

Philosophy of Mind = (analytische) wijsgerige antropologie; betreft de natuur van personen en onze essentiële kenmerken: de geest en de wil. De persoon is in de huidige opvatting een zelf (ook al is dit anders geweest in de geschiedenis). Deze opvatting komt in een spanningsrelatie te staan met het mechanistische wereldbeeld. De geest wordt tegenwoordig ook vaak fysicalistisch opgevat, waardoor we een manier moeten vinden om het subject van de wereldvisie in de wereldvisie zelf te incorporeren.

College 1 – Identiteit I

Locke definieert personen als volgt: “a thinking intelligent being, that has reason and reflection, and can consider itself as it self, the same thing, in different times and places”

Twee centrale vragen in het identiteitsprobleem:

- Wat zijn personen? (ontologisch)
- Wat is het één en dezelfde persoon te zijn doorheen de tijd?

LICHAAMSTHEORIE: “persoon x op tijdstip t is identiek met persoon y op tijdstip t*, asa, x en y één en hetzelfde lichaam hebben.”

Wat is dan een lichaam? Twee opties:

1. Een macroscopisch materieel ding [zeer breed]
2. Een biologisch organisme [ook anorganische delen lichaam mogelijk; biologisch \approx organisch]
3. Een machine (synthese van 1 en 2)

Levert twee opvattingen over identiteit op (identiteitsconcept 1 = 3)

1. **Materieel-ding identiteit:** “materieel ding x op tijdstip t is identiek met materieel ding y op tijdstip t*, asa, x en y spatio-temporeel continu zijn”
 - Elk materieel ding vertrekkend van het nu trekt een uniek, ononderbroken pad door ruimte en tijd zolang het bestaat.
 - Puzzelgeval: *schip van Theseus* (twee hedendaagse schepen zijn numeriek identiek met het origineel uit het verleden)
2. **Organisme-identiteit:** “organisme x op tijdstip t is identiek met organisme y op tijdstip t*, asa, x en y één en hetzelfde (biologische) leven hebben.”
 - Vereist geen (partiele) materiele gelijkheid, graduele verandering mag
 - Vereist stabiliteit in vorm, organisatie en levensfuncties

Simpele objecties:

1. Personen zijn meer dan hun lichamen, want personen zijn speciaal
 - ⇒ Repliek: de theorie dat we identiek zijn met lichamen stemt overeen met, en geeft een verklaring voor de manier waarop we feitelijk over onszelf denken (“waar mijn lichaam is, daar ben ik”)
2. Indien de liefde van persoon x gericht zou zijn op het lichaam van y, dan houdt x niet van een token-persoon maar van een type-persoon.
 - ⇒ Repliek: men zou hetzelfde kunnen zeggen als men deze objectie zou invullen voor een karakter, persoonlijkheid of een cluster aan karaktertrekken, namelijk dat je dan ook een zeer afgezwakte versie van een persoon krijgt. Williams zal ook benadrukken dat het lichaam noodzakelijk is om de identiteit van de token-persoon te bepalen, waar spatio-temporele continuïteit voor nodig is.
3. In ons normale taalgebruik geldt “Jan” \neq “Jan’s lichaam”.
 - ⇒ Repliek: uit louter linguïstische conventies kunnen we geen metafysische conclusies trekken.

Hoofdobjectie: ‘**Change of body**’-argument. Uitgangspunt: gedachte-experiment van geestelijke overdracht. Twee mannen, Brown en Robinson, worden geopereerd aan hun hersenen, waarna de dokters per ongeluk hun breinen verwisselen. Brown overlijdt, en Robinson leeft (met Brown’s brein).

“Brownson” gedraagt zich volledig als Brown, en heeft alle (mentale) eigenschappen die voor de operatie Brown karakteriseerden. Intuïtief zouden we dan zeggen dat Brownson, ondanks dat hij het lichaam van Robinson heeft, eigenlijk Brown is; maar als we lichaamsidentiteit zouden veronderstellen, dan zouden we zeggen dat hij nog steeds Robinson is.

⇒ *Brein Transplantatie Argument (BTA) [twee versies; tegen lichaamsidentiteit I en II]*

BTA-I (samengevat)

- (1) Het Brown-Brownson geval ontlokt de intuïtie dat (i) Brownson dezelfde persoon is als Brown en (ii) Brownson niet dezelfde persoon is als Robinson. Allicht heeft dit te maken met mentale continuïteit.
- (2) Indien lichaamsidentiteit voldoende was voor persoonsidentiteit, dan zou Brownson dezelfde persoon zijn als Robinson (want spatio-temporele continuïteit). Intuïtie ii ontkent dit consequent, dus geen voldoende voorwaarde.
- (3) Indien lichaamsidentiteit noodzakelijk was voor persoonsidentiteit, dan zou Brownson niet dezelfde persoon als Brown kunnen zijn (want geen spatio-temporele continuïteit). Intuïtie i ontkent dit consequent, dus geen noodzakelijke voorwaarde.
⇒ Dit argument is formeel geldig, maar is het ook solide?

Objectie via alternatieve beschrijving: Robinson leeft na de operatie verder als Brownson met het psychisch profiel van Brown.

- ⇒ Repliek (Williams): geestverwisseling veronderstelt dat persoonlijkheden geïdentificeerd kunnen worden zonder referentie aan lichamelijke trekken. Daarnaast zijn er implicaties van geestverwisselingen die we wellicht niet kunnen overzien, waardoor identificatie (door de omgeving) niet vanzelfsprekend is. (???)
- ⇒ Repliek op repliek (Quinton): Williams' bedenkingen zijn irrelevant: “if the character and memories in their new bodily surroundings really survived intact those closely concerned with them would say that the two switched bodies; that the original persons were where the characters and memories were”
- ⇒ Uiteindelijke repliek: de alternatieve beschrijving is implausibel, want het laat het cruciale punt van breintransplantatie weg (en vervangt het door een psychisch profiel). BTA-II zal

BTA-II (samengevat)

- (1) It is conceptually possible to transplant B_B of organism O_B into brainless O_R
- (2) B_B sustains a subject of experience with a characteristic psychological profile
- (3) Intuitively, post-operative O_R sustains the selfsame psychological profile with pre-operative O_B of which B_B was apart
- (4) $O_R \neq O_B$
- (5) Animalism can't be correct

MEMORY THEORY (MT): “Person x today is one and the same as person y at a particular moment in the past, if and only if, there exists memory connectedness between x and y.” – John Locke

Door middel van het geheugen hebben we een geprivilegieerde toegang tot onze eigen identiteit, want door middel van (experientieel) geheugen kunnen we teruggrijpen naar een gedachte op handeling in het verleden (een ervaring). Deze vorm van geheugen moet onderscheiden worden van factueel (herinneren “dat”) en habitueel (herinneren “hoe”), want dit zijn onpersoonlijke herinneringen. Het herinneren van ervaringen is het herinneren dat je die ervaringen had. Experientieel geheugen constitueert identiteit dan ook doorheen de tijd: twee persoon-fases behoren tot eenzelfde persoonlijke geschiedenis, *asa*, de latere persoon-fase een herinnering bevat van het hebben van een ervaring in de eerdere persoon-fase.

Probleem van circulariteit: het concept van geheugen veronderstelt persoonsidentiteit; we kunnen immers slechts onze eigen ervaringen herinneren van binnen, niet de ervaringen van anderen.

- ⇒ Oplossing: introductie van quasi-geheugen, wat een louter causaal concept is. Het proces van ontwikkeling van dit concept wordt hieronder gereconstrueerd:
- i. Volgens het non-causale concept van herinneren kunnen we zeggen: “*I remember having experience e, if and only if, (1) I apparently remember having e and (2) I had e*”. Deze twee voorwaarden, waarvan de tweede de ‘witness condition’ wordt genoemd, zijn niet voldoende.
 - ii. Hiertoe voegen we een derde conditie toe: “*(3) my apparent memory of having experience e is caused in an appropriate way by my having experience e in the past*”. Deze causale versie van het geheugenconcept is nog circulair vanwege de persoonlijke voornaamwoorden → vervangen + q-memory introduceren.
 - iii. “*I q-remember having experience e, if and only if, (1) I apparently remember having experience e; (2) someone had experience e; and (3) my apparent memory of having experience e is caused in an appropriate way by someone having experience e in the past*”. Merk op: i.t.t. het normale concept van geheugen kunnen we van binnenuit andermans ervaringen herinneren. Normale (eigen) herinneringen zijn een speciaal geval van q-herinneringen.
 - iv. Dit leidt tot de onpersoonlijke ervaringsanalyse: “*a memory of a past experience e occurs, if and only if, (1*) an apparent memory of experience e occurs; (2*); an experience e occurred; (3*) the apparent memory is caused in an appropriate way by experience e which occurred in the past.*” Stap iii. volstaat voor het oplossen van het circulariteitsprobleem; Parfit zal de onpersoonlijke analyse in college 3 gebruiken.
 ⇒ Aanpassing MT(1): “...if there exists q-memory connectedness...”

Overige objecties:

- Actualiteit: personen zijn niet altijd in een actuele staat van van q-geheugen, en dan verdwijnt de identiteit samen met het bewustzijn (bv. slapen)
 ⇒ Aanpassing MT(2): “...if there exists possible q-memory connectedness...”
- Absurditeit: geheugens kunnen volledig verdwijnen, en zo zijn we alleen identisch met de persoonsfasen die we kunnen herinneren. Memory connectedness is dus niet transitief, i.t.t. de identiteitsrelatie. Dit probleem wordt opgelost door het onderscheid tussen *directe verbondenheid* en *indirecte verbondenheid*, waarvan de eerste niet transitief is maar de tweede wel. De indirecte relatie wordt ook wel mentale continuïteit genoemd.
 ⇒ Aanpassing MT(3): “... if there exists continuity of possible experiential q-memory...”

De geheugentheorie legt de basis voor een uitgebreidere theorie van mentale continuïteit, die in college 2 aan bod zal komen.

College 2 – Identiteit II

Twee soorten identiteit:

- *Kwalitatieve identiteit*: fysieke eigenschapsidentiteit
- *Numerieke identiteit*: persoonsidentiteit; zelfidentiteit (door de tijd heen)

Wet van Leibniz (temporeel geïndexeerd): “indien x en y numeriek identiek zijn, dan moet elke eigenschap van x op tijdstip t een eigenschap van y op datzelfde tijdstip zijn, en vice versa.”

Identiteit volgens de wet van Leibniz karakteriseert men als identiteit in *strikte en filosofische zin*, terwijl men de zwakkere versie die men vindt in het dagelijks taalgebruik identiteit in *losse en populaire zin* noemt.

Het dubbele probleem van de numerieke persoonsidentiteit:

1. Epistemologisch probleem: hoe kunnen we weten of een persoon numeriek identiek is?
2. Ontologisch probleem: wat maakt een persoon numeriek identiek?
 ⇒ De eerste vraag is van belang voor het dagelijkse identificatieproces, waar we voornamelijk op lichaamskenmerken kunnen afgaan, maar de tweede heeft betrekking tot de natuur van

de persoonsidentiteit, wat veel diepere consequenties heeft voor zelfinteresse, liefde en morele verantwoordelijkheid. Met betrekking tot deze vraag kan lichaamsidentiteit niet voldoende zijn, zoals we al gezien hebben a.d.h.v. de BTA-argumenten.

DE ‘EMPIRISCHE BUNDEL’-THEORIE: “persoon x op tijdstip t is identiek met persoon y op tijdstip t*, *asa*, persoon x op tijdstip t mentaal continu is met persoon y op tijdstip t*.”

Omdat het innerlijke zelf niet kan worden waargenomen door de externe zintuigen, moet men gebruik maken van het innerlijk zintuig (introspectie). Dit ‘geestesorg’ ziet niets meer dan een complex van ervaringen, ofwel de *mentale bundel* (niet enkelvoudig of ondeelbaar door de tijd heen).

Wat constitueert dan zelfidentiteit in empirisch perspectief?

Dubbele vraag naar de eenheid van de geest:

1. *Synchrone eenheid*: constitutie van de persoonsfase (eenheid van het bewustzijn)
2. *Diachrone eenheid*: constitutie van de persoonsgeschiedenis (eenheid van de bewustzijnsstroom); 2 vooronderstelt 1.

Constitutieve voorwaarden:

1. Reflexiviteit van het bewustzijn: meerdere ervaringen kunnen tegelijk object van één bewustzijnsstaat zijn
2. *Mentale continuïteit*: seriële mentale verbondenheid; beter dan geheugen: ook karakter, talenten, intenties, ambities, overtuigingen, verlangens, et cetera (indirect, bestaat in een serie van overlappende schakels van mentale verbondenheid)
⇒ Mentale continuïteit veronderstelt causale continuïteit en de oorzaak hiervan is het brein (in functionalistische zin, niet-reductionistische zin om breinidentiteit te voorkomen)

In het geval van de ‘mentale bundel’-theorie spreekt is er sprake van identiteit in losse en populaire zin. Twee redenen:

- Het empirische zelf is een deelbare entiteit. Dit komt doordat de relatie van mentale verbondenheid afhangt van de kracht van de relatie, wat er toe leidt dat persoonsidentiteit een gradueel concept wordt, ook al is mentale continuïteit een alles-of-niets relatie (is immers gebaseerd op verbondenheid).
- Door de graduele natuur is de zelfidentiteit verwant aan de objectidentiteit, wat conventioneel is, want men schrijft (linguïstisch) identiteit toe aan een gradueel veranderend object, wat niet verenigbaar is met identiteit in strikte zin.

De common sense visie van identiteit bestaat in de strikte en filosofische zin. In existentiële kwesties is het antwoord op de identiteitsvraag namelijk altijd absoluut ja of nee. Existentiële praktijken (bv. verantwoordelijkheid, liefde) zouden niet kunnen bestaan zonder perfecte identiteit. De ‘mentale bundel’-theorie is dus catastrofaal hiervoor (terwijl dit theoretisch niet zo lijkt op het eerste gezicht); de praktische gevolgen van deze opvatting zijn ernstig.

BREINTHEORIE: “persoon x op tijdstip t is identiek met persoon y op tijdstip t*, *asa*, het ding dat x’s brein is op tijdstip t spatio-temporeel continu is met het ding dat y’s brein is op tijdstip t*”.

Twee soorten van essentie:

- *Nominale essentie*: bepaald door observeerbare, externe eigenschappen
- *Reële essentie*: bepaald door intern constitutieve eigenschappen (ligt ten grondslag aan ^)

Volgens Mackie is de eenheid van de bewustzijnsstroom de nominale essentie en is het centrale zenuwstelsel de reële essentie. Er is hier geen sprake van volledige, maar partiele lichaamsidentiteit. Deze conclusie vloeit voort uit empirische wetenschappen (dus contingent) en met de common sense hebben we alleen maar toegang tot de nominale essentie. Als we Mackie dan verder interpreteren, dan ligt de a

posteriori eenheidsrelatie van hersencontinuïteit feitelijk ten grondslag aan de a priori eenheidsrelatie van causaal-functionele psychologische continuïteit.

Objectie: als gevolg is persoonsidentiteit een vorm van fysieke objectidentiteit, en kan dus net als de lichaamstheorie geen verklaring geven voor ons common sense concept van persoonsidentiteit.

‘METAFYSISCH EGO’-THEORIE: “persoon x op tijdstip t is identiek met persoon y op tijdstip t*, *asa*, persoon x op tijdstip t zielsubstantieel identiek is met persoon y op tijdstip t*”.

Empiristen en metafysici zijn het eens over de noodzakelijkheid van introspectie om het zelf te vinden, en dat dit zelf de essentiële constituent van de persoon is (dus niet het lichaam!). Metafysici stellen echter dat een persoon bestaat uit een materieel deel (lichaam) en een immaterieel deel (ziel), waarvan de tweede wordt opgevat als de essentie. Deze visie geeft steun aan het religieuze geloof in leven na de dood, maar ook aan de bevreemdende gedachte dat ik iemand anders had kunnen zijn (gezien de ziel afzonderlijk van het lichaam kan blijven voortbestaan).

De hoofdprestaties van de ‘metafysische ego’-theorie zijn:

- De ziel is ondeelbaar, dus geen graduele persoonsidentiteit (=> strikte zin)
- Geeft onderbouw aan de common sense visie op identiteit
- Geen conventionele fictie, maar ontologische realiteit

Kritiek op empirisme: constitutie van identiteit heeft niets te maken met mentale aspecten zoals mentale continuïteit (wel consumenten, maar geen producenten). Gelijktijdige en verschillende ervaringen worden synchroon en diachroon verenigd omdat één en dezelfde zielsubstantie alle ervaringen heeft. De eenheid van het empirische zelf wordt geconstitueerd door de identiteit van het metafysische zelf.

Problemen:

1. Regressie van ego’s: de synchrone individuatie en diachrone identiteit van het ego zijn niet te bewijzen door de zelfbewuste ervaring (je kan namelijk meerdere ego’s hebben). Om de eenheid te garanderen moet men beroep doen op een supra-ego, maar dan moet men hiervan ook weer de eenheid bewijzen.
2. Zielsubstanties zijn onderling niet te differentiëren want de essentie heeft niets te maken met fysieke/mentale kenmerken
3. Zielsubstanties zijn niet-observeerbaar (ook door introspectie)
⇒ Ontologische fictie

De ‘mentale bundel’-theorie leidt tot existentiële ficties en de ‘metafysische ego’-theorie is een ontologische fictie. Geen van beiden visies heeft een bevredigende kijk op zowel de natuur als het belang van de persoonsidentiteit. => aporie (=besluiteloosheid)

College 3 – Identiteit III

Is het identiteitsprobleem onoplosbaar? Parfit geeft een oplossing: een radicale versie van de ‘mentale bundel’-theorie.

ONPERSOONLIJKE THEORIE

Indien een persoon in essentie een bundel van ervaringen is, dan bestaat persoonsidentiteit in niets anders dan mentale continuïteit (reductie tot) en verbondenheid. Dit heeft twee belangrijke gevolgen:

- ⇒ Personen bestaan niet, want ze bestaan niet afgescheiden en verschillend van het bestaan van ervaringen. Ze behoren niet tot de ‘make-up’ van de werkelijkheid. Personen bestaan slechts in de taal.
- ⇒ Soms is persoonsidentiteit ongedetermineerd (*lege vraag*). Dit toont Parfit aan met het gedachte-experiment van ‘**het spectrum**’. Het gaat als volgt: Er is sprake van een gedeeltelijke (procentuele) vervanging van de cellen van persoon x voor die van persoon y. Als

(bijna) alles of (bijna) niets vervangen is lijkt het antwoord op de identiteitsvraag duidelijk, maar bij 50% vervanging kan men om het even welke kant kiezen. Deze vraag is leeg, want ondanks dat men alle empirische feiten kent, kan men de vraag nog steeds niet beantwoorden (er bestaat geen speciaal identiteitsfeit). Een antwoord veronderstelt een metafysische achtergrond. Metafysici zullen, ondanks dat we het antwoord niet kunnen kennen (epistemologisch), toch zeggen dat er een antwoord is op de identiteitsvraag (ontologisch).

De lege identiteitsvraag druist in tegen het common sense besef van het belang van de persoonsidentiteit voor het (existentiële) zelfbehoud. Volgens de reductionist kunnen de vragen naar het zelfbehoud en de persoonsidentiteit echter worden ontkoppeld. Dan heeft de identiteitsvraag naast geen inhoud ook geen belang meer. Dit doet hij aan de hand van het gedachte-experiment van ‘**de verdubbeling**’. Het gaat als volgt: net voordat Medius overlijdt worden er twee exacte replica’s van hem gemaakt, Sinister en Dexter. Beide replica’s staan in een relatie van mentale continuïteit tot Medius die de 100% benadert. Hier is dus sprake van een vertakking van identiteit (logisch is dit denkbaar).

⇒ De vraag: “is Sinister of Dexter numeriek identiek met Medius?” is volgens de reductionist een lege vraag, maar voor de metafysicus niet. Hij stelt dat er drie opties zijn:

1. Sinister is Medius \wedge Dexter is Medius
2. Sinister is Medius \vee Dexter is Medius
3. Medius sterft

De reductionist heeft bezwaren tegen elk van de drie opties:

1. Over de tijd is dit onhoudbaar, want Sinister en Dexter zullen elk steeds meer verschillende ervaringen hebben en zullen steeds meer verschillende personen zijn van Medius.
2. De mentale continuïteitsrelatie van om het even welke replica met Medius is identiek, dus deze disjunctie is arbitrair.
3. In dit geval betekent verdubbelen sterven, maar waarom zou een enkele overdracht wel overleving inhouden? Dit is implausibel, dus verdubbeling betekent dubbele, synchrone overleving.

Het derde antwoord van de metafysicus is implausibel als overleving identiteit veronderstelt, dus moeten we deze twee vragen losmaken. Omdat in beide gevallen er sprake is van mentale continuïteit, dus er is geen kwalitatief verschil, spreken we in het geval van verdubbeling ook van overleving, maar dan zonder identiteit. Er kan dus sprake zijn van zelfbehoud zonder strikte, enkelvoudige identiteit. Dit lost ook het probleem van het spectrum op: 50% van mij overleeft in het hybride wezen, zonder dat ik identiek met hem ben.

Persoonsidentiteit is ook niet van belang in enkele overlevingsgevallen, ook al spreken we wel van identiteit (dat dit het geval is). Dit is alleen het geval omdat we in de logica van de identiteitsrelatie spreken van een 1-op-1 relatie, maar dit is slechts een kwestie van de logische taal in het geval van niet-vertakte mentale continuïteit, wat wel een reële kwestie is.

Het absolute concept van de common sense persoonsidentiteit kan dus rationeel niet gerechtvaardigd worden, maar deze intuïtie kan wel causaal verklaard worden binnen het naturalisme. De zgn. ‘fictie van de persoonsidentiteit’ wordt door het reductionisme verklaard d.m.v.:

- (1) Rationele verklaring van het belang van mentale continuïteit: het garandeert de relaties van *project-verwerkelijking* en *identificatie*. In andere woorden, mijn toekomstige zelfinteresse is rationeel gerechtvaardigd, niet omdat dat zelf identiek zal zijn met mij, maar wel omdat het toekomstig zelf mijn projecten ten harte neemt en het mijn object van identificatie is (en dit zal zo zijn als ik mentaal continu ben met mijn toekomstig zelf).
- (2) Causale verklaring van het belang van persoonsidentiteit: het feit dat de common sense een intrinsieke waarde toekent aan persoonsidentiteit kan oorzakelijk verklaard worden door *gewoontevorming* en *sociale versterking*. Het is een gewoonte om persoonlijke identiteitstermen te gebruiken in plaats van mentale continuïteitstermen. In normale gevallen garandeert de identiteitstheorie (in afgeleide zin) ook projectverwerkelijking en identificatie, maar dit heeft geen intrinsieke waarde omdat de echte grond ligt in de mentale continuïteit (volledig). Daarnaast is het in sociale, praktische zin makkelijker om te spreken over een “ik” dan een “mentale bundel

van ervaringen die mentaal continu is". Dit "ik" is echter slechts formeel en niet essentieel, waarvoor het wel wordt aangezien.

Persoonsidentiteit in strikte zin moet dus vervangen worden door de graduele, zwakke versie. De existentiële attitudes en het zelfbeeld van de common sense zijn weerlegd en zullen gereviseerd moeten worden:

- Het unitaire zelfbeeld moet vervangen worden door een pluralistische; een opeenvolgende 'familie van zelden' die verbonden zijn door (graduele) mentale continuïteit. Deze manier van spreken is enigszins analoog aan het spreken over opeenvolgende generaties. Mijn toekomstige zelf (bv. als 70-jarige) en mijn verleden zelf (bv. bij geboorte) zijn bij benadering totaal verschillende zelden.
- Graduele '*disconto*'-principe: hoe zwakker een relatie is, hoe minder ze van belang is. Hoe meer tijd er tussen twee 'zelden' zit, hoe zwakker de relatie van mentale continuïteit wordt. Dit heeft morele, emotionele en rationele gevolgen. De 'discontovoet' kan ook verklaard worden in termen van project-realisatie en identificatie.

De aporie wordt bijgevolg een impasse: de keuze wordt onmogelijk gemaakt. De gewone man moet kiezen tussen het mirakel van de metafysische persoonsidentiteit, wat strookt met de intuïtie maar niet rechtvaardigbaar is, of het concept van persoonsidentiteit weggooien, met alle gevolgen van dien.

College 4 – Mind/Body Problem I

Folk psychology (FP) is de common sense strategie om gedrag te beschrijven, verklaren en voorspellen in intentionele termen – overtuigingen, verlangens en andere propositionele attitudes. M.a.w., we gebruiken mentale concepten om gedrag van anderen en onszelf inzichtelijk te maken (dus niet via wetenschap & experimenten).

Het 'mentaal vocabularium' van de FP bestaat uit drie categorieën:

1. Ervaring/sensatie; zelfbewust, hebben 'qualia'
2. *Propositionele attitude* (PA): een bepaalde mentale houding hebben t.o.v. een bepaalde propositie; intentioneel
3. *Handeling*: lichamelijk bewegen met een intentie

Hybride categorieën: percepties, emoties, herinneringen

Dichotomie van het mentale: sensaties zijn fenomenaal en propositionele attitudes zijn intentioneel.

Fenomenaliteit is het subjectieve, kwalitatieve karakter van sensaties ("hoe het is om iets te voelen"). Sensaties zijn intrinsiek bewuste gebeurtenissen en worden vaak opgevat als objecten van kennis door vertrouwdheid (acquaintance; als je een sensatie niet kan hebben kun je de natuur ervan niet kennen). Omdat sensaties subjectief en kwalitatief zijn, kan een beschrijving in objectieve, kwantitatieve termen hun natuur niet vatten (bv. aan een blinde uitleggen wat kleur is).

⇒ Wat is (*zelf*)bewustzijn?

- *Bewustzijn* is breed verbonden met perceptie, proprioceptie en introspectie
- *Introspectie/zelfbewustzijn* is reflexief of tweede-orde bewustzijn, gericht op bewustzijn zelf als inhoud
- *Qualia*: eigenschappen van bewuste ervaringen, die hen hun bepaalde kwalitatieve karakter geven ("what it's like to..."); hoe de sensatie aan het bewustzijn verschijnt; veronderstelt introspectie.

Intentionaliteit van een PA is haar gerichtheid op een object. Het is een criterium om het fysische te onderscheiden van het mentale: een fysisch object is niet gericht op iets. Sensaties zijn ook niet intentioneel. PA's zijn extrinsiek bewust; ze worden ook wel mentale disposities genoemd. Voortdurend bewustzijn is niet belangrijk voor hun voortbestaan. Ze hebben een representatieve i.p.v. een kwalitatieve inhoud.

Volgens Brentano is elk mentaal fenomeen gekarakteriseerd door de ‘intentionele (mentale) inexistentie’ van een object, m.a.w., een *immanente objectiviteit*. Belangrijk is om het onderscheid te maken tussen het object en de inhoud van het denken:

[overtuiging \Rightarrow intentioneel object (inhoud)] \rightarrow reëel object

Brentano’s theorie is een manier om non-existerende objecten van het denken mogelijk te maken, want zijn intentionele *in*-existentie staat toe dat iets mentaal bestaat, tussen reële existentie en non-existentie in. Men noemt deze objecten ook wel mentale representaties.

Wat is de relatie tussen PA’s en handelingen? FP zegt dat PA’s handelingen rationaliseren en veroorzaken. Maar wat is een handeling?

Handelen is actief of intentioneel bewegen. Maar intentioneel \neq intentionaliteit in deze context.

Er zijn 3 mogelijke interpretaties van intentioneel:

1. Gerichtheid op een toekomstige handeling (voornemen om te handelen)
2. Opzettelijkheid (vrij en bewust handelen)
3. Doelgerichtheid (handelen met de bedoeling om een stand van zaken te bewerkstelligen)

De focus zal liggen op interpretatie 3. Volgens Donald Davidson kan men intentioneel handelen gelijkstellen met “handelen met redenen” (want men kan vragen waarom iemand handelt). Redenen kunnen een handeling (op verschillende manieren) rationaliseren. De belangrijkste reden voor een handeling noemt hij de *primaire reden*. Deze reden bestaat uit een *pro-attitude* (wil) en een overtuiging. Beiden zijn PA’s.

Davidson’s model van intentioneel handelen (doelgericht) lijkt een causale verklaring voor menselijk gedrag uit te sluiten. Overtuigingen en verlangens die een handeling rationaliseren zijn geen causale antecedenten en daarnaast heeft de causale (neurofysiologische) verklaring slechts betrekking op lichamelijke bewegingen. Dit noemt men de antinomie tussen doelgericht en causaal verklaren, waardoor (volgens voorstanders) de intentionele psychologie onverenigbaar is met de natuurwetenschappen.

\Rightarrow Davidson’s repliek poogt om de antinomie af te zwakken: de causaliteit van PA’s op lichamelijke handelingen speelt wel degelijk een rol. Het hebben van een reden is niet genoeg om de effectiviteit van de handeling te begrijpen (waarom ik het ook écht deed), er moet ook een causale relatie tussen reden en handeling zijn (want alleen de effectieve reden verklaart ook dat ik de handeling deed; er hadden ook andere mogelijke rationaliserende redenen kunnen zijn).

Om een handeling te veroorzaken, moet er naast interne causaliteit ook sprake zijn van **mentale causaliteit**. Algemeen geformuleerd: “een lichaamsbeweging is intentioneel, indien ze op de juiste manier veroorzaakt wordt door pro-attitudes en overtuigingen die ze rationaliseren”. Mentale causaliteit is echter breder dan het geval van handelen; we kunnen drie soorten onderscheiden:

- i. Fysisch \rightarrow Mentaal (bv. perceptie)
- ii. Mentaal \rightarrow Mentaal
- iii. Mentaal \rightarrow Fysisch (handelen)

Het **geest-lichaam probleem** is een ontologisch probleem (“ontologisch” \approx “metafysisch”). Twee initiële karakterisaties:

- De *geest (mind)* is de verzameling van mentale toestanden en gebeurtenissen (“mentale eigenschappen”) waarnaar het mentaal vocabularium van de FP verwijst.
- Een *eigenschap* is een ontologisch element dat in de taal wordt gerepresenteerd door een *predikaat*, wat een logico-linguïstisch element is. Aan niet alle predikaten beantwoorden ook daadwerkelijk eigenschappen, maar andersom wel. Dit geldt ook voor het onderscheid tussen mentale eigenschappen en predikaten.
- Een *lichaam* (Descartes) is een macroscopisch materieel ding dat spatiaal uitgebreid en zintuiglijk observeerbaar is.

Personen zijn enerzijds biologisch-fysische wezens, maar anderzijds hebben ze blijkbaar ook een psychisch, mentaal of geestelijk leven. In andere woorden, de mens heeft een lichaam en een geest. Over het bestaan van fysische eigenschappen zijn we zeker, en ook redelijk zeker over het bestaan van mentale eigenschappen. Voorlopig gaan we ervan uit dat aan mentale predikaten ook eigenschappen beantwoorden. Dan kunnen we het volgende onderscheid maken:

- *P-predikaat*: psychische/mentale eigenschappen (Ψ)
- *M-predikaat*: fysische/materiele eigenschappen (Φ)

Lichaam en geest hebben een aantal verschillen, waardoor we uiteindelijk aan Ψ een speciaal karakter zullen toeschrijven, omdat ze alleen aan personen toekomen, i.t.t. Φ . De verschillen:

LICHAAM	GEEST
Spatiaal	Niet-spatiaal
Publiek observeerbaar	Private, geprivilegieerde toegang
Niet-bewust	Bewustzijn (qualia)
Niet-intentioneel	Intentioneel

De twee centrale (ontologische) vragen van het geest-lichaam probleem zijn dan:

- (OV1) Bestaan Ψ s? Hebben Ψ s existentie?
 (OV2) Wat is de essentie van Ψ s en de drager van Ψ s?

OV1 nemen we voorlopig voor waar aan, dus als we OV2 meer specificeren krijgen we het volgende:

- (GL') Zijn mentale eigenschappen, Ψ s, qua natuur, niet-fysische immateriële eigenschappen, (totaal) verschillend van fysische eigenschappen, Φ s?
 (GL*) Is het subject van mentale eigenschappen, Ψ s, qua natuur, een niet-fysische immateriële substantie, (totaal) verschillend van een fysische substantie die subject is van fysische eigenschappen, Φ s?

College 5 – Mind/Body Problem II

Het **CARTESIAANS INTERACTIONISTISCH DUALISME (CID)** ziet de geest als reëel onderscheiden van het lichaam. Er bestaat een mentale en een fysische wereld waar respectievelijk de mentale en fysische substanties verblijven. Het subject van mentale eigenschappen is een niet-fysische, immateriële substantie. Ofwel: de mentale en fysische eigenschappen bestaan los van elkaar. Het cartesianisme is ook een *interactionisme*, want er is een intieme, causale band tussen geest en lichaam. Deze interactie vindt plaats in de pijnappelklier (voor de rest zijn geest en lichaam volledig gescheiden). Er is dus sprake van mentale causaliteit

Argumenten vóór CID:

- Het is denkbaar (en dus logisch mogelijk) dat mijn geest (“ik”) besta zonder mijn lichaam, dus ben ik niet identiek met mijn lichaam.
- Mijn lichaam heeft deelbaarheid als eigenschap, en mijn geest niet (dus geest \neq lichaam)
- Mijn lichaam heeft de eigenschap dat zijn bestaan rationeel betwijfeld kan worden, en mijn geest niet; cogito ergo sum (dus geest \neq lichaam)

(Simpel) argumenten tegen CID:

1. Nietszeggendheid: het is een negatieve en lege theorie, omdat ze alleen zegt dat Ψ s niet- Φ s zijn, maar niks zegt over de natuur en structuur van Φ s.
 \Rightarrow Repliek: de theorie is mysterieus omdat de werkelijkheid zelf mysterieus is
2. Oneconomisch: waarom 2 substanties postuleren als men ook 1 kan postuleren (Scheermes van Ockham).
 \Rightarrow Repliek: methodologie is geen ontologie; de werkelijkheid zelf is complex.

3. Onwetenschappelijk – drievoudig:
 - a. Het cartesianisme voegt een surplus aan immaterialiteit toe aan het wetenschappelijke, naturalistische wereldbeeld toe.
 - ⇒ Repliek: onbetwiste aanvaarding van de autoriteit van de natuurwetenschap
 - b. Waar treden zielen op in de ontwikkeling van substanties in de darwinistische evolutietheorie? En waarom slechts in selecte wezens?
 - ⇒ Repliek: dieren zijn slechts machines; God
 - c. Mentale processen zijn duidelijk afhankelijk van onze (fysische) hersenen, hoe zijn Ψ s dan reëel onderscheiden van Φ s?
 - ⇒ Repliek: interactionisme, het homo sapiens brein is slechts een coördinatensysteem van bewegingen en een schakel tussen het fysische en het psychische.
4. Conceptuele verwarring (Ryle): het cartesianisme maakt een categoriefout, wat het foutief onderbrengen van concept X in categorie C1, terwijl het eigenlijk hoort in C2 behelst. Het brengt namelijk mentale predikaten onder in de categorie van mechanistische concepten (bv. 'ding', 'eigenschap') terwijl ze eigenlijk tot de categorie van gedragsdisposities horen. Omdat de geest wordt opgevat als een ding, komt het naast het lichaam te staan. Hierdoor wordt de geest een soort para-mechanisch concept en wordt het ook causale relaties met het fysische toegeschreven.
 - ⇒ Repliek: taal filosofie en conceptuele analyse zijn geen ontologie en kunnen niet beslissen over de interne structuur van de realiteit

Het fundamentele bezwaar tegen het cartesianisme is het **probleem van mentale causaliteit**. Het probleem van de mysterieuze relatie tussen Ψ en Φ is drievoudig:

- i. Heterogeniteit van relata: hoe kan een volledig immateriële substantie een volledig materiele substantie bewegen?
- ii. Mechanisme: als de causale relatie bestaat, hoe werkt hij dan (causaal mechanisme)? Hoe werkt de omzetting in de pijnappelklier dan precies?
- iii. *Causale geslotenheid* (CG): de oorzaak van een fysische gebeurtenis moet een andere fysische gebeurtenis zijn. Dit is een plausibele vooronderstelling van de moderne natuurwetenschap.
 - ⇒ Het cartesianisme kan het concept van mentale causaliteit niet verhelderen; het blijft mysterieus. Volgens het cartesianisme is dit probleem in essentie mysterieus en onoplosbaar.

Het probleem van mentale causaliteit gaf de opvolgers van Descartes twee opties m.b.t. een oplossing: of men geeft (1) het dualisme op, of (2) het interactionisme. Gezien dat het bestaan van de ziel destijds een dogma was, kozen veel filosofen voor (2) of een aanpassing van het dualisme.

Het *niet-interactionistisch dualisme* stelt dus dat er geen rechtstreekse, causale interactie tussen Ψ s en Φ s bestaat. Hoe bestaat deze relatie dan wel? Het antwoord is co-variatie van Ψ s en Φ s. Twee vormen:

1. **OCCASIONALISME:** God veroorzaakt telkens een lichamelijke beweging ter gelegenheid van een geestelijke act en omgekeerd (Malbranche & Geulincx)
2. **PARALLELISME:** Het verloop van geestelijke akten/toestanden en het verloop van fysische/lichamelijke feiten door de tijd heen werden harmonieus ingesteld door God en van in den beginne lopen ze parallel zonder elkaar rechtstreeks causaal te beïnvloeden (Leibniz).

Kritiek: 'deus ex machina'-verklaring; ze doen beroep op toevallige, bovennatuurlijke ad hoc hypothesen en zorgen niet voor bevredigende verheldering van mentale causaliteit.

Een andere manier om de problemen met het cartesianisme op te lossen is om het substantie-dualisme te vervangen door een *eigenschap-dualisme*. Twee vormen:

1. **EMERGENTISME:** Alle substanties (en hun aggregaten) zijn stoffelijk, en wanneer fysische aggregaten een bepaald niveau van complexiteit bereiken dan emergeren er 'werkelijk nieuwe eigenschappen', en deze eigenschappen zijn niet fysisch. Emergente eigenschappen zijn niet-identiek en niet-reduceerbaar tot de fysische eigenschappen waar ze uit voortkomen. Ze zijn

sui generis en bezitten een autonome causale kracht (downward mental causation). Emergentie is een onverklaarbare sprong, en is daarmee onvoorspelbaar.

⇒ Kritiek: doorbreekt de geslotenheid van het causale domein

2. **EPIFENOMENALISME:** niet-interactionistisch eigenschap-dualisme. Mentale eigenschappen zijn hier niet-fysische, immateriële eigenschappen die geproduceerd worden door fysische substanties (brein) maar zelf niets produceren. Ψ s zijn een louter bijproduct; ze zijn causaal impotent.

⇒ Kritiek: respecteert de causale geslotenheid van Φ , maar ontkent mentale causaliteit.

Dan lijkt er niets anders op te zitten dan het verwerpen van het dualisme. Door het dogma van de niet-spatiale en niet-publieke geest moet men het **IDEALISTISCH MONISME** van Berkeley aanvaarden. Het idealisme ontkent het onafhankelijk bestaan van het fysische domein en bijgevolg moet het totale universum uit immateriële substanties en eigenschappen bestaan. M.a.w., alles is stoffelijk.

Het bestaan en natuur van fysische objecten is volgens Berkeley volledig afhankelijk van een waarnemende geest (fenomenalisme). Het bestaan van de geest zelf ligt in zijn actief waarnemen. Daarnaast onderscheidt Berkeley twee soorten geesten:

- i. De *goddelijke geest* (oneindig) die de continuïteit en objectiviteit van de ‘fysische’ objecten met hun immateriële natuur en ook de oorzaak is van de percepties van de objecten in de eindige geesten.
- ii. De *eindige, menselijke geesten*.
⇒ Het probleem van mentale causaliteit verdwijnt in het idealisme. Mentale causaliteit bestaat alleen in de vorm ‘mentaal → mentaal’ en dit is geen probleem omdat het totale ontologische domein geestelijk is. Als men deze conclusie niet accepteert (en de FP najaagt) dan is de laatste optie het materialistisch monisme.

College 6 – Mind/Body Problem III

Het *materialistisch monisme* vormt geen onsplitsbaar geheel; er is een spectrum van sterkere en zwakkere versies. Wat ze allemaal postuleren is een bepaalde vorm van een ‘stoffelijke geest’. Eerst zullen we de intelligibiliteit van dit concept uitdiepen vooraleer we op de verschillende opvattingen in zullen gaan.

Het ‘dogma van het sciëntisme’ is een belangrijke vooronderstelling van het materialistisch monisme. Hedendaagse philosophy of mind is een erfgenaam van de moderne empirische gedragswetenschappen. Bijgevolg is het materialisme een mechanische theorie van de geest. Soorten invloed:

- Vorm: er wordt veel aan theorievorming gedaan, i.t.t. de filosofen van de gewone taal, die anti-theoretisch en anti-wetenschappelijk zijn. Volgens hun is het geest-lichaam probleem een schijnprobleem, veroorzaakt door foutief taalgebruik. Zij geloven dat het volstaat om een adequate beschrijving te geven van de logische geografie van mentale concepten en te zeggen dat een verklaring geven of een herziening van mentale concepten niet mogelijk is omdat de gedragscriteria om mentale concepten toe te schrijven vastliggen in het normale taalgebruik. De taalfilosofie eigent zich een soort alleenrecht toe om het probleem te bespreken, wat eigenlijk zeer pretentief en onaanvaardbaar is. De theorievorming schuift van a priori naar a posteriori, en zo wordt de hedendaagse philosophy of mind deel van de cognitieve wetenschappen.
- Inhoud: het geldende mensbeeld en wereldbeeld wordt mechanistisch en deterministisch, en de natuur van de geest wordt verklaard binnen het kader van de fysica (fysicalisme). Het dualisme is hiermee definitief uitgesloten.

Men kan zich afvragen: vanwaar het privilege van het sciëntisme? De wetenschap is wellicht niet feilloos, maar heeft wel als enige een rationele consensus over controversiële vraagstukken bereikt in de zoektocht naar de waarheid. De hoop is nu om de eenheid van de wetenschappen te kunnen waarborgen

(ook een vooronderstelling) door de kloof tussen de psychologie (Ψ) en de neurofysiologie (Φ) te kunnen dichten. Men zou, zoals David Armstrong, het project van de wetenschap de “incorporatie van het subject van de mechanistische wereldvisie in de wereldvisie zelf” kunnen noemen.

De materialistische, mechanistische wereldvisie botst met de ‘common sense’-opvatting dat mentale fenomenen bestaan en een speciale aard lijken te hebben. Dit is op twee manieren op te lossen: of men ontkent het bestaan van mentale fenomenen, of men probeert ze in te passen in de wereldvisie.

Initieel lijken de concepten van bewustzijn en intentionaliteit al problematisch voor het materialisme: hoe kan het subjectieve karakter van een bewuste ervaring geheel en al geïdentificeerd worden met een objectieve activering van een op zichzelf niet-bewust neuronaal netwerk? Hoe kan een overtuiging die gericht is op een object letterlijk identiek zijn met een fysio-chemische hersentoestand die zelf op niets gericht is? Hoe kan de inhoud van een gedachte identiek zijn met een op zichzelf inhoudsloze neuronale configuratie?

Het project van het materialisme bestaat erin om de mogelijkheid van een materialistische theorie van de geest intelligibel te maken, wat een conceptueel en geen empirisch project is. Indien het bewustzijn en intentionaliteit in fysicistische termen kunnen worden geanalyseerd, dan is een materialistische theorie van de geest mogelijk en kunnen er daarna empirisch testbare hypothesen opgesteld worden. Deze mogelijkheid berust wel op drie fundamentele, methodologische vooronderstellingen:

- Verdeel en heers: begin theorievorming niet met het moeilijkste, maar het makkelijkste; de materialist focust eerst op niet-bewuste mentale disposities en (eventueel) daarna pas op bewuste.
- Derde persoon: Vervanging van het epistemologisch standpunt door het behaviouristische, wetenschappelijke standpunt waardoor mentale fenomenen a.d.h.v. lichamelijk gedrag en hersenactiviteit gedetecteerd kunnen worden. Subjectief lijken mentale fenomenen te onlichamelijk en mysterieus.
- Conceptuele analyse: Vervanging van een fenomenologische analyse van mentale concepten door een conceptuele; in plaats van de discussie over de natuur van fenomenen te laten gaan (waar meningen over uiteenlopend zijn), is het veiliger om over te stappen op algemeen geaccepteerde mentale concepten van FP, en deze door middel van de cognitieve wetenschappen te verklaren. Deze conceptuele analyse verschilt van die van de taalfilosofen omdat zij zich beperken tot beschrijving m.b.v. de gewone taal.

Het materialistische project is ontologisch, en vereist een herformulering van de vragen van het geest-lichaam probleem. Als men terugkijkt naar OV1/OV2 dan zien we het volgende:

- De eliminativist zegt “nee” tegen het bestaan van mentale fenomenen, waardoor de vraag naar de natuur irrelevant wordt.
- De verzoenende materialist zegt “ja” tegen OV1, maar “nee tegen OV2. Het project van het verzoenend materialisme bestaat er echter in om de mogelijkheid van Ψ s in een volledig fysisch universum aan te tonen, en vooral hoe dat mogelijk is. De centrale vragen zijn dan als volgt:

(KV1) Zijn Ψ s mogelijk?

(KV2') Hoe zijn Ψ s mogelijk (in een fysisch universum)?

(KV2*) Hoe is het mogelijk dat Ψ s en Φ s, uitgaande van het feit dat Ψ s speciale eigenschappen bezitten, equivalent kunnen zijn, ofwel dat Ψ s in termen van Φ s gedefinieerd kunnen worden?

⇒ KV1 is eigenlijk al beantwoord (want dat wordt verondersteld)

ELIMINATIVISME (Churchland): Mentale eigenschappen bestaan niet omdat mentale predikaten en FP algemeen onwaar zijn. Drie stappen in hun argumentatie:

1. FP is een empirische theorie: het gehele mentaal vocabularium is te vervangen door empirische equivalenten;
2. FP is onwaar: FP kan psychische fenomenen niet verklaren, want het is (i) stagnerend en onvruchtbaar, en (ii) theoretisch geïsoleerd;

3. FP moet opgevolgd worden door de neurowetenschap, die hedendaags de beste optie is om menselijk gedrag en de werking van het brein te verklaren.
 - ⇒ Eliminativisme is zelf-weerlegend, want ze veronderstelt concepten in de eigen theorie waarvan ze beweert dat ze geëlimineerd moeten worden.

BEHAVIOURISME (Ryle): “x heeft dispositionele eigenschap D, *asa*, indien initiërende conditie C gerealiseerd werd voor x, dan zou er een manifestatie M van x zijn”

Reductionistisch, perifeer materialisme. Het mentale leven speelt zich af aan de ‘periferie’ van het fysisch lichaam. De geest is identiek met gedrag (en veroorzaakt het niet; black box) en het behaviourisme is een materialisme omdat gedrag gebaseerd is op lichamelijke bewegingen in een spatio-temporeel kader. De geest is activiteit. Gedrag moet hier breed opgevat worden: het mentale wordt gereduceerd tot mogelijk (talig) gedrag, m.a.w., een *gedragsdispositie*.

Maar wat zijn dan mentale disposities? Deze moeten opgevat worden als dispositionele Φ -eigenschappen (zie definitie boven). Het toeschrijven van een dispositionele eigenschap komt neer op het toeschrijven van een tegenfeitelijke propositie (‘indien... dan...’-zin). Een complexe mentale dispositie is dan een verzameling van tegenfeitelijke zinnen. Dit wordt ook wel een *multi-track dispositie* genoemd, omdat er meerdere mogelijke stimuli en responsen zijn. Ryle’s analyse van mentale disposities is dus niet causaal maar tegenfeitelijk. Causaliteit veronderstelt namelijk een gebeurtenis in het feitelijke, spatio-temporele referentiekader (zie college 5), anders spreken we weer van een para-mechanisme.

Objecties:

1. Het mentale wordt exclusief gedefinieerd in termen van gedrag, maar wat is gedrag? Het is of een handeling (intentioneel), maar dit is zelf een mentaal fenomeen, leidend tot circulariteit, of een lichamelijke beweging (niet-intentioneel), maar dan is er sprake van onbepaaldheid omdat een lichamelijke beweging overeen kan komen met meerdere gedragsdisposities (en andersom). We hebben dus een groter complex van mentale toestanden nodig om gedrag te verklaren, maar dat is wederom circulair.
2. Het behaviourisme druist in tegen de common sense intuïties. Het is namelijk denkbaar dat pijn (mentale dispositie) zonder pijngedrag (manifestatie) mogelijk is (\Rightarrow geen noodzakelijke voorwaarde) en andersom ook (\Rightarrow geen voldoende voorwaarde). Daarnaast is pijn intuïtief een interne oorzaak voor pijngedrag.
3. Het behaviourisme ontkent mentale causaliteit, maar hoe moet de intieme band tussen het psychische en het fysische dan verhelderd worden?

College 7 – Mind/Body Problem IV

Omdat behaviourisme en eliminativisme geen goede opties meer zijn (en dualisme uitgesloten is) wordt de keuze beperkt tussen de identiteitstheorie of functionalisme (reductief vs. niet-reductief).

IDENTITEITSTHEORIE: Het brein is identiek met de geest, en dit is dan ook de relatie tussen de geest en het brein: identiteit (geen correlatie of causaliteit). Sensaties en PA’s zijn strikt identiek met neurofysiologische gebeurtenissen en toestanden in het brein. Omdat mentale toestanden volledig in termen van fysische gebeurtenissen zijn geconceptualiseerd, is er geen sprake van circulariteit/regressie. De geest-brein identiteitshypothese vereist wel een verduidelijking van wat ze bedoelt met “geest is brein”. De “is” wordt hier gebruikt in de vorm van compositie, als een verduidelijking van de essentie (i.t.t. het gebruik als predikatie (eigenschap) of definitie (synonymie)). Het is m.a.w. een hypothetische identificatie en geen fenomenologische.

Objecties:

1. Een verbaal rapport van een subjectieve (pijn)sensatie heeft een andere semantische waarde dan een verbaal rapport van een neurofysiologische gebeurtenis.

- ⇒ Repliek: de identiteitshypothese is ontologisch, niet semantisch; beide rapporten verwijzen wel naar dezelfde gebeurtenis.
- 2. Een mentaal beeld heeft bepaalde kwalitatieve eigenschappen, maar een hersenproces niet (wet van Leibniz; strikte identiteit)
 - ⇒ Repliek: niet het mentale beeld, maar de ervaring van het hebben van een mentaal beeld wordt geïdentificeerd met een hersenproces.
- 3. Mentale toestanden zijn niet-spatiaal en niet-publiek, en hersentoestanden wel (WvL dn)
 - ⇒ Repliek: Zolang de identiteitshypothese niet gefalsificeerd is kunnen we (voorwaardelijk) aannemen dat mentale toestanden identiek zijn met hersentoestanden en is het perfect mogelijk dat mentale toestanden wel publiek en spaciaal zijn (begging the question)
- 4. De identiteitstheoreticus lost het probleem van fenomenaliteit niet op, want hij stelt het probleem in topisch-neutrale (plaatsneutrale) termen, die zodanig abstract zijn dat mentale toestanden compatibel zijn met zowel de materialistische ‘plaats’ als de dualistische ‘plaats’. Hij doet dit met ‘topisch neutrale rapporten’ (TNR) van sensaties/quale, die de vorm van “there is something going on...” hebben. De identiteitstheorie ontwijkt dus eerder het probleem op slinkse wijze.
- 5. Het probleem van mentale causaliteit lijkt op het eerste gezicht niet fataal omdat het in de identiteitstheorie een soort van specifiek geval is van fysische causaliteit. Het mentale verliest echter qua mentale zijn kracht, want het is slechts qua fysische dat het causale kracht kan uitoefenen. Een intentie speelt dan geen enkele rol bij het tot stand brengen van een beweging.

De identiteitstheorie wordt ook wel het *reductionistisch materialisme* genoemd, maar wat is reductionisme (\neq reductie)? Hiervoor hebben we het type/token onderscheid nodig:

- *Token*: individu, particulier ding
- *Type*: soort, klasse, eigenschap

Dit onderscheid is op alle entiteiten in het spatio-temporele systeem toepasbaar (en in het fysicalisme dus ook op mentale toestanden). De relatie tussen types en tokens is instantiatie: elk token instantieert (is een instantie van) één type.

⇒ Onderscheid token/type-fysicalisme:

- *Token-fysicalisme* (zwak): een mentaal token is identisch met een neurofysiologisch token.
- *Type-fysicalisme* (sterk): een mentaal type is identisch met een neurofysiologisch type.

Type-fysicalisme is sterker, omdat het een theorie wilt zijn met algemene, universele zinnen en wetmatigheden (en dit kan alleen met types omdat tokens particularia zijn). Het type-fysicalisme tracht bijgevolg om *psycho-fysische wetmatigheden* (PFW) te formuleren van de volgende vorm:

“Alle personen met type-sensatie S/propositionele attitude PA bevinden zich (noodzakelijk) in type-breinproces N, en vice versa”

Type-fysicalisme is een voldoende voorwaarde voor token-fysicalisme en het token-fysicalisme is een noodzakelijke voorwaarde voor het type-fysicalisme. Bij type-fysicalisme is er een enkelvoudige relatie tussen types, maar bij token-fysicalisme is deze relatie meervoudig.

Hier komen we terug op het reductionisme. Omdat de type-type identiteitstheorie PFW's tracht te formuleren, is deze theorie ook een reductionistisch materialisme. Om reductie te kunnen voltrekken is het namelijk nodig om PFW's te hebben om de kloof tussen psychologie en neurofysiologie te dichten. Ψ wordt dus gereduceerd tot Φ .

Wat is (theoretische) *reductie*? De herleiding van een theorie op een hoger niveau T_2 in de hiërarchie van de wetenschappen tot een theorie op een lager niveau T_1 , wat alleen mogelijk is indien alle wetten van T_2 logisch afleidbaar zijn uit de wetten van T_1 , en deze logische afleidbaarheid steunt op brugwetten (BW), die in dit geval systematisch het psychologisch vocabularium met het neurologisch vocabularium verbinden. De PFW's zijn BW's.

Argumenten vóór de type-type identiteitstheorie:

- In het licht van het wetenschappelijk wereldbeeld beantwoordt de theorie aan het desideratum van strikt wetenschappelijke, nomologische theorievorming en aan het project van de eenheid van de wetenschappen

Argumenten tegen de type-type identiteitstheorie:

- Op basis van de FP is de type-type identiteitstheorie onbevredigend, want zij ontkent de mogelijkheid en de wenselijkheid van PFW's. Dit is echter te vroeg om te concluderen, want hoewel er nog geen exhaustieve lijst van PFW's opgesteld is, is het reductionistische onderzoeksprogramma in volle gang en ligt haar waarheid wellicht in de toekomst.
 - De theorie is niet empirisch adequaat, wat voorspruit uit het fenomeen van **meervoudige realiseerbaarheid van het mentale**. PFW impliceert dat één mentaal type altijd identisch is met één type neuronaal substraat bij alle personen. Maar dit wordt empirisch tegengesproken, want a) er bestaan meerdere causale hersengeschiedenissen onder personen met hetzelfde resultaat, en b) door bijvoorbeeld hersenplasticiteit kan het neuronale substraat binnen dezelfde persoon veranderen over de tijd. De reductie tot hersenprocessen sluit ook de mogelijkheid van niet-menselijk mentaal leven uit.
- ⇒ De identiteit tussen mentale types en neuronale types is bijgevolg niet strikt, dus een meervoudige realiseerbaarheid van het mentale (type) in het fysische (type) lijkt meer plausibel. Indien het mentale meervoudig realiseerbaar is, dan is het mentale verbonden met een zeer heterogeen fysisch domein, zodanig dat er geen gemeenschappelijk fysisch kenmerk meer gevonden kan worden om het mentale als geheel te karakteriseren. De meervoudige realiseerbaarheid van het mentale maakt het type-fysicalisme implausibel, maar ze is echter wel verenigbaar met het token-functionalisme.

Het **FUNCTIONALISME** identificeert de geest als een topisch neutrale, abstracte, causale rol/functie. Mentale type-toestanden zijn bijgevolg niet substantieel, maar essentieel relationeel. Mentale types worden gedefinieerd in termen van typische patronen van macroscopische causale relaties tussen (i) zintuiglijke input, (ii) gedragsoutput en (iii) andere tussenliggende mentale types. Het wordt ook wel gezien als de synthese van het behaviourisme en de identiteitstheorie: het combineert de input-output relaties van de ene met de inwendige brein-causaliteit van de ander.

Om het basisconcept van *causale rol* te specificeren en het analysans te preciseren, doet de functionalist een beroep op het onderscheid tussen natuurlijke en functionele soorten:

- *Natuurlijke soorten*: biologisch; concreet, microscopisch en intrinsiek
- *Functionele soorten*: artificieel; abstract, macroscopisch en extrinsiek/relationeel

Het functionalisme (i.t.t. identiteitstheorie) formuleert een 'functionele soort'-definitie van het mentale: "een *mentale toestand* is iets dat een bepaalde functie heeft of een bepaalde rol speelt in het tot stand komen van gedrag". De concepten 'functie' (1) en 'causale rol' (2) worden verder uitgeklaard door twee versies van het functionalisme:

1. **Turing machine functionalisme**: computeranalogie; de geest staat tot het brein hoe software tot hardware staat. Van een Turing machine weten we alle mogelijke toestanden die met elkaar en met de in- en outputs van het systeem staan. We weten dus alleen het 'gedrag' van de machine, niet van de fysische realisaties. De output ('gedrag') van de machine is een functie van de input en de toestand. Mentale toestanden moeten we analoog opvatten aan functionele machinetoestanden van dezelfde structuur: "voor een interne mentale toestand geldt $\{(input) \wedge (andere\ mentale\ toestanden)\} \rightarrow (output)\}$ ". Deze vorm van functionalisme is in strikte zin niet eens causaal, maar puur logisch. De volgende versie sluit iets meer aan bij de common sense.
2. **Causale rol functionalisme**: een conceptueel causale analyse (geen empirische) van mentale concepten; er worden dus i.p.v. causaal-nomologische verklaringen noodzakelijke en voldoende voorwaarden opgesteld. Een mentale toestand vervult een causale rol, als hij in causale relatie staat met zintuiglijke stimulaties, gedragsoutputs en andere mentale toestanden. Ofwel: een mentale type-toestand moet begrepen worden tegen de achtergrond van (a) het relationele karakter en (b) het holistische karakter van het mentale.

Objectie: de definitie van een mentale type-toestand van het causale rol functionalisme is circulair, want het gebruikt “andere mentale toestanden” in de definitie van mentale concepten.

⇒ Oplossing: **Ramseyficatie**. Drie stappen:

1. Formuleer psychologische theorie T in één zin (structuur F_a) die alle causale relaties tussen Ψ s, Is en Os specificiert.
⇒ T: causale relaties ($I_1, \dots, I_n, \Psi_1, \dots, \Psi_n, O_1, \dots, O_n$)
2. Pas existentiële instantiatie toe op de zin.
⇒ $(\exists x_1, \dots, x_n)(\exists y_1, \dots, y_n)(\exists z_1, \dots, z_n)[T(x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n)]$
3. Indien “ y_i ” de variabele is die “type-toestand Ψ ” vervangt, dan kan Ψ als volgt niet-circulair gedefinieerd worden.
⇒ “S heeft type-toestand $\Psi =_{\text{def}} (\exists x_1, \dots, x_n)(\exists y_1, \dots, y_n)(\exists z_1, \dots, z_n)[T(x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n)] \wedge S$ heeft y_i ”

De functionalistische hypothese beantwoordt bijgevolg aan het fenomeen van meervoudige realiseerbaarheid van het mentale, en is bijgevolg empirisch adequaat. Omdat het ook mogelijk is voor niet-menselijke entiteiten om mentale toestanden te hebben, moet het functionalisme een restrictie op de hypothese plaatsen, namelijk dat de causale rolvervullers fysisch moeten zijn.

Hoewel de ontologie van het functionalisme fysisch is, zijn mentale type-toestanden niet identisch met fysische type-toestanden. Deze relatie wordt verklaard door het onderscheid tussen ontologie en metafysica:

- *metafysica* heeft het over de natuur/essentie van de dingen. In het functionalisme is dat de functionele type-toestand (abstract, topisch neutrale eigenschap)
- *ontologie* heeft het over welke dingen er bestaan. Hierin is het functionalisme een materialisme, maar wel van de token-variant.

Functionalisme is dus een niet-reductionistisch materialisme; psychische eigenschappen zijn niet identisch met fysische eigenschappen. Er bestaat een dualiteit op het type-niveau, waardoor het mentale een (zwakke) ontologische autonomie bezit (cf. meervoudige realiseerbaarheid van het mentale)

*Onderscheid in (mentale) autonomie:

- i. *robuuste* autonomie: cartesiaans substantie-dualisme
- ii. *sterke* autonomie: emergentistisch eigenschap-dualisme
- iii. *zwakke* autonomie: functionalisme
- iv. *heteronomie*: identiteitstheorie

Hoe moeten we deze zwakke mentale autonomie interpreteren? Hoewel mentale eigenschappen niet reduceerbaar zijn tot fysische, zijn ze er wel van afhankelijk. De functionalist introduceert hiervoor de **these van superveniëntie** (ST), volgens welke mentale eigenschappen bovenop fysische eigenschappen komen. De superveniëntie-relatie is een ontologische relatie (geen causale of identiteits-), gespecificeerd als volgt: “de gelijkheid van Φ -types impliceert noodzakelijk de gelijkheid van Ψ -types, maar niet omgekeerd.”

ST garandeert materialisme; twee personen met dezelfde fysische toestand moeten dezelfde psychische toestand hebben, maar niet omgekeerd. Superveniënte eigenschappen zijn dus (i) covariant (veranderen mee met Φ), (ii) afhankelijk, en (iii) niet-reduceerbaar tot hun basiseigenschappen. Het functionalisme neemt dus ook een tussenpositie in tussen emergentie en reductie, wat autonomie betreft.

Argumenten vóór het functionalisme:

- Het probleem van mentale causaliteit lijkt opgelost. Het functionalisme respecteert de causale geslotenheid maar veroorzaakt op superveniënte wijze een fysische gebeurtenis, *asa*, de fysische gebeurtenis op strikte wijze wordt veroorzaakt door een neuronale toestand. Zonder de fysische basis valt de superveniënte causale band ook weg.
- Het functionalisme slaagt erin om de FP en het mechanistisch wereldbeeld te verzoenen (als het ware supervenieert de FP op het wetenschappelijke wereldbeeld).

Toch blijkt de positie van het functionalisme instabiel, want na de **kritiek van Jaegwon Kim** blijkt het toch niet mentale causaliteit in termen van het wetenschappelijk wereldbeeld te kunnen verklaren. Volgens hem aanvaardt het functionalisme de volgende 4 elementen:

- Het principe van causale geslotenheid
- Het principe van causale exclusie
- ST
- Niet-reduceerbaarheid van Ψ tot Φ

Zijn argument kunnen we als volgt samenvatten:

1. Stel dat mentale gebeurtenis M mentale gebeurtenis M^* veroorzaakt, dan heeft M^* superveniëntiebasis F^* .
2. M veroorzaakt M^* door (indirect) F^* te veroorzaken.
3. M heeft zelf superveniëntiebasis F, en dus veroorzaakt ook $F \rightarrow F^*$, omdat M op F supervenieert en M^* op F^* .
4. F^* 's optreden wordt op twee verschillende manieren gedetermineerd, wat indruist tegen het principe van causale exclusie (ook al is de lijn tussen M en F^* geen strikte dualistische causaliteit).
5. De enige uitzondering op causale exclusie is overdeterminatie (twee mogelijke, beide voldoende oorzaken voor één fenomeen), maar hiervan is in het geval van mentale causaliteit geen sprake. Dit betekent dan dat F de enige voldoende oorzaak is van F^* . De M- F^* causale relatie vervalt dus, maar daarmee ook de M- M^* relatie omdat deze daarvan afhangt.
 \Rightarrow Mentale causaliteit d.m.v. superveniëntie is onintelligibel.

College 8 – Determinisme I

Wat is *vrijheid*? Onder vrijheid verstaan we wilsvrijheid, waaronder (1) keuzevrijheid en (2) handlingsvrijheid vallen:

- (1) *Keuzevrijheid* wordt opgevat als het hebben van *alternatieve* mogelijkheden (AP). Het komt erop neer dat ik anders had kunnen kiezen; als je vrij bent heb je meerdere keuzemogelijkheden maar onder dwang heb je er maar één. Op moment t zijn er alternatieve mogelijkheden, maar na je keuze is er ook slechts één. We karakteriseren AP als volgt: "S is vrij in het doen van A, betekent dat hoewel S in feite A doet, had S ook $\neg A$, B, C, ... kunnen doen."
- (2) Een alternatieve opvatting van AP is de "could have done otherwise"-vrijheid, waar niet gesproken wordt over een tegenfeitelijke situatie, maar over het vermogen, of in staat te zijn om zowel A als B te doen.

Wat is *determinisme*? Onder de noemer van het fysisch-causaal determinisme kunnen we verschillende vormen van determinisme (sociaal, biologisch en psychologisch) samenvatten. Deze vorm van determinisme vindt zijn oorsprong bij Laplace, wie zegt dat "alles wat er gebeurt noodzakelijk voortvloeit uit voorgaande, fysische oorzaken". Dit veronderstelt het fysicalisme, ofwel alle mentale gebeurtenissen zijn identiek met fysische gebeurtenissen in tijd en ruimte. Het universum bestaat dus alleen uit fysische gebeurtenissen, onderworpen aan causale wetmatigheden.

Wat is *causaliteit*? Hiervoor gebruiken we Hume's regulariteitsanalyse: causaliteit veronderstelt (i) prioriteit (opeenvolging in tijd), (ii) contigüiteit (aangrenzendheid) in tijd en ruimte, en (iii) constante conjunctie (nomologische noodzakelijkheid). Het empirische verband dat hieruit volgt noemt Hume een *regulariteit*.

Een verdere precisering van het determinisme veronderstelt de volgende intuïtie (die uiteindelijk zullen leiden tot een formulering van de D-hypothese):

- (0) Het verleden determineert wat er nu gebeurt en zal gebeuren;
- (1) Er is voldoende grond voor alles wat gebeurt;
- (2) Elke gebeurtenis die optreedt heeft een oorzaak;
- (3) De wereld is bepaald door determinisme, *asa*, gegeven de stand van zaken op tijdstip t, de daarop volgende standen van zaken vastgelegd zijn door natuurwetten.

⇒ (D) Gegeven de initiële condities van het universum en de natuurwetmatigheden, is er slechts één geschiedenis fysisch mogelijk. $[(P_0 \wedge L) \rightarrow P_t]$

In strikte zin is D geen causaal determinisme (want dit wordt niet gespecificeerd) maar we zullen wetmatigheden opvatten als strikt noodzakelijk waardoor het wel causaal wordt. Ook is er strikt geen sprake van een fysicalisme, maar ook hier gaan we wel uit van het reductionistisch materialisme.

Wat is *toeval* in een deterministisch universum? J.J.C. Smart onderscheidt twee soorten:

- i. *accidenteel toevallige gebeurtenis*: coïncidentie (samenloop van omstandigheden); compatibel met determinisme.
- ii. *puur toevallige gebeurtenis*: onvoorspelbare gebeurtenis; niet compatibel met determinisme.

Determinisme bevat oorzaak en wetmatig gevolg, maar geen voorspelbaarheid (of lotsbestemming). Wat is *voorspelbaarheid*? Er is sprake van een voorspelling, *asa*, a) ik op een eerder moment omschreven heb dat x op een (exact) later moment zou plaatsvinden, b) ik op dat moment bewijsgronden had om die voorspelling te doen en c) ik de voorspelling ook had afgeleid van die bewijsgronden.

- Voorspelbaarheid is een epistemologisch concept (bewijsgronden), terwijl determinisme een ontologisch concept is. Voorspelbaarheid impliceert determinisme, niet andersom.

Is er een probleem van vrijheid en determinisme? De common sense gaat ervan uit dat we vrij zijn om A te doen, tenzij we gedwongen of fysiek belemmerd worden. Dit gevoel van vrije keuze te hebben, gepaard met de overtuiging dat we vrije wil hebben, is dat niet genoeg?

- ⇒ Deze ervaringen van het hebben van vrije wil hoeven niet ontkend te worden, maar deze ervaringen kunnen illusoir zijn. Het probleem is metafysisch, niet psychologisch. We moeten dus een onderscheid maken tussen het subjectieve gevoel/ervaring vrije keuzes te maken en het objectief feit om keuze- en beslissingsvrijheid te hebben.

De common sense komt normaal niet in aanraking met de determinisme-hypothese; dit probleem stelt zich pas op een voldoende reflexief niveau. Dan lijkt determinisme in conflict te staan met AP, want jouw keuze is volgens het determinisme al lang bepaald op basis van natuurwetten en het verleden, ongeacht of je het gevoel van keuzevrijheid hebt. Het debat binnen het determinisme gaat dan uitmaken of dit prima facie conflict ook daadwerkelijk een conflict is.

Als de vrije wil een illusie blijkt te zijn, heeft dit sterke gevolgen voor praktische en morele contexten die vrije wil vooronderstellen. Morele verantwoordelijkheid veronderstelt namelijk het hebben van alternatieve mogelijkheden volgens het *principe van alternatieve mogelijkheden* (PAP):

“S is moreel verantwoordelijk voor handeling A, *asa*, S anders had kunnen handelen.”

Twee kampen, genaamd het compatibilisme en incompatibilisme, hebben een verschillende kijk op de vraag op vrijheid (AP) onder de aanname van de hypothese van determinisme wel of niet mogelijk is. Volgens het compatibilisme kunnen vrijheid en determinisme samengaan, volgens het incompatibilisme niet. De één staat een vertakkende-wegen beeld van de toekomst toe, de ander slecht een enkel, lineair verloop van de geschiedenis.

TRADITIONEEL COMPATIBILISME: De compatibilist zal verdedigen dat de natuurlijke vrijheidsopvatting (AP) werkelijk verenigbaar is met het determinisme. Maar hoe dan? Het antwoord hierop is de *conditionele analyse* (CA) van AP. A.J. Ayer geeft 3+ voorwaarden voor AP:

1. Conditiona liteit (als anders gekozen, dan anders gehandeld)
2. Vrijwilligheid (niet-compulsief handelen)
3. Afwezigheid van dwang

Hier geeft hij nog aanvullende voorwaarden: (i) natuurwetmatigheden zijn descriptief, niet prescriptief; (ii) determinisme \neq fatalisme; (iii) vrije handelingen zijn intentionele handelingen.

Hume's 'liberty of spontaneity' sluit goed aan bij Ayer's derde voorwaarde. Vrijheid moet namelijk niet tegengesteld worden aan causaliteit, maar aan dwang. De tweede voorwaarde stelt verder dat de

handeling onder “normale” mentale toestanden veroorzaakt wordt. Dus handelingsvrijheid is wel onverenigbaar met speciale (dwang, compulsie) oorzaken. De incompatibiliteit van D en AP lijkt te wijten aan conceptuele verwarring. Voorwaarde 1 geeft echter antwoord op de hoe-vraag. CA toont zich door de term “could” in AP te splitsen in 2 varianten:

- i. Categorisch (“kunnen/kan”)
- ii. Hypothetisch (“had kunnen/kon”)

Hoewel i incompatibel is met D, is ii dat niet omdat AP afhankelijk wordt van een tegenfeitelijke conditie (if). Hoewel ik A niet gedaan heb, had ik A wel kunnen doen mits ik voor A gekozen had. Dit is de tegenfeitelijke interpretatie van AP. De D-hypothese houdt in dat het verleden nomologisch het heden veroorzaakt. Maar als in het verleden een andere oorzaak opgetreden had, dan zou in het heden ook een ander effect opgetreden zijn. Deze mogelijkheid van compatibilisme wordt ook beaamd door de compatibilist.

Kritiek op CA: het analyseert handelingsvrijheid in termen van de wil, maar dit veronderstelt ook wilsvrijheid. Maar als men veronderstelt dat dit weer opgelost kan worden door een “could have done otherwise”-interpretatie te geven aan de wil, dan beantwoorden we de vraag: “had S ook anders kunnen willen?”. Hier doet zich echter hetzelfde funderingsprobleem voor, leidend tot een oneindige regressie. Daarnaast is het psychologisch onmogelijk om te kiezen om te kiezen (Russell).

College 9 – Determinisme II

Eén van de centrale argumenten van het incompatibilisme is het **consequentie-argument** van Van Inwagen, die stapsgewijs formaliseerbaar is. CA is hoogstwaarschijnlijk onwaar, dus de incompatibilist wil d.m.v. een argument de prima facie incompatibiliteit van AP en D formeel aantonen en zo ook een argument formuleren tegen de compatibilistische these. Het argument gaat (in losse zin) als volgt: “Indien het determinisme waar is, dan zijn onze handelingen de consequenties van de natuurwetten en gebeurtenissen in het verre verleden. Maar over het verleden (voor onze geboorte) hebben we geen macht, noch over de natuurwetten. Daarom hebben we geen controle over de consequenties van deze dingen (en ook onze handelingen)”.

⇒ Omdat het verleden en natuurwetten gefixeerd zijn, en D waar zou zijn, is AP onmogelijk.

Van Inwagen geeft ook een logische, losse falsificatie van AP:

- (1) $(P \wedge L) \rightarrow A$ hypothese D
- (2) $\neg A$ aanname compatibilist
- (3) $\neg(P \wedge L)$
- (4) $\neg(P \vee L)$
- (5) $\neg(\neg P)$ aannames dat P & L gefixeerd zijn
- (6) $\neg(\neg L)$
- (7) $P \wedge L$
- (8) A contradictie met (2), (dus ongeldig)

Hij zal ook PAP in het consequentie-argument incorporeren en formele en modale versies geven van het consequentie-argument, maar vanwege hoge complexiteit volstaat het (volgens mij) om hier te stellen dat ze bestaan.

Kritiek op consequentie-argument: Als ‘kunnen’ in “we kunnen L en P niet veranderen” opgevat wordt in de hypothetische zin, is de conclusie van het argument onjuist, want dan hadden we wel degelijk anders kunnen handelen als determinisme waar was (volgens CA). De premissen waren echter ook juist in deze hypothetische zin (want fixatie van verleden en wetten). Het argument heeft de sterkere, categorische variant van ‘kunnen’ nodig. Terwijl de hypothetische variant het directe verleden d.m.v. condities (soms) kan veranderen (hier dus niet), vereist de categorische versie dat S anders kan handelen gegeven dezelfde condities in het verleden. Het is duidelijk dat van Inwagen de tweede versie preferert, waardoor het argument wel geldig is.

Maar dan lijkt de waarheid van het compatibilisme of incompatibilisme af te hangen van de interpretatie van 'kunnen', waardoor we in een soort impasse terecht komen.

Als we echter stellen dat het consequentie-argument logisch en solide is, dat CA onwaar is en PAP waar is, dan kunnen we stellen dat er sprake is van een rationele rechtvaardiging van de incompatibiliteit van morele verantwoordelijkheid en determinisme. Maar het incompatibilisme kan ook weer gesplitst worden in twee kampen:

- Hard determinisme: bevestiging van D en negatie van V (zie college 12)
- Indeterminisme: bevestiging van V en negatie van D

Het indeterminisme, ook wel bekend als libertinisme, kampt met een intelligibiliteitsprobleem. Een vrije keuze als een ongedetermineerd verschijnsel wordt niet bepaald door motieven en redenen, en lijken bijgevolg arbitrair, irrationeel, ongecontroleerd en onverklaarbaar te zijn. Kunnen vrijheid en verantwoordelijkheid dan wel samengaan met determinisme? Er zijn drie theorieën met een antwoord:

1. Causaal indeterminisme: vrije handelingen worden wél veroorzaakt door redenen, maar de causale band tussen reden en handeling is probabilistisch
2. Agent-causation: hoewel vrije handelingen niet veroorzaakt worden door redenen (en motieven), worden ze gedetermineerd door het onbeïnvloede zelf.
3. Noncausal indeterminisme: vrije handelingen kunnen niet-causaal verklaard worden in de termen van de rationaliteit van de redenen zelf.

Wat is *indeterminisme*? Twee opvattingen:

- i. Afwezigheid van oorzaak; toeval; niet-causaal; breuk in de causale keten
⇒ Contra-causale vrijheid
- ii. Probabilistische oorzaak; het is mogelijk dat gebeurtenissen niet-deterministisch causaal veroorzaakt worden; geen breuken in de causale keten, slechts probabilistische vertakkingen.

Er zijn twee beroemde tegenargumenten geformuleerd tegen het libertinisme (en dus voor het compatibilisme). De eerste ervan is het **luck-argument**, wat toont dat indeterminisme geen noodzakelijke voorwaarde is voor AP-vrijheid.

- (1) Indien S's handeling A ongedetermineerd is, dan is A een kwestie van geluk (voor zowel indeterminisme-opvatting i als ii).
 - (2) Indien A een kwestie van geluk is, dan heeft S geen controle over A.
 - (3) Indien S geen controle heeft over A, dan is S's handeling niet vrij/S niet moreel verantwoordelijk voor A.
 - (4) Indien S's handeling A ongedetermineerd is, dan is S's handeling A niet vrij/S niet moreel verantwoordelijk voor A
⇒ De conclusie is dus dat indeterminisme de afwezigheid van vrijheid en morele verantwoordelijkheid impliceert (dus precies tegenovergesteld).
- Korte verduidelijking m.b.t. premisse 1: het maakt dus niet uit of men indeterminisme opvat als niet-causaal of probabilistisch. In de eerste opvatting lijkt het evident te zijn dat er sprake is van geluk, maar in de tweede is daar ook sprake van omdat er niks aan het verleden of de actor is die de gang van zaken bepaalt. Het is pure kans.

Het tweede argument, het **mind-argument**, is de logisch geldende contrapositie van het luck-argument. Volgens dit argument is vrijheid niet alleen compatibel met determinisme, maar vereist het zelfs determinisme (dus incompatibel met indeterminisme). Hobart's uitgangspunt is dat moreel handelen vrije, intentionele en rationele handelingen vooronderstelt. Anders geformuleerd, het intentioneel en rationeel zijn van handeling A is een noodzakelijke voorwaarde voor het eventueel vrij zijn van A waarvoor S dan moreel verantwoordelijk is. Het argument gaat als volgt:

- (1) Indien S's handeling A vrij is en S moreel verantwoordelijk is voor A, dan is A intentioneel en rationeel.
- (2) Indien S's handeling A intentioneel en rationeel is, dan is A geproduceerd door S zelf en S's intenties en redenen.

- (3) Indien S's handeling niet geproduceerd is door S zelf of S's intenties/redenen, dan is A niet intentioneel en rationeel.
- (4) Indien S's handeling A ongedetermineerd is, dan is A niet geproduceerd door S zelf of S's intenties en redenen.
- (5) Indien S's handeling A ongedetermineerd is, dan is A niet intentioneel en rationeel.
- (6) Indien S's handeling A intentioneel en rationeel is, dan is A gedetermineerd.
- (7) Indien handeling A vrij is en S moreel verantwoordelijk voor A, dan is A gedetermineerd/

Van belang is dat handelingen tot stand worden gebracht door een actor zelf en niet louter lichamelijke bewegingen zijn. Indeterminisme brengt 'out of the blue' een handeling tot stand, en dus speelt de actor hier geen rol in. Vrijheid en morele verantwoordelijkheid impliceren dus determinisme.

College 10 – Determinisme III

Uit het luck-argument en het mind-argument, samen met het consequentie-argument zouden we de conclusie kunnen trekken dat vrijheid en morele verantwoordelijkheid onmogelijk zijn in zowel een deterministisch als een indeterministisch universum. Maar de libertinist bevestigt het bestaan van vrijheid in een indeterministisch universum als onloochenbaar feit. Zo is de uitdaging aan de libertinist, om, met de tegenargumenten in het achterhoofd, die aanname intelligibel te maken. Er zullen dus extra voorwaarden geformuleerd moeten worden die gezamenlijk toch voldoende en individueel noodzakelijk zijn voor hun project (en die de tegenargumenten onschadelijk maken)

GEBEURTENIS-CAUSAAL LIBERTINISME: i.t.t. het klassieke compatibilisme, waar gebeurtenissen deterministisch-causaal met elkaar verbonden zijn, zal deze vorm van libertinisme hier aan toevoegen dat bepaalde cruciale gebeurtenissen die handelingen veroorzaken dat doen op een niet-deterministische manier. Niet alle gebeurtenissen hoeven dus niet-deterministisch veroorzaakt te worden; de vrije wil treedt alleen op in het domein van het moreel conflict (waar de actor verdeeld is door enerzijds zijn sterkste verlangen en anderzijds zijn morele plicht). Zo wordt het gehele domein van morele verantwoordelijkheid toch beslagen. Robert Kane is de meest vooraanstaande vertegenwoordiger van deze vorm van libertinisme.

Kane meent dat het consequentie-argument te zwak is om een basis voor het indeterminisme te geven, en overstijgt de discussie over "kunnen" door de AP-definitie van vrijheid te vervangen door de '**Ultimate Responsibility/Origine'-definitie (UR/O):**

"S is vrij in het doen van A, alleen als S ultiem verantwoordelijk is voor om het even wat een voldoende oorzaak (motief, reden) is voor A".

UR moet begrepen worden als sterke controle, waarna we deze in verband kunnen brengen met vrijheid.

Op basis hiervan geeft hij zijn argument voor de incompatibiliteit van vrije wil en determinisme: "we hebben alleen een vrije wil als we ultiem verantwoordelijk zijn voor onze handelingen (verantwoordelijk voor voldoende voorwaarden). Als determinisme waar is, liggen de voldoende voorwaarden in het verleden (of natuurwetten), dan liggen de voldoende voorwaarden in het verleden (of natuurwetten). Bijgevolg zijn wij niet ultiem verantwoordelijk, dus hebben we geen vrije wil".

- ⇒ Vrije wil is alleen mogelijk als sommige van onze handelingen geen voldoende voorwaarde in het verleden hebben liggen, en dus ongedetermineerd zijn.

In een deterministische wereld zou UR leiden tot een oneindige regressus, dus in het indeterminisme noemt Kane deze regressus-stoppende akten '*zelf-vormende akten*' of '*wil-zettende akten*'. Deze akten hebben geen voldoende voorwaarden en moeten dus tot stand komen door een wilsinspanning. Dit impliceert dus niet dat alle gebeurtenissen tot stand komen op deze manier, maar juist die cruciale gebeurtenissen die dat niet zijn vormen ons karakter.

Kane geeft het voorbeeld van "Els", een zakenvrouw die in een conflictsituatie terecht komt waar ze moet kiezen tussen de morele plicht en haar eigenbelang. De wilsinspanning bestaat hier

in het in overeenstemming brengen van haar wil met de morele reden (terwijl ze neigt naar de egoïstische reden). Of ze hierin slaagt of niet maakt niet uit volgens Kane, want het gaat erom dat het slagen of mislukken van de wilsinspanning altijd op een vrijwillige, intentionele, rationele manier zal gebeuren.

Deze duale intentionaliteit zet Kane om in een *plurariteitsvereiste* van de zelf-vormende akten (en daarmee ook vrije wil). Aan de hand hiervan leidt hij dan een alternatieve interpretatie af van AP: “Indien S vrijwillig, intentioneel en rationeel anders had kunnen beslissen of kiezen (dan S in feite besliste of koos), dan had S ook (vrijwillig, intentioneel en rationeel) anders kunnen handelen dan S in feite handelde”. Zo impliceert UR niet alleen indeterminisme, maar ook AP.

Kane heeft echter nog wel af te rekenen met het luck-argument, want anders is de wilsinspanning ook louter een kwestie van geluk. Dit argument gaat er immers vanuit dat probabilistische processen ook slechts een kwestie van toeval inhouden, maar Kane zegt dat probabilistische processen geenszins controle en morele verantwoordelijkheid uitsluiten, omdat de wil etting, ondanks de kansverdeling, toch kan slagen.

ACTOR-CAUSAAL LIBERTINISME: ondanks Kane’s repliek, lijkt het luck-argument toch de overhand te houden, en lijkt een niet-causale of probabilistisch causale variant van het indeterminisme toch niet de actor de vereiste controle te kunnen geven. Dit leidt ons tot Chisholm’s theorie, waar de actor ongedetermineerd blijft en zo zijn handelingen veroorzaakt. Zo hebben we een theorie waarbij een gebeurtenis niet veroorzaakt wordt door een andere gebeurtenis of dat een gebeurtenis helemaal niet (of probabilistisch) veroorzaakt zou zijn. Chisholm maakt hiertoe een onderscheid tussen ‘*transeunt causation*’ (tussen gebeurtenissen) en ‘*immanent causation*’ (tussen actor en gebeurtenis).

Chisholm’s theorie gaat over de wilsvrijheid eerder dan handelingsvrijheid, en hij neemt eigenlijk een tussenpositie in tussen het determinisme en indeterminisme (de actor determineert namelijk de handeling causaal, maar is zelf ongedetermineerd). Chisholm doorbreekt de causale keten van gebeurtenissen én ontwijkt de willekeuropvatting van het indeterminisme. Zo kunnen we een neurologische formulering geven van zijn theorie:

“S is vrij in het doen van A, *asa*, er een brein-gebeurtenis B is, zodanig dat (i) B niet veroorzaakt is door een andere gebeurtenis, (ii) B actor-veroorzaakt is door S en (iii) er een ketting van gebeurtenissen X_1, X_2, \dots, X_n is zodanig dat B X_1 gebeurtenis-veroorzaakt, $X_1 X_2$ gebeurtenis-veroorzaakt, \dots , en $X_n A$ gebeurtenis-veroorzaakt.”

De vrijheid van de mens in deze zin sluit de *wetmatigheden van immanente causaliteit* (WICs), en daarmee de gehele menswetenschappen, uit, welke de volgende vorm hebben:

“Indien op een bepaald tijdstip t persoon S zich in condities X, Y en Z bevindt, dan actor-veroorzaakt S een gebeurtenis van type G.”

- ⇒ Een reden-verklaring is nooit voldoende voor Chisholm’s theorie, want de actor heeft altijd de mogelijkheid om redenen te negeren en iets anders te doen. Er is bijgevolg geen logische connectie tussen willen en kiezen, en ook geen causale, dus hoe is motivationele beïnvloeding mogelijk in dit model?

Actor-causaal libertinisme wordt ook wel hard libertinisme genoemd, omdat het veel sterkere metafysische implicaties heeft dan Kane’s variant. Ook is het een radicale breuk met het compatibilisme, waarvoor de beperking tot gebeurtenis-causaliteit een dogma is. Chisholm’s theorie lijkt het luck-argument te ontcrachten omdat de actor het optreden van de handeling determineert (en zo de eerste premisse als onwaar bestempelt). Of het luck-argument ook daadwerkelijk ontcracht wordt is nog steeds een onderwerp van debat, omdat men het niet eens is over of de actor-causatie een kwestie van geluk of wil etting is (er is namelijk niets aan de wereld dat deze vorm van causatie bepaalt). Maar het karakter van actor-causaliteit is mysterieus en doet beroep op een extra metafysische factor.

College 11 – Determinisme IV

Het traditioneel compatibilisme en libertinisme zullen eerst uitgaan van vrijheid, en van daaruit zal pas een concept van morele verantwoordelijkheid ermee in verband worden gebracht. Twee structuren:

1. AP → PAP
2. O → UR → PO

Het traditioneel compatibilisme ging nog uit van structuur 1, maar in het nieuwe compatibilisme zal morele verantwoordelijkheid op de eerste plaats gezet worden en zal vrijheid van daaruit geanalyseerd worden. Alle vrijheid is bijgevolg “morele vrijheid”. Hierdoor is de nieuwe compatibilist niet afhankelijk van AP, en zal morele verantwoordelijkheid eerder in termen van controle gedefinieerd worden, in het onderstaande *principe van vrijheid/controle* (PC/PV)

“S is moreel verantwoordelijk voor het doen van A, *asa*, S controle heeft over het doen van A”.

Controle is een neutraler concept van vrijheid en is verbonden met het concept van “kunnen”. Eén of andere vorm van controle dient als een noodzakelijke voorwaarde voor morele verantwoordelijkheid. Het nieuwe compatibilisme wordt ook wel een “volledig compatibilisme” genoemd, omdat zowel morele verantwoordelijkheid als vrijheid compatibel (moeten) zijn met determinisme.

“Could have done otherwise”-vrijheid kan ook als een vorm van controle opgevat worden, namelijk zogenaamde *duale controle*. Frankfurt zal straks uitgaan van een andere vorm.

Wat is *morele verantwoordelijkheid*? In essentie, het vatbaar zijn voor blaam. Cruciaal is het verschil met legale verantwoordelijkheid. Morele verantwoordelijkheid...:

- ...heeft zowel met normatieve afkeuring als goedkeuring te maken (i.t.t. slechts afkeuring)
- ...kan niet verbonden zijn met een artificieel, normatief (mogelijk arbitrair) systeem.
- ...is niet afhankelijk van sanctionerende, uitwendige reacties van anderen.
- ...wordt geïmpliceerd door wettelijke verantwoordelijkheid, niet andersom.

Morele verantwoordelijkheid veronderstelt een *normatieve* actor: iemand die in staat is tot (i) intentioneel handelen en (ii) rationeel delibereren, en daarnaast ook (iii) in bezit is van een evaluatief schema met normatieve principes. Daarnaast betekent iemand moreel verantwoordelijk te houden ook dat we die persoon zien als het doelwit van onze *reactieve attitudes*. Deze attitudes zijn echter slechts contingent verbonden met morele verantwoordelijkheid, en zegt ook niets over de constitutie ervan.

Aristoteles’ opvatting maakt plaats voor nog twee bijkomende voorwaarden:

“S is moreel verantwoordelijk – verdient blaam of lof – voor het doen van A, *asa*, S willens en wetens A doet.”

De ‘willens’-voorwaarde is met name relevant voor het debat over vrijheid en determinisme. Dwang ondermijnt volgens Aristoteles laakbaarheid. Als we wilsvrijheid vatten als controle, zijn we terug bij de voorwaarde voor morele verantwoordelijkheid van het nieuwe compatibilisme.

HIERARCHISCH COMPATIBILISME (Frankfurt) neemt een centrale rol in in het debat over vrijheid en determinisme vandaag de dag. Eerst nemen we aan dat de incompatibilist PV zal interpreteren in termen van PAP, om zo te concluderen dat vrijheid en determinisme incompatibel zijn. Frankfurt zegt echter, als PAP onwaar blijkt te zijn, dan vereist morele verantwoordelijkheid geen AP, en dan kan morele verantwoordelijkheid wél samengaan met determinisme. Dit doet hij aan de hand van een gedachte-experiment, later bekend als zogenaamde *Frankfurt-voorbeelden*:

- Mr. Bean is op een cocktailpartij, en hoewel hij weet dat het moreel verkeerd is om te liegen, liegt hij toch omwille van ‘goodwill’-redenen. Echter, buiten zijn weten om, had hij niet anders kunnen handelen door een “mastermind” die hem strikt in de gaten houdt en had ingegrepen zodra Mr. Bean op het punt had gestaan om niet te liegen. Tijdens deze cocktailpartij hoeft mastermind alleen niet in te grijpen, want Mr. Bean loog al geheel op zichzelf en vanuit zichzelf op basis van eigen redenen.
- ⇒ Mr. Bean had geen AP, maar is toch moreel verantwoordelijk. Dus dit Frankfurt-voorbeeld vormt een tegenvoorbeeld voor PAP. De opvattingen van het incompatibilisme die uitgaan van AP zijn bijgevolg ongeldig. Mr. Bean is in het gedachte-experiment moreel verantwoordelijk voor zijn leugen, omdat hij ondanks het gebrek aan AP wel actuele controle had

over zijn handeling. Hier gaat het dan niet om duale, maar Frankfurt's compatibele notie van controle.

Zijn concept van hiërarchische controle zal hij ontwikkelen door een opvatting over de wil, een vervolgens geformuleerde theorie over de vrijheid van die wil, die tenslotte compatibel is met het determinisme.

Wat is de *wil*? (Frankfurt): *volitie*. Dit is een effectief verlangen; een verlangen dat zich daadwerkelijk realiseert in een handeling. Volities zijn geen ongerealiseerde intenties, en hoeven zelfs niet moreel of rationeel opgevat te worden (\approx Hobbes' daad van het willen).

Frankfurt's opvatting is echter ook hiërarchisch (gebaseerd op het verschil tussen bewustzijn en zelfbewustzijn):

1. *Eerste-orde verlangen*: handeling/stand van zaken als object
2. *Tweede-orde verlangen*: eerste-orde verlangen als object.

Onder de eerste vallen volities, maar ook niet-effectieve verlangens, maar de tweede dient ter motivatie van de eerste; tweede-orde verlangens verlangen om een eerste-orde verlangen effectief te maken.

⇒ Het vermogen om tweede-orde volities te vormen is essentieel verbonden met wilsvrijheid.

Wat is *wilsvrijheid*? (Frankfurt): hij focust op de relatie tussen verlangens en identificeert wilsvrijheid als volgt: "S's wil is vrij, alleen als S onbelemmerd effectief verlangt wat S verlangt effectief te verlangen". Dit houdt dus in dat S alleen vrij is als hij de eerste-orde volities heeft die zijn tweede-orde volities bepalen. De eerste-orde volitie is op te vatten als wil. Men is echter niet vrij als de wil niet bepaald wordt door tweede-orde volities. I.t.t. Hobbes' handelingsvrijheid, waar "S vrij handelt als S onbelemmerd doet wat S verlangt te doen", waar de dwangneuroticus ook zogenaamd vrij was, is dit niet het geval bij Frankfurt.

⇒ Deze opvatting over wilsvrijheid is compatibel met het determinisme. Zowel de eerste- als tweede-orde volities kunnen gedetermineerd zijn, alsmede de relatie tussen de twee, en nog kan de wil vrij zijn (want dit is een structurele eigenschap die alleen bestaat in de verhouding). Mutatis mutandis geldt dit ook voor Hobbes' handelingsvrijheid.

Hoe is morele verantwoordelijkheid dan compatibel met determinisme? Controle in PC/PV wordt dan **hiërarchische controle (HC)**. Wilsvrijheid is een voldoende voorwaarde, naar geen noodzakelijke voor HC. Frankfurt's definitie van wilsvrijheid kan opgevat worden als een soort "could have willed otherwise"-vrijheid, en ook deze vorm van AP ziet Frankfurt niet als voldoende. HC moet dan ook begrepen worden in termen van *zelf-identificatie*. Wanneer S een tweede-orde volitie vormt m.b.t. een eerste-orde verlangen, identificeert S zich met dit verlangen waardoor het een eerste-orde volitie wordt. Dit sluit niet uit dat de wil ook bepaald kan worden door andere krachten dan de kracht van de zelf-identificatie. Dit kan zelfs tegelijk gebeuren (overbepaling). HC kan dan zo geformuleerd worden:

"S heeft hiërarchische controle over handeling A, alleen als A veroorzaakt wordt door een wil waar S zich mee identificeert, ook al is S's wil overbepaald."

⇒ Alleen dwang waar men zich niet mee identificeert, heeft tot gevolg het wegvallen van morele verantwoordelijkheid.

Om een aantal kwalijke gevallen te voorkomen, moet men HC aanvullen met het *principe van nasporen*. Als S op het moment van handelen geen HC had, maar dit controleverlies een direct gevolg is van een handeling waar S wél HC over had, dan is S alsnog moreel verantwoordelijk.

Een alternatieve opvatting van controle is te zien in de interpretatie van PV/PC in termen van *reden-responsieve controle*. Volgens deze theorie heeft een persoon zonder duale controle toch genoeg controle om vrij een verantwoordelijk te handelen wanneer zijn deliberatieve mechanismen ontvankelijk zijn voor redenen. Zijn praktisch redeneren is vatbaar voor redenen, als er een wereld denkbaar is waarin er een voldoende reden bestaat om anders te handelen en hij dan ook anders handelt. De vrijheid van de wil is dan afhankelijk van het wel of niet responsief zijn voor redenen van het praktisch redeneren.

Frankfurt's hiërarchisch compatibilisme stoot echter op het **probleem van manipulatie**. De context van het debat is te vinden in de discussie tussen het structuralisme en het historicisme over de vraag naar de noodzakelijke en voldoende voorwaarden voor 'zwakke' controle. Het *structuralisme* (ook Frankfurt) pleit dat de vrijheid van een verlangen puur afhankelijk is van dispositionele/psychologische identificatie en dus totaal onafhankelijk is van (a) de moraliteit en rationaliteit, alsmede (b) de causale voorgeschiedenis en zijn en de causale effecten (na het handelen). Wilsvrijheid is bijgevolg puur afhankelijk van structuur. Volgens het *historicisme* zijn causale voorwaarden noodzakelijk om 'illegitieme' identificatieprocessen te voorkomen. Vrije verlangens ontstaan dan alleen als het historisch vormingsproces niet onder manipulatieve invloed veroorzaakt werd. Cruciaal is hier dat manipulatie dus ook tweede-orde verlangens (naast de eerste) kunnen beïnvloeden. Zo lijkt een persoon die noodzakelijke en voldoende structurele voorwaarden bezit, intuïtief toch niet vrij te zijn.

College 12 – Determinisme V

Het negatieve antwoord op de existentie-vraag van wilsvrijheid noemen we het *hard determinisme*. In de filosofie leidt dit tot de verwerping van zowel de zwakke (compatibilistische controle) als de sterke variant van vrijheid (AP). Vrijheid en determinisme sluiten elkaar uit, maar deze vorm van incompatibilisme aanvaardt ook dat determinisme waar is, en daarmee ondermijnt het ook wilsvrijheid. De hedendaagse variant van hard determinisme beschouwt wilsvrijheid als zowel incompatibel met determinisme als indeterminisme. Dit dubbele compatibilisme noemt men ook wel *hard incompatibilisme*. De waarheid van determinisme wordt bijgevolg irrelevant.

Wat volgt er dan uit deze inexistentie van wilsvrijheid? Twee mogelijke bewegingen:

- A) Optimistisch: de meeste morele principes en emoties overleven het niet-bestaan van wilsvrijheid.
- B) Pessimistisch: het is moreel noodzakelijk om, ondanks dat dit onwaar is, de illusie van vrijheid hoog te houden, omdat de consequenties van de algemene verspreiding van de waarheid te erg zouden zijn.

Men vindt hard determinisme ook terug in de moderne wetenschap. De wetenschappelijke thesis van het vrije-wil scepticisme wordt als volgt geformuleerd: "vrije handelingen of vrije menselijke gedragingen bestaan (waarschijnlijk) niet; vrije wil is (hoogstwaarschijnlijk) een illusie."

Deze thesis is in twee verschillende varianten te vinden in zowel de (biologische) neurowetenschap als de (psychologische) sociale wetenschap. Beide perspectieven delen 3 uitgangspunten:

1. *Bewustzijnsprincipe* (BP): "S's handeling A is vrij, alleen als A, ten minste gedeeltelijk, veroorzaakt wordt door S's bewuste beslissing/keuze/intentie/reden.
2. De kwestie is niet of beslissingen/keuzes/intenties/redeën reduceerbaar zijn tot of superveniëren op hersentoestanden of niet-bewuste mechanismen vooronderstellen en daarom onvrij zijn.
3. Vrijheid kan zowel in zwakke als in sterke zin ingevuld worden in BP.

Neurowetenschap – Benjamin Libet experiment: De beslissing om een pols te buigen wordt niet-bewust al "veroorzaakt" (lees: meetbaar in hersenactiviteit) voordat de proefpersoon bewust wordt van zijn beslissing om zijn pols te buigen. Het *neurofysiologisch argument* gaat als volgt:

- (1) De proefpersonen veroorzaken hun beslissingen niet door bewuste beslissingen; met een hoge graad van waarschijnlijkheid kan men op voorhand op basis van hersenactiviteit de beweging al voorspellen.
- (2) Alle handelingen worden op fundamenteel dezelfde manier veroorzaakt.
- (3) Geen enkele handeling wordt veroorzaakt door een bewuste beslissing
- (4) BP
- (5) Geen enkele handeling is (waarschijnlijk) vrij.
 - ⇒ Formeel is het argument geldig, maar het is niet solide. De experimentele data rechtvaardigen de (eerste) empirische propositie niet, want (i) de voorspelling is probabilistisch (dus niet deterministisch) en (ii) de experimentele data sluiten alternatieve interpretaties van de situatie niet uit. De eerste premisse is dus implausibel. Premisse 2 is echter ook implausibel, omdat arbitraire keuzes (zoals in het experiment) wel degelijk verschillen van weloverwo-

gen beslissingen/keuzes. Om deze gelijk te trekken is een verwaarlozing van de complexiteit van de tweede, dus is de generalisatie ongegrond. Premisse 3 is bijgevolg ook ongehoofwaardig, dus we kunnen concluderen dat het argument niet solide is.

Situationisme: menselijk gedrag wordt niet (zozeer) bepaald door bewuste redenen en stabiele karaktertrekken, maar (veeleer) door de onbewuste effecten van de situatie waarin (proef)personen zich bevinden. Voorbeelden: bystander-effect; Milgram-experiment; Stanford prison-experiment. Het *situationistische argument* kan als volgt geformuleerd worden:

- (1) Sommige menselijke gedragingen (in experimentele situaties) worden volledig bepaald door deze situaties en de effecten van deze situaties op menselijke gedragsautomatismen.
- (2) Alle menselijk gedrag wordt bepaald op fundamenteel dezelfde manier
- (3) Geen menselijk gedrag wordt veroorzaakt door bewuste redenen
- (4) BP
- (5) Alle menselijke gedrag is onvrij
 - ⇒ Het argument is geldig, maar wederom niet solide. Premisse 1 is waar, als men “volledig bepaald door” niet-deterministisch opvat en het vervangt door “aanzienlijk beïnvloed” (omdat het probabilistisch is). Premisse 2 is echter implausibel omdat men zich dan onder exact dezelfde omstandigheden exact hetzelfde zou moeten gedragen, en dit is zelfs in de experimentele situatie niet het geval. Premisse 3 wordt bijgevolg ook dubieus, dus ook dit argument is implausibel.

Door inductie en de probabilistische natuur van experimenten is het zeer moeilijk, al dan niet onmogelijk om het (hard) determinisme aan te tonen.