

MIND-BODY PROBLEMETS OLÖSBARHET FRIGÖR VILJAN

Jan Scheffel, 29/3-2016

ABSTRACT

Mind-body problemet analyseras i ett reduktionistiskt perspektiv. Genom att kombinera emergensbegreppet med algoritmisk informationsteori visas i ett tankeexperiment att ett starkt epistemiskt emergent system kan konstrueras utifrån en relativt enkel, icke-linjär process. En jämförelse med hjärnans avsevärt mer komplexa neurala nätverk visar att även medvetandet kan karakteriseras som starkt epistemiskt emergent. Därmed är reduktionistisk förståelse av medvetandet inte möjlig; mind-body problemet har alltså inte en reduktionistisk lösning. Medvetandets ontologiskt emergenta karaktär kan därefter konstateras utifrån en kombinatorisk analys; det är därmed principiellt oreducerbart till lägre-nivå-fenomen. I perspektivet av en modifierad definition av fri vilja diskuteras den fysiska växelverkan som äger rum i hjärnans neurala system. Trots att enskilda neurala lägre-nivå-processer är deterministiska, kan globala processer visas vara icke-deterministiska i ontologisk mening. Vi argumenterar för att detta leder till viljans frihet.

1. INTRODUKTION

Att förstå medvetandet är naturligtvis en central strävan inom filosofin. Litteraturen som genom århundradena producerats inom 'mind-body' - problemområdet är också enorm; en delmängd av över 3000 artiklar har exempelvis samlats av Chalmers (2015). En uppenbar motsättning ligger i det faktum att medan vi normalt söker vetenskaplig förståelse utifrån ett reduktionistiskt perspektiv, i vilket helheten förstås utifrån sina beståndsdelar, har medvetandet under årtusenden evolverat till ett extremt komplext system med avancerade högre-nivå-egenskaper. De teoretiska svårigheter som hittills mötts antyder att någon ny tanke, en ny ingrediens, måste tillföras för att mind-body problemet ska kunna avgöras. I detta arbete argumenteras för att denna ingrediens är emergens i kombination med algoritmisk informationsteori. Dessa senare begrepp kommer strax att diskuteras närmare; vi kan här helt kort konstatera att emergens relaterar till komplexa system med egenskaper som är svåra eller omöjliga att reducera till systemens delar och algoritmisk informationsteori handlar om relationer mellan information och beräkningskapacitet. Vi når slutsatsen att medvetandet är starkt epistemiskt emergent, vilket definitionsmässigt har konsekvensen att mind-body problemet inte kan lösas reduktionistiskt. Reduktionistisk förståelse av de subjektiva aspekterna av medvetandet, som introspektion eller qualia, är därmed inte möjlig. Begreppet 'explanatory gap' (Levine, 1983) är motiverat.

McGinn kommer, i sitt inflytelserika arbete "Can we solve the mind-body problem?" (McGinn, 1989) också till slutsatsen att medvetandet inte kan förstås, men på andra grunder. Han fokuserar på möjligheten att förstå det fenomenologiska medvetandet (Block, 1995) och drar slutsatsen att vi människor, på grund av 'cognitive closure' inte har förmåga att lösa detta 'hard problem of consciousness' (Chalmers, 1995). Med reservationen "the type of mind that can solve it is going to be very different from our" utesluter McGinn inte helt att det fenomenologiska medvetandet kan ges någon slags förklaring, vilken optimism alltså inte stöds i detta arbete.

Men hur kan viljans frihet relatera till mind-body problemet? I ett tankeexperiment visar vi här på en process, som i epistemisk mening är starkt emergent. Medvetandets neurologiska funktion har likheter med denna process och kan därför också visas vara starkt epistemiskt emergent. Dess extrema komplexitet leder dessutom till att det, som

globalt system betraktat, är ontologiskt emergent. De individuella processerna i hjärnan, som överföring av information mellan neuronerna, är deterministiska men hjärnan som system betraktat är icke-deterministiskt. Chalmers (2006) argumenterar utifrån intuitiva snarare än formella argument för att medvetandet är 'starkt emergent'; ett begrepp som används i samma betydelse som 'ontologiskt emergent'. Medvetandets ontologiskt emergenta karaktär löser upp den motsättning vi länge sett för viljans frihet; att å ena sidan måste våra viljehandlingar ske deterministiskt just så som vi önskar se dem utförda, men å andra sidan får determinismen inte tvinga oss till att ha de tankar och avsikter vi har. Detta förklarar det till synes märkliga valet av titel för detta arbete. Avgörande här är användningen av en omformulering av den vanligt förekommande definitionen "att kunna handla annorlunda" till en vetenskapligt testbar definition av fri vilja. Omformuleringen är nödvändig eftersom en individ med fri vilja måste kunna konstatera att exempelvis dess muskelrörelser utförs deterministiskt korrekt av det somatiska perifera nervsystemet trots att hjärnan, som del av centrala nervsystemet, har ett icke-deterministiskt beteende. Vi diskuterar här också den omständigheten att om vi verkligen kunde förstå medvetandet följer att viljan inte kan vara fri.

Definitioner är viktiga i detta arbete. Det finns åtminstone två skäl för att ge i hög grad explicita och tydliga definitioner inför resonemang om medvetandet och viljans frihet. Det första är att de begrepp som används inom områdena inte definieras på samma sätt av olika filosofer. Förutom att olika författare använder olika definitioner av medvetande, kan det ibland även vara oklart vilken definition som avses. Detta är delvis förstäligt utifrån att medvetandet, inte minst semantiskt sett, är ett svårfångat begrepp, men inte acceptabelt för hållbarheten i resonemang och slutsatser. Ett försök görs här att behandla mind-body problemet och viljans frihet utifrån en formaliserad grund i huvudsak relaterad till matematiken och fysiken. För att i möjligaste mån undvika placering i någon specifik filosofisk tradition eller skola, undviker vi emellertid i fortsättningen specificeringar som exempelvis 'fenomenalt medvetande' och 'qualia'. Den medvetandedefinition vi använder oss av är inte på något sätt uttömmande avseende medvetandets funktioner, utan försöker snarare isolera vad som skiljer medvetandet från alla andra fenomen.

Det andra skälet till behovet av förtydligande av definitioner är helt enkelt att vi, för att kunna framlägga bindande argument, kräver precision (Carnap, 1950, Scheffel, 2001 och 2010). Följden av en sådan precisering kan naturligtvis bli att vissa författares definitioner utesluts; resultaten i denna artikel måste alltså ses i det perspektivet.

Vi inleder med att diskutera vilka krav som måste ställas på en hållbar lösning av mind-body problemet. Därefter argumenteras att en sådan lösning inte kan finnas. Kärnan i argumentet är att medvetandets starkt *epistemiskt* emergenta karaktär omöjliggör förståelse i reduktionistisk mening. I nästkommande avsnitt visar vi att medvetandets *ontologiskt* emergenta karaktär löser upp den deterministiska motsättning vi länge sett för viljans frihet. Slutligen följer diskussion och sammanfattning.

Några nyckelbegrepp definieras i Appendix. Första gången de används i texten skrivs de som medvetandet^A och emergens^B.

2. VAD KRÄVS AV EN LÖSNING TILL MIND-BODY PROBLEMET?

Målet för forskningen rörande mind-body problemet är att finna en *teori* som förklarar relationen mellan mentala och fysiska egenskaper. Det delproblem som rönt avgjort mest intresse berör frågan hur vi ska förstå medvetandet^A. Vi kan inledningsvis ställa oss frågan: vad krävs då av en hållbar teori?

Chaitin (1987) har förtydligt innebörden i det nödvändiga kravet att en teori måste i sig vara mindre komplex än det den beskriver; i hans terminologi måste den i någon mån vara 'algoritmiskt kompressibel' i relation till vad den ska förklara. Låt oss illustrera med ett

exempel. Ett samband $y = f(x)$ har konstaterats för ett fenomen, men det precisa beroendet är oklart. En experimentserie, som genererat data om N punkter $(x_i, y_i); i: 1 \dots N$ har därför utförts. Det står helt klart att ett polynom $Y(x)$ av graden $N-1$ (med N koefficienter) alltid kan anpassas så att det kan dras genom samtliga datapunkterna i ett xy -diagram. Är det en teori? Svaret är nej, av den enkla anledningen att $Y(x)$ inte förklarar något; det går alltid att dra ett polynom av graden $N-1$ genom N datapunkter. Hade vi däremot kunna anpassa ett polynom av lägre gradtal genom samtliga punkter, säg ett andragradspolynom genom 10 punkter, då hade vi haft en teori värd namnet; den förutsäger mer än den definitionsmässigt måste. Att den är algoritmiskt kompressibel innebär att den kan formuleras med färre bitar information än $Y(x)$. Enkelt uttryckt: en teori måste vara enklare än det fenomen den beskriver, annars förklarar den inget.

Vi kommer i detta arbete använda den *diskreta logistiska funktionen* för jämförelser med de neurologiska funktioner som utgör grunden för medvetandet. Denna utgörs av det diskreta rekursiva sambandet $x_{n+1} = \lambda x_n(1-x_n)$, $n: 0 \dots n_{max}$, där det positiva heltalet n_{max} kan väljas fritt. Den logistiska ekvationen genererar alltså iterativt nya tal x_{n+1} för allt större tal n . Parametern λ och startvärdet x_0 måste först väljas. Vi kan nu fråga: finns det en explicit teori för värdet x_{n+1} , det vill säga finns det en funktion $u(k)$ som uppfyller $x_k = u(k)$ och som dessutom är algoritmiskt kompressibel jämfört med upprepad användning av det iterativa sambandet $x_{n+1} = \lambda x_n(1-x_n)$? Vi kan ju bilda $x_1 = \lambda x_0(1-x_0)$, $x_2 = \lambda x_1(1-x_1) = \lambda \lambda x_0(1-x_0)(1-\lambda x_0(1-x_0))$ och så vidare. Denna senare väg är inte framkomlig; för stora n blir x_{n+1} extremt komplext, så detta sätt att söka $u(k)$ resulterar inte i en giltig teori. Alternativt formulerat: de digitala bitar som behövs för att representera dessa uttryck är av minst samma storleksordning som de bitar som representerar talen $x_1, x_2, x_3 \dots$ själva! Dessvärre kan det för det allmänna fallet visas att frågan måste besvaras negativt; hur vi än försöker går det inte att härleda ett kompakt, explicit uttryck för $u(k)$.

Orsaken till problematiken är att den diskreta logistiska funktionen är icke-linjär. Låt oss i stället betrakta den enklare, linjära iterativa funktionen $x_{n+1} = A + \lambda x_n$, $n: 0 \dots n_{max}$, för vilken den allmänna termen x_k kan skrivas *explicit* helt enkelt som $x_k = x_0 \lambda^k + A(1-\lambda^k)/(1-\lambda)$ då $\lambda \neq 1$ och $x_k = x_0 + Ak$ då $\lambda = 1$. Denna lösning har uttryckts med några få matematiska tecken; den är alltså algoritmiskt kompressibel (kan representeras av färre digitala bitar information) jämfört med den lösning för x_k man får iterativt genom att bilda $x_1 = A + \lambda x_0$, $x_2 = A + \lambda x_1 = A + \lambda(A + \lambda x_0)$, $x_3 = A + \lambda x_2 = A + \lambda(A + \lambda(A + \lambda x_0))$ och så vidare. Lösningen för x_k har härletts matematiskt genom användandet av välkända axiom och teorem; vi kan därför säga att vi *teoretiskt* kan *förklara* värdena x_k .

Vi har hållit oss till exempel inom matematiken men resonemanget gäller generellt då vi söker formella förklaringar eller teorier för alla slags fenomen. Medvetandet kan alltså inte förstås genom att, i någon form av teori, relateras till system av samma komplexitet (som andra medvetanden); förståelse nås endast utifrån en teori som relaterar till redan välkänd kunskap och som är mindre komplex än medvetandet självt; det vill säga är algoritmiskt kompressibel relativt medvetandet.

3. EMERGENS STÅR I VÄGEN FÖR MIND-BODY PROBLEMETS LÖSNING

Medvetandets emergenta karaktär debatteras fortfarande i den filosofiska litteraturen (Kim 1999, Chalmers 2006, Kim 2006). Vi ska emellertid i det följande lyfta fram argument för att medvetandet är starkt epistemiskt emergent^{Be} och även ontologiskt emergent^{Bo}. Medvetandet har därmed egenskaper som inte är reducerbara till egenskaperna hos dess komponenter. Med det i sammanhanget vanliga uttrycket 'reducerbar till' avses att lägre-nivå-komponenternas egenskaper på ett tillräckligt sätt förklarar högre-nivå-fenomens egenskaper. Detta är givetvis av central betydelse för mind-body problemet, eftersom det avgör frågan om 'explanatory gap' (Levine, 1983); ett oöverbryggbart avstånd existerar

mellan de teorier vi kan formulera utifrån hjärnans fysiologi och medvetandets funktion. En konsekvens är att medvetandets beteende därmed är oförutsägbart, ett förhållande som är av betydelse då vi närmar oss problemet om viljans frihet. Vi återkommer till detta. Först ska vi diskutera medvetandets emergenta karaktär.

Emergens är kopplad till komplexitet. Hjärnan har cirka 100 miljarder nervceller (neuroner) som var och en är kopplad till tusentals andra nervceller via dendriter. En modell av medvetandet måste kunna hantera en motsvarande komplexitet. Algoritmisk informationsteori visar, som vi nyss såg, att 'modeller' eller 'teorier' som inte algoritmiskt kan komprimeras till lägre komplexitet än det de beskriver inte håller måttet. Det är emellertid inte helt klart hur emergenta egenskaper uppstår i komplexa system. Vår metod blir att, utifrån ett tankeexperiment, ge exempel på ett komplext system som har starkt emergenta egenskaper och som är väsentligt enklare än hjärnan. Detta leder i sin tur till medvetandets starkt epistemiskt emergenta karaktär.

Vårt tankeexperiment är följande. Vi föreställer oss ett antal robotar som är utplacerade på en isolerad ö. Alla robotar är konstruerade på samma sätt. De är programmerade för att fritt kunna gå omkring på ön och uträtta vissa sysslor. Robotarna kan kommunicera med varandra och är dessutom utrustade med instruktioner om att så effektivt som möjligt utföra sina sysslor. Om en robot blir mer effektiv genom utförandet av en viss handling ska den 'memorera' den och 'lära' de andra robotarna samma egenskap. Låt oss koncentrera oss på beteendet hos en av dessa robotar; vi kallar därför tankeexperimentet för 'Den hoppande roboten'.

Antag att det finns anledning att robotarna ska kunna röra sig relativt fritt. Vi låter nu deras rörelsemönster delvis styras av den logistiska ekvationen, som vi nyss bekantade oss med. Den iterativa ekvationen $x_{n+1} = \lambda x_n(1-x_n)$ genererar nya tal x_{n+1} i intervallet $[0,1]$ då x_0 (också i intervallet $[0,1]$) och λ är bestämda. Dessa får utgöra underlag för hur roboten ska koordinera sina leder, muskler och kroppsdelar, men roboten är programmerad att endast använda resulterande instruktioner som leder till att den säkert kan ta sig fram, utan att ramla. Låt oss sätta $\lambda = 4$. Det kan då visas matematiskt att, för så gott som varje val av x_0 , genereras en kaotisk talföljd i intervallet $[0,1]$ redan för måttligt stora n . En sådan karakteriseras av att den inte är algoritmiskt kompressibel (Chaitin, 1987); det finns ingen teori som kan förutsäga värdet på x_k för stora k . Betraktar vi konsekutiva x_k, x_{k+1}, x_{k+2} och så vidare, kommer dessa tal att te sig fullkomligt slumpmässiga. Men det intressanta och viktiga är att de faktiskt är deterministiska; varje tal i talföljden är entydigt definerat av det tidigare och så vidare i en lång kedja. Låt oss sammanfatta: i det allmänna fallet finns det inget inget algoritmiskt kompressibelt explicit uttryck $x_k = u(k)$ för den logistiska ekvationen och för fallet $\lambda = 4$ finns det heller ingen empirisk möjlighet att finna något samband för de genererade värdena.

Låt oss anta att det skulle vara av stort värde om robotarna kunde utföra hopp utan att ramla. Av den anledningen prövas en enskild robot. Utifrån ett stort antal x_0 genereras talsekvenser i förhoppningen om att någon av dessa sekvenser ska motsvara rörelser, som när de sammansätts resulterar i att roboten utför ett hopp. Vi bortser här från att förfarandet uppenbarligen är omständligt; komplexiteten orsakas av att robotarna består av en mycket stor mängd leder, muskler och andra kroppsdelar som ska koordineras men också av att det är oklart vilka rörelser som roboten behöver utföra för att genomföra ett lyckat hopp samt att den logistiska ekvationen inte möjliggör kontroll av rörelserna. Efter ett stort antal försök ges företaget upp; robotarna kan inte 'läras' att hoppa.

Initiera nu i stället robotarna (med x_0) och lämna dem för sig själva en tid på ön, varefter vi återvänder. Till vår förvåning ser vi nu att robotarna i många fall tar sig fram inte bara genom att gå, utan även genom att hoppa över hinder. De har alltså utvecklat en emergent egenskap. Vi kan inte förklara hur en eller flera av robotarna fick egenskapen att kunna

utföra hopp; det finns ingen teori för det. Som vi såg, kunde vi heller inte simulera beteendet. I så fall hade detta varit svagt epistemiskt emergent. Robotens beteende är därmed *starkt epistemiskt emergent*. Däremot är beteendet inte ontologiskt emergent; vi vet ju att robotens egenskap att kunna hoppa verkligen är reducerbar till dess delar. Liknande slutsatser angående emergens för ickelinjära system har nåtts av andra författare (Silberstein och McGeever, 1999).

Vad har detta exempel för relevans för mind-body-problemet? Förvisso har vi givit ett exempel på starkt epistemiskt emergenta system, men vi skulle ju kunna studera hoppande robotar i detalj och bygga kopior som verkligen utför samma rörelsemönster och som därför också kan hoppa. Kan vi inte på samma sätt konstruera ett medvetande? Nja, i fallet med den hoppande roboten är det tydligt vilken emergent *egenskap* som det emergenta *tillståndet* leder till, eftersom egenskapen 'att hoppa' är externt mätbar. Men de emergenta tillstånden i hjärnan kan dessvärre inte på samma sätt sammankopplas med externt mätbara egenskaper. Vi kan inte, som för roboten, "räkna baklänges" utifrån medvetandets funktioner (exempelvis uppfattning av smärta eller introspektion) för att kartlägga de emergenta tillstånd som ger upphov till dessa. Detta måste vi ju kunna göra om vi ska ta fram en teori för medvetandet, om vi vill förstå det. En adekvat teori måste ha prediktiva egenskaper. För att sammanfatta: medvetandet som emergent tillstånd låter sig inte kopieras eftersom vi inte har extern tillgång till dess associerade emergenta egenskaper; de senare uppfattas endast internt. Medvetandets funktion är en sammansättning av ett mycket stort antal emergenta tillstånd, vilka vi alltså inte kan fånga i en teori.

Låt oss nu diskutera graden av medvetandets emergens. Den mänskliga hjärnan är ofantligt mycket mer komplex än den logistiska ekvationen. I huvuddrag kommunicerar de cirka 100 miljarder neuronerna som följer. Via så kallade dendriter kan varje neuron nås av elektrokemiska signaler från tiotals till tiotusentals (i medel cirka 7000) närliggande neuron. Bidragen från dessa signaler viktas samman i neuronens cellkropp till en elektrisk potential. När denna når ett visst tröskelvärde, skickar neuron ut en puls i en nervtråd, som sedan kan koppla via synapser till flera andra cellers dendriter. Den utgående signalen från en neuron är alltså en stegfunktion snarare än en kontinuerlig, ickelinjär funktion av den inkommande signalen. Detta motiverar valet av den diskreta logistiska ekvationen snarare än den kontinuerliga i tankeexperimentet. Neuroner kan skicka cirka 100 signaler i sekunden med en hastighet av upp till 100 meter per sekund. Beteendet varierar något från neuron till neuron. Det har visat sig att icke-linjära funktioner benämnda sigmoider, med ett S-format beroende av insignalen, ger realistiska modeller för nätverk av neuron (Wikipedia 1, 2016). Sigmoider är, matematiskt sett, nära besläktade med den logistiska ekvationen. Summerar vi detta, står det klart att nätverket av hjärnans neuron kommunicerar ickelinjärt med en komplexitet som vida överstiger vår enkla logistiska ekvation. Medvetandet, och för all del vårt undermedvetna, är *starkt epistemiskt emergent*. På samma sätt som vi inte kunde förklara den hoppande robotens beteende kan vi inte heller förklara medvetandets funktion; mind-body-problemet är olösbart.

Vi kan nu fråga oss om medvetandet också är *ontologiskt emergent*, det vill säga: är det så att medvetandets beteende i princip inte kan reduceras till de processer på lägre nivå som utgör grunden för medvetandet, som medvetandet supervenierar på? För att besvara den frågan behöver vi titta lite närmare på vad som avses med 'inte kan reduceras till'. Låt oss återvända till exemplet med Den hoppande roboten. Egenskapen att kunna hoppa var i detta fall inte ontologiskt emergent av den anledningen att i 'objektiv mening' fanns detta som ett alternativ för systemet, trots att möjligheten var okänd för oss. Med 'objektiv mening' avser vi att de olika möjliga sekvenser av tal, varav minst någon leder till ett hoppande beteende, som kan genereras av den logistiska ekvationen motsvarar en informationsmängd som är avsevärt mindre än den som principiellt kan hanteras. Det har

visats (Lloyd 2002, Davies 2004) att universums informationslagringskapacitet är begränsad utifrån de tillgängliga kvanttillstånden i den materia som befinner sig innanför den kausala horisonten, det vill säga det avstånd som begränsas av ljusets ändliga hastighet. Utifrån flera argument visas att storleksordningen 10^{120} digitala bitar information ryms inom denna horisont. Att detta resultat endast är en uppskattning är ju inte väsentligt; det avgörande är att informationslagringskapaciteten är begränsad till en storleksordning av denna karaktär. För att ett beteende ska kunna karakteriseras som ontologiskt emergent måste det alltså motsvara en komplexitet som övergår storleksordningen 10^{120} bitar; först då kan vi säga att det inte finns någon principiell *möjlighet* att 'reducera' detta till de lägre-nivå-fenomen som det bygger på.

Det är svårt att avgöra om tänkande som företeelse är ontologiskt emergent eller inte. Men specifika tankar har den egenskapen. Säg att en tusendel av hjärnans 10^{11} neuroner är delaktiga i formerandet av en viss tanke. Säg också att det finns en viss redundans, eller klusterbeteende, så att endast en tusendel av dessa i sin tur behöver aktivt styras för att generera denna tanke. Vi kan exempelvis tänka oss att specificera en process där en individ på goda grunder ska välja en viss rätt från en 20-rättersmeny. Då återstår $N = 10^5$ neuroner att initieras. Säg att dessas tillstånd är binära; endast två tillstånd existerar för var och en. Det innebär att vi har 2^N olika tillstånd att initiera på ett korrekt sätt så att individen genomför den avsedda tankeprocessen. Detta antal är mycket större än 10^{120} , vilket alltså innebär att specificerade tankar är ontologiskt emergenta. Detta möjliggör en 'nedåtriktad kausalitet' (den engelska termen är 'downward causation'). Vi återvänder strax till detta.

Det är intressant att emergens sällan associeras med resultat av mänsklig verksamhet, med *design*, utan snarare med *evolution*; den utveckling som sker i naturen. Evolutionen har genom det naturliga urvalet tillgång till en oerhörd mångfald av frihetsgrader och har därmed fantastiska förutsättningar att generera emergenta system. Ett exempel från kemin utgörs av myoglobin, en viktig syrebindande molekyl i muskelvävnad (Luisi, 2002). Här har 153 aminosyror sammankopplats i en så kallad polypeptidkedja. Eftersom det finns 20 olika aminosyror, utgör antalet möjliga kombinationer av kedjor det enorma talet $20^{153} \approx 10^{199}$, som motsvarar ett antal digitala bitar mycket större än 10^{120} . Myoglobin är alltså ett ontologiskt emergent fenomen i naturen; molekylen är inte reducerbar till sina lägre-nivå-fenomen. Utifrån sin funktion kunde den endast evolvera, inte designas.

Sammanfattningsvis kommunicerar hjärnans neuronnätverk i ickelinjära processer för att generera kognitiva funktioner som exempelvis förmåga att känna smärta och möjlighet till introspektion. Om processerna vore linjära skulle deras beteende, som vi såg ovan, möjligen kunna reduceras till en teori. En sådan skulle verkligen ha lägre komplexitet än vad den beskriver eftersom den skulle vara algoritmiskt komprimerbar. Ickelinjära system kan emellertid, med få enkla undantag, inte ges en explicit lösning. Eftersom neuronnätverket associerat med medvetandet har konstaterats vara ickelinjärt och inte bara starkt epistemiskt emergent utan även ontologiskt emergent kan vi heller inte konstruera en teori för det. Medvetandet kommer inte att kunna förstås.

Det kan vara intressant att kort diskutera en helt annan besvärande omständighet för förståelsen av medvetandet. Vi har hittills talat om teorier för medvetandet, om försök till förståelse av medvetandet självt. När denna väg visat sig vara oframkomlig, kan det i stället vara av intresse att på artificiell väg försöka konstruera medvetna processer. Inom neurovetenskapen söker man efter "neural correlates of consciousness" (NCC), som utgörs av nervprocesser i hjärnan som är direkt kopplade till individens aktuella tankeverksamhet. Låt säga att dessa verkligen kan kartläggas i tillräcklig utsträckning för att försök att skapa medvetna processer i artificiella hjärnor kan utföras. Vid varje sådant experimentellt försök måste funktionen säkerställas; systemet måste diagnosticeras. Annars finns naturligtvis möjligheten att vi har konstruerat ett avancerat system som

externt beter sig som ett medvetande men som saknar mentala processer. Ingen principiell gräns finns exempelvis för icke-kognitiv "intelligens" hos avancerade dataprogram. Dessa skulle alltså kunna passera varje form av Turingtest. I ett sådant godkänns varje maskin som ger liknande responser på tilltal som en människa; maskinen är dold så att den som utför testet vet inte om den talar med en människa eller en maskin.

Men hur skulle en tillräcklig diagnosticering kunna se ut? Gångse medvetandedefinitioner ger svaret: vi ska kunna säkerställa att systemet kan uppleva sig självt. Men eftersom all mätning av medvetandets funktioner måste ske externt, det vill säga av laboratoriepersonal som använder diagnostisk apparatur, kan inte systemets interna erfarenheter av avancerade medvetandefunktioner som exempelvis introspektion mätas i experiment-situationen. Det finns helt enkelt ingen information från systemet som på ett säkert sätt skulle kunna skiljas från den som kan produceras av ett avancerat, men omedvetet, datorprogram. Vi skulle ju kunna ha byggt en intelligent robot, som saknar möjligheten till medvetna beteenden som introspektion. Detta är alltså ingen framkomlig väg för mind-body problemets lösning. Vi kan summera detta argument: förståelse av ett system implicerar möjlighet att konstruera det, med alla dess funktioner. Men eftersom den avsedda funktionen (att vara medvetet) hos det system vi har konstruerat inte säkert kan experimentellt verifieras, kan vi inte heller med säkerhet säga att vi förstår det.

4. KONSEKVENSER FÖR VILJANS FRIHET

Är vi tvungna att ha de tankar vi har? Händer endast våra tankar, snarare än att vi själva skapar dem? Har determinismen försatt vår vilja i ett järngrepp? Fri-vilje-problemet är det kanske viktigaste existentiella problemet och har genererat hyllkilometer av litteratur genom århundradena. En orsak till problematiken utgör faktiskt redan den vanligaste definitionen av fri vilja: "möjlighet att handla annorlunda". Det finns nämligen knappast någon möjlighet att med vetenskapliga metoder avgöra sanningsvärdet av den satsen. Hur ska vi få kunskap om huruvida en individs handlingar är autonoma eller förutbestämda? Och varför skulle ens ett fritt handlande medvetande agera annorlunda om två situationer det ställs inför är identiska? Många argument om viljan leder ut i sankmarker. Ett steg i en mer konkret riktning utgörs av varianten: 'möjlighet till val som inte är kausalt nödvändiga eller gudomligt bestämda', även om det är oklart vad som avses med icke kausala val.

Vi väljer här att försöka ge en vetenskapligt avgörbar definition av fri vilja: *En medveten individ har fri vilja om dess beteende äger rum utifrån dess intentioner, om intentionerna inte är undermedvetet genererade och om individens beteende är ontologiskt indeterministiskt.*

Definitionen kan motiveras på följande sätt. Determinism på "lägre-nivå" (se definitionen av emergens) antas gälla. Men om den medvetna individens handlingar är epistemiskt eller *ontologiskt deterministiska* är det uppenbart att dess vilja lyder under lagar utanför dess medvetna kontroll. Detta förhållande ingår i det klassiska, deterministiska argumentet mot den fria viljan. Med 'ontologiskt indeterministisk' menas att beteendet inte är reducerbart till det aktuella fysiska, neurala tillstånd som medvetandet supervenierar på. En sådan individs agerande är alltså ontologiskt obestämt på förhand. Definitionens relation till kompatibilism diskuteras i avsnitt 5.

Vidare får ett fritt medvetandes önskade handlingar inte förvandlas till något annat än de avsedda; beteendet måste ske *utifrån dess intentioner*. Vill jag fundera över vad jag ska äta till middag, måste en sådan fundering vara möjlig. Därför måste också mitt handlande på ett konsekvent och adekvat sätt följa min vilja. Lydelsen 'äger rum utifrån dess intentioner' är avsiktligt något vagt formulerad; den precision i våra handlingar vi ofta önskar och eftersträvar uppnås ibland inte; detta är inte ett resultat av att viljan inte åtlöds

utan av våra fysiska och psykologiska begränsningar. Notera att vi också förutsätter att vi diskuterar medvetna individer; det är inte meningsfullt att tala om 'vilja' för andra system.

Villkoret att ”*intentionerna inte är undermedvetet genererade*” behövs för att säkerställa att individens 'hjärna' inte innehåller något dolt omedvetet system, som styr det på så vis att medvetandet, trots styrningen, upplever intentionerna som sina egna. Så kallade 'karaktärsbeslut', det vill säga sådana beslut vi fattar, utan egentlig aktiv reflektion, utifrån våra samlade erfarenheter och konsoliderade ståndpunkter behandlar vi i det här sammanhanget som medvetna. Vi återkommer strax till dessa.

Sammanfattningsvis ger vi 'möjlighet att handla annorlunda' en alternativ, vetenskapligt avgörbar formulering som skapar bättre metodiska förutsättningar att behandla fri-vilje-problemet. Uppgiften innebär nu inte minst att ta itu med den hämmande faktorn att våra medvetanden måste ha en *deterministisk* karaktär för att tankeprocesserna ska vara tillräckligt koherenta och för att det vi bestämmer oss för att göra verkligen utförs på avsett vis, men samtidigt måste de kunna agera *icke-deterministiskt* utifrån de givna, externa förhållandena så att vi inte var tvungna att bestämma oss för just den handlingen. En fri vilja förutsätter att denna potentiella motsättning upplöses.

Medvetandets emergenta karaktär kommer nu att vara av avgörande betydelse. Vi kan i sammanhanget också konstatera att om det verkligen hade varit fallet att en teori för medvetandet kunde konstrueras så skulle viljan inte kunna vara fri. *Fri vilja förutsätter att medvetandet är starkt epistemiskt emergent*, ett faktum som borde få mer ljus i litteraturen. Anledningen är naturligtvis att om individens beteende vore epistemiskt beräkningsbart eller simulerbart så vore beteendet också deterministiskt förutsägbart.

Begreppet *ontologisk* determinism måste användas här; medan *epistemisk* determinism avser beräkningsbarheten av framtida tillstånd utifrån vår nuvarande kännedom av dem, avser ontologisk determinism snarare just de entydigt bestämda, faktiska framtida tillstånd som resulterar, oberoende av vår beräkningsförmåga. Bollar som kolliderar på ett biljardbord är ett enkelt exempel; även om vi kan approximativt beräkna bollarnas banor gör faktorer som underlagets molekylära beskaffenhet det omöjligt för oss att få kunskap om den exakta dynamiken. Ska vi vara precisa, bör vi ta hänsyn till att kvantmekaniken visar att determinism inte gäller på mikronivå; naturen är i någon mån 'suddig' när enskilda partiklar studeras. Men för större anhopningar av partiklar, som de molekyler som bygger upp neuronerna, får denna effekt mycket mindre betydelse, på grund av så kallad kvantdekoherens. Begreppet 'adequate determinism' (Wikipedia 2, 2016) har myntats för att understryka att den statistiska determinism som resulterar och används här, i huvudsak är korrekt i den makroskopiska världen, även om kvantfenomen kan vara viktiga för mycket små system.

Vi ska nu argumentera för att medvetandet, som globalt system, är ontologiskt indeterministiskt. Vi återvänder alltså till att studera hjärnans neurala processer. Hur kan de vara indeterministiska på högre-nivå? Människan och medvetandet utgör ju en del av den fysiska världen. Låt oss ägna en stund åt att studera hur den är fundamentalt beskaffad.

Vår erfarenhet visar att kausalitet gäller. Det innebär att ett fysiskt systems tillstånd, i termer av dess mikroskopiska beståndsdelars lägen och hastigheter, i sig är tillräckligt för att föra det till något påföljande tillstånd; orsak ger verkan. Om det påföljande tillståndet är entydigt förutbestämt utifrån det aktuella fysiska systemets tillstånd, talar vi om determinism. Vardagshändelser, som exempelvis när biljardkön träffar köbollen som sedan slår ned den gula bollen i hålet, får oss lätt att tro att kausalitet och determinism är samma sak. Så är det emellertid inte. Ett fysiskt systems påföljande tillstånd kan vara obestämt, även om vi bortser från den statistiska karaktären av kvantmekaniken som exempelvis visar sig vid radioaktivt sönderfall.

Vardagserfarenheten får oss lätt att tillskriva naturen egenskaper som den inte har. Som att krafter verkar ögonblickligen. All kraftpåverkan i naturen, det vill säga förändringar av tillstånd, sker för de fyra grundläggande formerna av kraftväxelverkan genom utbyte av partiklar som kallas bosoner. Denna växelverkan tar en ändlig tid, vars nedre gräns ges av en viktig relation inom kvantmekaniken; Heisenbergs osäkerhetsprincip (Lloyd, 2002). Gränstiden är proportionell mot Plancks konstant och omvänt proportionell mot systemets medelenergi ovanför dess grundtillstånd, vilket för ett enkilos system innebär högst $5 \cdot 10^{50}$ möjliga tillståndsförändringar under en sekund. I hela det synliga universum, som uppstod för cirka 10 miljarder år sedan med massan 10^{53} kg har totalt sett maximalt cirka 10^{121} tillståndsförändringar genomförts. Även om detta är ett enormt tal är det inte oändligt. Universums "beräkningskapacitet" är alltså begränsad av Plancks konstant och den tillgängliga energin. Ett materiellt system med ett relativt måttligt antal frihetsgrader kan alltså inte "förutse" sin egen framtid, inte ens i ontologisk mening. Resultatet är rimligt utifrån det faktum att universums informationskapacitet är begränsad; hade naturen kontinuerligt rört sig mellan olika tillstånd hade detta ju motsvarat en oändlig mängd information. De krafter som verkar i naturen har alltså endast varit aktiva ett ändligt antal gånger. Varje kraftpåverkan är associerad med en fördröjning. Krafter verkar inte kontinuerligt. Detta är viktigt när vi talar om determinism.

Vi har konstaterat att ontologiskt emergenta fenomen, som medvetandet, kan utvecklas. Emergenta fenomen är komplexa och deras fysiska växelverkan involverar ett mycket stort antal komponenter. Låt oss nu betrakta ett fysiskt förlopp i naturen där växelverkan med ett emergent fenomen kan ingå. Antag att systemet är isolerat från omvärlden. Vi frågar oss: beter sig systemet ontologiskt deterministiskt, det vill säga är systemets utveckling i tiden objektivt förutbestämt utifrån dess tillstånd vid en viss tidpunkt? Svaret är att det beror på. Vi har all anledning att förvänta oss deterministiskt beteende så länge inga överraskningar dyker upp, det vill säga för lägre-nivå-fenomen där emergenta fenomen inte spelar någon roll. Men är ett ontologiskt emergent fenomen verksamt i systemet kan dess dynamik utvecklas på ett sätt som inte är kausalt förutbestämt i den meningen att systemet i ontologisk mening definierar en unik utveckling. Detta är, som vi diskuterat tidigare, ytterst sett ett resultat av universums ändliga informationskapacitet. Vid varje ögonblick av kausal kraftpåverkan kan alltså ontologiskt emergenta fenomen inverka och på så sätt utöva 'downward causation'.

Låt oss nu återgå till frågan om medvetandet är ontologiskt deterministiskt och hur de överväganden vi nyss gjort kommer in. Till en början kan vi konstatera att öppna system, det vill säga system som har kontakt med någon omgivning, som regel är indeterministiska - detta eftersom de hela tiden kan störas av inkommande påverkan. För att diskutera fri vilja kan det vara lämpligt att föreställa sig beteendet hos en enskild människa som placeras i ett slutet rum, utan kontakt med omvärlden. Speciellt undrar vi om personens handlingar är deterministiska. Hade vi i stället placerat en enkel maskin i rummet, hade den varit epistemiskt determinerbar; givet dess initialtillstånd, skulle vi kunna beräkna dess framtida tillstånd. En avancerad maskin som Den hoppande roboten hade inte betett sig epistemiskt deterministiskt, men däremot ontologiskt deterministiskt eftersom dess agerande i princip är bestämbar redan från den första iterationen. Men varför skulle då människans beteende inte vara ontologiskt deterministiskt? Vi inser att ett nödvändigt krav är att det åtminstone delvis styrs av ontologiskt emergenta system; annars är det ju determinerbart på samma sätt som roboten.

Här kan det hjälpa med ett par analogier. 1: Vatten strömmar via ett antal flodfåror in i en underdimensionerad damm. Kommer dammväggen att brista? 2: Snö faller mot en snötäckt bergstopp. Kommer en lavin att utlösas? Och kommer den i sin tur att utlösa fler laviner? Vi vet att *något* kommer att hända; dammen håller eller brister, snön ligger stilla eller en lavin inleds. Men för att avgöra *vad* som kommer att ske måste en enorm mängd

olika möjliga fall av icke-linjära processer tas hänsyn till där var och en av mer än miljoner miljoner molekyler är inblandade. 'Vet' naturen själv vad som kommer att hända?

Låt oss återgå till tankeexperimentet med den logistiska ekvationen och Den hoppande roboten. Vi konstaterade att varje iterationssteg för den logistiska ekvationen var fullkomligt deterministiskt. Ändå resulterade, efter ett stort antal sådana händelser, ett oväntat emergent beteende i epistemisk mening. Men ontologiskt sett var beteendet inte emergent; det var i princip bestämbar redan från den första iterationen. Den dansande robotens beteende är alltså ontologiskt deterministiskt även om det inte är epistemiskt deterministiskt. Det verkar finnas något slags 'gap' som gör att efter ett antal iterationer övergår det deterministiska beteendet för de enskilda iterationsstegen i epistemisk obestämbarhet eller indeterminism för det tidsutvecklade systemet.

När vi sedan studerade medvetandets fysiologi, upptäckte vi att den enorma komplexiteten på neural nivå tilldelar medvetandet en ontologiskt emergent natur. Men åter, emergens är en utveckling över tid, medan determinism handlar om förutbestämbarhet vid övergången från ett tillstånd till nästa. Ontologisk emergens borde inte automatiskt leda till ontologisk indeterminism. Utifrån vår fysikaliska utvidgning ovan kan vi dock konstatera att sett till sina kausala resultat är de komplexa och emergenta hjärnprocesser, som exempelvis associeras med självreflektion, obestämda i det ögonblick vi går in i denna. På ett liknande sätt som för dammen och den snöklädda bergstoppen, vars framtid är indeterministiska trots att kausalitet gäller, är det ontologiskt obestämt vilken process som leder till nästa tillstånd. Varje neuron kommunicerar i genomsnitt med 7000 andra neuroner. Neuronen A:s framtida tillstånd bestäms av alla dessa neuroner och hade varit deterministiskt om växelverkan skedde på låg-nivå. Men nu vet vi att ontologiskt emergenta fenomen kan komma att påverka situationen. Dessa fenomen är oförutsägbara och därför är medvetandeprocesserna inte deterministiska på högre-nivå trots att de är kausala.

För att summera möjliggörs alltså ontologiskt indeterministiska processer på högre-nivå i hjärnan av den ontologiskt emergenta karaktären hos processerna i det komplexa neurala nätverket i kombination med begränsningarna i naturens kapacitet för informationshantering och tillståndsförändring.

Fri vilja kräver också, enligt definitionen ovan, att individens beteende äger rum *utifrån dess intentioner*. Detta villkor är inte problematiskt; det uppfylls utifrån våra erfarenheter. Individens vardagsfunktion är helt beroende av att hon konsekvent gör det hon bestämmer sig för. Vill hon göra sig en kopp kaffe så gör hon också det. De undantag vi kan anföra, som att det kanske inte finns något kaffe hemma eller att hon inte orkar, handlar inte om mentala begränsningar utan om fysisk oförmåga eller begränsningar i omvärlden.

Hittills har vi presenterat stöd för att det kombinerade medvetandet / undermedvetandet uppfyller fri-vilje-kraven. Men få skulle se detta som tillräckligt; om våra viljebeslut omedvetet dikterades för oss, med en upplevelse av att vara våra egna, är det svårt att tala om en fri vilja. Vi kommer nu att argumentera för att medvetandet, åtminstone i vissa situationer, utövar sin vilja utan avgörande inflytande från det undermedvetna. Erfarenhetsmässigt kan det undermedvetna arbeta oberoende av medvetandet; drömmar visar på det. Det undermedvetna och medvetandet kan också samarbeta, som exempelvis under normal bilkörning. Och medvetandet kan, åtminstone synbarligen, arbeta oberoende av det undermedvetna; under intensiva diskussioner exempelvis. Men medvetandets oberoende roll har ifrågasatts på många håll under de senaste decennierna och flera författare talar om "illusionen av fri vilja". Stöd för detta hämtas från neurovetenskaplig forskning. En "beredskapspotential", som aktiveras omedvetet i god tid innan vi fattar medvetna beslut, förefaller skvallra om att det huvudsakliga beslutsfattandet sker bortom medvetandet. En föregångare inom området var Libet (1985). De experiment som utförts

inom området har emellertid många möjliga felkällor och har fått utstå allvarlig kritik på flera håll (Klemm, 2010).

I vissa praktiska situationer, inte minst ur evolutionär synpunkt, är det avgörande att medvetandet får agera ostört. Snabba beslut då vi kör bil och plötsligt behöver undvika en olycka är ett bra exempel. Vi måste inte endast snabbt utföra ett stort antal överväganden; det undermedvetna hinner inte, med sin fördröjda verkan, inhämta all relevant information för att överblicka situationen och på kort tid leverera adekvata beslut som inte konflikterar med medvetandets uppfattning och hantering av situationen. Vidare är det välkänt att, vid inläring av kunskaper och färdigheter, utförandet alltmer tas över av det undermedvetna under det att vi blir kunnigare eller skickligare. Men för en nybörjare som sätter sig ned vid ett piano, är det undermedvetna helt oförberett. Det finns absolut ingen möjlighet för det undermedvetna att kontrollera fingerrörelserna eftersom det inte 'vet' vad som ska utföras (Klemm, 2010). Det behövs mer forskning för att identifiera undermedvetandets inverkan på våra handlingar, men det finns alltså exempel där det undermedvetna rimligtvis inte kan ha någon betydande roll.

Samarbetet mellan medvetandet och det undermedvetna leder till det andra argumentet för att medvetandet inte behöver vara deterministiskt kontrollerat av det undermedvetna. Neurovetenskapen visar att en betydande del av de 'processorer' i hjärnan som används för medvetet tänkande också används för undermedvetna processer (Dehaene, 2014). Detta ger alltså stöd för att inte endast de globala neurala medvetandeprocesserna utan även de globala undermedvetna processerna är både ontologiskt emergenta och indeterministiska. Kopplingen mellan medvetandet och det undermedvetna är därmed inte ontologiskt deterministisk; det undermedvetna kan utgöra en förutsättning för medvetandets aktivitet, men medvetandets agerande är därmed inte fullständigt bestämt av det undermedvetna. Vår erfarenhet säger också att vi medvetet kan avbryta impulsstyrda intentioner.

Vi har diskuterat interaktionen mellan medvetandet och det undermedvetna. Men vi behöver inte nödvändigtvis göra en distinktion mellan två tydligt separerade system. Även enskilda neurologiska delsystem direkt associerade till medvetandet självt är tillräckligt komplexa (ontologiskt emergenta) för att deras interagerande, som system betraktat, bör kunna karakteriseras som ontologiskt icke-deterministiskt. Inom spelteorin har, intressant nog, liknande resultat framkommit. Emergent beteende har observerats vid simuleringar av den ickelinjära interaktionen mellan två spelare, som dels agerar utifrån att optimera sitt spelande, dels försöker bete sig oförutsägbart relativt opponenter, om spelarna tillåts utnyttja spelets historia (West och Lebiere, 2001).

En komplikation relaterad till distinktionen mellan undermedvetna och medvetna beslut utgörs av vad vi kan kalla karaktärsbeslut. Utifrån sina erfarenheter och tidigare reflektioner ackumulerar människor olika dispositioner eller karaktärer som kan leda till rutinmässigt agerande i vissa situationer. Inför exempelvis en uppdykande fara väljer vissa människor som regel att fly medan andra företrädesvis stannar för att bekämpa faran. Detta utgör alltså inte ett aktivt medvetet val av den typen vi hittills diskuterat utan snarare ett passivt rutinmässigt val, grundat på individens disposition för handlande i sådana situationer. Eftersom individen rimligtvis är medveten om sin karaktär väljer vi här att betrakta dessa karaktärsbeslut som medvetna snarare än undermedvetna.

Vår vilja är alltså inte deterministiskt förutbestämd. Våra tankar och viljeyttringar händer oss inte bara. De utvecklas ontologiskt indeterministiskt i ett samarbete mellan medvetandet och det undermedvetna. Även om neurovetenskapen idag ännu inte klarlagt graden av det undermedvetnas betydelse står det klart att medvetandet i vissa situationer inte konsulterar någon omedveten process och att det, i kraft av sin ontologiskt indeterministiska karaktär, har möjlighet att oberoende korrigera eller avfärda den påverkan som utövas av det undermedvetna; det kan utöva "free won't" (Libet, 1985).

Men hur kan våra tankar och viljeyttringar formas på ett strukturerat och koherent sätt om processen för detta inte är deterministisk ur ett reduktionistiskt perspektiv? Hur kommer det sig att individen på ett globalt indeterministiskt sätt kan genomföra sina intentioner utan att kontrolleras av det undermedvetna? Utgör karaktärsbeslut exempel på de ontologiskt emergenta val som kännetecknar en fri vilja? Frågorna är viktiga men analyseras inte här; det som är av intresse för frågan om viljans frihet är att tankar och viljeyttringar uppstår i en miljö där de inte var tvungna att uppstå.

5. DISKUSSION

Slutsatserna i detta arbete pekar otvetydigt mot icke-reduktionistisk fysikalism; mentala tillstånd supervenierar på fysiska tillstånd men kan inte reduceras till dem. Analysen är ny på så sätt att den bygger på resultat inom matematiken och fysiken. Vi kan undra vilka konsekvenserna blir för kausal slutenhet, det vill säga tesen att inga fysiska händelser har orsaker utanför den fysiska världen. Svaret blir att den fysiska världen verkligen är kausalt slutet i ontologisk mening eftersom vi har all anledning att uppfatta den som just kausal; varje materiellt tillstånd leder, tillsammans med naturlagarna, till något nytt tillstånd. För enklare system, på låg nivå, är de nya tillstånden även deterministiska, alltså principiellt förutbestämda och ibland även beräkningsbara. Människohjärnan använder exempelvis deterministiska lägre-nivå-processer på neural nivå för tankeprocesser, för utförande av handlingar (somatiska nervsystemet) och för reflexer (autonoma nervsystemet). Men det innebär, som vi visat, inte att alla system i den fysiska världen är deterministiska. Emergens kan ändra på det förhållandet. Medvetandet, som visats vara ontologiskt emergent, är just ett sådant högre-nivå-system. Det innebär ur ett epistemiskt perspektiv att möjligheten till kausala slutsatser om mentala systems funktioner är begränsade. Situationen påminner om den som gäller för matematiken för vilken Gödel visat att det finns sanna satser inom systemet som är obevisbara på grund av sin komplexitet.

Vi kan nu förklara varför emergens inte innebär överbestämmdhet vad gäller medvetandets kausala situation (Kim, 2006). Det kan ju argumenteras att om medvetandets dynamik är entydigt determinerat av dess aktuella tillstånd plus naturlagarna så kan inte emergenta fenomen existera oberoende; de måste i så fall kunna beskrivas av de fakta vi redan har om situationen. Annars förefaller vi stå inför ett överbestämt problem. Lösningen till dilemmat är att de emergenta egenskaperna är av samma karaktär som de nya villkor som kan presentera sig då ett slutet system övergår till ett öppet system. I exemplet med personen i det slutna rummet skulle detta motsvaras av att dörren öppnas. Emergenta egenskaper har alltså, vad determinismen beträffar, samma betydelse för systemets utveckling som extern påverkan i ett öppet system. Lösningen till problemet med överbestämmdhet förklarar också hur "downward causation" (Kim, 2006) kan äga rum. Samverkande emergenta fenomen determinerar utvecklingen av systemet (medvetandet) utan att styras av den kausala situationen.

Var hamnar då kompatibilismen, det vill säga ståndpunkten att determinism är förenlig med fri vilja? Intressant nog behöver inte frågorna om kompatibilism och inkompatibilism diskuteras eftersom även om neurala lägre-nivå-processer är deterministiska så är högre-nivå-fenomenet medvetandet ontologiskt indeterministiskt. Hur medvetandet går från ett tillstånd till nästa är, som vi har försökt visa i denna artikel, ontologiskt obestämt.

Slutligen, hur förhåller sig dessa resultat till epifenomenalismen, det vill säga uppfattningen att mentala tillstånd endast är biprodukter av de fysiska och att mentala tillstånd därmed saknar möjlighet till kausal påverkan av fysiska tillstånd? För att besvara denna fråga behöver vi konstatera att den icke-reduktionistiska fysikalism som här impliceras inte är en form av egenskapsdualism. Visserligen är mentala tillstånd oreducerbart mappade till neurobiologiska tillstånd, men de korresponderar entydigt till

fysiska tillstånd, ehuru mycket komplexa. Icke-reduktionismen följer av de mentala tillståndens emergenta karaktär, inte av att en korrespondens skulle saknas mellan fysiska och mentala tillstånd. Fysikalism måste, som i detta arbete, rimligtvis vara icke-reduktionistisk för att kunna utesluta den epifenomenalistiska uppfattningen.

6. SAMMANFATTNING

I en reduktionistisk analys av mind-body problemet används emergensbegreppet och algoritmisk informationsteori för att visa att problemet är principiellt olösbart. Hjärnans neurala komplexitet är alltför stor; i ett tankeexperiment visas att redan ett mycket enklare, ickelinjärt system kan uppvisa starkt epistemiskt emergenta egenskaper. Reduktionistisk förståelse av medvetandet är därmed inte möjlig. Neurovetenskapen kommer att fortsätta göra stora framsteg; vi kommer med stor säkerhet kunna veta och predicera exempelvis vilka kognitiva centra som är aktiva vid vissa stimuli eller under vissa tankeprocesser. Men emergenta kognitiva fenomen som introspektion eller upplevelserna av lukt och smärta kommer inte att vara uttryckbara i en teori. 'Explanatory gap' kan inte överbryggas. Utifrån en, vetenskapligt sett, mer fruktbar definition av fri vilja än den traditionella visas därefter att den växelverkan som sker i det fysiska, neurala system som inbegriper medvetandet och det undermedvetna inte är deterministisk, trots att de underliggande fysikaliska lagarna är deterministiska. Därmed kan de tre krav vi har formulerat för en fri vilja uppfyllas; att individen kan konstatera att dess agerande äger rum utifrån dess intentioner, att dessa intentioner inte har tvingats fram undermedvetet och att individens beteende som globalt system betraktat är ontologiskt indeterministiskt.

7. FÖRFATTARENS TACK

Ett stort tack går till Keith Elkin för att ha delat sina kunskaper inom neuro-vetenskap med mig, för alla relaterade begrepp och idéer han medvetandegjort mig om och för sina kloka synpunkter under många filosofiska diskussioner. Tack också Erik J. Olsson för insiktsfull och konstruktiv diskussion rörande flera aspekter av arbetet.

LITTERATUR

- Ned Block. 1995. On a confusion about the function of consciousness. *Behavioral and Brain Sciences*. Vol. 18, s. 227-247.
- Rudolf Carnap. 1950. *Logical Foundations of Probability*. University of Chicago Press.
- Gregory J. Chaitin. 1987. *Algorithmic Information Theory*. Cambridge University Press.
- David J. Chalmers. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies*. Vol 2, s. 200–219.
- David J. Chalmers. 2006. Strong and weak emergence. *The Re-Emergence of Emergence*. Oxford University Press.
- David J. Chalmers. 2015. MindPapers. Se <http://consc.net/mindpapers/>
- Paul C. W. Davies. 2004. Emergent biological principles and the computational properties of the universe. *arXiv preprint astro-ph/0408014, 2004 - arxiv.org*
- Stanislas Dehaene. 2014. *Consciousness and the Brain*. Penguin Books, New York.
- Jaegwon Kim. 1999. Making sense of emergence. *Philosophical Studies*. Vol. 95, s. 3-36.
- Jaegwon Kim. 2006. Emergence: Core ideas and issues. *Synthese*. Vol. 151, s. 547-559.
- W. R. Klemm. 2010. Free will debates: Simple experiments are not so simple. *Advances in Cognitive Psychology*, Vol. 6, s. 47-65.
- Joseph Levine. 1983. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, Vol. 64, s. 354-361.

- Benjamin Libet. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*. Vol. 8, s. 529–566.
- Seth Lloyd. 2002. Computational Capacity of the Universe. *Physical Review Letters*. Vol. 88, 237901-1-4.
- John Locke. 1690. *An Essay Concerning Human Understanding*.
- Pier Luigi Luisi. 2002. Emergence in chemistry: chemistry as the embodiment of emergence. *Foundations of Chemistry*, Vol. 4, s. 183-200.
- Colin McGinn. 1989. Can we solve the mind-body problem? *Mind*. Vol. 98, s. 349-366.
- Thomas Nagel. 1974. What is it like to be a bat? *Philosophical Review*. Vol. 83, s. 435-450.
- Jan Scheffel. 2001. *Tankens Villkor*. Nykopia Global Print AB, Solna.
- Jan Scheffel. 2010. Att förstå - betraktelser utifrån en ny teori. *Filosofisk Tidskrift*, Nr 4.
- Michael Silberstein och John McGeever. 1999. The search for ontological emergence. *The Philosophical Quarterly*. Vol. 49, s. 182-200.
- Robert L. West och Christian Lebiere. 2001. Simple games as dynamic, coupled systems: randomness and other emergent properties. *Journal of Cognitive Systems Research*, Vol. 1, s. 221-239.
- Wikipedia 1. 2016. https://en.wikipedia.org/wiki/Artificial_neuron.
- Wikipedia 2. 2016. <https://en.wikipedia.org/wiki/Determinism>

APPENDIX - DEFINITIONER

Definitionerna nedan är inspirerade av Carnaps (1950) analys, nämligen att eftersträva 1) likhet mellan det förvetenskapliga begrepp som ska definieras och den resulterande definitionen, 2) exakthet, 3) fruktbarhet och 4) enkelhet. Speciellt är kravet på 'fruktbarhet' av stor vikt för slutsatserna i detta arbete. Det blir särskilt tydligt avseende 'fri vilja' som ofta definieras på sätt som utesluter vetenskaplig testbarhet, något vi avser att försöka lösa här.

Medvetandet

[A] - Medvetandet är det fenomen som karakteriseras av att kunna uppleva sig självt.

Förtydligande: Det finns ett inslag av cirkularitet i denna definition, men det är oundvikligt på grund av medvetandets unika karaktär. Att relatera medvetandet till enbart från medvetandet skilda begrepp är inte möjligt. Det är av stort intresse att visa var självreferensen ligger. Lockes (1690) tidiga och inflytelserika, karakterisering var "Consciousness is the perception of what passes in a man's own mind". Den av filosofer idag förmodligen mest omhuldade definitionen utgörs av Nagels (1974) "there is something that it is like to be that organism - something it is like for the organism", men även här refererar "is like" naturligtvis till subjektet, det vill säga till medvetandet självt. Den cirkulära karaktären hos definitioner av medvetandet sänder en varningssignal. Vi känner igen problemet från försök att definiera fundamentala fenomen i naturen som exempelvis tid, rum, massa och laddning. Det som kännetecknar dessa fenomen är att de vart och ett är unika; de kan inte förklaras i termer av andra kända fenomen. Om medvetandet är unikt på ett liknande sätt, så kan det alltså inte förklaras fysikaliskt eller reduktionistiskt.

Kommentar: Definitionen av medvetandet ovan är naturlig och torde vara tillräckligt täckande av följande skäl. Jag har, som individ, förmågan att internt kunna reflektera över mina tankeprocesser, att uppleva dem. Tankeprocesserna och reflektionen utgör, tillsammans med andra processer som exempelvis varseblivning, ett fenomen hos mig som jag kan kalla medvetande. De flesta människor har liknande funktioner som jag. Det finns inget som föranleder mig att tro att dessa människor är sämre rustade. Alltså har andra människor också medvetande. För dessa människor gäller därmed också att de har möjlighet att reflektera över sina tankeprocesser. Definitionen blir relevant även för dem.

Vi kan också konstatera att definitionen inte omfattar något annat fenomen som vi inte skulle karakterisera som medvetet. Förmågan till självreflektion är unik för medvetandet.

Emergens

[Be] - *Ett högre-nivå-fenomen är starkt epistemiskt emergent med avseende på fenomen på lägre nivå om dessa senare utgör grunden för högre-nivå-fenomenet och om de teorier som beskriver lägre-nivå-fenomenen inte kan förutsäga högre-nivå-fenomenets beteende.*

Förtydligande: Emergens är ett begrepp som används för att urskilja högre-nivå-fenomen eller beteenden som inte återfinns i de aktuella systemens konstituerande lägre-nivå-delar eller som inte heller på något uppenbart sätt kan härledas från dessa. Svärbeteende hos fiskstim, fågelflockar och myrstackar utgör enkla exempel. Emergens kan uttryckas i ontologisk eller epistemologisk mening. Svag epistemisk emergens definieras här gälla för ett högre-nivå-fenomen om lägre-nivå-fenomenen utgör grunden för högre-nivå-fenomenet och om de teorier som beskriver lägre-nivå-fenomenen inte med nödvändighet förutsäger högre-nivå-fenomenets beteende. Notera formuleringarna "inte kan" avseende stark och "inte med nödvändighet" avseende svag epistemisk emergens. Svagt epistemiskt emergenta system kan alltså simuleras på en dator, vilket inte gäller för starkt emergenta system. Med 'lägre' respektive 'högre' nivå avses systemets delar respektive det sammansatta systemet.

Ontologisk emergens definierar vi på följande vis:

[Bo] - *Ett högre-nivå-fenomen är ontologiskt emergent med avseende på fenomen på lägre nivå om dessa senare utgör grunden för högre-nivå-fenomenet och om dettas beteende inte är reducerbart till de processer som karakteriserar lägre-nivå-fenomenen.*

Förtydligande: Ontologisk emergens är kraftfullare än epistemisk emergens i den meningen att här relateras till *kausala* egenskaper hos systemet snarare än till vår kunskap och våra teorier om dessa fenomen. En distinktion mellan olika former av ontologisk emergens (som exempelvis svag och stark) är därför inte meningsfull.