

Lunds universitet
Sociologiska institutionen

Integration av artificiella agenter i mänskliga organisationer

**en teoretisk analys
av de hybrida systemens problematik**

Simon Flyman

Uppsats Soc 446, 41-80 p
VT 2001
Handledare: José F. Pacheco

ABSTRACT

Författare: Simon Flyman

Titel: Integration av artificiella agenter i mänskliga organisationer
– en teoretisk analys av de hybrida systemens problematik

Uppsats Soc 446 41-80 p

Handledare: José F. Pacheco

Sociologiska institutionen, vårterminen 2001

Problem/Bakgrund: Utvecklingen av distribuerad artificiell intelligens och sofistikerad agentteknik skapar nya utmaningar för sociologin. I Tyskland har denna utmaning antagits av en grupp sociologer som grundat ett nytt forskningsfält kallat *Sozionik*. En riktning inom detta fält sysselsätter sig med frågor kopplade till de hybrida systemen, d v s de sociotekniska system som uppstår då agentsystem införs i mänskliga kontexter. Med utgångspunkt i denna riktning avhandlas i uppsatsen möjligheterna att integrera artificiella agenter i mänskliga (och i regel komplexa) organisationer. På mikronivå uppkommer då frågor kring människors acceptans av artificiella agenter som kooperationspartners. På makronivå är frågan hur roll- och funktionsfördelningen mellan mänskliga aktörer och artificiella agenter ser ut i organisationer. Slutligen uppkommer också nya problem kring risker och ansvarsfördelning i de hybrida systemen.

Syfte: Syftet med uppsatsen är att undersöka de sociologiska förutsättningarna för integration av artificiella agenter och agentsystem i mänskliga organisationer. Tanken är att skapa ett teoretiskt underlag till integrationsproblematiken genom att jämföra och analysera olika sociologiska angreppssätt.

Slutsatser: Genom att anlägga ett pragmatiskt perspektiv på integrationsprocessen är det möjligt att fånga in dess sociala och kontextuella aspekter. Studier har visat att datorers roll och funktion i mänskliga system kan betraktas som en följd av interaktionsprocesser, där den mänskliga användarens tolkning och upplevelse av datorns egenskaper står i centrum. På makronivå visar nyare ansatser inom organisations- och tekniksociologi att tekniska artefakter har viktiga integrativa funktioner att fylla och därmed tillmäts agenter en potential att bli en integrerad aktör i mänskliga organisationer. Vad det gäller ansvarsfördelningen mellan artificiella agenter och mänskliga aktörer inom organisationer är det idag svårt att ge några klara svar. På detta område krävs sannolikt framtida forskningsinsatser.

Nyckelord: aktör-nätverk teori, artificiella agenter, tekniksociologi
distribuerad artificiell intelligens, socionik

Förord

Arbetet med denna uppsats har till stora delar varit en isolerad sysselsättning. Delvis beror det på att jag suttit på annan ort och delvis beror det på att jag i mångt och mycket rört mig på, åtminstone för svensk sociologi, okänd terräng. Resan har pendlat mellan å ena sidan osäkerhet och förvirring och å andra sidan nyfikenhet och spännande läsupplevelser. Helt ensam har jag dock inte varit och därför vill jag ta tillfället i akt att framhålla ett par människor som bidragit till att jag kunnat genomföra mitt uppsatsarbete på distans.

Först vill jag rikta några tacksamhetens ord till ”socionikerna” Jörg Strübing och Ingo Schulz-Schaeffer för navigeringshjälp, i form av litteraturtips och allmänna råd, inom ett nytt och okänt fält.

Vidare vill jag också tacka min handledare José Pacheco som bidragit med värdefulla tips och konstruktiv kritik samt på kort varsel avsatt tid för mig under de få visiter jag gjort i Sverige under arbetets gång.

Tack också till Jenny Hansson och Henning Janson för korrekturläsning och kritik.

Slutligen vill jag tacka mina föräldrar, utan vars stöd och uppmuntran det inte hade varit möjligt för mig att genomföra mina studier och denna uppsats.

Simon Flyman

Berlin, maj 2001

Innehållsförteckning

1. Inledning.....	6
1.1 Problemformulering.....	7
1.2 Syfte.....	8
1.3 Frågeställningar	9
1.4 Disposition.....	9
2. Bakgrund och forskningsläge.....	11
2.1 Distribuerad artificiell intelligens – en översikt	11
2.1.1 Historia och utveckling.....	11
2.1.2 Uppdelning och avgränsning.....	13
2.1.3 Nyckelbegrepp.....	13
2.1.4 Användningsområden.....	17
2.2 Socionik	18
2.2.1 Forskningsfältets framväxt	18
2.2.2 Forskningsperspektiv och inriktningar	19
2.2.2.1 Den distribuerade artificiella intelligensens sociologi	21
2.2.2.2 Datorsimulation som sociologisk metod	22
2.2.2.3 Hybrida system.....	24
2.2.3 Teoretiska problemområden.....	25
3. Agenten som interaktionspartner – ett mikroperspektiv	30
3.1 De artificiella agenternas handlingsförmåga	30
3.1.1 Det sociologiska handlingsbegreppet	30
3.1.2 Handlingsförmåga hos människor och maskiner	32
3.1.3 Agentmodeller och handlingsförmåga	33
3.2 Social acceptans av artificiella agenter.....	34
3.3 Interaktion mellan mänskliga aktörer och artificiella agenter	36
3.3.1 Det sociologiska interaktionsbegreppet.....	37
3.3.2 Artificiella agenter och interaktion.....	38

4. Integration i komplexa organisationer – ett makroperspektiv	40
4.1 Organisationer och komplexitet.....	40
4.2 Artificiella agents roll i mänskliga organisationer	44
4.2.1 Agenter som enheter i komplexa sociotekniska system	44
4.2.2 Tekniska system som sociala konstruktioner	46
4.2.3 Agenter som aktör-nätverk	47
4.2.4 Hybrida system – ett organisatoriskt Cyberspace?.....	50
5. En metodisk reflektion kring integrationsproblematiken.....	53
5.1 Turing-testet.....	53
5.2 Durkheim-testet	54
5.3 Testernas tillämpningsmöjligheter.....	55
6. Diskussion.....	57
6.1 Risker och ansvarsfördelning i hybrida system	57
6.2 Integrationsvägar – teoriernas potential och begränsningar	59
6.2.1 Handlingsbaserad integration	60
6.2.2 Interaktionsbaserad integration	61
6.2.3 Aktör-nätverksteorin – en syntes	62
6.2.4 Agentmodellernas integrationspotential.....	64
6.3 Agentteknikens samhälleliga konsekvenser – en sociologisk utblick	64
Litteratur	68

Bilaga – Ordlista

1 Inledning

Listan på saker som är möjliga att artificiellt återskapa blir allt längre. Material, substanser, mat, ja till och med livet själv verkar vara möjligt att fabricera. På det senare området blev framställningen av fåret Dolly på konstgjord väg något av en vattendelare och framstår idag som ett epokgörande genombrott för gentekniken. Mer eller mindre lyckade försök att återskapa intelligens på konstgjord väg går att spåra årtionden tillbaka i tiden. Denna trend av artificiellt återskapande och nyskapande verkar nu även vara på väg att inbegripa den sociala världen och därmed sociologin på ett högst markant sätt.

Utvecklingstendenser på den artificiella intelligensens område rör sig mot en mer fördelad och samarbetsorienterad problemlösningsprocess. Målet är att skapa konstgjorda sociala system och därmed kanske också i förlängningen artificiella samhällen. Det kan till en början låta som science fiction och så har det hittills också mestadels behandlats från sociologins håll. Det som är skrivet om artificiella sociala världar inom sociologin rör sig i regel i fotspåren av cyberpunkförfattare som William Gibson och Neil Stephenson. Ofta handlar det om ganska spekulativa och utopiska/dystopiska betraktelser om framtida cybersamhällen.

Faktum är dock att artificiella sociala kontexter redan nu är på väg att bli högsta realitet. Inom forskningsfältet distribuerad artificiell intelligens (DAI) håller man på att modellera datorsystem baserade på agentteknik där de nämnda agenterna modelleras på grundval av sociala mänskliga kompetenser och beteendemönster. Primitivare varianter av sådana multi-agent system (MAS) förekommer redan idag i organisationer och system av olika slag.

Den hittills förda diskussionen om de nämnda artificiella sociala systemen har mest handlat om dess tekniska möjligheter och begränsningar. Alldeles nyligen har dock även intresse från sociologiskt håll aviserats. Det rör sig än så länge om en relativt liten skara teknik- och organisationssociologer i Tyskland. De har samlats kring ett gemensamt intresse för denna utvecklingsgren av artificiell intelligens och döpt forskningsprojektet till socionik¹. Till skillnad från mycket av den tidigare teknikforskningen inom sociologin stannar inte intresset vid teknikens inverkan på och följderna för samhället och dess olika fält. Istället föreslås att sociologin bör ingripa betydligt tidigare i utvecklingen och samarbeta med informatiken redan

¹ Det tyska begreppet är Sozionik. Jag har tagit mig friheten att försvenska detta begrepp.

i modelleringsfasen. Som man kan utläsa av namnet socionik, är delprojekten också i regel tandemprojekt, med deltagare från såväl sociologin som informatiken.

1.1 Problemformulering

Socioniken skulle kunna sägas behandla två olika koordinations-/integrationsproblem. Först och främst handlar det om att med hjälp av sociologiska teorier och modeller skapa en fungerande koordination och interaktion de artificiella agenterna emellan. I en andra fas uppkommer problemet att implementera ett MAS i en mänsklig organisation, d v s integrera de artificiella agenterna i en mänsklig kontext. I det här arbetet har jag tänkt koncentrera mig på det senare området. Det som intresserar mig är alltså hur förutsättningarna att implementera system av samarbetande artificiella agenter i mänskliga organisationer ser ut.

Frågan är då hur de sociologiska förutsättningarna för integrerade agentsystem ser ut. I samband med de hybrida systemen uppkommer frågor som i vilken utsträckning de konstgjorda agenterna kommer att betraktas som jämlika aktörer och interaktionspartners av de mänskliga medlemmarna i de hybrida systemen. Kort sagt frågor kring social integration. Det gäller alltså att studera och teoretiskt grunda förutsättningarna för en fungerande integration inom de hybrida systemen.

Utbredningen av distribuerad artificiell intelligens och multi-agent system gör det troligtvis nödvändigt att sociologiskt definiera om såväl teknikbegreppet i sig som relationen mellan människa och teknik samt därmed också teknikens roll i sociala kontexter. Teknik och tekniska artefakter har länge betraktats som miljövariabler i mänskliga kontexter. De har m a o setts som utomstående krafter och därmed har man i regel sociologiskt sett gjort en strikt åtskillnad mellan den tekniska och den sociala sidan av ett sociotekniskt system (ex Grint & Woolgar, 1997:11ff).

Frågan är om det i fallet med intelligenta agenter är intressant och relevant att låta den strikta distinktionen mellan skapade – artificiella – och naturliga – mänskliga – agenter fortleva. Sedan några år tillbaka finns det vissa teoretiska riktningar som låter prefixet försvinna och enbart talar om agenter i egenskap av dess funktioner. Tekniskt sett är målet med

agenttekniken att integrera tekniken i sociala system och därmed skapa s k hybrida system, där ingen åtskillnad görs mellan människa och maskin i problemlösningsprocessen.

1.2 Syfte

Jag stötte på begreppet socionik för första gången våren 2000 under utbytesstudier i Berlin. Eftersom jag kombinerat mina sociologistudier med informatik fann jag det genast väldigt intressant. Jag har länge sökt en möjlighet att kombinera dessa båda – relativt olikartade – fält. Jag är också av den uppfattningen att det är viktigt för sociologin att aktivt intressera sig för den snabba och i förlängningen radikala utvecklingsprocessen på informations- och datorteknikens område. Vidare lockas jag av idén att teknikorienterade inriktningar inom sociologin inte enbart stannar vid ett kritiskt betraktande av tekniken och dess följdverkningar utan även vågar ge sig in på teknikens hemmaplan. Jag tror att sociologin mycket väl kan bidra på ett konstruktivt sätt till utvecklingen av distribuerad artificiell intelligens och dess integration i olika mänskliga kontexter. Dock är det viktigt att påpeka att det här inte enbart handlar om att utföra beställningsjobb åt informatiken där sociologer skapar artificiella sociala mekanismer för de nämnda multi-agent systemen. Den sociala världen och studiet av denna är för komplex och mångfaldig för att kunna reduceras till ett antal universella mekanismer. Socionikern kan aldrig bli en social ingenjör. Snarare tror jag på ett reflexivt, kontextmedvetet och kritiskt inflytande från sociologin i denna sfär.

Mitt huvudsyfte med denna uppsats är att skapa någon form av sociologisk plattform för utveckling av hybrida socio-tekniska system. Jag kommer m h a olika sociologiska teoribildningar se hur förutsättningarna ser ut för integration av artificiella agenter i mänskliga organisationer. Arbetet är tänkt att fungera som ett orienterande underlag, i form av en genomgång och jämförelse av olika möjliga sociologiska angreppssätt, för senare praktisk användning.

Uppsatsen som följer kommer att vara av teoretisk art. Jag har tänkt att reflektera kring de uppställda problemen och försöka hitta olika sociologiska angreppssätt. Vidare har jag tänkt ställa dessa olika teorier mot varandra och försöka se vilka konsekvenser, brister och begränsningar de har. Där det är möjligt kommer jag också att illustrera och klargöra framställningen med hjälp av exempel och resultat från andra näraliggande undersökningar.

1.3 Frågeställningar

Problemställningen är möjlig att bryta ned i ett antal frågeställningar som alla på något sätt är kopplade till integrations- och implementeringsprocesser i samband med skapandet av hybrida system. På mikronivå uppkommer frågor om huruvida en agent kan sägas vara kapabel att utföra mänskliga handlingar. Vidare är frågan om de mänskliga aktörerna accepterar artificiella agenter som jämställda interaktionspartners? På makronivå är den huvudsakliga frågan vilken roll eller funktion agenterna har i förhållande till människorna inom organisationen. I samband med denna diskussion uppkommer också frågan hur man skall handskas med de risk- och ansvarsfrågor som uppkommer inom hybrida organisationer/system när viss aktivitet överläts åt softwareagenter. Vad uppstår det för risker i organisationer där vissa mänskliga funktioner utförs av konstgjorda agenter? Vem skall ställas till svars när negativa effekter av organisatoriska processer uppkommer?

Utifrån ett sociologiskt perspektiv är naturligtvis även kopplingen i andra riktningen ytterst intressant. Här handlar det om vilka effekter sofistikerad agentteknik kan tänkas ha för framtidens organisationer och arbete och även för samhället i stort. Av utrymmesskäl kommer jag inte att kunna utvidga detta resonemang, men jag kommer ändå försöka att kommentera frågan i slutkapitlet.

1.4 Disposition

I kapitlet som följer försöker jag skapa en bild av hur de båda forskningsfälten distribuerad artificiell intelligens och socionik har växt fram och utvecklats. Kapitlet är tänkt att fungera som en introduktion till mitt studiefält och som en bakgrund till problemområdet. Vidare är min avsikt i bakgrundskapitlet att skissa upp debatterna och positionerna, d v s forskningsläget inom respektive fält samt också invända en del egen kritik. I de två följande kapitlen tänker jag avhandla min problemställning utifrån två skilda perspektiv. Allra först tänker jag utgå från ett mikroperspektiv där interaktionen mellan mänsklig aktör och artificiell agent diskuteras, framför allt i socialpsykologiska termer. Därefter tänker jag i kapitel 4 förflytta analysen upp på makronivå och med utgångspunkt i teknik- och organisationssociologiska termer betrakta de artificiella agenternas roll i organisationen.

I kapitel 5 återfinns en diskussion av möjliga metodiska angreppssätt. Jag kommer här främst att utgå från det klassiska Turing-testet. Detta test har vållat mycket debatt och en del alternativ har föreslagits. Dessa kommer att kommenteras och vidare diskuteras här en föreslagen anpassning av Turing-testet – det s k Durkheim-testet – till DAI-förhållanden. I det därpå följande kapitlet försöker jag knyta ihop diskussionen. Här finns det anledning att återkomma till frågeställningarna om risk- och ansvarsförhållanden i de hybrida systemen. I det avslutande kapitlet kommer också utrymme att ges för en kort utblick över möjliga utvecklingslinjer som kan tänkas följa i agentteknikens spår.

2 Bakgrund och forskningsläge

Forskningsområdet distribuerad artificiell intelligens (DAI) samt dess nybildade gren socionik fungerar i mångt och mycket som ett ramverk för min uppsats. Jag tror därför det är nödvändigt att ha en viss kännedom om dessa områden för att kunna följa det fortsatta resonemanget i uppsatsen. Min avsikt är att i detta bakgrundskapitel beskriva utvecklingen inom fälten, ringa in några av dess nyckelbegrepp och huvudområden samt problem och lösningsstrategier.

2.1 Distribuerad artificiell intelligens – en översikt

2.1.1 Historia och utveckling

Den artificiella intelligensens (AI) utveckling under senare år har av den tyske tekniksociologen Werner Rammert sammanfattats som en tredelad process. Den har genom nätverk och modem blivit mobil. Numera är det möjligt för intelligenta program att kommunicera och utbyta information med andra intelligenta program även då avståndet är av betydande art. Vidare har AI blivit medial. Med det menar han att datorsystem utnyttjas i allt högre grad som medium eller kommunikationsverktyg, mot att som tidigare nästan uteslutningen använts för stora beräknings- eller slutledningsprocesser. Slutligen går det också att urskilja en social utvecklingslinje, där den sociala interaktionen fungerar som utgångspunkt vid formgivningen av datorsystem och dess gränssnitt (Rammert, 1998:92). De här tendenserna har skapat utrymme för nya inriktningar inom vetenskapen om artificiell intelligens. En av de viktigaste av dessa är DAI.

Rötterna till den distribuerade artificiella intelligensen går att finna i USA för runt tjugo år sedan. Bakom dess framväxt kan man spåra en rad olika utvecklingslinjer liknande de Rammert beskriver. Delvis var de av teknisk art, såsom utvecklingen och utbredningen av kraftfulla, samverkande datorer och datornätverk. Viktigare ändå var kanske den (kunskaps-) teoretiska omorientering som uppkom inom vissa AI-kretsar. Framför allt var det diskussionerna om mänskliga problemlösningsprocesser och dess form som styrde in den artificiella intelligensen på nya vägar. Problemlösningsprocessens sociala komponent började uppmärksammas i allt högre grad (Martial, 1992:6; Bond, 1988).

Följden blev att socialteoretiskt tänkande fick ett allt större inflytande vid modellerings- och konstruktionsprocesser i samband med AI. Rammert nämner ett par punkter inom DAI där detta går att märka. Först av allt utgår man här från att kunskap och information distribueras, fördelas, till flera självständiga, intelligenta agenter. Vidare läggs en ökad vikt vid kommunikationens roll vid kunskapsutveckling. Också olika sociala strukturers påverkan på kunskapsutvecklingsprocessen framhävs (Rammert, 1998:93).

Framväxten av DAI kan betraktas som en kritik mot klassisk AI och dess syn på hur intelligens uppstår. Kritiken går ut på att kunskapsutveckling och -användning inte kan reduceras till individuella, kognitiva processer utan att den också har sociala kvaliteter. Kunskap uppkommer, omformuleras och utvecklas genom olika sociala processer. Med andra ord hävdar kritikerna att kunskap är socialt konstruerad (Malsch, 1998:12). Så istället för att, som inom den klassiska AI-forskningen, utgå från en ”omnipotent” problemlösare, en superhjärna, uppstod inom DAI nya modeller där problemen delades upp och delegerades ut till flera problemlösande agenter. Problemlösningsprocessen övergår från att vara individuell till att bli kollektiv (Strübing, 1998:61f).

En viktig förgrundsgestalt för denna nya riktning var Carl Hewitt. Han förespråkade en intelligensmodell baserad på ett samhälle av kommunicerande experter. Inspiration till modellen kommer från den vetenskapliga världen. Han använder *the scientific community*, forskarsamhället, som metafor. Vetenskapen har genom sin pluralism – det finns ingen central ”skiljedomare” som sätter upp kriterier för sanning – och parallellism – många forskare kan arbeta med samma eller liknande problem samtidigt – varit mycket lyckosam i hanterandet av komplexa problem. Översatt i DAI-termer framstår problemlösningsprocessen för Hewitt som ett mönster av meddelandeförmedling mellan agenter i ett agentsamhälle. På grundval av dessa idéer skapade Hewitt ett s k *actor-system* bestående av agenter, d v s programelement, sammankopplade genom beroendeförhållanden till andra agenter. Carl Hewitts idéer innebar en ny bas för problemlösningsprocessen, bestående av kommunicerande agenter istället för att som tidigare utgjorts av ett urval av alternativ inom ett viss beslutsrum (Hewitt, 1977; Kornfeld & Hewitt, 1980; Sundermeyer, 1993).

2.1.2 Uppdelning och avgränsning

Martial definierar ett DAI-system som ett system bestående av minst två agenter, som förfogar över individuell information och/eller autonomi och som också har någonting att bidra med, exempelvis kompetens till slutledning eller planering, i ett AI-system (Martial, 1992:6). I övrigt görs i regel en åtskillnad mellan två huvudområden inom DAI: distribuerad problemlösning (DPS) och multiagent system (MAS). Som framgår av namnet sysselsätter man sig inom DPS med hur problemlösningsprocessen kan uppdelas i subprocesser och sedan fördelas på ett antal olika moduler. Modulerna samarbetar i så motto att de delar med sig av sin kunskap och gemensamt bidrar i utvecklingen av en lösning. Arbetsgången rör sig uppifrån och ned i en hierarkisk organisation. Ett MAS är ett löst sammankopplat nätverk av autonoma, och ofta heterogena, agenter som samarbetar för att lösa problem som ligger utanför dess egen förmåga. Det handlar alltså om att koordinera parallella processer av kognitiv karaktär som utförs av löst sammankopplade program/agenter. Agenterna i ett MAS måste inte nödvändigtvis samarbeta utan även konkurrenssituationer kan uppstå i vissa lägen (Moulin och Chaib-Draa, 1996:4f; Bond, 1988:3; Rammert, 1998:97; Müller, 1993:22).

Det finns ett antal andra subdiscipliner inom AI som uppvisar likheter med DAI och som därför ibland bidrar till missförstånd. Konnektionismen är en sådan. Inom konnektionismen betraktar man också intelligens som ett fenomen som uppkommer genom sammankoppling av flera element (Malsch, 1998:31). Den stora skillnaden här är att DAI utgår från antagandet att de enskilda element/enheterna besitter intelligens medan man inom konnektionismen menar att intelligens uppstår först i och med att de olika (icke-intelligenta) enheterna kopplas samman. Ett annat närbesläktat område är parallell AI. Inom detta område diskuteras frågor som parallell datorarkitektur och algoritmer för AI. Det är framför allt tekniska frågor, ofta rörande prestandaproblem i AI-system, som står i centrum (Martial, 1992:7; Bond, 1988:3f).

2.1.3 Nyckelbegrepp

Det verkliga nyckelbegreppet inom DAI är agentbegreppet. En allmänt hållen, handlingsbaserad definition av vad en agent är erbjuds av Kiss: *"Agents are the 'active ingredients' in nature. They are the sources of changes of states, i.e., events. Action is the central characteristic of agents. They bring about change through action"* (Kiss, 1996:248). Utifrån ett informatiskt perspektiv kan man definiera agent som: *"an entity that may change*

state when it communicates with another agents. Complex agents that encapsulate a collection of cooperating agents may also change state due to hidden internal communications" (Pflüger, 1997:433). Här framhävs alltså agenter dynamik och förmåga till lärande. Moulin och Chaib-Draa väljer att definiera en agent som ett (mer eller mindre sofistikerat) kunskapsbaserat system (Moulin och Chaib-Draa, 1996). Hughes och Singh betraktar agenter som "*active, persistent (software) components that perceive, reason, act and communicate*" (Hughes och Singh, 1997:1). Som synes är det svårt att fånga in begreppet agent i en klar och allmängiltig definition. Intressantare är därför kanske att titta på vilka olika former och kategorier av agenter som förekommer i DAI-litteraturen. I ett MAS görs ofta en åtskillnad mellan artificiella agenter, softwaremoduler, och mänskliga agenter, användare. Artificiella agenter kan uppvisa olika grad av heterogenitet när det gäller områden som tillgängliga resurser och problemlösningsmetoder. Spännvidden är stor och varierar från ytterst primitiva stimuli-respons agenter till agenter med sociala och kognitiva egenskaper (Sundermeyer, 1993:28). Agenter kan alltså skilja sig åt på en rad plan. Man kan kategorisera agenter på grundval egenskaper som placering/situation i systemet, grad av autonomi, rationalitet, intelligens och social förmåga.

Det går att finna två grundläggande riktningar av agentmodellering i DAI-litteraturen. De är till viss del byggda på de ovan givna egenskaperna. Den första, reflektiva skolan förespråkar att man modellerar intentionella agenter, d v s agenter med en kognitiv komponent eller ett mentalt tillstånd bestående av kunskap, avsikter, mål. En sådan agent besitter explicit kunskap om sin omgivning och skapar på grundval av denna kunskap planer, enligt vilka den så småningom handlar. Till grund för agentens intentionalitet ligger det faktum att dessa agenter förfogar över ett språk byggt på symboler. Den andra, reaktiva skolan har ett betydligt enklare agentkoncept. En reaktiv agents beteende och handlande bygger på stimuli-respons scheman. Agentens handlande består endast av förprogrammerade reaktioner på förändringar i omgivningen eller på meddelanden från andra agenter (Sundermeyer, 1993:26; Moulin och Chaib-Draa, 1996:8).

Wooldridge och Jennings gör en liknande åtskillnad med begreppen *the weak notion* och *the stronger notion*. *The weak notion* utrustar agenterna med autonomi, social förmåga, reaktionsförmåga och *pro-activeness*. *The stronger notion* adderar därutöver ett mentalt tillstånd och "känslor" (Wooldridge och Jennings, 1995).

Samarbete agenterna emellan är ett av de främsta kännetecknen för DAI. Graden av samarbete är olika hög i olika system. Skalan sträcker sig från system med fullt ut samarbetsvilliga agenter (*fully cooperative agents*) till antagonistiska system där agenterna överhuvudtaget inte samarbetar utan t o m kan blockera varandras mål (Moulin och Chaib-Draa, 1996:23).

I vissa situationer kan det uppkomma konflikter i planeringsprocessen. Agenterna är i regel utrustade med olika förmågor och preferenser och kan därför ha olika mål som tenderar att kollidera med varandra. I sådana lägen tvingas agenterna till förhandling för att lösa de uppkomna konflikterna. Genom en förhandling försöker parterna lösa konflikten dels genom att modifiera de lokala agenternas planer och dels genom att försöka hitta andra situationer där interaktion är möjlig. Den mest utbredda förhandlingsmetoden är *the contract net protocol*, där problem och konflikter agenter emellan löses genom budgivning (Moulin och Chaib-Draa, 1996:24f; Müller, 1996:211f).

En annan viktig dimension handlar om hur man kan skapa sammanhållning, koherens, i ett MAS. Det gäller att försöka koordinera de olika agenternas handlande. Tre villkor måste vara uppfyllda för att man skall kunna tala om koherens i ett DAI-system. Först måste varje nödvändig del av det övergripande problemet täckas av åtminstone en nod/enhet i systemet. Vidare måste noderna interagera i sådan utsträckning att dess aktiviteter kan integreras i en slutlig lösning. Slutligen måste de båda första kriterierna gå att uppfylla med de tekniska resurser som finns tillgängliga i nätverket (Bond & Gasser, 1988:19).

Jennings definierar koordination som "*the process by which an agent reasons about its local actions and the (anticipated) actions of others to try and ensure the community acts in a coherent manner*" (Jennings, 1996:187). Koordination behövs för att kunna möta globala begränsningar, hantera beroenden mellan agents handlingar samt därför att ingen ensam agent besitter tillräcklig kompetens, information eller resurser för att kunna lösa ett helt problem. Koordinationsprocesserna i ett MAS bygger i regel på förpliktelser – löften om att handla på ett visst sätt – och konventioner – regler för att kontrollera förpliktelserna under förändrade betingelser. Dessa båda grundkomponenter är tänkta att bidra till förutsägbarhet och flexibilitet i DAI-systemen (Jennings, 1996:188ff).

Ett av de stora problemen inom DAI är hur koherens och koordination skapas i ett MAS utan centraliserad kontroll. Vad finns det för koordinations-/designstrategier att tillgå?

En viktig strategi går ut på att genom planering få agenterna att styra sina handlingar mot ett gemensamt mål. Planering i MAS återkommer i olika former. De flesta texter urskiljer dock två huvudformer. Den första av dessa, multi-agent planering, utser en agent till "huvudplanerare". Denne undersöker hur olika agents handlingar påverkar/stör varandra och försöker därefter utarbeta en plan för parallell handling. Den andra formen, distribuerad planering, innebär en uppdelning av planeringsprocessen på flera agenter. Varje agent, som förutsätts vara rationell, skapar en egen subplan. Utifrån subplanerna förhandlas sedan en övergripande, gemensam plan ut (Bond, 1988:21f; Moulin och Chaib-Draa, 1996:27f; Rich och Knight, 1991:434ff).

Alan Bond ger vidare en rad andra exempel på koordinationsstrategier. Som vi redan sett kan man införa problemlösningsrutiner som begränsar handlingsutrymmet i givna situationer. En annan väg går ut på att hålla alla agenter uppdaterade om alla händelser som bidrar till att förändra nätverkets tillstånd. Därigenom möjliggörs ett mer rationellt beslutsfattande från agenternas sida. Bond pekar också på vikten av att öka agenternas lokala förmåga genom att göra de medvetna om sin position i nätverket (Bond, 1988:21ff).

För att olika agenter skall kunna samarbeta och lösa gemensamma problem krävs det att de kan kommunicera med varandra. I MAS är det främst två olika modeller för informationsutbyte som används.

- Blackboard systems: Kommunikationen äger här rum via ett mellanled, en "svart tavla", bestående av ett gemensamt minne där meddelanden postas, läses och leder vidare till nya handlingar.

- Message-passing systems: En andra variant är ett kommunikationssystem där agenterna kan rikta kommunikationen och välja vilka agenter som skall fungera som mottagare (Rich & Knight, 1991:439).

Utöver dessa båda former finns det bl a en ny, ännu relativt outvecklad riktning, byggd på s k högnivå kommunikation. Forskning pågår här på teman som förståelse av naturligt språk och intentioner i kommunikationen. Utgångspunkten här är *the speech-act theory*. Sociala agenter uttrycker förfrågningar, löften och order i form av språkliga utsagor. Genom användandet av ett naturligt språk är det möjligt att resonera kring och påverka andra agents förväntningar och intentioner (Bond, 1988:16f; Moulin och Chaib-Draa, 1996:6).

Arbetsdelningen, uppdelningen av problemlösningsprocessen på flera agenter, framstår som ett annat nyckelmoment i den distribuerade artificiella intelligensen. En arbetsdelning kräver att huvudproblemet delas upp i flera subproblem som sedan fördelas ut till agenter/noder med tillräcklig kompetens, information och kapacitet för att kunna lösa sin uppgift. Till grund för en distribuerad problemlösningsprocess ligger alltså en viktig distributionsprocess, liknande den tidigare beskrivna planeringsprocessen. Distributionsprocessen kan man dela upp i tre steg. Till att börja med måste det övergripande målet fastställas och huvudproblemet formuleras. När problemet väl är definierat är det i regel möjligt att dela upp lösningssprocessen i ett antal deluppgifter. Det sista steget i distributionsprocessen går ut på att allokera uppgifterna och resurserna så att huvudproblemet blir löst på effektivast möjliga vis (Bond, 1988:10ff).

2.1.4 Användningsområden

Översikten över DAI och MAS har hittills varit av teoretisk art och jag tror därför det kan vara en idé att försöka illustrera och klargöra begreppen genom att ge några exempel. Ett relativt enkelt exempel på agentteknik är sk assistentagenter (Maes, 1994). Denna redan idag relativt utbredda form av artificiella agenter bistår sina användare på en rad olika sätt. De kan dölja svåra uppgifters komplexitet, de kan undervisa användaren, de kan hjälpa olika användare att samarbeta och de kan övervaka olika procedurer. I praktiken innebär det, med dagens teknik, framförallt att agenterna utför olika sorters rutinartade, repetitiva uppgifter. Exempelvis kan det röra sig om informationsfiltrering, sökarbeten på Internet, e-mailsortering och mötesbokning.

De två viktigaste användningsområdena för DAI i samband med implementering i stora komplexa organisationer är i form av kontroll/övervakningssystem samt i system för beslut/planeringsstöd. Dessa återkommer i en rad olika verksamheter och områden. Bl a finner vi DAI-baserade kontroll- och övervakningssystem för flygtrafik, fartygstrafik och kärnkraftverk. Beslut- och produktionsplaneringssystem i form av MAS återfinns i olika former av komplexa organisationer (Müller och Wittig, 1993; Martial, 1992).

Ett exempel på ett DAI-baserat beslutstödssystem är logistiksystem MARS som skapats för samarbetande speditorsfirmor. MARS är en simulation av transportfirmor och den omgivning de arbetar i uppbyggd som ett MAS. Syftet med systemet är att skapa ett

planeringsunderlag för att kunna optimera verksamheten. Verksamhetens mål är att med de transportmedel som står till buds uppfylla de från kunderna efterfrågade uppdragen så billigt som möjligt. Omgivningen – infrastruktur i form av av-/pålastningsstationer, städer, färdvägar o s v – är statistiskt modellerad. I övrigt har systemet en dynamisk struktur. Det betyder att antalet agenter, resurstillgången och antalet uppdrag kan ändras efter hand. Det finns olika former av agenter i systemet. Speditionsagenterna sköter uppdragsbearbetning och resursfördelning. En andra agentform utgör lastbilarna, som fungerar som experter på områdena vägplanering och sammanställning av last. Det finns vidare specialagenter som uppfyller vissa uppgifter inom systemet och inte har någon egentlig motsvarighet i verkligheten. Dessa är i MARS medlaragenter, som arbetar med uppdragsfördelning, och världsagenter, som skapar representationer av speditörernas och lastbilarnas verksamhet samt dess påverkan på "världen". För att uppnå det globala målet, en optimering av transporter, måste agenterna samarbeta och förhandla med varandra. I MARS skiljer man mellan vertikalt, inom en spedition, och horisontellt, mellan olika speditioner, samarbete. På så sätt försöker man sammanlänka uppdragen och resurserna på ett sätt som gynnar alla parter. De mottagna uppdragen uppdelas i regel i olika deluppdrag, som sedan fördelas ut via ett *Contract Net* (Müller och Wittig, 1993:275ff).

2.2 Socionik

2.2.1 Forskningsfältets framväxt

DAI har fram till för något år sedan främst varit ett gebit för ingenjörer och informatiker. Vissa av dessa har förvisso förespråkat ett rotande av DAI i sociologisk teori, men i realiteten har inte samarbetet med sociologer inletts förrän alldeles nyligen. De sociala grunderna i DAI-modellerna har därför i regel varit klart förenklade och ofta byggts på samhällsmodeller bestående av individualistiska, rationella aktörer.

Socionik är namnet på ett nytt tvärvetenskapligt forskningsfält som växt fram i gränslandet mellan sociologi, informatik och artificiell intelligens. Initiativet till forskningsdisciplinen har tagits av ett antal tyska sociologer. De har gemensamt framfört en kritik mot det allt för naiva och alldagliga umgänget med sociologiska begrepp och modeller inom forskningsdisciplinen DAI (Malsch et al, 1998). Rammert var tidigt ute med att förespråka en mer omfattande roll

inom AI-forskning än bara ett studium av AI-projektens sociala följder och ansvarsfrågor. Redan ett par år innan socionikprojektet officiellt såg dagens ljus menade han att sociologin kunde bidra med värdefulla inlägg i AI-debatten. Framförallt fanns det ett utrymme för sociologisk kritik mot utveckling och användning av begrepp och modeller inom AI (Rammert, 1995). Med inspiration från Steve Woolgar föreslår Rammert en sociologi som även sysselsätter sig med forsknings- och utvecklingsfrågor samt gränslinjer mellan människa och maskin inom den artificiella intelligensens forskningsfält (Woolgar, 1985:558f; Rammert, 1998:100f).

Begreppet socionik myntades av den tyske sociologen Thomas Malsch 1995. Precis som man inom bioniken använder förebilder i naturen vid modellering och konstruktion av tekniska artefakter, sysselsätter sig socioniken med hur man med hjälp av förebilder och mekanismer i den sociala världen kan lösa koordinations- och kommunikationsproblem i MAS. Egentligen är dock fältet äldre än så. En grupp amerikanska DAI-forskare ringade in området redan i slutet på 1980-talet men valde att låta namnfrågan ligga på is och kallade fältet helt enkelt *the Unnamable* (Malsch et al, 1998). Dessa forskare delade sedan tidigt 80-tal ett gemensamt intresse för frågor om distribuerad problemlösning, modellering av datorprogram som sociala organisationer samt decentraliserat samarbete i datorsystem. Åsikten att man kan profitera från ett användande av sociologisk modell- och teoribildning går att finna relativt tidigt i DAI-kretsar. Les Gasser argumenterar 1991 i en artikel för att det behövs ett sociologiskt fundament för att kunna skapa storskaliga MAS (Gasser, 1991). Andra DAI-begrepp, såsom koordination, organisation, kommunikation och förhandling, skvallrar också om att vi rör oss inom ett område där målet är att skapa social samverkan på ”konstgjord” väg. I det socioniska forskningsprogrammet fungerar utforskningen och modelleringen av artificiella sociala system, exempelvis MAS, som ett uttalat mål (DFG-Schwerpunktprogramm Sozionik). Man skulle alltså kunna betrakta socioniken som en utvidgning av forskningsprogrammet distribuerad artificiell intelligens.

2.2.2 Forskningsperspektiv och inriktningar

De gynnsamma effekterna av ett samarbete mellan DAI och sociologi rör sig enligt många av socionikens företrädare i två motsatta riktningar. Å ena sidan innebär ett sociologiskt intresse för och bidrag till DAI att dess grund blir fastare och DAI-systemen på sikt kan komma att

fungera bättre. Sociala system besitter i regel just sådana egenskaper som DAI-forskningen eftersträvar i skapandet av MAS. Sociala system är robusta och feltoleranta eftersom de har en förmåga till självreparation och självrevolution. Sociologin kan, enligt förespråkarna, erbjuda modeller och begrepp för hur sociala system och dess komponenter fungerar (Malsch, 1998:31; Rammert, 1998:103). Å andra sidan finns det också vinster att göra för sociologin. DAI kan erbjuda instrument för att testa komplexa teorier och modeller. Med MAS är det möjligt att simulera sociala processer av olika slag. På så sätt går det att explorativt testa modellbildningar som är för abstrakta och/eller komplexa för att styrkas/falsifieras av empirisk data (Malsch, 1998:37; Rammert, 1998:103).

Tanken är alltså att när ett tekniskt system, exemplifierat av MAS, beter sig som ett socialt system så ligger det nära till hands att använda de tekniska systemen för att simulera sociala processer. Socioniken har metaforiskt beskrivits i form av en jacka som går att vända ut och in (Braun & Imhof, 1999). Resonemanget implicerar alltså att den beskrivna ömsesidiga vinsten av ett samarbete likväl kan tolkas som ett ömsesidigt beroende.

I argumentet – en ömsesidig vinst – för etablerandet av socionik som forskningsdisciplin tycker jag mig se något av ett cirkelresonemang. För att kunna skapa ett väl fungerande system av samarbetsdugliga, autonoma artificiella agenter behöver man ta hjälp av sociologiska teorier och modeller rörande abstrakta sociala begrepp som samarbete, organisation och förhandling. Dock är dessa sociologiska teoribygggen² i regel så abstrakta att de inte fullt ut går att testa med hjälp av dagens sociologiska metodarsenal. För att kunna legitimera dessa teorier skulle man, menar de socioniska företrädarna, behöva ta hjälp av simulationer byggda på DAI-teknik. En DAI-teknik som dock blir verklighetstrogen och sociologiskt trovärdig först när den grundats i dessa nämnda, abstrakta sociologiska teorier. Kort sagt abstrakta *grand theories* används som underlag för MAS. MAS används som underlag för att testa samma abstrakta *grand theories*. Jag tycker det går att ana en paradox här någonstans.

I vilket fall som helst utgör dessa två rörelser eller riktningar de två första subdisciplinerna inom socioniken. Den tredje centrala subdisciplinen är studiet av de hybrida system – system

² Jag syftar här på teoribildningar som varit uppe till diskussion inom socioniken och som i regel är högst abstrakta och empiriskt svårtestade. Exempelvis rör det sig om Giddens struktureringsteori och Luhmanns systemteori. En närmare diskussion om dessa teories roll inom socioniken återkommer jag i 2.2.3.

bestående av såväl mänskliga som artificiella agenter – som blir en effekt av att MAS implementeras i organisationer av olika slag. För att kunna räta ut den ovan nämnda paradoxen och i övrigt reda ut förhållandena mellan de olika subdisciplinerna har jag nedan försökt teckna en bild av den diskussion som förs inom respektive delområde.

2.2.2.1 Den distribuerade artificiella intelligensens sociologi

Det finns som sagt kritiska invändningar att göra mot de tendenser inom socioniken som strävar mot att skapa allmängiltiga sociala mekanismer för MAS. ”Socionikern” framstår härigenom som en sociologisk ingenjör, som m h a universella sociologiska teorier om samhällsliga processer skapar samhällen av konstgjorda agenter. Det går inte att kringgå det faktum att de flesta teoretiska modeller som föreslås i konstruktivt syfte i socioniklitteraturen är *grand theories*, d v s abstrakta och allmängiltiga teorier, exempelvis i form av Giddens struktureringsteori och Parsons strukturfunktionalism (jfr 2.2.3). I det officiella forskningsprogrammet talas det om samhället som en reservoar med förebilder för modellerande av sociala mekanismer (Schwerpunktprogramm Sozionik). Sociologin framstår utifrån det ovan beskrivna synsättet som en homogen och sammanhållen vetenskap där det är möjligt att objektivt beskriva samhället enligt allmänna lagbundenheter. På samma sätt är det nu tänkt att man inom socioniken skall kunna skapa förutsägbara, om än dynamiska, artificiella samhällen baserade på dessa *grand theories*. Det är svårt att inom de egna leden hitta uttalad kritik mot denna hållning. Dock varnar såväl Jörg Strübing som Ingo Schulz-Schaeffer för faran att betrakta sociologin som en sammanhållen och enhetlig vetenskap. De menar att det är felaktigt att tala om en sammanhållen sociologisk grund för DAI. Möjligtvis kan sociologin erbjuda ett antal olika plattformar att utgå från (Strübing, 1998: 60; Schulz-Schaeffer, 1998:136). Vidare, menar Schulz-Schaeffer, måste man kritiskt reflektera över om det överhuvudtaget är möjligt och till nytta att använda sociologiska teorier och modeller vid MAS-modellering (Schulz-Schaeffer, 1998:136f).

Thomas Malsch, socionikens grundare, varnar för svårigheten att rakt av överföra sociologiska begrepp och teorier till tekniska vetenskaper. Istället förespråkar han en ”metaformigration”. Med detta begrepp avser han en stegvis ombyggnad av förebilder från den sociala världen till tekniska artefakter (Malsch, 1998:48). Malsch får kritik från Strübing som menar att begreppet är för enkelspårigt. Socioniken handlar inte enbart om att med reformerade sociologiska teorier lösa DAI/MAS-problem. Istället förespråkar Strübing en

interaktiv organisation av disciplinen som möjliggör lösning av gemensamma problem. Socioniken handlar alltså enligt Strübing snarare om att på grundval av gemensamma problem och uppgifter skapa ihållande samarbetsstrukturer (Strübing, 1998).

Malsch talar vidare om en tredelad sociologisk utmaning. För det första kan sociologin fungera som en konkurrent i teoridiskussionen. Det innebär ett kritiskt granskande av DAI-modeller och även erbjudande av egna bidrag som kan betraktas som mer relevanta. Vidare kan sociologin fungera som samarbetspartner och gemensamt med vetenskapen om DAI försöka förenkla och formalisera sociologiska begrepp. En tredje möjlig roll för sociologin är betraktarens. Här handlar det om att observera och beskriva arbetsprocessen inom DAI från ett sociologiskt perspektiv (Malsch, 1998:28).

2.2.2.2 Datorsimulation som sociologisk metod

Datorsimulationer av samhällsvetenskapliga fenomen är egentligen ingen ny metod. De har förekommit i olika sammanhang såväl som ett sätt att testa teorier som att prognostisera sociala skeenden. Ofta har det handlat om relativt primitiva modeller som i regel varit alltför reduktionistiska och koncentrerade antingen helt till mikro eller makronivå.

Förespråkarna för datorsimulationer som en sociologisk metod menar att sociala simulationer kan fylla en lucka mellan empirisk forskning och teoretiskt arbete samt också fylla ett tomrum som den existerande sociologiska metodarsenalen uppvisar. I vissa fall är sociala fenomen inte direkt åtkomliga m h a av sociologisk mainstream-metodologi. Vissa makrofenomen går inte att fånga in och undersöka empiriskt, då dess strukturer är alltför komplexa och omfattande. I andra situationer, exempelvis vid studier av historiska fenomen, då det inte finns möjlighet till insamling av data kan datorsimulationer utgöra ett intressant alternativ (Gilbert och Conte, 1995).

Detta område har dock stött på mycket kritik, inte minst inom socioniska kretsar. En kritikpunkt är att de metaforer som ligger till grund för sociala datorsimulationer i regel är alltför primitiva. Även simulationer som utger sig för att motsvara fenomen i ett modernt samhälle bygger ofta på beteendemodeller hämtade från djurvärlden (Schulz-Schaeffer och Malsch, 1998:235). Vidare kritik påpekar att den sociala världen är alltför komplex för att överhuvudtaget kunna modelleras oberoende av simulationsprogrammets komplexitet. Sociala

sammanhang kan inte framställas som syntaktiskt konsistenta och slutna modeller. Möjligtvis kan man skapa modeller av sådana sociala situationer som kännetecknas av högt standardiserade och genomrationaliserade handlingsstrukturer (Malsch, 1998:42f). Mot detta kan dock invändas att denna kritik i så fall också skulle gälla den sociologiska teorin och modellbildningen i allmänhet. Det är som Harry Collins hävdar att: "...*neither regular science nor machines can model social life*" (citerad i Malsch, 1998:44). Så länge man betraktar datorsimulationer som ett metodiskt verktyg i mängden med dess möjligheter och begränsningar faller de ovan anförda invändningarna. Men att överföra begreppet artificiell intelligens till detta fält och börja tala om artificiella samhällen är dock högst tveksamt (jfr Gilbert och Conte, 1995). Malsch påpekar också att det är ett klart missvisande begrepp. Han menar att DAI-modeller på sin höjd kan representera sociala fenomen men de kan aldrig ersätta den sociala verkligheten (Malsch, 1998:45).

Nigel Gilbert anför andra invändningar mot kritiken att den sociala verkligheten och dess strukturer skulle vara allt för komplexa för att kunna representeras och testas i datorsimulationer. Han tar sin utgångspunkt i kaosteorin, som han menar påvisat att högkomplexa och oförutsägbara system ofta uppstår utifrån ett par enkla principer. Diverse datorsimulationer har uppvisat ett liknande mönster. Simuleringar där agenterna är väldigt primitiva – reaktiva och endast utrustade med två eller tre handlingsalternativ – leder ofta fram till högkomplexa strukturer. Det tyder, enligt honom, på att det visst kan vara så att också några relativt enkla principer ligger till grund för människors beteende och sociala strukturers uppkomst. Utvecklingen av högkomplexa strukturer följer dock inte några enkla orsak-verkansamband och går därför inte att spåra med matematiska, statistiska sociologiska metoder. Simulationer skulle då enligt Gilbert vara det enda metodiska verktyg som återstår (Gilbert, 1995).

Socioniken företräder i regel en pragmatisk hållning till sociala simulationer. De pekar på de brister som system för sociala simulationer uppvisar idag. De sociologiskt grundade simuleringarna har hittills förbisett MAS möjlighet och använt sig av konventionell AI-teknik. Å andra sidan har de ansatser som börjat använda sig av MAS som simulationsverktyg varit allt för dåligt grundande i sociologin. Genom ett samarbete mellan sociologin och DAI menar dock företrädare för socioniken att dessa brister går att övervinna. De menar vidare att MAS-tekniken lämpar sig väl för att undersöka sambandet eller samspelet mellan den sociala mikro- och makronivån (Schwerpunktprogramm Sozionik). Det kräver dock ett gemensamt studium

av begreppet *emergence*, alltså sociala strukturers uppkomst och påverkan på aktörerna. Jag återkommer framöver till de teoretiska diskussionerna därom.

2.2.2.3 Hybrida system

MAS utveckling och utbredning innebär att så kända hybrida system uppstår i olika former av organisationer där multiagent-tekniken implementeras. Per definition är ett hybrid system ett system som ”... *do not distinguish between interaction with humans and interactions with artificial agents*” (Parunak, 1996:150). Denna definition implicerar att det principiellt råder en jämlik relation mellan mänskliga och artificiella agenter i hybrida system. En vidare egenskap i hybrida system är dess heterogenitet. Dess komponenter – människor och artificiella agenter – har olika former och funktioner (Burkhard och Rammert, 2000).

Van Dyke Parunak menar att de hybrida systemens utbredning har dramatiska konsekvenser för människans respektive teknikens roll i organisationerna. Han väljer t o m att betrakta dess genomslag som ett brott i den mänskliga arbetsorganisationens historia. Fram till industrialismens genombrott hade verktyg i allmänhet enbart en manuell funktion. Människan var tvungen att bidra med såväl kraft (energi) som intelligens. Nya energikällor och snabb teknisk utveckling bidrog i den industriella revolutionens släptåg till att maskiner ersatte den mänskliga, fysiska kraften. Människans roll i arbetsorganisationen växlade från hantverkare till operatör. Nu, menar Parunak, tillfogas också maskinerna den sista komponenten, nämligen intelligensen. Maskinerna kommer därmed, med MAS-teknik, snarare att fungera som kollegor till människorna (Parunak, 1996:151). Denna utvecklingsbeskrivning framstår förvisso som hårt tillspetsad och har hittills svag empirisk grund. Men faktum kvarstår att DAI och MAS säkerligen kommer förändra människans (och teknikens) roll i (de hybrida) organisationerna.

Detta faktum har man inom socioniken börjat uppmärksamma. De hybrida systemens framväxt och utbredning ger upphov till många sociologiska problemställningar. Rammert påpekar att den strukturella kopplingen och översättningen mellan de olika – mänskliga sociala och tekniska agentbaserade – systemens logik kräver en omfattande socionisk forskningsinsats. Människa-maskin interaktionen har såväl en teknisk som en social sida och utveckling av hybrida system kräver därför samarbete över disciplinerna. För att kunna skapa DAI-baserade hybrida system som överlever i realiteten, krävs det att man utifrån ett

sociologiskt perspektiv studerar beroendena och relationerna mellan mänskliga och icke-mänskliga agenter (Rammert, 1998).

2.2.3 Teoretiska problemområden

Det finns framför allt två övergripande teoretiska inriktningar inom socioniken i dag. Den första inriktningen sysselsätter sig med frågor om de artificiella sociala systemens dynamik. Här handlar det om, för sociologin bekanta frågeställningar som dialektiken mellan aktör och struktur samt förhållandet mellan ett systems/samhälles mikro och makronivå. Det andra övergripande teoretiska problemområdet behandlar frågeställningar som uppkommer i samband med de hybrida systemen.

DAI kritiseras från socioniskt håll för att ha haft en allt för reduktionistisk hållning i förhållandet mellan agent och struktur i de utkast till MAS-modeller som hittills modellerats. De två huvudsakliga agentmodeller som hittills stått som förebild har antingen överbetonat den övergripande strukturens inflytande – reaktiv agentmodellering – eller utgått från antagandet att agenten besitter en total handlingsfrihet, som är fallet vid reflektiv agentmodellering (Rammert, 1998:107f). Den stora frågan kvarstår dock: hur är uppkomsten av strukturer i MAS förenbar med antagandet om autonoma agenter? Rammert föreslår här en lösning av problemet m h a Giddens struktureringsteori.

Anthony Giddens försöker genom sin struktureringsteori lösa upp den dualism som råder mellan aktör och struktur i stora delar av den sociologiska teoridiskussionen. Han hävdar istället att det råder en strukturdualitet, d v s det föreligger en dialektisk relation mellan aktör och struktur där de båda delarna är ömsesidigt beroende av varandra. Ett nyckelbegrepp i teorin är handlingen, *agency*. En handling skapar betingelser för nya handlingar. Dessa nya handlingar skapar i sin tur nya betingelser. Detta kretslopp, som försiggår i en social praxis, benämner Giddens struktureringsprocess. Strukturer existerar därmed endast virtuellt i social praxis och i människors minne. Strukturernas – eller snarare de strukturella egenskaperna som Giddens väljer att benämna dem – dualitet ligger därmed däri att de utgör såväl ett medel för som ett resultat av agents sociala praxis. De är vidare såväl möjlighetsskapande – *enabling* – som handlingsbegränsande – *constraining* (Ritzer, 1996:528ff; Giddens, 1988).

I struktureringsteorin ser man på många håll inom socioniken en potential att förbinda agenternas handlingsförmåga med uppkomsten av sociala strukturer i MAS. I sociala "agentsamhällen" formulerar och följer autonoma agenter egna mål men bygger samtidigt sina mål delvis utifrån påverkan från den sociala omgivningen. Denna *bounded autonomy* som man länge betraktat som en svårlöst paradox i DAI går kanske att lösa genom att applicera struktureringsteorin. Förutsättningarna verkar vara likartade och struktureringsteorin tillåter också att nya egenskaper hos agenterna utvecklas. Flera sociologiska företrädare menar också att det riktiga svaret på paradoxen återfinns i Giddens teorier (Rammert, 1998; Conte & Castelfranchi, 1995).

Det förekommer vidare andra sociologiska bidrag inom socioniken och dess problemställningar mellan mikronivå – den enskilda, autonoma agenten – och makronivå – den sociala koordinationen av agenternas handlingar. Schulz-Schaeffer och Malsch (1998) utgår i en artikel från antaganden som förekommer i Hewitts öppna MAS. Agenten betraktas där som en *black box*, dvs dess mentala och kognitiva operationer är inte siktbara för systemets andra agenter. Vidare antas kommunikationen agenter emellan äga rum på ett "armslängds avstånd" (Hewitt, 1991). I ett sådant system där även handlingsfrihet och tillgång till olika handlingsalternativ förutsätts uppkommer en situation av dubbel kontingens (osäkerhet). Begreppet är lånat från Talcott Parsons och betyder att en agents handling är beroende av såväl de tillgängliga alternativen som de reaktioner handlingen kan tänkas uppbringa hos de andra i interaktionsprocessen deltagande agenterna. Eftersom agenterna inte kan förutsäga de andra agenternas handlingsval och reaktioner uppstår här en situation präglad av osäkerhet. Schulz-Schaeffer och Malsch tar hjälp av ett annat Parsons-begrepp – symboliska generaliserade interaktionsmedier – för att lösa detta koordinationsproblem.

Enligt Parsons har olika former av symboliska generaliserade interaktionsmedier skapats för att handskas med koordinationsproblem i det moderna samhället. De återkommer i olika former, såsom pengar och makt. De överbrygger problemet med dubbel kontingens genom att förmedla utvecklade och stabiliserade ömsesidiga förväntningar hos de berörda aktörerna (Ritzer, 1996:248f). Schulz-Schaeffer och Malsch kritiserar de försök som hittills gjorts inom DAI för att koordinera agenter handlande. I regel har man använt sig av fördefinierade restriktioner, i form av fixa globala strukturer eller handlingsprogram, för att begränsa valmöjligheterna i givna situationer. Sådana åtgärder riskerar, menar Schulz-Schaeffer och Malsch, reducera agenternas handlingar och interaktion till ett realiserande av *pre-established*

harmonies. M a o riskerar MAS dynamik därigenom gå förlorad. Ett alternativ skulle här kunna vara att modellera koordinationsmekanismer som fungerar som symboliska generaliserade interaktionsmedier. Genom att använda generaliserade interaktionsmedier som koordinationsresurser är det möjligt att minska kommunikationen och därmed göra systemet effektivare. Det finns redan ansatser som pekar i den riktningen. I s k *Contract-Net Protocol* används en form av ”pengar” för att föra lösa konflikter och fördela uppgifter. Det kvarstår dock att betänka att agenterna i ett MAS, liksom aktörerna i samhället i stort, har möjligheten att låta bli att använda de symboliska generaliserade interaktionsmedierna. Konsekvensen blir då att interaktionsmedierna visserligen kan reducera osäkerheten i interaktionsprocessen men de kan inte helt eliminera den (Schulz-Schaeffer och Malsch, 1998).

En annan sociologisk inriktning som delvis figurerat i DAI-sammanhang under en ganska lång tid är symbolisk interaktionism (SI). Strübing argumenterar i en artikel för ett utökat samarbete mellan SI och DAI (Strübing, 1998). Han menar att de båda inriktningarna uppvisar likheter såväl vad gäller forskningsstil och metodik som i dess teoretiska begreppsapparat. Han spårar också, både implicit och explicit, tankar kopplade till SI hos tidiga företrädare för DAI som Carl Hewitt och Les Gasser.

Utgångspunkten för Strübing är Hewitts och Gassers teoretiska arbeten om grunderna för DAI. De båda talar om MAS och distribuerade system som öppna nätverk eller system (open systems). Öppna system är inte som de flesta andra datorsystem rumsligt och tidsligt slutna utan kan växa och utvecklas på egen hand. De saknar vidare en central beslutande instans (Hewitt, 1991:81f; Hewitt, 1986:272ff). De öppna systemen liknar organisationsformer som har sin grund i SI. I *the ecological approach*, en subdisciplin inom SI, betraktas organisationer som variabla miljöer. De uppkommer och formas i interaktionssammanhang. De sociala strukturerna synliggörs endast i interaktionen agenter emellan och kan därmed bara förändras genom socialt handlande.

Agenter/aktörer handlar i organisationen/systemet utifrån olika perspektiv. De tolkar problemen på olika sätt, använder sig av lösningsstrategier som avviker från varandra samt uppvisar olikheter i tillvägagångssättet (metodik). Just denna mångfald av aktörperspektiv hoppas företrädare för DAI kunna bidra till en uppkomst/framväxt, *emergence*, av kreativa lösningar på komplexa, flerdimensionella, problem. Frågan som uppstår här är hur de

heterogena perspektiven binds samman? Vad finns det för förhandlings- och koordinationsmodeller? Strübing (1998:64ff) tar här upp några möjliga lösningsvägar.

Den ekologiska inriktningen pratar om *commitments* – åtaganden eller förpliktelser – som ett sammanhållande element i organisationer. Den sociala tillhörigheten till en problemlösningssituation är problem- och processberoende. Fujimora, en företrädare för nyare SI, menar på att komplexa problem måste omvandlas till *doable problems* (rutinarbete). Han försöker med dessa tankar förklara hur vetenskapliga och teknikutvecklingsprojekt lyckas bli genomförda trots heterogena aktörer och en växande komplexitet. Transformationsprocessen är en interaktiv process där problemen struktureras och delas upp av olika instanser på olika nivåer. En sådan process bygger på förhandlingar där aktörerna/agenterna försöker komma fram till en gemensamt kompatibel (men ej nödvändigtvis identisk) tolkning av problemet (Strübing, 1998:70ff).

DAI-forskare som Hewitt och Gasser har låtit sig inspireras av dessa idéer och utvecklat liknande tankegångar. Problemlösningssituationen i en organisation eller i ett MAS bygger på olika agenter mikroteorier. Dessa mikroteorier är i själva verket små, idealiserade, matematiska teorier som bygger på logisk deduktion (Hewitt, 1986:278). Olika agenter använder olika mikroteorier och det innebär att de kan vara inkonsistenta med varandra. För att lösa dessa konflikter förespråkar Hewitt och Gasser s k *due processes*. En sådan process definieras av Hewitt som den organisationella aktivitet som äger rum för att skapa en relevant och pålitlig beslutsbas inom gränserna för de tillgängliga resurserna (Hewitt, 1986:275). Denna aktivitet består, enklare uttryckt, däri att man i en förhandlingssituation gör en prövning av olika lösningsvägar och dess bakomliggande regler (Gasser, 1991).

Den andra teoretiska huvudfrågan, de hybrida systemen, kommer jag behandla i senare avsnitt. Därför tänker jag här inte gå in djupare på möjliga teoretiska ingångar för detta område. Dock tror jag ändå det kan vara en idé att ge en första kort orientering av möjliga teoretiska angreppssätt. I den socioniska litteraturen förekommer det framförallt frågor om tekniska artefakter i form av artificiella agenter är kapabla till handling och huruvida de kan betraktas som jämlika interaktionspartners.

Strübing kommenterar frågan utifrån ett SI-perspektiv. Aktörerna i handlings-/problemlösningssituation ses här som socialt konstruerade. De existerar endast som

problemlösande agenter genom att vara en del av kommunikations- och interaktionsprocessen. SI utgår alltså från ett pragmatiskt synsätt enligt devisen: *"If men define situations as real, they are real in their consequences"* (Thomas & Thomas citerade i Strübing, 1998:76). I en handlingskontext består därför principiellt ingen skillnad mellan mänskliga och icke-mänskliga agenter, så länge aktörer själva inte uppfattar en sådan. Det betyder m a o att om maskiner – artificiella agenter – kan uppbringa signifikanta gester och uttryck som mänskliga aktörer kan tolka och handla efter, så kan artificiella agenter anses besitta en handlingsförmåga (Strübing, 1998:73-77).

Ett annat teoretiskt bidrag som återkommer i socioniklitteraturen är aktör-nätverkteorin. Inte heller här görs det någon åtskillnad mellan tekniska artefakter (exempelvis artificiella agenter) och sociala (mänskliga) aktörer vad det gäller handlingsförmåga. Istället betraktas alla element i ett sociotekniskt system, vars handlande påverkar andra elements handlande som "aktanter", d v s jämställda deltagare i handlings-/problemlösningsprocessen (Schulz-Schaeffer, 1998).

Jag har här bara gett exempel på två huvudspår inom den teoretiska debatten om interaktion och handling i hybrida system. I mitt fortsatta arbete kommer jag att utveckla dessa båda riktningar och samtidigt presentera en rad andra uppfattningar, framförallt hämtade från tekniksociologin, som jag anser vara relevanta för min problemställning.

3 Agenten som interaktionspartner – ett mikroperspektiv

Framställningen av softwareagenter är förknippad med två huvudproblem. Det första rör agentens kompetens. Agenten måste veta när, med vad och hur den skall hjälpa användaren. Det andra problemet handlar om användarens tillit till agenten. I vilken utsträckning är användaren beredd att delegera problem till en softwareagent (Maes, 1994)? Dessa två problem utgör en lämplig utgångspunkt för en analys av integrationsproblematiken på mikronivå. Det första problemet utgår från agenten själv och handlar om i vilken utsträckning agenten är kapabel till handling. Det andra problemet rör de mänskliga aktörerna och deras acceptans av artificiella agenter som samarbetspartners. Dessa båda problem möts slutligen i en analys av interaktionen mellan mänskliga användare/aktörer och artificiella agenter.

3.1 De artificiella agenternas handlingsförmåga

För att kunna avgöra under vilka betingelser artificiella agenter är kapabla till handling, är det först nödvändigt att definiera vad handling respektive handlingsförmåga är.

Handlingsbegreppet behandlas på en rad olika sätt av olika discipliner. Det allmänt utbredda handlingsbegreppet inom klassisk AI, *the planning view*, innebär att handlingar förstås utifrån dess bakomliggande planer. En plan ses här som en sekvens av handlingar utformad för att uppnå ett visst slutmål. Handlingarna är delkomponenter i en längre problemlösningsprocess. Detta perspektiv är väl utbrett i AI-kretsar, framför allt eftersom det är relativt lätt att formalisera och modellera. Planer fyller ungefär samma funktion för organismer som program gör för datorer. De innehåller en hierarkisk uppsättning instruktioner för hur ett visst problem skall lösas (Suchman, 1987:27ff; Wooldridge & Jennings, 1995:10).

3.1.1 Det sociologiska handlingsbegreppet

Det kognitivt orienterade handlingsbegreppet inom AI möter dock kritik från sociologiskt håll. Lucy Suchman använder begreppet *situated action* för att klargöra att handlingar inte utgår från rationella planer där omgivningen abstraherats bort. Hon hävdar istället att varje handling är beroende av den omgivande kontexten och dess sociala förutsättningar. En handling kan inte betraktas isolerad från det sociala sammanhang den äger rum i (Suchman, 1987:49). Suchmans kritik mot hur handlande betraktas hos många företrädare för AI innebär

ett handlingsbegrepp som orienterar sig i sociologisk riktning. Talcott Parsons framställer i *The Structure of Social Action* (1937/1968) en referensram för mänsklig handling där handlingens kontext framhävs. Den ”analytiska enheten” handling (act) innehåller enligt honom fyra olika analytiska komponenter. Förutom att handlingen förutsätter någon form av aktör samt riktar sig mot någon form av slutstadium/mål, är en handling också alltid beroende av den omgivande situationen. Slutligen föreligger det ett samband mellan situationens förutsättningar och mening. Parsons benämner detta samband för handlingens normer.

Även George H. Meads handlingsbegrepp distanserar sig från det kognitiva synsättet. Enligt Mead formas handlingar i samverkan med den omgivande världen. För att kunna handla måste aktören först identifiera vad han/hon vill uppnå. Aktören är en aktiv organism som handlar med utgångspunkt från de objekt hon/han själv indikerar. En handling ses som ett beteende konstruerat av aktören själv. Mead delar vidare upp en handling i fyra steg. Handlingen uppkommer genom att organismen befinner sig i ett *state of disequilibrium*, orsakat genom någon form av impuls. Ett begrepp som impuls kan lätt förleda en att tro att det rör sig om en kraft/stimuli som påverkar individen/organismen i en särskild riktning och framkallar en särskild respons. Meads handlingsmodell bör dock snarare ses som en reaktion mot sådana behavioristiska tankegångar. Mead tillskriver individen ett större handlingsutrymme. Individen definierar situationen och uppställer ett mål. I ett tredje steg försöker individen, utifrån sitt uppsatta mål, att manipulera omgivningen och fysiskt handla mot dess objekt. En ideal handlingsakt avslutas med att målet uppnås. Det sker dock inte alla gånger. I vissa fall uppkommer en impuls och individen påbörjar en ny handlingssekvens innan den förra avslutats (Blumer, 1969; Charon, 1979:118ff; Shibutani, 1987:64-70).

Meads handlingsbegrepp ligger också till grund för symbolisk interaktionism. En handling betraktas här som en del av en längre ström av handlingar. En enstaka handling är därför svår att urskilja och avgränsa från strömmen. Handlingen ses snarare som en analytisk enhet. Handlingar utförs i situationer av olika slag. Individen definierar mål för att handskas med problem som uppfattas i den aktuella situationen. Med målet i sikte uppställer individen vidare en handlingsplan. Målen och därmed också planerna tenderar att orientera sig mot de objekt, människor inkluderade, som individen indikerar i den föreliggande situationen (Charon, 1979:111ff).

Anthony Giddens väljer att belysa skillnaden mellan olika former av handling. Han utgår från en skiktad handlingsmodell. Giddens beskriver en agents handlande utifrån tre nivåer. På den understa nivån styrs agentens handlande av omedvetna motiv. Vidare finns det en praktisk medvetandenivå hos agenten där handlingarna är rutiniserade och automatiserade. Slutligen kan agenten också handla utifrån ett explicit uttryckbart motiv. Giddens talar här om ett diskursivt medvetande (Giddens, 1988:55ff).

3.1.2 Handlingsförmåga hos människor och maskiner

Handlingsförmågan är kopplat till handlingsbegreppet och definieras därför utifrån de villkor som tillskrivs den handlande aktören. Schulz-Schaeffer ställer upp två kriterier för handlingsförmåga. För att aktören skall kunna sägas besitta en handlingsförmåga förutsätts för det första att aktören skulle kunna ha handlat annorlunda än hon/han gjorde i en viss situation. För det andra bör handlingen vara intentionell, d v s aktören bör kunna ange en orsak eller ett skäl till varför hon/han handlade som hon/han gjorde (Schulz-Schaeffer, 1998:133). Liknande tankegångar återkommer även inom symbolisk interaktionism (SI). Människor utför inte en handling enbart som en förprogrammerad respons på en viss stimuli utan gör val baserade på den föreliggande situationen och dess element. M a o agerar människor snarare än bara reagerar. De har också en förmåga att ta hänsyn till tidigare erfarenheter och möjliga framtida konsekvenser i valet av handling (Charon, 1979:126).

Med liknande termer väljer Harry M. Collins att diskutera maskiners handlingsförmåga. Utgångspunkten är att det föreligger en åtskillnad mellan handling (act) och beteende (behaviour). En handling kräver ett medvetet subjekt som uppvisar en intention bakom handlingen. Maskiners avsaknad av intentionalitet innebär att de förvisso kan utföra mänskligt arbete – även intelligent sådant, exempelvis i form av problemlösning – men per definition är de inte kapabla att handla utan uppvisar bara ett beteende. Collins hävdar vidare att vissa mänskliga handlingar lämpar sig bättre än andra att utföras av maskiner. Dessa *machine-like acts* är handlingar som människor alltid försöker utföra med samma beteende. Det kan även gälla högkomplexa uppgifter som kan lösas av s k expertsystem genom att de uppdelas i många enklare rutinartade deluppgifter (Collins, 1990:30-45). Vid rutinhandlingar spelar subjektiviteten ingen större roll, eftersom de ofta låter sig reduceras till schematiska funktioner. Sådana funktioner kan utföras av såväl maskiner som människor. Om handlingar däremot också kräver en förmåga till reflektion uppstår en skillnad mellan maskiner och

människor (Rammert, 1998:113ff). Dessa tekniksociologiska reflektioner kring maskiner och handling utgår från klassisk AI och studier av expertsystem. Dessa AI-system har endast, i någon mån, en kognitiv problemlösningförmåga och är inte per sociologisk definition förmögna att handla. Frågan är hur det ligger till med de nya agentmodeller som föreslagits.

3.1.3 Agentmodeller och handlingsförmåga

Reaktiva agentmodeller, byggda på *the weak notion of agency*, innebär kortfattat att agenten besitter autonomi – den kan utföra operationer utan direkt intervention från människor – och reaktionsförmåga – den kan uppfatta förändringar i omgivningen och svara mot dessa.

Vidare kan reaktiva agenter kommunicera med andra agenter m h a ett ”agentspråk” och kan i vissa fall även ta initiativ till ett beteende utan direkt påverkan från omgivningen.

Dessa egenskaper pekar mot att den reaktiva agentens uppförande i viss mån orienterar sig mot Meads handlingsbegrepp. De besitter en viss situativ förmåga och kan således reagera på och lära sig av förändringar i omgivningen (jfr Maes, 1994; Scheuermann, 2000:21). Dess initiativförmåga kan också vid en snabb blick göra att man lockas tillskriva reaktiva agenter handlingsförmåga. Agenten besitter i regel ett antal olika, ofta inlärd, handlingsalternativ och skulle på så sätt kunna sägas uppfylla det första kravet på handlingsförmåga. Men någon intentionell förmåga går inte att finna hos den här typen av agenter och därför är dess operationer ej att betrakta som handlingar i sociologisk mening. Dess initiativförmåga innebär då snarare att de kan utföra vissa programmerade eller inlärd beteenden utan att någon ger order om det.

De reflektiva agenterna, modellerade enligt *the stronger notion of agency*, uppvisar ett mentalt tillstånd. Det mentala tillståndet består i en uppsättning övertygelser och önskningar eller BDI (Belief, Desire, Intention) som det ofta benämns i DAI-litteraturen. Reflektiva eller intentionella agenter borde med denna utgångspunkt i större utsträckning vara kapabla att utföra handlingar i Mead och Parsons mening. M h a normer och intentioner definierar agenten situationen och uppställer därefter mål som ligger till grund för dess operationer. Den intentionella komponenten skulle också enligt Collins innebära att agenten besitter en handlingsförmåga.

Det går också att dra paralleller mellan vissa ansatser av agentmodellering inom DAI och Giddens skiktade handlingsmodell. Kiss skissar upp en dynamisk agentmodell där såväl en

affektiv, reaktiv, som en kognitiv komponent inryms (Kiss, 1996). Kiss försöker alltså skapa en agentmodell genom att kombinera det bästa från två riktningar – reflektiv och reaktiv modellering. Han föreslår analogt med Giddens en treskiktad modell för artificiella agenter. Termerna för agentens olika handlingsnivåer – *fully deliberative actions*, *complex skilled routines* och *simple reflex actions* – indikerar att handlingstyperna ligger nära de av Giddens föreslagna (Rammert, 1998:109; Sundermeyer, 1993:26).

3.2 Social acceptans av artificiella agenter

Nästa fråga som uppkommer på mikronivå är i vilken mån människor har förtroende för artificiella agenter som interaktionspartners. Perspektivet skjuts alltså här över till den mänskliga sidan. För att agenter skall kunna integreras på ett smidigt sätt i mänskliga sammanhang och där, i samarbete med människor, utföra ”mänskliga” uppgifter krävs det att agenterna accepteras av de mänskliga aktörerna. Jag tänker i det följande först försöka rota resonemanget i pragmatisk teori för att sedan kort gå vidare och se på hur datorer och artificiell intelligens upplevs av människor i olika sammanhang.

Enligt ett pragmatiskt synsätt existerar det ingen verklighet ”*out there*”, utan verkligheten konstrueras av mänskliga aktörer då de verkar i och gentemot sin omvärld. Människorna definierar m a o de sociala och fysiska objekt de stöter på utifrån vilken nytta eller betydelse de har för dem. Pragmatismen har vidare påverkat den symboliska interaktionismen på ett markant sätt genom att betona interaktionen mellan aktören och världen som den viktiga analytiska komponenten (Ritzer, 1996:328f). Pragmatismens dynamiska synsätt lägger stor vikt vid hur aktören tolkar den sociala världen och aktörernas tolkning skapar också förutsättningar för situationen och dess konsekvenser. Utifrån ett sociologiskt perspektiv tycker sig Rammert kunna se en potential i ett pragmatiskt betraktande och menar att det därigenom är möjligt att undvika en strikt och förenklad uppdelning mellan subjekt och objekt. Vad som betraktas som subjekt respektive objekt är här snarare en fråga om hur den föreliggande situationen upplevs av de inblandade deltagarna (Rammert, 1998:113).

Även Strübing företräder ett pragmatiskt, interaktionistiskt synsätt. Han menar att alla aktörer, artificiella som mänskliga, är socialt konstruerade. De existerar som intelligenta, problemlösande väsen endast genom deltagande i kommunikations- och interaktionsprocesser.

För att en aktör överhuvudtaget skall kunna tillerkännas handlingsförmåga krävs ett deltagande i sociala processer. Att resonera kring handlande och handlingsförmåga utan referens till den sociala kontexten ter sig enligt Strübing meningslöst. Åtskillnaden mellan mänskliga aktörer och tekniska artefakter är i slutändan en fråga om hur de inblandade (mänskliga) aktörerna uppfattar och definierar situationen (Strübing, 1998:74ff).

Sherry Turkle har undersökt olika datoranvändare i olika sammanhang. Hennes resultat stödjer ett pragmatiskt betraktande av relationen mellan dator och användare. Turkle menar att datorn, eller snarare människors förhållande till densamma, går att beskriva i olika dimensioner. Först och främst är datorn ett verktyg som understödjer oss vid olika arbetsuppgifter. Viktigare än det är dess funktion som ett medium för projektion av ens föreställningar och fantasier. På senare tid har datorn också börjat erbjuda ännu en dimension, nämligen en virtuell värld – Cyberspace.

I en första omfattande studie från början av 1980-talet ger Turkle en bild av datorn främst som ett projektningsmedium, där en mångfald av förbindelser möjliggörs. Den kan fungera som en projektion av olika delar av sig själv och som en spegel för ens tankar. I motsatt riktning bildar datorn, dess egenskaper och funktioner, mall för reflektion över sig själv och andra. Människor tenderar börja prata om sina problem och handlingar i datortermer. Vidare betraktas och definieras datorn på olika sätt av olika ålders- och användargrupper och får följaktligen därför annorlunda betydelser för olika grupper av människor (Turkle, 1986). Datorprogram som uppvisar någon form av reaktionsförmåga kombinerat med ett naturligt språk får många människor att överskatta datorns intelligens. Särskilt går detta att iakttä i samband med datorprogram för samtalsterapi som exempelvis Eliza. Användarna tillskriver programmet intentioner och komplexitet och behandlar det som en mänsklig samtalspartner (Turkle, 1998:159;194f).

Turkle beskriver även mer specifikt människors förhållande till artificiella, intelligenta agenter. Hon ser intelligenta artefakter, här i form av agenter, som ännu en i den långa raden av kränkningar mot människans självbild som universums härskare, där hon har en upphöjd ställning i förhållande till alla andra väsen. Hon drar paralleller till Kopernikus och Darwins revolutionära upptäckter. Turkle förutspår därför att agenterna till en början kommer mötas av motstånd från användarna. Men på sikt kommer troligen en anpassningsprocess inträda och agenterna långsamt accepteras av de mänskliga användarna (Turkle, 1998:158f).

3.3 Interaktion mellan mänskliga aktörer och artificiella agenter

Jag tänker i det följande knyta ihop analysen på mikronivå. Teoretiskt sett tänker jag göra ett försök att koppla samman agentens handlingsförmåga och den mänskliga aktörens uppfattning av densamma. Dessa båda aktörer möts och handlar gentemot varandra i olika interaktionssammanhang. Just interaktionsbegreppet fungerar som ett sammanfattande begrepp för min mikroanalys. Som vi tidigare uppmärksammat betraktas inom symbolisk interaktionism handlingen vara beroende av interaktion med omgivningen. Vidare pekar Turkles studier på att även acceptansen av och förtroendet för agenten avgörs i interaktionsprocesser.

Sett från ett pragmatiskt perspektiv har, enligt Hans Geser, de konstaterade tendenserna att läsa in mänskliga egenskaper i datorn konsekvenser även för datorns roll i interaktionen. Geser menar att relationen mellan PC och användare i allt större utsträckning tenderar att bli social. Datorn betraktas som ett subjekt och uppnår därmed deltagarstatus i ett ”intersubjektivt växelförhållande” – interaktionen. Datorn bör därmed inte längre ses som en utomstående påverkande faktor eller ett objekt. Datorn är snarare en aktiv deltagare och ett subjekt i sociala processer (Geser, 1989).

Agentteknikens utveckling och dess följder för människa-dator-interaktion ger ytterligare underlag för Gesers hållning. Patti Maes jämför människors interaktion med datorer respektive autonoma agenter. Den sistnämnda bygger på *indirect management*. D v s istället för att interaktionen, som tidigare, initieras av användaren m h a kommandon och direkt manipulation, så är användaren numera del av kooperativa processer där såväl mänskliga som artificiella agenter initierar kommunikation, kontrollerar händelser och utför uppgifter. Hon använder metaforen ”personlig assistent”, vilket implicerar att agenter samarbetar med användaren i en gemensam arbetsmiljö (Maes, 1994). Frågan i vilken utsträckning agenter är att jämföra med en mänsklig interaktionspartner kvarstår dock. Låt oss ta utgångspunkt i sociologisk teori.

3.3.1 Det sociologiska interaktionsbegreppet

Schülein ser en klar koppling mellan handling och interaktion. Han definierar interaktion som system av handlingar som är relaterade till varandra.³ Mellan handling och interaktion råder ett dubbelriktat konstitutionsförhållande. Å ena sidan utgör interaktionen sammanhanget där handlingar får mening. Å andra sidan uppkommer interaktion först i samband med handling (Schülein, 1983:85). Inom symbolisk interaktionism poängterar man att interaktion bygger på att människor orienterar sina handlingar mot andra aktörer. ”*Interaction is mutual social action: action where individuals take each other into account when they act and are thus at least partially affected by each other*” (Charon, 1979:130).

Mead menar att social interaktion utspelar sig på två nivåer. Han gör en åtskillnad mellan icke-symbolisk och symbolisk interaktion. Vid icke-symbolisk interaktion framkallar en aktörs gester eller handlingar direkt respons hos andra aktörer. Symbolisk interaktion innebär att aktörerna tolkar varandras gester och handlingar samt förmedlar indikationer till andra aktörer hur de tänkt handla. Symbolisk interaktion innebär alltså en aktiv ”skapandeprocess”, där deltagarna bygger upp och reformerar sina handlingslinjer på grundval av upprepad tolkning av andra aktörers handlingar. Meads synsätt innebär att några egentliga handlingsmönster eller strukturer ej existerar. Istället kan man tala om en kontinuerlig användning av tolkningssystem. Deltagarna skapar sina handlingar utifrån tolkningar och definitioner av andras handlingar (Blumer, 1969:65ff). Inom symbolisk interaktionism lägger man – som framgår av namnet – tonvikten vid den symboliska sidan av interaktionen. Människor handlar i relation till varandra. Merparten av de mänskliga handlingarna är sociala och symboliska och merparten av de sociala handlingarna är interaktion. En viktig del av interaktionen utgår från hur aktörerna definierar situationen (Charon, 1979:130ff). En idealtypisk interaktionssituation skulle kunna beskrivas i ett antal steg. En aktör hamnar i, *enters*, en situation. Aktören definierar situationen för sig själv genom att sätta upp mål, ta andras roller och jämföra med tidigare erfarenheter. Aktören bestämmer därefter en *line of action* (handlingslinje), d v s hur hon/han tänker handla mot objekt och aktörer i situationen. Utifrån sin handlingslinje handlar aktören socialt. Aktörens handling sätter igång liknande processer hos andra aktörer. De i sin tur tolkar handlingen, definierar situationen, bestämmer sig för en handlingslinje och utför en (social) handling. ”Utgångsaktören” tolkar därefter dels

³ ”Systemen von Handlungen, die aufeinander bezogen sind” (Schülein, 1983:85).

egna handlingar i ljuset av andras handlingar och dels andras handlingar och dess betydelser. I denna fas har rollövertagande en stor betydelse. I ett sista steg reviderar aktören sina perspektiv, definition av situation och handlingslinjer (jfr Charon, 1979:135).

3.3.2 Artificiella agenter och interaktion

Reflektiva agenter går att dela upp i två kategorier. Intentionella agenter av första ordningen besitter övertygelser och önskningar (belief & desire) som fungerar som guide i olika situationer. Intentionella agenter av andra ordningen har såväl övertygelser och önskningar om egnas som om andras övertygelser och önskningar. Schulz-Schaeffer gör en uppdelning på samma premisser och skiljer mellan intentionella och interaktionistiska agenter (Wooldridge och Jennings, 1995; Schulz-Schaeffer, 1998).

Intentionella agenter av första ordningen är liktydiga med de reflektiva agenter vi talat om hittills. Denna typ av agenter är förvisso kapabla att delta i en symbolisk interaktionsprocess, men de är oförmögna att förändra och anpassa sina egenskaper efter de andra deltagarnas handlingar i interaktionssituationen. De från början fastställda egenskaperna (BDI) innebär att dess handlingslinjer är statiska och förändras således ej i interaktionsprocessen (Scheuermann, 2000:29f).

De interaktionistiska agenterna stämmer bättre överens med den interaktionsdeltagare som modelleras av företrädare för symbolisk interaktionism och har faktiskt också hämtat viss inspiration från SI. Agenternas egenskaper och trosuppfattningar uppkommer och utvecklas här i interaktion med andra agenter och även mänskliga aktörer (Schulz-Schaeffer, 1998:133ff). Dessa agenter har också en förmåga att skapa bilder av andra agenter genom att ta över deras roller och på så sätt skapa en orientering för sina handlingar (Scheuermann, 2000:28).

Vid en första anblick är det svårt att se hur reaktiva agenter skulle kunna vara kapabla till en jämbördig interaktion med mänskliga användare. De uppfyller knappast de krav som ställs på aktören i interaktionssituationen enligt ett symbolisk interaktionistiskt synsätt. Dock har lyckosamma försök gjorts där en reaktiv modell kombinerats med ett interaktionistiskt synsätt. Maes tidigare nämnda assistentagenter kan kategoriseras in i denna grupp. Agenten skapas här för utföra en viss funktion, exempelvis handha mötesbokning, och denna funktion

förutsätter att agenten lär sig en viss roll. Precis på samma sätt som en ”mänsklig” assistent behöver en viss tid för att bli bekant med arbetsgivarens och organisationens normer och funktionssätt, måste agenten genomgå en tillvänjnings-/lärandeprocess innan den är kompetent nog att bistå sin uppdragsgivare i interaktiva problemlösningsprocesser. Att tala om en socialisationsprocess i agentens fall är kanske något överdrivet. Snarare rör det sig om en habitueringsprocess där agenten lär sig vissa rutinartade handlingssekvenser. Denna lärandeprocess visar upp ett interaktivt mönster. Agenten lär beteende – man kan knappast tala om handlande i fallet med assistentagenter – av användaren och andra agenter. Agenten studerar hur användaren utför uppgifter och härmar beteendet. Vidare kan agenten ändra sitt beteende genom direkt och indirekt feedback från användaren, genom att ta emot instruktioner i form av exempel från sin ”uppdragsgivare” samt slutligen genom att fråga andra agenter om råd (Maes, 1994; Braun, 1998).

4 Integration i komplexa organisationer – ett makroperspektiv

På makronivå, inom en organisation, uppstår i samband med skapandet och införandet av MAS en dubbel koordinationsproblematik. Dels gäller det att få till stånd ett fungerande samarbete mellan – de åtminstone semi-autonoma – agenterna. Därutöver måste också arbets- och rollfördelningen mellan agenter och människor fungera på ett tillfredsställande sätt. Det är främst den senare problematiken jag kommer att avhandla här.

4.1 Organisationer och komplexitet

Inom socioniken utgår man från att organisationer är komplexa. Organisationerna kännetecknas av brist på inre sammanhållning, skillnader mellan olika avdelningars preferenser och handlingsmönster samt en inkonsistens vad det gäller den sammantagna organisationens mål och rationalitetsbegrepp (Buckhard och Rammert, 2000; Schulz-Schaeffer, 2000a).

De socioniska antagandena stödjer sig på de inriktningar inom sociologisk organisationsforskning som problematiserat rationaliteten och sammanhållningen inom organisationer.⁴ Grovt sett tar sig komplexiteten uttryck i konflikter på två nivåer. Dels återkommer den i form av inkonsistens i individens tänkande, som gör att hon/han har problem att göra val och fatta beslut på rationell basis. Vidare råder det en konflikt mellan olika grupper och individer med inkonsistenta mål och preferenser (jfr March & Simon, 1958:135).

Herbert A Simon betraktar organisationer som öppna system bestående av begränsat rationella handlingar och beslut. Synsättet baseras på ett antal antaganden om organisationen. Enligt ett första antagande kännetecknas aktörernas handlande av begränsad rationalitet. Vidare är informationen och kunskapen som ligger till grund för beslut i regel ofullständig. Det antas också föreliggande svårigheter att förutsäga värdet eller nyttan av framtida händelser. Slutligen anses det omöjligt att i en beslutssituation ta alla föreliggande handlingsalternativ i beaktande.

⁴ Dessa komplexitetsorienterade inriktningar inom organisationssociologi kategoriseras på tyska under begreppet *Verhaltenswissenschaftliche Entscheidungstheorie* (Berger & Bernhard-Mehlich, 1999). Jag har inte kunnat hitta något motsvarande svenskt begrepp.

Handlingsval utgår snarare från ett begränsat urval av handlingsalternativ (Simon, 1961:80ff). Beslutssituationer kännetecknas vidare av en påverkan från en komplex och föränderlig omgivning samt att beslutsfattarna, situationens subjekt, enbart besitter en begränsad kognitiv förmåga (Berger & Bernhard-Mehlich, 1999:133f).

Chester I Barnard betraktar en organisation som ett öppet system av medvetet koordinerade handlingar. Handlingarna utförs av olikartade deltagare med olika intressen. Fokus ligger på handlingarna i sig och aktörerna – de som utför handlingarna – ses snarare som miljöfaktorer än en del av organisationen. Organisationerna framstår därigenom som opersonliga och enbart bestående av de utförda handlingarna (Berger & Bernhard-Mehlich, 1999:134ff).

Inkonsistensen mellan de olika deltagarnas preferenser och mål gör det svårt att enas om gemensamma handlingsmål. Målbildningsprocessen är en konfliktladdad process där deltagarna m h a förhandlingar försöker nå fram till ett gemensamt mål (Berger & Bernhard-Mehlich, 1999:145ff).

Liknande organisationsmodeller har sedan länge också figurerat i samband med DAI. De benämns här ”öppna system” (Hewitt, 1986; Gasser, 1991). Mänskliga organisationer betraktas även här som sammansatta av olika intressegrupper och olika rationalitetsbegrepp. Informationen som ligger till grund för beslut inom organisationer är i regel konfliktladdad och osammanhållen. Med inspiration från SI hävdar Gasser vidare att kunskapen i öppna system alltid är lokal och situationsbunden (Gasser, 1991:116). Vidare kännetecknas öppna system av egenskaper som samverkan mellan flera komponenter (databaser, nätverk etc), asynkronitet (i meningen att omgivningens beteende ej är möjligt att förutsäga samt att komponenterna är fysiskt åtskilda), lokala maktbefogenheter, arbetsdelning, specialisering samt att kommunikationen sker på ett armlängds avstånd (Hewitt, 1991:81f; Hewitt, 1986:272ff). Strübing tycker sig i de öppna systemen kunna se en organisationsform som påminner om den som förekommer inom vissa delar av SI, särskilt *the ecological approach*. Organisationer (och sociala aggregationsformer i allmänhet) betraktas då – från ett aktörsperspektiv – som variabla miljöer. De bildas och formas i interaktionskontexter. Sociala strukturer är enbart siktbara i interaktionen agenter emellan och därmed endast möjliga att förändra genom socialt handlande. Organisationer framstår därmed som arenor för tolkningar av problem och skeenden samt för förhandlingar om problemlösningstrategier (Strübing, 1998). Denna bild påminner om Hewitts: ”*Whatever consistency exists among the beliefs within an organization is constructed and negotiated by the participants*” (Hewitt, 1986:274).

Med utgångspunkt i antagandet om organisationernas komplexitet går det att spåra en kritik mot det rådande sättet att modellera datoriserade informations- och beslutsstödssystem. En stor del av de datoriserade informationssystem som återfinns inom organisationer idag är baserade på en hierarkisk logik. Ofta rör det sig om globala och centralstyrda system, exempelvis expertsystem, byggda på antagandet om en universell rationalitet inom organisationen (Gasser, 1986:212ff). Följden har ofta blivit att människorna och organisationerna tvingats anpassa sig och sitt handlande till informationssystemets struktur snarare än tvärtom. Risken är då att de sociala mekanismer inom organisationen som handskas med inkoherens – såsom förhandlingar och koalitionsbildningar etc – spelas ut (Schulz-Schaeffer, 2000a).

Les Gasser har studerat detta fenomen närmare och redovisar sina resultat i en artikel (Gasser, 1986). Han pekar på att de arbetsprocesser inom organisationen som datorsystemen är utformade för att understödja ofta i hög grad är oförutsägbara. Skälen därtill hittar han i organisationernas komplexitet. Resurserna liksom deltagarnas rationalitet är begränsade. Informationen inom organisationen är ojämnt fördelad. Gasser menar vidare att betydelser och tolkningar av olika organisatoriska processer skapas i interaktionen med medlemmarna emellan.

För att kunna använda sig av centraliserade datorsystem för sina egna ändamål måste användarna därför överbrygga ett flertal så kallade *computer slips*. De tvingas tillämpa anpassningsstrategier av olika slag. Det kan handla om att förändra arbetsstrukturen så att den passar samman med systemet. I andra fall tvingas användare ta till *augmenting work*, extraarbete, för att med hjälp av datorsystemet lösa sina arbetsuppgifter. En tredje strategi går ut på att lura systemet genom att använda det på ett annat sätt än det är tänkt att användas. Utan dessa strategier, menar Gasser, skulle många centralistiska datorsystem troligen inte fungera alls.

Det finns allmänt sett olika vägar att handskas med komplexiteten inom organisationer. En möjlighet är att anpassa målbildnings- och beslutsprocessen till den komplexa omgivningen och på så sätt nöja sig med tillfredsställande snarare än optimala lösningar. Det talas här om en justering av anspråksnivån. En annan strategi riktar in sig på att i olika avseenden försöka stabilisera omgivningen/miljöfaktorerna. Ett tredje sätt att handskas med komplexiteten går ut

på att fördela vissa målsättnings- och beslutskompetenser till bestämda subgrupper (Berger & Bernhard-Mehlich, 1999:145ff).

I fallet med agentbaserade besluts-/problemlösningssystem som understöd för komplexa organisationer lutar man sig snarast åt det senare alternativet, d v s en distribution av problemlösning-/informationshanteringsprocessen. Ett MAS fungerar då som en resurs i olika komplexa organisatoriska processer i form av beslutsstöd, informationshantering och kontrollfunktioner. Varje intresse-/subgrupp representeras av en agent. MAS kan sägas fungera som en parallell organisation där agenterna interagerar (Schulz-Schaeffer, 2000a).

Hewitt har arbetat ut ett teoretiskt underlag för att kunna modellera agentsystem enligt denna logik (Hewitt, 1991). Han menar att man i komplexa, öppna organisationer når fram till beslut genom olika former av förhandlings- och argumentationsprocesser, s k *due processes*. I dessa *due processes* möts olika deltagare med olika preferenser och intressen för att enas om ett gemensamt beslut. Hewitt hävdar vidare att deltagarnas preferenser går att modellera matematiskt i s k ”mikroteorier”. Mikroteorierna är internt konsistenta och uppbyggda enligt logisk deduktion. Varje aktör representeras av en mikroteori, vilken kan modelleras och implementeras i en agent.

Det finns dock anledning att förhålla sig kritisk till denna modell. Frågan är om det är möjligt att representera grupper eller ens individer med hjälp av konsistenta mikroteorier. Den organisationsteoretiska utgångspunkten är ju att det råder en inkonsistens såväl vad det gäller subgruppernas som individens målsättningar och preferenser. Det torde alltså vara ganska svårt att formalisera aktörernas preferenser. Vidare föreligger det också problem i samband med representationsfrågan. Frågan är vilka subgrupper och individer som tillmäts aktörstatus och därmed gjort sig förtjänta av representation.

4.2 Artificiella agenter roll i mänskliga organisationer

Från ett tekniksociologiskt perspektiv ligger tyngdpunkten snarare på agentens roll och funktion i organisationer samt dess förhållande till människor och mänskliga kontexter. Framväxten av ”intelligent teknik”, exempelvis i form av agenter, har fått en del forskare att dra radikala slutsatser om följderna för mänskliga organisationer och kontexter. Parunak tycker sig se ett paradigmskifte där den nya teknikens artefakter blir en form av kollegor till de mänskliga aktörerna inom organisationen (Parunak, 1996:151). Liknande, men än mer radikala visioner återkommer i cyberpunkromaner och diverse science-fictionfilmer. I det följande tänker jag utifrån tekniksociologisk teori betrakta agenternas roll och funktion i mänskliga organisationer.

Inom klassisk sociologisk teori har i regel tekniska artefakter setts som miljöfaktorer som – på ett negativt eller positivt sätt – påverkat sociala aktörer. Vidare har artefakterna också betraktats som ”svarta lådor” besjälade med vissa inneboende egenskaper. Tekniken har utifrån dessa premisser utövat någon form av deterministiskt tvång på människorna. En sådan teknikuppfattning går bl a att finna hos företrädare för Frankfurtskolan och inom tidig sociologisk teknikforskning (Rammert, 2000). På senare tid har det växt fram ett antal riktningar inom sociologin där teknikens roll i organisationer samt människors och tekniska artefacters förhållande till varandra omvärderats. Å ena sidan har tekniska element tillmätts en mer aktiv betydelse än tidigare. Synen på teknik har växlat från ett manipulativt objekt till en integrerad aktör. Å andra sidan läggs tonvikten vid hur teknikens funktion och betydelse förhandlas fram i olika sociala kontexter.

4.2.1 Agenter som enheter i komplexa sociotekniska system

En första representant för denna omvärdering av teknikens roll har jag funnit i organisationssociologen Charles Perrow. Hans analytiska angreppssätt där han betraktar teknik utifrån ett systemperspektiv är intressant för min frågeställning av två skäl. Först och främst gör Perrow ingen strikt åtskillnad mellan mänskliga aktörer och tekniska artefakter när han analyserar organisatoriska processer. För det andra är det just komplexa system eller organisationer, där MAS ofta är förekommande, som är Perrows studieobjekt.

Enligt Perrows teori är det varken den mänskliga organisationen eller en viss teknik, exempelvis MAS, som står i centrum för analysen. Ett system ses snarare som en kombination av flera organisationer och tekniska system som samverkar i en gemensam process. Systemet går att dela upp i ett antal nivåer. Ett system består av flera subsystem. Ett subsystem i sin tur är sammansatt av flera enheter som samverkar. Enheterna är slutligen sammansatta av en uppsättning delar (Perrow, 1987:99). Ett MAS skulle man kunna tänka sig som ett subsystem inom ett större, övergripande system. Dess heterogena agenter fungerar i sådant fall som subsystemets enheter. En agent har i sin tur en rad olika delar, exempelvis i form av handlings- och kommunikationskomponenter. Detta synsätt innebär en upplösning av gränserna mellan tekniska artefakter – som ej nödvändigtvis behöver vara intelligenta eller handlingsförmögna – och människor inom en organisation.

Ett systems beteende är enligt Perrow snarare beroende av relationerna mellan de heterogena komponenterna/enheterna än de enskilda aktörernas/agenternas handlingar.

Subsystemen och enheterna är sammankopplade och interagerar med varandra. Perrow gör en åtskillnad mellan linjär och komplex interaktivitet. Den komplexa interaktiviteten ökar med antalet enheter eller komponenter i ett system. Många sammankopplade komponenter får till följd att många interaktioner sker samtidigt i systemet. Därigenom skapas ömsesidiga beroenden och olika sammankopplade processer. Följden blir en mängd återkopplingsprocesser vars effekter är svåra att förutsäga och som därför gör ett system ytterst komplext (Perrow, 1987:114f, 128f). Exempel på system där interaktiviteten är komplex återfinns bl a i högteknologiska system som kärnkraftverk men även organisationer som universitet och ministärer uppvisar komplex interaktivitet (Perrow, 1987:138).

Vidare skiljer Perrow också mellan tät och lös koppling enheterna emellan. Om enheterna och subsystemen är löst kopplade till varandra existerar det olika former av buffertar och tröghetsfaktorer dem emellan. Dessa ger tid och utrymme att åtgärda uppkomna oönskade effekter. För att återvända till våra exempel kännetecknas i det här fallet kärnkraftverk i regel av tät koppling mellan enheter och subsystem medan kopplingen inom universiteten snarare är lös (Perrow, 1986:148ff; Perrow, 1987:131ff).

4.2.2 Tekniska system som sociala konstruktioner

I direkt polemik med det tidigare nämnda teknikdeterministiska perspektivet har en socialkonstruktivistisk riktning växt fram. Den är inriktad på att rekonstruera den gällande bilden av tekniska utvecklingsprocesser. Riktningen innebär ett avsteg från den utbredda myten om en ensam genial uppfinnare som plötsligt får en lysande idé som han mer eller mindre på egen hand lyckas genomföra. Den socialkonstruktivistiska utvecklingsprocessen är snarare en icke-linjär process bestående av många små utvecklingssteg och upptäckter som sammanfogas till en slutlig produkt. Riktningen innebär också en kritik mot bilden av att den artefakt som visar sig vara tekniskt och ekonomiskt mest rationell mynnar ut i den slutgiltiga produkten. Utvecklingsprocessen är enligt socialkonstruktivismen en maktkamp mellan olika aktörer med olika perspektiv och rationalitetsbegrepp. Riktningen kan ses som ett försök att öppna den "svarta lådan" och se vad som finns inuti (Bijker, Hughes & Pinch, 1987; Rammert, 2000).

Thomas P Hughes företräder den socialkonstruktivistiska synen och definierar tekniska system som komplexa, problemlösande komponenter som samverkar i en interaktiv process. Tekniska system är å ena sidan socialt konstruerade men utövar samtidigt också ett inflytande på det omgivande samhället/organisationen. Systemets komponenter kan utgöras av såväl fysiska artefakter som förordningar och organisationer. Komponenterna interagerar med varandra för att uppnå ett gemensamt systemmål. Även människor räknas till komponenterna. Hughes gör dock en åtskillnad mellan artefakter och människor. Människorna har till skillnad från artefakterna en viss grad av handlingsfrihet (Hughes, 1987). Precis som är fallet med Perrows komplexa system och i aktör-nätverksteorin kan även artefakter – exempelvis agenter – utöva kontroll och påverkan på systemet. "*Technological systems are bounded by the limits of control exercised by artifactual and human operators*" (Hughes, 1987:54).

Ett socialkonstruktivistiskt angreppssätt torde innebära att själva utvecklingsprocessen av DAI och MAS ägnas större uppmärksamhet. Agenternas roll och funktion bestäms genom sociala "förhandlingar" delvis i utvecklingsprocessen men också vidare i organisationen där de implementeras. Med socionikens inträde är en utvidgad utvecklingsprocess att förvänta. Sociologin – i den mån den kan agera sammanhållet – framträder som en ny aktör i utvecklingsprocessen av DAI och agentsystem. Vidare är också agentteknikens funktion och

roll i den organisatoriska kontexten beroende av hur de mänskliga användarna och organisationsmedlemmarna upplever och definierar agenterna (jfr kap 3.2).

4.2.3 Agenter som aktör-nätverk

Aktör-nätverksteorin (ANT) är en annan riktning som omvärderat teknikens roll och dess förhållande till människor. Tekniska artefakter betraktas här inte längre som passiva objekt utan som ”medverkande” agenter. Riktningen har växt fram som en kritik mot den klassiska analytiska uppdelningen av samhälle och teknik och åtskillnaden mellan subjekt och objekt. Motivet bakom ANT är att skapa en mer omfattande förklaringsmodell, i vilken tekniken inte betraktas som en utomliggande kraft som påverkar samhället och dess aktörer. Snarare är såväl teknik som samhällets aktörer delar av en integrerad interaktionsprocess, där de ömsesidigt påverkar såväl varandra som de uppkomna handlingsstrukturerna (Rammert, 2000:48).

En agent definieras av Michel Callon – en av riktningens mest kända företrädare – som en förmedlare eller ett mellanled⁵ som kan sätta andra mellanled i rörelse. Agenten fungerar som en entitet som kan associera andra entiteter. Agenten eller aktören går alltså inte att skilja från sina kopplingar till andra entiteter och är således snarare att betrakta som ett nätverk än en fristående enhet. Aktörerna som nätverk är i sin tur kopplade till mer omfattande nätverk (Callon, 1991). Den metodiska följden därav blir att sociala agenter och maskiner inte kan analyseras fristående från sina kopplingar till omvärlden och andra aktörer. Agenten analyseras som ett nätverk av heterogena relationer. ANT är på så sätt relations- och processorienterad. Agenter, organisationer och redskap/teknik ses som interaktiva effekter (Law, 1991).

En aktör benämns inom ANT oftast aktant – ett begrepp hämtat från språkvetenskapen. Med en aktant avses en potentiell handlingsenhet, oavsett om det råkar vara en människa eller en teknisk artefakt (Braun, 2000:8). Med aktantbegreppet vill man påvisa att det råder symmetri mellan natur och samhälle, mellan mänskliga och icke-mänskliga entiteter i handlingsprocessen. Såväl människor som teknik och naturliga entiteter bidrar till de

⁵ Det engelska begreppet är *intermediary*

uppkomna handlingsstrukturerna. Såväl människor som teknik är aktiva delar i samhälleliga integrationsprocesser.

Aktör-nätverken hålls samman med hjälp av "översättningar" och mellanled/förmedlare. En översättning är en (re-)definition av någon entitets identitet, egenskaper eller beteende för att frambringe stabila associationer av entiteter, dvs nätverk (Schulz-Schaeffer, 1998:140). Relationerna mellan aktanterna i nätverket är dubbelriktade. Det betyder att en aktant definierar och översätter andra entiteter/aktanter men är likaledes ett resultat av andra aktanters översättningar. ANT uppvisar härigenom ett interaktionistiskt synsätt genom att hävda att en agents egenskaper uppkommer i interaktion med nätverk(-sförhållanden) (Schulz-Schaeffer, 1998:156f). En agent kan vara kopplad till flera olika nätverk. Nätverken är dynamiska och förändras med tiden genom associationer och nya översättningar. Olika former av förmedlare eller mellanled definierar relationerna mellan aktörerna, som i själva verket också kan vara förmedlare. Förmedlare kan ta sig uttryck i såväl texter och pengar som människor och tekniska artefakter (Callon, 1991).

ANT fungerar som en opposition mot det mainstreambegrepp av aktör som vi stött på i samband med diskussionen om handling och handlingsförmåga. Denna handlingsbaserade aktör definieras utifrån sin intentionalitet och förmåga att kunna välja mellan olika handlingsalternativ i en bestämd situation. Handlingsbegreppet inom ANT ser helt annorlunda ut. Handlingar uppstår när människor och icke-mänskliga artefakter stöter på varandra. Handlingarna är emergenta storheter som uppstår genom förbindelser av flera aktanter. De är sås produkter av associationer mellan mänskliga och icke-mänskliga element (Braun, 2000). ANT intresserar sig alltså snarare för förutsättningarna eller associationerna som krävs för att en viss handling eller ett visst tillstånd skall uppkomma än hur varje enskild aktörs handlingar ser ut.

Olika former av aktanter har olika roller. Tekniska artefakter och fysiska objekt bidrar i regel till att stabilisera relationerna i nätverken. De försöker framtvinga att vissa regler och sociala konventioner efterlevs. Aktanter kan besitta olika former av integrations- eller moralmekanismer. Tekniska objekt försöker utöva ett sorts tvång på individer genom att skapa ett motstånd i en viss situation vilket bidrar till att påverka och begränsa individens handlingsrum. Bruno Latour betraktar teknik som inkapslade förmedlingar mellan mänskliga och icke-mänskliga element (Rammert, 2000:48). Teknik kännetecknas vidare av en social

asymmetri. Den skapar system som ”stänger av” vissa (handlings-)alternativ men skapar samtidigt nya och ibland oväntade (handlings-)alternativ (Callon, 1991). Den socialintegrativa moralen – reglerna och normerna som krävs för att ett visst nätverk skall fungera – är så fördelad mellan människor och objekt. Moralerna efterföljs antingen genom människors självdisciplin eller genom att vara inkorporerad i tekniken (Braun, 2000:13).

För att kunna förstå sociala processer måste man alltså ta objektens socialintegrativa funktion i beaktande. Objekt möjliggör och påverkar olika sociala processer. Latour illustrerar resonemanget med ”det berlinska låset” som exempel. I Berlin finns det en särskild sorts låsordning i vissa hyreshus där man måste skjuta nyckeln igenom låset och låsa dörren från insidan. På så sätt är det omöjligt att lämna dörren på glänt, vilket medför att inga objudna (eller ens inbjudna) gäster kan leta sig in utan att få hjälp av någon av bostadsgästerna (Latour, 1994). I andra fall kan tekniska artefakter ta över roller i ett nätverk som tidigare utförts eller borde utföras av mänskliga aktörer. Exempelvis fyller den automatiska dörrstängaren en viktig roll i ”nätverket husingång”. Den ser till att dörren stängs efter det att någon människa passerat in eller ut genom ingången och bidrar på så sätt till att hålla ute kylan (Latour, 1988).

ANT innebär alltså en rekonstruktion av begreppet handlingsförmåga. Även objekt utan intentionalitet kan enligt ANT besitta handlingsförmåga. Objekt definieras som aktörer om de bidrar med något till en process. Vidare är en aktör/handling beroende av den organisatoriska kontexten/nätverket och går därför ej att skilja från densamma (Braun, 2000). Samhällets sammanhållning och processer är i stor utsträckning beroende av materiella artefakter. Latour ser samhället som mer än summan av sina sociala delar. Han menar *”(a)tt få samhället att hänga ihop med endast sociala element är som att försöka göra majonnäs utan vare sig ägg eller olja – dvs med bara snack”* (Latour, 1998:55). På samma sätt skulle man kunna tänka sig att mänskliga organisationer är beroende av artificiella agenter och andra tekniska artefakter för sin sammanhållning. Artefakterna förstärker och stabiliserar relationerna inom nätverken. De materialiserar konventioner och normer och tvingar fram beteende i enlighet med dessa.

Schulz-Schaeffer anger tre skäl till varför ANT skulle kunna utgöra en relevant teoretisk grund för en tekniksociologisk analys av MAS. För det första görs inom ANT, liksom vid agentorienterade riktningar inom DAI, ingen kategorisk skillnad mellan tekniska artefakter

och sociala aktörer med avseende på handlingsförmåga. Alla element vars beteende bidrar till sammanhangets/nätverkets funktionssätt och därmed till handlingars ”uppkomst” tillmäts aktantstatus. För det andra utesluter de tre agentbegreppen (reaktiv, intentionell och interaktionistisk) inte varandra när en organisation betraktas utifrån ett ANT-perspektiv. De motsvarar istället olika processer som bidrar till nätverksbyggande. Slutligen pekar ANT på att ett MAS endast utgör en del av ett större problemlösande nätverk (Schulz-Schaeffer, 1998:137f).

Handlings-/Agentbegreppet liknar det som återfinns inom vissa SI-riktningar. Gemensamt är att aktörstatusen utgår från deltagandet i problemlösningssammanhanget. Den sociala tillhörigheten till en sådan kontext styrs inte av fasta, formella medlemsregler utan utgår från problem- och processberoende *commitments*. Det betyder att koncentrationen ligger på huruvida en viss aktör, eller agent för den delen, uppfyller vissa förpliktelser gentemot en kontext (Strübing, 1998:65ff). Enligt ett sådant synsätt kan såväl människor som datorer betraktas som en integrerad del av en arbetsprocess eller produktionskedja (Gasser, 1986:223).

4.2.4 Hybrida system – ett organisatoriskt Cyberspace?

Såväl Perrow som föreläsarna för ANT ser alltså tekniken som en integrerad del i sociala system. Tekniska artefakter är enligt dem i lika stor utsträckning som människor medverkande aktörer i en organisation. Den strikta åtskillnaden mellan artificiell agent och mänsklig aktör har enligt dem spelat ut sin roll och är inte längre analytiskt relevant. En rumslig kontext där det inte längre anses relevant, eller ens möjligt, att skilja mänskliga från datorframställda aktörer brukar benämnas Cyberspace. Cyberspace är ett modebegrepp som återkommit allt oftare i såväl populär som vetenskaplig litteratur under senare år för att beskriva kontexter där dator- och informationsteknik upplöst rumsliga gränser och samtidigt gjort reella och artificiella aktörer allt svårare att skilja åt. Jag finner därför att det är på sin plats att i en kort exkurs kommentera förhållandet mellan MAS/hybrida system och Cyberspacebegreppet.

De sociologiska diskussioner och teoribildningar som tar sin utgångspunkt i Cyberspacebegreppet återfinns framförallt inom kultursociologi och Cultural Studies. Begreppet Cyberspace myntades av science-fictionförfattaren William Gibson och är en sammanskrivning av cybernetiskt rum. Rummet är ej geometriskt mätbart och kännetecknas vidare av en dimensionslös tid. Cyberspace är alltså m a o ”ett platslöst rum av tidlösa

dataflöden” (Geier, 1999:276). Ett begrepp som brukar sammanblandas med cyberspace är Virtual Reality (VR). VR framställs i regel som något radikalare än cyberspace. Med VR förstår Achim Bühl en teknik som skapar möjligheter för människor att integreras omedelbart i datorgenererade, tredimensionella miljöer. Användaren upplever miljön genom flera sinnen och hennes/hans handlande sker i realtid. Hon/han kan också röra sig i och interagera med omgivningen och andra aktörer som befinner sig i rummet (Bühl, 1997:49f). De normala sensoriska retningarna ersätts här med datorproducerade artificiella retningar. Ett kriterium är att den datorgenererade, virtuella verkligheten ej skall vara möjlig att skilja från den ”reella verkligheten”. De virtuella objekten är dock varken illusoriska eller fiktiva utan går att faktiskt manipulera genom ingripande från aktörerna (Esposito, 1995:187ff).

Cyberspace används numera inte enbart i denna strikta betydelse. Cyberspacebegreppet har snarare orienterat sig mot en betydelse av ett sorts ”kroppslöst icke-rum” (Geier, 1999:277). Rammert beskriver Cyberspace som ett artificiellt erfarenhetsrum med visualiserade datastrukturer och agenter som konstrueras genom ett samspel mellan datorer, program, nätverk och mänskliga handlingar (Rammert, 2000:116f). Denna senare form av Cyberspace finner sitt mest kända exempel i Internet och dess grafiska användargränssnitt World Wide Web. Cyberspace är så ingen alternativ verklighet utan ett virtuellt rum som existerar i verkligheten (Geier, 1999:279).

Ett Cyberspace i form av Internet kan betraktas som ett hybrid system eller rum befolkat av såväl artificiella agenter som människor (jfr Geier, 1999:20). På Internet är troligen sökmaskinerna den mest utbredda formen av artificiella agenter, men även agenter för t ex informationsfiltrering går att finna. Agenterna handlar på Internet och i liknande virtuella miljöer som representanter för mänskliga användare (Bühl, 1996:182). Agenterna arbetar i det dolda och dess handlingar är därför svåra att följa. De visar sig bara när tjänster erbjuds eller förfrågningar görs. Om vissa handlingar utförs av agenter eller människor är ofta svårt att avgöra.

Ett närbesläktat begrepp som fått mycket genomslag i den postmoderna debatten är Donna Haraways cyborgbegrepp. I begreppet cyborg låter Haraway olika, tidigare strikt åtskilda kategorier, såsom människa-djur och teknik-natur, smälta samman. Cyborg är en förkortning av cybernetisk organism. Begreppet kan definieras som en hybrid av människa och maskin. En cyborg besitter såväl levd verklighet som fiktion (Haraway, 1995:36f). En aktör i ett

hybrid system fungerar också som en sorts sammansmältning av människa och maskin. En person kan ha många alter ego i form av agenter (Maes, 1995). I hybrida organisationer är tanken att användaren/gruppen och dess agenter skall integreras till en aktör. Agenterna är tänkta att fungera som en förlängd men integrerad del av sin användare.

5 En metodisk reflektion kring integrationsproblematiken

I min uppsats har jag hittills enbart rört mig på ett teoretiskt plan och utifrån sociologisk teori diskuterat möjligheterna att integrera agentsystem i mänskliga organisationer. Jag kommer i detta kapitel kort beröra min problemställning från metodiskt perspektiv och diskutera några möjligheter att testa huruvida en agent eller ett agentsystem kan sägas vara integrerat i en mänsklig kontext. Hur är det möjligt att gestalta ett sådant test? Vad finns det för brister i de tester/metodiska angreppssätt som hittills används i samband med AI?

5.1 Turing-testet

Det i särklass mest kända testet i de här sammanhangen är det s k Turing-testet. Turing-testet framställdes egentligen för att testa huruvida en maskins eller en dators beteende var intelligent eller ej. Det presenterades för första gången i en artikel av matematikern Alan Turing (Turing, 1950). Turing ställer i början av artikeln frågan: ”Kan maskiner tänka?”. Denna fråga kom så småningom att bli ett startskott för en helt ny forskningsinriktning kallad Artificiell Intelligens. För att få svar på sin fråga konstruerade Turing ett test – Turing-testet. Detta test bygger på ett s k *imitation game* som utförs av tre människor, en man, en kvinna och en försöksperson. Försökspersonen, som sitter åtskild från de båda andra, skall genom att ställa frågor till de båda andra personerna försöka lista ut vem som är kvinna respektive man.

Med en liknande logik, menar Turing, är det möjligt att avgöra om en maskin kan uppvisa intelligent beteende. Försökspersonen sitter även i det här fallet fysiskt åtskild från de två ”svarsenheter” – en datamaskin och en människa. Försökspersonen ställer sedan under en begränsad tid ett antal valfria, skriftliga frågor till de båda ”svarsenheter” via en ledning. På grundval av dessa frågor gäller det för henne/honom att avgöra vilken av ”svarsenheter” som är dator respektive människa. Testet görs sedan om ett antal gånger av olika försökspersoner. Om testdeltagarna ej lyckas skilja dator och människa åt anses datorn/maskinen vara intelligent (Turing, 1950:433ff).

5.2 Durkheim-testet

Turing-testet kan ses som ett förslag till hur man skulle kunna mäta agenters integration på mikronivå. Vad det gäller makronivå och integration av DAI och MAS i organisationer skulle man kunna utgå från ett av Susan Leigh Star (1989) föreslaget alternativt testförfarande.

Utgångspunkten är Emile Durkheim och hans begrepp sociala fakta. Star menar att ett fenomen – exempelvis intelligens eller problemlösning – som existerar på systemnivå går ej att reducera till lägre nivåer. Intelligens och problemlösningsprocesser i öppna system är en form av sociala faktum som inte kan brytas ned och avgränsas till enskilda individer eller noder i ett system. Följden blir att ett test för DAI bör utgå från sociala metaforer snarare än psykologiska. Star menar att testet bör mäta systemets förmåga att svara mot eller uppnå de gemensamma organisatoriska målen.

Star påpekar vidare att Turing-testet bygger på en ”stängd-världsmodell” och att dess intelligensbegrepp grundas på en fast uppsättning regler för mänskligt beteende. Ett sådant test, hävdar hon, kan ej analysera motsatta synsätt och dess uppkomst i öppna system. Turing-testet utförs av individer på mikronivå och ej av grupper eller organisationer. Testet lämpar sig enligt Star därför inte att mäta i vilken grad ett DAI eller MAS integrerats i en organisatorisk kontext (Star, 1989:39f).

Turing-testet utgår från en datorkontext som är stängd, centraliserad och asocial. Ett test av intelligens i ett distribuerat öppet system bör – enligt Star – istället vara av ekologisk art. Att testa en nod i systemet kommer som sagt ej att ge ett tillförlitligt resultat. Att testa hela det öppna systemet är å andra sidan omöjligt. Tanken med Durkheim-testet är istället testa hur pass användbart systemet är för den sociala kontext – d v s interaktionssammanhanget – där systemet utvecklas. Genom att studera hur ett systems utformning, acceptans, användning och modifiering förhandlas fram i en social gemenskap – exempelvis i en organisation – är detta möjligt. Kriterierna för ett systems intelligens handlar då snarare om hur användbart det är för organisationens arbete, dess förmåga till förändring och anpassning samt dess förmåga att omfatta de olika synsätten och perspektiven i en organisation (Star, 1989:41). Fungerar ett system enligt dessa förutsättningar samtidigt som ett förtroende för tekniken uppstår, då är det möjligt att tala om en framgångsrikt institutionaliserad intelligent teknik (Rammert, 1995:19).

5.3 Testernas tillämpningsmöjligheter

Durkheim-testet är, som framgått, till viss del en kritisk reaktion eller anpassning av Turing-testet. Turing-testet har dock även mött kritik från många andra håll. Framförallt inom AI-kretsar har debatten om testets möjligheter och begränsningar varit livlig.

Företrädare för den så kallade Hollowshell-kritiken menar att datorprogram som uppfyller Turing-testets kriterier visserligen kan imitera intelligent mänskligt beteende men det innebär inte att man kan tillskriva dem intelligens. De är mer eller mindre tomma skal utan kognitiva och intentionella färdigheter (Scheuermann, 2000:10). Den mest kända företrädaren för detta perspektiv, John Searle, menar att de maskiner som tillskrivs intelligens i Turing-testet besitter en syntax men ingen semantik. De förstår vare sig de inmatade frågorna eller de angivna svaren utan imiterar enbart en människa eller en mänsklig roll/funktion utifrån vissa fasta regler. Den semantiska komponenten ligger snarare hos den mänskliga "utfrågaren" som utifrån den givna kontexten tillskriver maskinens utsagor vissa betydelser (Heinz, 1993:277f).

Turing-testet har vidare kritiserats för att det egentligen bara mäter en maskins interaktiva kompetens i ett visst begränsat sammanhang (Collins, 1990:190). Testsituationen är som Collins och andra kritiker påpekar i stor utsträckning laboratoriemässig. Den klart avgränsade testsituationen gör att datorns chanser att klara testet ökar. Turing-testet uppvisar här paralleller med det tidigare nämnda datorprogrammet Eliza. Eliza är ett språkanalysprogram som möjliggör konversation mellan en dator och en människa. Programmet består av en språkanalyskomponent som analyserar de inmatade satserna och ett variabelt "script" med (situationsspecifika) regler som används för att generera svar. Det innebär att programmet enbart fungerar när interaktionssituationen liknar ett terapeutiskt samtal. Då frågepersonen utgår från en sådan situation har programmet stora chanser att "vinna" Turing-testet (Heinz, 1993:276f). En maskin (eller en agent) stöter på problem när kommunikationen tar oväntade vändningar. Framför allt går det att märka då någon form av störning uppkommer. Maskiner av idag har i regel problem med att tolka och urskilja felstavningar och nonsenskommunikation (Collins, 1990:196f).

Hollowshell-kritiken mot Turing-testet utgår i mångt och mycket från ett essentialistiskt synsätt där testet kritiserats för att det inte kan mäta huruvida maskinen eller datorprogrammet

besitter intelligens eller ej. Koncentration ligger på maskinen i sig och dess kognitiva förmåga. I samband med integrationsproblematiken skulle jag dock vilja hävda att denna fråga inte är lika relevant. I integrationsprocessen är det snarare artefaktens roll i en social kontext som är av intresse. Det innebär att det är upplevelsen och tolkningen av maskiners (eller agenters) intelligens som står i centrum.

Kritiken mot att Turing-testet är koncentrerat till en bestämd och eventuell konstgjord situation går att bemöta genom att hävda att agenterna faktiskt implementeras i en viss miljö för att utföra vissa uppgifter. Vad som är intressant för min problemställning är hur pass väl integrerade agenterna är i en viss specifik miljö. *"A Turing-type test [...] is a test of the extent to which a machine can be located in an interactional network without stain"* (Collins, 1990:190). Som Collins skriver verkar Turing-testet vara bra lämpat för att mäta en agents eller ett agentsystems integration i en viss organisatorisk kontext.

Sett från ett pragmatiskt perspektiv tror jag alltså att Turing-testet, trots alla kritiska invändningar, kan vara intressant då det gäller att studera agentintegration på mikronivå. Den sociala acceptansen av agenterna är betydligt viktigare för integrationsprocessen än huruvida agenter verkligen – kognitivt sett – besitter intelligens.

Durkheim-testet ter sig vid en första blick väl lämpat att använda för att undersöka ett agentsystems integration i en organisatorisk kontext, d v s integrationen på makronivå. Problemet med Durkheim-testet eller snarare Stars beskrivning av detsamma är att det fortfarande är kvar i ett utkaststadium. Testet är luddigt och ospecificikt beskrivet. Det ges inget förslag till hur testet skulle kunna gestalta sig i praktiken. Star pekar enbart på vilka kriterier som bör sättas upp för att en implementering av DAI skall kunna anses vara lyckosam. Någon operationalisering av kriterierna eller något förslag till praktisk utformning, som i Turings fall, ges ej av Star. Durkheim-testet skulle enligt min mening behövas utvecklas ytterligare för att kunna vara till en nytta vid en undersökning av DAI och dess integration i en organisation.

6 Diskussion

I detta avslutande kapital tänker jag till att börja med kort beröra de nya risker och ansvarsfrågor som uppkommer i samband med intelligent teknik. Detta är ett studiefält som fortfarande är relativt outforskat och därför är det svårt att idag ge några klara svar på dessa frågor. Det är dock möjligt att göra vissa reflektioner och ge förslag till angreppssätt. Vidare tänker jag ägna kapitlet åt att kritiskt diskutera de teoretiska angreppssätt jag presenterat i uppsatsen. I anslutning därtill är min tanke att förlänga diskussionen till agentmodellerna och utifrån de sociologiska teorierna diskutera deras lämplighet. Avslutningsvis tänker jag kort spekulera kring agentteknikens följder i ett bredare, samhällsligt perspektiv.

6.1 Risker och ansvarsfördelning i hybrida system

Hybrida sociotekniska organisationer skapar nya risker och nya problem kring ansvarsfördelningsfrågan. Tanken med DAI och MAS är att delegera och distribuera vissa beslut eller delar av problemlösningsprocesser till artificiella agenter. På liknande sätt har man sedan två årtionden tillbaka använt sig av integrerade expertsystem i komplexa organisationer för att skapa effektivare problemlösnings- och beslutsprocesser. I samband med användning av expertsystem har det dock uppkommit olika oönskade effekter och olyckor. I vissa fall t o m med dödlig utgång. Det finns bl a exempel på där medicinska expertsystem feldoserat läkemedel till patienter, vilket i vissa fall lett till döden (Arnesjö et al, 1998). Situationer som denna skapar frågor kring hur oönskade effekter uppkommer samt hur ansvarsfrågan skall fördelas i de hybrida systemen. Frågan är om man skall lägga skulden på systemkonstruktörerna (de som modellerat och konstruerat systemet), systemoperatörerna (de som handhar systemet, i det här fallet sjukvårdspersonal) eller kunskapsbasen (d v s de läkare, vars kunskap ligger till grund för expertsystemet). Går det kanske t o m att ställa expertsystem eller agenter till svars för sina handlingar?

Orsaken till att en olycka eller en oönskad effekt uppkommer i en organisation låter sig oftast förklaras genom ett komplex av organisatoriska, tekniska och ekonomiska faktorer som samverkar (Berner, 1999:128f; Ericson et al, 1998). Charles Perrow har försökt förklara uppkomsten av olyckor i socio-tekniska system genom att koncentrera sig på hur systemet är sammankopplat och hur dess komponenter samverkar. Han menar att risken för att oväntade

effekter och olyckor skall uppkomma inom ett system är beroende av två faktorer – kopplingen och interaktiviteten mellan systemets komponenter. Komplex interaktivitet kombinerad med tät koppling innebär hög risk. I system med sådana egenskaper uppkommer ett svårlöst dilemma. Komplexiteten fordrar en decentraliserad organisation för att kunna handskas med oväntade problem. De som arbetar nära problemen är i regel också mest kompetenta att lösa dem. Å andra sidan saknar de helhetsbilden och har därför svårt att se effekterna av den lokala handlingen (Perrow, 1986:150f; Berner, 1999:130ff).

Paradoxen ovan påminner till stor del om det ofta återkommande dilemmat angående agenternas grad av frihet i samband med modellering av MAS och dess implementering i organisationer. Även här råder det ett konfliktfyllt förhållande mellan autonomi och kontroll. För att multi-agent systemet skall kunna fungera effektivt i en komplex miljö krävs det att agenterna kan handla självständigt, d v s utan att först få ett kommando om att utföra handlingen. Agentautonomi innebär å andra sidan en risk att agenterna tar initiativ till och utför oönskade eller direkt felaktiga och skadliga handlingar. Ett sätt att förebygga denna risk är att skapa vissa kontrollfunktioner, exempelvis genom begränsningar av agentens handlingsutrymme eller att använda dialogfönster där användaren kan tillåta eller avvisa en handling. Ökade kontrollfunktioner i ett system kan dock få en paradoxal effekt. Enligt Perrow bidrar sådana funktioner till ökad interaktivitet och återkoppling mellan systemets komponenter. Därmed ökar systemets komplexitet och med komplexiteten också riskerna (Perrow, 1987:107f).

Diskussionerna kring teknik och ansvar rör sig framför allt i juridisk och etisk riktning. Frågan om en agent kan hållas ansvarig för en handling är framför allt en moralfilosofisk fråga. Jag tänker därför här inte behandla frågan särskilt djupt. Utifrån en moralfilosofisk utgångspunkt finns det en utbredd föreställning hämtad från miljödebatten att lägga ansvaret på ”förorsakaren”. Även en maskin eller teknisk artefakt skulle då kunna ses som ”förorsakare” av en önskad effekt. Om maskinen/agenten behandlas som en handlande aktör, så vore det också juridiskt sett lättare för den skadade parten att utkräva ersättning (Lenk, 1994:71f). Enligt Hans Lenk är en nödvändig förutsättning för ansvarstagande att aktören/agenten har en möjlighet att kunna gripa in i kausalkedjan som ledde fram till ett beslut. M a o krävs en handlingsförmåga hos agenten (Lenk, 1994:78f). Också Schulz-Schaeffer pekar på en liknande koppling mellan handlingsförmåga och ansvarstagande (Schulz-Schaeffer, 1998:133).

Som diskussionen inom aktör-nätverksteorin (ANT) och även Perrows systemteori visar är tekniska utvecklingsprocesser och även handlingsförlopp inom organisationer snarare ett resultat av komplex samverkan mellan komponenterna i ett system eller nätverk. Med utgångspunkt i ett sådant synsätt blir svårt att fastställa "förorsakare" i komplexa, hybrida system. Som påpekats är agenters verksamhet vidare ofta dold och svår att spåra (Rammert, 1998:119).

Problemet är alltså huvudsakligen att finna ett ansvarssubjekt, d v s en aktör som är ansvarig för ett visst ansvarsområde gentemot en ansvarsinstans. Med utgångspunkt från tekniksociologiska teorier som Perrows systemteori och ANT framstår handlingssubjektet alltså snarare som kollektivt, medan ansvarssubjektet fortfarande i regel anses vara individuellt (jfr Zimmerli, 1987). En möjlig väg vore att göra alla som varit med och gestaltat systemet medskyldiga eller att skapa någon form av modell där en övergripande institution fungerar som ansvarssubjekt. Sådana ansvarsbegrepp har dock visat sig svåra att operationalisera och omsätta i praktiken (jfr Lenk, 1994; Ropohl, 1987:164ff). Det finns idag inga givna svar på frågan hur ansvaret skall fördelas i en hybrid organisation. Frågan är dock viktig och här finns ett stort forskningsbehov i framtiden.

6.2 Integrationsvägar – teoriernas potential och begränsningar

Det är framförallt två nyckelbegrepp som utkristalliserat sig i de teorier jag valt att gå igenom för att utforska möjligheterna att integrera artificiella agenter eller agentsystem i mänskliga organisationer. Den ena riktningen skulle kunna sägas utgå från handlingsbegreppet och dess betydelse i integrationsprocessen. Vad som då framförallt betraktas som intressant är i vilken utsträckning agenter är kapabla att utföra mänskliga handlingar och mänskligt arbete. Genom att kunna ta över och ersätta mänskligt arbete skulle då agenter kunna ses som integrerade i organisationen. Ett annat övergripande angreppssätt är att lägga tonvikten vid hur agenterna påverkar mänskliga aktörer och uppfattas av desamma i sociala interaktionsprocesser.

6.2.1 Handlingsbaserad integration

En möjlighet att angripa systemintegrationen från mikronivå är att som Collins dra en skiljelinje mellan handling och beteende. Människor påstås då besitta en handlingsförmåga medan maskiner eller tekniska artefakter enbart klarar av att utföra eller imitera ett beteende (jfr 3.1.2). Det torde då råda ett segregerat förhållande inom organisationen mellan de handlande människorna och de artificiella agenterna som bara utför ett beteende. Ett problem med Collins och liknande definitioner av handlingsförmåga är att de tenderar reducera bort den sociala kontextens betydelse på bekostnad av inomliggande faktorer hos aktören. Att lägga koncentrationen på den kognitiva förmågan innebär att de sociala sammanhangen riskerar glömmas bort.

Att basera en integrationsanalys på Collins begreppspår innebär vidare en förenklad syn på människors respektive maskiners handlingsförmåga. För det första har jag i arbetet konstaterat att det finns agentmodeller som skulle kunna uppfylla kriterierna för (mänsklig) handlingsförmåga. Intentionella agenter kan besitta såväl handlingsalternativ, kognitiv förmåga som (åtminstone viss) autonomi. För det andra skulle jag, precis som Giddens (jfr 3.1.1), vilja hävda att det mänskliga handlandet består av flera dimensioner, varav intentionella handlingar endast är en av dessa. I många fall handlar även människor reflexartat och automatiskt. Denna problematik har också visat sig i samband med Turing-testet, där datorn påstås ha två möjligheter att ta hem spelet. Antingen genom att datorn/programmet själv lyckas uppföra sig som en människa eller att människan uppför sig som en maskin, d v s handlar linjärt och reflexmässigt.

Perrow använder ett makroorienterat, emergent handlingsbegrepp för att peka på tekniska artefaktors integration i komplexa system. Betraktat från makronivå anser Perrow (1987) att tekniska artefakter genom att bidra till ett systems verkan/handlande är integrerade i systemet/organisationen. Även i Perrows modell intar alltså därmed handlingsförmågan i någon mån den centrala platsen. Precis som ANT urskiljer dock inte Perrow de enstaka enheternas handlingar utan menar att systemets beteende uppkommer genom samverkan av ett nätverk av enheter. Han gör därmed heller ingen åtskillnad mellan beteende och handling. Perrow lägger alltså den analytiska tonvikten på strukturen eller systemet snarare än på den enskilde aktören/enheten. Detta angreppssätt mynnar ut i ett sorts deterministiskt schema där han bl a menar att alltför komplexa system oundvikligen tycks leda till olyckor eller

katastrofer. Perrow ger i sin teori inget utrymme för att förklara hur en teknisk artefakt lyckas bli en integrerad del av ett sociotekniskt system. Integrationsstatusen orienterar sig snarare i positivistisk riktning och betraktar agenternas/agentsystemets integration som mer eller mindre given. Något utrymme för tolkningar inom komplexa system, d v s olikartade mänskliga uppfattningar av systemets processer och tekniska artefakters integration, finns det heller inget utrymme för i Perrows modell.

6.2.2 Interaktionsbaserad integration

Som jag påvisat i kapitel 3 utgår de flesta sociologiska handlingsteorier från att aktörers handlingar är beroende av den omgivande kontexten och uppkommer samt får sin mening först i interaktion med omgivningen. Att använda ett objektiva handlingsbegrepp som måttstock för integration verkar därför ur sociologisk synvinkel inte vara vidare fruktbart.

Intressantare är då kanske snarare teorier som sätter interaktionen i centrum och lägger mer vikt vid själva integrationsprocessen. Enligt interaktionistiska teorier (Mead, symbolisk interaktionism etc) skulle då de artificiella agents integrationsstatus utgå från hur de mänskliga aktörerna upplever och definierar agenternas handlande och funktion. Integrationen äger rum i interaktionsprocesser. Beroende hur pass trovärdigt agenterna kan imitera mänskliga uttryck eller mänskligt beteende tilldelas de sin status inom organisationen.

Socialkonstruktivistiska tekniksociologiska betraktelser av sociotekniska system kan sägas vara rotade i dessa interaktionsprocesser på mikronivå. Ett sociotekniskt system består enligt denna riktning i princip av interaktion mellan olika (tekniska och mänskliga) komponenter. Komponenterna i ett sociotekniskt system påverkar varandra och definierar varandra ömsesidigt. Artificiella agents roll och integration i en organisation avgörs således i en förhandlingsprocess. Denna inriktning distanserar sig från Perrows systemteori genom att betona tolkning och interaktionen mellan aktörer eller subgrupper inom en organisation som de avgörande komponenterna i integrationsprocessen.

Den omgivning som agentsystem normalt sett implementeras i kännetecknas av komplexitet. Olika personer och subgrupper med olika mål och bakgrund försöker i dessa organisationer enas om en gemensam tolkning. Utifrån ett sådant maktperspektiv ter sig ett socialkonstruktivistiskt synsätt vara intressant. Integrationsprocessen kan då analyseras som

en maktkamp mellan olika aktörers/intressens tolkningar. Vissa grupper är säkerligen inte lika benägna att acceptera ett integrerat agentsystem. Exempelvis kan man tänka sig att vissa grupper och människor känner sig och sitt jobb hotade av artificiella agenter och därför motarbetar en integration. I andra fall, exempelvis hos en äldre generation, är det lätt att föreställa sig en viss skepsis mot ny datorteknik vilket kanske innebär att de är mindre benägna att fördela uppgifter vidare till artificiella agenter.

Det går att invända en del kritik mot det interaktionistiska integrationsbegreppet.

Interaktionssammanhanget tenderar bli alltför aktörsorienterat. Strukturer och andra objektiva element reduceras bort. I den mån strukturer finns är de inget annat än summan av de olika aktörerna och deras interaktioner. Jag skulle dock vilja hävda att det i de flesta organisationer finns institutionaliserade element och kvarlevande kulturer som ligger utanför organisationens aktörskrets och som påverkar möjligheterna till integration. Den materiella kontexten har en oinflytande över vilka tolkningsvägar står öppna.

Vidare tycker jag mig kunna se vissa metodiska problem här. Går det överhuvudtaget att operationalisera begreppen och därmed skapa empiriska testmöjligheter utifrån en socialkonstruktivistisk grund? En invändning är också att socialkonstruktivistiska modeller inom tekniksociologin främst har använts för att beskriva de tidigaste faserna av teknikutveckling, dvs utvecklingen fram till färdig produkt (Rammert, 2000:50ff). I implementerings- och integrationssammanhang har de så vitt jag vet inte använts lika ofta.

6.2.3 Aktör-nätverksteori – en syntes

Som vi sett går det att finna brister både hos de positivistiskt inriktade handlingsbaserade angreppssätten och den socialkonstruktivistiska interaktionsbaserade integrationssynen. Finns det då någon lämplig medelväg mellan dessa båda poler? Till viss del tror jag ANT kan erbjuda en sådan. Enligt ANT framstår tekniska artefakter såväl som objektiv verklighet som socialt konstruerade produkter. Alla komponenter, såväl tekniska som sociala, är på samma sätt aktörer i etableringen och upprätthållandet av socio-tekniska system.

Integrationsprocessen skulle med ANT-termer kunna förstås som att artificiella agenter kopplas till ett eller flera problemlösande nätverk inom eller i samband med den aktuella organisationen. Komponenterna/artefakterna får sin identitet i nätverksbildningsprocesser, i

interaktion med andra aktanter. Såväl människor som den materiella omgivningen räknas som potentiella handlingsobjekt i sådana processer.

Materiella och tekniska artefakter erbjuder alltså visst "handlings/tolkningsrum" i integrationsprocessen. Interaktionen mellan en artificiell agent och människa innebär en tolkning från människans sida men även ett aktivt tvång från artefaktens sida. Aktanter och objekt kan därmed anses ha vissa givna, om än något amorfa, egenskaper. Inom ANT talar man om inskriptioner. Inskriptionerna skapar förutsättningar för andra aktörers beteende eller anslutning till handlingsnätverket (Schulz-Schaeffer, 2000b: 91ff).

Ett nätverk uppnår i regel en viss stabilitet efter ett tag. Stabiliteten kännetecknas av konvergens och irreversibilitet. Nätverket utvecklas till en "black box", där ett oberoende i förhållande till den omgivande kontexten utvecklats och där "... *many elements are made to act as one...*" (Bruno Latour citerad i Schulz-Schaeffer, 2000b:118). Aktanterna har så integrerats i nätverket. Med konvergens avses då att nätverkets aktanter passar samman trots heterogeniteten. Irreversibilitet avser aktanternas och dess relationers motståndskraft mot förändringar som försvagar de inbördes relationerna.

Det finns självklart även kritik att invända mot ANT. Teorin är relativt oprövad då det gäller att studera integration av teknik. Den har egentligen framställts som förklaringsmodell för innovationers uppkomst. Å andra sidan har faktiskt Latour i tänkta exempel, som i fallet med dörrstängaren och det berlinska låset (jfr 4.2.3), också pekat på tekniska artefaktens roll som integrerande element i ett nätverk. En annan kritikpunkt är att teorin till viss del framstår som alltför vag och abstrakt. Vissa delar är svåra att förstå, delvis beroende på slarvigt användande av begrepp. I vissa fall används aktör, i andra agent, i ett tredje sammanhang talas det om aktanter. Teorins komplexitet och oklarhet gör också att jag kan tänka mig att man stöter på svårigheter då man skall operationalisera begreppen och omsätta teorin metodiskt.

Trots dess brister och problem finner jag ändå ANT vara den intressantaste teoretiska förklaringsmodellen. Som jag försökt påvisa tror jag ANT besitter potential att förstå såväl integrationsprocessen som ett agentsystems integrationsstatus i en organisation. ANT påvisar att tekniska artefakter har en potentiell integrativ funktion att fylla. Den öppnar upp för olika empiriska angreppssätt där såväl aktörens definition av kontextens element och dess kopplingar till nätverket som översättningarna och hur aktören definieras genom dessa kan

studeras. Den ger m a o en möjlighet att förklara integrationen från såväl ett aktör- som ett systemperspektiv.

6.2.4 Agentmodellernas integrationspotential

På mikronivå går det att konstatera att agentens ”reella” intelligens och handlingsförmåga är av mindre vikt för integrationsmöjligheterna. Som studier av Turkle (1986;1998) och Sagawe (1994) visat är agenternas intentionella förmåga inte avgörande utan viktigare är agentens kommunikativa kompetens. Agenten tillmäts sin betydelse utifrån hur pass trovärdigt den uppträder i interaktion med människor. Interaktionistiska agenter (jfr 3.3.2) har en förmåga att dra lärdomar av interaktionen med människor och är därmed sannolikt bättre på att anpassa sig till en viss social kontext. Troligen ökar då möjligheterna för en lyckad integration.

På makronivå, ur problemlösningssynpunkt, menar de nyare tekniksociologiska perspektiv (jfr 4.2) jag haft uppe till diskussion att även icke-intelligenta tekniska artefakter kan fungera som integrerade delar av sociala sammanhang. Ur integrationssynpunkt utgör då agenttekniken inte något drastiskt brott jämfört med andra tekniska artefakter. Oavsett intelligensförmåga besitter agenter en potential att medverka till de organisatoriska handlingsprocesserna. I synnerhet inom ANT betraktar man tekniken utifrån ett integrativt sammanhang. Tekniken bidrar till att den sociala kontexten förblir sammanhållen. Agenter har m a o en social och integrativ funktion att fylla oavsett hur intelligenta och eller sofistikerade de är. Genom att betrakta tekniska artefakter på ett sådant sätt öppnar ANT därmed upp en möjlighet för olika former av agenter med olika kompetenser. Olika agenttyper har olika integrativa funktioner att fylla och kan mycket väl verka tillsammans sida vid sida.

6.3 Agentteknikens samhälleliga konsekvenser – en sociologisk utblick

Softwareagenterna är mitt ibland oss och det är ett faktum vi måste lära oss leva med. Agenttekniken skapar många nya möjligheter men samtidigt också nya problem. Problem som kräver forskning och diskussion. Jag tänker avslutningsvis antyda några tänkbara komplikationer och följder som agenttekniken kan föra med sig. Hur agenttekniken och därmed de påföljande problemen kommer att gestalta sig i framtiden kan vi idag bara gissa

oss till. De utvecklingslinjer jag beskriver är därför med all nödvändighet spekulativa till sin art.

Det kanske intressantaste kvalitativa brottet gentemot tidigare teknik ligger i agenternas autonomi. Agenterna är utrustade med en förmåga att själva fatta beslut och välja handlingsvägar. Autonomi innebär att agentsystemens utvecklingsvägar inte kan förutsägas fullt ut. Detta problem har uppmärksammats i debatten om hur fördelningen mellan autonomi och kontroll ska se ut. Risker för att mänskliga tekniska innovationer utvecklas till okontrollerbara monster har uppmärksammats i romaner och filmer långt innan artificiell intelligens som forskningsdisciplin ens var påtänkt. Dystopierna har i regel gått ut på att de av människor skapade maskinerna eller robotarna utvecklats på egen hand i helt oväntade riktningar och senare tagit makten från sina tidigare herrar och blivit världens härskare.

En så drastisk utveckling kommer väl knappast de artificiella agenterna att medföra. Men att agentsystemen tar egna oväntade utvecklingsvägar är nog ett faktum vi måste leva med. Med utgångspunkt i kaosteori har försök gjorts att skapa sk artificiella världar i form av datorprogram (jfr 2.2.2.2). Även då förutsättningarna och agenternas handlingsalternativ varit väldigt enkla och klart begränsade har efter ett tag väldigt komplicerade och oförutsägbara mönster utvecklats (jfr Turkle, 1998:238ff). Likaså måste man räkna med att agentsystemen kan utvecklas på de mest oväntade sett även då människorna skapat förutsättningarna och handlingsalternativen för agenterna. Ändock tror jag att man genom att bygga modelleringen och uppföljningen av MAS på en sociologisk grund kan skapa agentsystem som bättre liknar mänskliga sociala system och därmed blir mer förutsägbara.

Agenttekniken kan också betraktas som ännu ett steg i den kontinuerliga externaliseringsprocessen av mänsklig kunskap och kompetens. Med externalisering avser jag då överföring av kunskap och kompetens till medier och artefakter som ligger utanför människan. Externalisering av mänsklig kunskap är långt ifrån en ny företeelse. Redan genom böckernas utbredning överfördes viss kunskap och vissa färdigheter till materiella medier. På senare tid har utbredningen av exempelvis expertsystem och fickräknare bidragit till att såväl kunskap som vissa problemlösningsprocesser lagras utanför människorna (Sinding-Larsen, 1987). I nuläget är det framför allt relativt okomplicerade administrativa rutinarbeten som delegeras till agenter, men i takt med att agenterna blir alltmer sofistikerade lär de ta över allt fler avancerade uppgifter. Precis som är fallet med tidigare externaliseringsprocesser riskerar

vissa kompetenser glömmas bort av människorna genom att ligga inkapslade i maskiner eller andra materiella artefakter. Detta faktum riskerar öka de sociala systemens sårbarhet. Om vissa (tekniska) komponenter i ett system slås ut, exempelvis ett multiagentsystem i en organisation, kan stora svårigheter uppkomma. Problem som dessa illustreras tydligt i aktör-nätverkbaserade analyser. Organisationer och arbetsprocesser är idag helt enkelt beroende av tekniska (artificiella agenter inkluderade) och materiella artefakter för att fungera.

I samband med externaliseringsprocessen och delegerandet av uppgifter till artificiella agenter uppkommer nya frågor kring människors tillit till agenter. I uppsatsen har jag diskuterat vikten av tillit i organisatoriska integrationsprocesser. I framtiden kan man även tänka sig att vissa tjänster och arbeten kommer tas över av agenter och därmed blir vanliga människor kanske förvisade till artificiella rådgivare i samband med bankkäranden, medicinska problem o s v. Här handlar det om den "lilla människans" tillit till agenterna. Kommer kunder och patienter acceptera att bli betjänade av artificiella intelligenser istället för mänskliga intelligenser? Kommer de över huvudtaget ha ett val eller en möjlighet att skilja människa från maskin?

Agenttekniken får förmodligen också arbetsorganisatoriska konsekvenser. M h a MAS och DAI kan vissa mänskligt arbetsprocesser automatiseras genom att delegeras till artificiella agenter. Vilket omfång denna automatisering kommer att få är idag svårt att säga. Likaså är det inte helt klart vilka arbetstillfällen och yrken som därigenom riskerar att försvinna. Artificiella agenter kan redan idag exempelvis överta stora delar av administrativa uppgifter och ersätta mänskliga assistent- och övervakningsjobb av olika slag. Även mer avancerade jobb i form av planering och rådgivning kommer troligen på sikt kunna utföras av artificiella agenter. Frågan är här om vi enbart ser en fortsättning av en redan löpande rationaliseringsprocess som datoriseringen och automatiseringen bidragit till eller om det rör sig om en kvalitativt helt ny process.

I vilken utsträckning agenttekniken verkligen innebär ett radikalt brott gentemot andra former av teknik är svårt att säga och frågan i sig känns egentligen inte särskilt relevant. Som påpekats i uppsatsen är teknikens följderna något som uppkommer i interaktion med människor och kulturer. Sammanfattningsvis är det nog ändå rimligt att anta att de artificiella agenternas utbredning i samhället innebär att det blir än svårare att dra gränser mellan människa och teknik. Agenterna har ofta mänskliga egenskaper som gör de svåra att skilja från människor.

Dessutom handlar de i det dolda vilket gör det besvärligt följa dess handlingsprocesser. Agenternas autonomi gör även att de på ett aktivt sätt kan påverka och transformera sociala strukturer. Av dessa skäl tror jag att det är rimligt att analytiskt sett tillmäta agenterna en aktörstatus. Från ett sociologiskt perspektiv bör man då intrikta sig på att studera interaktionsprocesserna mellan agenter och människor, samt även agenterna emellan, om man skall kunna göra sig en rättvis bild av utvecklingsprocessen.

Litteratur

- Arnesjö, B, Collste, G, Helgeson, B & Lundberg, L (1998): Datoriseringens risker, i Grimvall, G, Jacobsson, P & Thedéen, T (red): *Risker i tekniska system*, Stockholm: Utbildningsradion.
- Berger, U & Bernhard-Mehlich I (1999): Die Verhaltenwissenschaftliche Entscheidungstheorie, i Kieser A (red): *Organisationstheorien*, Stuttgart: W. Kohlhammer.
- Berner, B (1999): *Perpetuum Mobile? teknikens utmaningar och historiens gång*, Lund: Arkiv.
- Bijker, W E, Hughes, T P & Pinch, T (red.) (1987): *The Social Construction of Technological Systems*, Cambridge, MA: MIT Press.
- Blumer, H (1969): *Symbolic Interactionism*, Englewood Cliffs, NJ: Prentice Hall.
- Bond, A H & Gasser, L (red) (1988): *Readings in Distributed Artificial Intelligence*, San Mateo: Morgan Kaufman.
- Braun, H (1998): The Role-Taking of Technology, i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.
- Braun, H & Imhof, P (1999): Künstliche Gesellschaften, *Telepolis*, <<http://www.heise.de/tp/deutsch/inhalt/co/5219/1.html>> (990823)
- Braun, H (2000): *Soziologie der Hybriden. Über die Handlungsfähigkeit von technischen Agenten*, Berlin: Institute for Social Sciences, Technische Universität Berlin.
- Bühl, A (1996): *Cybersociety – Mythos und Realität*, Köln: PapyRossa Verlag.
- Bühl, A (1997): Die virtuelle Gesellschaft, i Gräf L & Krajewski M (red): *Soziologie des Internets*, Frankfurt/Main: Campus Verlag.
- Burkhard, H-D & Rammert, W (2000): *Integration kooperationsfähiger Agenten in komplexen Organisationen – Möglichkeiten und Grenzen der Gestaltung hybrider offener Systeme*, <<http://www.ki.informatik.hu-berlin.de/inka/antragstext.html>> (citerad 000928)
- Callon, M (1991): Techno-Economic Networks and Irreversibility, i Law J: *A sociology of monsters: essays on power, technology and domination*, London: Routledge.
- Charon J (1979): *Symbolic interaction : an introduction, an interpretation, an integration*, Englewood Cliffs : Prentice-Hall.
- Collins, H M (1990): *Artificial Experts. Social Knowledge and Intelligent Machines*, Cambridge, Mass.: MIT Press.
- DFG-Schwerpunktprogramm Sozionik: Sozionik: Erforschung und Modellierung künstlicher Sozialität, <<http://www.tu-harburg.de/tbg/SPP/spp-antrag.html>> (citerad 000724).
- Durfee, E H, Lesser, V R & Corkhill, D D (1987): Cooperation through communication in a distributed problem solving network, i Huhns, M N (red), *Distributed Artificial Intelligence*, London: Pitman.
- Ericson, M & Mårtensson, L (1998): Den mänskliga faktorn, i Grimvall, G, Jacobsson, P & Thedéen, T (red): *Risker i tekniska system*, Stockholm: Utbildningsradion.
- Esposito, E (1995): Illusion und Virtualität: Kommunikative Veränderungen der Fiktion, i Rammert, W (red): *Soziologie und Künstliche Intelligenz. Produkte und Probleme einer Hochtechnologie*, Frankfurt/Main: Campus.
- Florian, M (1998): Die Agentengesellschaft als sozialer Raum, i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.

- Gasser, L (1986): The Integration of Computing and Routine Work. *ACM Transactions on Office Information Systems* 4(3), 205-225.
- Gasser, L (1991): Social conceptions of knowledge and action: DAI foundations and open systems semantics, *Artificial Intelligence* 47, 107-138.
- Geier, M (1999): *Fake: Leben in künstlichen Welten. Mythos, Literatur, Wissenschaft*, Hamburg: Rowolt.
- Geser, H (1989): Der PC als Interaktionspartner, *Zeitschrift für Soziologie*, 18, 3, 230-243.
- Giddens, A (1988): *Die Konstitution der Gesellschaft*, Frankfurt/Main: Campus.
- Gilbert, N (1995): *Simulation: an emergent perspective*, Centre for Research on Simulation in the Social Sciences (CRESS), Department of Sociology, University of Surrey, UK
<<http://www.soc.surrey.ac.uk/research/simsoc/tutorial.html>> (citerad 000903)
- Gilbert, N & Conte, R (red.) (1995): *Artificial societies – The computer simulation of social life*, London: UCL Press.
- Grint, K & Woolgar, S (1997): *The machine at work - technology, work and organization*, Cambridge: Polity Press.
- Haraway, D J (1995): *Die Neuerfindung der Natur – Primata, Cyborgs und Frauen*, Frankfurt/Main: Campus Verlag.
- Heintz, B (1993): *Die Herrschaft der Regel. Zur Grundlagengeschichte des Computers*, Frankfurt/Main: Campus Verlag.
- Hewitt, C E (1977): Viewing Control Structures as Patterns of Passing Messages, *Artificial Intelligence* 8, 323-364.
- Hewitt, C E (1986): Offices are Open Systems, *ACM Transactions on Office Information Systems* 4(3), 1986, 271-287.
- Hewitt, C E (1991): Open Information Systems Semantics for Distributed Artificial Intelligence, *Artificial Intelligence* 47, 79-106.
- Hughes, T P (1987): The Evolution of Large Technical Systems, i Bijker, W E, Hughes, T P & Pinch, T (red.): *The Social Construction of Technological Systems*, Cambridge, MA: MIT Press.
- Huhns M & Singh M (red) (1997): *Readings in Agents*, San Francisco: Morgan Kaufman.
- Jennings N R (1996): Coordination Techniques for Distributed Artificial Intelligence, i O'Hare G M P & Jennings N R (red): *Foundations of Distributed Artificial Intelligence*, New York: John Wiley&Sohns.
- Kiss, G, (1996): Agent Dynamics, i O'Hare G M P & Jennings N R (red): *Foundations of Distributed Artificial Intelligence*, New York: John Wiley&Sohns.
- Kornfeld W A & Hewitt C E (1980): The Scientific Community Metaphor, *IEEE Transaction on Systems, Man and Cybernetics*, SMC 11, 1, 24-31.
- Latour, B (1988): Mixing Humans and Nonhumans Together. The Sociology of a Door-Closer, *Social Problems* 35 (3), 298-310.
- Latour B (1994): *Der Berliner Schlüssel*, Berlin: WZB.
- Latour B (1998): *Artefaktens återkomst: ett möte mellan organisationsteori och tingens sociologi*, Stockholm: Nerenius & Santérus.
- Lenk, H (1994): *Macht und Machbarkeit der Technik*, Stuttgart: Reclam.
- Law J (1991): *A sociology of monsters: essays on power, technology and domination*, London: Routledge.

- Maes, P (1994): Agents that Reduce Work and Information Overload, *Communications of the ACM* 37, 7, 1-40+146.
- Malsch, T (red.) (1998): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.
- Malsch, T, Florian, M, Jonas, M & Schulz-Schaeffer I (1998): Sozionik, i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.
- March, J G & Simon, H A (1958): *Organizations*, New York m fl: John Wiley & Sons.
- Martial, F. v. (1992): Einführung in der Verteilte Künstliche Intelligenz, *KI* 1, 6-11.
- Moulin B & Chaib-Draa B (1996): An overview of Distributed Artificial Intelligence, i O'Hare G M P & Jennings N R (red): *Foundations of Distributed Artificial Intelligence*, New York: John Wiley&Sohns.
- Müller H J (red) (1993): *Verteilte Künstliche Intelligenz*, Mannheim: BI-Wissenschaftsverlag.
- Müller H J (1996): Negotiation Principles, i O'Hare G M P & Jennings N R (red): *Foundations of Distributed Artificial Intelligence*, New York: John Wiley&Sohns.
- Müller H J & T Wittig (1993): Anwendungen von Multi-Agenten Systemen, i Müller H J (red): *Verteilte Künstliche Intelligenz*, Mannheim: BI-Wissenschaftsverlag.
- Parsons T C (1937/1968): *The Structure of Social Action*, New York : Free Press.
- Parunak, H V D (1996): Applications of Distributed Artificial Intelligence in Industry, i O'Hare G M P & Jennings N R (red): *Foundations of Distributed Artificial Intelligence*, New York: John Wiley&Sohns.
- Perrow, C (1986): *Complex Organizations. A Critical Essay*. 3. uppl, New York. McGraw-Hill.
- Perrow, C (1987): *Normale Katastrophen. Die unvermeidbaren Risiken der Grosstechnik*, Frankfurt am Main: Campus.
- Pflüger, J (1997): Distributed Intelligence Agencies, i Warnke, M., Coy, W. & Tholen, G. (red): *Hyperkult. Geschichte, Theorie und Kontext digitaler Medien*. Basel: Stroemfeld.
- Rammert, W (red) (1995): *Soziologie und Künstliche Intelligenz. Produkte und Probleme einer Hochtechnologie*, Frankfurt/Main: Campus.
- Rammert, W (1998): Giddens und die Gesellschaft der Heinzelmännchen, i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.
- Rammert, W (2000): *Technik aus soziologischer Perspektive 2*, Wiesbaden: Westdeutscher Verlag.
- Rich, E & Knight, K (1991): *Artificial Intelligence*, New York: McGraw-Hill.
- Ritzer, G (1996): *Sociological Theory*, New York: McGraw-Hill.
- Ropohl, G (1987): Neue Wege, die Technik zu verantworten, i Lenk, H & Ropohl, G: *Technik und Ethik*, Stuttgart: Reclam.
- Sagawe, H (1994): *Einfluss "intelligenter" Maschinen auf menschliches Verhalten*, Opladen: Westdeutscher Verlag.
- Scheuermann, K (2000): *Menschliche und technische 'Agency': Soziologische Einschätzungen der Möglichkeiten und Grenzen künstlicher Intelligenz im Bereich der Multiagentensysteme*, Berlin: Institute for Social Sciences, Technische Universität Berlin.
- Schüle J A (1983): *Mikrosoziologie - ein interaktionsanalytischer Zugang*, Opladen: Westdeutscher Verlag.
- Schulz-Schaeffer, I (1998): Akteure, Aktanten und Agenten, i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.

- Schulz-Schaeffer, I (2000a): *Enrolling Software Agents in Human Organizations. The Exploration of Hybrid Organizations within the Socionics Research Program*, Berlin: Institute for Social Sciences, Technische Universität Berlin.
- Schulz-Schaeffer, I (2000b): *Sozialtheorie der Technik*, Frankfurt am Main: Campus-Verlag.
- Schulz-Schaeffer I & Malsch T (1998): Das Koordinationsproblem künstlicher Agenten aus der Perspektive der Theorie symbolisch generalisierter Interaktionsmedien, i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.
- Simon, H A (1961): *Administrative Behavior*, New York: The Macmillan Company.
- Shibutani, T (1987): *Society & Personality – An Interactionist Approach to Social Psychology*, New Brunswick: Transaction Books.
- Sinding-Larsen, H (1987): Informationsteknologi och kunskapsförvaltning, *Dialoger*, 5, 9-19.
- Star, S L (1989): The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving, i Huhns, M & Gasser, L (red): *Distributed Artificial Intelligence, Vol. 2*, London: Pitman.
- Strübing, J (1998): Multiagenten-Systeme als "Going Concern", i Malsch, T (red): *Sozionik – soziologische Ansichten über künstliche Sozialität*, Berlin: Edition Sigma, Rainer Bohn Verlag.
- Suchman, L A (1987): *Plans and Situated Actions. The Problem of Human-Machine Interaction*, Cambridge: Cambridge University Press.
- Sundermeyer, K (1993): Modellierung von Agentensystemen, i Müller H J (red): *Verteilte Künstliche Intelligenz*, Mannheim: BI-Wissenschaftsverlag.
- Turing, A (1950): Computing Machinery and Intelligence, *Mind*, 50, 433-460.
- Turkle, S (1986): *Die Wunschmaschine. Vom Entstehen der Computerkultur*, Reinbek: Rowolt.
- Turkle, S (1998): *Leben im Netz – Identität in Zeiten des Internet*, Hamburg: Rowolt.
- Wooldridge, M J & N R Jennings (1995): Agent Theories, Architectures, and Languages: A Survey, i Wooldridge, M J & N R Jennings (red): *Intelligent Agents. ECAI-94 Workshop on Agent Theories, Architectures, and Languages, Amsterdam, The Netherlands, August 8-9, 1994, Proceedings*, Berlin: Springer-Verlag.
- Woolgar, S (1985): Why not a Sociology of Machines? The Case of Sociology and Artificial Intelligence, *Sociology*, 14, 4, 557-572.
- Zimmerli, W C (1987): Wandelt sich die Verantwortung mit dem technischen Wandel?, i Lenk, H & Ropohl, G: *Technik und Ethik*, Stuttgart: Reclam.

Bilaga – Ordlista

Artificiell/Intelligent/Software agent:

Ett datorprogram som på uppdrag av en individ eller en organisation utför olika handlingar kontinuerligt och autonomt. Kallas även softbot ("software robot") i andra sammanhang.

Artificiell Intelligens (AI):

En dators eller en datorstyrd robots förmåga att utföra uppgifter som normalt sätt förknippas med högre intellektuella processer karakteristiska för människor, såsom förmåga att resonera logiskt, upptäcka betydelse, generalisera eller lära från tidigare erfarenheter. Begreppet används även för den gren inom datorvetenskapen som sysselsätter sig med utveckling av system utrustade med sådana förmågor.

Distribuerad Artificiell Intelligens (DAI):

En riktning inom AI som utgår från att intelligenta lösningar på komplexa problem för det mesta inte är en enskild individs verk, utan är ett resultat av flera handlande enheters interaktion och sysselsätter sig med att utforska möjligheterna att skapa AI-program baserade på detta antagande.

Expertsystem:

Datorprogram som används för att utföra AI-uppgifter av olika slag. Designade att manipulera och hantera symbolisk information i extrem hög hastighet och på så sätt uppväga dess brist på mänsklig kunskap och selektiva förmågor.

Multi-agent system (MAS):

Ett löst sammankopplat nätverk av problemlösare (agenter) som arbetar tillsammans för att lösa problem som ligger utanför deras egen förmåga.

Reaktiv agent:

Agent utrustad med autonomi, social förmåga, reaktionsförmåga samt förmåga att ta egna handlingsinitiativ. Saknar dock kognitiv förmåga och dess handlande består av förprogrammerade alt. inlärd reaktioner framkallade av förändringar i omgivningen eller av meddelanden från andra agenter.

Reflektiv/intentionell agent:

Agent utrustad med en kognitiv förmåga och ett "mentalt tillstånd" bestående av kunskap, avsikter, mål.