevat) een leenterm. Semantiek staat voor 'betekenis'. Maar wat betekent 'semantiek' zoals Searle het gebruikt? Voor welke begrippen gebruikt Searle het woord semantiek? *Minds, Brains and Programs*, het oorspronkelijke Chinese Kamerartikel, is de bron van deze vergelijking; in dit artikel komen dan ook de belangrijkste termen waar Searle op doelt aan bod (Searle, 1980). Het is helaas niet zo dat er slechts één concept

is waarvoor semantiek als referentieterm wordt gebruikt; er zijn meerdere concepten die een soort grote gemene deler, 'semantiek', hebben. Dit zijn de volgende:

1) "Understanding" (p. 235 en verder) heeft met stip de nadruk

2) 'meaning' (p. 238), omdat het over het begrijpen van de *betekenis* van taal gaat

3) 'intentionality' (p. 247) '[..] that I am able to understand English and have other forms of intentionality' en

4) 'mental content of intentional states' (p. 249)

Waarom ligt de focus op begrijpen, en daarmee op intentionaliteit (de Engelse taal begrijpen is zoals geciteerd een vorm van intentionaliteit)? Bedoelt Searle niet eigenlijk ook bewustzijn of alle mentale fenomenen tezamen? Over bewustzijn heeft hij het niet in Minds, Brains and Programs, dit geeft hij later in een interview zelf nogmaals expliciet aan (1995, p. 210)! Bewustzijn is een ander fenomeen dan intentionaliteit, en duidelijk een ander, maar op een of andere manier samenhangend, probleem. Het is een van de vier kenmerken van het mind-body probleem van de filosofie van mind (1984, p. 17), die niet identificeerbaar met of reduceerbaar tot elkaar zijn. De correlatie van intentionaliteit en bewustzijn ligt in het feit dat het beide biologische mentale fenomenen zijn. De onderlinge relatie tussen deze fenomenen is ook voor Searle een open vraag die een goede theorie van mind allemaal moet kunnen plaatsen:

> 'These four features, consciousness, intentionality, subjectivity, and mental causation are what make the mind-body problem seem so difficult. Yet, I want to say, they are all real features of our mental lives. Not every mental state has all of them. But any satisfactory account of the mind and of the mind-body relations must take account of all four features. If your theory ends up denying any one of them, you know you have made a mistake somewhere' (1984, p. 17).

3.3.1. Intrinsieke intentionaliteit en alsof-intentionaliteit

De Chinese Kamer redenering en zeker ook de syntax-semantiekredenering, gaan over intentionaliteit. 'I originally presented the Chinese Room Argument as an argument about intentionality' (1995, p. 210). Intentionaliteit is niet te reduceren tot een ander begrip (1994a, p. 380). Welk concept (intentionaliteit hebben, mentale inhouden hebben, bewust zijn) het meest fundamenteel is, daar doet Searle geen uitspraken over. Intentionaliteit, intrinsiek en niet te reduceren, krijgt met nadruk de grootste aandacht in de kritiek op het computationalisme. De nadruk ligt in deze discussie niet op wat intentionaliteit dan *is*, hoe het ontstaat, wat het inhoudt of bewerkstelligt (zie voor Searles opvattingen hierover zijn *'Intentionality'* (1983)),

maar op de *aanwezigheid van intentionaliteit als intrinsieke eigenschap*, waarvoor het eerste persoonsperspectief essentieel is.

'Intrinsiek' betekent 'het echte ding', als in 'niet slechts de verschijning van het ding' (*alsof-intentionaliteit*) of 'afgeleide intentionaliteit' (1992, p. 80). 'Intrinsiek' staat *niet* voor 'mysterious, ineffable, and beyond the reach of philosophical explanation or scientific study' (Searle, 1992, p. 80). Dus wat dit *schijnbaar* met mysterie omhulde fenomeen intrinsieke intentionaliteit ook is, het bestaat wel, en het moet daadwerkelijk aanwezig zijn*; niet slechts onder interpretatie. Het oordeel 'wel of niet intrinsiek aanwezig' moet niet afhangen van de waarnemer*. Dit maakt het oordeel uiteraard bijzonder lastig:

> 'One of the hardest – and most important – tasks of philosophy is to make clear the distinction between these features of the world that are intrinsic, in the sense that they exist independent of any observer, and those features that are observer-relative, in the sense that they only exist relative to some outside observer or user' (Searle, 1992, p. xii).

Het aannemen van een 'intentional stance' (hiermee duidelijk verwijzend naar Dennett) (Searle, 1992, p. 81), waarin je een verklaring van gedrag op basis van intentionele toestanden geeft, ten opzichte van een systeem, zegt niets over de intrinsieke intentionele toestanden van dat systeem. Het verschil hier is het bovengenoemde verschil tussen *alsof-intentionaliteit* en *intrinsieke intentionaliteit* (Searle, 1992, pp. 78-81). Dit onderscheid is geen ingewikkeld onderscheid; de uitleg van Searle is intuïtief zeer aanvaardbaar: als een uitspraak over het toeschrijven van intrinsieke intentionaliteit waar is, is er daadwerkelijk een intentionele toestand in het object waaraan de intrinsieke intentionaliteit toegeschreven wordt. Als je intentionaliteit figuratief of metaforisch toeschrijft, heb je het over 'alsof'-intentionaliteit. Het is echter niet zo dat alsof-intentionaliteit een echte soort (variant van) intentionaliteit is. Het (kunnen) toeschrijven van alsof-intentionaliteit aan een systeem geeft aan dat het systeem te beschrijven is alsof het daadwerkelijk intentionaliteit heeft, omdat dat bijvoorbeeld handig is:

> 'It is very convenient to use the jargon of intentionality for talking about systems that do not have it, but that behave as if they did. I say about my thermostat that it *perceives* changes in the temperature [..]' (Searle, 1992, p. 79).

Het kunnen gebruiken van een dergelijke beschrijving zegt echter niet dat er daadwerkelijk intentionaliteit aanwezig is in 'de thermostaat' (het systeem).

3.3.2. Ontkennen van het onderscheid

Als je dit onderscheid ontkent, leidt dit tot een reductie tot in het absurde: alles kan dan 'intentionaliteit' hebben onder de juiste beschrijving of interpretatie: '[..] everything behaves as if it were following a rule, trying to carry out a certain project, acting in accordance with certain desires, etc.' (Searle, 1992, p. 81). Maar het oordeel over of een system wel of geen mentaliteit 'bezit', moet gebaseerd zijn op de intrinsieke aanwezigheid ervan:

> 'And the mental-nonmental distinction cannot just be in the eye of the beholder but it must be intrinsic to the systems; otherwise it would be up to any beholder to treat people as nonmental and, for example, hurricanes as mental if he likes' (Searle, 1980, p. 242).

Als je intentionaliteit als een 'marker' van mentaliteit ziet, geldt deze redenering ook voor intentionaliteit.

Er zijn wetenschappers die dit verschil niet als gegeven zien. Larry Hauser meent dat Searles dit onderscheid 'baseless' is, en geen enkele vorm van voorspellende of verklarende basis biedt (2002, p. 140).

> 'I submit here there are no compelling intuitive reasons for accepting the ambiguity between 'intrinsic' and 'as if' (attributions of) intentionality Searle alleges. Intuitive tests for ambiguity yield no evidence of ambiguity in such contexts. Tests, for instance, which enable us to 'hear' ambiguity [..] in certain contexts yield no sense of [ambiguity] [..] when applied to mental predications of computers. There are, it seems, then, compelling intuitive grounds to suppose that such predications are *unambiguous* literal predications. Theoretical grounds Searle *does* offer for positing an ambiguity here, where intuition recognizes none, are woefully inadequate' (2002, p. 133).

Hauser vindt het onderscheid dus zeker niet intuïtief aanvaardbaar; hij vindt dat het in tests waarin het onderscheid duidelijk zou moeten worden, geen sprake is van ambiguïteit in welke 'vorm van' intentionaliteit er toegeschreven wordt. Searle:

> 'I hope the distinctions I have been making are painfully obvious. However, I have to report, from the battlefronts as it were, that the neglect of these simple distinctions underlies some of the biggest mistakes in contemporary intellectual life. A common mistake is to suppose that because we can make *as-if* ascriptions of intentionality to systems that have no intrinsic intentionality, that somehow or other we have discovered the nature of intentionality'[10] (1992, p. 82).

Ik ben het eens met deze uitspraak van Searle; en ik zie inderdaad hoe vaak deze 'fout' gemaakt wordt (door bijvoorbeeld Hauser). Volgens Hauser zijn er duidelijke intuïtieve redenen om in te zien dat dit onderscheid niet zo gemaakt wordt in tests. Volgens Searle is het 'pijnlijk duidelijk' dat dit onderscheid hoe dan ook wel bestaat.

[10] Aan het eind van deze zin verwijst Searle door middel van een voetnoot naar Dennetts 'The Intentional Stance'.

Deze contrasterende intuïties geven reden tot een ‘welles’-‘nietes’ dialoog van eenzelfde soort als we eerder hebben gezien tussen Dennett en Searle (al is Dennett wat subtieler en diepgaander in de manier waarop hij tegen Searle ingaat). Hauser geeft hier een ‘tegenargument’ op basis van tests die het *gedrag* van mensen die mentale predicaten aan systemen toeschrijven onderzoeken. Hij vindt in deze tests geen theoretisch bewijs voor het door Searle aangegeven *intuïtieve* onderscheid! Natuurlijk niet, zou Searle kunnen zeggen, omdat een dergelijke test niet op ‘compelling intuitive grounds’ uit kan komen. Dit intuïtieve onderscheid is echter ‘painfully obvious’ voor Searle. Ik denk dat hij dit zo ziet, omdat deze intuïtie een oorsprong vindt in zijn visie, die intrinsieke eigenschappen die vanuit het eerste persoonsperspectief als essentieel en reëel beschrijft. In dit debat is mijn persoonlijke keuze voor de ‘welles’-kant (van Searle en anderen) nu te beschrijven als een logisch gevolg van het delen van deze, nu explicieter beschreven, oorsprong.

Ook Dennett is het er (letterlijk) niet mee eens dat er zoiets als intrinsieke intentionaliteit bestaat (1987, p. 337 en uitleg in hoofstuk 8), vooral niet als een privé-eigenschap voor het subject die alleen via het bewustzijn bereikbaar is. De uitgebreide discussie tussen Dennett en Searle hierover en in het algemeen is ook bijzonder interessant, maar reikt verder dan waar deze scriptie over gaat[11]. De bron van de discussie tussen Searle en Dennett (en Hauser) lijkt, onder andere, te bestaan uit het wel of niet willen erkennen van de speciale status van de mens als ‘eigenaar’ van intrinsieke intentionaliteit. Als je niet wilt erkennen dat intrinsieke intentionaliteit bestaat, is het natuurlijk niet lastig om mogelijkheden te zien voor de Harde KI om een mind te maken; het is alleen dan niet hetzelfde concept ‘mind’ waar je over spreekt: juist datgene wat constitutief is (volgens de hier uiteengezette positie) voor minds mist dan! De vragen, ‘Wat is een mind precies?’, ‘Wat wordt er veroorzaakt door welke causale krachten in het brein?’ zijn het terrein van de filosofie van mind en de wetenschap van de hersenen, en zijn relevant voor de opvattingen van de Harde KI. De intuïtieve plausibiliteit van het erkennen van een intrinsiek fenomeen als intentionaliteit (en de causale krachten van de hersenen), die nu hopelijk iets explicieter uitgewerkt zijn, zorgen voor de botsende ideeën tussen filosofie en Harde KI voor sommige filosofen binnen de filosofie van mind.

[11] Haugeland probeert de twee te verenigen in “*Having Thought*” (1998), hoofdstuk 12: ‘*Understanding: Dennett and Searle*’, middels een beschrijving van de fundamentele overeenkomsten van hun uitgangsposities. Deze poging is eerder in dit hoofdstuk al kort aan bod gekomen.

Een tussentijdse samenvatting: de causale krachten komen voort uit het brein, en zijn de veroorzakers van mentale (en tevens biologische) fenomenen als intentionaliteit en bewustzijn. *Semantiek hebben* is een teken van of een voorwaarde voor intentionaliteit en mentaliteit – een 'marker' van de juiste causale krachten hebben. Hoe deze onderlinge relaties volgens Searle en volgens anderen precies in elkaar zitten is een open vraag, maar niet de vraag van dit onderzoek. De 'details' doen er minder toe dan het feit dat semantiek en intentionaliteit, mentaliteit, allemaal met elkaar samenhangen. Of een van de begrippen fundamenteler of essentiëler is, en zo ja, hoe, is een bijzonder lastige vraag. Een theorie in de filosofie van mind moet al deze fenomenen een plaats kunnen geven, volgens Searle (deel uit een reeds gegeven citaat):

> '**But any satisfactory account of the mind and of the mind-body relations must take account of all four features.** If your theory ends up denying any one of them, you know you have made a mistake somewhere' (1984, p. 17) (mijn nadruk).

Volgens het biologisch naturalisme van Searle *kunnen* andere systemen dan de hersenen (ook a priori) dezelfde krachten hebben. De juiste causale krachten kunnen in welk systeem dan ook zorgen voor 'semantiek'. Er bestaat geen algemene consensus over wat de juiste causale krachten zijn; de meningsverschillen in dit debat komen neer op zeer fundamentele kwesties over wat de intrinsieke en speciale eigenschappen van de hersenen precies zijn.

3.4. Wat is de relevantie van semantiek?

Een nieuwe vraag bij de voorafgaande informatie is nu: Waarom zouden we eigenlijk willen dat andere systemen (systemen in de KI) semantiek (intentionaliteit, mentaliteit) hebben? Waarom is dit nodig voor 'persoon zijn'? Wat is er essentieel aan semantiek hebben voor 'persoon zijn'? Searle stelt deze vragen niet op deze manier centraal. Dat is begrijpelijk, omdat hij zich waarschijnlijk niet gedwongen voelt om dit te doen: de redenen hiervoor zijn vanuit zijn intuïtie al overduidelijk. Daarom kan ik alleen op basis van het voorafgaande conclusies trekken over de relevantie, en steun zoeken bij andere auteurs. In deze paragraaf zal ik daarom de opvattingen van John Haugeland en Joseph Rychlak gebruiken om meer (expliciete) diepgaande steun voor de positie van Searle te creëren.

3.4.1. Haugeland: relevantie van intrinsieke intentionaliteit

Haugeland ondersteunt Searles conclusie over Harde KI, maar niet alle aannames; hij heeft een andere manier van het centraal stellen van 'het menselijke' van de mens (Haugeland, 2002). Haugeland stelt in de introductie van zijn boek '*Having Thought*':

> 'Understanding is the mark of the human. This is a better way to make the point, and for two reasons. On the one hand, understanding is *not* exclusively *mental* but is essentially corporeal and worldly as well; but, on the other, it *is* exclusively (and universally) *human*. Accordingly, intentionality, rationality, objective knowledge, and self-consciousness, properly understood, are likewise exclusively human. By 'human', I don't mean specific to homo sapiens. Humanity is not a zoological classification, but a more recent social and historical phenomenon – one which happens, however, so far as we know, to be limited to homo sapiens' (1998, p. 1).

Haugeland ziet intentionaliteit als een eigenschap die exclusief is voor mensen, en 'menselijkheid' als een sociaal en historisch fenomeen dat – zo ver als wij weten – alleen weggelegd is voor de soort 'homo sapiens'.

Haugelands antwoord op de vraag waarom semantiek en de ermee samenhangende fenomenen ertoe doen, is de *normativiteit* van intentionaliteit. De opvatting dat intentionaliteit normatief is, delen Dennett en Searle zelfs (Haugeland, 1998, p. 294). Volgens Haugeland is iedereen het erover eens dat intentionaliteit *op de een of andere manier* normatief is. Haugeland beschrijft dat deze normativiteit volgens Searle gerelateerd is aan het hebben van de bij intentionele toestanden horende 'satisfaction conditions':

> 'For Searle, the normative element shows up in what he calls the *satisfaction conditions* for intentional states; and, needless to say, he maintains that the having of satisfaction conditions is itself also intrinsic to intentional mental states' (Haugeland, 1998, pp. 294-295).[12]

Het probleem is volgens Haugeland, dat Searle en Dennett beiden niet genoeg te zeggen hebben over *hoe* een fysisch systeem deze normatieve eigenschappen *intrinsiek* zouden kunnen hebben (1998, p. 295). De vraag (waarvan een adequaat antwoord Dennett en Searle kan verenigen, volgens Haugeland) is dan ook:

> 'How can naturally evolved physical brain configurations be normative in this way *intrinsically*? This is the question that I want to sketch an answer to – an answer that superficially sides with Searle against Dennett, but more deeply brings the two of them together' (1998, pp. 295-296).

Haugeland beschrijft deze schets van zijn kijk op 'intrinsicness' en 'intrinsic intentionality' als volgt:

[12] Zie ook Searle over normativiteit in '*The Rediscovery of the Mind'* (1992, pp. 51-52 en 238).

> ‘**Commitment to standards** is the very foundation and essence of intrinsic intentionality. To see this, we must examine what is meant by ‘intrinsic’. Obviously, the intentionality of some particular belief or desire cannot be intrinsic to that individual state all by itself. That would be incompatible with the holism of intentionality and its essential dependence on the background – two theses which Searle (among others) has espoused for years (1992, 175-178). Rather, the intrinsicness must pertain at once to all the intentional states of a single system; and it means that the intentionality of those states is independent of any other intentionality – that is, the intentionality of another system. [...] [their] intentionality is *derivative* from ours, and in that sense *observer* (or user) *relative*. Intrinsic intentionality is not thus derivative (78-82 and 211-212).
>
> That, however is merely a negative characterization: it says what intrinsic intentionality is *not*. A genuine positive characterization would have to show how intentionality that is not derivative is even possible – that is, *how it is possible* for any system to have intentional states *on its own*. (This amounts to showing how intentionality is possible at all, since derivative intentionality only makes sense if there is some nonderivative intentionality for it to be derivative from.) It’s worth remarking in passing that what counts as a single “system” is so far a free parameter in the account; Searle often speaks as if the relevant systems are individual brains, but that’s not built into the theory. What matters is that the unity of the system follow from the same positive account of how such a system could have intentional states on its own – that is, intrinsically.
>
> But that’s precisely what commitment to constitutive standards provides. The unity of the system is the unity of a single consistent commitment in terms of which a plurality of intentional states can be normatively beholden to their constituted satisfaction conditions. Moreover, it is the basis of the necessary *subjectivity* of intentional states. ‘Subject’ here cannot just mean grammatical subject: even an adding machine is the “subject” if its states in that sense. Rather, ‘subject’ means something like “author and owner” – someone who is responsible for the states in question, and to whom they matter. Surely that subject is non other than the one who is committed to the very standards that render these states intentionally normative in the first place. My commitment to getting my intentional states “right” is what *makes* their intentionality my *own* – that is, *intrinsic* to me.* In other words: *subjectivity* as such is constituted’ (1998, pp. 299-300) (mijn bold nadruk).
>
> *Haugelands voetnoot (p 304): ‘I compare this discussion of Searle and intrinsic intentionality to chapter 7, note 9, above. I am in effect suggesting here that the term ‘intrinsic’ should be reserved for that special case of original intentionality (in my old sense) in which the intentional system is *committed* to a constituted domain – otherwise I call it ‘ersatz’ (see sections 10 and 11 of the present essay). I have, however, no reason to believe that Searle (or Dennett) would accept this suggestion [Note added 1997].

Een positieve karakterisering van intentionaliteit (wat is intentionaliteit *wel*?) zou moeten kunnen beschrijven hoe welk systeem dan ook *op zichzelf* intentionele toestanden *kan* hebben. Een dergelijke beschrijving van een niet-afgeleide vorm van intentionaliteit is de enige beschrijving die inzicht kan geven in wat ‘afgeleide’ intentionaliteit zou kunnen betekenen. Volgens Haugeland geeft het hebben van intentionele toestanden hebben een norm voor de eenheid van een systeem. Een systeem is een *eenheid* omdat het zich als geheel intrinsiek intentioneel verbindt (committeert) aan bepaalde standaards. Deze verbintenis (via intrinsieke

intentionaliteit) is dus constitutief voor de standaard van *eenheid* van een systeem. Deze standaard zorgt voor de 'normen' voor intentionele toestanden. Het systeem als eenheid kan dus normatieve intentionele toestanden hebben vanwege die ene vorm van verbintenis. De verbintenis van een systeem om zijn eigen intentionele toestanden 'op orde' (genormaliseerd) te hebben is hetgeen dat ervoor zorgt dat de intentionaliteit van de toestanden aan het systeem toe te schrijven zijn; dit is hoe volgens Haugeland de intrinsieke status van intentionaliteit in het systeem, en daarmee *subjectiviteit* tot stand komt.

Deze uitleg van Haugeland is 'nieuw' en verre van algemene consensus: ik vind hem echter bijzonder sterk als een stap in een interessante richting voor het fundamenteel bespreken van intentionaliteit, subjectiviteit en de 'persoonstatus' van een systeem: dat wat het systeem een intentionele eenheid maakt. Dit is een mogelijke, expliciete en unieke verantwoording van waarom intrinsieke intentionaliteit essentieel is voor 'mens zijn' op meer dan een 'welles'-'nietes'-niveau; het is een poging tot filosofische verantwoording (waarmee verder geredeneerd kan worden).

3.4.2. Rychlak: relevantie van semantiek voor redeneren

Joseph Rychlak, een psycholoog die zich voornamelijk op het gebied van de filosofie van de psychologie richt, heeft een eigen 'opvatting' over de *aard* van personen. In ten minste twee van zijn boeken ((Rychlak, 1991) en (Rychlak, 1997)) doet hij een pleidooi voor de aparte status van menselijk redeneren en menselijk bewustzijn. In '*Artificial Intelligence and Human Reason: a Teleological Critique*' (1991) doet hij een aanval op KI door de bijzondere of eigenaardige status van het menselijk redeneren te benadrukken. In het eerste hoofdstuk beschrijft hij de implicaties van de Chinese Kamer, en voorziet deze van zijn eigen onderbouwingen. Samenvattend volgt hieronder zijn redenering voor het verschil tussen menselijk redeneren en redeneren in computers, en wat de relevantie van semantiek (betekenis) van proposities voor menselijk redeneren is.

Searle herinnert ons eraan dat betekenisvolle informatie niets te doen heeft met de pure machine-eigenschappen van een computer (1991, p. 6). Betekenisvolle informatie bestaat vooral of alleen op het gebied van introspectief begrijpen, dit is iets dat (perse) totaal voorbij gaat aan de machine, vanwege de aard van de redeneerprocessen in machines. Rychlak maakt dan ook het onderscheid tussen

logische en mechanische processen, om aan te tonen dat deze niet (noodzakelijk) identiek zijn. Betekenis is fundamenteel voor menselijk redeneren, gelooft Searle, volgens Rychlak. Het volgens regels redeneren van computers is niet hetzelfde als (het volgens regels redeneren) van mensen (1991, pp. 6-7) (met verwijzing naar Dreyfuss' '*What computers can't do*' uit 1979).

Redeneren volgens regels zoals *mensen* dat doen is afhankelijk van assumpties en oordelen die buiten de regels die we leren toe te passen vallen. Dit buiten de regels vallen suggereert dat er altijd een bredere context van betekenis is waarin de regels gelegen zijn. De suggestie van Rychlak is nu dat we, wanneer we het over redeneren volgens regels bij mensen hebben, we betrokken zijn bij een *logisch proces* dat geen (mogelijk) *mechanisch proces* is. Regels zijn altijd op een bepaalde manier onderhevig aan vooronderstellingen, en vooronderstellingen belichamen altijd *predicaties*. Predicatie heeft te maken met het ordenen van betekenissen, met de cognitieve daad van bevestigen, ontkennen, of kwalificeren van bredere patronen van betekenis in relatie tot smallere (beperktere) of meer doelgerichte patronen van betekenis. Het categoriseren van algemene informatie naar kleinere categorieën (verzamelingen) middels het gebruiken van predicaten is, korter gezegd, wat predicatie inhoudt volgens Rychlak:

> 'The Greek word *kategorein* means "to predicate," so in a true sense, when we seek to organize, categorize, and classify information we are predicating it in some fashion' (Rychlak, 1991, p. 7)

De persoon in de Chinese Kamer past geen predicatie toe op de figuren die hij koppelt: koppelen is niet hetzelfde als predicaten toeschrijven. Om predicaten te kunnen toeschrijven, heeft de persoon de bredere context van betekenis nodig waarin de *squiggles* en *squoggles* die slechts optreden als tekens voor de semantische informatie die naar binnen gestuurd wordt, te situeren:

> 'In order to predicate, the person requires a wider context of meaning within which to situate the squiggles and squiggles acting as mere signs for the semantic information being sent inward. This context turns such signs into meaningful symbols' (Rychlak, 1991, p. 7).

Binnen een context kunnen die tekens wel betekenisvolle symbolen worden. Zodra een bredere betekenis een beperktere, specifiekere betekenis met zich meebrengt, is de uitbreiding van de bredere naar de bedoelde (beperktere) betekenis ogenblikkelijk een feit (p. 8). Het predicatieproces is een logisch proces, en onafhankelijk van de specifieke woorden die in het proces gebruikt worden. In een predicatieproces ga je

altijd van een bredere naar een beperktere spreiding van betekenis. Menselijk redeneren behelst deze predicatieprocessen; het 'omgaan' met de betekenis van gedachten en proposities. Het is niet slechts het 'afhandelen' (*mediating*) van informatie (zoals de processen in een computer) maar het *werken met* de informatie. Dit is het verschil tussen 'mediation' en 'predication' (p. 8). Door semantiek te 'vatten' zijn mensen in staat om te *werken met de informatie* (to 'predicate'). Dit houdt onder andere de capaciteit tot het maken van 'opposities' in (p. 10). 'Opposition' is het redeneren van 'wat gezegd is' naar 'wat niet gezegd is' en daar de implicaties van inzien. De begrippen 'goed' en 'slecht' zijn, bijvoorbeeld, intrinsiek aan elkaar verbonden; het zijn oppositioneel gerelateerde betekenissen. Op deze manier bestaan er 'bipolaire koppelingen'. Een computer is niet in staat tot oppositioneel redeneren (zie voor Rychlaks uitleg: hoofdstuk vier).

Menselijk redeneren gebeurt essentieel via de semantiek – de inhoud van proposities. Dit maakt (onder andere) oppositie mogelijk, een essentieel verschillende capaciteit dan appositie (een 'minder krachtige' capaciteit die computers wel bezitten, dit komt ook in hoofdstuk vier aan de orde). Menselijk redeneren (via semantiek, bijvoorbeeld in predicatie) is echter essentieel voor 'persoon zijn' of op zijn minst voor het 'als persoon beschouwd worden'. Daarom is de capaciteit 'menselijk redeneren' nodig voor systemen om in de buurt van de status 'persoon' te komen; hiervoor is semantiek (mentale inhouden, intentionaliteit) nodig, om ook oppositioneel te kunnen redeneren.

3.5. Conclusie

De aanname dat hersenen de juiste causale krachten bezitten om mentale inhoud te veroorzaken, de 'substelling' uit hoofdstuk 2, is een bijzonder belangrijke basis voor Searles uitleg van zijn aanname dat minds semantiek hebben. Het is echter niet Searles bedoeling om deze aannames te bewijzen; hij is overtuigd dat het bewijs voor deze stellingen ligt in de neurobiologie (volgens zijn biologisch naturalisme). De relevantie van mentale fenomenen ziet hij als 'gegeven' en wat hij kan zeggen over het ontstaan van die mentale fenomenen is dat het ontologisch subjectieve, maar toch objectief te bestuderen fenomenen zijn. Door deze ontologische subjectiviteit staat het eerste persoonsperspectief centraal, omdat dit het enige perspectief is dat 'toegang' tot objectieve feiten over de fenomenen kan bieden.

Searle wil aangeven dat het in handen van de neurofysiologie en –biologie is om te laten zien hoe mentale fenomenen, waaronder intentionaliteit, plaats kunnen vinden in de hersenen. Hij zegt niet dat andere systemen dan de hersenen niet ook semantiek kunnen hebben, maar wel dat deze dan de causale krachten van het brein moeten kunnen evenaren of dupliceren of dat er 'krachten' moeten zijn die hetzelfde resultaat kunnen bewerkstelligen.

'Semantiek' staat Searle voor een algemene notie die de intrinsieke inhoud van mentale fenomenen aanduidt: intrinsieke intentionaliteit is de belangrijkste 'capaciteit' (die via de juiste causale krachten in de hersenen kan ontstaan) die hiermee samenhangt. Mentale inhoud is gerelateerd aan intrinsieke intentionaliteit. Hoe dit samenhangt, kan hij (nog) niet beschrijven: het antwoord hierop moet uit neurobiologisch onderzoek van de hersenen komen.

De vraag waar semantiek voor staat, waarom het relevant is voor personen, is voor deze discussie belangrijker dan de vraag of personen 'semantiek hebben'; volgens de redenering tegen Harde KI is een systeem zonder de juiste semantiek niet als persoon te beschouwen, omdat het systeem niet op dezelfde manier kan redeneren als een persoon. Semantiek is relevant, omdat het een voorwaarde *of* een gevolg van intentionaliteit is. Intrinsieke inhoudelijke toestanden (en hiermee 'semantiek hebben') zijn essentieel voor menselijk redeneren (in navolging van Rychlak) en de eenheid van een systeem (in navolging van Haugeland). Om persoon te zijn, moet een systeem 'menselijk' kunnen redeneren en zich als systeem aan bepaalde standaards verbinden (om normatieve intentionele toestanden te kunnen hebben). Zonder intrinsieke inhoud is het niet mogelijk om iemand een persoonstatus (als redenerend, 'verbonden' persoon) toe te kennen. Harde KI die niet streeft naar een mind met intrinsieke inhoud, kan geen persoon 'maken'.

4. Searle: syntax, semantiek en KI

In dit hoofdstuk combineer ik de bespreking van de aannames 3 en 1, over de aard van computers en de onmogelijkheid om op basis van syntax semantiek te kunnen genereren. Hoe gebruikt Searle een bepaalde standaardopvatting uit de logica en taalkunde en combineert hij deze met opvattingen over computers om zo tot de conclusie te komen dat computers geen semantiek hebben? Heeft hij zelf zijn syntax-semantiekredenering aangepast of van fundamenteel aanvullende uitleg voorzien na het verschijnen van het Chinese Kamergedachte-experiment en de SSR?

4.1. Searle en 'Syntax <=/> semantiek als logische waarheid' voor Harde KI

Om Searles positie te bekijken, zijn de vragen in deze paragraaf hoe hij de aanname 'syntax is niet voldoende voor semantiek' 'onderbouwt' (4.1.1) en vervolgens toepast op de KI (4.1.2). Hoe past Searle de opvattingen over syntax en semantiek vanuit de logica en taalkunde toe op de KI, en waarom is syntax in computers als onvoldoende voor semantiek (benodigd voor minds) te zien?

4.1.1. Searle en de 'conceptual truth' (aanname 3)

De derde aanname van Searle luidt: 'Syntax (alleen) is niet voldoende voor semantiek.' Het is bijna onmogelijk om teveel nadruk te leggen op het belang van deze relatief simpele aanname voor de syntax-semantiekredenering (Preston, 2002, p. 34). Wat houdt deze stelling in, waar komt deze aanname vandaan? Zijn Searles uitspraken erover te verantwoorden?

Het onderscheid tussen syntax en semantiek en de barrière die ertussen bestaat is voor Searles redenering van bijzonder groot belang.

> '*Syntax is not sufficient for semantics.*.That proposition is a conceptual truth. It just articulates our distinction between the notion of what is purely formal and what has content' (Searle, 1984, p. 39).

Preston geeft toe dat deze aanname voor het computationalisme binnen de cognitiewetenschap vitaal is:

> 'Thinkers from outside computational cognitive science, such as some followers of Wittgenstein, have occasionally tried to challenge the integrity of this distinction. But that move is not available to computationalists' (Preston, 2002, p. 34).[13]

[13] Preston geeft geen verantwoording van deze uitspraak! Ik denk dat hij het simpelweg eens is met Searle in dit debat over cognitiewetenschappen en het probleem van de barrière (het onderscheid) tussen syntax en semantiek. Cognitiewetenschappers kunnen het onderscheid en daarmee de barrière niet zomaar ontkennen.

In het originele Chinese Kamerartikel komen de termen ‘syntax’ en ‘semantiek’ niet vaak voor. De redenering wordt in dat artikel vooral uiteengezet in termen van vorm en inhoud (Searle, 1980), (Preston, 2002, p. 34 voetnoot). In de later uiteengezette versies van het ‘Brutally Simple Argument’ worden de termen syntax en semantiek echter wel centraal gesteld (Searle, 1984) en (Searle, 1980). Zoals in hoofdstuk drie al is bekeken, is het gebruik van de term semantiek *intuïtief* aanvaardbaar als refererend aan ‘intentionaliteit’ en ‘mentale inhoud’. De kracht en het nut van het gebruiken van het woord ‘semantiek’ schuilt in de *strikte scheiding* van semantiek en syntax, samen met de opvatting dat syntax *niet voldoende* is voor semantiek. Met het Chinese Kamer scenario probeert Searle ons aan de evidentie van dit onderscheid te herinneren (zie hoofdstuk twee). Hij noemt het ‘a very simple logical truth’[14].

Het probleem bij het bespreken van deze stelling is dat Searle zich totaal baseert op deze waarheid uit de taalkunde en logica, en niet veel energie steekt in een verdediging van deze aanname.

> ‘The Churchlands complain that I am “begging the question” when I say that uninterpreted formal symbols are not identical to mental contents. Well, I certainly did not spend much time arguing for it, because I take it as a logical truth’ (Searle, 1990, p. 25) en (Moural, 2003, p. 260 voetnoot).

Hij voelt zich niet gedwongen tot het verdedigen van de stelling, het is namelijk een ‘conceptual truth’, ‘which, admittedly, is a well established view in both logic and linguistics’ (Moural, 2003, p. 249). Het is dus geen onbekende of ongeaccepteerde stelling voor de logica en de taalkunde.

In *Artificial Intelligence: a modern approach*, beschrijven Russell en Norvig in de eerste-orde logica als de taal waarmee ‘knowledge-based agents’ gemaakt kunnen worden (Russell & Norvig, 1995, p. 185). Over representaties en semantiek in logische talen schrijven ze:

> ‘In logic, the **meaning** of a sentence is what it states about the world, that the world is in *this* way and not *that* way. So how does a sentence get its meaning? How do we establish the correspondence between sentences and facts? Essentially, this is up to the person who wrote the sentence. In order to say what it means, the writer has to provide an **interpretation** for it; to say what fact it corresponds to. A sentence does not mean something *by itself*’ (1995, pp. 161-163).

[14] Preston geeft bij zijn citaat van deze zin maar liefst negen werken waarin Searle deze stelling geeft (Preston, 2002, p. 35).

Oftewel: de vorm of ‘verschijning’ van een zin alleen is niet voldoende om de betekenis ervan te ‘openbaren’. Het is, zoals uit het citaat van Moural ook blijkt, een feit dat in taalkunde en logica de syntax en de semantiek van talen altijd apart van elkaar, in verschillende systemen van uitleg of definitie, beschreven en bekeken worden. Logische talen zijn nooit ‘zonder semantiek’: de semantiek geeft aan hoe de symbolen die gemanipuleerd worden door de syntax geïnterpreteerd dienen te worden. Zodoende gebeurt dit ook zo in de logische talen die in de KI veelvuldig gebruikt worden. Deze consensus bestaat in zo goed als alle gebieden van de KI, bij voorstanders (zoals Russel & Norvig) en tegenstanders (zoals Searle) van de Harde KI, en bij neutrale ‘beschouwende’ filosofen (zoals Moural).

In de taalfilosofie wordt uiteraard ook aandacht besteed aan semantiek; als de betekenis van taal en van mentale inhoud van gedachten in het bijzonder. ‘Wat is betekenis?’ is een vraag die bijzonder uitgebreid wordt behandeld in de taalfilosofie (onder andere door Searle). Een beknopte samenvatting van de meningen in dit debat is naast zo goed als onmogelijk, ook geen doelstelling voor deze scriptie. Het is wel relevant voor het begrip van Searles positie, dat hij betekenis van taaluitingen in de taalfilosofie (in het geval van mensen) ziet als sterk afhangend van intentionaliteit (Searle, 1994a, pp. 380-383) en ook (Searle, 1983, p. vii).

De stelling is, vanuit de logica en taalkunde gezien, inderdaad een conceptuele, logische waarheid. De overkoepelende en relevantere kwestie is, hoe dit inzicht in de filosofie van mind en filosofie van de KI bekeken en gebruikt kan worden. Zijn Searle en anderen in de filosofie van mind gerechtvaardigd dit inzicht toe te passen op mentale inhoud *in minds* (hier gaat het weer over de discussie of semantiek in minds daadwerkelijk aanwezig is, zie hoofdstuk drie over de discussie van Searle en Dennett), **en** hiermee conclusies te trekken over de mogelijkheid van dergelijke mentale inhoud *in computers*?

4.1.2. Searle: syntax is voldoende noch noodzakelijk voor semantiek voor KI

De eerste aanname van Searle luidt: ‘computers hebben slechts syntax, en geen semantiek’. De stelling ‘syntax is niet voldoende voor semantiek’ zoals hierboven beschreven is een essentiële aanname om ook deze uitspraak te kunnen doen. Wat is de reden van Searle om te zeggen dat computers ‘alleen syntax’ hebben, en daarom geen semantiek? Waarom kan die semantiek er niet toch zijn, bijvoorbeeld veroorzaakt door iets anders dan de syntax alleen? Om dit te bekijken, is het handig

om Searles relevante uitspraken over digitale computers (gebruikt in het computationalisme) te noemen en te zien hoe deze beschrijvingen zijn redenering ondersteunen.

4.1.2.1 Basis van computatie en Harde KI - Searle

Searles redenering richt zich tegen de opvatting van het computationalisme, dat 'all there is to having a mind is having the right program (running, implemented on a system).' De notie van computatie in digitale computers is hiervoor een belangrijk basisbegrip; deze moet het gebruik van computationele systemen om een *mind* te kunnen maken ondersteunen. Deze ideeën komen samen in de *computationele theorie van mind*. De uitleg van deze theorie is het onderwerp van deze paragraaf.

De (standaard) digitale computer is gebaseerd op het principe van de Turing Machine, en het uitvoeren van algoritmes. In 'Computation and mental processes' (2004a, pp. 46-52) geeft Searle een korte beschrijving van de belangrijkste begrippen die bij digitale computatie horen, en de computertheorie van mind volgens Searle beschrijven. Deze begrippen zijn: algoritme, Turing Machine, Church' Thesis, Turings theorema, de Turing test, niveaus van beschrijving, multipele realiseerbaarheid en recursieve decompositie. Ik zal hier de volgende begrippen zeer kort beschrijven: algoritme, Turing Machine, niveaus van beschrijving en multipele realiseerbaarheid. Het zijn basisbegrippen die niet echt 'afhankelijk van wat Searle vindt'; zijn conclusies erover zijn belangrijker.

Een **algoritme** is een methode om een probleem op te lossen door een gespecificeerde serie van handelingen uit te voeren. Deze specificatie garandeert een oplossing, het zijn 'effectieve procedures' (2004a, pp. 46-47). Een **Turing Machine** is een abstract 'apparaat' dat berekeningen (algoritmes) uit kan voeren met gebruik van slechts twee types symbolen. Het is een abstract, mathematisch concept, en geen 'echte' machine (die je kunt aanraken). De computers die wij tegenwoordig gebruiken zijn wel gebaseerd op deze abstracte computers (2004a, p. 47). Het manipuleren van 0-en en 1-en tot op zeer abstract en hoog niveau is nog steeds de essentie van een digitale computer. Dit vormt het grote probleem voor de Harde KI volgens Searle: computer programma's zijn formeel, syntactisch:

> 'Axiom 1. *Computer programs are formal (syntactic).* This point is so crucial that it is worth explaining in more detail. A digital computer processes information by first encoding it in the symbolism that the computer uses and then manipulating the symbols through a set of precisely stated rules. These rules constitute the program. For example,

> in Turing's early theory of computers, the symbols were simply 0's and 1's, and the rules of the program said such things as, "Print a 0 on the tape, move one square to the left and erase a 1." The astonishing thing about computers is that any information that can be stated in a language can be encoded in such a system, and any information-processing task that can be solved by explicit rules can be programmed' (Searle, 1990, p. 21).

In dit citaat beschrijft Searle tevens dat de mogelijkheid van het coderen van alle typen informatie die talig kan worden weergegeven, computers zulke 'astonishing' apparaten maakt.

Voor elk complex systeem bestaan meerdere **niveaus van beschrijving**. Vooral op computers is het goed mogelijk deze verschillende manieren van beschrijving[15] toe te passen. Zo kunnen op een 'laag' niveau twee computers verschillende typen processoren (en dus verschillende beschrijvingen) hebben, terwijl op een abstract niveau dezelfde beschrijving (van bijvoorbeeld het programma dat ze uitvoeren) van toepassing kan zijn (Searle, 2004a, p. 49). De notie van niveaus van beschrijving bevat impliciet een andere notie die essentieel is voor de computationele theorie van mind; die van **multipele realiseerbaarheid**. Hetzelfde computerprogramma kan op verschillende manieren geïmplementeerd zijn in verschillende soorten hardware:

> 'Indeed, this is the feature of digital computers that makes them so powerful. One and the same type of hardware, if it is appropriately designed, can be used to run an indefinite range of different programs. And one and the same program can be run on an indefinite range or different types of hardwares' (Searle, 1984, p. 31).

Voor de computationele theorie van de mind betekent dit dat een mentale toestand ook in 'verschillende soorten hardware' (breinen, computers?) gerealiseerd zou moeten kunnen worden. Dit is, als we kijken naar het biologisch naturalisme, niet onmogelijk, maar het laat het hebben van de juiste causale krachten buiten beschouwing.

4.1.2.2 Searles kritiek op de computertheorie van mind

De aard van een digitale computer en de rol van de symbolen beschrijft Searle als volgt:

> 'It is essential to our conception of a digital computer that its operations can be specified purely formally; that is, we specify the steps in the operation of the computer in terms of abstract symbols – sequences of zeroes and ones printed on a tape, for example. [..] *But* the symbols have no meaning; they have no semantic content; they are not about anything. They have to be specified purely in terms of their formal or syntactical structure. The zeroes and ones, for example, are just numerals; they don't even stand for numbers' (Searle, 1990, pp. 30-31)(mijn nadruk).

[15] Het woord 'niveaus' wordt meestal gebruikt om deze verschillen te typeren.

Searle ziet de kracht van computers, het abstractieniveau van syntax zonder semantiek, dus tegelijk als het zwakke punt. Om een mind te kunnen genereren, is het niveau van abstractie van de syntax niet voldoende; de symbolen hebben op dat niveau geen betekenis.

> 'But this feature of programs, that they are defined purely formally or syntactically, is fatal to the view that mental processes and program processes are identical. And the reason can be stated quite simply. There is more to having a mind than having formal or syntactical processes. [..] Minds are semantical, in the sense that they have more than a formal structure, they have a content' (Searle, 1984, p. 31).

Voor een mind is meer nodig dan **slechts formele symboolmanipulatie**, omdat een mind semantiek heeft en semantiek nodig heeft. Symboolmanipulatie in computers is niet voldoende om semantiek te kunnen genereren. Met 'semantiek' duidt Searle op *intentionaliteit*:

> 'But the main point of the present argument is that no purely formal model will ever be sufficient by itself for intentionality because the formal properties are not by themselves constitutive of intentionality, and they have by themselves no causal powers except the power, *when instantiated*, to produce the next stage of the formalism when the machine is running' (Searle, 1980, p. 247)(mijn nadruk).

Uit dit citaat kunnen we het volgende concluderen:

1) **Formele modellen** zijn niet constitutief voor intentionaliteit en hebben op zichzelf niet de juiste causale krachten.

   Dit is het probleem dat aangeduid wordt met: formele modellen hebben *'slechts'* symboolmanipulatie.

2) Searle heeft het ook over instantiaties van programma's, en dus niet slechts over ongerealiseerde of ongeïmplementeerde programma's.

Bij deze redenering zijn twee aanmerkingen over de aard en het gebruik van **symbolen** belangrijk:

> '**First**, symbols and programs are purely abstract notions: they have no essential physical properties to define them and can be implemented in any physical medium whatsoever. The 0's and 1's, qua symbols, have no essential physical properties and a fortiori have no physical, causal properties. I emphasize this point because it is tempting to identify computers with some specific technology – say, silicon chips – and to think that the issues are about the physics of silicon chips or to think that syntax identifies some physical phenomenon that might have as yet unknown causal powers, in the way that actual physical phenomena such as electromagnetic radiation or hydrogen atoms have physical, causal properties. The **second** point is that symbols are manipulated without reference to any meanings. The symbols of the program can stand for anything the programmer or user wants. In this sense the program has syntax but no semantics' (Searle, 1990, p. 21) (mijn nadruk).

Symbolen in de computer hebben geen fysische, causale eigenschappen; deze kunnen ook niet ‘overgeërfd’ of ‘geleend’ worden van de fysische basis waarin ze gerealiseerd worden: deze basis is namelijk niet definiërend voor de symbolen (het abstracte niveau is definiërend, aldus het computationalisme). De symbolen hebben geen eigen betekenis doordat ze op zo’n hoog abstractieniveau behandeld worden. De enige betekenis die eraan gegeven kan worden is *onder interpretatie van de programmeur of gebruiker.* In het computationalisme is er dus absoluut geen sprake van ‘intrinsieke’ semantiek, en daarmee geen mogelijke veroorzaking van intrinsieke intentionaliteit; de juiste causale krachten missen:

> ‘*[..] formal symbols have no physical, causal powers*. The only power that symbols have, qua symbols, is the power to cause the next step in the program when the machine is running’ (Searle, 1990, p. 24).

De causale krachten moeten er zijn om semantiek te bewerkstelligen: ‘een vliegtuig hoeft geen veren te hebben om te kunnen vliegen, maar wel dezelfde causale krachten om de zwaartekracht te kunnen verslaan’ (1994b, p. 547).

In hoofdstuk drie kwam de stelling van Rychlak aan bod, dat een computer niet in staat is tot *oppositioneel* redeneren, vanwege de computationele (mechanische) basis. In mensen zijn symbolen en concepten *bipolair* (1991, p. 11). In computers (en in de Chinese Kamer) zijn de symbolen alleen maar *unipolair*. Rychlak biedt een (mogelijke) verklaring voor waarom computers hiertoe niet in staat zijn: de ‘wel-of-niet’ en niet ‘mogelijk allebei’ Booleaanse logica staat dit niet toe (p. 12). De machine heeft slechts de capaciteit tot appositie, patronen van items (betekenissen, etc.) *naast* elkaar plaatsen (p. 13).

> ‘The upshot of this inability to reason oppositionally is that a machine cannot learn what was *not* input. If we send the Ten Commandments inward to a machine it would record ten ways in which to behave. A person, on the other hand, would have *at least* twenty possibilities suggested based on precisely the same input. Even if a machine is programmed to reason oppositionally, this will have to be accomplished based upon the software or the “rule book” in Searle’s example. The hardware processing is never oppositional. As a result, the machine will be **simulating** oppositional reasoning with as little understanding of oppositional reasoning as Searle’s closeted person has of the Chinese figure. Moreover, the machine will never be capable of reasoning to the opposite of the programmed injunctions, negating what it is instructed to do by deciding to affirm the contrary course’ (1991, p. 13) (mijn bold nadruk).

De computer is dus hooguit in staat om oppositioneel redeneren te *simuleren*, zonder de oppositionele eigenschappen van het redeneren in te zien.

Het is hopelijk duidelijk voor de lezer dat de redenering van Searle niet afhankelijk is van het huidige stadium van technologie, zolang als de basis van deze technologie dezelfde is; namelijk die van symboolmanipulerende abstracte Turing Machines, geïmplementeerd in steeds snellere of betere technologie:

> 'And there is no question of waiting on further research to reveal the physical, causal properties of 0's and 1's. The only relevant properties of 0's and 1's are abstract computational properties, and they are already well known' (Searle, 1990, p. 24)

> 'It is important to emphasise this point because the temptation is always to think that the solution to our problems must wait on some as yet uncreated technological wonder. But in fact, the nature of the refutation is completely independent of any state of technology. It has to do with the very definition of a digital computer, with what a digital computer is' (Searle, 1984, p. 30).

In het geval van het computationalisme en Harde KI die gebruik maakt van 'standaard' digitale computers is de redenering dus 'afgesloten' en niet afhankelijk van verdere resultaten. De praktijk van de KI is dus geen plek om naar tegenargumenten te zoeken. Deze moeten dus meer uit filosofische beschouwingen van het onderwerp komen!

4.1.3. Conclusie

Een digitale computer mist volgens Searle een belangrijk ingrediënt om de in de algemene consensus bestaande 'logische barrière' tussen syntax en semantiek te overbruggen: **causale krachten.** Deze missen in de digitale computers, omdat ze niet kunnen ontstaan door 'slechts' syntax. Het niveau van syntax is het niveau waarop computerprogramma's beschreven worden; digitale computers en hun programma's zijn gebaseerd op de abstracte notie van een Turing Machine. Enige vorm van semantiek volgt (volgens de logische scheiding tussen syntax en semantiek) niet uit deze basis, ook niet uit 'running' programma's of in op Turing machine gebaseerde computationele systemen die in de toekomst nog gemaakt zullen worden. Zolang de basis van specificatie van programma's formeel en abstract is, is elke implementatie van een programma hooguit voldoende voor een simulatie van menselijk redeneren, van minds, van 'personen'. Semantiek is nodig voor redeneren, omdat menselijke redeneerprocessen ook via semantiek verlopen, zoals bijvoorbeeld oppositieprocessen.

4.2. Searle en 'intrinsic features in nature'

Searle heeft naast zijn 'eerste' redenering tegen computationalisme (syntax is niet voldoende voor semantiek) nog een tweede redenering 'gevonden' om zijn eerdere redenering kracht bij te zetten of aan te vullen, namelijk dat syntax niet intrinsiek is aan de fysica. (1994b, p. 547). Hij geeft zelf toe dat hij dit in 1980 nog niet besefte, terwijl hij het toen wel had kunnen gebruiken (1992, p. 210) en (1995, p. 210). Dit nieuwe argument heeft te maken met 'intrinsic features of the world'; wat zijn intrinsieke fysische feiten over een systeem die het systeem computationeel maken? Searles antwoord hierop maakt de computationele theorie van mind incoherent (1994b, p. 548). Hieronder zal ik zijn verdere uitleg (met behulp van citaten) geven om inzicht te geven in de kracht van deze nieuwe redenering. Deze (heldere) redenering, die mijns inziens gelukkig niets met de vaagheid van het Chinese Kamer gedachte-experiment van doen heeft, is impliciet gebleven binnen de syntax-semantiekredenering, maar vormt wel een versterkende en verklarende filosofische fundering ervoor (1995, p. 209).

4.2.1. Computatie, syntax en natuurwetenschappen

In zijn interview in 'Speaking Minds' beschrijft Searle het probleem van computationalisme; dit ligt in het identificeren van mind met een computerprogramma. Deze identificatie is een fout van de KI en de cognitiewetenschappen. De verwarrende opvatting over computatie is dat computatie een proces is dat intrinsiek aanwezig is in de natuur (zoals zwaartekracht):

> 'such notions as implementing a program, being a computer, and manipulating symbols *do not name intrinsic processes of nature in the way that gravitation, photosynthesis, or consciousness are intrinsic processes of nature*. Being a computer is like being a bathtub or a chair in that it is only relative to some observer or user that something can be said to be a chair, a bathtub or a computer' (1995, p. 205) (mijn nadruk)..

Computatie bestaat echter alleen onder interpretatie: computatie wordt 'toegeschreven' aan fysische processen. Dit vereist een interpretator, en maakt een computationele beschrijving waarnemerrelatief (relatief aan de interpretator).

> 'The consequence of this is that there is no way you could discover unconscious computational processes going on in the brain. The idea is incoherent, because computation is not discovered in nature; rather, computational interpretations are assigned to physical processes' (1995, p. 205).

Computatie wordt niet ontdekt in de natuur, maar eraan toegeschreven. Computatie is niet intrinsiek aan de natuur. De bijzondere status van intrinsieke fysische processen is echter essentieel. In de natuurwetenschappen wordt een wereld beschreven die waarnemeronafhankelijk is, de intrinsieke eigenschappen van deze wereld zijn onafhankelijk van een waarnemer:

> 'The natural sciences describe features of reality that are intrinsic to a world that is independent of any observer. That is why gravitation, photosynthesis, and so on can be subjects of natural science – because they are intrinsic features of reality' (1995, p. 209).

Computatie is gedefinieerd in termen van symboolmanipulatie, maar een 'symbool' is geen notie binnen de natuurwetenschappen: een symbool is alleen een symbool omdat een waarnemer het als een symbool opvat. Er zijn geen fysische eigenschappen van een symbool die bepalen dat het een symbool is. 'Symbool zijn' is geen intrinsieke eigenschap van entiteiten in de natuur. Symboolmanipulatie en daarmee computatie bestaan dus alleen relatief aan een waarnemer die een computationele interpretatie aan een fenomeen 'opleggen': 'So computation exists only relative to some agent or observer who imposes a computational interpretation on some phenomenon' (1995, p. 209). Als we nu willen weten of mentale verschijnselen als bewustzijn intrinsiek computationeel kunnen worden opgevat, is het antwoord dat niets intrinsiek computationeel kan worden opgevat, vanwege de waarnemerrelativiteit van het interpretatieproces dat nodig is om computatie toe te schrijven:

> 'But if the question is whether consciousness is intrinsically computation, then the answer is that nothing is intrinsically computational; it is observer-relative. This now seems to me an obvious point. I should have seen it ten years ago, but I did not. It is devastating to any computational theory of mind' (1995, p. 209).

Elke intrinsieke eigenschap of intrinsiek fenomeen dat je wilt dupliceren, kan niet worden gedupliceerd door het op het abstracte (niet-essentiële) niveau van computatie beschrijven of definiëren. Dit niveau kan niet essentieel zijn, als het waarnemerrelatief is.[16] In een recentere bron waarin Searle dit argument aanhaalt, schrijft hij (zichzelf corrigerend) het volgende: 'Except for cases where a person is actually computing in his own mind there are no intrinsic or original computations in nature' (2004a, p. 64). Hij gebruikt een beschrijving van rekenmachines om aan te geven dat zelfs die niet aan 'intrinsieke computatie' doen, en trekt deze door naar commerciële 'computers':

[16] Searle verwijst in zijn verschillende werken naar hoofdstuk 9 van '*The Rediscovery of the Mind*'' (1992), voor zijn 'echte uitleg' (eerste uitleg) van dit argument.

‘The electrical state transitions are intrinsic to the machine, but the computation is in the eye of the beholder’ (2004a, p. 64) .

Terugkomend op de vergelijking tussen brein en computer, is de conclusie die Searle uit deze nieuwe redenering trekt, dat de vraag of het brein gelijk te stellen is aan een digitale computer een onmogelijke en geen valide vraag is:

> For this reason you could not discover that the brain is a digital computer, because computation is not discovered in nature, it is assigned to it. So the question, Is the brain a digital computer? is ill defined. If it asks, Is the brain intrinsically a digital computer? the answer is that nothing is intrinsically a digital computer except for conscious agents thinking through computations. If it asks, Could we assign a computational interpretation to the brain? the answer is that we can assign a computational interpretation to anything’ (2004a, p. 64).

De computationele interpretatie die we kunnen toeschrijven aan het brein, is geen aanduiding van daadwerkelijke intrinsieke processen in het brein. Het ‘intrinsiek zijn’ van een proces is dus niet alleen voor intentionaliteit, maar voor alle natuurlijke fenomenen essentieel en constitutief. Intrinsieke computatie bestaat wel (in mensen), maar wordt niet gedupliceerd door digitale computers. Computatie onder interpretatie is slechts een simulatie van de intrinsieke processen in mensen. Zoals eerder aangegeven is het nu nog steeds mogelijk voor een ander systeem dan een brein om intrinsieke processen zoals die in het menselijk brein te hebben:

> ‘This point has to be understood precisely. I am not saying there are a priori limits on the patterns we could discover in nature’ (Searle, 1992, p. 211).

Searle wil de nadruk leggen op de intrinsieke eigenschappen van verschijnselen in de natuur. Computatie is waarnemerrelatief (in bijna alle gevallen) en daarmee niet het niveau waarop de gewenste intrinsieke fenomenen van mentaliteit benaderd kan worden.

4.2.2. Conclusie: Searles nieuwe redenering

De nieuwere redering van Searle over computatie en syntax in relatie tot fysica, is een fundering en versterking van zijn eerdere redenering over syntax en semantiek en digitale computers. In digitale systemen bestaat alleen computatie onder interpretatie. Iets wat onder interpretatie bestaat, bestaat niet (noodzakelijk) gegarandeerd voor iets wat intrinsiek aanwezig is. En volgens Searle kunnen we niet ‘tevreden zijn’ met een systeem dat een mind moet kunnen dupliceren als de semantiek niet zeker weten intrinsiek aanwezig is. En omdat computatie zelf, volgens Searle, waarnemerrelatief is, is het veroorzaken van semantiek middels die computatie al helemaal twijfelachtig.

Als de semantiek (en alles wat eraan gerelateerd is) er niet is, is dat geen catastrofe, maar dan kunnen we niet spreken over een ‘persoonstatus’.

Deze bijzonder sterke claims over de aard van computatie en de relevantie van het intrinsiek zijn van processen of fenomenen is een bijzonder kenmerkende positie voor Searle, en is een betere basis voor een fundamenteel debat dan de bespreking van het Chinese Kamergedachte-experiment. Dit fundamentele debat wordt hopelijk verder en uitgebreider gevoerd in de nabije toekomst. Het is een lastig debat, omdat het (wederom) lijkt neer te komen op een ‘overtuiging’ die religieus van aard is (zoals Searle eerder over KI sprak). Deze redenering van Searle zorgt mijns inziens voor nieuwe stof voor het debat, waardoor het debat een het ‘welles’-‘nietes’-niveau kan ontstijgen. De redenering omvat de verwijzing naar ‘intrinsieke eigenschappen’ in de natuur, en is daardoor heel anders, fundamenteler en veel breder van aard dan de SSR is.

4.3. Conclusie

Searle gebruikt de kracht van de terminologie van syntax en semantiek en de consensus die over deze begrippen in de logica en taalkunde bestaat om uitspraken te doen over het domein van de Harde KI. Deze kracht wordt ontleend aan het (logisch correcte) onderscheid tussen syntax en semantiek. Syntax alleen is niet voldoende om semantiek te beschrijven. Door ‘syntax’ in computers te gebruiken om de symboolmanipulerende processen in computers aan te duiden, en ‘semantiek’ om (het missende begrip van) de inhoud van die processen aan te duiden, ziet Searle een duidelijk probleem voor de Harde KI: met slechts symboolmanipulatie kan geen interne semantiek worden bereikt. Deze kan alleen worden toegeschreven onder interpretatie: er kan vanwege het abstractieniveau van de syntax, waarop de programma’s essentieel gedefinieerd zijn, geen sprake zijn van intrinsiek aanwezige eigenschappen als intentionaliteit, of andere mentale fenomenen. Door slechts te simuleren (of computers als mentale systemen te interpreteren) wordt datgene dat nodig is voor een mind niet *gedupliceerd*.

Searle heeft deze redenering verdiept door aan te geven hoezeer hij belang hecht aan intrinsieke eigenschappen in de natuur: het probleem van computatie is, dat computatie bijna altijd slechts onder interpretatie (en niet als intrinsieke eigenschap in de natuur) als ‘computatie’ kan worden gezien. Zeer zelden is er sprake van ‘intrinsieke computatie’ (wanneer wij mensen bijvoorbeeld echt aan het berekenen

zijn). Computatie zelf is dus waarnemerrelatief: dit beschrijft het probleem van het abstractieniveau van computatie. Definiëren van programma's op een niveau dat al waarnemerrelatief is, kan geen intrinsieke eigenschappen 'aansturen'. Uiteraard heeft een geïmplementeerd programma als zodanig intrinsieke fysische eigenschappen, maar dat zijn niet degene die wij als definiërend voor het programma gebruiken.

Deze opvattingen over semantiek (intrinsieke mentale eigenschappen), syntax (computatie) en intrinsieke verschijnselen in de natuur (fysica) beschrijven dus meerdere barrières of problemen voor de Harde KI en geven aan waarom deze door de fundamenten van de Harde KI (computationalisme) niet overbrugd kunnen worden.

5. Kritiek op Searle: syntax en semantiek in de KI

In dit hoofdstuk wil ik twee verschillende typen van kritiek op de stelling van Searle over syntax en semantiek in de KI bespreken. Hiervoor heb ik gekozen voor William Rapaport, die tegen de stelling dat syntax niet voldoende is voor semantiek ingaat. Daarnaast bespreek ik de opvatting van John Haugeland over de mogelijkheid van semantiek in computers. Deze twee kritieken heb ik gekozen omdat ze beide kiezen voor een specifieke aanval op (elk een van de) twee van de drie aannames van Searle.

5.1. Rapaport: Syntax is wel voldoende voor semantiek

William J. Rapaport is duidelijk een computationalist binnen de KI: hij gebruikt een symbolisch systeem (SNePS), om een computationele mind te maken (1995, p. 52). Dit is dus het 'echte' type tegenstander van Searle. Wat zijn Rapaports redenen om tegen Searle in te gaan, waar gaat hij precies tegenin en welke conclusies kunnen wij trekken over deze dialoog?

5.1.1. Basisbegrippen bij Rapaport

In deze paragraaf bespreek ik de opvattingen die Rapaport heeft over de belangrijke aspecten van syntax en semantiek in Harde KI. Een van de belangrijke noties die Rapaport uitlegt is het verschil tussen het **systeem, het programma en het proces** van het werkende programma:

> 'Rather, the question is whether a computer that is *running* (or executing) a suitable program – a (suitable) program being executed or run – can understand natural language. A program being actually being executed is sometimes said to be a "process" [..]. Thus, one must distinguish three things: (a) the computer (i.e., the hardware; in particular, the central processing unit), (b) the program (i.e., the software), and (c) the process (i.e., the hardware running the software)' (1988, p. 81).

Het *proces* datgene is waarvan we mogelijk kunnen zeggen dat het *begrijpt*. We hebben gezien dat Searle dit onderscheid niet negeert, maar ook niet bijzonder belangrijk acht. Searle heeft het ook 'gewoon' over 'running' programma's. In hoofdstuk zes bespreek ik deze kwestie als algemene kritiek op Searle.

Zoals in het bovenstaande citaat al zichtbaar is, spreekt Rapaport vooral over (de tweede belangrijke notie van) het begrijpen van **natuurlijke taal**. De relevantie van het kunnen begrijpen van natuurlijke taal is een afgeleide relevantie (in de bespreking van Searle): taal is een vorm van cognitie, een vorm van vertonen van

intentionaliteit (zie hoofdstuk drie). De verantwoording van Rapaport om nadruk te leggen op natuurlijke taal houdt in dat het begrijpen van natuurlijke taal op zijn minst een ‘mark’ voor intelligentie is, en volgens Rapaport zelfs ook een voldoende voorwaarde voor intelligentie:

> ‘So, understanding natural language [is a necessary condition for passing the Turing Test, and to that extent], at least, it is a mark of intelligence. [..] I think, by the way, that it is also a sufficient condition’ (1988, p. 83).

Dus wanneer hij het over het begrijpen van natuurlijke taal heeft, is dit voor hem een ‘marker’, en zelfs een voldoende voorwaarde van meer algemene menselijke capaciteiten (‘to imitate a human’ (1988, p. 83), intelligentie, wellicht ook ‘mindedness’).

Rapaport spreekt, zoals Searle, over de mogelijkheid van ***minds*** in computationele systemen. Hij beschrijft minds echter op een eigenaardige ‘KI’-manier, zijn notie van mind is hierop gebaseerd:

> ‘To do all of this, a cognitive agent who understands natural language must have a “mind” – what AI researchers call a ‘*knowledge base’*. [..] For convenience and perspicuousness, let us think of the knowledge base or mind as a propositional semantic network, whose nodes represent individual concepts, properties, relations, and propositions, and whose connecting arcs structure atomic concepts into molecular ones (including structured individuals, propositions, and rules). The specific semantic-network theory we use is the SNePS knowledge representation and reasoning system (see par. 1.2), but you can think in terms of other such systems, [..]’ (1995, p. 51) (mijn nadruk).

Hier geeft hij aan dat een mind in zijn ogen gelijk is aan een ‘knowledge base’, die in meerdere en verschillende systemen gerealiseerd kan worden. ‘The knowledge base, expressed in a knowledge-representation language augmented by an inferencing package, is (at least a part of) the “mind” of the system’ (1988, p. 85). Een dergelijke ‘knowledge base’ is dus op zijn minst een deel van, zo niet simpelweg alles wat nodig is voor een mind. Strikt gesproken, Searle verdedigend, kunnen we hier al opmerken dat deze opvatting van een mind wel erg ‘mager’ lijkt. Is dit concept van mind wel te vergelijken met het concept dat Searle handhaaft? De aparte status ervan, die tot stand komt door de intrinsieke status van mentaliteit, lijkt in het concept van Rapaport niet aan de orde te zijn. In de verdere bespreking van Rapaport kunnen we zien wat zijn aannames en uitwerkingen zijn en kijken of deze ‘hypothese’ juist is.

5.1.2. Rapaport over Searle

Rapaport richt zich expliciet tegen de opvatting van Searle dat syntax niet voldoende is voor semantiek en daarmee tegen zijn stelling dat computers geen semantiek kunnen 'bezitten'. De opvatting van Searle over betekenis beschrijft hij als volgt, met zijn eigen commentaar erop volgend:

> 'Meaning is a relation between symbols and things in the world that the symbols are supposed to represent or be about. This "aboutness", or intentionality, is supposed to be a feature that only minds possess. [...] But there is another way to provide the link between symbols and things in the world *[to provide meaning for the symbols]*: Even if the system has sensor and effector organs, it must still have internal representations of the external objects, and – I shall argue – it is the relations between *these* and other symbols that constitute meaning for it. Searle seems to think that semantics must link the internal symbols with the outside world and that this is something that cannot be programmed. **But if this is what semantics must do, it must do it for human beings, too,** so we might as well wonder how the link could possibly be forged for us. Either the link between internal representations and the outside world *can* be made for both humans *and* computers, or else semantics is more usefully treated as linking one set of internal symbolic representations with another. On this view, semantics does indeed turn out to be just more symbol manipulation' (1988, pp. 87-88) (mijn bold nadruk).

Rapaport ziet dus geen essentiële verschillen in de mogelijkheden voor mensen en de mogelijkheden voor computers om betekenis aan interne representaties te geven: of het is een probleem voor beide 'systemen', of het is geen probleem omdat het gewoon een kwestie van symboolmanipulatie is. Dat laatste is wat hij ons duidelijk wil maken.

De 'causale krachten' die Searle op een voetstuk plaatst, ziet Rapaport niet op dezelfde manier:

> 'What, then, are these "causal powers"? All Searle tells us [..] is that they are due to the (human) brain's "biological (that is, chemical and physical) structure". But he does not specify precisely what these causal powers are. (In Rapaport 1985b and 1986b, I argue that they are not even causal)' (1988, p. 88).

Rapaport is het duidelijk niet eens met het bestaan van *dergelijke* causale krachten.

> '[..] *any* device that "implements" (in the technical sense of the computational theory of abstract data types) an algorithm for successfully processing natural language can be said to *understand* language, no matter how the device is physically constituted [..]. My intent here is to argue, [..], that a purely syntactic entity *is* sufficient for understanding natural language' (1988, p. 88).

Zijn stelling is dat de juiste manier van puur syntactische symboolmanipulatie van de 'knowledge base' van een system voldoende is om begrip van natuurlijke taal mogelijk te maken (1988, p. 85). Hij wil de lezer duidelijk maken dat de causale krachten zoals Searle ze beschrijft niet nodig zijn voor het verkrijgen van semantiek

en daarmee een mind; een syntactisch (symboolmanipulerend) systeem is voldoende voor het maken van een mind.

5.1.3. 'Syntax suffices'

Om te kunnen begrijpen waarom syntax voldoende is voor semantiek, legt Rapaport ook uit welk type semantiek hij bedoelt:

> 'Briefly, my thesis in this essay is that *syntax suffices*. I shall qualify this somewhat by allowing that there will also be a certain causal link between the computer and the external world, which contributes to a *certain kind* of nonsyntactic semantics, but not the kind of semantics that is of computational interest' (1988, p. 84).

Het type semantiek waarmee computatie gepaard gaat en hoeft te gaan, is dus niet 'nonsyntactic', maar syntactisch. Syntax is voldoende voor het benodigde type semantiek. Hij maakt een verschil tussen typen semantiek of 'semantic points of view', (C = cognitive agent, O = natural-language output of another agent):

> 'Two semantic points of view must be distinguished. The *external* point of view is *C*'s understanding of *O*. The *internal* point of view is *C*'s understanding of itself. **There are two ways of viewing the external point of** view: the "third-person" way, in which *we*, as external observers, describe *C*'s understanding of *O*, and the "first-person" way, in which *C* understands its own understanding of *O*. Traditional referential semantics is largely irrelevant to the latter, primarily because external objects *can* only be dealt with via internal representations of them. **It is first-person and internal understanding that I seek to understand and that, I believe, can only be understood syntactically**' (1995, pp. 51-52) (mijn bold nadruk).

Eerste persoons-, interne semantiek is dus het type semantiek dat puur syntactisch begrepen kan worden. De relevante 'toegang' (interne semantiek) voor een machine naar de betekenis van symbolen, waarvan de mogelijkheid betwijfeld wordt door Searle (en Fred Dretske, Rapaport bekritiseert ook hem (1988, p. 94)), is volgens Rapaport, als het een vraagstuk is, voor mensen hetzelfde vraagstuk:

> 'All of this is what I shall call *internal* semantics: semantics as an interconnected network of internal symbols – a "semantic network" of symbols in the "mind" or "knowledge base" of an intelligent system, artificial or **otherwise**' (1988, p. 94) (mijn bold nadruk).

Samenvattend schrijft Rapaport over zijn opvatting over hoe begrip van taal, en de relevante interne semantiek tot stand komt:

> 'To understand language is to construct a semantic interpretation – a model – of the language. In fact, we *normally* understand something by modeling it and then determining correspondences between the two domains. In some cases, we are lucky: We can, as it were, keep an eye on each domain, merging the images in our mind's eye. In other cases, notably when one of the domains is the external world, we are not so lucky – [..] – and so we can understand that domain *only* in terms of the model. Lucky or not, we understand one thing in terms of another by modeling that which is to be

> understood (the syntactic domain) in that which we antecedently understand (the semantic domain). But how is the antecedently understood domain antecedently understood? In the base case of our recursive understanding of understanding, a domain must be understood in terms of *itself*, i.e., syntactically' (1995, p. 74).

Het basisgeval van ons (recursieve) begrijpen (de basis voor onze interne semantiek) komt dus neer op een domein dat in termen van zichzelf, d.w.z. syntactisch, begrepen moet en kan worden. Deze elementen van dit domein, ook wel aan te duiden met 'primitieven', krijgen hun betekenis door hun gebruik en hun rollen in het netwerk met andere primitieven:

> 'Another thing that using parts of the syntactic domain to understand the rest of it might mean is that those parts are primitives. How are *they* understood? What do *they* mean? **They might be 'markers' with no intrinsic meaning**. But such markers *get* the meaning the more they are used – the more roles they play in providing meaning to *other* nodes' (1995, p. 77)(mijn bold nadruk).

Hier stuiten we op een interessante opmerking: deze primitieven *zouden* 'markers' zonder intrinsieke betekenis *kunnen zijn*. De betekenis van de primitieven wordt functioneel of gerelateerd aan andere primitieven opgebouwd. Oftewel, of ze nu wel of geen intrinsieke betekenis hebben, doet er niet toe. Uit dit citaat en de eerdere citaten wordt duidelijk dat het 'intrinsieke' aspect er niet toe doet voor de 'internal semantics' volgens Rapaport. Doet het er wel toe, dan vormt het voor computers *en* mensen hetzelfde probleem. *Mensen hebben dus geen 'voorsprong' op computers wat intrinsieke eigenschappen betreft.*

Het algemene probleem waar Rapaport nu op stuit is dat van **symbol grounding**, dat hij met een uitleg van Harnad benoemt:

> 'Alternatively, the fixed points or the markers (or, - for that matter – any of the nodes) *[de primitieven]* are somehow "grounded" in another domain. This, of course, is just to say that they have meaning in the correspondence sense of semantics, [..]. The symbol grounding problem, according to Harnad (1990), is that without grounding, a hermetically sealed circle of nodes can only have circular meaning. And, presumably, circles are vicious and to be avoided' (1995, pp. 77-78).

Het symbol grounding probleem beschrijft het probleem van het circulair 'moeten' vastleggen van betekenissen, omdat symbolen (volgens het probleem) zonder 'gegrond te zijn' in de werkelijkheid geen betekenis kunnen hebben. Rapaport concludeert dat de 'oplossing' die symbol grounding volgens Harnad biedt ook niet uit de vicieuze cirkel kan komen, en merkt hier bij op dat dit niet bezwaarlijk is:

> ‘Symbol grounding, thus, does *not* necessarily get us out of the circle of words – at best, it widens the circle. That is my point: Syntactic understanding – the base case of understanding – is just a *very* wide circle’ (1995, p. 79).

Maar zoals uit dit citaat en eerdere citaten blijkt, is dit het symbol grounding probleem voor Rapaport dus niet noodzakelijk een probleem. Zijn punt is: we hoeven ook niet uit die vicieuze cirkel te komen, want de relevante en enige noodzakelijke basis ligt *binnen* het syntactische netwerk, binnen de cirkel, en heeft geen ‘uitbraak’ nodig. Er komt niets ‘extra’s’ kijken bij het verkrijgen van betekenis, buiten syntax.

5.1.4. Conclusies van Rapaport:

De basisaanname over de benodigde vorm van interne semantiek van Rapaport, is dat deze semantiek in het basisgeval neerkomt op syntax, en niets daarbuiten. Voor digitale computers maar ook voor mensen is dit het geval: de primitieven die voor semantiek zorgen hebben geen intrinsieke betekenis *nodig*. Nu we weten wat Rapaports basisaannames over de interne semantiek zijn, kunnen we zijn kritiek op Searle begrijpen. Hij legt zelf zijn aanpak van kritiek als volgt uit:

> ‘This is one of the flaws in Searle’s Chinese-Room Argument. Part of his argument is that computers can never understand natural language because (1) understanding natural language requires (knowledge of) semantics, (2) computers can only do syntax, and (3) syntax is insufficient for semantics. I take *my* argument to have shown that (3) is false, and that, therefore, (2) is misleading, since the kind of syntax that computers do *ipso facto* allows them to do semantics)’ (1995, pp. 80-81).

Rapaport heeft kritiek op stelling (3), om daarmee aan te tonen dat stelling (2) misleidend (en dus fout) is. Zijn meningsverschil met Searle laat hij neerkomen op een pessimistische houding van Searle die hij zelf niet inneemt:

> ‘**Searle holds**, however, that the links [..] are necessary for understanding, that humans have (or that only biological entities can have) such access, that computers lack it, and, hence, that computers cannot understand. By contrast, **I hold**, that *if* such access *were* needed, then computers could have it, too, so that Searle’s pessimism with respect to computer understanding is unsupported’ (1995, p. 120) (mijn bold nadruk).

Hier wordt nogmaals duidelijk dat Rapaport geen verschil ziet tussen de mogelijkheden tot ‘understanding’ van mensen dan wel computers; mensen zijn niet bijzonderder (in dit opzicht) dan computers.

> ‘I **also** hold that such access is *not* needed, that, therefore, humans don’t need it either (here is where methodological solipsism appears), so that, again, there’s no support for Searle’s conclusion. I agree with Searle that semantics is necessary for understanding natural language, but that the *kind* of semantics that’s needed is the semantics provided by an internal semantic interpretation, which is, in fact, syntactic in nature and hence, computable. Syntax suffices’ (1995, p. 120) (mijn bold nadruk).

De reden waarom hij de ‘bijzonderheid’ van mensen op deze manier niet onderschrijft, is omdat hij ervan overtuigd is dat er geen sprake is van een bijzondere ‘toegang’ tot de benodigde semantiek (interne semantiek); deze semantiek heeft namelijk een syntactische basis. Die basis is ‘computable’ (berekenbaar), en dus mogelijk voor alle systemen die computationeel zijn.

Uit deze conclusies, uit Rapaports beschrijving van ‘minds’ en waarom semantics uiteindelijk syntactisch kan zijn, komt dus naar voren dat hij ‘onze’ aanname 1 (van Searle), ‘minds hebben semantiek’ op een andere manier interpreteert: er is niets ‘intrinsieks’ nodig voor het begrijpen van taal (en hiermee het vertonen of hebben intelligentie). De vereiste causale krachten zijn niet bijster bijzonder. Of Rapaport zijn beschrijving van ‘mind’ als voorwaarde voor ‘persoonstatus’ ziet, wordt niet expliciet duidelijk; waarschijnlijk ziet hij het als een noodzakelijke voorwaarde, en is hij geneigd te zeggen dat het ook een voldoende voorwaarde is (1988, p. 85). Maar hij ziet geen probleem in het niet intrinsiek zijn van semantiek: zoals geciteerd, als er een probleem is voor computers, is dat probleem er ook voor mensen. Mensen hebben hierin dus geen ‘aparte’ status. Het contrast met Searle is hier bijzonder groot: juist datgene wat Searle als uitgangpositie beschrijft, de bijzondere causale krachten van het menselijk brein, ziet Rapaport als onbelangrijk. Met Searle kunnen we zeggen, dat op deze manier juist datgene wat essentieel is voor minds en personen, buiten beschouwing wordt gelaten; voor een serieuze aanpak van het maken van echte minds met intrinsieke intentionaliteit in Harde KI is dit daarom een verkeerde benadering.

5.1.5. Conclusie: Rapaport en Searle

Rapaport gaat in tegen de aanname dat syntax in digitale systemen niet voldoende is voor semantiek, waarbij hij met ‘semantiek’ vooral het begrijpen van natuurlijke taal bedoelt; deze capaciteit is echter een voldoende voorwaarde voor intelligentie volgens Rapaport. Searle stelt semantiek als equivalent aan intrinsiek menselijke fenomenen (intentionaliteit, maar ook ‘mentale fenomenen’ in het algemeen). Rapaport ziet semantiek als een minder ‘beladen’ concept: het relevante type semantiek is ‘gewoon’ een kwestie van syntax (in het uiteindelijke basisgeval). Door deze aanpak kan Rapaport de aanname ‘syntax is niet voldoende voor semantiek’ aanvallen en daarmee de aanname ‘computers hebben slechts syntax’ als irrelevant bestempelen: mensen hebben namelijk ook ‘niets anders dan syntax’, want dat is alles wat nodig is.

Intrinsieke eigenschappen van mensen bestaan niet, of zijn in ieder geval zeker niet alleen voor mensen weggelegd. Rapaport ziet mensen niet als 'bijzonder': natuurlijk valt zo het probleem voor KI dat Searle ziet grotendeels weg. Als we minds echter wel zien zoals Searle ze beschrijft, met alle mentale fenomenen die minds zo bijzonder maken, is Rapaports aanpak van het maken van minds voor Harde KI totaal 'besides the point'.

5.2. Haugeland: semantiek in computers is wel mogelijk

Haugeland geeft in zijn *'Syntax, semantics, physics'* (2002) een commentaar op Searle, op de aanname van Searle over mogelijkheid van semantiek in digitale computers. Hij is het in zoverre met Searle eens dat hij de aannames 2 (minds hebben semantiek, zie hiervoor ook zijn boek *'Having Thought'*) en 3 (syntax is niet voldoende voor semantiek) onderschrijft. Hij is van mening dat aanname 1, dat computers 'slechts' syntax hebben, niet hoeft te gelden (2002, p. 386). Wat is zijn redenering voor deze opvatting? In deze paragraaf zal ik proberen duidelijk te maken hoe Haugeland semantiek in computers als mogelijkheid ziet; hij begint hiermee door een beschrijving van een mogelijk model voor systemen in digitale computers die in zijn ogen de juiste uitgangspositie voor Harde KI kan zijn (2002, p. 382).

5.2.1. Paradox van mechanisch redeneren: een uitweg

In zijn boek *'Artificial Intelligence: The very idea'* (1985) bespreekt Haugeland semantiek in computers ook, maar op een andere manier dan in het artikel uit 2002: het artikel uit 2002 is gedetailleerder uitgewerkt. In *'The Very Idea'* beschreef hij echter al wel de 'paradox of mechanical reason':

> 'if the manipulators pay attention to what the symbols mean, then they can't be entirely mechanical because meanings exert no mechanical forces; but if they ignore the meanings, then the manipulations can't be instances of reasoning because what is reasonable depends on what the symbols mean' (1985, p. 117) (*manipulators* zijn symboolmanipulerende systemen).[17]

Hij geeft aan dat om deze paradox te laten 'verdwijnen', een fundamenteel andere strategie nodig is:

> 'A different fundamental strategy is required here: not analysis but redescription, seeing the very same thing from radically different points of view and thus describing it in radically different terms' (1985, p. 118).

[17] Denk hierbij ook terug aan Rychlaks opvattingen over redeneren uit hoofdstuk drie en vier.

Deze strategie is precies de strategie die hij later, in zijn artikel uit 2002, toepast, om een vorm van semantiek in computers te kunnen beschrijven. In 1985 beschrijft Haugeland dat deze kwestie uitkomt op 'the mystery of original meaning' (1985, p. 119), dat volgens hem ouder en diepgaander is. 'The questions remain: Which symbolic systems have their meanings originally (nonderivatively) and why?' (1985, p. 119). Of computers ook originele betekenis hebben, is dus de vraag (ervan uit gaande dat mensen dit ipso facto hebben). In de onderstaande beschrijvingen uit het artikel uit 2002 probeert Haugeland een basis voor *originele* betekenis zoals hij eerder als gewenst beschreef (1985, p. 121) in computers te geven: ik zal met citaten en parafrases deze beschrijving uiteenzetten en proberen uit te leggen.

5.2.1.1. Eerste en tweede manier van beschrijven

De eerste manier van beschrijven die Haugeland beschrijft, is in overeenstemming met de beschrijving die Searle gebruikt:

> 'Searle is right that one way of describing any digital computer program or data is as a formal structure of formal tokens – that is, purely, syntactically. It is an essential feature of them that they can be so described; and this has many important consequences, including that they can in principle be implemented or realized in an open-ended variety of different physical media. There are also, **however**, **substantial *further* constraints** on any adequate implementation – that is, on any concrete instance of them in which they can actually be the program data that they are supposed to be' (2002, p. 382) (mijn bold nadruk).

Het (inerte) programma heeft inderdaad als essentiële beschrijving een formele (syntactische) beschrijving. Maar de daadwerkelijke implementatie is onderhevig aan meer voorwaarden – voorwaarden die de ervoor zorgen dat de data 'echte' data worden:

> 'Above all, any implementation *must* be such that the operations on the data, including input and output operations, that are explicitly prescribed in the program, are reliably carried out as prescribed. Since in general these operations will involve modifications to the (concretely implemented) data structures, carrying out the operations must be a *causal* process' (2002, p. 383).

De formele specificatie is belangrijk: deze zorgt ervoor dat het volgens strikte regels verloopt en de datastructuren in een *causaal* proces terechtkomen.

> 'So, *another* way to describe computer programs and data structures is that there must be possible **concrete implementations of them such that they interact causally in the right way** – in other words, such that they have 'the right causal powers'. That there can be such implementations is just as *essential* a feature of them, *qua* programs and data, as that they can be described formally. Indeed, in the light of this second essential feature, it becomes clear that the purely syntactical descriptions (which must

> also be possible) must, strictly speaking, be regarded as *abstractions* from the various possible concrete causal implementations' (2002, p. 383) (mijn bold nadruk).

Er is dus een andere, tweede, essentiële manier om computerprogramma's en datastructuren te beschrijven, die de causale interactie en daarmee de juiste causale krachten inbedt in de beschrijving. Deze manier van beschrijven houdt rekening met de daadwerkelijke implementaties van de programma's en datastructuren. Dit kenmerk, dat programma's ook zo te beschrijven zijn, is net zo essentieel als de abstracte beschrijving (die puur syntactisch is).

Searles kritiek hierop is: zo lang het abstracte niveau het essentiële niveau is, heb je niet de (intrinsieke) essentie van de causale krachten te pakken die je moet dupliceren. Dan kun je dus alleen maar 'hopen' dat in de implementatie deze krachten opeens emergeren uit de abstractie. Haugeland zegt echter dat het abstracte niveau niet het *enige* essentiële niveau is, omdat we anders computers niet als computers kunnen beschrijven:

> 'Without *actual* causal implementation there is no *actual* program or data, but only an abstract specification of the kind of causal implementation that it would take to actualize them. This is exactly like the relation between engineering drawings or diagrams for a pump or an electronic circuit and the various possible actualizations of such pumps or circuits. The only difference (so far) is that, given the kind of causal system being specified, the abstract specification itself (at one level of abstraction) is a formal syntactical structure rather than a drawing or a diagram' (2002, p. 383).

Zonder daadwerkelijke implementatie is een programma niet echt op te vatten als een programma; dit doet ook denken aan het onderscheid dat Rapaport maakte. Zoals Rapaport, wil Haugeland aanduiden dat het daadwerkelijke proces van het 'running program' het enige interessante fenomeen is. Natuurlijk is alleen het ongeïmplementeerde programma niet voldoende, maar, zoals we in een eerder citaat hebben gezien, had Searle het ook niet slechts over 'inerte', niet-werkende programma's. De essentie van de verschillen van meningen tussen Searle en tegenstanders lijkt hier dus vaker op uit te draaien. De volgende stap van Haugeland beschrijft hoe hij deze mogelijkheden tot het beschrijven van geïmplementeerde programma's gebruikt om verder te gaan in zijn aanpak van 'fundamenteel anders herbeschrijven', namelijk richting een semantische manier van beschrijven van digitale systemen.

5.2.1.2. Derde manier van beschrijven: semantisch

Nu komt er voor Haugeland een derde, essentiële manier, waarop programma's beschrijfbaar moeten zijn in het spel: 'semantically' (2002, p. 383)! In '*The Very Idea'* benadrukte Haugeland een soortgelijke kwestie door te zeggen dat symbolen in gedachteprocessen en computationele processen 'semantisch actief' (in tegenstelling tot semantisch inert) zijn, en daardoor kandidaat voor het hebben van 'original meaning' (1985, p. 121). Niet alle soorten processen zijn echter zo beschrijfbaar: 'only semantic activity of a certain sophisticated or advanced kinds, or with certain distinctive characteristics, can suffice for genuine original meaning' (1985, p. 122). De 'substantial further contraints' van een implementatie van een programma doen er dus toe: deze kunnen de 'distinctive characteristics' voor computationele systemen vormen. Deze beschrijving geldt voor abstracte specificaties *en* voor concrete implementaties, en komt dus als het ware bovenop deze twee eerdere manieren van beschrijven.

> 'Programs are written in a *general* code with a *compositional* structure – a structure that can be, and often is, recursive. The components out of which basic program elements are built include general terms for the operations to be performed, singular terms and variables for their operands, second-order devices for forming indirect or complex singular or general terms from given ones, conditional devices for letting what is prescribed next be a function of the current state, and so on. So, though what can be specified in a program is limited to operations of formal tokens, the possibilities for what these can be are in principle unbounded' (2002, pp. 383-384).

Dat de programmaspecificaties beperkt zijn tot het beschrijven van symbolen, bepaalt niet dat deze symbolen geen hoge mate van diversiteit en complexiteit kunnen 'bevatten'.

> 'And whatever is specified is what the processor (or processors) should *do*. [..] If programs were not specifications of what to do, the notion of implementation would make no sense, and there would be no such things as computers or computer programs. But, of course, there are. This is why I used the word 'prescription'.
> **Specifications and prescriptions as such, however, have a *semantics*** – that is, meanings – complete with modes of presentation, **conditions of satisfaction**, and directions of fit. In particular, singular terms ***denote*** operands (which can also be denoted in other, non-equivalent ways), general terms *denote* operations to be performed, program elements have the **illocutionary force** of declarations or imperatives, and conditional devices permit genuine conditional declarations or imperatives. If programs are not understood in this way, they are not intelligible *as programs* at all' (2002, p. 384) (mijn bold nadruk[18]).

[18] Met het gebruik van de termen 'conditions of satisfaction' en 'illocutionary force' verwijst Haugeland letterlijk naar Searles eigen jargon van intentionaliteit.

Wij moeten de processor dus zien als een ‘iets’ dat de voorschriften van programma’s als betekenisvol opvat; als de processor dat niet zou doen, zouden we niet kunnen spreken van een ‘*programma op een computer*’. het geen programma op een computer zijn: wij verwachten dat de processor de opdrachten als opdrachten opvat. De opdrachten en specificaties hebben een semantiek die doorwerkt in de elementen van het programma.

> ‘ [..] to whom (or what) are *[the symbolic expressions]* meaningful?’ [..] ‘Now, one can certainly maintain that the processor doesn’t ‘understand’ in the same sense that the programmer does; and there is surely something right about that. For instance, the processor has no clue as to the point of the operations it is carrying out, or even any inkling of the meta-concepts of prescription, operation, operand, and the like. And, no doubt, there are other important differences as well.
> But **it would be a mistake to dismiss**, on such grounds, the first-order semantics of computer programs as merely ‘as-if’ semantics – an anthropomorphizing metaphor – [..]. The difference in the computer case is that we have explicit prescriptions, expressed in a *general* symbolic code, with denotation, conditionals, and all the rest, that the processor responds to *as prescriptions with those semantics*. [..]
> Here is another way to put the point that may be better. **The *only* way that we can *make sense of* a computer *as executing a program* is by understanding its processor as responding to the program prescriptions as meaningful.** This is a level of description without which computers *as such* would be unintelligible to us’ (2002, pp. 384-385) (mijn bold nadruk).

Alhoewel de processor de opdrachten niet ‘begrijpt’ zoals wij, is er wel een niveau van eerste orde semantiek in computers die we niet als slechts ‘alsof’-semantiek kunnen afdoen. Zonder deze vorm van semantiek in onze beschrijving van de programma’s en de beschrijving van de *uitvoering* van een programma mee te nemen, is die beschrijving ‘loos’, en is een computer ‘als zodanig’ voor ons niet te beschrijven of begrijpen.

> ‘**Certainly, syntax *by itself* is never sufficient for semantics. But neither is it sufficient for computer programs as computer programs.** What else is required is (at least) implementability in a system with the right causal powers. Moreover, these must be the right causal powers *for semantics*, because, if the system as implemented isn’t intelligible – and, at the relevant level, *only* intelligible – as *actually having* those semantics, then it isn’t intelligible as a programmed computer’ (2002, p. 386) (mijn bold nadruk).

Hoewel syntax niet voldoende is voor semantiek, kan er in digitale systemen wel degelijk sprake zijn van enige vorm van interne semantiek. Of deze vorm voldoende is of genoeg op die van ons lijkt om ‘semantiek’ zoals Searle deze aanduidt te benaderen (gerelateerd aan intentionaliteit), bespreekt Haugeland niet. Het is waarschijnlijk te betwijfelen of dit soort semantiek ‘het juiste soort’ kan benaderen of

dupliceren. Wat Haugeland tot nu toe zegt is dat het niet per definitie waar is dat er geen enkele vorm van semantiek in computers kan bestaan.

5.2.2. De relevante en problematische semantiek voor KI

De hierboven beschreven 'interne' semantiek is volgens Haugeland niet datgene waar het in de Harde KI om draait:

> 'The semantics of the computer program itself – with its singular and general terms denoting data tokens and operations – is not really what is at issue in serious AI. **Rather, what is at issue is the possibility of semantics about the 'outside world'**. These meanings, if any, would be attributed not to the *program* tokens (which already have their semantics) but rather to the *data* tokens – including symbolic inputs and outputs, if any – that the program prescriptions denote. Nevertheless, the discussion of the internal semantics of computers has been worthwhile for two reasons. First, it demonstrates that computer technology really can have the right causal powers for semantic, at least of a certain sort. And secondly, it can serve as a kind of model for how the states of computer systems in a larger sense might be able to have semantics of a further sort' (2002, p. 387) (mijn bold nadruk).

Het relevante type semantiek, is de semantiek 'naar de buitenwereld toe'. Om te beschrijven hoe deze tot stand kan komen, geeft Haugeland een model als suggestie (2002, p. 387). Hij beschrijft een complexe opbouw van een systeem dat als 'processor' ook weer een geheel systeem heeft, en waarin de data van een deelsysteem kunnen dienen als programma in het overkoepelende systeem. 'Accordingly, though these internal data have no semantics when considered only as data for the narrower system, it is precisely they that must be understood semantically – namely, as denoting objects and properties in the world' (2002, p. 387). Dit type van semantisch begrijpen is dus *gebaseerd op de interne semantiek* die hij eerder beschreef:

> 'In other words, the causality that this **internal semantics** depends on is really an elaborate pattern of causal interactions among the processor, the meaningful program symbols, *and* the data tokens that they are about. Accordingly, **if the internal case is to serve as a model for a larger system with semantics about objects in the outside world**, then those objects themselves will have to be included in a correspondingly larger pattern of causal interactions. **But** since these outer objects are not themselves tokens within the narrow system (the computer itself), the necessary causal interactions with them will have to be mediated by special facilities called 'transducers' (2002, pp. 387-388).

De basis van de 'interne semantiek' is dus volgens Haugeland niet het probleem, maar juist de 'perifere' stap naar de semantiek over de buitenwereld: de objecten in de buitenwereld zijn geen deel van het 'beperkte' systeem, de computer. Om de laatste stap naar de 'buitenwereld' mogelijk te maken, zijn speciale faciliteiten nodig,

‘transducers’, die zorgen voor een adequate en bruikbare representatie van de externe objecten. Het probleem in de huidige Harde KI is nu volgens Haugeland dat de systemen niet zo geconstrueerd worden, dit type ‘transducers’ wordt niet gebruikt. Er wordt gebruik gemaakt van handige ‘shortcuts’ (2002, p. 388), d.w.z. zelfbedachte manieren om deze representaties ‘symbolisch’ als input aan een systeem te geven. Hierdoor mist de computer de juiste causale interactie met de objecten: namelijk die interactie die noodzakelijk is om ‘echte semantiek’ over de objecten te kunnen verkrijgen.

5.2.3. Haugeland over (Searles conclusie voor) KI

‘Serieuze’ KI is het volgens Haugeland *eens* (2002, p. 388) met de claims die Searle maakt aangaande

1. de vereiste materialistische basis voor een mind en semantiek

(Searle: alleen materialistisch)

2. de relevantie de aard van die basis: of deze wel of niet kunstmatig kan zijn

(Searle: de basis kan zowel een natuurlijk als een kunstmatig systeem zijn)

3. de relevantie van causale krachten

(Searle: deze zijn relevant)

Het twistpunt is volgens Haugeland (2002, p. 388) *wat* de juiste causale krachten zijn, en wat een voldoende voorwaarde voor het hebben van deze krachten is. Dit is de hoofdvraag van Serieuze KI:

> ‘Indeed, Serious AI is nothing other than a theoretical proposal as to the genus of the requisite causal powers, plus a concrete research program for homing in on the species. Therefore, the observation that syntax *by itself* (without causal powers) is insufficient for semantics is, though true, entirely besides the point’ (2002, p. 388).

Haugeland wil dus aangeven dat de juiste causale krachten niet per definitie (zoals Searle claimt) *niet* te dupliceren zijn in digitale computers. Het model dat hij uiteen heeft gezet, is hier een ondersteunende suggestie voor. Haugeland merkt vervolgens ‘just for the record’ op dat hij zelf niet gelooft in de mogelijkheid van het maken van een model zoals hij heeft beschreven; hij is het eens met Searles conclusie dat computers geen mind kunnen hebben, maar om andere redenen (2002, p. 388) (zie voor zijn eigen beschrijvingen: ‘*Having Thought*’). Dit maakt de kracht van zijn voorstel iets minder groot, maar het is wel een model of een voorstel dat echt tegen de basisredenering en aannames van Searle in gaat. Does anyone actually deny this?’

wordt hier dus beantwoord, wat zelden gebeurt. Dit maakt zijn positie erg sterk en bruikbaar; het is echter wel een beschrijving van een mogelijk model (en niet van een reeds bestaand systeem). Voor nu is het dus een suggestie in een goede richting, maar kunnen we, ook namens Searle, grote vraagtekens zetten bij het soort semantiek dat we in computers kunnen aantreffen. Is deze basis voldoende om semantiek in minds (voor personen) te kunnen genereren? Searle zou deze vraag met een 'waarschijnlijk niet' beantwoorden: volgens Searle kunnen de relevante intrinsieke kenmerken van minds, met gebruik van slechts computationalistische (abstracte) systemen, niet gedupliceerd worden. Het zal de lezer niet verbazen dat Haugeland het niet eens is met de beschrijving van intrinsieke processen in de fysica zoals Searle deze centraal stelt (4.2), Haugeland vindt niet dat computatie waarnemerrelatief is (2002, pp. 391-392). Zijn ontkenning van dit argument van Searle luidt als volgt:

> 'So, is syntax an essentially 'observer-relative' notion – in a way that, for instance, being a pump is not? No, not at all; and here's why. Something is a concrete token of a certain syntactical type, or a pump of a certain type, if and only if it can (in principle) be *correctly* described as a token or a pump of that type. Whether any such description is correct is subject to empirical test, based on more or less stringent specifications of what it takes to be such a token or pump – in effect, definitions of them. And, whether any given description would actually pass that test is not relative to any observers' (Haugeland, 2002, p. 391).

Volgens Haugeland is het niet-waarnemerrelatief zijn van syntax een kwestie van de *juiste manier van beschrijven* van syntactische eenheden ('token') (zoals semantiek in computers ook onder de juiste beschrijving van computers wel mogelijk is). Een eenheid is een syntactische eenheid als de juiste beschrijving, die empirisch te testen is (de juiste definitie) op deze eenheid van toepassing is. Of die beschrijving 'slaagt' in empirische testen, is niet afhankelijk van waarnemers. Het verschil tussen Haugeland en Searle is hier, denk ik (de kwestie is hiermee zeker niet afgedaan), het punt wat ze willen maken. Haugeland zegt dat de *validiteit* van mogelijke beschrijvingen niet 'afhankelijk' van waarnemers hoeft te zijn; deze is 'objectief' en empirisch te testen. Searle kan hierover echter zeggen, dat het doen van die beschrijvingen alleen al (valide of niet), een waarnemer (beschrijver) vereist. Dit maakt de kenmerken die beschreven worden gevoelig voor beschrijvingen van waarnemers. Natuurlijk kan er ook een valide beschrijving (van een intrinsiek kenmerk) worden gevormd, maar dit geeft geen reden om alle mogelijke kenmerken die onder een bepaalde beschrijving 'geconstitueerd' worden, als intrinsiek te zien. De beschrijvingen zijn dus hooguit hypotheses die empirisch getest kunnen worden (zoals

Haugeland ook aangeeft). Computatie en syntax zijn in zoverre waarnemerrelatief, dat het lastig is om te testen in hoeverre er ‘intrinsiek’ sprake is van computatie dan wel syntax. Een computationele of syntactische beschrijving van een object zegt niets over de intrinsieke kenmerken van het object dat beschreven wordt, maar geeft slechts suggesties over deze kenmerken. Deze beschrijvingen zijn dus an sich geen correcte beschrijvingen, maar mogelijk correcte beschrijvingen. Searle zegt nu dat computatie in digitale systemen geen intrinsieke eigenschap is, omdat deze computatie slechts onder beschrijving (interpretatie) computatie is, en niet meer dan dat. Haugeland zegt, dat deze mogelijkheid wel bestaat; het feit dat er waarnemers nodig zijn om de beschrijving te doen, betekent niet dat de beschrijving per se onjuist is. Dit debat is komt nu dus neer op de open vraag: ‘Is er een empirisch te valideren beschrijving van digitale systemen mogelijk op grond waarvan we intrinsieke fenomenen als semantiek of mentaliteit aan deze systemen kunnen toeschrijven’? Haugeland ziet hier een mogelijk ‘ja’. Searle kan zeggen dat dit op zijn minst bijzonder lastig is, omdat ten eerste de interpretatiegevoeligheid van het ‘toeschrijven’, en ten tweede de bestaande vragen rond de oorsprong van deze intrinsieke fenomenen bij mensen, voor grote problemen zorgen.

5.2.4. Conclusie: Haugelands opvattingen over semantiek in computers

Haugeland maakt een onderscheid tussen soorten semantiek: originele, interne (in de computer mogelijke) semantiek en aan de buitenwereld gekoppelde (op basis van interne semantiek mogelijke) externe semantiek. Hij zegt dat een computer niet per definitie *geen* interne semantiek kan hebben, omdat een daadwerkelijk geïmplementeerd en ‘running’ programma meer omvat dan ‘slechts’ syntax: er zijn voorschriften (met semantiek) nodig om de processor van een computer de opdrachten te laten uitvoeren, en de processor moet deze opdrachten op een of andere manier als opdrachten opvatten. Als wij de processor niet als zodanig beschrijven, kunnen we de processor niet als processor en de computer niet als computer ‘als zodanig’ zien. Deze basis van interne semantiek is ook een mogelijke basis voor een uitgebreid en complex model (complexer dan nu in de KI gebruikt wordt, want daarin worden ‘shortcuts’ gebruikt) dat semantiek over de externe wereld kan toestaan.

Haugelands model is het enige ‘model’ dat ik kan vinden dat aansluit op de aannames van Searle. De vraag ‘Does anyone actually deny this?’ van Searle wordt hier dus op de juiste manier met een ‘ja’ beantwoord. “I could deny it”, zou

Haugeland kunnen zeggen. Vervolgens zegt hij wel dat hij Searles conclusie ondersteunt (maar om andere redenen): het is dus ook volgens Haugeland geen realistisch, uitgevoerd (uit te voeren) model. *Als* serieuze, Harde KI een poging wil wagen, zou het volgens het model van Haugeland geprobeerd moeten of kunnen worden. Of hiermee minds gedupliceerd kunnen worden, *blijft* echter de hamvraag. De problematiek van menselijke minds is voor beide (zij het om verschillende redenen) een knelpunt, Searle ziet daarbij de waarnemerrelativiteit van het interpreteren van digitale systemen als probleem, en Haugeland niet.

5.3. Algemene beschouwing en conclusie: Rapaport, Haugeland, Searle

Twee critici van Searle die zijn opvattingen zoals in hoofdstuk vier aan de orde zijn gekomen, bekritiseren, zijn William Rapaport en John Haugeland.

Rapaport stelt dat syntax wel voldoende kan zijn voor semantiek. Hij ziet ‘syntax is niet voldoende voor semantiek’ niet als logische waarheid voor digitale systemen (of voor mensen), omdat syntax wel voldoende is voor semantiek: de syntax is de essentie van het basisgeval van semantiek. De basis van semantiek is gegrond in syntax. Het symbol grounding ‘probleem’ is geen probleem, omdat er geen ‘buitensyntactische’ ‘gronding’ nodig is. Rapaport ziet semantiek dan ook niet perse als ‘intrinsieke intentionaliteit‘: als semantiek bij mensen voor iets intrinsieks zou staan, is die mogelijkheid er ook voor computers. Hun basis is namelijk ook de syntax. Een syntax-semantiek barrière zoals Searle die ziet, bestaat in Rapaports ogen niet.

Haugeland ziet door een adequate beschrijving van computers te geven, die meerdere niveaus van beschrijving omvat, een mogelijkheid tot een niveau van semantiek in een computer. Een deel van Haugelands kritiek is, dat deze andere manier van beschrijven juist datgene is wat Searle zou moeten doen (maar niet doet): Searle bekijkt de computer te simplistisch. Deze mogelijke vorm van interne semantiek *kan* de basis zijn voor meerdere vormen van semantiek; deze mogelijkheid is niet vooraf uit te sluiten. Een model waarin deze semantiek tot meerdere vormen van semantiek kan leiden, zou wel een complexere constructie nodig hebben dan systemen in de KI (tot nu toe) hebben en realistische ‘transducers’ tussen systeem en buitenwereld moeten hebben. Als deze ‘transducers’ voor geprogrammeerde shortcuts zijn, mist de computer een ‘ingrediënt’ om de interne semantiek te kunnen gebruiken voor semantiek over de buitenwereld. Deze opvatting is er een die het symbol

grounding probleem, zoals aangestipt bij Rapaport, wel serieus opvat en een link naar de buitenwereld als noodzakelijk ziet voor het benaderen van semantiek zoals we deze in mensen vinden. Het minpunt aan Haugelands beschrijving is dat deze niet meer dan hypothetisch is, en hij zelf niet denkt dat het een realiseerbaar model is.

De kritieken van Rapaport en Haugeland zijn beide kritieken die echt inhaken op aannames van de SSR van Searle. De ene kritiek komt neer op het niet erkennen van een syntax-semantiek barrière, zelfs bij mensen en minds. De andere kritiek komt neer op het wel degelijk zien van een kleine mogelijkheid tot semantiek in computers, die de basis kan zijn voor meer ingewikkelde vormen van semantiek. Of deze vormen voldoende zijn voor een ‘persoonstatus’ is hierbij de vraag, die Haugeland in dit geval ook met een ‘nee’ beantwoordt, maar op basis van geheel verschillende opvattingen over ‘persoon zijn’ dan Searle uit (zie voor deze opvattingen ook hoofdstuk 3).

6. Waarom kan de geldigheid van de SSR toch in het geding komen?

Dat de conclusie ‘computers kunnen geen mind hebben’ uit de premissen wordt getrokken zoals Searle ze opvat, is geen logische inconsistentie. Maar, zoals gezien, zijn de interpretaties van de opvattingen en de toepassingen ervan de oorzaken van de twijfelachtige status van de aannames en daarmee van de conclusie. Welke analyses kunnen we nu bij de redenering van Searle en de kritiek op zijn aannames maken?

In de vorige hoofdstukken hebben we enkele keren opvattingen van tegenstanders van Searle naar voren zien komen die kritiek uitten op hoe Searle zijn aannames gebruikt. In dit hoofdstuk wil ik, concluderend over de gehele redenering, drie kwesties bespreken die fundamenteel zijn voor de redenering van Searle, en waarin de kritiek op Searle gerechtvaardigd is. Mijns inziens is dit de logische volgorde van deze kwesties, en zijn het ook de belangrijkste kwesties in het debat. Het was mijn bedoeling om Searle te verdedigen, en dat doe ik nog steeds. Daarom is het nu relevant om te vragen *waarom* de redenering over syntax en semantiek zo’n groot probleem vormt: welke fundamentele kwesties zijn de ‘achillespees’ van Searle? Omdat het beschrijven en ondersteunen van Searles intuïtief plausibele redenering een van de doelstellingen van mijn onderzoek is, zal ik bij deze zwakke punten proberen aan te geven in hoeverre de bewijslast ten opzichte van deze punten bij Searle (en zijn redenering) of juist bij anderen ligt.

6.1. Opvatting van Harde KI vanuit filosofie + de praktijk van de KI

Er zijn wetenschappers die de stelling ‘syntax is niet voldoende voor semantiek *in computers*’ zoals Searle deze aanneemt ontkennen, terwijl het in principe een ‘logische waarheid’ uit de theoretische filosofie is. Hebben deze wetenschappers nu zulke aparte ideeën, omdat ze daar tegenin kunnen gaan? We zagen bij Rapaport dat hij ertegenin kon gaan, omdat hij eigenlijk niet alleen voor computers, maar ook voor mensen (ook voor mensen) syntax als basis voor semantiek ziet. De andere tegenstanders van Searle die we zijn tegengekomen hebben andere redenen om ertegenin te gaan. Waarom en hoe doen ze dat? Welke ‘versie’ van de stelling staat er dan ter discussie, waarop passen ze de stelling precies toe?

6.1.1. Haugeland over de fundamenten van KI

Haugeland geeft aan dat Searle een verkeerde **vooronderstelling** doet in zijn beschrijving van Harde KI als de opvatting dat wat er nodig is om een mind the hebben, het juiste programma is.

> 'This [*de beschrijving waarin Searle vooronderstellingen doet]* is misleading because it insinuates that what one might call serious AI is committed to denying the claim about syntax and semantics put forward in that passage. And it's irresponsible to the extent that Searle believes he is entitled to discount serious AI on the basis of his ability to refute the straw position that he here attributes to strong AI. In undertaking to raise fundamental questions about a major intellectual or scientific point of view, it is incumbent on one to confront that point of view in its most credible version – even if, as Descartes did for epistemological skepticism, one has first to articulate that version oneself' (2002, p. 382).

Haugeland zegt dus dat Searle Harde (serieuze) KI beschrijft alsof Harde KI resoluut tegen het syntax-semantiek onderscheid zou moeten zijn, maar dat Searles beschrijving van Harde KI hiertoe teveel op maat gemaakt is (dus precies zo om er zijn eigen kritiek over te kunnen ventileren). Een realistischere beschrijving van de fundamenten van Harde KI is mogelijk die niet een ontkenning van de stelling 'syntax is niet voldoende voor semantiek' inhoudt. De beschrijving van Searle is dus verkeerd, en in zijn eigen voordeel. Het is goed mogelijk voor de Harde KI om de propositie 'syntax is niet voldoende voor semantiek' in stand te houden, omdat deze niet relevant is voor het onderzoek van Harde KI, als je de Harde KI anders (in het voordeel van de Harde KI) beschrijft (2002, p. 382). De meer realistische beschrijving in het voordeel van Harde KI, is volgens Haugeland een mogelijke basis om computers wel degelijk semantisch te kunnen beschrijven (zie hoofdstuk vijf). Hij bedoelt dus niet dat de 'logische waarheid' niet waar is, maar wel dat deze niet als zodanig geldt voor computers. Preston zegt daarentegen, dat het computationalisme de duidelijke splitsing van syntax en semantiek niet *kan* betwijfelen (2002, p. 34). Preston geeft helaas geen expliciete onderbouwing voor deze uitspraak. Het lijkt erop dat Preston de beschrijving van Harde KI zoals Searle die geeft accepteert en daarom deze uitspraak kan doen.

6.1.2. Churchland en Churchland over de empirie van KI

Churchland en Churchland geven een redenering die lijkt op die van Haugeland: zij willen aangeven dat deze vraag, 'is syntax voldoende voor semantiek?' in de Harde KI een *empirische* vraag is, en dus anders van aard dan de vaststaande logische waarheid in de logica. Hun argumentatie is als volgt te parafraseren:

> ‘Searle’s third axiom cannot be decided in advance or scientific research. Searle’s third axiom assumes that syntax cannot produce semantics. But the question of whether syntactic machines can be used to produce semantic understanding is exactly what is at issue in classical AI. It is an empirical question that cannot be decided in advance of scientific research’[19] (Can Chinese Rooms Think? (Map 4)).

De Churchlands doelen hier op het volgende probleem: *het vraagstuk van de mogelijkheid om semantiek te verkrijgen door middel van syntax als logische waarheid in de logica en taalkunde is niet hetzelfde als het vraagstuk van de mogelijkheid om semantiek te verkrijgen door middel van syntax als empirische vraag voor de KI.* De Churchlands ontkennen de stelling niet in de vorm die vaststaat in de logica en taalkunde, maar wel in de vorm of op de manier waarop Searle hem gebruikt voor de KI. Blijkbaar verandert in de transitie van het vraagstuk van taalkunde en logica naar (empirie van) Harde KI de lading (betekenis) van het vraagstuk.

Searle antwoordt op het bezwaar van de Churchlands dat het *wel* een logische waarheid en *geen empirische* vraag is. Hij gebruikt hiervoor weer de Chinese Kamer, waarin Searle in de kamer de Chinese taal niet kan begrijpen (Can Chinese Rooms Think? (Map 4)). Natuurlijk gaat het *daardoor* dus weer over het domein van taal en betekenis van taal. In de taalkunde is de consensus aanwezig dat syntax niet voldoende is voor semantiek, maar in de empirie (van de KI) hoeft men daar niet per definitie boodschap aan te hebben, als wij de Churchlands zouden geloven. De empirie moet dit dus (nog) uitwijzen. Hiermee wordt echter (alleen) de empirie van de KI bedoeld, en die is blijkbaar niet afhankelijk of identiek aan de ‘empirie’ van logische waarheden in de logica of taalkunde. De conclusies uit andere wetenschappen hoeven blijkbaar niet geldig te zijn voor de empirie van de KI. Dit is een fundamenteel andere opvatting dan die Searle heeft, maar dit is wellicht te verklaren door de verschillende invalshoeken van beide ‘kampen’ wetenschappers in acht te nemen. Searle is een filosoof, terwijl de Churchlands zich ook veel met de praktijk van KI bezig houden (naast het nadenken over de filosofische kant van de KI).

Enerzijds kunnen we bij dit stuk van het debat de vraag stellen, op welke manier de (algemene) invloed van filosofen op KI gezien zou moeten worden, dat is echter uiteraard een grotere overkoepelende discussie in ‘filosofie van de KI’. De relatie tussen filosofie over KI en de wetenschappelijke praktijk van de KI zal ik kort aan de

[19] Dit is een samenvatting van een redenering van de Churchlands door de makers van ‘Map 4’.

orde stellen (in 6.1.3) via enkele citaten van objectieve beschouwingen van de problematiek. Anderzijds kunnen we de vraag stellen, waarom in dit geval Searles gebruik van de 'logische' (theoretische filosofie) waarheid voor de KI zo problematisch is (paragrafen 6.2 en 6.3).

6.1.3. Moor: huidige praktijk van de KI

James Moor geeft een commentaar bij Searles onderscheid tussen Harde en Zwakke KI, hiermee de types wetenschappers binnen de KI specificerend:

> 'Searle's distinction between weak and strong AI is a little misleading in that it invites an image of two conflicting camps doing research within the AI community [..]. In fact, the strong AI camp is inhabited by philosophers and researchers in their philosophical moments, whereas actual scientific research in AI is done in the weak AI camp. Thus, one effective way of defending AI against Searle is to agree with him that the computer is only a tool. Because scientific AI really is weak AI, an attack on strong AI is irrelevant to the discipline at least as it currently exists and will exist for the foreseeable future' (1988, p. 37).

Moor maakt ons duidelijk dat het huidige onderzoek in KI alleen 'Weak AI' is, en dat alleen filosofen over de mogelijkheden van Harde KI 'nadenken'. 'Though strong AI is not essential for current scientific AI, it is, nevertheless, a possible conceptual foundation for scientific AI of the distant future' (Moor, 1988, p. 39). Ik denk dat deze vreemde relatie tussen filosofie, praktijk en 'toekomstmuziek' een van de spillen van de kracht van Searles argumentatie is; er is (nog) geen onderzoek dat er daadwerkelijk tegenin gaat of tegenin kan gaan, ook al beweren de filosofen die de Harde KI verdedigen dat dit wel *gaat* gebeuren.

Moor maakt dezelfde opmerking als Haugeland (6.1.1) over de overname van de syntax-semantiek 'barrière' in de Harde KI:

> 'Strong AI need not deny that there is a logical distinction between syntax and semantics. The thesis of strong AI is that it may be possible to construct *high*-level semantics from *low*-level syntax. In other words, although at a low level of analysis a digital computer's operation is syntactical, at higher levels of organization semantic structures may emerge. These semantic structures are composed of nothing but syntactic units but are semantic in that they are causally connected to the world in the right way. It is not a conceptual truth that this view is false' (1988, p. 43).

Hierin staat (bijzonder kernachtig) beschreven *dat* en *waarom* de 'conceptual truth' vanuit het logische kamp niet geldt voor de KI (volgens Moor): semantiek op een hoger niveau kan uit de syntax (op een laag niveau) *emergeren*. De laatste zin van Moors conclusie vertelt ons dat de toekomst ons computers met minds kan brengen:

> ‘Causal computationalism permits a reasonable course between neural chauvinism and panpsychism. All nonbiological entities we know about lack minds. But one day, far in the future, the appropriately programmed computer may have one’ (1988, p. 51).

Deze conclusie is echter een geval van pure toekomstmuziek, en is (volgens Searle) geen bron van refutatie van de syntax-semantiekredenering. Daarnaast is het hopen op emergentie van semantiek niet erg geruststellend: zeggen dat iets zal emergeren, is als zeggen of hopen dat iets ‘vanzelf’ zal gaan, of dat je iets zal ‘laten emergeren’. Wellicht is en dergelijke vorm van emergentie, als die al mogelijk is, nauwelijks of niet te herkennen (alleen onder interpretatie, en dan weet je nog niet of er intrinsiek ‘iets’ aanwezig is).

Pollock, die in het eerste hoofdstuk aangestipt is als verdediger van Harde KI (‘*How to build a person*’), weet ook accuraat te beschrijven waarom de problematiek van filosofie en KI als zodanig bestaat:

> ‘My claim in this book is that the failure *[of achieving the dream of AI]* is not intrinsic to the task, but stems from the fact that many of the problems involved are essentially philosophical, while researchers in AI have not usually been trained in philosophy. Training in philosophy is not by itself sufficient to solve the problems, because they are hard problems and have difficult non-philosophical ingredients as well, but input from philosophers is probably a necessary condition for their solution’ (1989, p. viiii).

De onderzoekers in de KI hebben niet voldoende filosofische training ondergaan om via hun onderzoek de filosofische problemen aan te kunnen pakken: zij hebben filosofen dan wel filosofie ‘nodig’ om de problemen echt in de praktijk aan te kunnen pakken. Filosofie alleen is uiteraard ook niet voldoende, omdat er ook niet-filosofische ingrediënten meespelen.

Searle kan als filosoof tegen het Harde KI-kamp (nu al) zeggen waarom hij denkt dat deze basis die óók te maken heeft met ‘non-philosophical ingredients’, nooit voldoende kan zijn voor een mind; hij kan tegen alle ‘toekomstmuziek’ (met of zonder hoop op emergentie) in gaan. Waarom kan hij dat, op grond van welke, eventueel verkeerde, redeneringen?

6.2. Dynamiek van programma’s in implementatie

Een mogelijke redenering die uitlegt waarom de transitie van de syntax-semantiekredenering vanuit logica en taalkunde naar relevantie voor KI-programma’s zoals Searle die maakt, niet opgaat, ligt in het feit dat syntax en semantiek in logica anders van aard zijn dan ‘in KI-programma’s’. Het probleem is, dat syntax en

semantiek in talen geen dynamische noties zijn; natuurlijk is een programma, een syntax, op zich, 'inert':

> 'The property of being a process is not, then, a purely formal or syntactic property but includes, essentially, a non-syntactic element – an element of dynamism – besides.' (Hauser, 2002, pp. 126-127).

Het is niet zo dat Searle het over 'niet-running' programma's heeft, in hoofdstuk vier bleek uit een citaat uit het oorspronkelijke Chinese Kamer artikel al dat hij dit kenmerk van programma's niet negeert. Maar het lijkt erop dat Searle hier dit kenmerk niet voldoende in overweging neemt, en zo dus een verkeerde (incomplete) opvatting over computers gebruikt. Moor merkt ook op dat Searle een fout maakt: 'Picking appropriate realizations is not nearly as arbitrary as Searle suggests' (1988, p. 49). Searle denkt te simplistisch over de daadwerkelijke implementatie en realisatie van 'running' programma's. Preston merkt over Hausers artikel op: 'Could Searle have failed to take into account the fact that program runs have (causal and dynamic) properties that inert programs don't?' (2002, p. 36). Wellicht wel!

Dennetts verwerping van de stelling "syntax is niet equivalent aan of voldoende voor semantiek" stoelt op een soortgelijke gedachte:

> 'This *[proposition]* may still be held true, if we make the simple mistake of talking about syntax on the shelf, an unimplemented program. But *embodied, running* syntax – the 'right program' on a suitably fast machine – *is* sufficient for *derived* intentionality, and that is the only kind of semantics there is, as I argued in chapter 8 (see also the discussion of syntactic and semantic engines in chapter 3). So I reject, with arguments, Searle's proposition 2 *[Dennett verwijst naar: syntax is niet equivalent aan of voldoende voor semantiek]*' (1987, p. 336).

Zoals we kunnen lezen in het citaat, is Dennetts positie op meerdere redeneringen gestoeld, maar dit is zijn verantwoording van het verwerpen van 'propositie 2'. Hij geeft hierin wel ook aan dat de stelling niet per definitie onwaar is, maar dat de stelling voor de syntactische en semantische *processen* in de KI *niet geldig hoeft te zijn.* Blijkbaar vindt ook hij dat er in de toepassing van de stelling op KI zoals deze in de filosofie van mind wordt gemaakt, niet gerechtvaardigd is, vanwege het niet inzien van de kenmerken van de dynamiek van programma's.

Ook bij Rapaport hebben we deze kritiek gezien, beschreven onder het kopje 'Systemen, programma's en processen' in 5.2.1. 'Rather, the question is whether a computer that is *running* (or executing) a suitable program – a (suitable) program being executed or run – can understand natural language. A program being actually being executed is sometimes said to be a "process" [..]' (1988, p. 81). Het proces met

al zijn verschillende aspecten en vereisten is datgene dat onder de loep moet worden genomen, moet worden gebruikt, en datgene dat kandidaat is voor het 'laten veroorzaken' van een mind.

De conclusie die Searle trekt over het missen van semantiek in computers omdat programma's slechts 'syntax' hebben, is een aanname die door velen op dezelfde manier wordt aangevallen; Searle maakt hier de fout om 'running', geïmplementeerde, realiseerde programma's, te eenvoudig ook onder 'slechts computatie te scharen'. Searle ziet dit te simplistisch. Bij Dennett, Rapaport, Haugeland, Hauser en Moor is deze kritiek te vinden. Zij hebben allen andere redenen om Searles redenering aan te vallen, maar ze maken allemaal gebruik van dit argument. Is het onderkennen van de dynamiek van programma's het zwakke punt van Searles redenering? Ja; volgens de *filosofen* die Harde KI verdedigen wel, zij kunnen zeggen dat dit de reden is dat de mogelijkheid tot minds in computers niet nu al bepaald kan worden, en dat dit nog moet worden uitgezocht, en in een sterker claimende opvatting, dat minds wel in computers kunnen ontstaan. De wetenschappers die werken in de *praktijk* van de KI werken, kunnen echter bijna geen kracht achter deze uitspraken zetten door een mogelijk gebrek aan training in de filosofie: de praktijk van de KI is op dit moment meer Zwakke KI dan Harde KI (zie citaat Moor).

Hiermee is Searles redenering niet direct ontkracht: hij dient eerder aangepast of versterkt te worden om deze kritiek te kunnen 'overwinnen'. Een dergelijke (expliciete) redenering is bijvoorbeeld te vinden in Fetzers boek '*Computers and cognition: why minds are not machines'* (2001, p. xv), "The dynamic difference".

6.3. Searles gebruik van de terminologie van syntax en semantiek

De opvatting over minds die Searle gebruikt, 'minds hebben *semantiek*', is een derde belangrijke bron van problematiek. Het gebruik van de krachtige terminologie vanuit de logica en taalkunde, vooral wat *semantiek* betreft, en Searles vertaling hiervan in minds en daarmee in computers is een bron van mogelijke kritiek. Want, waar staat semantiek in minds voor? Volgens Searle staat het voor *intrinsieke intentionaliteit*; dit is besproken in hoofdstuk drie. Het gebruiken van de term 'semantiek' uit de taalkunde en logica in een overgang naar een beschrijving van minds binnen de filosofie van mind, is echter lastig:

> ‘So the very difficult issue of how to describe the semantical relation carries over from the philosophy of language to the philosophy of mind. It is often called the ‘problem of intentionality’, though this label covers other issues as well’ (Guttenplan, 1994, p. 584).

Dat Searle deze terminologie zo gebruikt, heb ik in hoofdstuk eén een kwestie van *op zijn minst* ‘lenen’ van terminologie genoemd, maar het is een ingewikkeldere constructie dan slechts ‘het gebruiken van leentermen’. Zoals Guttenplan aangeeft, is het een algemene en moeilijke kwestie. Searle heeft hier wellicht ook wel weet van gehad:

> ‘Axiom 3. *Syntax by itself is neither constitutive of nor sufficient for semantics.* At one level this principle is true by definition. **One might, of course, define the terms syntax and semantics differently**. The point is that there is a distinction between formal elements, which have no intrinsic meaning or content’ (1990, p. 21)(mijn bold nadruk).

In dit citaat lijkt het alsof Searle zich bewust is van de aparte en zichzelf ondersteunende manier waarop hij deze termen ‘definieert’. Ook doet hij de volgende uitspraak over de argumenten van zijn critici:

> ‘All of these arguments share a common feature: they are all inadequate because they fail to come to grips with the actual Chinese room argument. **That argument rests on the distinction between the formal symbol manipulation that is done by the computer and the mental contents biologically produced by the brain, a distinction I have abbreviated – I hope not misleadingly – as the distinction between syntax and semantics**’ (1990, p. 24)(mijn nadruk).

Searles opmerking ‘I hope not misleadingly’ kunnen we nu dus zien als een bevestiging van de mogelijkheid dat zijn gebruik van de terminologie niet onproblematisch is! Hiermee geeft hij Haugeland een onderbouwing voor de opmerking dat Searles opvatting over KI misleidend is (zie boven) en (2002, p. 382).

De reden dat het misleidend is, is de vertaling van semantiek naar ‘intrinsieke intentionaliteit’; deze brengt een ‘kluwen’ van aannames en vaagheid met zich mee. Intrinsieke intentionaliteit, semantiek, de relatie met ‘causale krachten’; deze constructie die Searle, alhoewel zeer kundig, uiteenzet, is niet kraakhelder. De misleiding volgt onder andere uit de aanname van de noodzaak van bepaalde causale krachten. In hoofdstuk vier trok ik de volgende conclusie uit een citaat:

> ‘Formele modellen zijn niet constitutief voor intentionaliteit en hebben op zichzelf niet de juiste causale krachten.’

Nu kunnen we de vraag stellen: hoe moeten we de ‘oorzakelijkheid’ in deze conclusie opvatten?

> ‘Formele modellen zijn niet constitutief voor intentionaliteit en hebben **daarom** op zichzelf niet de juiste causale krachten.’

of als

> ‘Formele modellen zijn niet constitutief voor intentionaliteit, **want** ze hebben op zichzelf niet de juiste causale krachten.’

De afhankelijkheidsrelatie in deze conclusie wordt (door Searle) niet expliciet gemaakt.

Een andere conclusie die ik trok luidt:

> In het computationalisme is er dus absoluut geen sprake van ‘intrinsieke’ semantiek, en ‘**daarmee’** geen mogelijke veroorzaking van intrinsieke intentionaliteit.

Nu is de soortgelijke vraag: moet er staan ‘en daarom’ of ‘omdat’?

Is er geen intrinsieke semantiek omdat er geen intrinsieke intentionaliteit is, of is er geen intrinsieke intentionaliteit omdat er geen intrinsieke semantiek is? Er bestaat een onduidelijke relatie tussen deze begrippen.

Josef Moural geeft in zijn artikel over de Chinese Kamer van Searle een zelfde soort kritiek:

> ‘In the initial paper, the words “syntax” and “semantics” are introduced somewhat reluctantly as belonging to “the linguistic jargon”. [..] But Searle has found it convenient to formulate his point in terms of syntax and semantics, and we have no reason to follow him in that’ (2003, p. 249).

Het is handig voor Searle om deze terminologie te gebruiken, maar wellicht niet geheel gerechtvaardigd, omdat het zijn claims ongerechtvaardigd anders van aard maakt:

> ‘We have noticed that the claim about intentionality may be stronger than that about meaning, and it is not clear how the arguments based on the definition of computational operations supports the claim about intentionality’ (Moural, 2003, p. 250).

De claim over intentionaliteit vormt een ander argument (wellicht sterker dan het argument over semantiek). De twee argumenten zijn gerelateerd:

> ‘But while the semantics argument does this in a bottom-up fashion (that is, by focusing on the limits of what can be achieved solely by any number of syntactically defined operations on uninterpreted terms – the envisaged intentionality argument does it top-down – that is, by focusing on the closure of the realm of aboutness’ (Moural, 2003, p. 250).

Het argument gebaseerd op intentionaliteit is dus anders van aard dan het argument gebaseerd op semantiek. Dit kan leiden tot een contradictie van de argumentatie met Searles eigen biologisch naturalisme:

> ‘Specifically, the difficulty consists in the suspicion that, if valid, it *[argument gebaseerd op intentionaliteit]* would show not only that you cannot get intentionality from syntax, but also that you cannot get intentionality from anything that does not already have intentionality. This conclusion would be incompatible with Searle’s biological naturalism [..]. It is possible that Searle’s new argument about the observer-relativity of computation blocks this undesired conclusion, but to be clear about that requires (once again) mastering the difficult area of being a cause under a description’ (Moural, 2003, p. 251).

Het blokkeren van een ongewilde contradictie voor het biologisch naturalisme zou volgens Moural als volgt kunnen worden uiteengezet:

1) breinen veroorzaken minds, intentionaliteit (etc.) omdat ze de vereiste causale krachten hebben.

2) syntactische systemen kunnen dit niet, omdat ze slechts waarnemerrelatief zijn en op die basis geen eigen causale krachten hebben.

Een dergelijke uitleg is echter sterk afhankelijk van het ‘causale krachten’-argument; hier ontstaat weer een afhankelijkheid binnen de constructie van het begrippenapparaat. Causale krachten, semantiek en intentionaliteit worden dus teveel ‘op een hoop’ gegooid; het zou duidelijker zijn als Searle de onderlinge relaties (van veroorzaking of afhankelijkheid) van deze losse fenomenen beter zou beschrijven. Searle schrijft echter met dit begrippenapparaat, de wens om ‘iets intrinsieks’ (intentionaliteit, semantiek, causale krachten) te genereren aan Harde KI toe. Voor velen in de KI is de relevantie van de dit begrippenapparaat twijfelachtig. Semantiek is zeker iets dat gegenereerd moet worden in systemen van de Harde KI, maar waar semantiek voor staat, en om welke redenen (omdat het intentionaliteit veroorzaakt **of** aanduidt) het gegenereerd of veroorzaakt zou moeten worden, is dus het twistpunt.

Als je de benodigde causale krachten anders beschouwt, kun je deze vage constructie omzeilen (zoals Dennett bijvoorbeeld doet). De vraag is of met een andere opvatting dan die van Searle (over het ‘intrinsiek moeten zijn’) de kwestie van ‘persoonstatus’ op eenzelfde manier benaderd kan worden. Over deze kwesties schrijft James Fetzer ook in zijn ‘*Computers and cognition: why minds are not machines*’ (2001) (een recente aanrader). Zijn (zeer beknopte) samenvatting van het debat is als volgt:

> ‘Searle (1992) takes a different tack, contending that, if mental processes are nothing more than algorithmic operations on formal systems, the computational conception is not strong enough to encompass the nature of thought, because thoughts have a semantical dimension (they are meaningful), while manipulations of formal systems are purely syntactical (they do not possess any inherent meaning**). His position depends upon drawing a strong distinction between *syntax*, which concerns the relations**

> **that marks (or tokens or signs) bear to one another, and *semantics*, which concerns the relations between those marks and that for which they stand (or represent or signify)**.
> This position has been challenged by those, such as William J. Rapaport (1988), who want to maintain that pure syntax can be sufficient for semantic meaning. **Searle has sought to reinforce his position by adding the further thesis that, not only is semantics not inherent in syntax, but syntax is also not inherent in physics** (Searle 1992: 201-212). Computationalists could contend, however, that in some systems, at least, syntax is physical, namely, in the case of machines that are intentionally designed to operate on the basis of those marks, which is the case for digital computers. Thus, the views that Searle has advanced confront a variety of counterarguments.
> A crucial distinction must be drawn between marks (or tokens or signs) that are significant for the *users* of a system and those that are significant for *use* by a system (Fetzer,1988a, 1990c, 1991, 1992). Causal systems *can* be designed to operate on the basis of the sizes, shapes, and relative locations of various marks without those marks having any meaning for those systems themselves. **To this extent, Searle's critics seem to be right.** When those marks are envisioned as ***syntax*, however, they are viewed as the (actual or potential) bearers of meaning, which presupposes a point of view. In this sense, syntax *is* relative to an interpretation, interpreter or mind**. As Jahrens (1990) has observed, Rapaport's argument seems to beg the question by assuming that humans implement natural language "the same way it would be on a computer"' (2001, pp. 113-114)(mijn bold nadruk).

In dit citaat stipt Fetzer de kwestie van interpretatie en waarnemerrelativiteit aan, waarop Searle ingaat met zijn 'nieuwe redenering'. Searles gebruik van de termen syntax en semantiek vraagt om interpretatie, die de termen bijna onmogelijk maakt voor gebruik om iets intrinsieks aan te duiden. Als semantiek ingebed is in een vage constructie met intrinsieke intentionaliteit en andere begrippen, maakt deze constructie het, zoals Moural ook aanduidde, bijna onmogelijk om aan iets anders dan iets waarvan we al aannemen dat het intentionaliteit heeft, intentionaliteit toe te schrijven. Searle zegt echter wel dat het niet onmogelijk is dat een systeem anders dan de hersenen intentionaliteit heeft, maar dat dit alleen kan gebeuren als dat systeem de juiste causale krachten heeft. Meer antwoorden dan dat antwoord heeft hij (nog) niet, en hoeft hij volgens zijn biologisch materialisme, dat 'wacht' op de neurobiologie, ook niet te kunnen geven.

6.4. Conclusie 'Waardoor komt Searles conclusie in het geding?'

Er zijn uit mijn bespreking van Searle en sommige van zijn critici drie kwesties naar voren gekomen die de SSR *mijns inziens* en ook volgens anderen zo omstreden maken. Deze verzameling is waarschijnlijk niet uitputtend, maar het zijn wel de meest opvallende kwesties die uit mijn literatuurstudie naar voren kwamen, en waar ik nog iets verder op in ben gegaan.

Ten eerste, bestaat er de kwestie van de ‘botsende’ samenwerking (of het gebrek aan samenwerking) tussen de praktijk van KI, de filosofie van (Harde) KI en de filosofie van mind. De echte Harde KI bestaat alleen in de filosofische hoek van de KI (en (nog) niet in de praktijk; de discussie over Harde KI bestaat dus vooral op filosofisch vlak! Verwijzen naar de praktijk heeft voor de voorstanders minder nut dan voor de tegenstanders die kunnen zeggen dat de status van de techniek en de ‘toekomstmuziek’ er niet toe doet. Dit zorgt voor twee andere knelpunten (2 en 3).

Ten tweede bestaat er de kwestie van de wellicht te simplistische opvatting van Searle over de werking en de ‘causale krachten’ van computatie in computers. Searle zegt dat de status van de techniek er niet toe doet, omdat het vaststaat dat computatie gestoeld is op de reeds bestaande en uitgewerkte abstracte notie van Turing Machines – dit gaat niet veranderen. Zoals velen aanstippen, is echter die puur abstracte notie alleen niet datgene wat een computer *is*: de implementatie en de realisatie van de abstracte notie zorgt voor meer dan alleen een abstracte ‘machine’. Hiervoor moeten de filosofen die zo de Harde KI een warm hart toedragen, ‘kijken naar’ de computer zelf en er eigenlijk mee aan de slag gaan; zij geven dan ook toe dat een mind maken ‘op dit moment’ niet juist geprobeerd wordt. Maar, Searle kan dan nog steeds blijven verwijzen naar de basis van de abstractie van het computationalisme en computatie, en heeft daar ook een punt. De bewijslast zou dus moeten liggen bij diegenen die zeggen dat het wel mogelijk is, en zij moeten dit onder de juiste voorwaarden laten zien (die zij zelf ook onderschrijven!). Daarbij kan de redenering van Searle altijd weer aangepast of toegespitst worden, en is deze verre van ontkracht.

Ten derde is sprake van een problematisch ‘convenient’ gebruik van de terminologie van syntax en semantiek door Searle, maar ook door anderen (!) binnen de filosofie van mind. Deze krachtige terminologie zou wel eens ongeschikt kunnen zijn om alle facetten van een mind te kunnen beschrijven (een redenering over intentionaliteit is een andere dan over semantiek), en daarmee de mogelijkheden voor een mind in een computer verkeerd beschrijven. Maar, het feit is dat deze terminologie ook door voorstanders van Harde KI (binnen filosofie) wordt overgenomen en vervolgens van commentaar voorzien of omgebogen wordt naar een conceptenapparaat waarvoor de logische waarheid uit de logica en taalkunde niet meer geldt. Dit is op zijn minst verwarrend. Dit betekent dat Searles tegenstanders hem aanvallen, maar wel zijn begrippenapparaat eerst ‘accepteren’. Als je dit

vervolgens gaat aanpassen, kloppen de voorwaarden van de gehele redenering (inderdaad) niet meer, omdat je niet meer over hetzelfde praat!

Searles redenering en daarmee zijn eigenlijke conclusie staan hierom nog overeind, maar dus wel met de nodige op- en aanmerkingen en verwijzingen naar fouten die ‘buiten Searle’ vallen (door anderen ook worden gemaakt).

7. Conclusie

Harde KI is een door Searle zo gedoopt paradigma binnen de Kunstmatige Intelligentie, waarvan de basis het vertrouwen in het kunnen maken van minds in computers is. De systemen waarmee in de Harde KI wordt gewerkt hebben een computationalistische basis en zijn symboolmanipulerende digitale systemen. Het willen maken van een mind komt idealiter uit op het willen maken van een persoon. Deze sterke claim is exact waartegen Searle redeneert, op basis van zijn opvattingen over mensen (personen) en de relevante kenmerken van persoon zijn.

De syntax-semantiekredenering bestaat uit een aanname over mentale fenomenen bij mensen (semantiek) en de onmogelijkheid tot het hebben van mentaliteit voor computers. De hersenen van mensen veroorzaken mentaliteit, en met name intrinsieke intentionaliteit. ‘Semantiek’ in de redenering van Searle hangt samen met het hebben van mentale fenomenen: hoe semantiek hebben, mentale fenomenen hebben (intrinsieke intentionaliteit hebben) en de daarbij benodigde veroorzakende causale krachten aan elkaar gerelateerd zijn, is niet op te maken uit Searles opvatting *dat* deze allemaal gerelateerd zijn. Deze opvatting gebruikt Searle als fundamenteel gedachtegoed voor zijn positie binnen de filosofie van mind, het biologisch naturalisme. Het biologisch naturalisme beschrijft mentale fenomenen als realistische fenomenen in de natuur, en schrijft een sterke relevantie toe aan het eerste persoonsperspectief. Alleen dit perspectief is constitutief voor mentaliteit, mentaliteit is ontologisch subjectief; dit betekent niet dat de kennis over dit domein (mentaliteit) *epistemologisch* subjectief is: de kennis erover kan wel degelijk objectief zijn. De tak van wetenschap waar Searle naar verwijst om antwoorden te geven over hoe mentaliteit en alle begrippen uit zijn begrippenapparaat voor mind samenhangen, is de neurobiologie.

Dennett ziet niet dezelfde relevantie van het eerste persoonsperspectief en daardoor ook niets intrinsieks aan de benodigde causale krachten: deze zijn voor hem simpeler uit te leggen door naar de snelheid en capaciteit van de hersenen te verwijzen: de onmogelijkheid om die te evenaren zorgt volgens hem voor een probleem in de Harde KI, en niet de problemen zoals Searle die beschrijft.

Computers missen volgens Searle per definitie semantiek, omdat de essentiële definiëring van hun processen in termen van syntax uiteen wordt gezet. Alleen Haugeland geeft een hypothetische beschrijving van hoe computers wel semantiek zouden kunnen hebben. De semantiek die hierbij komt kijken bestaat, volgens Searle, slechts in de interpretatie van de programmeur of gebruiker van de processen en hun invoer en uitkomsten. Deze semantiek is niet 'zomaar' intrinsiek aanwezig en kan dat ook niet zijn, omdat de computatie een waarnemerrelatief (niet intrinsiek aan de fysische processen van de computer) begrip is. Bij mensen is het basisgeval van semantiek hebben een intrinsieke notie die samenhangt met het hebben van de juiste causale krachten (volgens Searle); semantiek hebben is niet tot een syntactische te reduceren. Rapaport is van mening dat het basisgeval van semantiek *wel* syntactisch is, en dat Searle mensen en computers hierin als gelijken moet behandelen. Haugeland is van mening dat er wel in principe enige vorm van semantiek in computers mogelijk is die de basis kan zijn voor uitgebreidere semantiek; hiervoor zou echter een ander soort opzet van systemen moeten worden gebruikt dan de Harde KI nu gebruikt.

De echte knelpunten van de redenering liggen, zoals blijkt uit de bestudering van Searle en enkele van zijn critici, niet in de redenering als zodanig, maar in de aparte status van de filosofie binnen de Harde KI, in de opvatting over computersystemen als (louter) abstracte systemen, en in het gebruik van de krachtige terminologie vanuit de logica en taalkunde voor zowel het domein van de menselijke mind als het domein van de Harde KI (digitale computers).

Nu komen we terug bij de hoofdvraag van dit literatuuronderzoek: Waarom is de syntax-semantiekredenering van Searle een probleem voor de Harde KI? Het antwoord luidt: Omdat de aannames zoals Searle ze doet geldig zijn (en de 'foutieve opvattingen' niet aan hem te wijten zijn), heeft Searle in zeker opzicht gelijk en is Harde KI een 'vertrouwen' in het kunnen bewerkstelligen van minds in computers dat, vanwege de in de Harde KI gebruikte manier van bewerkstelligen, een ongegrond vertrouwen kan zijn. De problematiek rond het ontkrachten van de redenering gaat verder dan de redenering zelf. Daarnaast bestaat er al de consensus dat de redenering ongeldig is, terwijl een genadeslag nog niet gegeven is! Dit is de bron van het debat. De redering is niet ongeldig, maar de opvattingen die gepaard gaan met de aannames

zijn de daadwerkelijke twistpunten en facetten die de redenering twijfelachtig kunnen maken.

Deze conclusie geeft mijns inziens aan dat de redenering op andere manieren moet worden bekeken om ‘opgelost’ of ‘ontkracht’ te kunnen worden. Slechts het aangeven dat er ‘iets niet klopt’ volstaat bij lange na niet om het tegendeel van de conclusie aan te tonen. De relevantie van mijn resultaten ligt in het kunnen verdedigen van Searle (wat tegen ‘de consensus’ ingaat), en het kunnen aangeven waar in het debat de ‘uitwegen’ en de knelpunten liggen.

Mijn suggesties voor het verder bekijken van de uitwegen, het verdiepen van de onderliggende redeneringen, uitwerken van meningsverschillen en uitbreiden naar andere, aanverwante onderwerpen, liggen in de volgende kwesties:

Aangestipte of tot op zekere hoogte besproken onderwerpen:

- Waarnemerrelativiteit van interpretaties en realisme over natuurlijke fenomenen (Haugeland, Dennett, Searle).
- Haugelands opvatting: computers hebben veel mogelijkheden (meer dan Searle erkent) maar mensen zijn weer zodanig ‘bijzonder’ (meer dan Searle beschrijft) dat het alsnog onmogelijk is om computers op mensen te laten gelijken.
- ‘The Dynamic Difference’ (Fetzer): SSR bekijken ten opzichte van de dynamiek van geïmplementeerde, ‘running’ programma’s.
- Rechtvaardiging of bestrijding van het gebruik van terminologie van syntax-semantiek voor beschrijvingen van minds.
- Analyse van de (hypothetische) onderlinge relaties tussen mentale verschijnselen (bijvoorbeeld door bestudering van Searle, die deze niet expliciet beschrijft).

Verwante (nauwelijks of niet besproken) onderwerpen:

- De kritiek van Searle (de SSR) in relatie tot het connectionisme (Searle en anderen vs. bijvoorbeeld Stevan Harnad).
- Symbol grounding en semantiek in (sub)symbolische netwerken (o.a. Stevan Harnad)
- Is Searles biologisch naturalisme niet reductionistisch (zoals hij zelf wil bepleiten)?
- Persoonstatus van mensen en minds in relatie tot ‘persoon zijn’
- Filosofie en wetenschap – filosofie in relatie tot praktijk van KI
- Syntax als notie die niet (eens) zonder semantiek kan bestaan!

Ik wens degenen die deze aanbevelingen zouden willen opvolgen veel plezier en succes met deze eigenaardige doch interessante en uiteenlopende mogelijkheden binnen de studie CKI. Ik bedank u, de lezer, voor uw aandacht!

8. Literatuur

Block, N. (1994). Functionalism (2). In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (pp. 323-332). Oxford: Blackwell Publishing Ltd.

Bringsjord, S. (1992). *What robots can and can't be.* Dordrecht: Kluwer Academic Publishers.

Copeland, J. (1993). *Artificial Intelligence: A Philosophical Introduction.* Oxford: Blackwell.
Copeland, J. (2002). The Chinese Room from a Logical Point of View. In J. Preston, & M. Bishop, *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence* (pp. 109-122). Oxford: Clarendon Press.

Cunningham, S. (2000). *What is a Mind? An Integrative Introduction to the Philosophy of Mind.* Indianapolis; Cambridge: Hackett Publishing Company, Inc.

Dennett, D. C. (1987). *The Intentional Stance.* Cambridge, Massachusetts: The MIT Press.

Driessen, Y. (2002, Januari). *Gedachte-experimenten.* Opgeroepen op Februari 2007, van preprints CKIscripties: http://www.phil.uu.nl/preprints/ckiscripties/SCRIPTIES/008_driessen.pdf

Fetzer, J. H. (2001). *Computers and cognition: why minds are not machines.* Dordrecht: Kluwer Academic Publishers.

Fodor, J. (1995). The Folly of Simulation (interview). In P. Baumgarter, & S. Payr, *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists* (pp. 85-100). Princeton: Princeton University Press.

Franklin, S. (1995). *Artificial Minds.* Cambridge: MIT Press.

Guttenplan, S. (1994). syntax / semantics. In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (pp. 583-584). Oxford: Blackwell Publishing Ltd.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea.* Cambridge, Massachusetts: The MIT Press.
Haugeland, J. (1998). *Having Thought. Essays in the Metaphysics of Mind.* Cambridge, Massachusetts and London, England: Harvard University Press.
Haugeland, J. (2002). Syntax, Semantics, Physics. In J. Preston, & M. Bishop (eds.), *Views into the Chinese Room. New Essays on Searle and Artificial Intelligence* (pp. 379-392). Oxford: Clarendon Press.

Hauser, L. (2002). Nixin' Goes to China. In J. Preston (ed.), *Views into the Chinese Room. New essays on Searle and Artificial Intelligence.* (pp. 123-143). Oxford: Clarendon Press.

Lycan, W. G. (1994). Functionalism (1). In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (pp. 317-322). Oxford: Blackwell Publishing Ltd.

Moor, J. H. (1988). The pseudorealization fallacy and the Chinese Room argument. In J. H. Fetzer (ed.), *Aspects of Artificial Intelligence* (pp. 35-53). Dordrecht: Kluwer Academic Publishers.

Moural, J. (2003). The Chinese Room Argument. In B. Smith (ed.), *John Searle* (pp. 214-260). Cambridge: Cambridge University Press.

onbekend. *Can Chinese Rooms Think? (Map 4)*. Opgeroepen op 2006-2007, van Mapping Great Debates: Can Computers Think? The History and Status of the Debate: http://www.macrovu.com/CCTWeb/CCT4/CCTMap4.html

onbekend. (2007, april 20). *Straw man*. Opgeroepen op april 21, 2007, van Wikipedia: http://en.wikipedia.org/wiki/Straw_man
onbekend. (2006, Oktober 4). *Stropopredenering*. Opgeroepen op April 21, 2007, van Wikipedia: http://nl.wikipedia.org/wiki/Stropopredenering

Pollock, J. L. (1989). *How to Build a Person: a Prolegomenon.* Cambridge: MIT Press.

Poole, D., Mackworth, A. K., & Goebel, R. (1998). What is Computational Intelligence? In D. Poole, A. Mackworth, & R. Goebel, *Computational Intelligence: A Logical Approach* (pp. 1-7). New York: Oxford University Press.

Preston, J. (2002). Introduction. In J. Preston, & M. Bishop (eds.), *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence* (pp. 1-50). Oxford: Clarendon Press.

Rapaport, W. J. (1988). Syntactic semantics: foundations of computational natural-language understanding. In J. H. Fetzer (ed.), *Aspects of Artificial Intelligence* (pp. 81-131). Dordrecht: Kluwer Academic Publishers.
Rapaport, W. J. (1995). Understanding Understanding: Syntactic Semantics and Computational Cognition. *Philosophical Perspectives, Vol 9, AI, Connectionism and Philosophical Psychology* , 49-88.

Russell, S. J., & Norvig, P. (1995). *Artificial Intelligence. A modern approach.* New Jersey: Prentice Hall International Inc.

Rychlak, J. F. (1991). *Artificial Intelligence and Human Reason: A Teleological Critique.* New York: Columbia University Press.
Rychlak, J. F. (1997). *In defense of human consciousness.* Washington: American Psychological Association.

Searle, J. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences* , 235-252 reprinted in Heil, John: Philosophy of Mind. A guide and anthology. 2004. Oxford: Oxford University Press.
Searle, J. R. (1983). *Intentionality. An essay in the philosophy of mind.* Cambridge: Cambridge University Press.

Searle, J. (1984). *Minds, Brains and Science. The 1984 Reith Lectures.* London: British Broadcasting Corporation.
Searle, J. (1990). Is the Brain's Mind a Computer Program? *Scientific American, vol 262* , 20-25.
Searle, J. (1992). *The Rediscovery of the Mind.* Cambridge: MIT Press.
Searle, J. (1994a). Intentionality (1) (companion entry). In S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (pp. 379-386). Oxford: Blackwell Publishers Ltd.
Searle, J. (1994b). Searle, John R. (companion entry). In S. Guttenplan (ed.), *A companion to the Philosophy of Mind* (pp. 544-550). Oxford: Blackwell Publishers Ltd.
Searle, J. (1995). Ontology is the question (interview with John Searle). In P. a. Baumgarter, *Speaking Minds. Interviews with Twenty Eminent Cognitive Scientists* (pp. 203-213). Princeton: Princeton University Press.
Searle, J. (1997). *The Mystery of Consciousness.* London: Granta Books.
Searle, J. (2001). *The Failures of Computationalism: I.* Opgeroepen op januari 2007, van Psycholoquy: http://psycprints.ecs.soton.ac.uk/archive/00000189/
Searle, J. (2004a). *Mind. A Brief Introduction.* New York: Oxford University Press.
Searle, J. (2004b). *Biological Naturalism.* Opgeroepen op maart 1, 2007, van Professor John Searle: http://ist-socrates.berkeley.edu/~jsearle/BiologicalNaturalismOct04.doc