Abstract

The following Master's thesis presents two articles on, how the philosophical concept of free will can be investigated within experimental neuropsychology.

Within the past few decades the philosophical concept free will has gaining serious attention within the scientific field of neuropsychology. The experimental findings from neuropsychology do however seem to clash with the common understanding of free will, in the sense that the scientific investigation seems to be undermining concept of a free conscious agent. The first article proposes that the apparent clash, between philosophy and scientific neuropsychology, not is a matter of science defying the notion free will, as much as it is an expression of different understandings of what free will entails between the two disciplines. By reviewing how free will often are operationalized in the experimental approach and comparing it to the philosophical definition of free will, it concluded that so far, merely aspects of free will, of the philosophical definition has been investigated. Based on this review a new framework is proposed for understanding free will, which is useful from both a scientific and a philosophical perspective. The aim of this framework is not to answer whether free will exists or not, nor to determine the ontology of conscious will or freedom. As far as possible this is left open for further scientific investigation. The framework proposes that free will is a highly complex phenomenon, depending on dynamics in the hierarchical organization of the brain. Connectionistic, naturalistic and contextualistic approaches are advocated. This breaks from the initial conceptualization of free will as a metaphysical property. Instead it is proposed to leave the metaphysical concept of freedom should in favor of autonomy and that conscious will is a natural occurring phenomenon belonging to the physical world.

The without doubt most influential experiment on free will, is the experiment by Benjamin Libet et al., which showed that the reported times of intention to move occurred 350 ms after the neural activity associated with the movements had begun. This has often been interpreted as being proof of that conscious intention is an illusion. Libet et al. did however only focus on intention in immediate relation to movement. Article II presents an experiment were the subjects a similar task as in the experiment by Libet et al., but instead of acting on the intentions immediately they had to delay performing the intended action. In this manner the intention to move was separated from the actual movement. The study used EEG to measure

the neuroelectrical signal called bereitschafpotential, as well as activity from other areas. The results showed that the signals did differ significantly when subjects were to plan the entire movement of pressing a key in advance, compared to not planning it. The results showed what resembled a “mini-bereitschafpotential” at the time where the prior intention was formed, as well as activity in frontal areas. The results indicate that sensormotoric areas are involved in planning of behavior as well as in executing the movements.

The overall conclusion to this thesis is that simply treating the philosophical concepts regarding free will as ontological truths will lead to erroneous conclusions. Instead both science and philosophy should try to differentiate the concepts as they relate to phenomena emerging from a complex system, therefore not singular entities. The goal of a neuropsychological approach to free will is therefore not to answer if free will exists, but how and what in the brain that makes the individual autonomous emerges.

Dispensation fra studieordning

I forbindelse med følgende specialerapport er følgende dispensation givet af Studienævnet for Psykologi:

Forslaget følger KASO §17, men der dispenseres for linjerne under ”prøve 9”, hvor der står:

> ”*Pensumramme: 2500 sider ikke for snævert afgrænset vejledergodkendt, selvvalgt litteratur, hvoraf højst 750 sider må have været anvendt som pensum ved prøver, som den studerende har gennemført eller indstillet sig til eksamen i.*
> *Sidetal: Mindst 35 og højst 70 sider pr. studerende, dog højst 80 sider ved individuelt udarbejdede specialer.”*

I stedet følges denne ordning:

> Den skriftlige fremstilling skal indeholde en *rammesættende del* og en *artikel-del*. Den rammesættende del skal være mindst 10 normalsider. Artikel-delen skal være mindst 15 normalsider.
>
> Pensumramme: Relevant pensum for den aktuelle problemformulering, fastsat af den studerende og godkendt af vejleder. Nærmere sideantal uspecificeret. Manglende kendskab til relevant pensum vil kunne komme den studerende til last i bedømmelse af opgaven.

Uddybelse af dispensationen (godkendt af studienævnet):

- **I den rammesættende del** præsenteres en problemformulering som skal være specifik og psykologisk væsentlig. Ydermere præsenteres den kontekst artiklen/artiklerne indgår i, herunder hvilket tidsskrift (eller tidsskrifter) den/de er målrettet mod samt relevante begreber og problemstillinger indenfor fagområdet.
- **I artikel-delen** besvares problemformuleringen i form af en eller flere artikler. Det er artikel-delen som skal bære den væsentlige faglige substans i kandidatspecialet. Den studerende vælger i samarbejde med vejleder antallet af artikler og de tidsskrifter, som artiklen/artiklerne rettes mod. Artiklen skal være kvalificeret til det valgte tidsskrift, men det er ikke et krav at den er submitted eller publiceret på afleveringstidspunktet.

  Sidetallet af hver artikel afstemmes til traditionen indenfor tidsskriftet. Artiklen/artiklerne skal stå som selvstændigt/selvstændige afsnit i den skriftlige fremstilling, mens den rammesættende del kan være spredt, fx som introducerende, sammenbindende og konkluderende.

Forord

Denne specialeafhandling følger ikke de i kandidatstudieordningen beskrevne rammer for udarbejdelse af speciale. I stedet er dispensation givet til at aflevere specialeafhandlingen i form af to artikler, bundet sammen af en rammesættende del (se *Dispensation fra studieordning*). Grunden til at denne form vælges, frem for den traditionelle form for specialeafhandling, er et ønske om at fortsætte med at arbejde inden for samme felt og samme problemstilling som specialeafhandlingen omhandler, hvormed evt. publicering af dette materiale vil være en fordel. For det andet ønskedes problemfeltet grebet an fra to forskellige tilgangsvinkler på samme tid, hvilket virkede mere overskueligt ved at dele opgaven op end ved en samlet specialeafhandling. Da emnet for specialeafhandlingen befinder sig i spændingsfeltet mellem videnskab og filosofi, virkede det naturligt at gribe emnet an med henholdsvis en eksperimentel del og en diskuterende/argumenterende del.

Den eksperimentelle del af specialeafhandlingen er en beskrivelse af et forsøg, som blev startet op under undertegnedes praktikforløb på uddannelsens 8. semester, som forskningspraktikant ved *Cognitive Neuroscience Research Unit* (CNRU). Forsøget i denne afhandling er udarbejdet og udført i samarbejde og med hjælp fra med folkene bag CNRU. Jeg vil derfor bringe en særlig tak til Mads Jensen (Ph.D. studerende ved CNRU) for hjælp med design, udførelse, databehandling og generelt sparring omkring projektet, samt en tak til resten af folkene ved CNRU, som har hjulpet på vej. Derudover vil jeg takke Frederik Schmidt og Magnus Vinding for hjælp med rettelser og kommentarer til denne specialeafhandling.

Indholdsfortegnelse

Indledning og ramme for specialet

Har vi *fri vilje*? Dette er utvivlsomt et af de mest omdiskuterede spørgsmål inden for filosofien. Spørgsmålet drejer sig om, hvorvidt de handlinger og beslutninger som et individ fortager sig, kan betragtes som et udtryk for individets egen vilje, eller i princippet er bestemt af faktorer uden for individets rækkevidde. Dette spørgsmål kan spores tilbage til de allertidligste filosoffer og er fortsat et emne som flittigt debatteres inden for alle filosofiens grene og visse videnskaber såvel (Kane, 2002).

Spørgsmålet om fri vilje blev særligt ekspliciteret i løbet af middelalderens teologiske filosofi. Her var det centrale spørgsmål, hvordan mennesker kunne handle som følge af deres egen frie vilje, hvis der eksisterede en alvidende gud. Da en alvidende gud per definition ved alt, så vil denne vide, hvordan fremtidige events vil komme til at foregå. Hvis denne gud i år 0 ved, hvad der vil ske i år 2011, er der intet, som mellemtiden vil få tingene til at udfolde sig på anden vis. Hvis dette var tilfældet, ville den alvidende gud ikke være alvidende. Menneskets handlinger og vilje er derfor logisk set uden betydning, da intet vil kunne ændres. Den position at menneskelige handlinger er uden kausal betydning kaldes *fatalisme* og må betragtes som antitesen til fri vilje (Bernstein, 2002). Selvom diskussionen om fri vilje har bevæget sig væk fra teologisk filosofi, har diskussionen om fri vilje ikke ændret sig betydeligt. Selvom den alvidende gud blev forkastet til fordel for et naturvidenskabeligt verdensbillede, hvor naturens orden ikke er bestemt af en gud, men er bestemt af fysikkens naturlove. Fatalisme stadig være resultatet, da den fysiske verden betragtes som kausalt lukket (Bernstein, 2002). Den moderne filosofis tilgang til spørgsmålet om fri vilje er derfor, om universet er fuldt ud determineret, eller om fri vilje kan eksistere. I givet fald hvordan er dette så foreneligt med et lovmæssigt styret fysisk univers?

Utallige forklaringer og grupperinger af positioner inden for filosofien har forsøgt at svare på dette spørgsmål. Uden at gå i dybden med alle disse positioner og argumenter for og imod, er det muligt at udlede tre overordnede hovedretninger. Den første hovedretning er *libertarianisme*, som går ud på, at universet ikke er kausalt lukket og udeterminerede handlinger derfor vil kunne finde sted, hvilket er en nødvendighed for fri vilje (Kane, 1999). Den anden hovedretning er *determinisme*, som i modsætning til libertarianisme ser universet som kausalt lukket, hvilket vil betyde at alt er bestemt af kausale sammenhænge og frie handlinger ikke kan eksistere. Den sidste hovedretning er *kompatibilisme*, som forsøger at sammenflette de to forrige hovedretninger, ved at påstå at fri vilje ikke er uforeneligt med et fult ud determineret univers styret af kausale love (Kane, 2002).

Vigtigheden af spørgsmålet om fri vilje understreges især af moralfilosofien, hvor spørgsmålet om handlinger er frie eller ej, har stor betydning for hvorvidt mennesker kan stå til moralsk og ikke mindst juridisk ansvar for deres handlinger. For det giver mening at stille et individ moralsk til ansvar for en handling, kræver det at individet kunne have handlet anderledes (Frankfurt, 1969). Hvis alle handlinger i princippet kan tilskrives naturens iboende orden, hvordan kan det da være muligt at stille individer til ansvar for deres handlinger? Undersøgelser har vist er den almene opfattelse af fri vilje er, at selvom universet er deterministisk og følger lovmæssigheder, opfatter folk stadig individet som årsag til handlinger og ansvarlig for disse handlinger. Kompatibilisme lader derfor til at være den almene opfattelse i forhold til fri vilje (Nichols, 2006; Nichols & Knobe, 2007). Dette gælder ikke kun kulturer, der normalt betegnes som individualistiske, men er vist at være gældende på tværs af kulturer, uafhængigt af de akademiske filosofiske diskussioner herom (Sarkissian et al., 2010).

Spørgsmålet om fri vilje er imidlertid ikke kun begrænset til filosofien. Inden for neurovidenskaberne (neuropsykologi, kognitiv-neurovidenskab, neurologi o. lign.) er der en stigende interesse for dette spørgsmål. Dette skyldes uden tvivl at neurovidenskaberne beskæftiger sig med hjernen, hvor det må antages at den subjektive vilje opstår. Samtidig er hjernen et biologisk system, hvilket betyder, at den følger naturens fysiske love. Dette paradoks fører derfor direkte tilbage til den filosofiske diskussion. Mange moderne filosoffer diskuterer da også de filosofiske implikationer af resultaterne fra neurovidenskaberne og hvordan subjektiv vilje skal forstås i forhold til processer i hjernen (f.eks. Dennett, 2003; Kane, 1996; Kim, 2005; Mele, 2009).

Det neurovidenskabelige felts interesse er imidlertid forholdsvis nyt i forhold til den filosofiske diskussions historie. Dette formegentlig behaviorismens afvisning af mentalitet, herunder subjektiv vilje, som genstand for videnskabelig undersøgelse, som ifølge behaviorismen burde fokusere på sammenhænge mellem målelig stimulus og respons (Bargh & Ferguson, 2000). Den utvivlsomt mest debatterede eksperimentelle undersøgelse af fri vilje er de kontroversielle studier udført af Libet et al. (Libet, 1985, 1999; Libet, Gleason, Wright, & Pearl, 1983; Libet, Wright, & Gleason, 1982, 1983). Disse studier var formegentlig det som i starten af 1980'erne startede den neurovidenskabelige interesse for fri vilje. Libet et al. viste at et neuroelektrisk signal, som det tidligere var vist, går forud for motoriske bevægelser, det såkaldte *bereitschaftpotential* (Kornhuber & Deecke, 1965), begyndte omkring 350 ms før forsøgsdeltagerne blev bevidste om at ville lave en bevægelse. Hvis hjernen er i gang med at udføre en handling, før den subjektive vilje om at lave denne handling opstår, kan den subjektive vilje ikke tilskrives som kausal årsag til handlingen. På baggrund heraf er det blevet påstået at når en handling udføres foregår der to processer i hjernen. Den første af den rent neuro-motoriske udførelse af handlingen og den anden er om, at den subjektive oplevelse af vilje til at lave handlingen. For subjektet vil det se ud som den subjektive vilje var årsagen til handlingen, hvilket imidlertid kun er en *illusion* som hjernen skaber (Wegner, 2003a, 2003b).

Har neurovidenskaberne hermed endelig modbevist fri vilje? Som følge af neurovidenskabernes fortsatte udforskning af mekanismerne bag, det viljestyrede subjekt, følger spørgsmålet om, hvad dette vil betyde for det spørgsmål om moralitet som fulgte debatten om fri vilje. Hvis al adfærd er determineret af hjernens biologi vil subjektet da kunne handle anderledes? Det er blevet foreslået, at siden resultaterne fra neurovidenskaberne strider imod den almene forståelse af individuel moral og juridisk ansvar og dermed potentielt underminerende for moralske værdier og juridiske love, er det nødvendigt at se bort fra resultaterne fra neruovidenskaberne og fortsætte med at følge de almene opfattelser af moral (Kaposy, 2010).

Formål med specialet

Ideerne til dette projekt udspringer, bland andet af denne åbenlyse konflikt mellem neurovidenskabernes resultater og den moralfilosofiske forståelse af fri vilje, især fordi konklusionen om at se bort fra neurovidenskabernes resultater. Dette går imod det idealt videnskaben står for, om at udforske og forstå den menneskelige natur. Ignorance over for og negligering af et videnskabeligt felt syntes på ingen måde at være en holdbar løsning på længere sigt. Dette gav anledning til en undren om, hvorvidt det virkelig er en nødvendighed at droppe neurovidenskaberne for at bevare individets autonomi og moralitet. Fri vilje er, som nævnt, et omdiskuteret begreb og der er mange definitioner af hvad fri vilje er, samt hvilke kriterier der afgør om en handling er fri og viljestyret. Inden for neurovidenskaberne er fri vilje blevet operationaliseret som både den subjektive fornemmelse af ejerskab over handlinger, evnen til beslutningstagen, intentioner og det at udføre selvbestemte handlinger (Roskies, 2010). Ydermere er det erkendelsesmæssige grundlag for henholdsvis neurovidenskaberne og filosofien forskellig, da neurovidenskabernes erkendelsesmæssige grundlag er eksperimentel induktion, hvorimod filosofien oftest er logisk deduktion. Det er dermed ikke nødvendigvis det samme grundlæggende begreb som de to discipliner behandler, når der inden for disse tilgange nævnes fri vilje, eller om filosofien og neurovidenskaberne taler forbi hinanden. Hvis dette er tilfældet, vil der opstå et problem, når resultaterne fra neurovidenskaberne tolkes inden for filosofien.

Formålet med denne specialeafhandling er, med udgangspunkt ovenstående kritik at undersøge, hvordan fri vilje kan undersøges ud fra et neurovidenskabeligt perspektiv. Dette gøres ud fra to tilgange. Den første tilgang er en teoretisk analyse af, hvad begrebet fri vilje indeholder og hvordan dette kan forstås ud fra de seneste års resultater inden for neurovidenskaberne. Den anden tilgang er, at undersøge fri vilje eksperimentelt med henblik på at yderligere udvide den konceptualisering af fri vilje, som bruges i Libet's eksperimentelle paradigme (Libet et al., 1982). Formålet med afhandlingen er ikke at svare på, om der er fri vilje, eller eksperimentelt vise/afvise dette, men at illustrere behovet for en yderligere konceptualisering af fri vilje inden for neurovidenskaberne, samt redegøre for hvordan dette skal forstås i en konceptuel-filosofisk diskussion af neurovidenskabernes resultater.

Tilgang til problemfelt

Dette problemfelt vil blive behandlet i to artikler, med hvert sit fokus. Den første artikel (*Understanding and Investigating Free Will: Translating a Concept between Philosophy and Neuropsychology*) er en teoretisk redegørelse og diskussion af en konceptuel forståelse af fri vilje på baggrund af neurovidenskabernes resultater, med det formål at skabe en fælles konceptualisering der er anvendelige både videnskabeligt og filosofisk. Den anden artikel (*Neural correlates of Free Will separated from Action*) er en præsentation af et neuropsykologisk eksperiment som bygger på Libet's eksperiment, men yderligere specificere begrebet intention i forhold til det oprindelige eksperiment og metode.

Specialeemnet befinder sig i sammenspillet mellem, mellem filosofi og neurovidenskab. Dette gav en udfordring i forhold til udvælgelsen af relevant litteratur og brugen heraf i artiklerne. For det første er diskussionen om fri vilje som tidligere nævnt omtrent ligeså gammel som filosofien i sig selv. Hvad der i forbindelse med dette projekt har været krav i udvælgelsen af litteratur inden for det filosofiske felt, er at det gerne skulle være nyere moderne filosofisk litteratur med fokus på interaktionen mellem psyke og krop, gerne specifikt i forhold til neurovidenskab. Hermed er visse grene af filosofien, som ellers traditionelt set beskæftiger sig med fri vilje blevet sorteret fra, så som eksistentielle og humanistiske retninger, der befinder sig inden for den kontinentale filosofi. I stedet er der fokus på den analytiske tradition, af to grunde: For det første er det som regel inden for denne filosofiske retning, at sammenhængen mellem krop og psyke, samt det metafysiske begrebet om frihed diskuteres. Meget relevant litteratur befinder sig derfor inden for denne tradition. For det andet er den analytiske traditions fokus på logisk konceptualisering og metafysik ideel i forhold til at diskutere konceptualisering af begreber over for operationaliseringer til brug i eksperimentel videnskab (Martinich & Sosa, 2006). Den eksperimentelle litteratur er udvalgt inden for eksperimentiel og teoretisk neurovidenskab. Fokus har været på studier, der specifikt undersøger fri vilje og hertil beslægtede begreber, så som fornemmelse af agens (eng: sense of agency), intentioner beslutningstagen og udførelse af selvbestemte handlinger, samt artikler der opsummer og tolker på sådanne resultater. Neurovidenskab er et bredt defineret felt, der beskæftiger sig med hjernen, fra lav-niveaus komponenter, så som enkelt neuroner og neurotransmittere, til højere-ordens funktionalitiet og adfærd. Det er hovedsageligt inden for den psykologiske del af neurovidenskaben, med fokus på kognitive og fænomenologiske aspekter i relation til neurale mekanismer og processer, at undersøgelse og diskussion af fri vilje finder sted. Enkelte kliniske studier er inkluderet i dette, men fokus er på normalområdet. Metoderne inden for neurovidenskaben, især de krav som stilles til statistisk analyse af neurofysiologiske data, er ydermere en videnskabelig gren i sig selv, som spænder mellem statistik og ingeniørkunst. Selvom sådanne metoder anvendes i den eksperimentelle tilgang, er fokus i dette projekt imidlertid ikke på det tekniske eller teoretiske bag metoderne, men på de tolkningsmæssige erkendelser som kan uddrages af resultaterne.

Kort præsentation of artikel I

Den første artikel er en teoretisk artikel, hvis formål er at give et bud på, hvordan fri vilje kan undersøges inden for eksperimentel neuropsykologi. Artiklen tager udgangspunkt i det tidligere nævnte spørgsmål om, hvorvidt de konceptualiseringer af fri vilje, der henholdsvist bruges inden for filosofien og neuropsykologien rent faktisk kan betegnes som det samme begreb. I artikel I påpeges nogle af de misforståelser og fejlagtige slutninger ud fra en analyse af de forskellige forståelser af fri vilje. Hovedformålet med artikel I er at fremstille en ny konceptualisering af fri vilje som både kan bruges redskab inden for neuropsykologien og konceptuelt forenelig med den bredere filosofiske tradition.

Artiklen starter ud, efter en kort introduktion, med en indledende beskrivelse af hvad den gængse forståelse af fri vilje indebærer rent konceptuelt. Herefter følger en kritisk gennemgang af resultater fra en række betydningsfulde neuropsykologiske eksperimenter og fund heraf. Vægten på denne gennemgang er at afklare, hvordan disse eksperimenter forstår fri vilje og dermed hvad det rent faktisk kan bruges til at sige om fri vilje. Dette leder over i en hypotesedannelse omkring den metafysiske baggrund for fri vilje, for dermed at forankre neuropsykologiens resultater i en filosofisk ramme. Ønsket er imidlertid ikke at give en a posteriori forklaring på fri vilje, men en a priori ramme for fremtidig undersøgelse. Konceptualiseringen lader derfor den metafysiske natur bag flere af de centrale begreber relateret til fri vilje stå åben, men giver forslag til i hvilken kontekst de skal forstås. Den foreslåede tilgang til undersøgelse af fri vilje kan i hovedtræk opsummeres som tre centrale metateoretiske tilgange. Det første metateoretiske perspektiv er *konnektionisme*. Dette går imod ideen om en samlet bevidst agent, til fordel for en agent der opstår på baggrund af distribuerede moduler i hjernen. Det andet metateoretiske perspektiv er *kontekstualisme*, som bryder med ideen om kausal uafhængighed som en nødvendighed for frihed. I stedet lægges der op til en klarer konceptuel skelnen mellem metafysisk frihed (som det for eksempel kommer til udtryk i kvantemekanik) og individuel autonomi. Det tredje metateoretiske perspektiv er *naturalisme*, hvor den bevidste vilje ses som et biologisk funderet fænomen. I forbindelse hermed bliver det pointeret at en tilgang der antager, at fri vilje kræver en immateriel agent (for eksempel i den kartesianske tradition), har forklaringsproblemer rent konceptuelt i forhold til den immaterielle agents kausale indvirkning, samt forklaring af neuropsykologiens resultater.

Artikel I kan betegnes som en teoretisk artikel eller en review-artikel, alt efter hvordan de forskellige tidskrifters forlag definere dette. Emnet for artiklen taget i betragtning, er denne artikel egnet til tidsskrifter som beskæftiger sig med grænsefalden mellem psykologi og filosofi, så som *Journal of Philosophical Psychology,* eller mere specifik i forhold til tidsskrifter dedikeret til videnskabelig tilgang til bevidsthed (i dette tilfælde vilje som bevidsthedsfænomen) så som *Consciousness and Cognintion* eller *Journal of Consciousness*

Studies. Da de tiltænkte tidsskrifter alle er engelsksprogede er artiklen på engelsk. Artiklen er endnu ikke indsendt.

Kort præsentation af artikel II

Artikel II præsenterer et eksperiment baseret på eksperiment af Libet et al. (Libet et al., 1982). I eksperimentet observerer forsøgsdeltagerne en modificeret urskive med en omdrejningstid på 2.56 s. På et vilkårligt laver de så en bevægelse med hånden og angiver herefter, ud fra urskiven, tiden hvor de blev bevidste om deres intention om at lave bevægelsen. Uden at gå yderligere ind i den tekniske beskrivelse af eksperimentet på nuværende sted i specialeafhandlingen, er formålet med det aktuelle eksperiment er at kigge på de neurale korrelater for intention separeret fra selve det, at lave den intenderede bevægelse. Det aktuelle eksperiment bruger samme opstilling, med den forskel i instruktionerne at deltagerne i stedet for at bevægelse hånden med det samme, når intentionen opstod, skulle vente til uret havde roteret en hel omgang med at bevæge hånden. På den måde blev intentionen i princippet separeret fra selve bevægelsen. Samtidig måles hjerneaktiviteten med elektroencefalografi (EEG). Resultaterne viste, at hjerneaktivitet i forbindelse med intentionen adskilt fra bevægelsen mindede om det motoriske signal bereitschaftpotential. Dette tolkes som et udtryk for, at det sensormotoriske cortex er involveret i intentionel planlægning af bevægelse, så vel som i selve udførelsen af bevægelsen. Formålet for artikel II skal ses i forlængelse af artikel I. Eksperimentet bruger en udvidelse af intentionsbegrebet, i forhold til måden intention operationaliseres på i Libet's eksperiment, for at undersøge forskellen i neural aktivitet når sådan operationalisering bruges. Eksperimentets mål er ikke at bevise de metateoretiske perspektiver, som foreslås i artikel I, men skal ses som bidragende til den konceptuelle forståelse og nuancering af især intentionsbegrebet i eksperimentel neuropsykologi og hvordan dette relatere til hjerneaktivitet.

Artiklen er opbygget som en typisk forskningsrapport, hvor metode og resultater nøjsomt rapporteres efterfulgt af en kort diskussion af resultaternes betydning. Denne artikel kunne i princippet bringes i mange forskellige tidsskrifter, både rent eksperimentelle neuropsykologiske tidsskrifter og mere specifikke tidskrifter relateret til eksperimentiel bevidsthedsforskning (f.eks. *Consciousness and Cognintion*). Eksperimenter, der ligesom dette studie bygger på Libet's metode, har været bragt så forskellige steder som *Nature Neuroscience, Experimental Brain Research* og *Electroencephalography and Clinical*

Neurophsiology, for bare at nævne et par stykker. Opstillingen af aktuelle artikel, i forhold til afrapportering af metode og resultater bygger på en skabelon og retningslinjer fra *European Journal of Neuroscience,* dog uden nogen yderligere begrundelse herfor. Da artikler af denne type typisk findes i internationale tidsskrifter, er artiklen på engelsk.

Artiklen er ikke indsendt og hvorvidt den overhoved vil blive er tvivlsomt. Der er visse aspekter af selve eksperimentet som trænger forbedring. For det første er antallet af forsøgsdeltagere for lavt (kun 9 deltagere) og aldersspredningen (33.33 år, SD = 6.48) stor hvilket uden tvivl har en indflydelse på resultaterne. Der var flere grunde til at der ikke blev anvendt flere forsøgsdeltagere på daværende tidspunkt. For det første opstod der i løbet af den uge forsøget fandt sted et teknisk problem omkring den software, som blev brugt i forsøget. Dette betød bland andet at en eksperimentel serie mangler for et enkelt subjekt. For det andet er der for mange variabler i eksperimentet som det blev udført. Der blev for eksempel spurgt ind til den oplevede grad af impulsivitet i udførelsen af bevægelsen. I den indledende analyse af data, viste dette dog ingen effekt (og behandles derfor heller ikke yderligere i artiklen). Denne del af eksperimentet er derfor udelukkende en confounder i forhold til forsøgets egentlige mål. Hvis forsøget gentages vil dette aspekt blive droppet, eller blive undersøgt i et separat eksperiment.

Noter om analyse af EEG-data:

Som nævnt er fokus med dette speciale ikke metodisk. De neurofysiologiske metoder brugt i artikel II, er dog forholdsvist komplekse, så en kort beskrivelse af metoderne, er på sin plads for at vise forståelsen af disse, da det ikke uddybes i selve artiklen.

Eksperimentet bruger EEG som metode til at måle hjerneaktivitet. EEG består af elektroder som monteres på skalpen, hvor de indirekte måler det neurofysiologiske spændingsfald som følge af synkron fyring af neuroner i det underliggende cortex. Specifikt bruges EEG i det aktuelle eksperiment til at se på såkaldte *event relateret potentialer* (ERP). Et ERP er et specifik signal som fremkommer under udførelse af en speciel opgave eller præsentation af stimuli. ERP måles inden for et begrænset tidsmæssigt vindue på få sekunder (Luck, 2005). Det elektrofysiologiske signal beritschaftpotential, som måles i Libet's forsøg, er et eksempel på et ERP (Shibasaki & Hallett, 2006). Ofte er det ført muligt at observere et ERP signal ved at tage gennemsnit af målt aktivitet over flere gentagelser af den gældende handling da signal-til-støj, forholdet som regel er lavt i EEG, da at elektroderne opfanger irrelevant hjerneaktivitet og muskelaktivitet, øjenblink og generel støj fra omgivelserne, så vel som det signal der ønskes undersøgt (Luck, 2005).

I det aktuelle eksperiment bruges *independent component analysis* (ICA), til at rense EEG-data for forskellige former for støj. ICA er et statistisk værktøj, som opdeler det samlede datasset i et antal statistisk uafhængige komponenter, baseret på størst mulig uafhængighed mellem komponenterne (Makeig, Bell, Jung, & Sejnowski, 1996). ICA antager at værdierne målt fra hver elektrode består af et antal uafhængige generatorer i hjernen, som er blandet sammen, når signalerne måles med elektroden på overfladen af huden. ICA-algoritmen[1] er is stand til, på baggrund af det samlede dataset for hver forsøgsdeltager, at udregne et antal uafhængige komponenter (*independent components* - IC). Der kan dog ikke udregne flere IC end antallet af kilder, som IC bliver udregnet ud fra, dvs. antallet af EEG-elektroder. Da elektrodernes placering er samme for hver forsøgsperson, kan den spatielle såvel som temporale udbredelse af IC udregnes (Groppe, Makeig, & Kutas, 2008; Makeig et al., 1996). I dette eksperiment bruges ICA ikke til at finde IC, som udtryk for specifik aktivitet, men kun til at fjerne øjenblink og anden støj, til hvilket ICA har vist sig at være en særdeles robust metode (Jung et al., 2000). Dette gøres ved, at de IC der markeres som støj fjernes, hvorefter den aktivitet som disse IC indholdet statistisk fjernes fra det samlede dataset (Jung et al., 2000).

Den statistiske analyse i forhold til forsøgets formål er udført med *statistical parametric mapping* (SPM). Denne metode er oprindelig udvilket til analyse af positron emission tomografi (PET), men er videreudviklet til også at kunne bruges til EEG (Litvak et al., 2011). Ved denne metode omdannes det samlede EEG-dataset for hver forsøgsdeltager til statistiske ”billeder” af aktiviteten fordelt over et todimensionelt kort af skalpen. Til dette tilføjes den tidsmæssige dimension, således at al data i den epoke der undersøges, bliver til tredimensionelle spatial-temporale kort over aktivitet. På disse kort kan da laves statistiske test af forskellig karakter alt efter den gældende datatype (Litvak et al., 2011). Fordelen ved at bruge SPM til analyse frem for analyse af kun gennemsnit over ERP er den rumlige udbredelse medregnes, hvilket betyder at det er muligt at give et mere kvalificeret billede af hvor præcis aktiviteten er lokaliseret. Da eksperimentet dog kun bruger 32 elektroders EEG er den spatialle lokation dog kun anvisende.

[1] Flere forskellige ICA-alogritmer findes. I det nærværende eksperiment er den i EEGLAB indbyggede algoritme *runica* anvendt (Delorme & Makeig, 2004)

Da billederne i SPM består af mange tusinde datapunkter, som sammenlignes, vil der statistisk set fremkomme et antal falske positive forskelle. Den kritiske p-værdi burde derfor korrigeres med en faktor der svarer hertil. I det nærværende eksperiment er det dog ikke tilfældet. I stedet er den kritiske p-værdi sat til $p = 0.001$. Dette lader desuden til at være den gængse praksis ved dataanalyse med denne metode.

ARTIKEL I:
Understanding and Investigating Free Will: Translating a Concept between Philosophy and Neuropsychology

Abstract

Free will is to gaining serious attention within the neuropsychology. The experimental findings from neuropsychology do however seem to clash with the common understanding of free will, in the sense that the scientific investigation seems to be undermining concept of a free conscious agent. The present article proposes that the apparent clash, between philosophy and scientific neuropsychology, not is a matter of science defying the notion free will, as much as it is an expression of different understandings of what free will entails between the two disciplines. By reviewing how free will often are operationalized in the experimental approach and comparing it to the philosophical definition of free will, it concluded that so far, merely aspects of free will, of the philosophical definition has been investigated. Based on this review a new framework is proposed for understanding free will, which is useful from both a scientific and a philosophical perspective. The aim of this framework is not to answer whether free will exists or not, nor to determine the ontology of conscious will or freedom. As far as possible this is left open for further scientific investigation. The framework proposes that free will is a highly complex phenomenon, depending on dynamics in the hierarchical organization of the brain. Connectionistic, naturalistic and contextualistic approaches are advocated. This breaks from the initial conceptualization of free will as a metaphysical property. Instead it is proposed to leave the metaphysical concept of freedom should in favor of autonomy and that conscious will is a natural occurring phenomenon belonging to the physical world.

1 Introduction

One of the oldest questions in science and philosophy is whether we have free will. Are we agents of our own actions, who can control our future or is everything determined? The concept of free will is found within a diverse range of fields from philosophy to studies in artificial intelligence. The question of free will has recently gained serious attention within the experimental psychological sciences, especially the neuro-cognitive branches of psychology. The experimental approach to free will from the neurosciences is fairly new, compared to the philosophical discussion of free will and mental causality, and does not go further back than Benjamin Libet's experiments in the 1980's (Libet, 1985).

Free will has been defined in numerous ways by scholars from a range of disciplines. In addition it seems like a concept that all people got some everyday conceptualization of, independent of the broader philosophical background and debate (Monroe & Malle, 2010).The belief in free will, and the belief that one is morally responsible for his or her own action, have been shown to be a universal belief shared cross-culturally (Sarkissian et al., 2010). The problem with scientifically investigating free will is that free will is a wildly used term with ambiguous meanings depending on which field it is defined and used within. It is relevant to ask, whether the various understandings of what *will* is and what is meant by *free,* is the same across the various academic fields. Is it even the same concept within each field, for example within psychology? The ambiguity of the concept is however a huge problem in putting forward meaningful investigations hereof, and how can an empirical science help the philosophical debate of the metaphysical nature of free will at all? In order to make sure that results from neuroscience provides useful information for the understanding of free will and is to be taken serious, a clarification of what is meant when claiming to study free will is necessary. Furthermore it is necessary to understand which questions regarding free will that can be answered experimentally and which questions belongs to abstract philosophical reasoning.

The need for a clarification is not just a scientific need. The implications of the findings from neuroscience on the question of moral and legal responsibly are currently widely debated (e.g. Anckarsater, 2010; Cashmore, 2010a; Sie & Wouters, 2010). It has been argued that since the findings from the neurosciences clashes with the common held view on free will and moral responsibility these findings should be ignored altogether in order for moral responsibility to prevail (Kaposy, 2010). First this claim is build on the assumption that when the cause of action can be traced to neural activity then no moral agent can be present. Secondly it seems like it is assumed that neuroscience have given an exhaustive account of the neural mechanisms involved in moral irresponsible behaviour, which definitely is not the case (Urbaniok et al., 2011).

The current debate on the moral implications of the scientific findings on ethics illustrates the need for a conceptual clarification on how free will should be understood, and how it can be investigated from a neuroscientific perspective. It has been proposed that free will is a purely philosophical construct and cannot be investigated scientifically (Hinsen, 2010). Many conclusions regarding free will from neuroscience is in fact not using the same conceptualization of free will as the philosophers of ethics, and the conclusions from neuroscience tends to take the concept of free will across the boundaries of what the experiments really is showing (Klemm, 2010; Montague, 2008). There is a need for a "*translation*" of the philosophical concept of free will into something that can be investigated scientifically, but in such a way that the result from the scientific investigation can be 'back-translated' into the philosophical theoretical debate. In order to reach this goal it is necessary to consider both conceptual, metaphysical and scientific problems and challenges attached to the concept of free will, in order to give an account of how the concept can be translated from philosophy to science and back again. The framework advocated here for investigating free will is aimed at making as few a priori assumptions regarding the metaphysical nature of consciousness as possible, as it has been proposed that such can only be settled a posteriori (Dowell, 2006). The purpose of the current article is not to answer whether we have free will or not, but to determine the way in which the concept of free will should be used and how it should be investigated in neuropsychology[1]. The aim is to focus on clarifying which aspects of free will that can be empirically investigated, and which that belongs to deductive philosophical reasoning.

[1] The term *neuropsychology* is here used as an umbrella-term to cover a wide range of scientific fields spanning across cognitive sciences, cognitive neuropsychology to neurology, since the problems and framework outlined in the present article is of mutual relevance for all fields. The main focus is the scientific approach contra philosophical deduction.

2 When is an action free and what is will?

From a philosophical perspective the idea of free will is related to performing actions. Action is a category among physical events which is characterized by being performed by an agent. Action can be expressed entirely in its physical properties, and can be compared to behaviour, as it is expressed by any possible organism (biological or artificial alike). The physical nature of actions means that all kinds of action can be reduced to physical properties such as time-space location, kinetics etc (Davidson, 1980). Davidson illustrates this by the example of a queen who moves her hand thereby signaling the execution of the king. The moving of the hand was the cause of the killing of the king, and therefore we can say that the queen caused the kings death, since the moving of the hand "*was identical with her doing something that caused the death of the king*" (Davidson, 1980, s. 58). This means that description of behavior can be reduced to an account of the causal chain of events following the kinetic actions by an agent.

In order for an action to be an expression of free will it has to be a *voluntary* action. An action is voluntary if it is caused by, and only by, an act of will. The causal relation among will and voluntary actions can be expressed as $W \rightarrow A$. If an action is an effect of any other cause S, so that $S \rightarrow A$, then it is not an voluntary action (Kane, 1996). An *act of will* is something an agent does. This is here used synonymous with *volition*. Voluntary actions can be distinguished from all other actions, by being caused by an agent's volition[2].

This initial definition views volition as a phenomenological property. Phenomenological properties are defined as; the "*what it is like to be"* for an organism to be that organism, and can per definition only be accessible from a first-person perspective (Nagel, 1974). Phenomenological properties, subjective experience and consciousness are here used somewhat interchangeable to refer to events only accessible from a first-person perspective. When considering the phenomenology of voluntary actions, it is the subjective experience of being initiator and executer of the actions that one performs. For volitions to be the cause of actions they need to have a causal effect in the physical world.

The problem of volition is whether it has a causal impact on the physical world, known as the *mind-body problem*. The physical world, to which actions belong, is generally viewed as causally closed. This means that for every physical event that has a cause, this cause has to be

[2] The terms will and volition is here uses synonymous.

physical as well (Kim, 2005). A central problem related to free will as well is therefore the question of the metaphysical nature of consciousness, and how it is related to neural processes, known as the *hard problem of consciousness* (Chalmers, 1995).

What exactly is required in the subjective experience in order to qualify it as volition is not clear (Roskies, 2010), whether it is a sense of ownership, decision-making, forming of intention or a combination of these. What all these concepts have in common, is that they are mental phenomena. These differences are actually a big deal when considering a scientific approach to free will, since it is not a priori given that these terms covers one and same ontological process. Rather there seems to be a conceptual difference between concepts as intention and deliberate decision-making and subjective sense of being an agent. An un-reflected use of these concepts could lead to a conceptual misunderstanding, if all these phenomena are investigated as being one type-identity.

The mind-body problem is however just one of the philosophical problems of free will. Shariff, Schooler and Vohn (2008) have outlined two central problems of free will. The first of a testable character regarding neural activity involved in voluntary actions, including which functions of the brain that is involved in volition and decision-making. The second problem is related to the metaphysical concept of freedom and does not necessarily include the phenomenological aspects of volition. This problem stems from the analytical tradition and debate related to the causal nature of the universe, between libertarianism, determinism or compatibalism; expressed as the question of whether the agent is *free* when performing an action or predetermined to do so. What is meant by determinism is that "*a universe is* deterministic *if there are transition rules (the laws of physics) that* determine exactly *which state description follows any particular state description*" (Dennett, 2003, p. 28 original emphasis). Freedom as a metaphysical concept means that an action or event can be undetermined by any prior cause (Kane, 1996). On the other hand, the agent could be entirely determined by causes in its environment, and all that the agent does, is to facilitate the outcome.

Determinism is often used synonymously with *fatalism*; the position that every event in the universe in principle is given from starts to infinity. If one then gained cosmic all-knowing insight (as Laplace's demon) then one will be able to tell how the universe falls out to the end of time. If this is a metaphysical possibility, then no truly free events exist, since all in principle can be calculated in advance. Fatalism is therefore considered the antithesis to freedom, since there would be no way in which the agent could have an impact on future events (Bernstein, 2002). In neuropsychology the problem of freedom can be reformulated to if and how freedom can it arises from the brain, assumed to be a deterministic system? Compatibalists holds that expressing *autonomy*, understood as the ability for any organism to follow intentions originating from the organism itself, is enough to make the agent architect of its actions (Frankfurt, 1971), whereas the libertarian position holds, that if any prior determinants can predict action then it is not free (Kane, 1999). This can be exemplified by the traditional Kantian concept of an agent who unaffected by context makes rational choices that guide action. The problem of freedom is in principle independent of the mind-body problem and metaphysical nature of subjective volition, but both freedom and mental causation is necessary for free will to exist. This initial definition of free will requires a conscious agent who *causes* action as a result of *own* intention, undetermined by any other causes in doing so. Volition can therefore also be reframed as the subject's *intention* to do an action. In this sense, intention is crucial for free will. Intention can be defined as an irreducible conscious event directed toward performing a certain action (Zhu, 2004).

3 Conceptualizing intention

The broad definition of intention encapsulates both the rational planning of complex actions and as the volition in accordance with the execution of simple motoric actions (Mele, 2009). With such inclusive definition it is possible that when investigating the role of intention, scientists would actually be studying different ontological phenomena. Mele (2010) divides intention into *proximal intention* and *distal intention*. Proximal intention concerns the immediate execution of action straightaway, whereas distal intention is concerned action which is executed later, comparable to planning an action. This qualitative and quantitative difference stresses the importance of conceptual clarification when making conclusions regarding the role of conscious intention.

In the model proposed by Pacherie (2008) different kinds of intention are distinguished by their functional roles. The distal intention is not only including specific information regarding an action to be performed later, but includes rational reasoning and abstract planning as well as the subject's beliefs and desires. In contrast the proximal intention is bound to the immediate context. The model proposes a dynamic hierarchical implementation of intentions, where the template of the distal intentions guides the proximal intentions in the context, which is converted into specific motoric intentions or commands by the sensor-motoric system, which mainly operates unconsciously (Pacherie, 2008). The dynamic view of intention is an argument against viewing intention as a singular entity. Conceptualizing intention as an irreducible entity a priori, would be wrong, given intentions could be many ontological phenomena.

The without doubt most debated and best known experimental approach to free will is the study by Libet (1985) on the timing of conscious intentions. In the experiment the subjects had to move their hand at a random time at their own will without any prior planning. Afterwards they reported the time they felt the *urge* to move their hands, using a fast rotating clock they watched during the task. The subjective times (labeled *W-time*) were then compared to the bereitchaftpotential (BP) an event-related negative potential, measured with electroencephalography (EEG), which occurs up to 1.5 s before a motoric movement (Kornhuber & Deecke, 1965). Libet found that the subjective intention to perform an action occurred approximately 350 ms *after* the onset of BP, and about 150 ms before actual movement. The conclusion was that unconscious neural processes are the causal initiator of actions, and conscious will then occurs afterwards. Still the conscious will could override the unconsciously initiated action, as the 150 ms from W to movement were enough for the subject to “veto” the initiated act, though Libet considered that this veto itself could be a result of unconscious processes as well (Libet, 1985). This experiment has been replicated several times, using both functional magnetic resonance imaging (fMRI) (Lau, Rogers, Haggard, & Passingham, 2004), and single-neurons recordings (Fried, Mukamel, & Kreiman, 2011), with results consistent with the findings by Libet regarding timing of mental events; that neural activity related to movement starts before the reported time of conscious intention. A study using a paradigm resembling Libet’s were able to predict the outcome of free decisions to move either left or right hand up to 8 s before the reported time of conscious intention, with a certainty of about 60% (Soon, Brass, Heinze, & Haynes, 2008).

Philosophers and psychologists have used these kinds of results to argue, that conscious will is an epiphenomenon to the neural mechanisms which is the real causation behind action, and thus making the experience of volition an *illusion* (Cashmore, 2010b; Wegner, 2003b). According to Wegner (2003a, 2003b; Wegner & Wheatley, 1999) the mechanisms causally responsible for initiating the action, is in fact unconscious neural processes, as showed by Libet. These mechanisms then gives rise to two further processes: execution of the action, and a conscious intention about the same action. If there is consistency between the conscious intention and outcome of the action, it induces an illusion of the conscious intention as being the source of the action.

Although it is possible that conscious will is an epiphenomenon to neural processes, founding this on Libet's experiment is a too prompt conclusion for two reasons: The first being methodological problems concerning the experiment. Without going into excessive detail of the experimental design, the experiment have been criticized for various methodological errors, e.g. systematic visual bias in the clock (Joordens, van Duijn, & Spalek, 2002), that the subjects are performing at least two task simultaneously in watching the clock and deciding to move which result in an interference that bias the data (Danquah, Farrell, & O'Boyle, 2008; Lau, Rogers, & Passingham, 2006) and a great amount of uncertainty in the reported W-time (Matsuhashi & Hallett, 2008). Furthermore is the use of BP as physiological measure of brain activity prior to movement controversial. BP is an averaged measure of electrophysiological signal measured by EEG, and the anatomical origin is in principle indeterminable, though evidence point to the sensorimotor cortex, especially supplementary motor area (SMA), pre-supplementary motor area (pre-SMA) and primary motor area (M1) as sources (Shibasaki & Hallett, 2006). The onset time of the BP has been shown not to correlate with the subjective experience of action initiation in an experiment using Libet's paradigm. Instead the W-time correlated with the onset of the lateralized signal (Haggard & Eimer, 1999), although the LRP still preceded the W-time. Studies using fMRI have shown not only activity from the sensorimotoric cortex and SMA is present when performing voluntary movement, frontal areas, especially dorsal prefrontal cortex (DPFC) and parietal areas, especially intraparietal sulcus (IPS), have been shown to be involved as well (Lau et al., 2004). Only focusing on motor cortex when investigating voluntary action is insufficient to account for the areas involved (Klemm, 2010). These methodological implications have seldom been addressed in the philosophical discussion of the findings by Libet.

The second problem of claiming that conscious will is an epiphenomena is that the *kind* of free will Libet-type experiments investigate does not cover all aspects of intention, but just the subjective sense of intending to perform a motoric movement. This operationalization leaves out aspects such as decisions involving other than deciding the time to move. What can be concluded from these experiments is, that the conscious intention in immediate relation to executing an action *appears* to occur after the neural mechanisms involved in generating the action are initiated. Libet assumed a dualistic theory of consciousness, in which the conscious will is substantially different from neural processes. Conscious will emerged when neural activity reached a certain level (Libet, 1994). That conscious will first happened 350 ms after the neural initiators of action, must be build upon the assumption that subjects must have been unconscious *zombies* before the intention emerged. It has been argued that the intention to move the hand does not occur at W-time at all, but is formed prior to the experiment when the subjects were instructed to perform the movement (Mele, 2010; Zhu, 2003). There is no reason to assume that the subjects did not experience several thoughts and sensations before explicitly forming conscious intentions to move their hand. What Libet investigated was in this way not the subject's intentions, but the voluntary conscious implementation of prior formed intentions[3]. This raises the question of what the intention measured at W-time in Libet's experiment is an expression of?

A series of experiments by Bargh shows that unconscious processing contributes to goal-directed behavior, in an adaptive way, illustrating that the concept of intentions as entirely conscious phenomena is inadequate (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001). The unconscious intentional system should not be viewed as suppressed intentions or merely computational processing, but as contributing to an organism's behavior by providing an organization of a complex behavioral system, that was impossible if every aspect of behavior required conscious processing. Experiments have for example shown that skilled typists, actually have poor knowledge of their fingers movement during typing (Crump & Logan, 2010; Liu, Crump, & Logan, 2010). Due to the complexity of the unconscious intentions working in parallel, errors does however occur. It has been proposed that in order to minimize such errors the organism requires a higher level of awareness, thereby giving rise to conscious intention, as a mechanism responsible for action control (Bargh & Chartrand, 1999; Bargh & Morsella, 2008)

If a dynamical understanding of intentions is accepted, then the causal power of intention does not depend on a unified conscious entity causing action, but on the constitution of the dynamic system, from which both conscious and unconscious intentions emerge. This solves

[3] Libet seems to use volition and intention synonymously and does not likely make a distinction between the two.

this problem by distinguishing between different kinds of intention on behalf of their temporal properties and level of abstractness. An intention formed at time T_1 does not have to cause an action at time T_2 in order to be considered causally relevant. It just has to contribute significantly among other causations to the probability that the intended action is performed. It has been shown that when subjects deliberately form an implementation strategy toward a given action, the execution of that action is more likely to occur (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006). To gain more qualified categories regarding types of intentions, would also give a more nuanced picture of the neural mechanisms involved in forming intentions.

4 Disassembling the conscious agent

If intentions can be both conscious and unconscious and at the same time not is a unified entity, it raises the question of what the metaphysical nature of conscious will is. If volitions have to originate from the agent and without any prior causal influence on the mind, as initially defined for there to be free will, the necessary solution would be to view consciousness as something ontologically different from the physical world. The first-person characteristic of consciousness contra the third-person physical nature does imply such Cartesian dualism[4]. An immaterial consciousness could undetermined make truly free decisions, without being influenced by the deterministic causes of the physical world equal to a Kantian rational agent. The conscious agent is the immaterial consciousness, which as a singular entity can act upon the physical world (so far ignoring the mind-body interaction problem). The Cartesian position holds that consciousness belongs to an immaterial world, contra the material physical world. Therefore only immaterial entities can be conscious, and physical entities cannot. Such conceptualization of the agent is however challenged by the previously point that intentions can occur unconsciously, though not as a metaphysical concept, since it takes some of the causal relevance of the conscious will. Although consciousness might appear as a unified conscious field from a phenomenological perspective, it cannot be a priori settled that this is the case, as it could be a “trick” or an “illusion” by the brain that the subject experiences (Wegner, 2003a, 2003b). A scientific

[4] Cartesian dualism is here used in a similar way as Dennett (1991) uses it. This includes not just the explicit position held by Descartes, but includes various philosophical positions which have in common the view that *mental* is ontological and functionally distinct from the *physical*.

approach to free will should therefore consider whether consciousness really is a unified entity, before claiming anything about its causal properties.

In a survey of the subjective experience related to tics in patients with Tourette's syndrome, Leckman Walker & Cohen (1993) found that the majority of subjects reported a subjective urge to perform the tics and that they act on their own will on behalf of this urge, illustrating that subjective urges can be formed on behalf of unwanted stimuli. This means that subjective urges do not have to be a result of rational subjective considerations. At the phenomenological level volition does not necessarily have to be experienced as a unified process. That conscious will is not a unified force that acts upon the world and thereby generating action, is supported by findings, showing that conscious intention to perform an action, can be overridden by a new conscious intentions. Extended activity in dorsal frontomedian cortex (dFMC) have been correlated with such inhibition of intended action (Brass & Haggard, 2007; Kühn, Haggard, & Brass, 2009). This does not disprove a unified will, since it could be a matter of the unified will having one intention at time T_1 and another at T_2; simply meaning that conscious will can change. What it does illustrate is a further need to investigate at which rate that phenomenological properties of voluntary actions can change, and more basic what the phenomenological properties of volition is. Although the W-time in Libet type experiments might just be implementation of prior formed intentions, it is more than just the subjective experience of performing the movement. When subjects were instructed to attend to either their intention or the movement, they are able to distinguish their intention from the time of their experience of movement in Libet-style experiments (Lau et al., 2004; Libet, 1985). When subjects were asked to monitor their movement fMRI showed more activity in the cingulate motor area compared to more activity in the pre-SMA when asked to monitor their intention (Lau et al., 2006), indicating that the process of intention and action implementation relies on different neural mechanisms. Though the relation between consciousness and neural mechanisms is still unknown, it seems reasonable to assume that the biological basis hereof is distributed in the brain. The current state of neuropsychological research points to a distributed system involved in generating voluntary action, with a causal connected path from prefrontal areas through pre-SMA and SMA to M1, which prior to the generation of action receives signals from basal ganglia and prefrontal cortex (PFC). Furthermore do pre-SMA receive inputs from parietal cortex and intraparietal sulcus assumed to contain information about current sensor-motoric properties of the limbs and to be responsible for object oriented actions and also to contribute to the sense of ownership of actions (Haggard, 2008, 2009). Disrupting this path with transcranial magnetic stimulation (TMS) showed that of areas involved in voluntary action pre-SMA were highlighted as crucial contributor to the sense of agency and prediction of action outcome (Moore, Ruge, Wenke, Rothwell, & Haggard, 2010). In addition, applying direct electrical stimulation to certain areas in parietal cortex has shown to induce a conscious intention to move and in some circumstances inducing a sense of having moved, even though no movement was performed (Desmurget et al., 2009).

Blakemore and Frith (Blakemore & Frith, 2003; Blakemore, Wolpert, & Frith, 2000) have proposes that a “copy” of the motor command is sent to the motor cortex, known as a *forward model*. The forward model contains both information about the initial positions of the limbs and expected positions of the limbs, and its function is to predict outcome of the action in preparation. In order for a sense of agency to occur when doing an action, there has to be correspondence between the forward model and the actual outcome of action, referred to as intentional binding (Haggard & Clark, 2003). As a result hereof does the outcome of a voluntary action influence the phenomenology related to the action. The sense of agency can in fact be manipulated, depending on post-action circumstances or by priming intention (Kühn & Brass, 2009; Aarts, Custers, & Wegner, 2005). This implies that the neural processes involved in the phenomenological sense of agency rely on a variety of distributed processes, making the idea that the phenomenology of agency is an ontological singular entity implausible. The idea that the conscious will precedes its effects, understood as W -> A is challenged as well.

However not just the idea of conscious intention as a unified entity is challenged by the results from the neurosciences, The mirror-box paradigm (Ramachandran & Rogers-Ramachandran, 1996) and rubber-hand illusion (Botvinick & J. D. Cohen, 1998) is showing that subjects can attribute sense of ownership to objects not even part of their own body. Subjects experience sensations as a result of stimuli given to the illusory hand, as if it was their own hand and subjects tend to rate the illusory hand as if it was a part of their body and that the illusory hand took over the real hands place (Longo, Schüür, Kammers, Tsakiris, & Haggard, 2008). It has been shown that the perceived size of the real hand can be altered (Schaefer, Flor, Heinze, & Rotte, 2007), as well as the physiological regulations of the real hand as well, in the skin temperature of the real hand (Moseley et al., 2008) when the illusory hand is experimentally manipulated. Sense of ownership not only apply to illusory hands resembling the subjects own hand; there is no effect of altering the physical properties of the rubber hand (e.g. skin colour). First when the hand was replaced by an object not resembling a hand (e.g. a wooden block) did the effect of the illusion decrease (Holmes, Snijders, & Spence, 2006; Haans, IJsselsteijn, & de Kort, 2008; Tsakiris & Haggard, 2005a).

That illusions can appear of what subjects experience to have causally influence over, means that the conscious experiences related to an action, not necessarily reflecting the real causal chain of events. If the conscious sensation of initiation an action in fact is caused by the action itself then it cannot logically be the source of the action, thereby giving a strong argument in favor of eliminating free will altogether. This would require that the concept of free will is identical to the sense of agency. This is not the case and the mere sense of agency cannot be used synonymously with free will. Free will includes processes not related to sense of performing actions, such as planning and decision-making. The sense of agency is only the subjective experience of being the one who initiates an action, including a sense of ownership of one’s body (Haggard & Tsakiris, 2009). Compared to the previous conceptualization of intention, the research on the sense of agency only covers the proximal intentions and leaves out the distal intentions. Sense of agency is still a useful conceptualization when investigating the conscious aspect of voluntary actions, since a subjective sense of agency is viewed as a requirement for actions to be viewed as voluntary and intentional (Roskies, 2010).

In the traditional Kantian sense, intentions to act arise from decision-making. Decision-making can conceptually be divided in three stages, the first being forming of preferences and intention, then execution of action and finally the action-outcome, involving a vast array of areas (Ernst & Paulus, 2005). In decision-making tasks such as economic risk-games it has often been shown that parietal cortex seems to be involved in processing and perception of probability, whereas activity in anterior cingulated cortex (ACC), has been hypothesised to be responsible of integration of task-relevant information and integration of conflicting opportunities, which is crucial in making choices (Botvinick, Braver, Barch, Carter, & J. D. Cohen, 2001). ACC have therefore been describes as responsible for cognitive control in relation to behavioural responses (Carter & van Veen, 2007). ACC is linked to a network of action selection receiving input from the dorsal-visual system through fronto-parietal and medial prefrontal cortex (Cisek, 2007) as well as receiving signals from the amygdala responsible for emotional processing (Bechara & Damasio, 2005) Though all these mechanisms are important in the generation of decisions, the research indicate that this processing to a great extend is unconscious (Brass & Haggard, 2010). This assumption is supported by studies showing that subjects likeliness for moving either right or left hand can be biased using TMS even without the subjects became aware of using the one hand more than the other (Brasil-Neto, Pascual-Leone, Valls-Solé, L. G. Cohen, & Hallett, 1992).

Based on the brief previous review, it is reasonable not to approach conscious will as a unified entity. Neuroscience shows that the action generation as well as the phenomenology in action generation is a result of various modules interacting. Instead a connectionist approach seems more adapt to explain the mechanisms involved in conscious will and voluntary actions. In accordance with this approach Zeki (2003) has proposed that integration of visual perception occurs through a hierarchical organization of processing sites in the brain. Distinct areas in the visual cortex have shown to process distinct features of the perceived object (e.g. colour) and giving rise to the phenomenological experience hereof (e.g. redness), termed micro-consciousness. In this way the coherent conscious experience M can be expressed as the sum of several minor phenomenological properties or qualia ($M = \sum M_1\ M_2\ M_3$ ...) each originating from different neural modules (P_1, P_2, P_3 ...). Several distinct micro-consciousnesses related to features of a perceived object are integrated into a macro-consciousness, resulting in a coherent conscious experience. Since the integration between modules relies on neural processing, it takes time for one module to be integrated with another. It is therefore not possible to give any specific point in either space or time where the unified phenomenological experience appears. There is not one phenomenology, but an integration of many "phenomenologies" described as a *heterophenomenology* (Dennett, 1991). The connectionist perspective should however not assume an equal connection between modules since the research of neural underpinnings of the sense of agency, and visual cognition, supports the hypothesis of a distributed dynamic hierarchical system. A replication of Libets experiment using patients with neural lesions showed that patients with parietal lesions who were able to move their hand and to report the time they moved their hand, were unable to report the time they intended to move their hand (Sirigu et al., 2003). This indicates that action preparation and awareness hereof happens in stages, where the patients were unable to consciously access the early stages in action generation, more specific hypothesized to be the internal model of the action.

A scientific approach to free will should not start out by assuming a unified self or phenomenology as a singular entity. Instead a connectionist approach to both the neural mechanisms and phenomenology seems more adequate. What a neuropsychological approach should focus on is; how the various modules interact in order to generate voluntary action, and then how the subjective experience of volition arises, as a result of the functional organization of the underlying modules. Thereby being able to describe which areas that are related to, and basis of the conscious experience and which areas that has more or less automatic processing.

5 Naturalizing conscious will

An objection to what has been argued so far is that, it is not really putting forward an approach to free will as much as post hoc accounting for what research already has shown. The metaphysical nature of consciousness, whether it is a feature of the physical world or immaterial, cannot be deducted a priori (Dowell, 2006). By a priori setting the metaphysical bounds a science of free will cannot be able to contribute to the philosophical debate regarding the metaphysics of free will. What have been done so far is to broaden the concept of free will and therefore still an a priori assumption. Based on the research voluntary action it is strongly suggested that no single site in the brain holds the entire conscious agent or that consciousness should be detachable from neural processes. A Cartesian mind does however have a problem regarding causal impact. As initially defined, actions can be described in terms of basic physical features and are therefore physical phenomena. Immaterial minds do not have any physical properties whatsoever, and if the physical world is causally closed, then there is no room for any causal relation from an immaterial mind to the physical world (Kim, 2005). In order for conscious will to be causally relevant in a physical world it needs to have spatiotemporal properties that can act in the physical world and cause action[5]. This leaves the immaterial mind as an epiphenomenon to the physical cause and effect. Furthermore are immaterial minds not in any way measurable by scientific inquiry, since it would leave no traces in the physical world. Even subjective methods, such as introspective reports, rely on neural mechanisms as well as any other behavior in the physical world. There is no action that could tell of the phenomenology and every action that would point toward expression of

[5] Though the *anomalous monism* proposed by Donald Davidson (1980) argues against this definition.

phenomenological states, would in principle be expressible by a zombie as well (Chalmers, 1996). This is however not an argument against the metaphysical possibility of immaterial minds but it is an a priori conclusion that Cartesian dualism gives no better understanding of mental causation than eliminativism from a scientific perspective, resulting in “zombianism”. The neural processing of a zombie is on behalf of the exact same sensory input, the exact same neuro-cognitive processing in exactly the same way, and has the same behavioral repertoire and acts in precisely as many ways as an ordinary human being. The only thing that the zombie is missing is phenomenological experience associated with the actions. The zombie is still as free or determined as humans, and must be considered the agent of own action, as much as humans. If moral actions depends on actions which originates from the agent, then the zombie is as moral responsible for its actions as the ordinary human. Assuming the universe is causally closed, it reduces human free will to the same kind of agency which the zombie has. Conscious will must be considered an epiphenomenon and can therefore be reduced from the investigation, due to the principle of causal exclusion (Kim, 2005). The idea of zombies is only a tool for illustrating the fallacy of Cartesian dualism, even if zombies cannot be disproved, giving only first-person characteristics separate them from ordinary humans, which per definition is inaccessible from a third person perspective. The zombie argument should therefore be interpreted as a need for naturalizing consciousness, by viewing phenomenology as based in biological process (Bailey, 2006).

A naturalistic approach is here used as an umbrella-term to cover various positions that views consciousness and physical processes as functionally connected. This can be done in several ways, either as monism or materialism holding that consciousness and physical processes is one and the same metaphysical entity, *the nothing over and above physical processes* approach (Smart, 1959), or as naturalistic dualism (Chalmers, 1996) which holds that consciousness and the physical is ontological different, but where consciousness logically supervenes on physical processes, thereby in principle inseparable from the physical processes. Although the materialists and naturalistic dualists disagree on the metaphysical nature of consciousness, they agree that consciousness is tied to functional properties in the physical world, as opposed to Cartesian dualism.

A critique of the proposed naturalistic approach is that when conscious will is tied to neural processes, then free will could be explained in terms of neuro-cognitive mechanisms alone without including phenomenology. The goal would then just be to describe the neural processes in order to understand the function of conscious will and thereby embracing reductionism and return to the same kind of zombianism as Cartesian dualism resulted in. Viewing consciousness and neural activity as identical or at least functionally supervening, then for each phenomenological state there must be one specific neural state ($P_x = M_x$) (Smart, 1959). If one could give an exhaustive picture of the brain which accounts for the exact amount of activity in each neuron at time T, then one would have an exhaustive picture of the phenomenological aspects at time T as well. If the state of P was changed by removing one neuron *N*, we would have changed the state of P to P'. The mental state would therefore change as well from M to M', so that $P' = M'$. This is not necessarily what is meant by the naturalistic approach that is proposed here. The advocated naturalism should be understood as a mild functionalism that merely states that the mind is a functional property of neural activity, and is not the same as claiming a strong form of computational functionalism (Piccinini, 2010). What is proposed is that the ontological nature of consciousness is unknown but somehow tied to physical processes and thereby belonging in the physical world. The idea that for each neural state there is one and just one mental state is problematic considering the plasticity of the brain shown in rehabilitation after brain damage. Lost functions and even phenomenological loss can be regained but will be relocated in a new neural area, challenging the assumption of strict identity and suggest phenomenological events are multiple realizable (Overgaard & Mogensen, 2011). However it may be that the reacquisition of a lost mental property M (e.g. the control of a arm) due to damage in a neural area P_1, is in fact not the same as mental property M but only a similar mental property M', when the function is relocated to a new area P_2, so that $P_1 = M$ and $P_2 = M'$. Since $P_1 \neq P_2$ then $M \neq M'$. Whether the regained mental ability is M or M', cannot be scientifically validated because it is a first-person characteristic. What this example illustrates is the need for focus on functionality of phenomenology in relation to brain areas, instead of a correlation of global states of neural activity and global phenomenological state. Since free will was defined as a process and thereby is a functional feature of the world, whether a phenomenological property is M or M' does not make any difference if they are functionally identical. The interesting part for a science of free will is when functionally properties are changed, such as alien hand syndrome, which occurs after inter-hemispheric disconnection due to damage in the mid-body of the corpus callosum (Geschwind et al., 1995). These patients still have functionally arm movements but the arm is perceived as outside their own control, as if controlled by another agent. Even though they logically are generating the movements themselves but are missing the phenomenological aspects of doing so, thereby losing their will over the arm.

The empirical findings and connectionist approach points, as previously presented, towards a functional understanding of consciousness when investigating conscious will. In the initial definition of free will it was a requirement that the agent’s volition s the causal starting point in a serial causal chain ($W \rightarrow A$). In a connectionist approach, such as previously described, it is difficult to set a functionalistic starting point, since different modules work in parallel in an inter-connected way. If many modules contribute with different aspects information in the phenomenological will, and the integration of this information takes time, then a serial account of mental causation is impossible. The causal power point depends on how the different modules are connected. In the dynamic approach to mental causation it is however impossible to set a functional causal starting-point, since there is always modules causally influencing other modules (assuming that brain-activity does not shut down occasionally). The problem of mind-body interaction becomes further qualified by not simply being *if* a conscious will can cause action, but which phenomenological properties that is associated with neural processes or modules, and how these causally act upon each other. What needs to be uncovered in order to understand mental causality, is how modules causally influence each other in the complex dynamic system. Taken together with the findings that much information processing is performed unconscious, the concept of *the agent* as a unified entity is not a useful conceptualization.

It has been shown experimentally that it is possible to distinguish functionally separated phenomenological states, when the agent is *passive* and when the agent is *active* (Gallagher, 2000; Gallagher & Marcel, 1999; Tsakiris & Haggard, 2005b). The passive agent is described as a sensory state in which the subject just perceives stimuli. The active agent is comparable to the sense of agency, and is involved in the generation of action. When an action is voluntary, the phenomenological properties of the action are different from the phenomenology of non voluntary or passive actions (Ebert & Wegner, 2010; Haggard & Tsakiris, 2009) These differences involve measurable changes in spatial and temporal experience of both one’s own body and external objects. For example in an experiment where subjects had to report the times of various mental events, when performing either an voluntary movement or received TMS over M1 to induce a similar but involuntary movement, showed that when a movement is self-initiated itis perceived to occur about 50 ms before the action actually takes place (Haggard & Clark, 2003; Haggard & Magno, 1999; Haggard, Clark, & Kalogeras, 2002). That the phenomenology related to motoric movement occurs before moving is explained as the forward model, generated when initiating a voluntary movement is the source of conscious awareness and not the movement itself (Ebert & Wegner, 2010; Haggard, Newman, & Magno, 1999). The shift in subjective experience is the result of the generation of an anticipatory forward model of what outcome to expect when an action is performed by an active agent. The temporal delay between the initiation of action and when the action-effect feedback is received is crucial for the intentional binding. Several studies have shown that this time delay varies somewhat between 200 ms to 300 ms after performing the action (Blakemore, Frith, & Wolpert, 1999; Blakemore, Wolpert, et al., 2000; Farrer et al., 2003; Franck et al., 2001; Lau, Rogers, & Passingham, 2007).

In accordance with the connectionist approach Moore and Haggard (2008) have proposed that the predictive process is modulated by recent experience of the action-effect relation and that the sense of agency therefore is a result of a multimodal integration of processes. This integration process takes time to occur and it is therefore impossible to pinpoint when the sense of agency appears on a smaller timescale (Dennett & Kinsbourne, 1992; Haggard & Tsakiris, 2009).

6 Voluntary action in context

Much psychological research can be criticized for investigating the object of interest, in a setting that is fundamental different than the context in which the object of interest normally exists, and hence the results doesnot necessarily reflect how it would behave under any other instance than the laboratory, rendering the result somewhat unimportant (Neisser, 1978). Libets experiment can be target of such criticism. The task of letting subjects focus on when they become aware of their intention to move a finger might strange for the subjects, since little conscious attention is directed to the movement of the fingers, compared to how much a subject is using their fingers in their everyday life. It has been shown that in the case of typing, attention is linked to a higher hierarchical level than finger movements, in this case the formulation of the word that were typed (Crump & Logan, 2010; Liu et al., 2010). Gallagher and Marcel (1999) have proposed a contextualistic framework in which intentional action is directed toward the higher-level goal. What subjects usually are conscious of during the performing an action, is the highest meaningful semantic level of their behaviour. The lower levels in the hierarchy do not receive much direct conscious attention. This does not mean that attention cannot be directed toward specific movements. The object of attention focus can be directed to external objects as well as internal mental events, but when *in-action*, the higher hierarchical semantic level becomes the focus of attention. Conscious will should therefore not only be viewed as intention to perform movements but will as intention directed toward external as well as internal objects. In this way the agent is in an interaction with the surroundings in forming conscious will. The agent receives information from its environment on which meaning and action is founded (Shotter, 2003). This is not a remarkable approach, since it is clear that the brain constantly receives enormous amounts of information through various sensory modules, both from the external world and the body, which must be processed, and these influence the outcome and behavior of the agent. The pathway that leads to a conscious choice is not a linear causal chain but should be approached from a dynamic and context-bound view of agency (Skewes & Hooker, 2009). The conscious will operates as a result of a complex dynamic system that constantly receives input from its environment and at the same time acts toward the same environment, where the outcome of the agent's actions then can become stimuli for further action and so forth. The further argument against at least simple computationalism in stimulus-respons generation is that the input the brain receives from the environment cannot be viewed as a single unified stimulus. The external influences are received through various sensory- and perceptual modules (e.g. visual, auditory, sensory etc.) which are integrated in the dynamic connections between the various modules. The number of inputs the agent receives is huge, and therefore it is not possible to identify a sole causal chain that accounts entirely for a given action. The decisions the subject makes must in part depend on these external stimuli, since it has been shown that the sense of agency can be influenced by external subliminal priming (Moore, Wegner, & Haggard, 2009; Aarts & van den Bos, 2011). Also social cues can influence the subject's actions. When a go/no-go reaction task was performed, and another person was doing the exact same task to the same stimuli next to the subject, the actions of the other person influenced the subject's responses to the task (Sebanz, Knoblich, & Prinz, 2003).

The argument against contextualism when studying free will is that if an agent's volitions are influenced by external factors, then the will is not free at all, and therefore this would eliminate free will. The problem with this objection is that it is assumed that the external stimuli fully determine the subjects will. This overlooks that the conscious will of the subject originates from a complex system. Even if the agent is determined by external stimuli, it is not in a linear fashion. The complexity and hierarchical organization of the agent's internal processes gives rise to a self-organization, thereby generating *autonomy* for the agent (Felsen & Reiner, 2011; Frankfurt, 1971). The self-organization as a result of the hierarchical organization means that when considering the causes of an action, it is not just the external stimuli, but the preliminary internal stimuli, conscious or not, contributes as well. When a significant degree of internal factors contribute to action, it is appropriate to consider an organism as autonomous from its surroundings (Baumeister & Vonasch, 2011; Frankfurt, 1971). This is not the kind of detached autonomy which initially was viewed as a requirement for free will. Instead it is the ability of an embodied agent to choose among a possible array of behavioral responses. Such organization would in fact be an evolutionary advantage for an organism over simply automated responses (Brembs, 2011; Dennett, 2003). From a theoretical- biological perspective it has been shown, through the use of game theory, that an organism that expresses unpredictable behavior will be favored in natural selection (Maynard Smith, 1982). Here autonomous behavior is a result of processing in the organism, which can generate several responses (R_1, R_2, R_3 *etc*.), contra automated responses where for one stimulus there is one response. The study where the outcome of a choice of moving either left or right hand , could be predicted before the conscious decision with about 60% accuracy by looking at brain activity (Soon et al., 2008), should not be interpreted as unconscious brain mechanisms outside conscious awareness causing action. Instead it should be viewed as activity within the system that *contributes* to the decision, is prevalent already eight seconds before (Klemm, 2010). This supports the claim, that unconscious prior processes have a role in forming conscious intention, illustrating that the traditional way of looking at mental causation is not an adequate explanation. Following this definition it is not just humans that possesses agency. It has been shown that relatively simple organisms such as fruit flies do not behave in a strictly deterministic fashion, but is capable of showing stochastic behavior (Maye, Hsieh, Sugihara, & Brembs, 2007).

A characteristic of human autonomy is the ability of the agent to distinguish own actions from others actions. Based on the finding related to mirror-neurons (Rizzolatti, 2005) Decety and colleagues did a series of experiments investigating the neural activity when imitating others actions (Decety, Chaminade, Grezes, & Meltzoff, 2002; Ruby & Decety, 2001). They found that when the subjects imitated an action done by the experimenter, the left inferior parietal areas showed greater activity than control conditions. When the subject's actions were imitated by the experimenter they found increased activity in the right homologous region. They therefore concluded that these regions are responsible for the ability to distinguish one's own action from actions made by others. In a later study they asked participants to either take a first- or third-person perspective on a presented case. This showed activity in the frontopolar cortex, somatosensory cortex and the inferior parietal cortex related to distinction between self and other, leading to the conclusion that these areas, and the right inferior parietal lobe, are crucial parts of generating a sense of self (Ruby & Decety, 2004). In a task of judging whether a moving hand presented on a screen was their own or not, apraxic patients with lesions in left parietal cortex made significantly more errors than a control group, when the movements were similar to the patient's own (Sirigu, Daprati, Pradat-Diehl, Franck, & Jeannerod, 1999). The recognition of own actions seems tied to parietal cortex, the area also associated with planning of motoric action. Not just patients with brain injury show an altered sense of agency. Schizophrenic patients with hallucinating symptoms made significantly more errors in pointing out their own hand than patients with non- hallucinating symptoms (Daprati et al., 1997), which implies that schizophrenic patients experiences delusions of influence and have trouble with self-attribution of actions. Frith (Frith, Blakemore, & Wolpert, 2000a; Frith, 2005; Frith, Blakemore, & Wolpert, 2000b) has hypothesized that the schizophrenic patients disturbed sense of agency, e.g. that they experience their actions as controlled by others, is a deficit in the mechanisms involved in action generation. It is proposed that the schizophrenic patients are lacking a connection between the intended motor command in the frontal areas and the sensory feedback from the sensory areas, and therefore they cannot properly integrate the two. This is supported by an experiment showing that schizophrenic patients in fact can tickle themselves, or at least perceives self-generated stimuli as more intense than normal subjects (Blakemore, Smith, Steel, Johnstone, & Frith, 2000), which implies a lack of forward model which normally would attenuate the perceived intensity of self-generated stimuli.

The contextualized naturalistic framework here proposed breaks from the traditional view of an idiosyncratic agent in favor of agency as a biological trait, in which autonomy is a functional property of an organism (Brembs, 2011; Skewes & Hooker, 2009). This breaks with the philosophical idea of a Kantian rational subject as a source of human action. It has in fact been shown that humans are bad at making rational decisions and that choices tend to be biased and based on heuristics rather that rational decision-making (Tversky & Kahneman, 1974). This does however not mean that humans are not able to make rational choices, but that this ability merely is a small part of action generation and depends, as all other behavioral response, on the organization of the dynamic system (Felsen & Reiner, 2011). Damasio (Bechara & Damasio, 2005; Damasio, 1996) have proposed that rational decisions are based on the emotional body-state of the subject. The ability to perform rational choices is build upon an internal regulation mechanism, evolved to provide the subject with somatic states, so called somatic markers, that gives rise to certain behaviors, e.g. fear of snakes will result in flight. Emotions are crucial contributors to decision-making processes, as they function as cues of whether an option is beneficial or potentially harmful for the organism. Somatic markers are flexible and generally adaptable, where learned states become stored in memory. Simply recalling a given stimulus will elicit a somatic state corresponding to that stimulus. Somatic markers are therefore both involved in rational choice-making, shown in economic games, as in any other decision-making context not based on rationality. The body-state is hypothesized to consist of an interaction between the amygdale and the ventromedial prefrontal cortex (Bechara & Damasio, 2005). Basing the concept of free will upon rational decision-making is narrowing the field down to an exclusive field among the broader field about how choices are made by subjects in everyday context would be missing. The automated processes are highly adopted to perform the task at hand, but they do so in parallel, making it possible that conflicts occur between opposing behavioural responses, making conscious awareness hereof an adaptive control mechanism (Bargh & Morsella, 2008). This can be illustrated by learning tasks, where conscious attention is directed toward internalizing the given task, which when learned, depending on the task, becomes automated process as well. Several studies have investigated the neural activity during learning tasks where subjects had to learn a given sequence, either verbally or by moving their fingers by trial and error, until the sequence could be performed automatically (Jenkins, Brooks, Nixon, Frackowiak, & Passingham, 1994; Rowe, Friston, Frackowiak, & Passingham, 2002). These studies showed extensive activity in PFC especially the prefrontal and anterior cingulated cortex, when the subjects were trying to learn the sequence, but not when the actions had become automated.

Suggesting that the ontology of conscious will is a complex system, is not the same as claiming that it cannot appear unified from a phenomenological perspective. In the connectionist-naturalistic framework presented here, a unified sense of agent, not reflecting the organization of the involved neural mechanisms, could be an adaptive mechanism from an evolutionary perspective. Acting in a context would be impossible, if every movement required conscious attention. Instead the conscious attention is kept at the highest or functional level, meaning the semantic meaning of the intended action and not the kinetics (Gallagher & Marcel, 1999). That motoric movements then mainly operate outside the conscious content, is not an argument against conscious will as being functional. The automatic processes are *guided* by the higher level semantic understanding of the action and the situation. A series of experiments had subjects perform various movements, and then a week later showed the recorded kinetic trajectories of those movements, or trajectories of movements done by someone else. In this task the subjects were able to correctly point out which movements were their own and which were not. When the speed of the shown movements later was manipulated, the subjects were unable to distinguish which movement that was their own. As a conclusion to these experiments it was suggested that simple kinetic properties of the action are stored in memory (Knoblich & Prinz, 2001; Knoblich, Seigerschmidt, Flach, & Prinz, 2002). Categorizing these actions as unconscious might therefore be misleading, Pre- and/or subconscious might be better terms, since the transition from automated behavioural response to consciously intended action is not well distinguishable.

7 The problem of freedom

So far the focus of this article has been on the problems regarding mental causation, which is only one of the problems regarding free will. While the problem of mental causation itself is a hard question, it has been argued that the real hard question for a science of free will is whether actions can be *free* (not just autonomous) if the universe is deterministic (Shariff et al., 2008). Even if mental properties interact with the physical world and the individual is autonomous, this interaction would itself be determined, no matter how complex it would be, and therefore un-free. The analytical problem of freedom is in principle unanswerable by scientific means, since it would be impossible to test whether an event could have turned out otherwise. There is no way to go back to observe it over again. Truly free actions would be unpredictable by any means, which per definition makes it impossible to scientifically investigate, since science can only study causal relations among events (Hinsen, 2010). These are metaphysical questions unanswerable by scientific investigation, but such investigation can support either the libertarian or deterministic hypothesis regarding the causal structure of the universe.

A promising case in favor of libertarianism is quantum-mechanics, which implies that the universe, at least at a sub-atomic level, expresses genuine random processes. Penrose (1989) argues that human consciousness and rational capacity is unexplainable by any algorithmic or rule based approaches, and the only physical mechanisms that allow this kind of phenomena is the quantum uncertainty; referring to quantum processes as a contributing factor in the neural processes underlying conscious will. Kane (1996) has proposed that neurons act as amplifiers for quantum processes which would allow the neural functioning and would allow behavioral responses to be an expression of genuine freedom for the subject. When the subject makes a choice no prior causal mechanism is the cause of this choice. A hypothesis on how quantum processes influence neural activity, it has been proposed that self-collapsible waves in the cytoskeletal microtubules of the neurons is the origin of consciousness (Hameroff & Penrose, 1996).

Although some interpretations of quantum mechanics imply that the universe is not deterministic, the hypothesis that human free will is explainable by quantum mechanics is widely criticized. The counterargument is that quantum mechanics does not prove any argument for libertarian free will, since the genuine randomness provided by quantum processes hinders a causal relation between mental events and actions. This leaves the role of conscious will as useless as under the kind of determinism the quantum approach tries to disprove (Dennett, 2003; Roskies, 2006). A further problem of the quantum mechanistic approach is that the timescale on which quantum processes operate and the timescale on which neurons fires, is different with a factor of 10^{10} (Seife, 2000). It is disputable whether the randomness of quantum processes at the sub-atomic level can create randomness at a molecular level and even more disputable at a behavioral level.

Another defense of libertarianism is that the function of the ion channels in neurons can show stochastic behavior (Dorval, 2006). Though this shows that seemingly random mechanisms can be present in the brain, without resorting to quantum physics, there is still an epistemic leap from stochastic behavior in neural-transmissions and to conscious will. It is however not a gap that cannot be bridged by neuropsychology. Others have shown that stochastic behavior of a neuron transforms into what seems like ordered timed sequences of spikes, which could be a part of the higher-order cortical processing of information (Mainen & Sejnowski, 1995). In order to resolve this problem further research have to give a better understanding of how the micro levels of the brain are related to the global functioning (Brembs, 2011).

Determinism is as mentioned often assumed to disqualify free will. Dennett (2003) has however argued that in a deterministic universe where simple elements[6] follow strict causal rules and always behave in accordance to these rules, these clusters are able to express what looks like self preserving qualities when clustered together. While the lower level components (e.g. neurons) follow the physical rules, the ability to be self-steered emerges when clustered together. The organism is still determined but its organization is so complex that it is practically impossible to predict the outcome, so "[…] *even though Laplace's demon knows exactly how each composition will end, there may be genuine drama and suspense for lesser intelligences, who cannot know, from their limited perspective, how the contest will end*" (2003, p. 91). In complex biological systems (such as the brain), that are constituted by a high number of components, with multiple effect upon eachother, the number of possible states so large that it appears to be unpredictable. So even though the organism is not free in the metaphysical sense, the complexity does that it has "freedom *enough*". It has been argued that what is usually understood by freedom, when conceptualizing free will, should be reframed as an internal self-organization of an organism which allows for autonomy of the organism. The crucial aspect is the organization of the system.

Dennett (2003) hypothesized that in order to understand the function of enormous dynamic systems, such as the brain, methods derived from chaos theory and nonlinear dynamics might be methods of analyzing such complexities. The appeal from chaos theory is, without going into further details, that chaotic systems *seems* to be driven by randomness, but are in fact governed by complex laws in a deterministic way (Lorenz, 1993). This approach to neuropsychology is underdeveloped, but so far neurons showing chaotic behavior, have been shown in the hippocampal and olfactory systems (Tsuda, 2001). The investigation of chaotic behavior of neural activity can in principle be investigated, but it requires complex mathematical modeling and it is complicated by the enormous number of neurons, and the even greater number of synapses they interact through.

[6] In his example Dennett uses a "life world"; a democritean universe consisted of finite elements following simple rules.

The idea of complex systems broadens the common understanding of determinism in relation to free will and mental causation. While causation can be expressed simply as; $X \rightarrow Y$, such a description applies only to basic physical components. When it is applied to higher-order processes, it becomes a violation of the principles of reduction, since higher-order processes are the result of numerous processes at the lower level causally acting upon each other. If a metaphysical claim about the exact causal relation on the higher level is to be made, then it would in fact have to be an exhaustive account of *all* causal relations on the low level. Therefore higher levels can *practically* at best be described as probabilistic. When moving the finger to pres a key, an immeasurable amount of differences could be present between two seemingly identical actions as the finger could hit the key slightly different. This would alter the physical identity of each press, and logically make the distinct actions. Assume that a subject in an experiment presses a key and does so in two identical ways, except the absence of a single neuron *N*. Except neuron *N*, every other physical property involved in the two actions A and action A' is identical. If described in terms of higher-order function they would in fact appear as similar actions, as it is practically impossible to measure the difference between A and A'. These two actions, as indistinguishable as they practically might be, are still, in the metaphysical sense, two different actions. Even if there logically are near infinite possible outcomes of behavior, a great deal of these can be clustered together and labeled as one and the same action on a higher descriptive level. Categories on higher descriptive levels contain several metaphysical possibilities which appear similar (not *identical*) even across time. Taking this into account, the scientific approach rises above the fatalist and Libertarian objection that science cannot deal with the question of whether events could have unfolded otherwise. While it is true that this question cannot be answered scientifically, whether a subject hits a key in a key-press task slightly different each time, is in fact unimportant, as it, at a higher descriptive level, has the same effects.

If action A' is further differentiated from A by the absence of two neurons firing, it still seems reasonable to consider the two actions in one category, if three neurons then the same and so forth. But at some point the difference between A and A' would be so different that it does not make sense to put these in the same category. This gives rise to an important consideration when merging a high number of metaphysical states into a single higher-order descriptive category. There is no simple solution to this practical problem, other than when considering the action in an experimental paradigm, there has to be some valid internal coherence that justifies that each specific action is considered as the same action at a higher descriptive level.

8 A neuropsychological approach to investigating free will

It has been proposed that free will is an outdated construct tied to the belief in a vital immaterial soul that acts upon the world, and since this is an outdated concept free will is not a topic that can be studied scientifically (Cashmore, 2010b). Such an eliminativist position, which a priori denies free will as a topic worth investigating, will miss out a central question; namely why people, no matter whether genuine free will exists or not, have a subjective experience of being the master of their actions. It could very well show up to be the case that the experience of volition and subjective sense of agency is an illusion we fool ourselves with (Wegner, 2003b), but this does not answer why a subjective experience of volition accompany certain actions.

The start of a neuropsychological approach to free will is to make the fundamental assumption that people experience something that can be described as volition regardless of metaphysical nature. From here there is a need for further conceptual clarification of exactly how the different kind of “acts” in generating volition should be defined, such as decision and intention. Making the claim that subjective will exits, no conclusion regarding the metaphysical nature of will is made. All that has been postulated is existence of a subjective *what it is like to have will*. First step to investigating free will is therefore to make a clarification of what the object of the investigation actually is. When concepts so diverse as subjective sense of agency, decision-making and intention all can be included in the concept of free will, it is clear that simply investigating one of the instances will leave something out (Roskies, 2010). When interpreting what neuropsychological experiments tell about free will, it is worth considering what *kinds* of will that are being investigated.

The purpose of the connectionistic, naturalistic and contextualistic approaches advocated here is to leave the arguments of simple mentally caused action, in favor of viewing the agent as a complex system, constantly receiving and interacting with information in the context. The ontology on conscious will, although unknown, is bound to this complex system.

Some departures have to be taken from the initial definition of free will. The relation among conscious will and action, where a singular agent is the cause of action ($W \rightarrow A$), have been discarded in favor of a connectionistic perspective, where conscious will is tied to neural processes in a complex dynamical system, in which the causal power of the agent depends upon the constitution of components involved. In addition the agents must be viewed as both dependent and influenced by external as well as internal factors in the generation. The proposed neuropsychological approach does not necessarily require the agent to be metaphysically free in order to perform actions on behalf of own intention or will. In the translation from philosophy to science the topic has changed slightly from *free will* into *autonomous will*. This redefinition is however not irrelevant from a philosophical perspective, quite the contrary.

The extreme libertarian version of free will, where any logical outcome is possible in the real world, is too idealistic, since the physical reality without doubt sets constrains on the possibilities of expressible behavior by the agent. Instead of trying to answer whether freedom exists, a neuropsychology of free will should focus on exactly how much autonomy the self-organization of the human brain can provide, and to show which processes that are outside the grasp of subjective will, by uncovering which mechanisms that are involved in generating autonomous behavior and morally guided actions. Thereby illuminating which actions that are "*freer*" than others (Baumeister, 2008). It is important to distinguish between being metaphysically free and pragmatically free, at least for analytical purposes.

Instead of debating if conscious will has a causal impact or not, the focus should be on investigating; how exactly this phenomenological experience is manifested, the mechanisms that give rise to these, and then the causal connections they are an integrated part of. A science of free will should not adopt the logical definitions of free will, such as "if *X* (or if not *X*) then free will does not exists". The quest of neuropsychology should be to investigate why people experience will and which mechanisms in the brain that lay behind this experience (Baumeister, 2008; Roskies, 2006, 2010). This can be described as a plea for investigating Chalmers' (1995) *easy problem* (instead of starting out by trying to answer the hard problem), which is to explain which areas that are associated what kinds of subjective experience.

Chalmers (2000) have formulated the requirements to describe neural correlates of conscious as "*a minimal system whose state is sufficient for a given conscious state and whose state is not wholly correlated with the state of any other system"* (Chalmers, 2000, p. 13). Specifically in relation to the neural correlates of *conscious will*, it becomes complicated, given that when investigating free will the focus is not on neural states, but on neural functionality. It cannot a priori be ruled out that the functional correlates of performing a voluntary action might involve mechanisms not specifically giving rise to conscious will, such as selection of task relevant information which does not necessarily become the content of conscious will, but still is a crucial contributor (Brass & Haggard, 2010; Egner & Hirsch, 2005). If a hypothetical module (P_1) not associated with consciousness contributes to a module (P_2) responsible for a certain phenomenological property (M), should module P_1 then be considered a functional correlate of M as well? This must lead to a practical distinction between the neural processes of generating voluntary actions and intentions, and the neural processes directly responsible for phenomenological properties, even through these two systems might overlap in functionality.

Such an approach would in fact be lucrative in relation to investigating autonomy, since it would clarify how conscious will is influenced by external factors and thereby how autonomy is manifested in the brain, both phenomenological and functionally.

A topic in the neuropsychological science of free will is to clarify exactly when and how the transition from sub-conscious and conscious occur, and which neural mechanisms they depend on. The question of what one is conscious of is in itself complicated. It might be that the seemingly unconscious action preparation in fact is conscious, but does not leave any traces in memory, or it might be that one is only faintly aware of this information which then could “drown” in the wider integration process. So far it illustrates a need for investigating what a subject actually is aware of when performing voluntary actions, and how this is different from non-voluntary movements. What aspects of the phenomenology of volition are being stored in memory is also a question worth addressing. This relates to the question of what and how phenomenal aspects are stored in memory (Block, 2005). It is likely that not all aspects of a conscious experience leave traces in memory and they are therefore impossible to recall post hoc, since brief phenomenal experiences simply slide out of memory. When Libet, and others using the same experimental paradigm, claimed to investigate the timing of mental and neural events and then compare the two timelines, the subjective reported times might not reflect the phenomenology as much as it has to be analyzed as an expression of reconstructed times on behalf of memory, which are easily manipulated (Aarts & van den Bos, 2011).

Since conscious will is a subjective phenomenon, it is necessary to include first person methodologies, such as introspective reports (Overgaard, 2006). Much critique regarding first-person methodologies take issues with the fact that subjects have limited or none at all insight into the underlying cognitive processes which guide behaviour and phenomenological reports therefore does not give any insight into the underlying mechanisms (Nisbett & Wilson, 1977). This point is an important consideration, but a misguided critique, since there is no a priori reason why conscious knowledge of action should be of the underlying mechanisms. It has been proposed that the function of conscious will is behavioural guidance and control, and as a result it does not necessarily include relatively simple processes. This should not be viewed as an *error* but as a *feature*. Instead of rejecting introspection, because it is not an insight into the processes itself, introspective methods should be used to clarify what the phenomenological properties of volition are, and then why the underlying processes make them appear as they do. This includes an account of different aspects of volition, such as conscious decisions and intentions that are accompanied by different kinds of subjective sensation.

When consciousness is viewed as a result of dynamic processes of integration modules, it becomes difficult to set up a starting point of an intention, since any event occurring would be a result of previous activity causally relevant for the intention to be formed. A possible way to search for the neural correlates of free will is to view the neural origin of consciousness as originating from a *dynamic core* (Edelman, 2003; Tononi & Edelman, 1998), which is to look for neural activity related to functional phenomenological properties distributed across the brain, and describe the various connections in which they are integrated and differentiated, thereby generating the phenomenological experiences of volition. The dynamical approach states that all neural- and phenomenological events in principle could be influenced by prior events, thereby eliminating any causal starting point. An objection would be that it would be impossible to investigate the causal relation in intentional action, since the causal chain would be impossible to map. Strict determinism applies to the lower level (e.g. neural level) only. Through the causal relation could hypothetically be fully accounted for at lower levels, the enormous complexity of this level makes it impossible to describe the higher descriptive level properties as causal; it would even be metaphysically incorrect. What can be described is how these relations *influence* each other. Still influences could in principle be traced back as far as possible.

In order to scientifically investigate voluntary action, it is necessary to reduce the amount of possible data to minor temporal and spatial sequences, and therefore define a starting point. There is however a difference between setting a pragmatic starting point (e.g. a ‘go’-signal in an experimental paradigm) and claiming a functionalistic starting point in the causal role of agency. The criterion of setting such a pragmatic starting point is that the prior influences become so insignificant that they practically do not contribute to the unfolding of the event of investigation. As a result it can be argued that the processes responsible for conscious will are in fact not only neural mechanisms, but processes in the external world as well, since the subject’s intention toward external objects logically must be present in order for the subjective intention to arise (Keijzer & Schouten, 2006).

9 Conclusion

So far it is shown that free will can be translated from a metaphysical concept to a feature that can be investigated through scientific inquiry. Neuropsychology has neither disproven the individual autonomy, rational decision-making or moral responsibility, but actually broadened the concept compared to common understanding hereof. To seriously investigate what is captured by free will, the question is not to ask the all-or-nothing question *if* free will exists and trying to prove or disprove such. Leaving the all-or-nothing approach whether free will and moral responsibility exist, in favor of asking *what* are the mechanisms involved in autonomy and moral actions. This would actually contribute to the understanding of these phenomena. This error is mainly due to the concept of a metaphysically free (immaterial) conscious agent as a requirement for free will, which seems both unapproachable and implausible from a neuropsychological perspective. Even if the universe is causally closed and humans are nothing more than a “*billion tiny robots*” (Dennett, 2003) who just act in accordance with their nature independent of moral convictions, they still constitute a complex system and are capable of moral actions when viewed on a higher level of analysis. The “thread” to moral responsibility from neuropsychology is overrated, since neuropsychology cannot disprove that morality exists, but in fact contribute understanding the mechanism that gives rise to morality. When considering free will in moral responsibility the interesting part is not whether the subject is metaphysically free, but what mechanisms that in some cases weakens the subjects autonomy, explaining why such social phenomena as *group-thinking* (Janis, 1972) or the obedience shown in the Milgram-experiment where subjects act against their moral beliefs (Milgram, 1963), by explaining the underlying neural mechanisms giving rise to such behavior.

ARTIKEL II:
Neural correlates of Free Will separated from Action

Abstract

The experiments on free will by Benjamin Libet et al. showed that the reported times of intention to move occurred 350 ms after the neural activity associated with the movements had begun. This has often been interpreted as being proof of that conscious intention is an illusion. Libet et al. did however only focus on intention in immediate relation to movement. In the present study the subjects were to perform the same task, but instead of acting on the intentions immediately they had to delay performing the intended action. In this manner the intention to move was separated from the actual movement. The study used EEG to measure the neuroelectrical signal bereitschafpotential, as well as activity from other areas. The results showed that the signals did differ significantly when subjects were to plan the entire movement of pressing a key in advance, compared to not planning it. The results showed an ERP at the time where the prior intention was formed, as well as activity in frontal areas. The results indicate that sensormotoric areas are involved in planning of behavior as well as in executing the movements.

1 Introduction

Benjamin Libet's experiments on volition (Libet, Gleason, et al., 1983; Libet et al., 1982; Libet, Wright, et al., 1983), is without doubt the most influential experimental approach to free will. The conclusion of the experiment was that unconscious neural mechanisms initiate an action before the subject becomes conscious aware of the intention, suggesting that our everyday sense of being in control of our actions is nothing but an illusion.

In the experiment subjects were seated in a comfortable chair facing a circular screen on which the dot from a cathode oscilloscope rotated clockwise, resembling a clock with a rotation time at 2.56 seconds. The subjects were instructed to focus on a fixation point at the center of the clock, while monitoring the position of the dot on the clock. The subjects were instructed to either flex their hand or move their fingers spontaneously, as soon as they felt the urge to do so. The subjects were specifically instructed not to do any planning in advance regarding when the movement was to take place. After a movement had been performed the subjects had to report the time of their subjective experiences through introspection, by specifying the position of the dot on the clock at the time when they felt the urge to make the movement (W-time). Meanwhile the electrophysiological activity from the motor cortex was recorded by electroencephalography (EEG) and the movement of the hand (zero-time) recorded with electromyography (EMG). The subjective reported times were then compared to the bereitschaftpotential (BP), an electrophysiological event related potential which emerges from up to 1.5 s before of a self-paced movement usually measured over motor cortex (the Cz location following the international 10-20 system) (Kornhuber & Deecke, 1965). BP is usually divided into early BP, which is a slow decrease of the electrical potential, and late BP which is a more rapid fall of the electrical potential followed by a rapid discharge when the movement is executed (Shibasaki & Hallett, 2006). Libet discarded trials where early BP were found, arguing that it in 8 of 9 cases were associated with the subjects accidentally planning the movement in advance, and looked solely at late BPs (Libet, Gleason, et al., 1983). In Libet's experiment the average starting time for BP were approximately 550 ms before zero-time, whereas W-time in average was around 200 ms before zero-time. The BP thereby preceded the subjective experience of initiating the movement by about 350 ms, leading to the conclusion that unconscious processes initiate action, and that the conscious intention emerges after the action is in reality unconsciously initiated (Libet, 1985).

The timing of mental events and comparing of these to neurophysiological data in these types of experiments have for several reasons been criticized. The use of subjective reports based on the rotating spot on the clock has been criticized for biasing the subjective reports. Observing a moving spot is strikingly similar to tasks used to show the representational momentum; that moving objects tend to be perceived ahead of their actual location, in a temporal delay of approximately 70-100 ms (van de Grind, 2002; Joordens et al., 2002; Joordens, Spalek, Razmy, & van Duijn, 2004). Another critique regarding the timing of mental events is that mental events does not happen at an instance but over a period of hundreds of ms due to the integration of different sensory input, making timing of specific mental events problematic or even impossible (Danquah et al., 2008; Dennett & Kinsbourne, 1992; Gomes, 2002). Third, that the reported times are recalled post-action and therefore can be influenced by events that occur after the action, making them somewhat unreliable (Banks & Isham, 2008; Lau et al., 2007; Rigoni, Brass, & Sartori, 2010). A study measuring intention as they were formed, by interrupting the subjects at random times in a task similar to Libet's experiment and then asking them whether they had formed any intention, found that these were present at an estimated 1.42 s before movement (Matsuhashi & Hallett, 2008), more than 1 s before the reported times in Libet's experiments.

The use of BP in relation to volition also has some controversies. A study found that the time of conscious will did not correlate with the BP, but correlated with the lateralized readiness potential (LRP), a signal that appears as activity in motor cortex that get lateralized right before movement (Haggard & Eimer, 1999). Keller and Heckhausen (1990) found that the onset time of BP did not differ, even if the movement was performed unconsciously. This indicates that although BP is a strong indicator of emerging movement, its relation to conscious will is unclear. Since it is a signal measured from the scalp and most likely a result of several distinct underlying components.

The generation of voluntary movements is viewed as a result of a distributed system where the primary motor cortex (M1) receives signals from the basal ganglia and prefrontal cortex (PFC) through the supplementary motor area (SMA) and pre-supplementary motor area (pre-SMA), usually attributed to intentional aspects and an interaction of M1, the primary sensory areas (S1), parietal cortex and the intraparietal sulcus as responsible for object oriented actions and sense of ownership (Haggard, 2008, 2009). Out of these areas are contributors to BP activity bilateral from primary sensorimotor area (SM1 and M1) and the medial frontal cortex, particular SMA (Toma & Hallett, 2003). The BP might therefore not capture the entire process of voluntary action; especially the frontal areas involved in decision.

In Libet's experiment conscious will was viewed as the immediate conscious sensation of initiating or intending a motoric movement in accordance with performing the movement. This is however not all there is to the concept of free will. Intentions are not always formed in immediate relation to the action, but can be formed in advance anytime before the action is performed. Mele (2010) distinguishes between *distal* intentions about things to do later, and *proximal* intentions related to the immediate action. What was studied in Libet-like experiments is only the proximal intention. Thus the critique is that the intention to move the hand was not formed at the W-time, but prior to the experiment when the instructions to move the hand was given and W-time only reflecting implementation of previously formed distal intentions (Mele, 2010; Zhu, 2003).

The purpose of the current study is to examine the neural activity related to intention separated from the actual movement, thereby making it a distal intention in Mele's conceptualization. The neural activity was measured during a task where the subjects had to plan their movement in advance, in a task similar to the one in Libet's experiment. This is somewhat similar to *veto*-tasks in Libet-style experiments (Kühn et al., 2009; Libet, Wright, et al., 1983), but instead of inhibiting the action it is postponed, thereby not terminating the intention, but explicitly forming a distal intention. Instead of instantly moving, when they experience an intention to do so, the subjects wait one rotation on the clock to move until executing the intended movement. The distal intention is then always approximately 2.65 s before the actual movement, although this time could vary about 100 ms due to the previous mentioned bias in the use of the clock. It is not the main goal to compare subjective and neurophysiological data. The clock is used primarily to standardize the time where the intention is formed by the subject. How exactly the neural activity would differ when forming prior intention compared to only proximal intention has, to the author's knowledge, not been investigated previously in relation to this kind of task. Therefore no expectation regarding how the neural activity would differ was given, other that somehow differences would be present. Since the relation between RP and intention is unclear, the way this signal responds at the time of forming prior intention is of interest. If no BP is present then it must be concluded that it is only a reflection of motor-preparation and not related to intention.

Since the W-time might only reflect implementation of previous formed intentions, it is of interest to investigate how deliberate the proximal intention formed at this time is. It has been shown that the perceived time of events is shifted in time depending on whether the subject was the initiator of action or externally triggered, where it is perceived as occurring earlier when a sense of ownership toward an action is present (Ebert & Wegner, 2010). If the action in the Libet experiment is mostly automatic, the sense of ownership might decrease, and thereby also shift the perceived time of intention closer to movement than it actually was. The W-time might therefore be different as a result of how automated the movement is perceived. This is a secondary objective of the present study.

2 Method

2.1 Participants

9 healthy right-handed subjects (6 females and 3 males, aged 21-69 years, mean-age = 33.33 years, SD = 6.48) volunteered to participate in the experiment. Participants volunteered on behalf of notes posted at the Department of Psychology at Aarhus University and Department of Communication and Psychology at Aalborg University. The participants gave their informed consent prior to participation after being informed of purpose and procedure of the experiment. The study was approved by the Ethical Committee of Central Denmark Region and in accordance with the Helsinki Declaration. Participation in the experiment took approximately 2 hours and 15 minutes per subject, including mounting the EEG equipment. No financial compensation was given for participating.

2.2 Materials

The experimental setting was similar to the experimental paradigm developed by Libet and colleagues (Libet et al., 1982) with the additions used by Haggard and Eimer (1999), where the subjects were to choose between both left and right hand. Measurement of movement were done by the subjects pressing one of two keys on a standard PC-keyboard, with either left index finger (“z”-key) or right index finger (“.”-key).

The stimulus consisted of a clock-face similar to the clock described in Libets original study (Libet et al., 1982). The clock was programmed and ran in E-prime v. 2.0. It consisted of a white circle on a black background, on which a white dot was moving with 2.56 s per rotation. Each “5 min” on the clock were marked by numbers and each “2.5 min were marked with smaller lines as if it was a real clock. Subjects sat approximately 120 cm from the screen and the diameter of the clock was 15 cm. A fixation mark at the center of the clock indicated where the subject had to fix their gaze during trials. Before the experiment the participants were instructed to find the seating position they found most comfortable while facing the screen of the stimulus computer. Subjects were instructed not to follow the dot with their gaze, but focus on the fixation mark in order to minimize noise in EEG due to eye movement, and only noting the position of the dot for later report. The subjects were told to proceed through the experiment at their own pace, and each trial was started by the subjects themselves, when they felt ready by pressing the SPACE key on the keyboard. Each trial began with the dot appearing at the 12 o’clock position on the clock, indicating that participants had to relax, blink if necessary within the next 1-3 seconds, and fix the gaze at the fixation point. The subjects were told that they should try not to blink from when the clock started running and until they had pressed a key. When the computer registered a key press, E-prime would register the location of the spot. After a key-press the spot would continue to move in a random time-interval of 500 - 800 ms and then disappear. Hereafter the participants had to press the SPACE key and the dot would reappear at the 12 o’clock position. The participants had to move the dot, using the arrow-keys, to the position on the clock that corresponded to the time at which they experienced certain mental events, depending on the procedure. In addition the participants had to rate their experience of the decision on a three point scale divided into *impulse*, *partly deliberate* and *fully deliberate*.

2.3 Procedure

The experiment is divided into three series labeled respectively A, B and C. All series had the same experimental setting but followed different procedures and had different instructions. The three series ran consecutively and the order of the series was the same for all subjects. The subjects were, before each series began, given a number of test-trials to become familiar with the task in the respective series. The test-trials continued until the experimenter believed that the subjects fully understood the instructions and where in fact performing the intended task. This was done by asking how the subject experienced the task after each test-trial. Because of the subjective nature of the task, the subjects were told to report if they accidently diverged from the instructions during the experiment.

2.3.1 Series A: Replication of the original experiment by Libet et al.

The procedure of the first series was a replication of Libet's original study (Libet et al., 1982), with the addition of using both hands. This series was used as a control-condition to which the following series are compared. Before each trial began, instructions on which hand the subject should use appeared on the screen; either left or right (fixed conditions) or if the subjects had to choose themselves which hand (free conditions). There was double the number of trials for choice-conditions in order to have approximately equal amounts of data from each hand in the free conditions as in the fixed conditions. The ratio between these conditions was left/right/choice: 15/15/30, giving a total amount of trials of 60[1]. The order of fixed or free conditions was randomized. The participants were told not to plan in advance which hand to use or when they were going to press the key. It was emphasized that the key-press had to be spontaneous. Participants were explicit told not to plan their movement in advance. There was not given specific instructions to follow any particular pattern of alternating between using right and left hand in the free condition. Participants were instructed to wait pressing a key until the dot had performed its first rotation. Participants were told to note the position of the dot on the clock at the time they felt the first *urge*[2] to make the key-press, and afterwards report this time.

2.3.2 Series B: Active pre-planning

Participants were told they had do a similar task as in the first series, but instead of pressing the keys as soon as they felt the urge to do so, they only had to note the position of the dot on the clock, and then wait until the clock had made one full rotation before performing the movement. When the dot then had made a full rotation, the subjects were to execute the movement when the dot passed the same location. Following the instructions the participant "spontaneously" planned the action about 2.56 s before executing the movement. This time will be labeled P-time. Similar to the first series, each trial began by showing instructions on which hand the subject should use; either left or right (fixed conditions) or had to choose themselves (free conditions). The ratio between these conditions was Left/right/choice:

---

[1] The first subject performed a total of 80 trials per condition due to an error in the script which ran the experiment, that later was fixed and the number of trials adjusted. All the anwers from this subject were used in the further analysis.

[2] The instructions were given in Danish, but kept as close as possible to the original formulation by Libet et al. (1982). The original uses of the wording "first urge to move" were in the present experiment translated into "første indskydelse ", literally meaning "first impulse". The semantic meaning of the instructions were however similar to the original study.

15/15/30, giving a total amount of trials of 60. The order was randomized within each series. When the subjects were to choose hand themselves, they were instructed to 'know' which hand to use when the first urge appeared, so that the entire action was clear at the planning time, so all they needed was to implement the movement. The subjects were afterwards to report the position of the dot, where they planned the movement. They were explicitly told to report the time of the planning, even though it was similar to the point where they executed the movement.

2.3.3 Series C: Active pre-planning – time only

Similar to series B the subjects had to plan the key-press one clock-rotation in advance while monitoring the clock, only this time they should not plan the entire movement at first, but only notice the time when they felt an urge to move, and postpone the decision of which hand to use until they were about to execute the movement. In this way only the time of movement was decided at P-time. No information on which hand to use was given in advance. Instead subjects were told to choose which hand to use in all trials right before the time when they had to execute the movement. A fixed condition where subjects knew in advance which hand to use would conceptually be the same condition as the fixed condition in series B. Subjects were explicitly told to report the first time of the planning and not the time they decided which hand to use.

2.4 Recordings

EEG was measured using a 32 channel EEG system (BrainVision actiCAP) placed on the scalp with electrode locations following the international 10-20 system. An electrode was placed at the tip of the nose of the subjects which was used as reference. Electrooculography (EOG) was measured by two electrodes next to- and below the left eye, to detect blinks and eye-movement. The data was collected with the matching software (BrainVision Recorder) as continuous data, with a sampling rate of 250 Hz. The stimulus computer was connected so the time of key-press (zero-time) was stored in the continuous EEG-data.

2.5 Analysis

2.5.1 Subjective and behavioral data

The subjective and behavioral data was analyzed using PASW statistics v. 18. Though the primary objective of the current study was not to compare the subjective reports to the neurophysiological, the difference between reported time and actual press was still calculated in order to be compared to the subjective rating, of how deliberate the voluntary action was perceived. This was analyzed by performing a one-way ANOVA.

Due to the uncertainty regarding the subjective timing of mental events, the reported time of where distal intention to move (P-time) is made, can vary on the scale of approximately +/- 100 ms or more. The assumption was that this decision occurred in a time-window of 200 ms around P-time, ranging from -2660 to –2460 ms. No analysis will be made on the subjective reported P-times. P-time is only used to standardize the time where the intention is formed and to indicate where EEG-data should be analyzed.

2.5.2 EEG-data signal processing

Due to an error in the script running the clock on the stimulus computer, series C was discontinued for the first subject and therefore missing in the following analysis.

The offline preprocessing of EEG-data was done using EEGLAB v. 9.0.4.4b (Delorme & Makeig, 2004) toolbox for MATLAB v. 7.11.0 (MathWorks Inc.). Continuous EEG data was epoched in a time-window ranging from 4000 ms before zero-time to 600 ms after zero-time. The data was filtered with a 40 Hz low-pass filter to remove systematic line noise. No high-pass filter was applied. Epochs were baseline-corrected by averaging the mean potential over the entire epoch. Non-stereotyped EEG artifacts were removed semi-automatically by rejecting epoch with extreme values exceeding +/- 150 μV, such as abrupt muscle artifacts, bad channels and line drift. Independent Component Analysis (ICA) (Makeig et al., 1996) was used to remove blinks, eye-movement and muscle noise from the data, by using ICA decomposition on the datasets for each subject. Components were identified as artifacts and rejected semi-automatically following the independent component identification tool ADJUST (Mognon, Jovicich, Bruzzone, & Buiatti, 2011), and the general recommendations for visually inspecting independent components (Jung et al., 2000).

The statistical analysis of EEG-data was performed in the EEG/MEG infrastructure in SPM8 (Litvak et al., 2011) running on MATLAB. The data for each subject was averaged for each series and condition (A-free, A-fixed, B-free, B-fixed and C). The comparison between conditions was performed by making a contrast-wave of either two conditions for comparison. The contrast-waves was converted into a statistical parametric map (SPM), turning the data into time-scalp maps interpolated from all 32 channels and thereafter smoothed. The scalp-map topographies from all 32 EEG-channels during all epochs were interpolated and converted into statistical parametric maps (SPM's) that were smoothed for analysis. The use of all 32 channels, instead of just activity from the Cz electrode where BP usually is measured (Shibasaki & Hallett, 2006), makes it possible to give a crude indication of the location of differences in activity across conditions. A one-sample t-test was then performed on the statistical parametric maps for each contrast image. The statistical tests were not corrected for family wise errors (FWE) that statistically appear in images containing a large number of voxels. Instead the default critical α ($p = 0.001$) for uncorrected comparison in SPM8 was used, which then manually was Bonferroni-corrected for multiple comparison. Comparison was performed over the entire epoch in all tests, though only activity before movement is of interest in the discussion. Since no a priori hypothesis regarding the direction of differences is present, the statistical tests were two-tailed and a F-contrast was applied to the images. Since all subjects were right-handed only data from right hand trials was used.

3 Results

A total of 15 trials were instantly discarded due to the subjects reporting that something went wrong or that they made an error in performing the task they were instructed to. P-time was calculated by subtracting the time of one clock rotation from zero-time. All times are here reported relative to zero-time.

3.1 Subjective and behavioural data

The mean time for the reported subjective urge to move in series A was -82.2 ms (SD = 256,1 ms). The mean reported time for series B was 29.4 ms (SD = 119.1) and for series C 25.4 ms (SD = 95.6). Due to the difference between the tasks in series A compared to B and C, the different results were expectable.

The subjective rating of the experience of the decision showed that over half of the ratings were in the impulse category (59.7 %) contra the fully deliberate category which only was used in 11.7 % of the cases. The variation between the categories in the reported times of intention (*Fully deliberate*: *M* = -40.1 ms, SD = 271.5; *Partly deliberate*: *M* = -106.6 ms, SD = 275.8; *Impulse*: *M* = -78.8 ms, SD = 242.4), were by the one-way ANOVA shown to be non-significant $F(2,535) = 1.580$, p = .207. Finding no significant difference between the categories, there was no reason to continue further analysis of the categories in relation to the EEG-data.

3.2 EEG-data

A total of 144 epochs (10.5 %) were rejected during the semi-automatic inspection of the data due to artifact exceeding the extreme values on either channel. The remaining epochs for all trials where right hand was used was averaged for each series and condition. The BP for the free conditions derived from grand averages of the activity at the Cz electrode for all subjects after artifact rejection by ICA is shown in *figure 1*.

3.2.1 Planning vs. instant movement

The SPM of the contrast between condition A and B (free hand conditions only) is shown in figure 2a. This shows the biggest difference in activity in mid-central areas with peak at -2476 ms ($F(1,8) = 114.91;\ p < 0.001$). This corresponds to motor cortex from where the BP is measured. In addition eas activity in frontal areas found about the same time with peaks at -2500 ms ($F(1,8) =\ 91.27;\ p < 0.001$), -2496 ms ($F(1,8) = 53.39;\ p < 0.001$) and -2496 ms ($F(1,8) = 42.0;\ p < 0.001$). No significant difference was found on the SPM of the contrast between the free hand conditions of A and C. The contrast between free hand conditions of B and C (fig. 2b) showed differences in activity about 3 s before movement in frontal areas with peak at -3048 ms ($F(1,7) = 159.11;\ p < 0.001$) and -3052 ($F(1,7) = 91.21;\ p < 0.001$); in right fontal regions around -2400 ms with peaks at -2380 ms ($F(1,7) = 201{,}00;\ p < 0.001$), and -2384 ms ($F(1,7) = 134.44;\ p < 0.001$). The largest cluster of activity was in the central areas with peaks at -224 ms ($F(1,7) = 130.81;\ p < 0.001$) and -228 ($F(1,7) = 95.87;\ p < 0.001$). Furthermore there was a significant difference about 20 ms after key-press around central to partial areas ($p < 0.001$).

3.2.2 Free hand vs. fixed hand conditions

The BP's for free hand and fixed hand conditions are shown in figures 3 and 4. The SPM of the contrast between free and fixed hand in the no planning condition (series A), showed a difference in frontal activity clustered around 1 s before movement peaking at -1008 ms ($F(1,8) = 73.15;\ p < 0.001$) and again around -500 ms in left frontal areas peaking at -536 ms ($F(1,8) = 114.92;\ p < 0.001$) and –520 ms ($F(1,8) = 50.16;\ p < 0.001$). A small difference was also found in occipital areas peaking at -616 ms ($F(1,8) = 73.32;\ p < 0.001$), though statistically belonging to another cluster than the frontal activity. The contrast between free and fixed conditions in series B did not show any significant difference around P-time where the planning supposedly took place, but showed two clusters of difference around -230 ms, one cluster in posterior frontal left-hemisphere peaking at -228 ms ($F(1,7) = 116.36;\ p < 0.001$) and -223 ms ($F(1,7) = 77.57;\ p < 0.001$), and the other cluster in the mid-frontal area peaking at -228 ms ($F(1,7) = 105.56; p < 0.001$). Right before the time of moving a significant difference in the parietal area was found peaking at -80 ms at ($F(1,7) = 44.77;\ p < 0.001$).

4 Discussion

The replication of Libet's original experiment showed activity in form of the BP, starting at approximately -0.5 s. The subjective reported times of intention were at -82.2 ms, which is later than the study by Libet. There is however a relatively large spread in the reported times, meaning only little interpretation should be derived from the W-time in this study. It was expected that when the movement was rated as impulsive, the intention would be reported later in time. Taking the perceptual shift as an indicator of the sense of ownership, then there is no differences in the sense of ownership between the different kinds of self-initiated actions. It could be that a conceptual differentiation between *fully deliberate* and *impulsive* movements is an invalid differentiation, when considering the phenomenology in self-generated movements. The categories could be meaningless to the subjects. Gallagher and Marcel (1999) have proposed that what subjects are conscious about during the performance of an action in everyday context is the highest functionally meaningful or semantic level of doing the given behaviour. The lower levels of conscious awareness might therefore not receive much conscious attention, such as the initiation of specific motoric actions that constitute the behaviour at a higher functional level. If this is the case, then the task of focusing on the awareness of intention to move a finger might be an unnatural task for the subjects. The large age-span of the subjects and the relatively large distribution in W-time could also be an explanation

4.1 What is a neural correlate of intention?

In the planning conditions the grand averages of the BP's showed during the period where the planning was expected to take place, an event related potential resembling a BP. The analysis showed that this difference around P-time was significant. The difference was primarily from activity in mid-central area, which fits with the areas conventionally viewed as responsible for movement and preparation of movement, such as pre-SMA, SMA and M1. This signal did however only differ significantly from the control condition, where subjects were to plan the entire movement in advance, and not in the condition where the decision regarding which hand to use was prolonged right up until the action was to be performed. The missing difference between the activities of series A and C might be attributed to the nature of the subjective tasks as they were very much alike. Both conditions required the subjects to decide which hand to use at the time right before the key-press. The only difference is the subjective *urge to move*, which were present around 82.2 ms before the key-press in condition A and one rotation before in condition C. It has been proposed that the *urge to move* in Libet-type experiments does not reflect intention, but should be understood as implementation of a previously formed intention (Zhu, 2003). The proximal intention before movement is the same, in the manner that subjects had to choose hand as part of forming this intention, with the only difference that the distal intention was explicitly formed one clock rotation before en condition C. There is only a minimal conceptual difference at the time right before movement in the subjective tasks of condition A and C. The missing significant difference might be due to the fact that intention formed at P-time is made on behalf of the time read of the clock, which the subjects were monitoring constantly in all conditions, and thereby generating the same activity that in relation to the present task is considered to be noise. This could make the difference in activity rather small and unable to meet the level of significance. It has previously been shown that the task of monitoring the clock in the Libet-paradigm can influence the BP (Miller, Shepherdson, & Trevena, 2010).

The difference in activity found in central brain areas, when the entire movement had to be planned in forming the distal intention and assumed to be related to motoric areas, indicates that activity similar to performing the intended action is involved in the forming of a distal intention toward motoric action. The BP cannot be concluded to be an expression of entirely unconscious action generation based on the present study. How BP relates to conscious intention is however still unanswered, since it is not possible to conclude where the conscious intention specifically is located in the brain. In addition to the activity at the central areas, differences in frontal areas were found as well around P-time. The activity around sensormotor cortex could be projecting signals to the frontal lobe, where the forming of intentions and choice selection usually is agreed to take place (Haggard, 2008). In a connectionistic framework consciousness in general is hypothesized to emerge from distributed areas in the brain, each related to specific processing of information and then integrated. An explanation to the activity in motorcortex could be that conscious wills toward performing a motoric movements involves activity in areas responsible for processing bodily information, since the subjects are conscious aware hereof. On the other hand it could be that the conscious awareness of intention arises from these frontal areas entirely, thereby making the BP from in the mid-areas, an unconscious preparation of action (as the name of the signal suggests). BP could be an automatic signal that accompanies movements, as a readiness state which helps execute the intended movements.

How the subjects exactly formed their intention at P-time, and which mechanisms that are involved is unanswered. Studies have shown that during imagining a movement a signal resembling BP occurs and triggers activity in primary motor areas (Caldara et al., 2004; Jankelowitz & Colebatch, 2002). In the present study the BP at P-time could be due to the subjects imagining the movement to be performed later. Whether this specific activity can be attributed to imagining the movement, or simply an altered level of awareness regarding the finger or hands, is it not possible to conclude anything about. Common for these interpretations is that involvement of motoric areas informing distal intentions makes sense from a functionalistic perspective. If the intention is viewed as a possible readiness state then the implementation of the planned action would be easier since the mortor-commands are already present. Abstract imagining of behaviour might reflect a complex behavioural mechanism that optimizes the organism (i.e. humans) for better performance. From an evolutionary perspective such an interpretation makes perfect sense.

4.2 Free vs. fixed

The difference between free and fixed conditions where movement was to be performed as soon as intention was formed around -0.5 s, which could correspond to areas usually describes as involved in selecting task-relevant information prior to performing an action (Brass & Haggard, 2010; Egner & Hirsch, 2005). The difference might therefore be an expression of selecting or choosing a specific hand in the free condition compared to recalling previous instructions in the fixed hand conditions. How the difference 1 s before movement should be interpreted is unclear, since the temporal location is odd, but it might represent the same as at -0.5 s.

The difference between free and fixed in the planning condition both showed the mini-BP around P-time, and there was no significant difference between free and fixed at this time. Since the subjects were instructed to have planned both time and which hand to use at P-time, the incorporation of the previously given instructions into a new distal intention might not differ from forming an entirely new distal intention enough to be measurable by the here used method. The difference in activity between free and fixed, showed difference in activity in frontal areas already 3 s before movement, about half a second earlier than the P-time. This is about the same time in advance as the difference between free and fixed conditions for condition A. The most distinct difference between these conditions was around -2380 ms in an area around the posterior frontal lobe in the right hemisphere. Further research could focus on how intentions are formed and the subjective content of the intention in reality differs, and investigating whether these differences are neuroanatomically distinguishable, using imaging methods to locate these differences.

5 Conclusion

How conscious will is related to brain processes is still unclear, but the mechanisms underlying different kinds of voluntary actions are slowly starting to be revealed. Much research has focused on the proximal intention, by using the Libet-paradigm to time conscious will. Conscious will is however more than just the immediate awareness or initiation of action. Distal intentions and planning of behavior is as much included in the concept of free will, but is seldom included when studying free will within the neurosciences. The present experiment showed an event related potential at the time of forming prior intentions to move a finger. This indicates that the forming of distal intentions is involving areas responsible for performing the action. The neural mechanisms involved in forming distal intentions might be similar to proximal intentions activity-wise, as they might involve specific information regarding bodily movements, and not just abstract higher-order thoughts. The forming of distal intention is however not limited to motorareas as a difference in frontal activity were observed as well. These results points in the direction, that the neural underpinnings of intention are much more complex than usually viewed. When studying the role of intention in relation to BP and brain activity in general, it calls for a further conceptualization of intention, in order to uncover of the neural mechanisms involved.

Figurer til artikel II

Figure 1

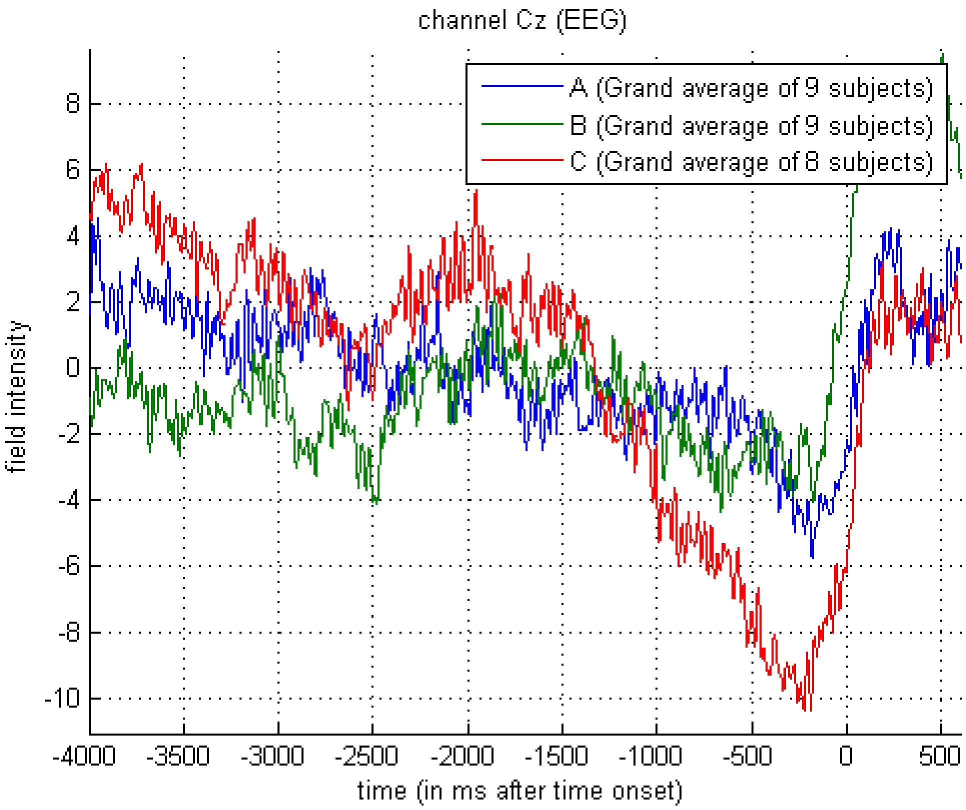

Figure 1: The bereitschaftpotentials of free-hand condition for all series, shown as the grand-average of activity at the Cz electrode from all subjects. The activity around -2560 ms (+/-200 ms) show a minor ERP when forming an intention to move later. Though only when the entire movement was planned were the signal at that time significantly different ($p < 0.000$ uncorrected).

Figure 2

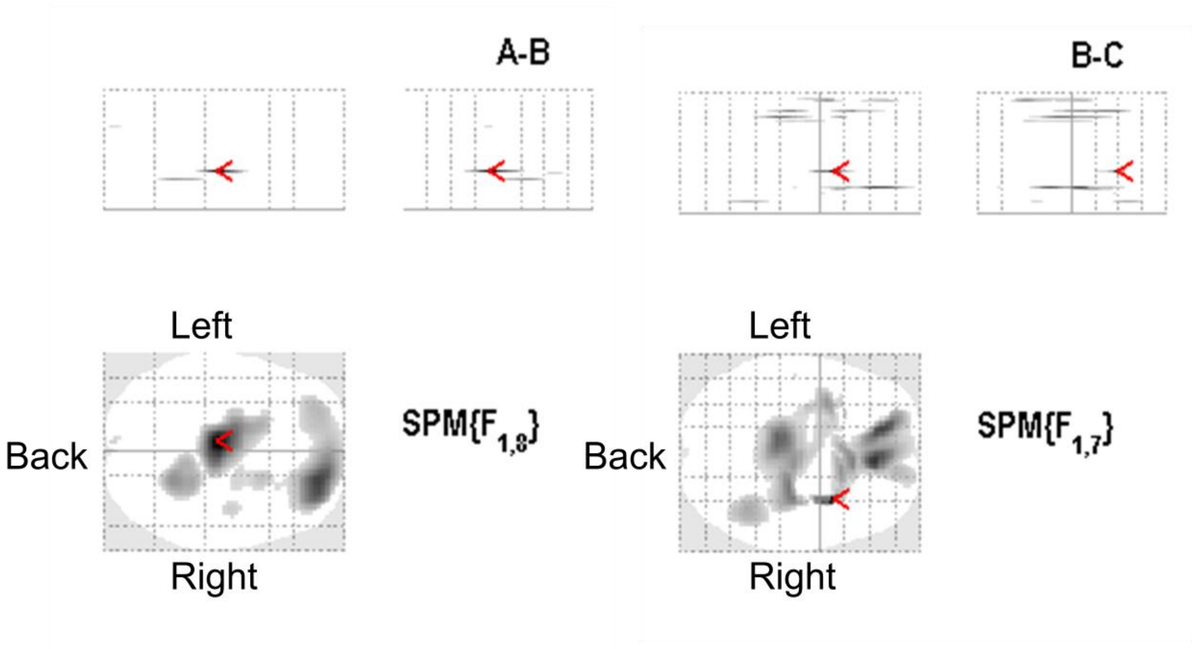

Figure 2: The SPM of the contrasts between condition A and B (left picture) show significantly ($p < 0.000$ uncorrected) activity around P-time at central areas as well as pre-frontal. The SPM of contrasts between condition B and C is shown to the right.

Figure 3

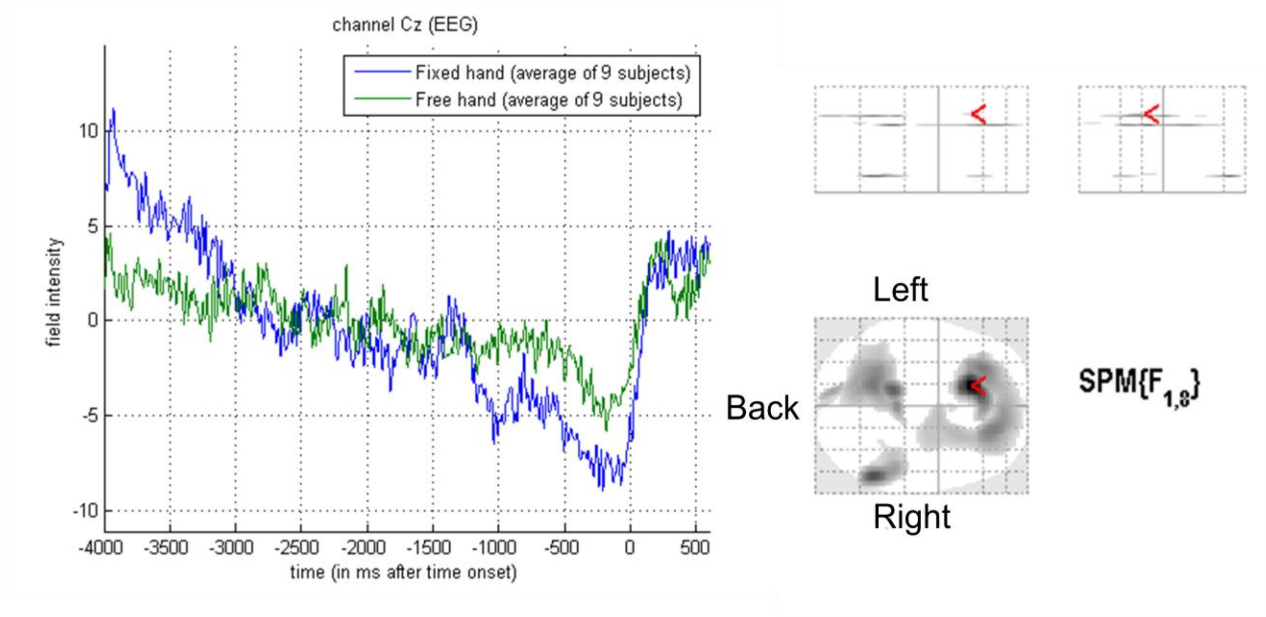

Figure 3: The grand-averages of the bereitschafpotential and the SPM of the contrast for the free/fixed conditions when no prior planning took place (series A).

Figure 4

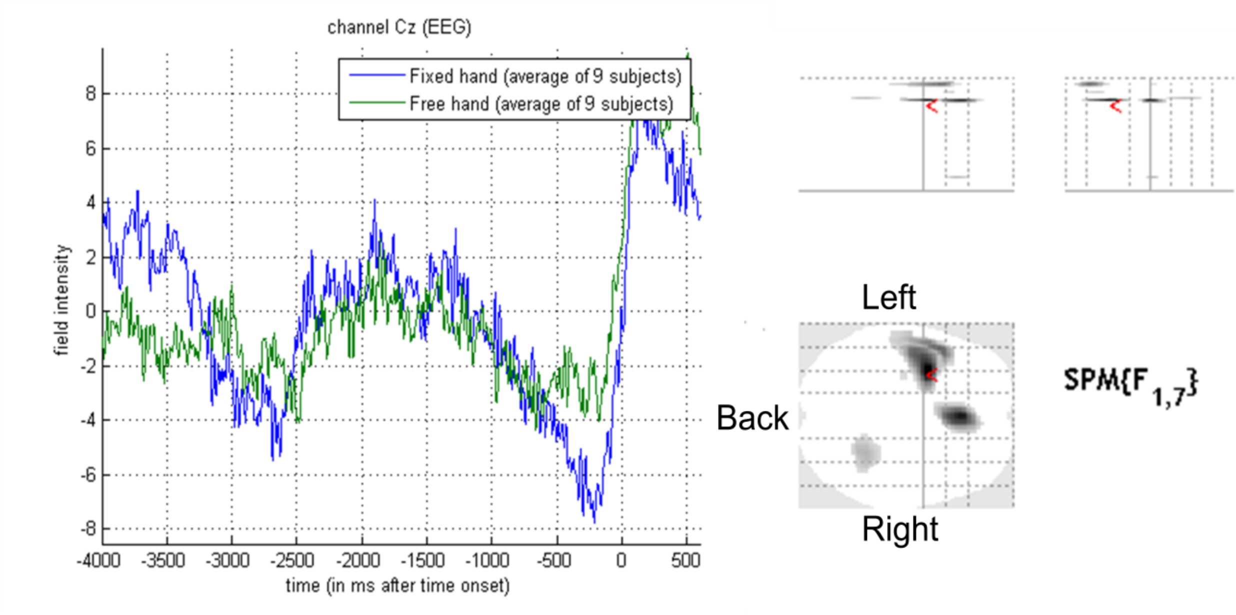

Figure 4: The grand-averages of the bereitschafpotential and the SPM of the contrast for the free/fixed conditions subjects had to form the intention to move one rotation on the clock in advance (series B).

Konklusion

De foregående to artikler lægger op til en mere nuanceret forståelse af begrebet fri vilje, ikke kun som undersøgelsesfelt for neuropsykologien, men i lige så høj grad inden for filosofien. For det første ved at skelne mellem hvilke begreber, som er metafysiske koncepter og hvilke som er praktisk forekommende fænomener, som kan studeres videnskabeligt. Dette er dog ikke en skelnen forstået som to forskellige domæner. Det er tværtimod et behov for yderligere konceptualisering af en række begreber i forhold til hvad den ontologiske baggrund for disse er. Et eksempel herpå er begrebet *intention*, som spiller en vigtig rolle i forståelsen af fri vilje. Ved at behandle intention som en singulær entitet ud fra en ontologisk forståelse, vil nuancer gå tabt. Forsøget i artikel II viste f.eks. forskellig aktivitet, alt efter hvordan instruktionen til at forme en forudgående (*distal*) intention blev givet. Da mentale fænomener blev defineret som nødvendigvis knyttet til neural aktivitet, må det derfor være to forskellige fænomener. I stedet er der blevet argumenteret for en dynamisk forståelse, ikke kun af intention, men af lignende mentale fænomener relateret til fri vilje. For kort at opsummere må *vilje* afhænge af et komplekst system, bestående af forskellige delkomponenter, som hver især bidrager til, hvad der normalt betragtes som fri vilje. Hvad disse komponenter er og hvordan de arbejder sammen er endnu ikke fuld ud forstået eller kortlagt af videnskaben. Artikel I skal ses som en opfordring til at stille spørgsmål om de koncepter, der normalt forbindes med fri vilje og ofte tages som absolut gældende (f.eks. påstanden at hvis et valg er influeret af udefrakommende påvirkninger, så har individet ingen kausal relevans), i realiteten burde betragtes som hypoteser frem for universelle facts. Den dynamiske forståelse har dog den bagside, at det at opstille meningsfulde konceptualiseringer i realiteten bliver umulig, da en komplet redegørelse for de involverede komponenter og mekanismer er praktisk umuligt. Antages den kompleksitet, som antydes i artikel I, vil en fuldstændig redegørelse for alle involverede mekanismer ikke være den praktiske løsning på at studere fri vilje. Det forklaringsmæssige niveau som de psykologiske videnskaber befinder sig på, bør derfor ikke udelukkende beskæftige sig med de minimale komponenters indbyrdes kausalitet, men hvordan ”højere-ordens” fænomener (konstitueret af lavere-ordens fænomener) interagere og påvirker hinanden. Det er derfor muligt at opstille meningsfulde kategorier, der godt nok inkludere utallige ontologiske lav-niveaus fænomener, ud fra det kriterium at kategorierne giver funktionel mening for det forklaringsmæssige niveau. Dette er selvfølgelig lettere sagt end gjort. Artikel II illustrer netop, at selvom intention blev operationaliseret på to forskellige måder, involverede begge former for intention aktivitet fra samme områder. Dette er imidlertid ikke grund til at afvise denne distinktion. Den konnektionistiske tilgang, forslået i artikel I, betyder ikke at en type funktionalitet er ekskluderende, for andre typer funktionalitet kan være tilstedet ved lignende aktivitet. Det er i de indbyrdes forbindelser mellem moduler at funktionaliteten opstår. Selvom der opstilles den hypotese at enkelte komponenter er ansvarlige for specifik bearbejdning af information og eventuelt forskellige fænomenologiske egenskaber, er det i den indbyrdes interaktion den funktionelle mening skabes.

Formålet med specialet har ikke været at svare på spørgsmålet, om vi har fri vilje eller ej. Ved at rekonceptualisere fri vilje har det dog ikke været muligt, andet end at være nød til at forholde sig til dette spørgsmål. Ud fra den foreslåede konceptualisering er svaret på dette spørgsmål *nej*! Vi er formegentlig altid under påvirkning af ubevidste processer, både fra interapsykisk og fra omverdenen, hvilket rent logisk bryder med frihedskravet. Sådan som fri vilje blev defineret i starten af artikel I, betyder det, at fri vilje ikke findes. Derimod argumenteres det, at fri vilje i den ultimative forstand, ikke er en nødvendighed sfor at afvise fatalisme. Så længe det giver mening at beskrive en organisme som værende *autonom*, er den metafysiske frihed ikke en nødvendighed. Selvom universet er kausalt lukket (omend dette ikke endegyldigt kan konkluderes) vil menneskelige (og andre organismers) handlinger have kausal indvirkning på omverdenen, og årsagen til disse handlinger vil kunne oprinde fra individet selv. I den filosofiske diskussion er det derfor vigtigt at skelne mellem determinisme og fatalisme. At erstatte frihedsbegrebet med autonomi vil gøre det muligt at studere de situationer, hvor det er ”interessant” at snakke om fri vilje. Fra et psykologisk perspektiv er det interessante ikke om fri vilje findes, men hvorfor nogle handlinger er *frier* end andre (Baumeister, 2008). Det centrale spørgsmål til studiet af fri vilje fra psykologien er hvilke faktorer, der influerer de beslutninger vi tager og hvilke mekanismer som giver autonomi. Det centrale spørgsmål kan formuleres som *hvor meget* frihed vi har, modsat at spøge om vi *har* frihed i den absolutte forståelse. Som en afsluttende kommentar på baggrund af det i denne opgave præsenterede, kan det konkluderes, at “[the] *reports of the death of human freedom have been greatly exaggerated*” (Roskies, 2010, p. 117).

Reference- og pensumliste

Pensum udgør de behandlede artikler/bøger i de fremstillede artikler. Tekster markeret med ”*” er ikke opgivet som pensum, da disse er opslagsbøger.

Anckarsater, H. (2010). Has biology disproved free will and moral responsibility? *Proceedings of the National Academy of Sciences*, *107*(28), E114-E114. doi:10.1073/pnas.1006466107

Bailey, A. (2006). Zombies, Epiphenomenalism, and Physicalist Theories of Consciousness. *Canadian Journal of Philosophy*, *36*(4), 481-509. doi:10.1353/cjp.2007.0000

Banks, W. P., & Isham, E. A. (2008). We Infer Rather Than Perceive the Moment We Decided to Act. *Psychological Science*, *20*(1), 17-21.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462-479.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*(6), 925-945. doi:10.1037//0033-2909.126.6.925

Bargh, J. A., & Morsella, E. (2008). The unconscious mind. *Perspectives on Psychological Science*, *3*(1), 73–79.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*(6), 1014-1027. doi:10.1037/0022-3514.81.6.1014

Baumeister, R. F. (2008). Free will in scientific psychology. *Perspectives on Psychological Science*, *3*(1), 14.

Baumeister, R. F., & Vonasch, A. J. (2011). Self-Organization as Conceptual Key to Understanding Free Will. *AJOB Neuroscience*, *2*(3), 44-46. doi:10.1080/21507740.2011.584949

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior*, *52*(2), 336-372. doi:10.1016/j.geb.2004.06.010

Bernstein, M. (2002). Fatalism. In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 65-81). Oxford University Press.

Blakemore, S. J., & Frith, C. D. (2003). Self-awareness and action. *Current Opinion in Neurobiology*, *13*(2), 219-224. doi:10.1016/S0959-4388(03)00043-6

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2000). Why can’t you tickle yourself?

Neuroreport, *11*(11), R11.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychological Medicine, 30*(05), 1131–1139.

Block, N. (2005). Two neural correlates of consciousness. *Trends in Cognitive Sciences*, *9*(2), 46–52.

Botvinick, M. M., & Cohen, J. D. (1998). Rubber hands' feel' touch that eyes see. *Nature*, *391*(6669), 756–756.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624-652. doi:10.1037/0033-295X.108.3.624

Brasil-Neto, J. P., Pascual-Leone, A., Valls-Solé, J., Cohen, L. G., & Hallett, M. (1992). Focal transcranial magnetic stimulation and response bias in a forced-choice task. *Journal of Neurology, Neurosurgery & Psychiatry*, *55*(10), 964-966.

Brass, M., & Haggard, P. (2007). To do or not to do: the neural signature of self-control. *Journal of Neuroscience*, *27*(34), 9141-9145.

Brass, M., & Haggard, P. (2010). The hidden side of intentional action: the role of the anterior insular cortex. *Brain Structure and Function, 214*(5-6), 603-610. doi:10.1007/s00429-010-0269-6

Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1707), 930-939. doi:10.1098/rspb.2010.2325

Caldara, R., Deiber, M.-P., Andrey, C., Michel, C. M., Thut, G., & Hauert, C.-A. (2004). Actual and mental motor preparation and execution: a spatiotemporal ERP study. *Experimental Brain Research*, *159*(3), 389-399. doi:10.1007/s00221-004-2101-0

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 367-379.

Cashmore, A. R. (2010a). Reply to Anckarsäter: A belief in free will is based on faith. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(28), E115.

Cashmore, A. R. (2010b). The Lucretian swerve: The biological basis of human behavior and the criminal justice system. *Proceedings of the National Academy of Sciences*, *107*(10), 4499-4504. doi:10.1073/pnas.0915161107

Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, *2*(3), 200-219.

Chalmers, D. J. (1996). *The conscious mind : in search of a fundamental theory*. New York: Oxford University Press.

Chalmers, D. J. (2000). What is a neural correlate of consciousness. In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 17–40).

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1485), 1585-1599.

Crump, M. J. C., & Logan, G. D. (2010). Hierarchical control and skilled typing: Evidence for word-level control over the execution of individual keystrokes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(6), 1369-1380. doi:10.1037/a0020696

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical transactions: Biological sciences*, *351*(1346), 1413–1420.

Danquah, A. N., Farrell, M. J., & O'Boyle, D. J. (2008). Biases in the subjective timing of perceptual events: Libet et al.(1983) revisited. *Consciousness and cognition*, *17*(3), 616–627.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., & Jeannerod, M. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, *65*(1), 71-86. doi:10.1016/S0010-0277(97)00039-5

Davidson, D. (1980). *Essays on actions and events*. New York: Oxford University Press.

Decety, J., Chaminade, T., Grezes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage*, *15*(1), 265–272.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9-21. doi:10.1016/j.jneumeth.2003.10.009

Dennett, D. C. (1991). *Consciousness explained*. London: Penguin London.

Dennett, D. C. (2003). *Freedom evolves*. London: Penguin.

Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer. *Behavioral and Brain Sciences*, *15*(2), 183–247.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement Intention After Parietal Cortex Stimulation in Humans. *Science, 324*(5928), 811-813. doi:10.1126/science.1169896

Dorval, A. D. (2006). The Rhythmic Consequences of Ion Channel Stochasticity. *The Neuroscientist*, *12*(5), 442-448. doi:10.1177/1073858406290793

Dowell, J. L. (2006). The Physical: Empirical, not Metaphysical. *Philosophical Studies, 131*(1), 25-60. doi:10.1007/s11098-005-5983-1

Ebert, J. P., & Wegner, D. M. (2010). Time warp: Authorship shapes the perceived timing of actions and events. *Consciousness and Cognition*, *19*(1), 481-489. doi:10.1016/j.concog.2009.10.002

Edelman, G. M. (2003). Naturalizing consciousness: a theoretical framework. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5520.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature neuroscience*, *8*(12), 1784–1790.

Ernst, M., & Paulus, M. (2005). Neurobiology of Decision Making: A Selective Review from a Neurocognitive and Clinical Perspective. *Biological Psychiatry, 58*(8), 597-604. doi:10.1016/j.biopsych.2005.06.004

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *Neuroimage*, *18*(2), 324–333.

Felsen, G., & Reiner, P. B. (2011). How the Neuroscience of Decision Making Informs Our Conception of Autonomy. *AJOB Neuroscience*, *2*(3), 3-14. doi:10.1080/21507740.2011.580489

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., d' Amato, T., & Jeannerod, M. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, *158*(3), 454-459.

Frankfurt, H. G. (1969). Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, *66*(23), 829-839. doi:10.2307/2023833

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, *68*(1), 5–20.

Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition. *Neuron*, *69*(3), 548-562. doi:10.1016/j.neuron.2010.11.045

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000a). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1404), 1771-1788. doi:10.1098/rstb.2000.0734

Frith, C. D. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition*, *14*(4), 752-770. doi:10.1016/j.concog.2005.04.002

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000b). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews*, *31*(2-3), 357–363.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *TRENDS in Neurosciences*, *4*(1), 14-21.

Gallagher, S., & Marcel, A. J. (1999). The self in contextualized action. *Journal of Consciousness Studies*, *6*(4), 4–30.

Gollwitzer, P. M. (1999). Implementation intentions: strong effects of simple plans. *American Psychologist*, *54*(7), 493-503.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. *Advances in Experimental Social Psychology*, *38*, 69-119. doi:10.1016/S0065-2601(06)38002-1

Gomes, G. (2002). The interpretation of Libet's results on the timing of conscious events: a commentary. *Consciousness and Cognition*, *11*(2), 221-230; discussion 308-313, 314-325.

van de Grind, W. (2002). Physical, Neural, and Mental Timing. *Consciousness and cognition*, *11*(2), 241–264.

Groppe, D. M., Makeig, S., & Kutas, M. (2008). Independent component analysis of event-related potentials. *Cognitive Science*, *6*, 1–44.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934–946.

Haggard, P. (2009). The Sources of Human Volition. *Science*, *324*(5928), 731-733. doi:10.1126/science.1173827

Haggard, P., & Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Consciousness and Cognition*, *12*(4), 695–707.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, *126*(1), 128–133.

Haggard, P., & Magno, E. (1999). Localising awareness of action with transcranial magnetic stimulation. *Experimental Brain Research*, *127*(1), 102–107.

Haggard, P., & Tsakiris, M. (2009). The Experience of Agency. *Current Directions in Psychological Science*, *18*(4), 242-246.

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382-385. doi:10.1038/nn827

Haggard, P., Newman, C., & Magno, E. (1999). On the perceived time of voluntary actions. *British Journal of Psychology*, *90*(2), 291–303.

Hameroff, S., & Penrose, R. (1996). Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and Computers in Simulation*, *40*(3-4), 453-480. doi:10.1016/0378-4754(96)80476-9

Hinsen, K. (2010). A scientific model for free will is impossible. *Proceedings of the National Academy of Sciences*, *107*(38), E149-E149. doi:10.1073/pnas.1010609107

Holmes, N. P., Snijders, H. J., & Spence, C. (2006). Reaching with alien limbs: Visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. *Perception & psychophysics*, *68*(4), 685-701.

Haans, A., IJsselsteijn, W. A., & de Kort, Y. A. W. (2008). The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb. *Body Image*, *5*(4), 389-394. doi:10.1016/j.bodyim.2008.04.003

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.* Oxford, England: Houghton Mifflin.

Jankelowitz, S. K., & Colebatch, J. G. (2002). Movement-related potentials associated with self-paced, cued and imagined arm movements. *Experimental Brain Research*, *147*(1), 98-107. doi:10.1007/s00221-002-1220-8

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S. J., & Passingham, R. E. (1994). Motor sequence learning: a study with positron emission tomography. *Journal of Neuroscience*, *14*(6), 3775-3790.

Joordens, S., van Duijn, M., & Spalek, T. M. (2002). When timing the mind one should also mind the timing: biases in the measurement of voluntary actions. *Consciousness and Cognition*, *11*(2), 231-240; discussion 308-313.

Joordens, S., Spalek, T. M., Razmy, S., & van Duijn, M. (2004). A Clockwork Orange:

Compensation opposing momentum in memory for location. *Memory & cognition*, *32*(1), 39.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKEOWN, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163–178.

Kane, R. (1996). *The significance of free will*. Oxford University Press.

Kane, R. (1999). Responsibility, luck, and chance: reflections on free will and indeterminism. *The Journal of philosophy*, *96*(5), 217–240.

*Kane, R. (2002). *The Oxford handbook of free will*. Oxford University Press.

Kaposy, C. (2010). The Supposed Obligation to Change One's Beliefs About Ethics Because of Discoveries in Neuroscience. *AJOB Neuroscience*, *1*(4), 23-30. doi:10.1080/21507740.2010.510820

Keijzer, F., & Schouten, M. (2006). Embedded Cognition and Mental Causation: Setting Empirical Bounds on Metaphysics. *Synthese*, *158*(1), 109-125. doi:10.1007/s11229-006-9053-9

Keller, I., & Heckhausen, H. (1990). Readiness potentials preceding spontaneous motor acts: voluntary vs. involuntary control. *Electroencephalography and clinical Neurophysiology*, *76*(4), 351–361.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton University Press.

Klemm, W. R. (2010). Free will debates: Simple experiments are not so simple. *Advances in Cognitive Psychology*, *6*(-1), 47-65. doi:10.2478/v10053-008-0076-2

Knoblich, G., & Prinz, W. (2001). Recognition of self-generated actions from kinematic displays of drawing. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY HUMAN PERCEPTION AND PERFORMANCE*, *27*(2), 456–465.

Knoblich, G., Seigerschmidt, E., Flach, R., & Prinz, W. (2002). Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception. *The Quarterly Journal of Experimental Psychology Section A*, *55*(3), 1027–1046.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv European Journal of Physiology*, *284*(1), 1–17.

Kühn, S., & Brass, M. (2009). Retrospective construction of the judgement of free choice. *Consciousness and cognition*, *18*(1), 12–21.

Kühn, S., Haggard, P., & Brass, M. (2009). Intentional inhibition: How the "veto-area" exerts control. *Human Brain Mapping*, *30*(9), 2834-2843. doi:10.1002/hbm.20711

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions. *Journal of Neuroscience*, *26*(27), 7265-7271.

Lau, H. C., Rogers, R. D., & Passingham, R. E. (2007). Manipulating the experienced onset of intention after action execution. *Journal of cognitive neuroscience*, *19*(1), 81–90.

Lau, H. C., Rogers, R. D., Haggard, P., & Passingham, R. E. (2004). Attention to Intention. *Science*, *303*(5661), 1208-1210. doi:10.1126/science.1090973

Leckman, J. F., Walker, D. E., & Cohen, D. J. (1993). Premonitory urges in Tourette's syndrome. *The American Journal of Psychiatry*, *150*(1), 98-102.

Libet, B. (1985). Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action. *Behavioral and Brain Sciences*, *8*(04), 529-566. doi:10.1017/S0140525X00044903

Libet, B. (1994). A testable field theory of mind-brain interaction. *Journal of Consciousness Studies*, *1*(1), 119–126.

Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies, 6*, *8*(9), 47–57.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, *106*(3), 623-642.

Libet, B., Wright, E. W., & Gleason, C. A. (1982). Readiness-potentials preceding unrestricted "spontaneous" vs. pre-planned voluntary acts. *Electroencephalography and Clinical Neurophysiology*, *54*(3), 322-335.

Libet, B., Wright, E. W., & Gleason, C. A. (1983). Preparation- or intention-to-act, in relation to pre-event potentials recorded at the vertex. *Electroencephalography and Clinical Neurophysiology*, *56*(4), 367-372.

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., et al. (2011). EEG and MEG Data Analysis in SPM8. *Computational Intelligence and Neuroscience*, *2011*, 1-32. doi:10.1155/2011/852961

Liu, X., Crump, M. J. C., & Logan, G. D. (2010). Do you know where your fingers have been? Explicit knowledge of the spatial layout of the keyboard in skilled typists. *Memory & Cognition*, *38*(4), 474-484. doi:10.3758/MC.38.4.474

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, *107*(3), 978-998. doi:10.1016/j.cognition.2007.12.004

Lorenz, E. (1993). *The essence of chaos*. Seattle: University of Washington Press.

*Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique* (1st ed.). The MIT Press.

Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*(5216), 1503-1506.

Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, 145–151.

*Martinich, A. P., & Sosa, D. (Eds.). (2006). *A Companion to Analytic Philosophy*. Oxford, UK: Blackwell Publishing Ltd.

Matsuhashi, M., & Hallett, M. (2008). The timing of the conscious intention to move. *European Journal of Neuroscience*, *28*(11), 2344–2351.

Maye, A., Hsieh, C.-hao, Sugihara, G., & Brembs, B. (2007). Order in Spontaneous Behavior. (M. Giurfa, Ed.)*PLoS ONE*, *2*(5), e443. doi:10.1371/journal.pone.0000443

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.

Mele, A. R. (2009). *Effective intentions : the power of conscious will*. Oxford ;;New York: Oxford University Press.

Mele, A. R. (2010). Testing Free Will. *Neuroethics*, *3*(2), 161-172. doi:10.1007/s12152-008-9027-3

Milgram, S. (1963). Behavioral Study of obedience. *The Journal of Abnormal and Social Psychology*, *67*(4), 371-378. doi:10.1037/h0040525

Miller, J., Shepherdson, P., & Trevena, J. (2010). Effects of Clock Monitoring on Electroencephalographic Activity: Is Unconscious Movement Initiation an Artifact of the Clock? *Psychological Science*, *22*(1), 103-109. doi:10.1177/0956797610391100

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229-240. doi:10.1111/j.1469-8986.2010.01061.x

Monroe, A. E., & Malle, B. F. (2010). From Uncaused Will to Conscious Choice: The Need to Study, Not Speculate About People's Folk Concept of Free Will. *Review of Philosophy and Psychology*, *1*(2), 211-224. doi:10.1007/s13164-009-0010-7

Montague, P. R. (2008). Free will. *Current Biology*, *18*(14), 584.

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2010). Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency. *Proceedings of the Royal Society B*, 1-7.

Moore, J. W., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition*, *17*(1), 136-144. doi:10.1016/j.concog.2006.12.004

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, *18*(4), 1056-1064. doi:10.1016/j.concog.2009.05.004

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., & Spence, C. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences*, *105*(35), 13169 -13173. doi:10.1073/pnas.0803768105

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4), 435-450.

Neisser, U. (1978). Memory: What Are the Important Questions? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 3-24). London: Academic Press.

Nichols, S. (2006). Folk Intuitions on Free Will. *Journal of Cognition and Culture*, *6*(1), 57-86. doi:10.1163/156853706776931385

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*(4), 663–685.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231-259.

Overgaard, M. (2006). Introspection in Science. *Consciousness and Cognition*, *15*(4), 629-633. doi:10.1016/j.concog.2006.10.004

Overgaard, M., & Mogensen, J. (2011). A Framework for the Study of Multiple Realizations:

The Importance of Levels of Analysis. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00079

Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179-217. doi:10.1016/j.cognition.2007.09.003

Penrose, R. (1989). *The emperor's new mind : concerning computers, minds, and the laws of physicsed*. Oxford: Oxford University Press.

Piccinini, G. (2010). The mind as neural software? Understanding functionalism, computationalism, and computational functionalism. *Philosophy and Phenomenological Research*, *81*(2), 269–311.

Ramachandran, V. S., & Rogers-Ramachandran, D. (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings: Biological Sciences*, *263*(1369), 377–386.

Rigoni, D., Brass, M., & Sartori, G. (2010). Post-action determinants of the reported time of conscious intentions. *Frontiers in Human Neuroscience*, *4*, 1-9.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*(5), 419–421.

Roskies, A. L. (2006). Neuroscientific challenges to free will and responsibility. *Trends in cognitive sciences*, *10*(9), 419–423.

Roskies, A. L. (2010). How Does Neuroscience Affect Our Conception of Volition? *Annual Review of Neuroscience*, *33*(1), 109-130. doi:10.1146/annurev-neuro-060909-153151

Rowe, J., Friston, K., Frackowiak, R. S. J., & Passingham, R. E. (2002). Attention to Action: Specific Modulation of Corticocortical Interactions in Humans. *NeuroImage*, *17*(2), 988-998. doi:10.1006/nimg.2002.1156

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nature Neuroscience*, *4*(5), 546–550.

Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of cognitive neuroscience*, *16*(6), 988–999.

Sarkissian, H., Chatterjee, A., De Brigard, F., Knobe, J., Nichols, S., & Sirker, S. (2010). Is Belief in Free Will a Cultural Universal? *Mind & Language*, *25*(3), 346–358.

Schaefer, M., Flor, H., Heinze, H.-J., & Rotte, M. (2007). Morphing the body: Illusory feeling of an elongated arm affects somatosensory homunculus. *NeuroImage*, *36*(3), 700-705. doi:10.1016/j.neuroimage.2007.03.046

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, *88*(3), B11–B21.

Seife, C. (2000). Cold Numbers Unmake the Quantum Mind. *Science*, *287*(5454), 791-791. doi:10.1126/science.287.5454.791

Shariff, A. F., Schooler, J. W., & Vohs, K. D. (2008). The Hazards of Claiming to Have Solved the Hard Problem of Free Will. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free?: psychology and free will* (pp. 181-204).

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, *117*(11), 2341–2356.

Shotter, J. (2003). `Real Presences': Meaning as Living Movement in a Participatory World. *Theory & Psychology*, *13*(4), 435-468. doi:10.1177/09593543030134001

Sie, M., & Wouters, A. (2010). The BCN Challenge to Compatibilist Free Will and Personal Responsibility, *3*(2), 121-133. doi:10.1007/s12152-009-9054-8

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., & Haggard, P. (2003). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience*, *7*(1), 80-84. doi:10.1038/nn1160

Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N., & Jeannerod, M. (1999). Perception of self-generated movement following left parietal lesion. *Brain*, *122*(10), 1867 -1874. doi:10.1093/brain/122.10.1867

Skewes, J. C., & Hooker, C. A. (2009). Bio-agency and the problem of action. *Biology and Philosophy*, *24*(3), 283–300.

Smart, J. J. C. (1959). Sensations and Brain Processes. *The Philosophical Review*, *68*(2), 141-156. doi:10.2307/2182164

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*(5), 543-545. doi:10.1038/nn.2112

Toma, K., & Hallett, M. (2003). Generators of the movement-related cortical potentials and dipole source analysis. In M. Jahanshahi & M. Hallett (Eds.), *The Bereitschaftspotential: movement-related cortical potentials* (pp. 113-130). Springer.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science*, *282*(5395), 1846-1851.

Tsakiris, M., & Haggard, P. (2005a). The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology*, *31*(1), 80–91.

Tsakiris, M., & Haggard, P. (2005b). Experimenting with the acting self. *Cognitive Neuropsychology*, *22*(3), 387–407.

Tsuda, I. (2001). Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences*, *24*(05), 793–810.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Urbaniok, F., Laubacher, A., Hardegger, J., Rossegger, A., Endrass, J., & Moskvitin, K. (2011). Neurobiological Determinism: Human Freedom of Choice and Criminal Responsibility. *International Journal of Offender Therapy and Comparative Criminology*. doi:10.1177/0306624X10395474

Wegner, D. M. (2003a). The mind's best trick: how we experience conscious will. *Trends in Cognitive Sciences*, *7*(2), 65-69. doi:10.1016/S1364-6613(03)00002-0

Wegner, D. M. (2003b). *The illusion of conscious will*. MIT Press.

Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480-492. doi:10.1037/0003-066X.54.7.480

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Sciences*, *7*(5), 214-218. doi:10.1016/S1364-6613(03)00081-0

Zhu, J. (2003). Reclaiming volition: An alternative interpretation of Libets experiment. *Journal of Consciousness Studies*, *10*(11), 61–77.

Zhu, J. (2004). Intention and Volition. *Canadian Journal of Philosophy*, *34*(2), 175-193.

Aarts, H., & van den Bos, K. (2011). On the Foundations of Beliefs in Free Will: Intentional Binding and Unconscious Priming in Self-Agency. *Psychological Science*, *22*(4), 532-537. doi:10.1177/0956797611399294

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition*, *14*(3), 439-458. doi:10.1016/j.concog.2004.11.001