

# Mentala modeller och externalismen

*—Om förhållandet mellan Johnson-Lairds mentala modeller och Putnams externalistiska synsätt*

**Av Andreas Chatzopoulos, 2006**

C-uppsats i Kognitionsvetenskap

Kollegium SSKKII,

Göteborgs universitet

*Handledare: Pierre Gander*

# Innehållsförteckning

1) Inledning	3
2) Språkfilosofisk bakgrund:	
2.1) Äldre teorier	4
2.2) Naiv referensteori	5
2.3) Språkanvändning	7
2.4) Bestämda beskrivningar	9
2.5) Kripke:	
2.5.1) Kritik av bestämda beskrivningar	10
2.5.2) Kausala kedjor	12
3) Johnson-Lairds mentala modeller	12
4) Mentala modeller och externalismen	17
4.1) Putnams externalism	17
4.2) Johnson-Lairds synsätt	20
4.3) Diskussion	23
4.3.1) Narrow och wide	24
4.3.2) Slutdiskussion	25
5) Slutsats	26
6) Källor	28

# 1) Inledning

Uppsatsen är inspirerad av P.N. Johnson-Lairds teori om mentala modeller. Detta är i grunden en *psykologisk meningsteori*, det vill säga en teori med avsikt att beskriva de psykologiska processer som är kopplade till språklig mening. Huvudsyftet är att förklara hur de språkliga uttryckens mening representeras psykologiskt. Trots det faktum att det är en psykologisk teori och att Johnson-Laird mest är intresserad av de psykologiska aspekterna, så är det oundvikligt att en teori av detta slag kommer in på språkfilosofiskt område, det vill säga frågor om meningens natur, vad mening egentligen är, var man kan finna den, och så vidare. Det är denna aspekt som intresserar mig. En mycket utbredd uppfattning inom samtida filosofi är nämligen den att den språkliga meningen till stor del bestäms av *omvärlden* snarare än av föreställningar i någons medvetande. Denna uppfattning kallas ofta för *externalism*, och dess mest kända förespråkare torde vara filosofen Hilary Putnam. Min tanke med denna uppsats är därför att undersöka hur Johnson-Laird hanterar just Putnams externalism. Johnson-Lairds teori går nämligen ut på att mening psykologiskt representeras av mentala modeller (i medvetandet). Vad jag då vill undersöka är hur denna uppfattning är förenlig med teorier som ser mening som någonting som till stor del dikteras av omvärlden. Det är det som är mitt syfte: *att undersöka hur Johnson-Laird hanterar externalismen och se om det två synsätten är förenliga*. För att uppsatsen skall bli läsbar även för de som inte är insatta i språkfilosofi så kommer jag att ge en kortfattad sammanfattning av språkfilosofiska teorier från John Locke fram till idag, med tyngdpunkt på moderna teorier. Jag kommer sedan först att redogöra för teorin om mentala modeller och sedan för externalismen. Efter det följer en undersökning av vad Johnson-Laird har att säga om externalismen, och till sist en kritisk diskussion av Johnson-Lairds synsätt.

## 2) Språkfilosofisk bakgrund:

### 2.1) Äldre teorier

Hur får språket mening? Detta är en traditionell, grundläggande filosofisk fråga som det gjorts många försök att besvara genom åren. Ett tidigt lösningsförslag framlades till exempel av John Locke i hans bok, *An Essay concerning Human Understanding* (1689). Han ansåg att all kunskap har sin upprinnelse i erfarenheten, och att grunden för vår tankevärld utgörs av "inre bilder" som grundas i sinnesintrycken. Dessa inre bilder kallade han för *idéer*. Något som låg nära till hands var då att se språklig mening som uppbyggd av dessa idéer, vilket var precis vad Locke gjorde. Han tänkte sig att meningen hos *singulära uttryck* (egennamn) utgörs av idén av det som representerar, det vill säga den inre bild vi har av personen/objektet i fråga (en bild som vi från början erhållit genom våra sinnen). Vad gäller *predikat* (det vill säga egenskaper som flera olika individuella objekt kan dela) så ansåg Locke att dessa representerades av vad han kallade för *generella idéer*. Dessa skall ses som framabstraherade ur de enkla idéerna. *Exempel*: jag har genom observation av många olika träd (som ligger till grund för enkla idéer om individuella träd) abstraherat fram de egenskaper som alla träd har gemensamt.. Denna framabstraherade bild utgör då min generella idé om "träd", dvs. är en representant för predikatet "träd".

Den främsta kritiken mot denna världsbild torde vara Berkeleys kritik som han utvecklade i *A Treatise Concerning the Principles of Knowledge* (1710) och går ut på att generella idéer enligt Lockes modell inte kan ha några bestämda egenskaper över huvudtaget. Detta inser man om man betänker hur en generell idé om, exempelvis, en triangel bör se ut. Denna idé (inre bild) måste ju föreställa en triangel som antingen är rätvinklig eller inte är det. Om den då inte är rätvinklig, låt säga, så kan den ju inte representera alla trianglar som är rätvinkliga (och vice versa). Fler sådana här aspekter kan adderas; triangeln på bilden måste antingen vara liksidig eller icke liksidig, likbent eller icke likbent osv. Kontentan blir då att det inte kan finnas någonting sådant som en generell bild som inkluderar alla tänkbara egenskaper. Detta eftersom bilden *i sig* måste ha vissa specifika egenskaper. En relaterad invändning är

också att varje konkret föreställningsbild måste avbilda motivet på ett bestämt sätt (från en viss vinkel osv.); varje konkret föreställningsbild av en färg måste representera en specifik nyans osv. En och samma person representeras då av flera olika bilder beroende på betraktningvinkel med mera. Detta verkar inte riktigt kunna förklaras med Lockes teori.

Ett annat problem uppstår med ord som ”icke”, ”eller”, ”förutsatt att” osv. Vilken inre bild representeras av sådana ord? Ett i detta sammanhang klassiskt exempel är satsen: ”det står inte mat på bordet”. Vilken inre bild kan den tänkas representera? Kanske ett tomt bord? Men: ett tomt bord kan ju likväl representeras av satsen: ”det står inte en staty av Beethoven på bordet” och liknande konstruktioner. Det står helt klart att denna invändning skapar problem för Lockes bildteori.

Sist men inte minst drabbas även Lockes bildteori av samma kritik som alla teorier som hävdar att språklig mening är någonting mentalt, nämligen undran över hur meningen kan vara publik (åtkomlig för alla språkanvändare) om den är helt och hållet privat. För att ett språk skall fungera så måste ju meningen hos ord och satser vara (någorlunda) samstämmig; det fungerar inte om alla har sin egen privata mening som kan vara helt godtycklig, menar kritikerna. Denna kritik behöver inte vara förödande för en teori som Lockes, men frågan måste utredas.

## 2.2) Naiv referensteori

Om det inte, på Lockes vis, är möjligt att koppla språklig mening till inre idéer så är en närliggande lösning att associera mening med det som orden refererar till i omvärlden. Detta leder fram till ståndpunkten att meningen hos ett språkligt uttryck är det som det refererar till i omvärlden; mening *utgörs* alltså av referens. En sådan teori brukar benämnas *naiv referensteori* eller *namnteori för mening*. Skillnaden mellan singulara uttryck och predikat i en sådan teori är att meningen hos singulara uttryck (egennamn) utgörs av de individer som de refererar till, medan meningen hos predikat (egenskaper och relationer) normalt ses som den mängd av individer som predikatet kan appliceras på.

Den teori som här skissats upp är väldigt grovhuggen, och det är oklart om någon filosof verkligen har vidmakthållit en sådan teori. Den skall snarare ses som ett utkast, en utgångspunkt i sökandet efter en adekvat meningsteori, och som sådan skapar den en hel del frågor som bör besvaras av en god meningsteori. De viktigaste av dessa frågor ställs av Gottlob Frege som i sin klassiska text *Über Sinn und Bedeutung* (1892) klargör att om mening utgörs av referens så måste alla identitetsutsagor bli självklara—vilket de inte är. Hans klassiska exempel för att påvisa detta består av en analys av egennamnen "aftonstjärnan" och "morgonstjärnan"; dessa är olika namn på en och samma sak, nämligen planeten Venus. De två uttrycken har alltså samma referens. Om nu mening skulle utgöras av referens så borde satsen "morgonstjärnan=aftonstjärnan" vara självklar eller trivial, det vill säga inte ge någon ny information över huvudtaget, på samma vis som satsen "morgonstjärnan = morgonstjärnan". Nu är det ju uppenbarligen inte så; satsen "morgonstjärnan = aftonstjärnan" är ju i allra högsta grad informationsbärande eftersom den ger ny kunskap till alla som inte känner till att de två egennamnen betecknar samma objekt. Rent historiskt representerar denna sats en betydande vetenskaplig upptäckt och borde då verkligen klassas som informationsbärande och långt ifrån trivial. Freges slutsats blir att meningsbegreppet måste innefatta *mer* än bara referens. Detta leder honom till att utveckla begreppen Sinn och Bedeutung (översätts vanligen med *mening* respektive *betydelse*). Freges åsikt är att varje Sinn motsvaras av en bestämd Bedeutung, det vill säga Sinn bestämmer Bedeutung. Detta synsätt har haft djupa influenser på senare filosofi, även om man idag inte längre talar om just Sinn och Bedeutung, utan snarare om de motsvarande logiska begreppen *intension* respektive *extension*. På samma vis som Sinn bestämmer Bedeutung så bestämmer intensionen extensionen. Detta genom att intension och extension är definierade i termer av varandra: en extension är en *mängd* av någonting. *Vad* detta någonting är specificeras i den associerade intensionen. Med andra ord: intensionen specificerar *innehållskriterierna*, det vill säga kraven som ställs för att ett objekt skall få lov att bli inkluderat i den associerade extensionsmängden. Givet ett språkligt uttryck så sätter man ofta likhetstecken mellan dess mening och dess intension. Detta leder till en definition av extensionen som den mängd objekt som uttrycket kan appliceras på. Alltså: enligt

detta synsätt utgörs meningen hos (exempelvis) predikatet "röd" av mängden röda objekt i omvärlden. Med andra ord: intensionen för "röd" bestämmer att den associerade extensionen skall vara mängden av alla röda föremål.

Förutom Freges kritik så kan man även hitta andra invändningar mot en naiv referensteori, exempelvis Bertrand Russells välkända analys av satsen "kungen av Frankrike är skallig" (Russell, 1905). Denna sats säger ju någonting om kungen av Frankrike, nämligen att han är skallig. Men kruxet är nu att det ju *inte finns* någon kung av Frankrike, och det verkar då som att satsen måste vara falsk. Om man då negerar den så borde man då erhålla en sann sats, och denna negation borde se ut som: "kungen av Frankrike är inte skallig". Men denna sats är ju *inte heller* sann eftersom den *också* säger någonting om kungen av Frankrike (att han *inte* är skallig), som ju fortfarande inte existerar. Kontentan är nu att detta strider mot *lagen om det uteslutna tredje*, en allmänt vedertagen logisk princip som säger att det för varje entydig sats gäller att den antingen är sann eller att dess negation är det. Någon tredje möjlighet existerar inte. Tydligt gäller det inte här, och Russells slutsats blir då att teorin i sin ursprungliga form måste vara felaktig.

Det är således mycket som pekar på att denna grovskiss av en meningsteori inte håller eftersom den skapar en hel del frågor som den inte kan svara på. Man måste därför försöka finna en annan angreppsvinkel:

### 2.3) Språkanvändning

Ett annat sätt att se på saken grundar sig i insikten att vi för att förstå språkets natur, först måste förstå dess *funktion*. Denna inställning bygger på tanken att språkets natur är underordnad dess användning, och en meningsteori av detta slag bör därför uppfatta språket och dess olika delar utifrån detta synsätt. Resultatet blir då förhoppningsvis en större helhetssyn på frågan om språklig mening; istället för att endast fokusera på meningen hos de *individuella* orden så koncentrerar man sig på *hela* satsen och vad vi vill förmedla genom vårt användande av dem. Finn Collin och Finn Guldmann ger en utförlig sammanfattning av detta förhållningssätt i sin bok *Språkfi-*

*losofi* (2000). De talar om *syftet* med ett språkligt uttryck som en lämplig startpunkt för en analys. Vad är exempelvis syftet hos ett språkligt *påstående* (som hävdar någonting om världen)? Detta kan redogöras för i termer av *framgångsvillkor*, det vill säga en specifikation av vad som krävs för att min språkhandling skall lyckas:

- Att yttrandet är *sant*.
- Att åhöraren *tror* att yttrandet är sant.
- Att talaren har *goda skäl* att tro att yttrandet är sant.

Notera att *sanningsbegreppet* spelar en avgörande roll här. Det visar sig att det ibland är fördelaktigt att beskriva framgångsvillkoren i termer av *hävanden*, det vill säga *sanningsvillkor*. Detta angreppssätt kan även tillämpas på andra språkhandlingar än påståenden. Ett par exempel:

- Att *hävda* att det är fallet att  $p$ : Denna talhandling har som framgångsvillkor att  $p$  är sant. Med andra ord:  $p$  måste vara sant för att mitt hävdande att det är sant skall vara framgångsrikt.
- Att *beordra* att det skall bli fallet att  $p$ : Framgångsvillkoret hos denna talhandling är att ordermottagaren gör så att  $p$  blir sann.

Man ser här hur man kan beskriva olika talhandlingars framgångsvillkor med ett motsvarande hävdande, så att varje framgångsvillkor får ett motsvarande sanningsvillkor. På så vis erhålles en meningsteori baserad på *sanningsbegreppet*, där satsernas *mening* utgörs av deras associerade *sanningsvillkor*. Detta motsvaras av tanken att om man förstår meningen hos en sats så förstår man även vad som krävs för att satsen skall vara sann, det vill säga man förstår just sanningsvillkoren. Detta gör då sanningsbegreppet till meningsteoriens centrala begrepp.

Om man analyserar komponenterna hos en språklig sats (de ingående orden) så finner man att deras mening kan uppfattas som det bidrag var och en av dem lämnar till satsens mening. Med andra ord så kan deras mening ses som *deras bi-*

*drag till bestämmandet av sanningsvillkoren.* Detta är en uppfattning som verkar rimlig när man insett att de gamla teorierna inte håller, samtidigt som man har kommit på det klara med att det centrala begreppet för beskrivning av talhandlingar är sanningsbegreppet. Dock finns här en liten motsättning; det traditionella synsättet är att en meningsteori bör vara *kompositionell*. Detta innebär att varje ord har sin egna, individuella mening, och att meningen hos satsen som helhet härleds ur de ingående ordens mening parade med de grammatiska reglerna. Den teori som här skissats upp är dock snarare *abstraktiv*, det vill säga den beskriver ordens meningsinnehåll som en abstraktion från meningsinnehållet hos satsen som helhet istället för att se på ordens meningsinnehåll som *komponenter*. Collin & Guldman föreslår därför en kompromiss; de argumenterar för en abstraktiv teori med kompositionella drag, inspirerad av Michael Dummett (Collin & Guldman, 2000, kap.2). Enligt denna medelväg så har ord förvisso sina egna meningsinnehåll oberoende av satssammanhanget, men icke desto mindre är det så att ett ords meningsinnehåll måste definieras som dess bidrag till alla möjliga satser som det kan ingå i. Collin & Guldman jämför med en uträkning av en pengasumma från en kassa: varje sedel och mynt har ett värde i form av ett tal, och uträkning av det totala värdet sker genom en summering. Här är alltså situationen den att de ingående pengarnas individuella värde bidrar till totalsumman på samma vis som ordens individuella mening bidrar till satsens mening.

## 2.4) Bestämda beskrivningar

Hur får då dessa individuella ord sin mening? Vad gäller singulära uttryck så verkade det ju inte fungera att sätta likhetstecken mellan deras mening och deras referens, såsom försöktes inom den naiva referensteorin. För att lösa dessa problem inför Russell (1905) begreppet *bestämda beskrivningar* och identifierar språklig mening med dem. Han anser att det korrekta sättet att behandla satser som: "kungen av Frankrike är skallig" måste innebära en omskrivning av uttrycket till dess logiska form, vilket i detta fall skulle resultera i någonting i stil med:

Det finns en och endast en sak som är nuvarande kung av Frankrike, och varje sak som är nuvarande kung av Frankrike är skallig

Detta är vad Russell kallar för en bestämd beskrivning. Den påstår *först* att något existerar, och *sedan* appliceras ett predikat på detta något. Om det inte finns någon referent (som i fallet ovan) så är helt enkelt existenspåståendet falskt, dock utan att den bestämda beskrivningen blir meningslös. Vi skulle då få någonting i stil med:

Det är inte så att det finns en och endast en sak som är nuvarande kung av Frankrike, och varje sak som är nuvarande kung av Frankrike är skallig

Detta verkar lösa det aktuella problemet, och lagen om det uteslutna tredje kan bibehållas. Det visar sig att detta synsätt även löser problemet med fiktiva namn; Russell anser att dessa (och alla andra) egennamn *egentligen* är bestämda beskrivningar. "Sherlock Holmes" bör då ses som en "förkortning" för en bestämd beskrivning i stil med: "Det finns en och endast en person som är detektiv och bor på 221b Baker Street samt har egenskaperna  $x$ ,  $y$ ,  $z$ ..." (där  $x$ ,  $y$ ,  $z$  då naturligtvis motsvaras av de egenskaper som är signifikanta för Sherlock Holmes). Denna bestämda beskrivning skulle då vara *falsk* i och med att någon sådan person aldrig existerat ("Det är inte så att det finns..."), men den skulle aldrig vara *meningslös*. På så vis blir även fiktiva namn meningsfulla. Russells slutsats är här att det vi i vanligt tal kallar för egennamn inte är egentliga egennamn, de är bestämda beskrivningar. De *äkta* egennamnen, de Russell kallade för *logiska egennamn*, är ord som "denna" och "jag", det vill säga *demonstrativer*. Hos dessa uttryck menar han att meningen utgörs av referensen.

## 2.5) Kripke:

### 2.5.1) Kritik av bestämda beskrivningar

Russells bestämda beskrivningar löser till synes en hel del, men har under den senare halvan av 1900-talet blivit kritiserade, främst av Saul Kripke (1980) som bland annat gör följande invändningar:

- Man kan använda sig av ett egennamn utan att veta någonting om bäraren utan att det man säger för den skull blir meningslöst.
- Man kan ge flera olika bestämda beskrivningar för ett och samma namn.
- Olika personer ger olika bestämda beskrivningar på samma personer

Kripke förnekar också att egennamn är synonymt med bestämda beskrivningar (Russell ansåg ju att egennamn bör betraktas som *förkortningar* för bestämda beskrivningar). Namn pekar bara ut objekt, menar Kripke. De har ej mening *i sig själva*. Ett viktigt krav som som han ställer är även att namn skall vara *rigida designatorer*. Med detta menar han att ett namn skall peka ut *samma* individ i alla möjliga världar där objektet i fråga existerar. Om detta skriver han:

Let's call something a rigid designator if in every possible world it designates the same object, a nonrigid or accidental designator if that is not the case. Of course we don't require that the objects exist in all possible worlds. [...] When we think of a property as essential to an object we usually mean that it is true of that object in any case where it would have existed. (1980, s.48)

Om jag exempelvis talar om Napoleon så skall det gå att inkludera honom i en hypotetisk diskussion om hur världen skulle ha sett ut ifall han inte hade tagit makten i Frankrike. För att detta skall fungera krävs då att namnet fungerar som en rigid designator, det vill säga refererar till Napoleon även i de hypotetiska världar där han *inte* tog makten i Frankrike (eller saknade andra egenskaper som vi förknippar men honom). Detta är nu inte möjligt ifall namn är bestämda beskrivningar; i sådana fall fungerar ju inte Napoleons bestämda beskrivning ("den man som regerade Frankrike under tidigt 1800-tal...") i vår hypotetiska diskussion då den inte skulle referera till Napoleon i denna tänkta värld (den skulle referera antingen till någon annan eller till ingen alls, lite beroende på hur specifik den bestämda beskrivningen är). Med bestämda beskrivningar skulle det således endast vara möjligt att referera till individer i *denna* värld.

### 2.5.2) Kausala kedjor

Kripke föreslår istället en teori där egennamns referens förklaras genom vad han kallar för *kausala kedjor* (orsakskedjor). På sätt och vis innebär detta synsätt en slags återgång till den naiva referensteori som beskrevs tidigare; ett namn fästs vid en bärare vid en dopceremoni, och detta fastställer referensen för all framtida användning av namnet. Detta beskrivs kärnfullt av Collin och Goldmann då de skriver:

Tanken är att namnet ursprungligen fästs vid bäraren genom en dopakt, som kan vara antingen formell eller informell. Vid ett formellt, kyrkligt dop, får det lilla barnet ett namn knutet till sig, till exempel "Sven Jansson", varpå de närvarande ger sig ut i samhället och sprider namnet genom att berätta om barnet och dopet och på så vis foga namnet "Sven Jansson" till sina åhörare repertoar. Dessa åhörare talar i sin tur med andra som också tillägnar sig namnet; till slut är namnet utbrett inom en större krets av talare som inte behöver känna till någon särskilt identifierande beskrivning av Sven Jansson, men för vilka namnet "Sven Jansson" likväl är knutet till Sven Jansson genom den orsakskedja (en kedja av kommunikation) som från talare till talare leder tillbaka tills den till sist når tillbaka till dopsituationen. En talare använder då namnet "Sven Jansson" om Sven Jansson förutsatt att han använder det med avsikten att använda det på samma sätt som de han lärt sig det av—och förutsatt att hans tillägnande av namnet "Sven Jansson" ingår i den kausala/historiska kedja som leder tillbaka till Sven Janssons dop. (Collin & Goldmann, 2000, s.71)

Skillnaden mot den naiva referensteorin är då att namnets mening inte identifieras med *bäraren*, utan med det *historiska spår* som leder från det ursprungliga dopet fram till talarens användning av namnet. På så vis blir detta spår en rigid designator och utgör en kausal kedja som pekar ut samma individ i alla möjliga världar där individen existerar.

## 3) Johnson-Lairds mentala modeller

Teorin om *mentala modeller* (*mental models*) är huvudsakligen utvecklad av Philip N. Johnson-Laird vid Princeton University. Den är sprungen ur studier av de mentala processer som ligger bakom logiskt resonerande eller *deduktion*. I korthet går den ut på att människor skapar "inre modeller" av sin omvärld, vad Johnson-Laird kallar för

*mentala modeller* (mental models), detta genom en mental process som han menar sker automatiskt och undermedvetet. Logiskt resonerande sker sedan genom manipulerande av dessa modeller. Dessa tankegångar är, enligt egen utsago, inspirerade av Kenneth Craik som 1943 föreslog just att tänkande utgörs av manipulering av inre representationer av omvärlden. Tyvärr röntes denna uppfattning tydligen inget större inflytande, enligt vad Johnson-Laird erfar i prologen till sin bok *Mental Models* (1983).

Enligt Johnson-Lairds synsätt (1983, kap.11) så konstrueras mentala modeller genom en så kallad *procedurell semantik*, en process som relaterar språket till en *mental modell* snarare än till omvärlden. Denna process specificerar det sätt på vilket de ingående orden i en sats bidrar till satsens sanningsvillkor. På detta vis erhålles en fullständig beskrivning av språkets semantiska egenskaper, menar han. Angående konstruktionen av mentala modeller skriver han (1983, s.407):

The comprehension of discourse depends on three principal levels of representation: first, there is a phonemic (or graphemic) representation that encodes the sounds (or letters) of an utterance; second, there is a propositional representation, which is close to the surface form of the utterance; third, there is a mental model. The mental model is constructed on the basis of the truth conditions of the propositions expressed by the sentence in the discourse.

Vad gäller *naturen* hos dessa mentala modeller så anser han att de kan innehålla enheter som korresponderar med objekt i omvärlden. Dessa enheters egenskaper och relationerna mellan dem motsvarar vår uppfattning av de sakernas tillstånd som den mentala modellen representerar. Han talar här om flera olika typer av mentala modeller. Några kortfattade exempel:

1. *Relationella* modeller: Består av ramverk innehållandes enheter som representerar fysiska objekt i omvärlden, samt representationer av deras egenskaper och relationer.
2. *Spatiala* modeller: Består av relationella modeller i vilka alla relationer är rumsliga. Representeras i två- eller tredimensionella miljöer.

3. *Temporala* modeller: Består av sekvenser av spatiala modeller (med samma dimensionalitet).
4. Bilder: Består av tvådimensionella representationer av spatiala modeller sedda ur ett visst perspektiv. På detta vis skapas alltså bilder av de objekt eller sakernas tillstånd som de underliggande modellerna representerar.

Johnson-Laird menar att det förvisso är många som i sina studier av språk och tankar har tagit fasta på medvetandets användning av mentala representationer (bilder och/eller symbolsträngar), men att alla dessa tycks ha missat kärnfrågan: vad det är som gör en mental entitet till en representation av någonting. En konsekvens av denna försummelse är att dessa teorier ofta saknar en redogörelse för den referentiella delen; de säger tydligen inte mycket om hur språk och tanke förhåller sig till omvärlden. Han menar även att de flesta författare gör sig skyldiga till en ofullständig analys av naturen hos de mentala representationerna, vilket visar sig i det faktum att många antar existensen av en inre, mental logik som antas ligga bakom vårt logiska resonerande. Teorier baserade på sådant formellt regelföljande antar att verbala premisser översätts till interna representationer av den logiska formen. Johnson-Laird anser att de teorier som hävdar detta är ofullständiga därför att de inte talar om hur denna antagna översättning fungerar, eller hur slutsatsen sedan översätts tillbaka till vanligt språk. Han anser således att dessa teorier är felaktiga, och att naturen hos våra deduktiva slutledningar snarare skall sökas i manipulerande av mentala modeller än hos ”inbyggda” logiska mekanismer. Den teori om mentala modeller som han föreslår utgår istället ifrån att logiska resonemang har sin grund i *förståelsen* av en situation. Det vill säga: utgångspunkten är en modell, konstruerad utifrån perception och/eller språkförståelse. Dessa mentala modeller bör alltså uppfattas som strukturellt identiska med de sakernas tillstånd som de representerar; och de kan förutom rent faktiska situationer även direkt representera sådana abstrakta tillstånd som tänkta alternativa situationer och negationer (1983, kap.9). Själva det logiska resonemanget utgörs sedan av en process som beskrivs på följande vis:

After the construction of models of the premises, the next step is to formulate a putative conclusion. Because the process is based on models of the premises, it naturally maintains their semantic information. No conclusion can be drawn in some cases unless an implicit model has been "fleshed out" explicitly. If there is a conclusion, the final step is to search for alternate models that might refute it. The conclusion is valid if there are no such counterexamples. (Johnson-Laird & Byrne, 1991, s.196)

Den för denna uppsats viktigaste delen av Johnson-Lairds teorier är dock den som handlar om hur mentala modeller är relaterade till språklig mening och deras relation till omvärlden. Denna del av teorin beskrivs bäst i vissa centrala kapitel i boken *Mental Models* (1983) som både tar upp (1) mentala modellers relation modellteoretisk semantik, och (2) deras förhållande till filosofen Hilary Putnams teorier. Jag kommer i den resterande delen av detta avsnitt att redogöra för (1) för att i nästa avsnitt behandla (2).

Johnson-Laird anser att det finns starka psykologiska skäl för att anta att mentala modeller är uppbyggda av en semantik som opererar på propositionella representationer av premisserna. Han menar att detta synsätt gör det kompatibelt med logikernas så kallade *modellteoretiska semantik* som primärt utvecklades för att specificera mening hos ideala språk, det vill säga matematik, predikatlogik och dylikt. Dessa språk har en mycket exakt definierad *syntax* som inte säger någonting om vad de ingående symbolerna *betyder*, dvs. vad de refererar till (eftersom en sådan specificering inte hör hemma i en syntaxdefinition). Det är först när man ger språket en *tolkning* som det får mening eftersom en sådan tolkning innebär att man specificerar vad symbolerna betyder, vad de är tänkta att referera till i omvärlden. Om en sådan tolkning är *sann*, det vill säga med verkligheten överensstämmande, så kallas den för en *modell* av språket. En modell av ett formellt språk är dock en abstrakt konstruktion, menar Johnson-Laird; den består av en funktion från syntaktiskt välformade uttryck till element i en specificerad modellstruktur, till exempel när egennamn tilldelas en tolkning bestående av korresponderande individer i modellen. Några viktiga komponenter för appliceringen av dessa formella metoder på naturligt språk ges dels av Alfred Tarskis sanningsbegrepp och dels den modellteoretiska

semantiken för modal logik som utvecklades av Hintikka och Kripke med flera. Enligt den sistnämnda är en proposition *nödvändig* om den är sann i *alla* möjliga världar och *möjlig* om den är sann i *några* möjliga världar. Detta synsätt möjliggör ett klargörande av distinktionen mellan intension och extension: ett predikats intension (egen-skapen som det fångar) kan behandlas som en funktion från mängden möjliga världar till mängden individer. Dess extension blir då den partikulära mängd individer på vilka detta appliceras i den aktuella världen. Detta innebär att en sats intension (den proposition som den uttrycker) kan behandlas som en funktion från mängden möjliga världar till sanningsvärden (sant/falskt). Dess extension blir då sanningsvärdet i den möjliga värld som studeras (1983, kap.8).

Synen på mening som en modellteoretisk tolkning är en idealisering eller, som Johnson-Laird uttrycker det, "an abstract theoretical fiction" (1983, s.182). Hans teori om mentala modeller är ju primärt skapad i syfte att redogöra för hur mening *repre-senteras mentalt*. En mer grundläggande fråga är den mer filosofiska angående vad mening egentligen *är*, och *var* man skall leta efter den. För att det över huvudtaget skall vara någon idé att tala om meningens mentala representation så måste man ju först visa att mening faktiskt är någonting mentalt, något som inte på något sätt är självklart. Johnson-Laird talar här om olika filosofiska inriktningar: antingen kan man ha uppfattningen att mening helt och hållet består av mentala konstruktioner som appliceras på omvärlden genom språket, eller så är man av den åsikten att språklig mening till största delen bestäms av omvärldens beskaffenhet. Den sistnämnda ståndpunkten brukar, som nämnts, gå under benämningen externalism. Detta leder nu fram till kärnpunkten i denna uppsats; hur Johnson-Laird ställer sig till denna ståndpunkt. Faktum är att han förhåller sig något undvikande på denna punkt då han hävdar att han inte har några pretentioner att lösa denna dispyt, samt att hans syfte är något annorlunda, nämligen: "to resolve the controversy in so far as it affects the psychology of meaning" (1983, s.182). Låt oss se vad han menar med detta.

## 4) Mentala modeller och externalismen

En viktig fråga i sammanhanget är om det går att specificera nödvändiga och tillräckliga villkor för meningen hos de begrepp som våra ord representerar. Uppenbart är att vi både i språk och tankar delar in världen i kategorier. Traditionellt har man inom filosofi och psykologi ansett att objekten inom en viss kategori nödvändigtvis har en eller flera egenskaper gemensamt som rättfärdigar indelningen, och att specifikation av dessa egenskaper är tillräckligt för att definiera kategorin i fråga. Johnson-Laird nämner som ett exempel psykologen Clark Hull som 1920 skrev att meningen hos ordet det engelska ordet "dog" är "a characteristic more or less common to all dogs and not common to cats, dolls, and teddy bears" (Hull, 1920). En sådan uppfattning är baserad på antagandet att enkla begrepp är definierbara som en mängd nödvändiga och tillräckliga villkor, och att sammansatta begrepp är definierbara som funktioner av de enkla begreppen (Johnson-Laird, 1983, s.186).

### 4.1) Putnams externalism

I essän "The Meaning of 'Meaning'" (1975) analyserar filosofen Hilary Putnam dessa frågor och utvecklar en så kallad *externalistisk* teori, det vill säga ett synsätt som går ut på att en stor del av den språkliga meningen bestäms av *omvärldens* beskaffenhet. Han utgår från den av Frege influerade uppfattningen att mening är detsamma som intension, och att intensionen bestämmer extensionen. Han menar att denna syn på mening som intension traditionellt har implicerat att mening är någonting *mentalt*; att känna till ett ords mening är detsamma som att befinna sig i ett visst *psykologiskt tillstånd* (psychological state). Det vill säga; olika ord med olika mening motsvaras av olika psykologiska tillstånd. Kärnan i Putnams resonemang är nu att detta är *oförenligt* med uppfattningen att ett ords mening (dess intension) bestämmer dess extension. För att visa detta så konstruerar han ett argument som grundar sig i hans okunskap om trädarter. Han påstår sig inte känna till skillnaden mellan en bok och en alm, vilket rimligtvis borde innebära att han befinner sig i samma psykologiska tillstånd när han tänker på en alm som när han tänker på en bok, menar han. Dessa två ord har alltså samma mening *för honom*, vilket har till följd att *även deras*

*extensioner blir desamma* (om nu mening är intension och bestämmer extension). Men så är ju inte fallet; de två orden har ju i själva verket helt olika extensioner då det första refererar till alla boken i omvärlden och det andra till alla almar. Detta resonemang visar nu att uppfattningen att mening bör identifieras med psykologiska tillstånd är *oförenlig* med synsättet att ett ords mening bestämmer dess extension.

Detta leder Putnam till en diskussion som anknyter till de nödvändiga/tillräckliga villkor som vi talade om innan, en diskussion som sammanfattas bra av Johnson-Laird (1983, kap.9) som använder ordet "citron" som exempel: man skulle kunna tänka sig att meningen hos detta ord skulle kunna delas upp i sådana egenskaper som "sur", "gul", "har skal" osv. Ingen av dessa egenskaper är dock nödvändiga för att citronen skall vara en citron (en grön citron är fortfarande en citron, et cetera). Man kan därför inte ge en meningsdefinition av "citron" eftersom syftet med en sådan definition just är att specificera nödvändiga sanningar. Putnam menar att karakteristiken istället måste förklaras via en teori (vetenskaplig eller förvetenskaplig) om citroners *essentiella natur*. Med detta menar han att det är själva *beskaffenheten* hos det utpekade (det som teorin *förhoppningsvis* fångar) som spelar stor roll i bestämmandet av ordets mening, inte vår eventuella *kännedom* om denna natur. Ord med sådana drag kallar han för *natural kind terms* (hädanefter *NK-ord*). Putnam menar att man när man använder dessa ord till vardags så gör man det utifrån en förenklad teori om (exempelvis) citroners natur, ofta genom en beskrivning av en typisk citron (en normal medlem av "klassen Citron"). På så vis erhålles en *stereotyp* som hjälper åhöraren att identifiera objektet i fråga. Denna stereotyp är mental och utgör *en* del av meningen hos ordet "citron", men det är viktigt att komma ihåg att den inte specificerar den motsvarande *extensionen* (som utgör den *andra* delen av ordets mening). Denna bestäms istället av experter som kan hänvisa till provexemplar, utföra tester (utifrån teorier) med mera. Putnams poängen är här att språkanvändaren då inte *själv* behöver besitta den nödvändiga kunskapen utan vid behov kan hänvisa till experter i ämnet. På så vis kan bland annat det tidigare beskrivna "bok/alm-problemet" förklaras: orden *har* olika mening *även* om talaren inte känner till dem. Det spelar ingen roll att han inte känner till skillnaden mellan dem eftersom han i sin språkan-

vändning (implicit) refererar till experter med kunskap om hur de korrekta extensionerna skall specificeras. Detta är kärnan i Putnams teori om så kallad *lingvistisk arbetsfördelning* (division of linguistic labour).

Hur bestäms då den ”korrekta” extensionen för NK-orden? Vad avgör denna del av deras mening? Tanken är ju att experterna skall kunna garantera ackuratesen hos extensionsbeskrivningen. Putnam menar nu att detta inte går att åstadkomma med hjälp av nödvändiga/tillräckliga villkor. Istället måste det lösas på samma sätt som Kripke löser problemet med egennamnens mening, nämligen genom användning av *rigida designatorer*. Han förespråkar alltså att experternas ord får sin extension fastställd genom den ursprungliga ostensiva definitionen, det vill säga en definition som får sin kraft av de essentiella egenskaperna hos det objekt som ursprungligen pekades ut. (På samma vis som egennamn i Kripkes teori får sin mening genom en kausal kedja från dopet fram till användningen.) Detta gäller alldeles *oavsett* om den ursprungliga språkanvändaren kände till objektets essentiella egenskaper eller ej, och alltså även ifall det skulle råka vara så att experterna ”har fel”. Putnam menar alltså att ett NK-ord är en rigid designator, det vill säga det pekar ut samma sak i alla möjliga världar där det existerar, precis som Kripkes egennamn. Språkanvändarnas (även experternas) vetenskapliga kännedom om det utpekade blir då irrelevant i sammanhanget; det är den *faktiska naturen* hos det initialt utpekade som fastställer extensionen. Detta tillåter Putnam att hävda att samma stereotyp kan associeras med ord som har olika extensioner (som i bok/alm-exemplet). Hans resonemang är som följer: extensionen specificeras av intensionen (per definition). Stereotypen bestämmer inte extensionen (vilket visas av att samma stereotyp kan associeras med olika intensioner). Alltså kan inte stereotyp vara detsamma som intension (eftersom intensionen bestämmer extensionen). Slutsatsen blir då att intensionen inte kan vara ett mentalt objekt.

Som en sammanfattning på Putnams resonemang passar det bra med ett citat ur Donald Davidsons text ”Knowing one’s own mind” (1987, s.46):

Putnam argues persuasively that what words mean depend on more than ‘what is in the head’. He tells a number of stories the moral of which is that aspects of the natural history of

how someone learned the use of a word necessarily make a difference to what a word means. It seems to follow that two people might be in physically identical states and yet mean different things by the same words.

Eller, som Putnam själv uttrycker det: ”meanings’ just ain’t in the *head!*” (Putnam, 1975).

## 4.2) Johnson-Lairds synsätt

Johnson-Lairds syn på saken är att NK-ord som ”vatten” från början mycket väl kan ha introducerats utan att ursprungsanvändaren kände till den essentiella (underliggande) naturen. I sådana fall var inte heller intensionen känd (eftersom extensionen inte var det, och intension *per definition* specificeras av extension). Vanliga språk-användare har ännu idag inte denna kunskap; man använder sig istället av *ytlig likhet* för kategorisering, och genom den lingvistiska arbetsfördelningen hänvisar man till experter i de fall då man själv saknar kunskap (som i fallet med Putnams okunnighet om skillnaden mellan almar och bokar). Dessa experter kan bidra med de ”korrekta” identifikationskriterierna och *i bästa fall* pekar dom ut den korrekta extensionen. Detta är dock inget vi kan *veta säkert* eftersom vi inte kan vara säkra på att våra teorier är korrekta. Kontentan blir att man inte kan *garantera* att det är ”rätt” extension som pekas ut (det var just denna insikt som fick Putnam att vilja använda sig av kausala kedjor). Johnson-Lairds åsikt är också att det inte spelar någon roll om experterna pekar ut ”fel” extension. För att påvisa detta spekulerar han kring vad som skulle hända ifall det uppdagades att det ”egentligen” inte finns några NK-ord. Det vill säga att de objekt som vi kategoriserar under samma etikett egentligen *inte har* någon gemensam, underliggande natur? Vad skulle hända då? En sak är i alla fall säker, menar han: en dylik upptäckt skulle *inte* få folk att sluta använda sådana termer *som om* de hade någonting gemensamt. Han skriver angående detta:

Perhaps the intentions of those who originally coined the term would have been thwarted, but its usefulness would continue as long as there were bits of stuff around that possessed the requisite superficial characteristics. (Johnson-Laird, 1983, s.194)

Ovanstående scenario skulle endast innebära att det inte skulle gå att specificera extensionen mer exakt än vad som är möjligt med kriterier baserade på *ytliga egenskaper*. Det vill säga, inte ens experterna skulle kunna ge en närmare definition av vatten än att det är en transparent, smaklös vätska som är törstsläckande, et cetera. Han åsikt är nu att en dylik oförmåga att specificera extensionen är irrelevant eftersom vi ändå aldrig kan veta ifall vi använder oss av "rätt" extensionspecification (det kan vara så att våra vetenskapliga teorier är inkorrekta då de aldrig går att verifiera, endast falsifiera). Han anser dock att nyttan av att använda sig av en förklaringsmodell uppbyggd utifrån kausala kedjor är tveksam, och är därför skeptisk till Putnams konstruering av en meningsteori utifrån Kripkes teori för egennamn. Han anser att den sistnämnda mycket väl kan vara korrekt; det är mycket möjligt att egennamn får sin mening genom en kausal länk från dopet fram till användningen. Men vår kategorisering verkar fungera efter andra principer, menar han.

Ett skäl för hans förkastande av kausala kedjor utgörs av insikten att NK-ord kan ändra sin mening (och därmed sin extension) med tiden. Ett exempel som han anför (s.194) är ordet "cattle" som med tiden har ändrat sin mening enligt:

1. property
2. moveable property
3. livestock
4. bovine animals

Detta visar, menar han, att det finns ord som *idag* används som NK-ord men som *ursprungligen* hade en helt annan betydelse. "Cattle" myntades ursprungligen med ett annat fenomen än dagens "nötkreatur" i åtanke, men fick sin mening förändrad i och med att betydelsen med tiden mer och mer gled över till ett ord med NK-struktur. Alltså har inte NK-ord oföränderliga extensioner, menar han. Dessa kan förändras med tiden. Han anser då att detta pekar på att det är ett misstag att använda sig av kausala kedjor som förklaring av meningen hos NK-ord. Som nämntes så anser han att dessa termers "sanna" intension/extension inte är speciellt intressant i samman-

hanget. Antingen finns den i experternas huvuden (något man aldrig kan veta säkert) eller så känner ingen till den. Om detta skriver han (s.195):

All that is necessary to acknowledge is that some words are used to refer to entities without anyone knowing for certain what the essential nature of these entities is, or indeed whether they have one. If the extension of word is fixed ostensively or by reference to a sample, then a plausible empirical hypothesis is that each natural kind term refers to a class whose members bear a superficial perceptual resemblance to one another. The class must generally be sufficiently homogeneous to merit baptism, though the speaker introducing the term may intend to designate a class sharing some as yet unknown underlying structure. If this structure remains unknown, then the only factor likely to maintain the use of a term is superficial resemblance.

Sammanfattningsvis så anser han alltså att kännedom om den (eventuellt) gemensamma underliggande strukturen hos de entiteter som ett NK-ord refererar till är irrelevant. Han verkar mena att det enda som spelar roll för vår kategorisering är den ytliga likheten: *ett NK-ord refererar till en klass som endast har yttlig likhet som medlemskapskriterie.*

En annan av Johnson-Lairds poänger är att meningens natur skiljer sig mellan olika slags ord; "intensions are not uniform" som han uttrycker det (s.195). Bara inom gruppen substantiv finns det olika slags termer vars meningsspecifikation skiljer sig åt. Vissa substantiv är till exempel analytiska (som t ex. "ungkarl"), medan andra (som exempelvis "äpple", "silver", "vatten" osv.) är NK-ord. Som vi har sett så anser han att de sistnämndas *riktiga* intensioner är okända (eller rättare sagt: vi kan inte *veta* ifall någon känner till dem eller ej), men detta anser han inte spela någon roll då deras mentala representationer utgörs av mentala modeller som korresponderar till klasser som hålls ihop av yttlig likhet. Vidare finns andra slags substantiv som t.ex. "stol", "melodi" vars mening specificeras av vad han kallar för en *konstruktiv semantik* (a constructive semantics) (s.196). Med detta menar han att de har intensioner som består av *mentala konstruktioner* som påläggs omvärlden, och fastän dessa ord mycket väl kan referera till verkliga objekt så har dessa objekt ingen gemensam underliggande struktur som rättfärdigar deras medlemskap i kategorin. Deras mening

vilar helt på mänskliga konventioner och kan förändras med tiden. Här behövs därför ingen hänvisning till experter eller ostensiva definitioner över huvudtaget (s.196).

Vad drar han då för slutsatser av allt detta? Finns meningen i medvetandet eller i omvärlden? Hans slutsats blir att detta verkar vara en falsk antites. Meningen hos *vissa* ord utgörs av mentala konstruktioner som påläggs omvärlden, medan andra ord används till att referera till klasser av objekt vars underliggande struktur är okänd (men som kan uppvisa en yttlig likhet). Sådana ord refererar till någonting konkret, men deras intension är *i bästa fall* endast känd av ett fåtal experter (denna eventuella kännedom spelar dock ingen roll i sammanhanget). Åter andra ord har en semantik som är stipulerad utifrån nödvändiga/tillräckliga villkor som stöd för analytiska härledningar. Och så vidare. Den grundläggande åsikten är alltså att den här diskuterade dispyten rörande var språklig mening står att finna är misstagen; även om vi skulle anta en externalistisk semantik så skulle våra ord likväl ha en associerad mental modell, menar han (s.203-4).

### 4.3) Diskussion

Fungerar då Johnson-Lairds lösning? Kan hans teori lösa de problem som Putnams resonemang skapar? Låt oss igen se på ordet "vatten". När man yttrar detta ord idag så refererar man till samma extension som de som använde ordet på medeltiden gjorde, trots att vi idag känner till både modern atomteori och kemi. Putnam skulle här säga att ordet alltid refererar till samma extension via den kausala kedja som leder från den ostensiva definitionen fram till användningen av ordet. Johnson-Lairds skulle förmodligen dra en liknande slutsats, men hans förklaring skulle nog istället gå ut på att den psykologiska meningen är densamma idag som den var på medeltiden eftersom vi fortfarande klassificerar vatten utifrån samma ytliga egenskaper som de gjorde då. Förvisso skulle medeltidsfolkets uppfattning av ordets intension/extension vara annorlunda än vår, men eftersom vi aldrig kan veta vilken intension/extension som är den korrekta så spelar inte detta någon roll för den psykologiska meningen,

skulle han hävda. Detta indikerar en intresseskillnad i jämförelse med Putnam: för Putnam är extensionen en väldigt viktig del av meningsbegreppet. Olika extensioner ger upphov till olika meningar, och omvärldens beskaffenhet blir då väldigt viktig. Johnson-Laird, å andra sidan, är helt fokuserad på den psykologiska meningen (ungefär det som Putnam kallar för "stereotyper" och som han endast anser utgöra en *del* av meningsbegreppet som helhet). Denna insikt verkar vara pudelns kärna: *de är inte intresserade av riktigt samma typ av mening och talar om delvis olika saker när de diskuterar mening*. Man skulle kunna säga att Johnson-Laird mest är intresserad av den "inre" meningen medan Putnam ser "yttre" mening som mycket viktig.

#### 4.3.1) *Narrow och wide*

Denna distinktion mellan inre och yttre brukar gå under benämningarna *narrow* respektive *wide*. En bra definition ges av Putnam (1975) där han talar om att psykologiska tillstånd är *narrow* om de inte förutsätter existensen av någonting annat än subjektet självt. Den psykologiska mening som Johnson-Laird talar om representerar just sådana psykologiska tillstånd som är *narrow*. Till skillnad från Putnam bryr han sig inte om ordens extension, och detta indikerar som sagt en skillnad i intresseområden mellan de två: Putnam är inriktad på att ge en fullständig redogörelse för meningsbegreppet (vilket inkluderar både stereotyp och extension) medan Johnson-Laird mest vill undersöka den psykologiska meningen (den som är *narrow*). Vi har här två olika sätt att se på saken, och min bedömning är att bägge inriktningarna har sitt berättigande. Vilken man intresserar sig för beror på vad man önskar studera, precis som fallet är med Johnson-Laird och Putnam. För att ta ett annat exempel: om man vill studera *språklig kommunikation* så verkar det vara mening i betydelsen *narrow* som man skall fokusera på. Detta följer av teorierna för språkanvändning som jag redogjorde för tidigare: här sätter man ju fokus på *framgångsvillkoren* för ett yttrande. Om jag nu hävdar någonting om världen, vad är det då jag vill göra med mitt yttrande? Vad är det jag vill uppnå? Svaret är att jag vill att åhöraren skall ta till sig det jag säger och förstå vad jag menar. Vad innebär då detta? Om man accepterar Johnson-Lairds teorier så innebär det att jag har en mental modell som motsvarar det hävdande som jag vill framföra, det vill säga en modell av det förhållande ute i

världen som mitt hävdande utgör. *Syftet med mitt yttrande borde då vara att skapa en motsvarande modell hos åhöraren.* Om jag exempelvis hävdar att "Napoleon var kortvuxen" så motsvarar det en mental modell som jag vill att åhöraren skall överta; jag vill framkalla samma föreställning av en kortvuxen Napoleon som jag själv har. Alltså verkar det i detta sammanhang vara mest vettigt att studera Johnson-Lairds teorier om mening och hur dessa representeras i medvetandet. Är det å andra sidan en fullständig redogörelse för ordens mening som man är intresserad av så bör man istället vända sig till Putnam. Båda synsätten är viktiga eftersom de utvecklats som svar på olika frågor.

#### 4.3.2) *Slutdiskussion*

Jag skulle som avslutning vilja diskutera ett tredje alternativ. Istället för att som Putnam inkludera både inre stereotyper och yttre extensioner i ett gemensamt meningsbegrepp så anser jag att det vore lämpligare med en meningsteori som skiljer dessa åt. Detta eftersom jag anser att de inte bör vara delar av samma meningsbegrepp då de behandlar delvis olika fenomen utifrån olika frågeställningar. Som ett exempel på detta har vi ju Johnson-Lairds teorier som är inriktade på psykologisk mening utifrån en frågeställning rörande medvetandets representation av språklig mening. Detta visar för övrigt att det går att utveckla en fruktbar teori rörande en psykologisk mening, skild från de frågeställningar om den "verkliga" meningens natur som Putnam intresserar sig för. En sådan distinktion skulle kunna göras tydligare genom införandet av termerna  $mening_n$  och  $mening_w$  (där "n" står för "narrow" och "w" för "wide").  $Mening_n$  skulle då motsvara Johnson-Lairds intressesfär medan  $mening_w$  skulle utgöra de yttre fenomen som Putnam är intresserad av. En ledande tanke är då att ju närmare  $mening_n$  ligger  $mening_w$ , desto mer korrekt är vår uppfattning av omvärlden. Låt oss se hur detta föreslagna alternativ skulle hantera de problem som tagits upp:

Vad gäller fallet med "samma psykologiska tillstånd—olika extensioner" som motsvarades av bok/alm-situationen så skulle man kunna analysera detta som att samma  $mening_n$  passar in *både* på det  $mening_w$  som motsvarar alm-begreppet och på det  $mening_w$  som motsvarar bok-begreppet. Detta eftersom  $mening_n$  representerar ett psykologiskt tillstånd där man har otillräcklig kunskap om skillnaderna mellan

bokar och almar. Kriterierna blir därför så pass vaga att de är applicerbara på både bokar och almar. Det verkar alltså som att denna situation skulle gå att hantera med mitt föreslagna synsätt. Hur är det då med motsatsen; ”samma extension—olika psykologiska tillstånd”, som illustrerades av det faktum att vi talar om samma vatten idag som de gjorde på medeltiden? Här skulle jag nog vilja säga att medeltidsfolkets mening<sub>n</sub> motsvarar ett begrepp vars kriterier endast inkluderar vattnets ytliga egenskaper (eftersom de inte hade kännedom om mikrostrukturen hos vatten). Det är därför applicerbart på vattens mening<sub>w</sub> (likväl som på alla andra transparenta, smaklösa (etc.) vätskor). Vårt nutida mening<sub>n</sub> för vatten däremot, skulle inkludera *både* dessa ytliga egenskaper *plus* kännedom om atomstrukturen (och via våra experter, förmåga att utföra tester för att avgöra denna struktur). Detta är ett mer detaljerat vattenbegrepp än det som man hade på medeltiden och därför är det bara applicerbart på mening<sub>w</sub> för vatten (givet att våra vetenskapliga teorier är korrekta). Båda vattenbegreppen kan dock appliceras på vattenextensionen.

Vid en första anblick verkar det alltså som att den teori som jag har föreslagit kan hantera de problem som uppkommer. Det skall dock endast ses som en snabb skiss av en teori som kan vara värd att utveckla, och det kan hända att den inte fungerar fullt ut.

## 5) Slutsats

Vad blir då kontentan av allt detta? Mitt syfte var att undersöka hur Johnson-Laird hanterat externalismen och att se ifall de två synsätten är förenliga. Det sägs redan från början att Johnson-Lairds teorier är tänkta att behandla psykologisk mening, och detta visar sig stämma. Han kommer dock in på språkfilosofiska frågor, och efter att ha analyserat Putnams externalistiska resonemang kommer han fram till resultatet att det finns olika slags mening för olika typer av ord; ”intensions are not uniform”. Vissa ord, närmare bestämt de som går under benämningen ”natural kind terms”, har en mening som är okänd för oss i den bemärkelsen att vi inte kan veta ifall vi känner till den eller ej. Putnam anser att en viktig del av meningsbegreppet för dylika ord utgörs av en historisk orsakskedja från de som först myntade uttrycken fram till

användandet av dem. Johnson-Laird håller inte riktigt med om detta då han anser att det motsägs av det faktum att dylika ord kan ändra mening allt eftersom. Han vill även understryka att andra typer av ord har en annan semantik, och att det inte är så enkelt som att hävda att "meanings ain't in the head". Hans generella slutsats om dessa frågor är att hela dispyten är misstagen. Han verkar inte se några större motsättningar mellan uppfattningen att mening dikteras av omvärlden och synsättet att den representeras psykologiskt av mentala entiteter (som exempelvis mentala modeller). Hans synsätt är alltså av allt att döma förenligt med Putnams externalism (även om han gärna vill förkasta Putnams användande av kausala kedjor). En annan slutsats som jag drar är att Johnson-Laird och Putnam har olika intresseområden; mening ur ett psykologiskt perspektiv respektive ett abstrakt, språkfilosofiskt meningsbegrepp. Detta ligger till grund för att de delvis talar om olika saker när de diskuterar mening. Dock verkar båda ståndpunkterna som sagt vara i stort sett förenliga, och de bör ses som lika viktiga då de utvecklats som svar på olika frågor.

## 6) Källor

Berkeley, G. (1710/1946) *A treatise concerning the principles of knowledge*, Thomas J. McCormack (red.), La Salle: Open Court Publishing Company.

Collin, F. & Guldman, F. (2000) *Språkfilosofi. En introduktion.*, Nora: Bokförlaget Nya Doxa.

Davidson, D. (1987/2001) "Knowing one's own mind" i Quassim Cassam (red.), *Self-knowledge* (s.43-64). Oxford: Oxford University Press

Frege, G. (1892) "Über Sinn und Bedeutung" Första publicering i *Zeitschrift für Philosophie und philosophische Kritik*, vol.100 (1892), s. 25-50.

Hull, C. L. (1920) "Quantitative aspects of the evolution of concepts" i *Psychological Monographs*, 28 (1920).

Johnson-Laird, P. N. (1983) *Mental models*, Cambridge: Cambridge University Press.

Johnson-Laird, P. N. & Byrne, R. M. J. (1991) *Deduction*, Hove: Lawrence Erlbaum Associates Ltd., Publishers.

Kripke, S. A. (1980) *Naming and Necessity*, Oxford: Blackwell Publishers.

Locke, J. (1689/1975) *An Essay concerning Human Understanding*, Oxford: Oxford University Press.

Putnam, Hilary (1975/2002) "The meaning of 'meaning'" i Chalmers D. J. (red.), *Philosophy of mind: classical and contemporary readings* (s.581-96), Oxford: Oxford University Press.

Russell, B. (1905/1956) "On denoting" i Marsh R. C. (red.), *Logic and knowledge : essays 1901—1950* (s-39-56), London: Routledge.