

Fri vilje og ultimativt ansvar:
**Et forsvar af Robert Kanes libertarianisme imod Derk Perebooms,
Gregg Carusos og Robert Sapolskys *free will skepticism***

Af

Søren Rohmann-Sønderby

Kandidatafhandling i Anvendt Filosofi
Aalborg Universitet

Studienr.: 20216295

Vejleder: Antje Gimmler

Dato for aflevering: 1. september 2025

Antal normalsider: 80

Indholdsfortegnelse:

Forside:	1
0.1. Indføring i specialets ærinde:	4
0.2. Problemformulering:	6
0.3. Kort om den nyere filosofihistoriske baggrund for specialet:	7
0.4. Om præmisserne for at tilgå specialets emne:	8
0.5. Struktur for specialet:	9
Del 1: Udlægning af Kanes libertarianisme:	11
1.1. Indføring i Kanes libertarianisme og dennes grundtræk:	11
1.2. Sellars' udfordring til Kane og teorien om <i>complex dynamical systems</i> :	13
1.3. Alternative muligheder (AP) og ultimativt ansvar (UR):	17
1.4. Self-forming actions:	19
Del 2: Viser moderne neuroscience, at vi ikke har fri vilje?	23
2.1. Viser moderne neuroscience, at fri vilje er en illusion? – En kritisk analyse af Benjamin Libets eksperiment fra 1983:	24
2.1.1. Libets eksperiment fra 1983:	24
2.1.2. Carusos og Sapolskys tolkning af Libets forsøg og senere eksperimenter:	26
2.2. Meles analyse af Libets forsøg og senere forsøg:	30
2.2.1. Meles analyse af Libets forsøg:	30
2.2.2. Meles analyse af to nyere forsøg:	33
Konklusion på del 2:	35
Del 3: Forsvar af Kanes libertarianisme mod Galen Strawsons "Basic Argument":	36
3.1. Præsentation af Galen Strawsons "Basic Argument":	37
3.2. Forsvar af Kanes libertarianisme imod The Basic Argument:	40
3.2.1. Hvad menes der med <i>causa sui</i> i praksis – og er dette en rimelig betingelse for at være moralsk ansvarlig for sine valg og handlinger som menneske?	40
3.2.2. Kanes perspektiv på ultimativt ansvar synes at være mere brugbart og nyttigt:	43
3.2.3. Kanes libertarianisme synes ikke at være præget af det regresproblem, som Strawson peger på i The Basic Argument:	44
Konklusion på del 3:	46

Del 4: Har troen på fri vilje en moralsk skyggeside? – Om fri vilje, retfærdiggørelsen af straf og ”moralsk vrede”:	47
4.1. Om koblingen mellem troen på fri vilje og retfærdiggørelsen af straf:	48
4.1.1. Kritik af Perebooms, Carusos og Sapolskys tanker om fri vilje og retfærdiggørelse af straf:	53
4.2. Om troen på fri vilje og vores ”moralske vrede” mod andre mennesker:	59
4.2.1. Kritik af Perebooms, Carusos og Sapolskys tanker om fri vilje og moralsk vrede:	64
Konklusion på del 4:	68
Del 5: Afrundende refleksioner: Hvad er de dybere moralske implikationer af troen på fri vilje i Kanes forstand?	69
5.1. Menneskelige aktører er forfatterne til deres egen karakter gennem deres SFAs:	69
5.2. Troen på fri vilje underbygger vores iboende værdighed som menneskelige aktører:	73
Konkluderende betragtninger:	76
Bibliografi:	78

0.1. Indføring i specialets ærinde:

I løbet af de seneste årtier har vi set en voksende gruppe af filosofiske og offentlige intellektuelle tænkere, der bredt kan forstås som *free will skeptics*. Disse tænkere, der favner fagfilosofiske stemmer samt evolutionsbiologer, hjerneforskere og meget mere, ser kort fortalt fri vilje som en illusion, altså som noget, der ikke reelt findes, men som vi af forskellige grunde alligevel har en hang til at tro på eksistensen af i vort daglige liv.

Et eksempel på denne *free will skepticism* ses i bogen *Determined: The Science of Life without Free Will* (2023) af Robert M. Sapolsky, professor i biologi og neurologi ved Stanford University. I bogen argumenterer Sapolsky for, at alle vores valg og handlinger kan forklares på basis af biologiske, neurologiske, miljømæssige, kulturelle og historiske faktorer, der alt i alt er uden for individets kontrol, og at det på denne baggrund er meningsløst at sige, at vi skulle være indehavere af en fri vilje, idet alle vore valg og handlinger hermed kan siges at være determinerede af forudgående faktorer og forhold i verden. Andre eksempler på fremtrædende *free will skeptics* de senere år er den verdensberømte fysiker Stephen Hawking (2012), psykologerne Stephen Pinker (2018) og Daniel Wegner (2002), biologen Jerry Coyne (2018) og den kendte israelske historiker og forfatter Yuval Noah Harari (2018). Herudover støttes *free will skepticism* som nævnt også af en del filosofiske røster, hvoraf nogle af de mest indflydelsesrige i de seneste årtier har været Galen Strawson (2013), Derk Pereboom (2014) og Greg Caruso (2012).

Dette speciale er skrevet på basis af en *grundlæggende skepsis over for denne free will skepticism*, der synes at være mere og mere fremherskende i tidens offentlige og bredere intellektuelle debat. Nærmere bestemt er det mit ønske med dette speciale at undersøge og forsøge at forvare et i offentligheden nok ikke helt så kendt, i min optik betydningsfuldt og interessant perspektiv i debatten om fri vilje, nemlig den nyligt afdøde amerikanske filosof Robert B. Kanes position. Robert Kane har gennem de seneste ca. fem årtier forsøgt at forfine, udvikle og bestemme en særlig libertarianistisk¹ position i debatten om fri vilje, og i min egen udforskning af debatten om fri vilje er jeg blevet mere og mere optaget af Kanes perspektiv.

Mere konkret ønsker jeg at bidrage til den aktuelle debat om fri vilje ved at tage udgangspunkt i Kanes posthumt udgivne værk *The Complex Tapestry of Free Will: A Philosophical Odyssey* fra august/september 2024 – et værk, som rummer en nuancering samt et forsvar af grundsynet i hans tidligere hovedværk *The Significance of Free Will* fra 1996. Nærmere bestemt vil jeg have

¹ Se del 1 for begrebsafklaring.

fokus på Kanes teori om såkaldte "self-forming actions" (SFAs) og hans begreb om "ultimativt ansvar" (*ultimate responsibility*), og hvordan disse to begreber for Kane er intimt forbundet med selve forståelsen af fri vilje som "*the power to be the ultimate source and sustainer to some degree of one's own ends or purposes*" (Kane, 2024, s. 10). Jeg vil mene at dette er en interessant definition af fri vilje, hvis betydning for os som mennesker jeg er interesseret i at forstå lidt nærmere.

Målet med specialet er på denne baggrund at forsøge at indkredse og artikulere nogle af de nærmere implikationer af Kanes libertarianistiske position – og herudover vil jeg forsvare Kanes position imod en række af de indvendinger, som denne er blevet mødt med i løbet af de seneste årtiers debat. I denne forbindelse har jeg valgt at have særlig fokus på filosofferne Derk Pereboom og Gregg Caruso og biologen Robert Sapolsky. Jeg finder disse tre tænkere særligt relevante som opponenter til Kane, idet de er repræsentanter for dét, som ofte kaldes for "optimistisk *free will skepticism*". Dette indebærer (1), at de ikke tror på, at vi kan siges at have fri vilje som menneskelige aktører, da der synes at være for stærke *videnskabelige* og *filosofiske* argumenter imod dette. På denne baggrund mener Pereboom, Caruso og Sapolsky, at vi bør kaste tesen om menneskets frie vilje på historiens mødding; og videre mener de, at vi hermed også må bortkaste idéen om, at vi som mennesker skulle have såkaldt "*moral responsibility in the basic-desert sense*", dvs. den form for moralsk ansvar, der ville betyde, at vi fortjente ros eller dadel, belønning eller straf for vores handlinger, *alene* fordi vi havde udført disse af vores egen frie vilje.²

Endvidere mener Pereboom, Caruso og Sapolsky på denne baggrund, (2) at det alt andet lige ville være *godt for os*, hvis vi forsøgte at besinde os på det forhold, at vi *ikke* er udstyret med fri vilje og rent faktisk bestræbte os på at leve vore liv og indrette vores samfundsinstitutioner i overensstemmelse hermed. Sagt på en anden måde mener de alle tre, at troen på fri vilje faktisk synes at have en moralsk skyggeside eller slagside – og at der derfor er grund til at være *optimistisk* angående, hvordan en verden uden troen på fri vilje ville se ud (deraf navnet "optimistiske skeptikere"). Som Pereboom³ under henvisning til Spinoza bemærker, synes *free will skepticism* som en filosofisk doktrin at "*contribute to the social life insofar as it teaches us to hate no one, to disesteem no one, to mock no one, to be angry at no one*" (Spinoza, 1992[1677], s. 490). Derudover mener de, at det vil være muligt at bevare betydningen af og troen på fænomener såsom kærlighed, kreativitet, venskab, tillid, troværdighed, tilgivelse og meget mere,

² Se del 4 for nærmere definition af denne form for ansvarlighed.

³ (Pereboom, 2014, s. 181).

også selvom vi giver slip på tanken om, at vi som menneskelige aktører skulle have fri vilje. På denne baggrund konstaterer Pereboom samlet set, at “*a conception of life without this sort of free will would not be devastating to morality or to our sense of meaning in life, and in certain respects it may even be beneficial*” (Pereboom, 2007, s. 85).

Jeg vil mene, at der er nogle enormt spændende betragtninger at finde i Perebooms, Carusos og Sapolskys skrifter. Min overordnet tese i specialet vil dog være, at det synes at være muligt at give et forsvar af Kanes libertarianistiske position imod en række af deres mest centrale indvendinger imod hans position. For det første vil jeg imod Caruso og Sapolsky mene, at moderne neuroscience ikke har påvist at vi rent faktisk ikke kan siges at have fri vilje som menneskelige aktører.⁴ Og for det andet vil jeg mene, at Kanes teori godt kan forsvares imod en række mere *tekniske filosofiske indvendinger*, eksempelvis Galen Strawsons berygtede ”Basic Argument”.⁵

Endelig vil jeg mene, at Kanes libertarianistiske position præsenterer os for et filosofisk perspektiv, det alt andet lige synes at give anledning til en mere opbyggelig måde at se på os selv og hinanden på end Perebooms, Carusos og Sapolskys *free will skepticism*. Med andre ord vil jeg mene, at vi tilsyneladende har meget at miste ved at give køb på troen på fri vilje, som Kane forstår denne egenskab.⁶

På baggrund heraf er det min samlede overbevisning, at det i hvert fald *synes* at være muligt at give et fyldestgørende forsvar af Kanes position imod Perebooms, Carusos og Sapolskys forskelligartede indvendinger og kritikpunkter imod denne – og på basis af disse indledende overvejelser og tanker har jeg valgt at tage udgangspunkt i følgende problemformulering:

0.2. Problemformulering:

Hvilke implikationer har troen på fri vilje i Kanes forståelse heraf for vores syn på os selv og hinanden som mennesker – og er det muligt at forsvare Kanes libertarianisme mod en række af indvendingerne imod denne artikuleret af Derk Pereboom, Gregg Caruso og Robert Sapolsky?

⁴ Se del 2.

⁵ Se del 3.

⁶ Se del 4 og 5.

0.3. Kort om den nyere filosofihistoriske baggrund for specialet:

Dette speciale behandler som nævnt emnet fri vilje – et emne, der har meget dybe filosofihistoriske rødder, og som videre kan siges at være ét af de mere omdiskuterede filosofiske emner i nyere tid, især i den analytiske filosofi fra midten af det 20. århundrede og frem.

Mere specifikt kan man ifølge Derk Pereboom sige, at vi fra ca. 1960'erne og frem har set en række filosofiske nybrud i debatten om fri vilje. Disse nybrud er (1) P.F. Strawsons (1962) artikel "Freedom and Resentment", der introducerede begrebet om *reaktive attituder*,⁷ hvilke kan forstås som "*special emotion-laden responses to which human beings are naturally prone in their interactions with one another*" (McGeer, 2012). Strawsons pointe er kort sagt, at netop disse reaktive attituder udgør en del af kernen af den klassiske og moderne debat om fri vilje, og ifølge Pereboom er diskussionen af disse reaktive attituder blevet en meget essentiel del af den filosofiske debat om fri vilje i årtierne efter udgivelsen af Strawsons artikel. Herudover har perioden fra 1960 og frem ifølge Pereboom været præget af introduktionen og diskussionen af såkaldte "Frankfurt-style cases" (1969), dvs. en række hypotetiske cases, som udfordrer tanken om, at dét at have alternative muligheder skulle være nødvendigt for at være moralsk ansvarlig for en given handling. Og endelig har perioden været præget af Peter van Inwagens (1983) "consequence argument",⁸ der er et mere principielt argument for inkompatibiliteten af determinisme og fri vilje – og som dermed understøtter bl.a. Kanes position.

Samlet set mener Pereboom, at disse nybrud "*make the subsequent debate very different from the way it was before 1960*", og det overordnede landskab i debatten om fri vilje er anderledes som følge af introduktionen af disse filosofiske bidrag.⁹ Med andre ord er det Perebooms påstand, at meget af den filosofiske debat fra ca. 1960'erne og frem har været kendetegnet ved, at nyere filosofiske tænkere har måttet forholde sig til de ovenstående og øvrige filosofiske bidrag, og at nye filosofiske positioner og tanker er blevet udviklet på basis heraf – og at Kanes libertarianistiske position også kan siges at være et eksempel herpå.

Endvidere kan man ifølge Kane sige, at debatten om fri vilje i løbet af de seneste ca. fire-fem årtier har været kendetegnet ved, at en række videnskabelige discipliner er blevet en integreret del af den bredere debat om fri vilje i vores tid; og at vi som filosoffer således ikke længere har en særskilt eller privilegeret position i denne debat, men så at sige må begrunde vore idéer som

⁷ Jeg kommer i specialet især ind på *moralsk vrede* som en reaktiv attitude og dennes sammenhæng med troen på fri vilje. Se del 4.

⁸ For et nærmere indblik i dette argument, se (Inwagen, 1983, s. 16).

⁹ Jf. (Pereboom, 2016) – hør fra 9.00 og frem i interviewet.

deltagere på et intellektuelt felt, hvor mange forskellige fagligheder brydes, mødes og indgår i en mere kompleks dialog med hinanden. På denne baggrund mener Kane, at vi som filosoffer i stigende grad må forholde os til øvrige faglige perspektiver, som på forskellige måder både supplerer og udfordrer den filosofiske tilgang til fri vilje som emne og fænomen – og videre mener Kane, at tidens filosoffer gør klogt i at forsøge at sikre og etablere et videnskabeligt fundament for deres filosofiske postulater om fri vilje.¹⁰

På baggrund af disse nyere ”trends” i debatten om fri vilje er det målet med specialet at give et mindre bidrag til den aktuelle filosofiske diskussion ved at give et forsvar af Kanes libertarianisme mod nogle af Carusos, Perebooms og Sapolskys indvendinger mod denne. Mere specifikt er mit valg faldet på Kane, da han i løbet af de seneste årtier har været en indflydelsesrig tænk, som af astrofysikeren og filosofen Robert Doyle er blevet kaldt ”*the acknowledged dean of the libertarian philosophers actively writing on the free will problem*” (Doyle, 2011, s. 295), og da jeg som nævnt er interesseret i at undersøge implikationerne af hans forståelse af fri vilje for vores syn på os selv og hinanden som mennesker.

0.4. Om præmisserne for at tilgå specialets emne:

Man kan sige, at én af de ting der gør debatten om fri vilje relevant for os er, at vi *føler eller fornemmer*, at vi har en fri vilje forstået som denne magt til at være den ultimative kilde til og skaber af vores egne målsætninger og formål som personer; og videre kan man sige, at vi med Galen Strawsons ord føler, at vi har ”*radical, absolute, buckstopping up-to-me-ness in choice and actions*” (Strawson G. , 2004, s. 380), dette i hvert fald i alle de situationer, hvor vi skal træffe et konkret og bevidst valg om at gøre noget frem for noget andet i vores liv.

Jeg vil mene, at denne følelse eller fornemmelse af at have fri vilje er selve grundlaget for vores filosofiske behov for at *begrunde* vores tanke om os selv som frie skabninger, der selv kan definere vores målsætninger og formål og bestemme os for handle på bestemte måder i vores liv. Kort sagt kan man altså sige, at dette er det ”før-filosofiske” niveau af debatten – vi synes at have en *erfaring* af, hvad det vil sige at have fri vilje, og dette før-filosofiske niveau synes i sig selv at være en del af dét, der strukturerer den filosofiske debat om fri vilje i litteraturen.

Man kan sige, at Kanes libertarianisme i denne henseende kan forstås som et forsøg på at sammenflette (1) vores umiddelbare erfaring af at have en fri vilje og at kunne træffe frie valg i

¹⁰ Dette bestræber Kane sig på selv at gøre i sin filosofi, se del 1.

mange af livets situationer og sammenhænge og (2) vores videnskabelige forståelse af verden, således som denne kommer til udtryk i visse empiriske discipliner såsom biologi og neuroscience mv.¹¹ Min ambition i specialet er på dette grundlag at komme frem til en forståelse af, om det er en *rimelig* position at fastholde troen på fri vilje i Kanes forståelse heraf – dvs. om der er grund til at betragte Kanes libertarianisme som et både deskriptivt og normativt forsvarligt standpunkt i den bredere debat om fri vilje i vores tid.

Endelig er det en grundtanke i dette speciale, at dét at tro på fri vilje eller ej også er et spørgsmål, der har en vis *praktisk eller moralsk* relevans for os som personer – dvs. min overbevisning er, at vi faktisk vil ændre adfærd baseret på vores antagelser på dette område. Ud fra Sartres (1956) eksistentialisme kan vi f.eks. sige, at ”ond tro” er en form for selvbedrag, hvor individer benægter deres frihed og ansvar for deres valg, ofte ved at tilskrive deres handlinger til eksterne faktorer eller deterministiske kræfter. Denne benægtelse ses ifølge Sartre som et forsøg på at undslippe den angst og usikkerhed, der er forbundet med vores radikale frihed som mennesker.

Jeg vil mene, at Kane ville være meget enig i disse betragtninger, og at der på denne måde er en slags ”eksistentialistisk brod” eller undertone i Kanes libertarianistiske position; og videre kan man i denne forbindelse sige, at Kane selv kommer med nogle vurderinger af, hvordan tanken om vores frie vilje hænger sammen med vor *autonomi* og *værdighed* som aktører – hvilket er noget af dét, jeg håber at undersøge lidt nærmere i specialet.¹²

0.5. Struktur for specialet:

Med henblik på at bevare den ovenstående problemformulering har jeg valg at dele specialet ind i fem dele. I den første del giver jeg en udlægning og nærmere bestemmelse af Kanes libertarianistiske position.

I den anden del af specialet har jeg fokus på én af de mest brugte *empiriske* indvendinger mod Kanes teori, nemlig at forsøg i moderne neuroscience skulle have påvist, at vi slet ikke har fri vilje som menneskelige aktører.

I tredje del af specialet forsøger jeg at forsvare Kanes libertarianisme imod Galen Strawsons Basic Argument.

¹¹ Se del 1.

¹² Se del 5 for en dybere udfoldelse af disse tanker.

I fjerde del af specialet kigges der nærmere på spørgsmålet om, hvorvidt selve dét at tro på eksistensen af vores frie vilje kan siges at have nogle negative følgevirkninger, og om der på dette grundlag kan siges at være nogle *moralske eller pragmatiske grunde* til at forkaste selve troen på fri vilje. Pereboom, Caruso og Sapolsky argumenterer nærmere bestemt for, at det ville være gavnligt for os at bortkaste ideen om vores frie vilje og *"basic-desert moral responsibility"* som aktører – og to af deres argumenterer herfor er deres syn på (1) retfærdiggørelsen af straf og (2) legitimeringen af såkaldt "moralsk vrede". Jeg vil her undersøge, om Kanes libertarianisme tillader et lige så gunstigt eller måske mere farbart blik på disse størrelser.

Endelig kommer jeg i del 5 med nogle mere overordnede refleksioner over, hvad troen på fri vilje i Kanes forståelse heraf betyder for vort syn på os selv og hinanden som mennesker. Der afrundes efter disse dele med en samlet konklusion på specialets fund.

Del 1: Udlægning af Kanes libertarianisme:

I den følgende del af specialet gives en indføring i Kanes libertarianistiske position, som denne optræder i hovedværkerne *The Significance of Free Will* og *The Complex Tapestry of Free Will*. Kane mener, at de to værker deler samme grundsyn og komplementerer og supplerer hinanden, og at *Tapestry* indebærer en nuancering af visse teser og tanker i *Significance* – og sådan vil de også bruges her. Målet vil være at give en overbliksskabende præsentation af Kanes teori, således at denne kan forsvares mod Perebooms, Carusos og Sapolskys indvendinger i de øvrige dele af specialet.

1.1. Indføring i Kanes libertarianisme og dennes grundtræk:

Som Kane bemærker i indledningen til *The Significance of Free Will*, er problemet om fri vilje et ældgammelt problem i filosofien; og selvom Kane giver Susan Wolf ret i, at problemet om fri vilje¹³ må siges at være "*arguably the most difficult problem in philosophy*" (Wolf, 1990, s. vii), så tror Kane dog på, at "*our understanding of the great issues can modestly progress*" (Kane, 1996, s. 3), hvilket også gør sig gældende ift. problemet om fri vilje. Med en sådan fremgang forstår Kane først og fremmest, at det er muligt at forfine både vores argumenter og teorier på området, samt at vi kan udforske nye områder af og kortlægge nye stier i dét, som han kalder for den "labyrint", som debatten om fri vilje fra oldtiden til i dag kan siges at udgøre.

Nærmere bestemt mener Kane, at der fra ca. 1960'erne og frem er fremvokset en nyere debat om fri vilje i filosofien; og hans egen position må forstås som et forsøg på at udvikle og bestemme en mere gunstig og forsvarlig libertarianistisk position¹⁴ i denne debat. Med en "libertarianistisk" position forstår Kane i denne forbindelse en position, der hævder, (1) at fri vilje som fænomen er inkompatibelt med et helt igennem deterministisk univers, og (2) at vi som mennesker *har* fri vilje i én eller anden udstrækning – eller at vi i hvert fald kan argumentere for, at vi meget vel kunne have dette. I indledningen til *Tapestry* kommer Kane endvidere med følgende kommentarer om dét, han her kalder for "the problem of free will and necessity or determinism", og hvad vores bevidsthed om dette problem synes at udspringe af:

¹³ Herved forstås af Kane overordnet set problemet om, (1) hvorvidt vi som mennesker har fri vilje, (2) hvad denne rummer og forudsætter, og (3) hvad dette betyder for vort aktørskab og vor ansvarlighed og øvrige forhold.

¹⁴ Kane refererer til sin position som *agent-causal/event-causal libertarianism*, idet det antages at både aktører og mentale begivenheder spiller en kausal rolle ifm. frie valg. Se (Kane, 2024, s. 81;155).

I've often quoted the thirteenth-century Sufi Muslim poet and philosopher Jalal ad-Din ar-Rumi, who said, "there is a disputation that will continue till mankind is raised from the dead, between the necessitarians and the partisans of free will." The problem of free will and necessity or determinism of which he speaks has arisen in history whenever humans have reached a higher stage of self-consciousness about how profoundly the world may influence their behavior in ways unknown to them and not controlled by them. The rise of doctrines of determinism or necessity in the history of ideas is an indication that this higher stage of self-consciousness has been reached. People have wondered at various times whether their actions might be determined by Fate or by God, by the laws of physics or the laws of logic, by evolution, genes, or environment, by unconscious motives, by psychological or social conditioning, or, with the latest scientific threats from the neurosciences, by the activity of their brains of which they are not conscious.

Despite differences, there is a core idea running through all these historical doctrines of determinism, whether they are religious, secular, or scientific, that shows why many people have felt they are a threat to free will. They all imply that, given the past at any time and the laws governing the universe, there is only one possible future. Whatever happens is therefore inevitable, it cannot but occur, given the actual past and laws. Free will by contrast seems to imply that (i) the future is open, with multiple possible paths into the future, and (ii) it is sometimes "up to us," and no one and nothing else, which of these possible paths will be taken. (Kane, 2024, s. 2).

Som det fremgår af dette citat, mener Kane altså, at der umiddelbart synes at være en *konflikt* eller *kontradiktion* mellem på den ene side doktrinen om determinisme i dennes forskellige varianter og gevanter gennem historien og på den anden side vores mere intuitive eller gængse opfattelse af, hvad der ligger i at have *fri vilje* – nemlig (i) at fremtiden rummer alternative muligheder for os og (ii) at det i hvert fald i visse sammenhænge er "op til os", hvilken vej ind i fremtiden, vi vælger at gå. Det er altså denne umiddelbare eller øjensynlige konflikt mellem determinisme og fri vilje, der i høj grad er motiverende for Kanes forsøg på at udvikle en libertarianistisk filosofi, dvs. at hans libertarianistiske filosofi på mange måder må tænkes som en *reaktion* og et *svar på* denne øjensynlige konflikt mellem determinisme og fri vilje.

Videre bemærker Kane, at der er en særlig opfattelse af fri vilje, som synes at have været omdrejningspunktet eller udgangspunktet for de antikke og middelalderlige debatter om fri

vilje, nemlig tanken om fri vilje som *"the power to be the ultimate source and sustainer to some degree of one's own ends or purposes"* (Kane, 2024, s. 10) – en tanke om fri vilje, som ifølge Kane også er et frugtbart og forsvarligt udgangspunkt for den moderne filosofiske debat herom, og som videre synes at fastlægge eller indramme noget meget essentielt ved vores forståelse af os selv som frie væsner i verden. Navnlig understreger Kane, at det synes at være tilfældet, at det er en central del af vores selvbillede som mennesker, at vi rent faktisk er udstyret med en sådan fri vilje, da vi intuitivt *føler*, at vi selv er i stand til at definere de dybere målsætninger og formål og værdier, som vi ønsker at handle ud fra i livets forskellige situationer – og hermed mener Kane også, at der synes at gå noget tabt ifm. vores selvforhold som mennesker, hvis det viser sig at vi faktisk *ikke* har en sådan fri vilje.

Videre er det Kanes holdning, at der synes at være noget ved denne klassiske eller traditionelle begrebsliggørelse af fri vilje, som er inkompatibelt med universel determinisme, dvs. tesen om, at vi lever i et univers, hvor det med Peter van Inwagens ord er tilfældet, at *"there is at any instant exactly one physically possible future"* (Inwagen, 1983, s. 3). Problemet er ifølge Kane, at *hvis* vi levede i et sådant helt igennem deterministisk univers, så ville vi aldrig nogensinde kunne have handlet på en anden måde end den, som vi *faktisk* endte med at handle på i hver og én af vores livs samlede antal af situationer – og hermed ville vi aldrig kunne have påvirket vort eget livsforløb som aktører, og vi ville ikke kunne have bidraget til at forme vores egen karakter og vores viljes beskaffenhed som aktører gennem *self-forming actions*.¹⁵ Hermed mener Kane, at vi simpelthen bliver nødt til at forudsætte, at universet ikke er et helt igennem deterministisk sted, hvis der skal være plads til den ovenstående form for frie vilje, som den klassiske filosofiske debat synes at have kredset om.

Endelig mener Kane, at det er behørigt for moderne libertarianistiske filosoffer at undersøge, om en sådan libertarianistisk fri vilje kan siges at være kompatibel med et moderne videnskabelige verdensbillede, dvs. om den rent faktisk kan forsvares på et empirisk grundlag – og på hvilket grundlag vi nærmere kan svare herpå.

1.2. Sellars' udfordring til Kane og teorien om *complex dynamical systems*:

Dette leder os videre til næste punkt i denne gennemgang af Kanes libertarianistiske position, nemlig hvordan dennes *empiriske* grundlag nærmere kan tænkes, og på hvilket måde

¹⁵ Se pkt. 1.4.

libertarianistisk fri vilje kan begrundes inden for rammen af det gældende videnskabelige verdensbillede. Men inden vi tager et nærmere kig på dette, er det hjælpsomt at tage i betragtning, hvad der oprindeligt fik Kane til at interesse sig for debatten om fri vilje.

Som Kane bemærker i *Tapestry*, opstod hans interesse for debatten om fri vilje i begyndelsen af 1960'erne, hvor konturerne af debatten om fri vilje i filosofien var anderledes og måske simple end fra ca. 1990'erne og frem. Det var dengang ifølge Kane antaget, at hvis man havde videnskabelige aspirationer, så var man en *kompatibilist*¹⁶ i debatten om fri vilje. Der var også *free will skeptics* eller *hard determinists*,¹⁷ påpeger Kane, dvs. filosoffer, som mente at determinisme ikke var kompatibel med fri vilje, men siden verden efter alt at dømme *var* deterministisk indrettet, så måtte dette betyde, at fri vilje slet ikke kunne eksistere i den verden, vi lever i. Og endelig var der ifølge Kane en tredje gruppe af filosoffer, som troede på eksistensen af en form for fri vilje, som var inkompatibel med determinisme – og disse filosoffer havde grundet deres store iver efter at godtgøre eksistensen af en sådan libertarianistisk fri vilje en stærk tilbøjelighed *“...to appeal to some kind of mysterious forms of agency or causation to make sense of it – to uncaused causes, immaterial minds, noumenal selves outside space and time, prime movers unmoved”*, etc. (Kane, 2024, s. 3). Sådanne postulater blev ifølge Kane dengang alle tendentielt set som eksempler på dét, som P.F. Strawson berømt havde kaldt for *“the obscure and panicky metaphysics of libertarianism”* (Strawson P. F., 1962). Med andre ord var der altså ifølge Kane i 1960'erne en meget udbredt skepsis over for alle libertarianistiske positioner, og disse blev almindeligvis forbundet med obskure og dubiøse metafysiske postulater og spekulative filosofiske præmisser af forskellig karakter.

Som Kane forklarer i *Tapestry*, begyndte han at tænke over fri vilje lige efter, at Strawsons artikel blev udgivet – og hans daværende filosofiske vejleder Wilfred Sellars gav ham på dette tidspunkt den interessante udfordring at forlige en traditionel inkompatibilistisk eller libertarianistisk fri vilje med det moderne videnskabelige verdensbillede. Nærmere bestemt beskriver Kane denne udfordring som følger:

While Sellars granted that free will in some sense was an integral part of what he notably called the *manifest image* of the world, the world of our ordinary experience, he did not believe a traditional libertarian or incompatibilist free will

¹⁶ Dvs. én, der mener at fri vilje og determinisme er kompatibel.

¹⁷ Teknisk set er Pereboom, Caruso og Sapolsky *“hard incompatibilists”*, dvs. de mener at fri vilje er inkompatibelt med *både* determinisme og indeterminisme. Se (Pereboom, 2001, s. xiii-xxiii) (Caruso, 2012, s. 2-12) (Sapolsky, 2023, s. 10-14).

could be reconciled with what he called the *scientific image* of the world; and he challenged me to show otherwise. (Kane, 2024, s. 3).

Som Kane husker tilbage, troede han som ung filosof, at han kunne løse denne opgave på nogle uger – eller måske et semester. Men denne opgave skulle vise sig at blive startskuddet til mange årtiers refleksioner over, hvordan man egentlig kan forlige en traditionel libertarianistisk opfattelse af fri vilje med et mere moderne, videnskabeligt blik på den empiriske virkelighed, vi er indfældet i som menneskelige aktører.

Som det videre fremgår af *Tapestry*, mener Kane, at det aktuelt er et "*open scientific question*" (Kane, 2024, s. 143;268), om libertarianisme som en filosofisk position kan underbygges empirisk. Ikke desto mindre understreger Kane, at meget moderne videnskab inden for blandt andet neuroscience og biologi synes at pege i retning af, at det i hvert fald ikke er umuligt eller utænkeligt, at vi skulle være udstyret med en sådan libertarianistisk fri vilje – og på denne baggrund mener Kane, at det er en forsvarlig position at fastholde, at libertarianistisk fri vilje godt *kunne* eksistere i den virkelighed, vi er indfældet i som mennesker.¹⁸

I denne forbindelse knytter Kane mere specifikt an til såkaldt "dynamic systems theory" – et forskningsparadigme med rødder i matematik, som bruges til at beskrive en række komplekse fænomener inden for mange forskellige domæner.¹⁹ Ifølge Kane er den grundlæggende tanke bag "dynamic systems theory", at der i naturen eksisterer en lang række såkaldte *complex dynamical systems*, herunder også mange levende organismer, og at disse systemer er kendetegnet ved, at "*emergent capacities arise as a result of greater complexity*" (Kane, 2024, s. 81). Kort sagt er det altså Kanes tanke, at fri vilje skulle være en *emergent egenskab* ved menneskelige aktører set som sådanne *complex dynamical systems*. Ved "emergente egenskaber" forstår Kane her en række egenskaber, som udspringer af de *systemiske processer* udført af et sådant komplekst dynamisk system, og som kommer til udtryk gennem samspillet mellem systemets dele og delprocesser uden at kunne forklares eller forstås fyldestgørende på delenes eget niveau.

Videre mener Kane, at der er en kobling mellem teorien om sådanne *complex dynamical systems* og forskellige modeller for at forstå menneskets kontrol over vores egne valg inden for

¹⁸ I denne sammenhæng understreger Kane, at han ser sin tilgang som et alternativ til Kants *transcendental* filosofiske måde at forsøge at begrunde libertarianistisk fri vilje på, der forudsætter hypotesen om en noumenal sfære – se (Kane, 2024, s. 245). Se også (Kant, 1996[1788], Ak V: 94-100).

¹⁹ For nærmere info om dette forskningsparadigme, se bl.a. (Mitchell, 2009).

rammen af den moderne neurovidenskab. I denne forbindelse kommer han med følgende kommentar:

Such complex systems exhibit *teleological guidance control* (TGC) when they tend through feedback loops and error correction mechanisms to converge on a goal (called an attractor) in the face of perturbations. Such control, as neuroscientist Marius Usher argues (2006), *is necessary for any voluntary activity*, and he interprets it in terms of dynamic systems theory, as I would as well. Neuroscientists E. Miller and J. Cohen (2001) argue that such cognitive (guidance) control in human agents stems from the active maintenance of patterns of activity in the prefrontal cortex that represent goals and the means to achieve them. These patterns provide signals to other brain structures whose net effect, as Miller and Cohen describe it, is to guide the flow of activity along neural pathways that establish the proper mappings between inputs, internal states, and outputs. (Kane, 2024, s. 82).

Som det ses af dette citat, er det altså Kanes overbevisning, at den neurologiske videnskabs betragtninger og modeller er relevante med henblik på at belyse egenskaberne ved mennesket som et komplekst dynamisk system – og at vi gennem brugen af vores præfrontale cortex (altså pandelappen) er i stand til at kommunikere med resten af vores hjernes strukturer og dermed kortlægge forholdet mellem inputs, indre tilstande og outputs i form af fx et frit valg om at gøre *X* frem for *Y*.

Som det også fremgår af citatet, er disse processer ifølge Kane forbundet med dét, som hjerneforskeren Marius Usher (2006) kalder for ”*teleological guidance control*” (TGC). Dette dækker ifølge Kane over den form for kontrol, der er nødvendig for at sikre, at vores adfærd faktisk er styret af en given eller mulig intention. Dette sikrer altså ifølge Kane kort sagt aktøren den *kontrol* i forbindelse med vores frie valg og handlinger, som gør, at vi som aktører med rette kan betragtes som den egentlige kilde til vores frie valg og handlinger i praksis. Alt i alt kan man altså hermed sige, at teorien om menneskelige aktører som *complex dynamical systems* udøvende *teleological guidance control* over deres frie valg og handlinger er Kanes svar på, hvordan menneskets frie vilje kan forstås inden for rammen af det moderne videnskabelige verdensbillede – eller i hvert fald visse tolkninger heraf i f.eks. biologi og neuroscience.

Endelig mener Kane, at der er brug for at antage at visse *indeterministiske hjerneaktiviteter* optræder i forbindelse med visse af vores frie valg, navnlig ifm. *self-forming actions*.²⁰ Denne indeterminisme kunne ifølge Kane eksempelvis optræde på et mikrofysisk eller kvantefysisk plan;²¹ og det kunne være tilfældet, at denne form for kvantefysiske indeterminisme kunne være med til at sørge for, at vi som aktører har *alternative valg- og handlemuligheder*²² i forbindelse med disse *self-forming actions*. Kane indrømmer i denne forbindelse at der er tale om en noget spekulativ videnskabelig hypotese, som meget vel kunne være forkert; men han argumenterer for, at det er nødvendigt at læne sig op ad sådanne videnskabelige påstande, hvis vi skal forsøge at begrunde tanken om libertarianistisk fri vilje på et naturalistisk grundlag. Kane understreger i denne forbindelse endvidere, at selvom mange filosoffer og videnskabsfolk aktuelt mener, at vi tilsyneladende lever i et helt deterministisk univers, så synes der at være stor uenighed herom blandt fagfolk, og det synes at være tilfældet, at *"way to soon to prejudge these empirical matters"*, og at *"it is far from clear today that universal physical determinism is true."* (Kane, 2007, s. 181). På denne baggrund mener Kane, at det meget vel kunne være tilfældet, at vi lever i et univers, hvor den form for indeterministiske beslutningsprocesser, som SFAs kræver, kunne være empirisk mulige ud fra det gældende videnskabelige verdensbillede. Og som Kane sammenfattende bemærker i denne forbindelse: *"I would take my chances with such an incompatibilist view rather than go to some revisionist or other view before I have to."* (Kane, 2024, s. 242).

1.3. Alternative muligheder (AP) og ultimativt ansvar (UR):

Vi har nu kigget på hvordan Kane mere overordnet karakteriserer sin libertarianistiske teori og det empiriske grundlag for denne – og det er da tid til at dykke dybere ned i de mere tekniske nøglebegreber, som også er væsentlige for at forstå hans position. Dette bringer os til næste punkt i Kanes teori, nemlig hans syn på og forståelse af begreberne *alternative muligheder* (*alternative possibilities, AP*) og *ultimativt ansvar* (*ultimate responsibility, UR*) og deres rolle i forbindelse med diskussionen af fri vilje som fænomen.

Som Kane bemærker, er AP og UR de to egenskaber eller "features", som man klassisk har forbundet med idéen om, at en aktør skulle have fri vilje i en given eller mulig situation eller sammenhæng. Sagt på en anden måde er det ifølge Kane en almindelig antagelse, at dét at have

²⁰ Se pkt. 1.4.

²¹ Se bl.a. (Kane, 2024, s. 59-60;134-135;242-243).

²² Se pkt. 1.3.

fri vilje indebærer eller synes at indebære, (1) at vi i det mindste på udvalgte tidspunkter i vores liv har forskellige valg- og handlemuligheder (AP), og (2) at det alene er op til os at afgøre, hvordan vi ønsker at vælge blandt disse valg- og handlemuligheder i forskellige scenarier eller situationer, dvs. at det synes at være *os selv og ikke nogen andre eller noget andet*, som er *ultimativt ansvarlige for* (UR) at vælge og handle på den måde, som vi måtte gøre i sådanne situationer.

Som Kane bemærker, er det betingelsen om alternative muligheder (AP), som har fået klart mest opmærksomhed og er blevet diskuteret mest i den moderne filosofiske debat, fx i forbindelse med spørgsmålet om, hvorvidt en given aktør i en given situation kan siges at være *moralsk ansvarlig* for en given handling, også selvom aktøren måske teknisk set ikke kunne have handlet anderledes i situationen.²³ Dog bemærker Kane, at AP-betingelsen efter hans mening isoleret et ikke udgør et tilstrækkeligt grundlag for at forsvare en libertarianistisk position i debatten om fri vilje – dvs. at *”focusing on alternative possibilities alone is too thin a basis on which to rest the case for the incompatibility of free will and determinism.”* (Kane, 2024, s. 11).

Det er ifølge Kane her, at betingelsen om *ultimativt ansvar* (UR) kommer ind i billedet – og overordnet betragter Kane UR som en *mere central betingelse for fri vilje end AP*, da UR knytter sig mere direkte til forestilling om fri vilje som *“the power to be the ultimate source and sustainer to some degree of one’s own ends or purposes”* (Kane, 2023, s. 4). Mere konkret mener Kane, at vi kan forestille os betydningen af dét at have ultimativt ansvar på følgende måde:

Its meaning can be captured initially by an image: when we trace the causal or explanatory chains of action back to their sources in the purposes of free agents, these causal chains must come to an end or terminate in the willings (choices, decisions, or efforts) of the agents, which cause or bring about their purposes. If these willings were in turn caused by something else, so that the explanatory chains could be traced back further to heredity or environment, to God, or fate, then the ultimacy would not lie with the agents but with something else. (Kane, 1996, s. 4).

Som det ses af dette citat, er det altså Kanes tanke, at aktører er *ultimativt ansvarlige* for en handling eller et valg, hvis den kausale kæde af begivenheder, som valget eller handlingen synes at udspringe af, kan spores tilbage til aktøren selv – frem for til forudgående forhold, andre aktører eller faktorer. Nærmere bestemt mener Kane, at vi kan formulere betingelsen om UR på følgende måde: *“to be ultimately responsible for an action, an agent must be responsible to*

²³ Jf. de såkaldte ”Frankfurt-style cases” (Frankfurt, 1969).

some degree for anything that is a sufficient reason (a sufficient condition, cause, or motive) for the action's occurring" (Kane, 2024, s. 12). Som det ses heraf, er det altså kort sagt Kanes tanke, at for at have UR for en handling, så må en aktør være helt eller delvist ansvarlig for alt dét, der er en tilstrækkelig grund til, betingelse eller motiv for handlingens frembringelse.

Som Kane videre kommenterer, er hans forståelse af UR på dette punkt inspireret af Aristoteles' kommentar om, at hvis en mand er ansvarlig for de slette gerninger, der udspringer af hans aktuelle karakter, så må han også på et tidligere tidspunkt i livet have været ansvarlig for at have formet den karakter, fra hvilken disse slette gerninger udspringer.²⁴ Således mener Kane, at UR dybest set handler om, hvordan vi danner vort eget selv og karakteregenskaber som mennesker, og om vi i én eller anden grad har *friheden til dette* – eller om vores selv og vores personlighed i sidste ende må siges at være et fuldkomment produkt af faktorer uden for vores kontrol. For kun hvis vi i én eller anden grad har kontrol med, hvordan vores egen karakter gennem livet formes eller skabes, kan vi alt andet lige også siges at have det *ultimate ansvar* for, hvordan vi selv ender med at formes og blive som mennesker.

Alt i alt man ifølge Kane hermed mere konkret sige, at dét at være ultimativt ansvarlig for vores egen karakter som personer også forudsætter, at vi gennem livet har mulighed for rent faktisk at præge, danne, forme og omforme vores karakter gennem visse af vores frie valg og handlinger i livet. Og sådanne formative eller karakterdannende handlinger kalder Kane for *self-forming actions*.

1.4. Self-forming actions:

Dette leder os til at allersidste nøglebegreb vi må komme ind på i denne gennemgang af Kanes libertarianisme, nemlig *self-forming actions* (SFAs).

Som Kane bemærker, er der her tale om en form for handlinger, hvor det *alene* er op til os selv og et produkt af vores egen frie vilje og frie og uafhængige beslutninger, hvad vi rent faktisk ønsker og ender med at gøre i den pågældende situation. Mere teknisk definerer Kane SFAs som *"the undetermined, regress-stopping voluntary actions (or refrainings) in the life histories of agents that are required if [the condition of Ultimate Responsibility] is to be satisfied."* (Kane, 1996, s. 75). Kane kommenterer i denne forbindelse, at selvom der her er tale om en mindre portion af den samlede mængde af handlinger i vort liv, så er det hans vurdering, at SFAs

²⁴ Se *Den nikomakhæiske etik*, 1114s13-22. (Aristoteles, 2000).

optræder hyppigere i livet, end man måske umiddelbart skulle tro; og han understreger, at der både kan være tale om helt hverdagslige handlinger såsom at beslutte sig for at tage opvasken eller blive på sofaen og mere skelsættende beslutninger såsom dét at tilsidesætte sine egne behov med henblik på at forsøge at redde et andet menneskes liv i en bestemt situation. Videre kan sådanne SFAs ifølge Kane forme os på måder, som både kan være belejlige og ubelejlige, ondartede og godartede, utilsigtede og tilsigtede mv. Endelig bemærker Kane, at en aktør i en given SFA altid vil være i stand til at handle på mindst én af to måder, dvs. at betingelsen om ”alternative muligheder” altid nødvendigvis må være opfyldt for, at en aktør rent faktisk kan udføre en SFA.

Mere generelt er det Kanes påstand, at SFAs optræder i følgende situationer:

Perhaps we're torn between doing the moral thing or acting from ambition, or between present desires and long-term goals, or we are faced with difficult tasks for which we have aversions, or have to make important life choices or have to exert willpower to keep prior commitments and resolutions rather than break them. In all such cases and many others, we may be faced with competing motivations and have to strive or make an effort or exert willpower to overcome the temptation to do something else we also strongly want. (Kane, 2024, s. 61).

Som det fremgår af dette citat, forestiller Kane sig altså, at SFAs kan finde sted i en række forskellige situationer i vores liv, og at der grundlæggende set er tale om en form for handlinger, hvor vi må yde en *indsats* for at træffe et bestemt valg i en bestemt situation, idet der er forskellige impulser, drifter eller interesser til stede i os, som så at sige hiver os i forskellige retninger. Eksempelvis kan der være tale om, at jeg må yde en indsats for at overkomme min hang til at blive liggende i sengen en morgen, idet jeg har aftalt med min ven at mødes eller lovet mig selv at løbe en rask tur som det første for at få en god start på dagen. Et andet eksempel på en SFA kunne være en lille dreng, som får at vide af sin mor, at han ikke må tage flere småkager fra dåsen, fordi der skal være nok til gæsterne senere på dagen; men da moderen er optaget af noget andet i fem minutter, fristes drengen til at tage en småkage mere alligevel – og der opstår derved en splittelse i hans vilje (Kane, 2024, s. 88). Skal han tage en småkage mere for at tilfredsstille sit umiddelbare begær, eller skal han behovsudskyde for at modtage mors anerkendelse, billigelse og kærlighed? Dette vil være et øjeblik, hvor drengen i nogen grad er med til at forme og danne sin egen karakter, hvilket vil ske gennem en fri handling.

Som Kane kommenterer, er der i SFAs generelt tale om "*soul-seaching moments of self-formation*", hvor vi er "*of two minds, and yet one and the same person*". Sagt på en anden måde er det altid tilfældet i en given SFA, at vi ikke helt kan adskille os selv fra de værdispørgsmål og overvejelser, vi har – og det føles som om, at vi i disse situationer i en vis forstand *vælger*, hvem vi vil være som mennesker. Kane formulerer også denne pointe på den måde, at der i SFAs optræder eller åbenbarer sig en form for *flertydighed* i vores værdisæt – og at vi i al fald for en stund er ambivalente, dvs. vores vilje "vender flere veje", og vi føler en dyb *splittelse* i os selv, idet vi på ét eller andet intuitivt plan ved, at vi her er med til at formere vores egen karakter. Af disse grunde kalder Kane også "self-forming actions" for "will-setting actions", dvs. det er de øjeblikke i vores liv, hvor vores vilje ikke på forhånd er "*set one way*", men i stedet må vi selv *gennem* vores valg og handlen være med til at *definere* vores viljes støbning, idet vores valg på én eller anden vis vil eller kan danne grobund for, hvordan vi med en vis sandsynlighed, naturlighed eller selvfølgelighed vil handle i tilsvarende situationer fremadrettet.

Et godt eksempel på dette er ifølge Kane den lille dreng i den ovenstående case, idet denne dreng alt andet lige vil være mere tilbøjelig til igen at tage småkager fra dåsen, uden mor ser det, hvis han lykkes med dette den første gang, og hvis mor ikke finder ud af det og skælder ham ud for det eller gør andet for at korrigere hans adfærd. På denne måde mener Kane, at drengen over tid vil udvikle eller danne et "*backlog of self-formed character*" (Kane, 2024, s. 88), dvs. at drengen over tid vil blive *mere* ultimativt ansvarlig for sin adfærd i sådanne situationer, givet at han har fået *for vane* at handle på en bestemt måde i disse situationer, og hans egen vilje herved er ved at blive formet. Dvs. hvis drengen stjæler småkager fra dåsen fem gange frem for én gang, vil han ifølge Kane alt andet lige være *mere* ansvarlig for denne snu adfærd den femte gang end den første gang. Drengen er ifølge Kane gennem sin vedvarende handlen med til at definere sin egen karakter i én eller anden grad; og selvom denne karakter selvfølgelig altid defineres i *kombination* med visse baggrundsmæssige faktorer og øvrige forhold, som vil være givne i situationen og uden for drengens kontrol, så har drengen *over tid* én eller anden grad af *ultimativt ansvar*, ikke blot for hvordan han handler, men også for hvordan han *ér*, dvs. den person, som han gennem sine valg og handlinger beslutter sig for at blive. Det er kort sagt dette, der ifølge Kane er den helt essentielle pointe gemt i begrebet om *self-forming actions*, og som forbinder det til spørgsmålet om *ultimativt ansvar*.

Et tredje eksempel på en SFA er en ingeniør, der er på vej ud af et alkoholmisbrug, men som er presset over sin jobsituation, og som også oplever knas i sit ægteskab. En sen aftenime føler han derfor en trang til at ty til flasken på ny, men der er også en stemme i ham, der siger, at han må

lade være med at drikke alkohol, da dette blot vil vanskeliggøre hans livssituation og forværre alle hans problemer på sigt (Kane, 1996, s. 126). Et fjerde eksempel på en SFA er en forretningskvinde, som på vej til et vigtigt karrierefremmende møde pludselig opdager et nødlidende offer for et overfald i en nærliggende gyde – og spørgsmålet er da, om hun blot skal fortsætte til sit møde eller droppe mødet for at hjælpe denne person i nød. Og et femte eksempel på en SFA brugt af Kane er følgende case:

Consider a young woman, call her Claudia, who is about to graduate with honors from a law school and who is deliberating about which of two attractive job offers to accept. One is with a large corporate law firm in a big city, the other a smaller, up-and-coming, but less prestigious firm in a smaller city near where she grew up and hence to her family and many friends. She is torn because each firm has features that are deeply attractive to her. On the one hand, people at the smaller firm are more friendly. On the other hand, she is extremely ambitious and the chance to be part of this prestigious firm in a large city is very attractive to her. She is a small-town person with big-time ambitions. (Kane, 2024, s. 69).

Som det fremgår af citatet, viser denne case os, at Claudia ikke er *et simpelt væsen med en mere endimensionel vilje og et helt igennem uproblematisk og nemt forhold til sine omgivelser* – men som Kane kommenterer, er dette helt generelt netop *”the kind of complexity needed for genuine self-formation and hence free will rather than merely freedom of action.”* (Kane, 2024, s. 70). Dvs. at Kanes antagelse er altså helt generelt, at der er brug for en vis kompleksitet i vores viljes struktur, hvis vi skal være i stand til at forholde os til vores egne værdier og valgmuligheder på den måde, som kommer til udtryk gennem livets mange forskellige SFAs. Således kan man også sige, at Kane mener, at den usikkerhed, tvivl og ambivalens, som vi føler og fornemmer i livets mange SFAs, faktisk er en iboende og uundgåelig del af selve dét at have en fri vilje som menneske.

Del 2: Viser moderne neuroscience, at vi ikke har fri vilje?

I den ovenstående del af specialet har jeg givet en overordnet præsentation af Kanes libertarianistiske position, som den optræder i hans hovedværker på dette område. Alt i alt vil jeg mene, at Kanes libertarianisme synes at være en yderst interessant position i vores tids debat om fri vilje, og at der synes at være en logisk overensstemmelse mellem de forskellige dele af hans position og de begreber, som denne baserer sig på, herunder "self-forming actions", "complex dynamical systems", "ultimativt ansvar", "alternative muligheder" osv.

Ikke desto mindre består den store test for Kanes libertarianisme naturligvis i, om denne position kan forsvares imod en række af de indvendinger og kritikpunkter, som den er blevet mødt med i de seneste årtiers debat herom – og i denne forbindelse ønsker jeg i den indeværende del af specialet at kigge nærmere på dét, der nok må siges at være den mest udbredte *empiriske* indvending mod Kanes teori og øvrige teorier, der hævder eksistensen af fri vilje, nemlig postulatet om, at forsøg inden for moderne neuroscience skulle have påvist, at vi rent faktisk ikke kan siges at være udstyret med en fri vilje som menneskelige aktører. Er det med andre ord tilfældet, at forsøg inden for moderne neuroscience skulle have påvist, at fri vilje er en myte, vrangforestilling eller illusion?

Med henblik på at komme med et velbegrundet svar herpå vil jeg i denne del af specialet kigge nærmere på nogle af de neurovidenskabelige forsøg, der har været mest betydningsfulde i forbindelse med drøftelsen af menneskets frie vilje. Mere specifikt vil jeg kigge på det mest berømte forsøg i denne forbindelse, nemlig Benjamin Libets klassiske forsøg fra 1983. Jeg vil i denne forbindelse først gennemgå Libets forsøg og hans egen tolkning af dette; og dernæst vil jeg kigge på, hvordan Sapolsky og Caruso tolker implikationerne af dette forsøg samt senere forsøg inspireret heraf. Endelig vil jeg under inddragelse af filosofen Alfred Mele forsøge at påvise, at selvom disse neurovidenskabelige forsøg er interessante som indspark i debatten om fri vilje, så kan vi *ikke* ud fra moderne neuroscience dokumentere eller demonstrere, at vi ikke har fri vilje som mennesker.

Alt i alt er det hermed håbet med del 2, at denne del viser, at Kanes teori om fri vilje i hvert fald på visse punkter godt *kunne* være i tråd med det moderne videnskabelige verdensbillede, som dette tegnes af moderne neuroscience. Jeg mener ikke, at del 2 hermed udgør et *komplet* forsvar af Kanes libertarianisme mod diverse videnskabelige indvendinger – dette ville kræve en

diskussion af øvrige videnskabelige discipliner, fx biologi og fysik.²⁵ Men mit håb er dog, at del 2 potentielt ville kunne indgå i et bredere forsvar af Kanes position mod sådanne forskelligartede videnskabelige indvendinger.

2.1. Viser moderne neuroscience, at fri vilje er en illusion? – En kritisk analyse af Benjamin Libets eksperiment fra 1983:

Det første, vi skal fokusere på i det følgende, er det mest berømte neurovidenskabelige eksperiment i den nyere tids debat om fri vilje – nemlig psykologen Benjamin Libets epokegørende forsøg fra 1983. Som hjerneforskeren Uri Maoz kommenterer i den fagfilosofiske podcast *The Free Will Show*, kan man sige, at dette var *”the experiment that launched a thousand studies”*, og videre påpeger Maoz, at Libets forsøg *”casts a large shadow to this day”*, og at megen forskning i emnet fri vilje inden for moderne neuroscience til stadighed er præget af dét, som Maoz kalder for *”the Libet paradigm”*, dvs. den forskning og litteratur, der er inspireret af og bygger videre på Libets fund.²⁶

2.1.1. Libets eksperiment fra 1983:

Benjamin Libet (1916-2007) var en berømt amerikansk psykolog, der havde sin praksis ved University of California, San Fransisco. Han publicerede i 1983 en meget indflydelsesrig artikel, som dernæst også gav anledning til publiceringen af yderligere forskningsartikler i 1980'erne og fremefter, der diskuterede og byggede videre på hans fund. Libet ønskede navnlig at bruge en *eksperimentel* tilgang til problemet om fri vilje – hvormed han ikke mente en empirisk undersøgelse af, *hvordan* fri vilje fungerer, men i højere grad en undersøgelse af, *hvorvidt* vi har fri vilje.

Libets tilgang var på mange måder meget simpel og tog udgangspunkt i at notere sig *rækkefølgen* af forskellige begivenheder relativt til hinanden som basis for at postulere noget om de kausale mekanismer involveret i beslutningsprocesser. Libets klassiske studie fra 1983 baserede sig mere specifikt på, at Libet og hans forskningsassister bad testpersoner om at udføre en meget simpel bevægelse – et klik med en computermus udført med pegefingern eller en bevægelse med håndleddet. På et hvilket som helst tidspunkt inden for en bestemt tidsramme

²⁵ Se i denne forbindelse Sapolskys kritik af Kanes teori om SFAs (Sapolsky, 2023, s. 235-238).

²⁶ Jf. (Cyr, Flummer, & Maoz, 2021) – hør fra 47:26 og frem.

kunne testpersonerne selv vælge at udføre denne bevægelse – testpersonerne skulle med andre ord blot udføre denne bevægelse, når de selv *følte* eller *mente*, at det nu var tid til at udføre denne bevægelse med fingeren eller håndleddet. Herudover indebar forsøget, at Libet bad testpersonerne om at registrere på et specialdesignet ur, hvornår de mærkede en intention om bevæge deres håndled/pegefinger; og mens de gjorde dette, foretog Libet og hans forskningsassistenter en såkaldt EEG-måling²⁷ af testpersonernes hjerneaktiviteter for at finde ud af, om der i deres hjerner kunne spores en forudgående årsag til eller forløber for initieringen af den pågældende bevægelse med fingeren eller håndleddet – hvilket Libet kaldte for et ”readiness-potential” i hjernen, forstået som *”a slow buildup of electrical potential recorded at the scalp”* (Schurger, Hu, Pak, & Roskies, 2021).

Libets tanke med forsøget var hermed kort sagt, at *hvis* fri vilje kan siges at eksistere, så måtte den bevidste beslutning om at bevæge håndleddet eller fingeren alt andet lige *gå forud for* dette såkaldte ”readiness-potential” i hjernen, eller man kan i hvert fald sige, at de to fænomener i så fald som minimum måtte optræde samtidigt. Men dét som Libet fandt i forsøget var, at dette ”readiness-potential” i gennemsnit kunne registreres ca. 550 millisekunder før testpersonens bevægelse af håndleddet eller pegefingeren og ca. 200 millisekunder før testpersonens registrering af intentionen om at bevæge håndleddet/fingeren. Libet konkluderede herudfra, at testpersonerne faktisk *var* bevidste om deres ønske eller intention om at bevæge håndleddet eller fingeren, før de udførte den pågældende bevægelse; men samtidig mente Libet, at den korrelerede elektriske impuls i hjernen (”the readiness-potential”) kunne registreres flere hundrede millisekunder inden testpersonernes egen registrering af deres ønske eller intention om at bevæge håndleddet/fingeren. Og hvis dette var tilfældet, ræsonnerede Libet, så måtte det også være tilfældet, at det *ikke* var testpersonernes bevidste beslutninger, der forårsagede de pågældende fysiske bevægelser – i stedet synes den egentlige årsag til initieringen af sådanne fysiske bevægelser at være *ubevidste* hjerneprocesser, som gik forud for deres bevidste intention om at bevæge deres krop.

Libet mente på basis af dette forsøg, at fri vilje ikke lader til at spille en rolle i forbindelse med testpersonernes beslutning om at påbegynde og udføre en fysisk bevægelse; og han mente, at fundene i dette laboratorieforsøg kunne generaliseres, således at disse fund alt andet lige måtte være gældende for *alle* fysiske bevægelser. Og givet at dét at træffe et frit valg synes at forudsætte, at dette valg må være et *bevidst* valg, så mente Libet, at forsøget synes at underminere idéen om

²⁷ ElectroEncephaloGram – en undersøgelse, der registrerer hjernens elektriske impulser i forskellige centre.

fri vilje, givet at det ikke var de bevidste intentioner hos testpersonerne, som var den egentlige årsag til initieringen af deres fysiske bevægelser.

Libet mente dog også på basis sine fund, at der var et tidsrum på ca. 100 millisekunder, hvor testpersonerne faktisk kunne udøve en form for "conscious veto", som ville forhindre bevægelsen med fingeren eller håndleddet i at blive påbegyndt. Sagt på en anden måde mente Libet, at selvom forsøget viste, at testpersonerne ikke bevidst kunne *initiere* nogen bevægelse med deres egen frie vilje, så indikerede forsøget også, at det var muligt for testpersonerne gennem en bevidst og intentionel handling at *stoppe* en bevægelse, som *ellers* ville være blevet påbegyndt, og at dette kunne finde sted ca. 0,1 sekunder efter forekomsten af "the readiness-potential".

Med andre ord kunne vi altså – ifølge Libets egen tolkning af forsøgets implikationer – forestille os, at hvis jeg er i færd med at vælge mellem at spise en is fra min fryser eller lade være, så har jeg alt andet lige et vindue på ca. 0,1 sekunder, hvor jeg kan stoppe initieringen af dét, der *ellers* ville være blevet til frembringelsen af en fysisk bevægelse – fx at åbne min kumrefryser med henblik på at tage en is. Således kan man altså sige, at vi ifølge Libets egen tolkning af sit forsøgs resultater har mulighed for at have en form for "free won't" i vores levede liv, men at dette i og for sig ikke er ensbetydende med, at vi skulle have "free will" – som Obhi og Haggard formulerer dette (Obhi & Haggard, 2004).

2.1.2. Carusos og Sapolskys tolkning af Libets forsøg og senere eksperimenter:

Vi har nu kigget på Libets forsøg fra 1983 – og spørgsmålet er da: Hvordan skal vi nærmere opfatte Libets forsøg – og har Libet virkelig ret i, at disse eksperimenter dokumenterer, at vi ikke har fri vilje, men allerhøjst "free won't"? Og hvad siger Caruso og Sapolsky til dette?

Som Caruso bemærker i bogen *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will* fra 2012, synes Libets eksperiment fra 1983 samt neurovidenskabelige forsøg udført sidenhen at udgøre et stærkt empirisk argument for, at vi *ikke* kan siges at have fri vilje som aktører. Nærmere bestemt mener Caruso, at Libets forsøg synes at vise, at "*the conscious intention cannot be the true cause of the action*" (Caruso, 2012, s. 194), da denne intention jo kommer *efter* forekomsten af "the readiness-potential" i testpersonernes hjerner. Og hvis dette er tilfældet, så er dette ifølge Caruso et stærkt vidnesbyrd om, at vi *ikke* har den from for basale kontrol med initieringen af vore egne valg og fysiske bevægelser, som ville være påkrævet for at hævde, at vi har fri vilje.

Som Caruso bemærker, har senere neurovidenskabelige forsøg bekræftet Libets fund, men der er også nogle forskere, som har sat spørgsmålstegn ved visse aspekter af Libets egen tolkning af sit studie. I denne forbindelse refererer Caruso til nogle forsøg udført af hjerneforskeren Patrick Haggard i samarbejde med øvrige forskere, hvilke havde til hensigt at efterprøve resultaterne af Libets forsøg (Haggard & Eimer, 1999) (Haggard & Libet, 2001) (Haggard, 2005). I sin diskussion af Libets forsøg og øvrige eksperimenter inspireret heraf kommenterer Haggard, at der synes at være ikke ét, men to relevante fænomener forbundet med forekomsten af en bevidst beslutning om at bevæge et håndled eller en finger, nemlig et *”Lateralized Readiness Potential”* (LRP) og så dét, som Libet kaldte for *”the readiness potential”* (RP). Haggard skriver nærmere herom:

The LRP has a particular psychological significance in situations where the subject must choose between a left- and a right-handed action: once the LRP has begun, the selection of which action to make must be complete. That is, by LRP onset the intention has progressed from abstract stage (‘Do something or other!’) to drive a specific movement (‘Do precisely this!’). (Haggard & Libet, 2001, s. 50).

Som det ses af dette citat, er det altså Haggards påstand, at der synes at være mindst to relevante fænomener forbundet med initieringen af en given fysisk bevægelse i de pågældende testforsøg – nemlig RP og LRP. Haggard mener mere specifikt, at de nyere forsøg udført af ham selv og andre forskere påviser, at RP hermed ikke er en *tilstrækkelig* årsag til testpersonernes initiering af fysiske bevægelser, som Libet ellers hævdede i forbindelse med sit oprindelige forsøg fra 1983. Men ikke desto mindre bemærker Haggard, at selvom LRP altså optræder efter RP, så er det jo stadig tilfældet, at LRP optræder *før*, at testpersonen bliver bevidst om sit eget ønske om at udføre en given fysisk bevægelse.

På denne baggrund mener Haggard samlet set, at: *”However a scientist looks at all of these data, the brain is going full speed ahead well before a person experiences the conscious intention of moving”* (Obhi & Haggard, 2004, s. 360). Med andre ord er det altså Haggards påstand, at selvom der er visse tekniske udfordringer forbundet med tolkningen af RP i Libets klassiske forsøg, så er selve *hovedpointen* med Libets forøg stadig gyldig; og vi kan på baggrund af Libets forsøg samt senere forsøg inspireret heraf alt i alt bekræfte teorien om, at *før-* eller *ubevidste hjerneaktiviteter* er den reelle årsag til, at testpersonerne i disse forsøg handler, som de gør.

Caruso erklærer sig enig i disse påstande og mener, at vi alt i alt må konkludere følgende på basis af de hidtil udførte neurovidenskabelige forsøg om fri vilje:

What do these findings mean for free will? Well, I maintain that if our conscious intentions do not actually cause or initiate action then we can conclude that, at least in cases of spontaneous voluntary behavior like those studied by Libet, conscious will is an illusion. Our conscious selves are not in the driver's seat as we typically believe. This is a devastating blow to our everyday understanding of free will. Libets discovery undermines the strongly held belief, which is supported by phenomenology and the way we experience our own voluntary acts, that we consciously cause or initiate action directly through our conscious intentions, decisions, and willings. (Caruso, 2012, s. 194).

Tilsvarende kommenterer Sapolsky i bogen *Determined: The Science of Life without Free Will* (2023), at Libets eksperiment og senere forsøg inspireret heraf synes at udgøre et stærkt empirisk argument for, at vi ikke har fri vilje som aktører, da de viser, at dét, vi opfatter som vores bevidste beslutninger, synes at være determineret af helt igennem ubevidste eller førbevidste hjerneaktiviteter. For at understøtte sin argumentation refererer Sapolsky ud over ovenstående eksperimenter bl.a. til forskningen udført af hjerneforskeren Itzhak Fried fra UCLA, som i forbindelse med sit arbejde med svært epileptiske patienter i 2011 lavede et Libet-inspireret forsøg med elektroder placeret på patienternes pandelapper, således at det kunne måles, hvornår flere millioner neuroner aktiveredes ved patienternes initiering af fysiske bevægelser. Her var resultatet, at neuronerne i hippocampus blev aktiveret ifm. forberedelsen af en fysisk bevægelse op mod nogle sekunder inden, at denne bevægelse fandt sted. Dvs. dette resultat synes ifølge Sapolsky ligeledes at bekræfte Libets fund (Fried, Mukamel, & Kreiman, 2011).

Hermed mener Sapolsky, at vi kan konkludere, at "*our brains decide to carry out a behavior before we think that we've freely and consciously done so*" (Sapolsky, 2023, s. 36). Endvidere bemærker Sapolsky, at øvrige neurovidenskabelige studier kan hjælpe med at belyse dette emne. F.eks. viste et studie af epileptiske patienter ifølge Sapolsky, at hvis man aktiverer en bestemt hjerneregion relevant for at træffe beslutninger (parietallappen), så opfatter testpersonerne dette som et udtryk for, at de selv har truffet frit valg om at udføre en bevægelse – og dette uden rent faktisk at have løftet fingeren eller sågar spændt en eneste muskel i kroppen (Banks & Isham, 2009). Og et andet forsøg viste ifølge Sapolsky, at det omvendt også er tilfældet, at man kan aktivere visse hjernecentre forbundet med bevægelsen af en testpersons finger og hermed

få denne testperson til faktisk at bevæge fingeren, samtidig med at testpersonen nægter at have gjort dette! (Marchetti & Sala, 1998). På denne baggrund mener Sapolsky, at grundlæggende set ikke kan stole på vores egen fornemmelse af at have truffet en beslutning om at have udført en given bevægelse – eller sagt på en anden måde vil det ofte være tilfældet, at vi gennem påvirkning af de rette hjernecentre kan narres til at tro, at vi har udført en given bevægelse af vores egen frie vilje, uden at vi rent faktisk har gjort dette, og at vi omvendt kan have udført en bevægelse uden at ænse, at dette er noget, vi har gjort.

Disse og øvrige adfærdsmæssige forsøg indikerer ifølge Sapolsky samlet set, *"how our conscious sense of choosing can be fickle and subjective."* (Sapolsky, 2023, s. 25), og han mener, at disse fund kan forstås i sammenhæng med de ovennævnte "Libet-style experiments". Alt i alt knytter Sapolsky på basis heraf følgende kommentar til den Libet-inspirerede forskning fra 1980erne frem til i dag:

Collectively, what does this Libet literature, starting with Libet, show? That we can have an illusory sense of agency, where our sense of freely, consciously choosing to act can be disconnected from reality. We can be manipulated as to when we feel a sense of conscious control; most of all, this sense of agency comes after the brain has already committed to an action. Free will is a myth. (Sapolsky, 2023, s. 26-27).²⁸

Endvidere bemærker Sapolsky, at Libets tanke om, at vi skulle have en mulighed for at udøve et "conscious veto" og dermed forhindre dét, der ellers ville være blevet initieringen af en given bevægelse, synes at være en meningsløs idé. Denne tanke om såkaldt "free won't" forudsætter ifølge Sapolsky, at man videnskabeligt set kunne identificere nogle neuroner, som var *"equally autonomous in preventing an act [as in carrying out that act]"* (Sapolsky, 2023, s. 36), men Libets studier viser jo netop, at sådanne neuroner *ikke* er et udtryk for hjerneaktiviteter, som vi har bevidst kontrol med, førend vi udfører eller påbegynder en given bevægelse. Problemet er med andre ord ifølge Sapolsky, at der nødvendigvis også ville være en form for "readiness-potential" forbundet med ethvert sådant "conscious veto", hvormed vi afbrød dét, der ellers ville være blevet til initieringen af en fysisk bevægelse. Og følgelig mener Sapolsky, at selve tanken om "free won't" neurologisk set synes at være lige så *"untenable"* som tanken om fri vilje, hvis vi tager udgangspunkt i Libets forsøg.

²⁸ Sapolsky understreger dog, at Libet-inspirerede forsøg ikke i sig selv kan vise, hvor *intentioner* kommer fra, og at det derfor kræver et samspil mellem forskellige videnskaber endegyldigt at dementere tesen om fri vilje.

2.2. Meles analyse af Libets forsøg og senere forsøg:

Vi har nu set på, hvordan Libets forsøg og senere forsøg tolkes af Sapolsky og Caruso. Det store spørgsmål er da: Har Caruso og Sapolsky ret i deres vurdering af implikationerne af Libets forsøg? Ja, viser disse rent faktisk, at vi ikke har fri vilje som menneskelige aktører?

Jeg vil mene, at Libets eksperiment fra 1983 er interessant som et studie af, hvordan der er klare koblinger mellem visse neurologiske mekanismer og visse meget simple beslutninger udført i specifikke testsituationer, hvor de pågældende aktører ikke har noget at vinde eller tabe ved at træffe et bestemt valg. Dog vil jeg alt i alt mene, at Caruso og Sapolsky tager fejl, når de hævder at Libets forsøg og tilsvarende eksperimenter udført senere skulle udgøre en slags ”bevis” for, at vi ikke har fri vilje.

For at underbygge disse påstand vil jeg i det følgende inddrage Alfred Meles bog *Free: Why Science Hasn't Disproved Free Will* fra 2014. Jeg vil mene, at vi ud fra Meles argumentation som minimum har mulighed for at så tvivl om, hvorvidt moderne neuroscience skulle have påvist, at vi ikke har fri vilje.

2.2.1. Meles analyse af Libets forsøg:

I kapitel 2 af *Free: Why Science Hasn't Disproved Free Will* kommer Mele med en kritik af Libets forsøg. Mere overordnet mener Mele, at vi kan opstille Libets argument imod fri vilje på basis af sit oprindelige forsøg fra 1983 som følger:

1. The participants in Libet's experiments don't make conscious decisions to flex. (Instead, they make their decisions unconsciously about half a second before the muscle burst and become conscious of them about a third of a second later).
2. So people probably never make conscious decisions to do things.
3. An action is a free action only if it flows from a consciously made decision to perform it.
4. Conclusion: there are probably no free actions. (Mele, 2014, s. 24-25).

Mele mener dog, at der er problemer med alle præmisserne i dette argument – og at konklusionen derfor ”rests on very shaky ground” (Mele, 2014, s. 25).

For det første peger Mele på, at der synes at være problemer med den første af de ovenstående tre præmisser. Dette skyldes ifølge Mele, at dét, at en given bevægelse med kroppen skulle udspringe af en række ubevidste hjerneaktiviteter, ikke nødvendigvis betyder, at disse

bevægelser dermed ikke skulle være udtryk for et bevidst valg. Med andre ord er det altså Meles tanke, at det ikke nødvendigvis synes at udgøre et stort problem for os som aktører, at vores bevidste beslutninger så at sige kræver, at der er visse forudgående ubevidste hjerneaktiviteter som en del af deres kausale historik. Så længe vi overordnet set kan siges at være i stand til at initiere fysiske bevægelser gennem brugen af vores bevidste beslutninger, uanset om disse beslutninger i og for sig skulle være kausalt determinerede af forudgående ubevidste hjerneprocesser, så synes det altså ifølge Mele at være tilfældet, at vi *godt* kan sige, at testpersonerne i Libets forsøg udførte *bevidste* beslutninger om at bevæge deres håndled eller finger, etc. Herudover kommer Mele i denne forbindelse med følgende kommentar: *“If informed conscious reasoning leads to a decision that is then put into action, why should it matter if there is a bit of a lag—a couple hundred milliseconds—between when the decision is made and when a person becomes conscious of it?”* (Mele, 2014, s. 25). Med andre ord er det Meles påstand, at dét, at der er et lille tidsrum mellem aktiveringen af visse ubevidste hjerneaktiviteter og så vores bevidsthed om udførelsen af en fysisk bevægelse, ikke er ensbetydende med, at det ikke er os som aktører, der udfører den pågældende bevægelse som en fri handling. Vi behøver således ifølge Mele samlet set ikke være bekymrede for, at vi ikke har fri vilje, blot fordi det muligvis tager nogle hundredemillisekunder for os at blive bevidste om de beslutninger, som bliver igangsat af vores egen hjerne, før vi ved af det, *så længe vi kan stå inde for disse beslutninger*, når vi rent faktisk bliver bevidste om at vi har truffet dem.

For det andet peger Mele på, at det synes absurd at generalisere fra den type af beslutninger, der optræder hos testpersonerne i Libets eksperiment, til alle de beslutninger, som vi udfører i vores liv. Nærmere bestemt mener Mele, at der synes at være så stor forskel på de almindelige beslutninger i vores levede liv og de mere eller mindre tilfældige og meningsløse beslutninger optrædende i Libets eksperiment. Som Mele kommenterer:

As a participant in Libet’s experiment you would arbitrarily pick a moment to begin flexing in order to comply with part of your instructions. There is no need at all for conscious reasoning about what to do. But in real-life situations in which we do reason consciously about [what to do], the road to action seems quite different. It often seems a lot less arbitrary.[...] If we want to know whether conscious reasoning ever plays a role in producing decisions, we shouldn’t restrict our attention to situations in which people are instructed not to think about what to do. (Mele, 2014, s. 15-16).

Som det fremgår af dette citat, mener Mele, at testpersonerne i Libets forsøg ikke havde nogen særlige grunde til at træffe et bestemt valg; dvs. deres valg om, hvornår de skal klikke med musen eller bevæge håndledet, synes at være et fuldkomment vilkårligt valg uden relation til nogen motiver eller grunde til at handle på forskellige måder, som vi kender dette fra mange af livets sammenhænge. I Libets eksperiment ser vi ifølge Mele kort sagt, at det er helt *meningsløse beslutninger* vedrørende tidspunktet for visse bevægelser, som testpersonerne bliver bedt om at træffe; men i vores levede liv har vi ofte en langt mere kompliceret beslutningsproces, hvor vi overvejer forskellige handlemuligheder i kortere eller længere tid, og hvor der optræder en masse mere eller mindre vigtige *grunde* til at handle på forskellige måder.²⁹

På denne baggrund mener Mele videre, at det synes at være en meget forhastet konklusion, at Libets forsøg skulle have demonstreret, at bevidste beslutningsprocesser *aldrig nogensinde* skulle have nogen kausal indflydelse på vores faktiske valg og handlinger som aktører, og at vi dermed aldrig nogensinde træffer bevidste beslutninger om at gøre dette eller hint. Det kan ifølge Mele meget vel være tilfældet, at vi gennem videnskabelige forsøg såsom Libets eksperiment har mulighed for at afdække de ubevidste eller førbevidste hjerneaktiviteter forbundet med dét, som f.eks. for testpersonerne i Libets eksperiment fremtræder som frie beslutninger om at bevæge deres håndled eller finger – men dette er for Mele at se ikke ensbetydende med, at disse testpersoner hermed slet ikke har nogen mulighed for at præge deres egen adfærd i det pågældende eksperiment og for den sags skyld i livet generelt. Sagt på en anden måde synes det at være en grundløs og spekulativ påstand, at Libets eksperiment skulle vise, at vi i alle livets sammenhænge og situationer ikke skulle være i stand til at benytte vores bevidste overvejelser og tanker som et grundlag for at påvirke vores egen adfærd og træffe valg baseret på de grunde, som vi finder meningsfulde at handle ud fra i forskellige henseender.

Endelig påpeger Mele, at det synes at være en mærkelig antagelse, at vi *skal* være bevidste om alle hjerneprocesserne forbundet med vores forskelligartede beslutningsprocesser, for at vore bevidste beslutninger i sandhed kan siges at være *vores egne frie beslutninger*. Det er måske rigtigt nok, at vores bevidsthed bliver nødt til at spille en *vis* rolle for at et givent valg eller en given handling kan siges at være fri, mener Mele. Men spørgsmålet er dog ifølge Mele, om det er et meningsfuldt krav at sige, at vi bliver nødt til at have *fuld kontrol* med *alle* de førbevidste eller ubevidste komponenter i en beslutningsproces, for at denne proces kan lede til et givent eller muligt frit valg eller en fri handling i en bestemt sammenhæng.

²⁹ F.eks. ville dette være tilfældet med *self-forming actions* ud fra Kanes teori.

Navnlig kan man sige, at vi jo ikke normalt forudsætter, at det er nødvendigt at have kontrol med hver og én af vores egne muskelfibre eller celler i kroppen for at være ansvarlige for at udføre en given bevægelse – det synes at være tilstrækkeligt, at vi har en mere *overordnet* kontrol med vores egen krop og kan påvirke denne på en sådan måde, at vi kan udføre bestemte bevægelser i overensstemmelse med vores forestilling om, hvordan vores krop *kan* bevæge sig. Analogt hermed kunne man sige, at det synes at være en urimelig antagelse, at vi *skal* have kontrol med alle de aktiviteter i vores hjerner, som er involverede i initieringen af fysiske bevægelser, for at vi med rette kan siges at være ansvarlige for at have udført disse bevægelser ved brugen af vores egen frie vilje. Med andre ord kan man sige, at det virker til at være en unødvendig streng idé om kontrol, som er på spil i hele diskussionen af Libets eksperiment; ja, vi kan umuligt have kontrol med *alle vores ubevidste hjerneaktiviteter som aktører*, og hvis det er denne type af kontrol, som er nødvendig for at kunne siges at have en ”fri vilje”, så gøres fri vilje dermed til noget tilsyneladende praktisk og empirisk helt uopnåeligt, hvilket synes at være en absurd og urimelig måde at definere selve begrebet ”fri vilje” på.

Samlet set mener Mele på denne baggrund, at det er rimeligt at antage, at fri vilje godt kan være kompatibelt med en mere overordnet form for kontrol med vores beslutningsprocesser, uden at vi dermed kræves at være i stand til at kontrollere *alle* vores individuelle neuroner i forbindelse med frie valg. Alt i alt konkluderer Mele ud fra ovenstående refleksioner og analyse, at vi må sige, at *”Libet’s experiments come nowhere close to proving that we never act of our own free will.”* (Mele, 2014, s. 25).

2.2.2. Meles analyse af to nyere forsøg:

Vi har nu set, hvordan Mele kritiserer påstanden om, at Libets forsøg skulle have påvist, at vi ej har fri vilje som aktører. Men som Mele bemærker, er der udført mange neurovidenskabelige eksperimenter sidenhen. Hvad viser disse nyere forsøg?

I denne forbindelse kigger Mele nærmere på to nyere forsøg inspireret af Libets oprindelige eksperiment. Om det første af disse forsøg skrev videnskabsjournalisten Elsa Youngsteadt i en artikel fra 2008:

Researchers have found patterns of brain activity that predict people’s decisions up to 10 seconds before they’re aware they’ve made a choice.[...] The result was hard for some to stomach because it suggested that the unconscious brain calls the shots, making free will an illusory afterthought. (Youngsteadt, 2008).

Som Mele bemærker, blev forsøget, som Youngsteadt henviser til, udført af et team af hjerneforskere ledt af John-Dylan Haynes fra Humboldt Universitet – og disse forskere brugte såkaldt fMRI-scanninger til at måle testpersonernes hjerneaktivitet.³⁰ Med denne metode kunne forskerne måle ændringer i blodgennemstrømningen i hjernen. Testpersonerne blev hernæst bedt om at foretage en række simple beslutninger, mens deres hjerneaktiviteter blev målt med denne teknik. Deres muligheder bestod mere konkret i at trykke på én af flere knapper, som de frit kunne vælge imellem. På basis af fMRI-scanningerne mente forskerne, at de med ca. 60 % sikkerhed var i stand til at forudsige testpersonernes handlinger op mod 10 sekunder forinden udførelsen af disse handlinger.

Ifølge Mele siger dette eksperiment dog ikke så meget om, hvad fri vilje ér, og det siger heller ikke så meget om, hvorvidt kan siges at være i besiddelse af fri vilje eller ej. Nærmere bestemt kan man ifølge Mele sige, at der i dette eksperiment ligesom i Libets eksperiment er tale om en meget bestemt type af handlinger, som er fuldkomment afkoblet enhver praktisk eller normativ kontekst med relevans for vores liv som menneskelige aktører. Som Mele bemærker:

How similar is the arbitrary picking of a button to a decision to ask one's spouse for a divorce — or to change careers or start a small business — after protracted reflection on reasons for and against that decision? If arbitrary picking is not very similar to these other decisions, claiming that what happens in instances of arbitrary picking also happens in instances of complicated, painstaking decision making is a huge stretch. (Mele, 2014, s. 30).

Desuden påpeger Mele, at det ikke er særligt imponerende, at forskerne kunne forudsige testpersonernes gøren og laden med ca. 60 % sikkerhed – dvs. kun en my bedre, end hvis de havde baseret deres forudsigelse på at slå plat eller krone! Mele vurderer i denne sammenhæng, at disse 60 % kan skyldes, at testpersonerne i situationen nok overvejer, hvad de skal gøre frem mod valget af den ene eller anden knap, og at testpersonerne dermed har en mulig *bias*, dvs. tendens til at trykke på den ene knap frem for den anden. Men sådanne *biases* har ifølge Mele ikke noget at gøre med om testpersonerne rent faktisk kan handle på basis af deres frie vilje eller ej. Som Mele skriver: *“slight biases certainly don't seem to rule out free will. They don't dictate or compel behavior. They're nothing more than nudges.”* (Mele, 2014, s. 30). Dvs. vi kan ikke på basis af sådanne *biases* eller *nudges* konkludere noget som helst om, hvorvidt testpersonerne rent faktisk kunne træffe et frit valg i den pågældende situation, mener Mele.

³⁰ (Bode & Haynes, 2009).

Videre henviser Mele til ovennævnte studie med epileptiske patienter udført af Itzhak Fried i 2011.³¹ Selvom dette studie benytter sig af mere moderne teknologi til at udføre forsøget, så er dette studie dog ifølge Mele kendetegnet ved de samme principielle metodemæssige begrænsninger som Libets forsøg – og man kan sige, at dette er uundgåeligt, uanset hvad man bruger til at måle testpersonernes hjerneaktiviteter mv. (Mele, 2014, s. 38-39). Dét, at man kan bruge nyere videnskabelige metoder til at udføre laboratorieforsøg, er hermed ifølge Mele ikke ensbetydende med, at de oprindelige problemer med Libets forsøg hermed ikke gælder for de nyere forsøg. Og på dette grundlag mener Mele alt i alt, at neuroscience i hvert fald indtil videre *ikke* har påvist eller er i stand til at påvise, at vi som mennesker ikke har fri vilje.

Konklusion på del 2:

I denne del af specialet har jeg undersøgt, om forsøg i moderne neuroscience skulle have påvist, at vi ikke har fri vilje som menneskelige aktører. Jeg vil på basis af ovenstående kritiske analyser udfoldet af Mele tillade mig at konkludere, at Libets forsøg og senere eksperimenter frem til i dag i hvert fald ikke synes at have påvist, at vi ikke har fri vilje – og hermed heller ikke synes at udgøre nogen reel trussel mod Kanes libertarianistiske position.

³¹ Se pkt. 2.1.2.

Del 3: Forsvar af Kanes libertarianisme mod Galen Strawsons "Basic Argument":

I den ovenstående del af specialet har jeg kigget nærmere på én af de mest fremtrædende og brugte empiriske indvendinger mod Kanes position, nemlig at moderne neuroscience skulle underminere postulatet om menneskets frie vilje generelt – og jeg har under inddragelse af Alfred Meles kritiske analyse forsøgt at vise, at dette *ikke* synes at være tilfældet.

Ikke desto mindre kan man sige, at diskussionen af sådanne videnskabelige indvendinger og perspektiver kun er ét mere snævert aspekt af den bredere diskussion af fri vilje i vores tid; og at der også er mange andre indvendinger imod Kanes position, som vi bliver nødt til at forholde os til, hvis vi ønsker at vise, at hans teori kan forsvares imod disse.

I denne forbindelse vil jeg i det følgende dykke ned i dét, som i min optik synes at være den allervæsentligste *filosofiske indvending* mod Kanes libertarianistiske position, nemlig Galen Strawsons "Basic Argument". Some Kane bemærker i *Tapestry* (2024, s. 249), er der her tale om et argument, der opsummerer Strawsons egen skeptiske position i debatten om fri vilje; og videre er dette argument blevet diskuteret af mange øvrige filosofiske tænkere i tiden. Ifølge Caruso skyldes argumentets vide udbredelse, at det "*trades on some strong and commonsense intuitions*" (Caruso, 2018b). Pereboom og Caruso anser dette argument for at være et argument for *hård inkompatibilisme*, dvs. tesen om, at uanset om vi lever i et deterministisk eller indeterministisk univers, så er der ikke plads til den form for frie vilje, som ville betyde, at vi havde det ultimative moralske ansvar for vore egne handlinger – en position, som de sammen med Galen Strawson tilslutter sig.³²

I det følgende vil jeg først give en præsentation af The Basic Argument; og hernæst vil jeg komme med nogle kritikpunkter ift. dette argument. Alt i alt vil jeg hævde, at The Basic Argument giver anledning til en relevant diskussion af forudsætningerne for fri vilje og vores ultimative ansvar

³² Det er dog værd at notere sig, at Pereboom og Caruso til forskel fra Strawson mener, at *agent-causal libertarianism* rent hypotetisk godt kan give os den form for frie vilje, der ville underbygge *basic desert moral responsibility*, men at "*our best scientific theories indeed have the consequence that we are not morally responsible for our actions.*" (Pereboom, 2001, s. xiii). Herudover taler de om "*basic-desert moral responsibility*", mens Strawson taler om "*ultimate or true moral responsibility*", men jeg vil mene at disse gloser dækker over stort set samme ting. I begyndelsen af del 4 nævner jeg øvrige relevante filosofiske argumenter imod Kanes position, som man også kunne have dykket dybere ned i i specialet.

som aktører; men at det dog synes at være muligt at give et fyldestgørende forsvar af Kanes teori imod The Basic Argument.

3.1. Præsentation af Galen Strawsons "Basic Argument":

I artiklen "The Impossibility of Ultimate Responsibility" (1994) giver Strawson følgende indledende bestemmelse af dét, som han her kalder for "The Basic Argument":

There is an argument, which I will call the Basic Argument, which appears to prove that we cannot be truly or ultimately morally responsible for our actions. According to the Basic Argument, it makes no difference whether determinism is true or false. We cannot be truly or ultimately morally responsible for our actions in either case.

The Basic Argument has various expressions in the literature of free will, and its central idea can be quickly conveyed. (1) Nothing can be *causa sui* – nothing can be the cause of itself. (2) In order to be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects. (3) Therefore nothing can be truly morally responsible. (Strawson G. , 1994, s. 5)

Som det ses her, mener Strawson altså, at "The Basic Argument" er blevet artikuleret af mange forskellige tænkere i den filosofiske litteratur, men at det (i grove træk) kan opsummeres i form af de ovenstående tre punkter. Mere konkret er det Strawsons påstand, at dette Basic Argument baserer sig på tanken om, (1) at en menneskelig aktør ikke kan siges at besidde evnen til at være *causa sui*, dvs. sin egen årsag – da *intet* (ja, måske ikke engang Gud!) meningsfuldt kan siges at besidde denne evne. Men samtidig synes det at være tilfældet, (2) at for i sandhed at være *moralsk ansvarlig* for sine egne handlinger, så må man være i stand til at være årsagen til én selv, givet at det trods alt er én selv, der udfører disse handlinger og derved er årsagen til, at man handler, som man gør. Og hvis dette er tilfældet, (3) så synes det ifølge Strawson også at være tilfældet, at intet i denne verden i sandhed kan være *moralsk ansvarligt* for sine handlinger – hvilket også må gælde for menneskelige aktører, da disse aktører ligesom alt andet i verden umuligt kan være årsagen til sig selv.

På basis af denne indledende eller mere overordnede bestemmelse af The Basic Argument kommer Strawson også dét, han selv kalder for en mere udførlig eller "*tedious*" variant af argumentet:

1. Interested in free action, we are particularly interested in actions that are performed for a reason (as opposed to 'reflex' actions or mindlessly habitual actions).
2. When one acts for a reason, what one does is a function of how one is, mentally speaking. (It is also a function of one's height, one's strength, one's place and time, and so on. But the mental factors are crucial when moral responsibility is in question.)
3. So if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking – at least in certain respects.
4. But to be truly responsible for how one is, mentally speaking, in certain respects, one must have brought it about that one is the way one is, mentally speaking, in certain respects. And it is not merely that one must have caused oneself to be the way one is, mentally speaking. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and one must have succeeded in bringing it about that one is that way.
5. But one cannot really be said to choose, in a conscious, reasoned, fashion, to be the way one is mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, 'P1' – preferences, values, pro-attitudes, ideals – in the light of which one chooses how to be.
6. But then to be truly responsible, on account of having chosen to be the way one is, mentally speaking, in certain respects, one must be truly responsible for one's having the principles of choice P1 in the light of which one chose how to be.
7. But for this to be so one must have chosen P1, in a reasoned, conscious, intentional fashion.
8. But for this, i.e. (7), to be so one must already have had some principles of choice P2, in the light of which one chose P1.
9. And so on. Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice.
10. So true moral responsibility is impossible, because it requires true self-determination, as noted in (3). (Strawson G. , 1994, s. 6-7).

Som det fremgår af citatet, giver Strawson her en mere udførlig eller nidkært præsenteret version af The Basic Argument. Nærmere bestemt kan man sige, at The Basic Argument ifølge denne fremstilling af argumentet baserer på tanken om, at "*true moral responsibility*" kræver "*true self-determination*", men da sidstnævnte lader til at være umuligt, må førstnævnte også må være umuligt.

Mere konkret lyder argumentet her, at *hvis* vi skulle være i stand til "*true self-determination*", så skulle vi være i stand til at forårsage, hvordan vi selv var mentalt set (i de relevante henseender). Men dette ville ifølge Strawson forudsætte, at vi havde adgang til et sæt af principper ("*principles of choice*", P1), dvs. præferencer, værdier, attituder, idealer mv., som ville være bestemmende for, hvordan vi hermed kunne forårsage vores egen mentale støbning på en bestemt måde. Problemet er i denne forbindelse, at *hvis* det kræver et sådant sæt af principper at bestemme, hvordan vi vil være mentalt set, så synes dette blot at lede til et nyt spørgsmål – nemlig hvordan vi i så fald fandt frem til dette sæt af principper (P1) for, hvordan vi vil være mentalt set. Og dette synes at kræve, at vi havde et forudgående sæt af principper (P2), baseret på hvilke vi kunne udvælge det første sæt af principper. Og hvis dette er sandt, så synes hele denne proces at forudsætte en uendelig regres, da vi altid vil have brug for et forudgående sæt af principper med henblik at have udvalgt det sæt af principper, som oprindeligt havde bestemt, hvordan ville være mentalt set, osv. Og givet at dette er tilfældet, så bliver vi aldrig nogensinde i stand til at have det *ultimate* ansvar for at have dannet vores egen mentale støbning i overensstemmelse med P1, dvs. det førstnævnte sæt af principper for, hvordan vi ønskede at være mentalt. For dette ville have krævet, at vi på et tidspunkt kunne have bestemt, hvordan vi ville være mentalt set, *uden* at vi *på forhånd* var på en bestemt måde mentalt set – hvilket synes at være en fysisk umulighed!

Alt i alt kan man altså ifølge Strawson, at ovenstående er udtryk for samme Basic Argument, blot præsenteret på en mere teknisk måde. Dog bemærker Strawson, at vi også kan gengive The Basic Argument på følgende mere ligefremme og letforståelige facon:

1. Den måde, vi handler på, synes at afhænge af og være forårsaget af den måde, vi er på – mentalt set.
2. For at være ultimativt ansvarlige for vores frie handlinger, så bliver vi derfor nødt til at være ansvarlige for den måde vi er på (mentalt set), givet at den måde vi er på er bestemmende for, hvordan vi handler.
3. Vi er dog efter alt at dømme *ikke* ultimativt ansvarlige for den måde, vi er på – thi vi er ikke *causa sui*.

4. Ergo kan vi ikke siges at være ultimativt ansvarlige for den måde, vi handler på.

3.2. Forsvar af Kanes libertarianisme imod The Basic Argument:

Jeg har nu gennemgået Strawsons "Basic Argument", og hvordan Strawson formulerer forskellige varianter eller opstillinger af dette argument. Spørgsmålet er da: Er der her tale om et holdbart argument? Og hvordan kan vi ud fra Kanes teori forholde os til dette argument?

Overordnet set vil jeg mene, at The Basic Argument er et enormt interessant skeptisk argument i debatten om fri vilje og moralsk ansvarlighed; og at det er meget nemt at se, hvordan dette argument også problematiserer Kanes libertarianistiske position. Nærmere bestemt kan man sige, at hvis The Basic Argument er holdbart, så synes dette at underminere en grundantagelse i Kanes teori, nemlig at vi over tid kan blive ultimativt ansvarlige for vores egen personlighed og mentalitet gennem totaliteten af *self-forming actions* i vore liv. Herudover kan man sige, at Strawsons Basic Argument antyder, at vi umuligt kan være udstyret med den form for frie vilje, som ville betyde, at vi i praksis kunne være ultimativt ansvarlige for at træffe frie valg i livets situationer. På denne baggrund synes det alt i alt at være tilfældet, at *hvis* The Basic Argument er holdbart, så underminerer dette hele fundamentet for Kanes libertarianistiske position.

Dog mener jeg, at der set fra Kanes perspektiv synes at være følgende relevante indvendinger imod The Basic Argument, og at disse tilsammen bidrager til at underminere dette arguments styrke som en kritik af Kanes position.

3.2.1. Hvad menes der med *causa sui* i praksis – og er dette en rimelig betingelse for at være moralsk ansvarlig for sine valg og handlinger som menneske?

For det første vil jeg mene, at der synes at være et problem forbundet med Strawsons forståelse af, hvad det vil sige at være *causa sui* eller "årsag til sig selv" – en evne eller egenskab, der har en helt central rolle i The Basic Argument. Spørgsmålet er: Hvad mener Strawson mere konkret med dette ifm. menneskers faktiske livsførelse?

Jeg vil mene, at det ikke er helt klart, hvad Strawson faktisk mener med begrebet *causa sui* i relation til vores livspraksis som mennesker – og videre vil jeg mene at det på et mere overordnet plan synes at være et noget urimeligt krav, at vi skulle være i stand til at være "årsagen til os selv" for hermed at bære det moralske ansvar for vores egne valg og handlinger som menneskelige

aktører. Man kan sige, at den formelle betydning af dét at være *causa sui* naturligvis er, at man derved er ”årsagen til sig selv”, og at det slet og ret er dette, som dette latinske udtryk betyder; men jeg vil mene, at der herved ikke er blevet sagt særlig meget om, hvordan vi egentlig kunne tænkes at være eller blive *causa sui* i vores levede liv som menneskelige aktører, og hvordan dét at være årsagen til os selv hermed i praksis kunne være en reel mulighed for os i løbet af vort liv. Og så længe det ikke er afklaret, hvordan vi hermed rent lavpraktisk eller konkret kunne være eller blive årsagen til os selv som aktører, eller så længe det forbliver uspecificeret, hvordan det i det mindste rent hypotetisk ville være muligt at være årsagen til sig selv i denne forstand, så har jeg meget svært ved at se *relevansen* af The Basic Argument som et argument for, at vi aldrig kunne være ultimativt ansvarlige for vores egen mentale støbning som aktører og dermed heller ikke for de valg og handlinger, der måtte udspringe af vores mentale støbning, dvs. af den måde, vi ér på mentalt set på valg- eller handlingstidspunktet.

Nærmere bestemt vil jeg mene, at Kanes idé om ultimativt ansvar synes ikke at basere sig på tanken om *causa sui*, men på en mere opnåelig og rimelig betingelse for vores aktørskab som menneske. Som det fremgik af del 1, anser Kane mere præcist kriteriet for vort ultimative ansvar for vore egne handlinger på denne måde:

...to be ultimately responsible for an action, an agent must be responsible to some degree for anything that is a sufficient reason (a sufficient condition, cause, or motive) for the action's occurring (Kane, 2024, s. 12).

Som det ses af dette citat, kræver Kane altså *ikke* at vi skal være ”årsagen til os selv” i en mere absolut eller kategorisk forstand for at kunne have (en del af) det ultimative ansvar for vores egen karakter og selvskabelse som menneskelige aktører. I stedet forudsætter Kane blot, at vi *har én eller anden meningsfuld rolle at spille* i forbindelse med formningen og udviklingen af vores grundlæggende mentale egenskaber som aktører, og at vi gennem vores frie handlinger hermed i én eller anden grad kan være med til at præge vores egen karakter, dvs. hvem vi ér som personer. Navnlig mener Kane, at vi gennem *self-forming actions* i livet har evnen til at være med til at forme vores egen karakter ved at træffe frie valg i de livsøjeblikke, hvor vi skal vælge mellem inkommensurable værdisæt og præferencer, og at vi hermed kan være med til at omdanne vores egne karaktertræk som aktører – ikke ”ud af intet”, men på basis af det ”stof”, som vores genetik, tidligere opdragelse osv. har givet os i hænde som personer. Således er det altså Kanes antagelse, at vi ikke én gang for alle kan blive *ultimativt ansvarlige* for os selv, men at dette ultimative ansvar for os selv vokser frem *gradvist og over tid*, i takt med at mængden af

SFAs i vores liv er med til at danne og omdanne vores karaktermæssige og mentale grundtræk og dispositioner.

På denne baggrund vil jeg mene, at Kane synes at operere med en lidt anden, mere lavpraktisk og måske mere meningsfuld forestilling om "ultimativt ansvar" end Strawson. Strawsons tanke om ultimativt ansvar synes kort sagt at basere sig på tanken om *causa sui* og dermed at bygge på et så spinkelt metafysisk grundlag, at *intet* menneske nogensinde kan være ultimativt ansvarligt for sine egne beslutninger og handlinger. Til forskel herfra kan man sige, at Kane blot forudsætter, at vi må have *en del af* det ultimative ansvar for vores egen karakter og dispositioner som mennesker; og at dette kan opnås ved, at vi gennem SFAs kan være den tilstrækkelige årsag til i hvert fald *visse* af vores mentale og karaktermæssige anlæg. Sagt på en anden måde kan man sige, at Kane anskuer spørgsmålet om vores ultimative ansvar som ét, vi bør betragte i relation til *et menneskes samlede livsforløb*, og som det derved ikke nødvendigvis altid giver mening at se på i en isoleret situation – eller man kan sige, at et menneskes ultimative ansvar for sig selv og sin karakter ifølge Kane altid vil være et gradsspørgsmål, da intet menneske kan være *fuldt ansvarligt* for at have skabt sig selv fra bunden af.

Kane er altså enig med Strawson i, at intet menneske kan være *causa sui*, men han ser ikke, at dette skulle forhindre os i at opnå en relevant og tilstrækkelig grad af ultimativt ansvar for os selv som aktører, og at vi dermed også kan siges at være moralsk ansvarlige for vores valg og handlinger generelt. Mere præcist forudsætter Kane, at vores selvskabelse som aktører navnlig manifesteres og finder sted i de kritiske øjeblikke (SFAs), hvor vi netop føler os *splittede* som mennesker og må vælge mellem to (eller flere) handlemuligheder, som vil føre os lidt forskellige steder hen i livet, og som også vil sige noget forskelligt om os som mennesker. Og så længe vi har alternative handlemuligheder i disse afgørende livsøjeblikke, fastholder Kane, så *vil* vi være i stand til at opnå en tilstrækkelig grad af ultimativt ansvar for vores egen karakter og mentalitet til at kunne siges at være moralsk ansvarlige for vores handlinger generelt.

Kort sagt vil jeg alt i alt mene, at vi ud fra Kanes libertarianistiske teori herudfra kan opstille nogle ganske rimelige og meningsfulde kriterier for, hvordan og hvornår vi er ultimativt ansvarlige for vores egne valg og handlinger som aktører – og at der dermed ikke synes at være noget åbenlyst eller klart problem forbundet med den måde, som Kane tilgår spørgsmålet om ultimativt ansvar på i sin libertarianistiske position.

3.2.2. Kanes perspektiv på ultimativt ansvar synes at være mere brugbart og nyttigt:

Dette leder mig til den anden indvending mod The Basic Argument, nemlig at Kanes position på dette område har den fordel frem for Strawsons, at Kanes position rent faktisk tillader os at skelne i praksis mellem (1) personer med ultimativt ansvar for deres handlinger og (2) personer uden ultimativt ansvar for deres handlinger.

Ud fra Strawsons opfattelse må vi som sagt erkende, at *intet menneske* nogensinde kan siges at være ultimativt ansvarligt for sine egne beslutninger og handlinger, givet at intet menneske kan "skabe sig selv" fra bunden af eller ud af intet – dette er en umulighed, mener Strawson, da det ville forudsætte, at vi kunne være *causa sui*, årsagen til os selv. Men ifølge Kane kan vi sige, at *visse mennesker* har mere ultimativt ansvar for deres egen karakter samt deres beslutninger og handlinger end andre, og at det også i nogle sammenhænge vil være et kontekstafhængigt spørgsmål, hvor meget ultimativt ansvar vi har vores egne handlinger og disses konsekvenser. -

Et eksempel kunne være, at en lille pige på tragisk vis ender med at skyde sin egen far, da han glemte at lægge sin jagtriffel på plads efter brug, og da pigen ønskede at lege med denne riffel. I så fald ville vi næppe sige, at pigen havde det ultimative ansvar for sin handling og dens konsekvens, givet at hun næppe forstod det fulde omfang eller de reelle implikationer af sin handling – hun var for ung til at fatte og have det ultimative ansvar for, hvad hun gjorde. Men hvis denne mand af hævnthørst eller jalousi beslutter sig for at skyde sin ekskone med dette jagtgevær, og han lykkes med dette, så har han selvklart det ultimative ansvar for denne handling og fortjener dermed alt andet lige også en form for straf for denne handling. Ligeledes ville en udviklingshæmmet person eller en 10-årig dreng nok have *en vis grad* af ultimativt ansvar for et mord – men man kunne argumentere for, at en voksen, helt normalfungerende mand alt andet lige ville have en *højere grad* af ultimativt ansvar for et givent mord eller mordforsøg, givet at denne mand alt andet lige måtte antages at være mere bevidst om betydningen af dét at forsøge at slå et andet menneske ihjel, og idet denne mand har en anden neurologi og en anden livserfaring mv., som tillader ham at operere som et myndigt og autonomt menneske på en måde, som den 10-årige dreng og den udviklingshæmmede person ikke kan (og ikke med rimelighed må forventes at kunne), etc.

Pointen er altså denne: Selvom det måske ikke altid er *nemt*, kan vi ud fra Kanes libertarianisme meningsfuldt skelne mellem personer med ultimativt ansvar og personer uden ultimativt ansvar, og videre kan vi skelne mellem forskellige *grader* af ultimativt ansvar i forskellige

sammenhænge og henseender. På denne måde kan man altså sige, at vi kan *operationalisere* idéen om ultimativt ansvar ud fra Kanes teori, mens det samme synes at være vanskeligt, hvis ikke umuligt ud fra Strawsons standpunkt, givet at han jo helt benægter, at nogen mennesker nogensinde skulle være ultimativt ansvarlige for noget som helst.

Hermed vil jeg alt i alt mene, at der synes at være et *pragmatisk* argument for at foretrække Kanes position frem for Strawsons position på dette punkt, og videre synes det at være tilfældet, at det *rent moralsk* er nødvendigt at kunne skelne mellem folk, der alt andet lige ér ultimativt ansvarlige for deres adfærd og valg, og folk, der *ikke* er dette. Man kan også hævde, at det synes at være *vanskeligt* at se, hvordan vi som aktører kunne operere i en verden, hvor vi *ikke* kunne drage sådanne skel.

3.2.3. Kanes libertarianisme synes ikke at være præget af det regresproblem, som Strawson peger på i The Basic Argument:

Dette leder mig til en sidste indvending imod The Basic Argument, nemlig at Kanes teori om fri vilje ikke synes at være kendetegnet ved det regresproblem, som Strawson taler om i forbindelse med The Basic Argument.

Inden vi kigger nærmere herpå, så lad os komme med et hurtigt ”recap” af dette regresproblem:³³

1. For at være ultimativt ansvarlig for, hvad du gør, må du være ultimativt ansvarlig for, hvordan du ér – mentalt set.
2. Men for at være ultimativt ansvarlig for, hvordan du ér (mentalt set), må du have *valgt* dette.
3. Du kan dog ikke være ultimativt ansvarlig for, hvordan du ér (mentalt set), da:
 - a. Du selv ville skulle have valgt dette.
 - b. Men *for* at kunne have valgt det, så måtte du allerede være på en bestemt måde forud for dit valg.
 - c. Og for at være på denne måde forud for valget, så måtte du have valgt at være *på denne måde*, etc.

Og dette leder til en uendelig regres.

³³ Se også pkt. 4-9 af den udførlige variant af The Basic Argument.

Jeg vil mene, at ovenstående er en meget ligefrem og klar måde at formulere regresproblemet i The Basic Argument på. Men spørgsmålet er da: Hvad kan vi med Kane sige til dette regresproblem?

Jeg vil overordnet set mene, at dette regresproblem er en central del af The Basic Argument, og at det også er meget intuitivt at se, hvorfor dette regresproblem skulle opstå i forbindelse med vores forsøg på at begrunde vores eget ultimative ansvar for, hvordan vi ér og handler som aktører. Dog vil jeg samtidig mene, at dette regresproblem synes ikke at gøre sig gældende i forbindelse med Kanes libertarianistiske position; og jeg vil mere specifikt mene, at teorien om SFAs netop må ses som Kanes svar på, hvordan vi kan *undgå* at havne i det regresproblem, som Strawson påpeger.

Hvis vi kigger nærmere på Kanes forståelse af SFAs, så kan man netop sige, at disse har en såkaldt "*buck-stopping role*" (Kane, 1996, s. 114) i forbindelse med spørgsmålet om, hvordan vi bliver ultimativt ansvarlige for vores egne karaktertræk og mentale egenskaber som aktører. Kort sagt forudsætter Kane i sin teori om ultimativt ansvar, at vi ikke behøver at have ansvaret for *hele vores mentale støbning som mennesker* for at være moralsk ansvarlige for vores handlinger som aktører, men at det netop er nok, hvis vi blot er i stand til forme visse *dele eller aspekter* af vores egen karakter og mentalitet gennem nøgleøjeblikke i vores liv, og at disse nøgleøjeblikke netop er SFAs. På denne baggrund er det også Kanes påstand, at vi ikke behøver at være ultimativt ansvarlige for, hvordan vi oprindeligt er formet som aktører, for at vi kan være ultimativt ansvarlige for vores egne valg og handlinger i livet generelt. Som han skriver: "*We do not have to be responsible for how we are formed originally in order to be responsible now, so long as we are capable of taking responsibility for our future formation by self-forming actions.*" (Kane, 1996, s. 121).

På baggrund af denne forståelse af SFAs kommer Kane i *Tapestry* med følgende overordnede bemærkning om Strawsons påstand om en uendelig regres ifm. begrundelsen af vores ultimative ansvar:

[From the theory of self-forming actions, it follows that] no such unending regresses are required on an account of ultimate responsibility that involves such self-forming actions. For the agents' ultimate responsibility on such an account goes back to all the varied self-forming actions agents performed in the course of their lives. If agents also happen to be ultimately responsible for some of the motives or reasons involved in these self-forming actions ("for being the way they

are in key mental respects”), it will be because their ultimate responsibility for these motives or reasons also traces back to some other earlier self-forming actions. *All ultimate responsibility agents possess thus originates in self-forming actions, past and present, of the agents themselves and does not regress any further back beyond all their self-forming actions.* (Kane, 2024, s. 251).

Kort sagt kan man altså sige, at der på basis af teorien SFAs ifølge Kane selv ikke er noget belæg for at hævde, at der skulle optræde nogen uendelige regresser forbundet med begrundelsen af vores ultimative ansvar for vores egne beslutninger og handlinger, da dette ultimative ansvar altid vil kunne spores tilbage til summen af SFAs i vores liv som aktører, og da vi gennem disse SFAs opnår en grad af ultimativt ansvar for, hvordan vi selv ér mentalt set i vores liv, da vi netop gennem SFAs kan være med til at forme vores egne karaktertræk og mentale egenskaber i én eller anden grad, som ikke vil være forudbestemt og dermed fuldkommen ude af vores kontrol som aktører.

Konklusion på del 3:

Jeg vil alt i alt mene, at The Basic Argument er en interessant og relevant skeptisk indvending imod bl.a. Kanes position. Ikke desto mindre vil jeg mene, at der synes at være mulighed for at give et fyldestgørende forsvar af Kanes position over for The Basic Argument gennem de ovenstående tre punkter; og på dette grundlag vil jeg samlet set mene, at The Basic Argument ikke synes at udgøre en uigendrivelig indvending imod Kanes libertarianistiske position.

Del 4: Har troen på fri vilje en moralsk skyggeside? – Om fri vilje, retfærdiggørelsen af straf og ”moralsk vrede”:

I de forudgående dele af specialet har jeg kigget nærmere på (1) Kanes libertarianisme, (2) truslen mod fri vilje kommende fra moderne neuroscience og (3) hvordan Kanes teori om SFAs kan forsvares imod Galen Strawsons ”Basic Argument”.

Jeg vil ikke mene at dette har bevist, at Kanes libertarianisme er ”home safe” og kan forsvares imod alle relevante indvendinger imod denne; for at etablere dette ville det eksempelvis også kræve, at vi kiggede på om Kanes libertarianisme er i tråd med moderne fysisk teori, og om hans position kan forsvares imod bl.a. Peter van Inwagens (2000) *rollback argument*, Derk Perebooms (2014, s. 30-49) *disappearing agent objection* og Daniel Dennetts (2003, s. 108-126) kritik af det epistemologiske grundlag for SFAs. Men jeg vil mene, at de ovenstående dele har påvist, at Kanes position i det mindste kan forsvares imod nogle af de indvendinger, som den er blevet mødt med i de seneste årtiers filosofiske debat.

I de resterende dele af specialet vil jeg bevæge mig ind på et mere *praktisk eller normativt relevant område* af debatten om fri vilje. I denne forbindelse er jeg særligt interesseret i at undersøge, hvordan vi på basis af Kanes libertarianisme kan forholde os til nogle af de kritikpunkter ifm. selve troen på vores frie vilje, som Pereboom, Caruso og Sapolsky artikulerer som en del af deres *free will skepticism*. Med andre ord er de relevante spørgsmål her: Hvad indebærer troen på fri vilje mere praktisk og moralsk for os som menneskelige aktører – og er vi bedst tjent med at fastholde troen på fri vilje som en del af vores livssyn og verdensbillede? Eller ville det være bedre for os, hvis vi simpelthen gav køb på hele tanken om, at vi skulle være udstyret med sådan noget som fri vilje?

Jeg vil mene, at disse er nogle af de vanskeligste spørgsmål i hele debatten om fri vilje – og jeg gør mig i det følgende ingen forhåbninger om at komme med endegyldige eller absolutte svar på disse spørgsmål. Dog vil jeg mene at vi godt kan komme med nogle velunderbyggede svar herpå ved bl.a. at foretage nogle grundige analyser og drøftelser af en række af de skeptiske indvendinger, som Pereboom, Caruso og Sapolsky kommer med i forbindelse med de mere praktiske eller normative implikationer af selve troen på fri vilje. Mere specifikt vil jeg i denne del af specialet kigge nærmere på dét, som Caruso (2014) har kaldt for den ”mørke” eller ”dystre” side af debatten om fri vilje, nemlig hvordan selve troen på eksistensen af fri vilje synes at være beslægtet eller behæftet med en række meget *skadelige eller problematiske tendenser*, og at det

derfor alt andet lige ville være godt for os som samfund og individer, hvis vi eliminerede eller bortkastede tanken og tesen om, at vi skulle være udstyret med fri vilje som mennesker.

I det følgende vil jeg kigge nærmere på Perebooms, Carusos og Sapolskys vigtigste argumenter i denne forbindelse. For det første vil jeg kigge nærmere på deres påstand om, at der er en direkte kobling mellem troen på fri vilje og retfærdiggørelsen af straf – og hvordan vi med Kane kan forholde os til dette. For det andet vil jeg kigge på koblingen mellem troen på fri vilje og legitimeringen af dét, som Pereboom kalder for vores ”moralske vrede” som mennesker – og hvordan vi med Kane kan respondere herpå. Samlet set er det mit håb, at disse to punkter viser os, at der i hvert fald ikke *synes* at være nogen moralsk slagside eller skyggeside forbundet med at tro på vores fri vilje; og at det herved vil være muligt at påvise, at Kanes libertarianisme ikke leder til et mere problematisk syn på os selv og vore medmennesker end *free will skepticism*.

4.1. Om koblingen mellem troen på fri vilje og retfærdiggørelsen af straf:

Det første område, som jeg vil komme ind på i det følgende, er som nævnt koblingen mellem troen på fri vilje og retfærdiggørelsen af straf – og hvordan Pereboom, Caruso og Sapolsky anskuer dette. Spørgsmålet er altså: Hvad er denne kobling, og hvordan kan den forstås? Og hvad deres argumenter for, at det på dette område skulle være fordelagtigt at bortkaste troen på vores fri vilje som mennesker?

For at svare grundigt på disse spørgsmål er det nødvendigt først at tage et hurtigt kig på, hvordan Pereboom og Caruso tilgår forholdet mellem fri vilje og moralsk ansvarlighed. Ifølge Pereboom og Caruso kan vi definere moralsk ansvarlighed på forskellige måder, f.eks.. som (1) *tilregnelighed* – dvs. at en handling kan tilegnes en aktør, der hermed må siges at have *gjort noget*, uanset om dette så var et produkt af en bevidst og intentionel handling eller ej. Herudover kan man sige, at moralsk ansvarlighed kan bestå i (2) *at stå til regnskab* for en given handling eller beslutning, dvs. at aktøren hermed kan siges at skylde andre en forklaring på, hvorfor han eller hun gjorde dét, som hun gjorde i den pågældende situation. Og endelig kan man sige, at moralsk ansvarlighed kan handle om (3) dét, Pereboom og Caruso kalder for ”*basic desert*”, og som de opfatter på følgende måde:

[F]or an agent to be morally responsible for an action in [the basic desert] sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary. The desert here is basic in the sense that the agent

would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialist or contractualist considerations. (Pereboom, 2014, s. 127).

Videre er det Perebooms og Carusos påstand, at det netop er denne form for moralske ansvarlighed, som er omdrejningspunktet for debatten om fri vilje både historisk og aktuelt – og de mener begge to, at vi således bredt set kan forstå fænomenet ”fri vilje” som den *handle- og kontrolbetingelse*, der er nødvendig for at kunne sige at menneskelige aktører generelt må antages at have et sådant ”*basic-desert moral responsibility*” for deres egne handlinger i livet. Sagt på en anden måde mener Pereboom og Caruso altså, at den filosofiske debat om fri vilje ret beset handler om (og *bør* handle om!), hvorvidt vi som menneskelige aktører har den evne til at kontrollere vores egne beslutningsprocesser, som ville gøre at vi reelt set kan siges at være moralsk ansvarlige for vores handlinger ”*in the basic-desert sense*”. Dvs. når Pereboom og Caruso hævder at være *free will skeptics*, så mener dermed også, at de hermed benægter eller betvivler at vi som mennesker skulle være udstyret med den form for kontrol med vores handlinger og beslutningsprocesser, der generelt ville betyde, at det var muligt for os at være moralsk ansvarlige for i det mindste visse af vores handlinger ”*in the basic-desert sense*”.

På dette grundlag hævder Pereboom og Caruso mere principielt, at problemet synes at være, at *hvis og såfremt* vi tror på eksistensen af folks frie vilje, så vil vi dermed også komme til at antage og tro på, at alle mennesker alt andet lige har et sådant ”*basic-desert moral responsibility*” for deres egne handlinger – og at dette synes at have nogle meget skadelige eller problematiske implikationer for vores samkvem som aktører samt for indretningen af en række af vores samfundsinstitutioner. Sagt på en anden måde er det altså Perebooms og Carusos tese, at selve troen på fri vilje nødvendigvis er knyttet til en række antagelser om menneskers ”*basic-desert*” moralske ansvar for deres egne beslutninger og handlinger i livet – og at det er denne kobling, der i en videre forstand kan siges at gøre selve troen på eksistensen af vores frie vilje uheldig eller problematisk i en samfundsmæssig optik.

Dette ses ifølge Pereboom og Caruso klarest på retsområdet, hvor man kan sige, at der en særlig teori om retfærdiggørelsen af straf – nemlig den *retributivistiske teori om retfærdiggørelsen af straf* – der synes at være intimt forbundet med denne forståelse af fri vilje som kontrolbetingelsen for *basic-desert moral responsibility*. Nærmere bestemt kan man ifølge Pereboom sige, at denne retributivistiske teori baserer sig på tanken om, at ”*punishment of a wrongdoer is justified for the reason that he deserves something bad to happen to him – pain, deprivation or death, for example – just because he has done wrong.*” (Pereboom, 2014, s. 157).

På baggrund heraf mener Pereboom at man kun kan gå ind for denne retfærdiggørelse af straf, hvis man mener at mennesker generelt har *basic-desert moral responsibility* for deres egne handlinger, hvorimod man kan sige, at hvis man tilslutter sig *free will skepticism*, så vil man også betragte enhver retributivistisk opfattelse af retfærdiggørelsen af straf som byggende på et illegitimt og urimeligt grundlag, da denne form for retfærdiggørelse af straf netop forudsætter at vi har *basic-desert moral responsibility* for vores egne handlinger.

I forlængelse af disse tanker bemærker Caruso mere konkret, at hvis vi kan anlægge et mere skeptisk blik på menneskets fri vilje i et land som USA, så vil vi også være i stand til at benytte os af mere effektive og humane metoder inden for en række forskellige politiske og sociale områder. Nærmere bestemt bemærker Caruso, at: "*Once we relinquish the belief in free will, this will allow us to look more clearly at the causes and more deeply at the systems that shape individuals and their behavior, and this will allow us to adopt more humane and more effective policies in education, criminal justice, and social policies.*" (Caruso, 2014).³⁴ I denne forbindelse henviser Caruso bl.a. til retsvæsenet i USA, og han mener, at i stedet for at gøre amerikanske borgere mere trygge og beskytte deres liv samt reducere kriminalitet, så er det amerikanske retssystem først og fremmest gearret til at *straffe* lovovertrædere på basis af en idé om, at de altid har handlet af deres egen frie vilje og dermed også har den form for moralske ansvar for deres handlinger, der betyder at de generelt fortjener at blive straffet for deres lovovertrædelser. I dette regi kommer Caruso også med følgende bemærkning:

Free will is built into the mythology of the rugged individual, the self-made man, the *causa sui*. The person who can pull themselves from the bootstraps and overcome all of their life-circumstances. But because Americans are so committed to this myth, the United States is also a very punitive society. The US makes up 5 % of the world's population, but incarcerates 25 % of the world's prisoners.

The US Criminal Justice system is broken, not working. It is not making us safer; not reducing crime; by no means achieving desired social goals, and not reducing the rate of recidivism, i.e. repeat crime. (Caruso, 2014).³⁵

Som det ses af dette citat, er det altså Carusos overbevisning, at der er nogle grundlæggende problemer forbundet med det amerikanske retssystem, og at dette system ikke fungerer på en måde, som sikrer hverken resocialisering eller mere langsigtet kriminalitetsbekæmpelse. I stedet

³⁴ Hør fra 9.50 og frem.

³⁵ Hør fra 4.00 og frem.

er der tale om et system, der i det store og hele er begrundet gennem dét, vi kunne kalde *fri vilje-ideologien* i USA, altså tanken om, at et individ *altid* vil være i stand til at gøre brug af sin egen frie vilje til at ændre sine omstændigheder på en positiv facon; og hvis individet ikke lykkedes med at gøre dette og i stedet tyr til kriminalitet, så er det altid helt og holdent individets egen skyld og ultimative moralske ansvar.

Som Caruso (2019, s. 50) i denne forbindelse videre bemærker, er der i dagens USA er over 10 gange så mange indsatte pr. indbygger i landets fængsler som i skandinaviske lande som Norge, Danmark og Sverige. Som en mulig forklaring herpå nævner Caruso en stærkere tro på individuel frihed og individets fundamentale moralske ansvar for sin egen skæbne i samfundet samt den brede folkelige tilslutning til "*retributivism and severe punishment*" (Caruso, 2019, s. 44). For at gøre op med USA's aktuelle tilgang til kriminalitet mener Caruso derfor også, at man først og fremmest må opgive tanken om individet som en aktør med fri vilje og det *fulde eller komplette* moralske ansvar for sin egen livssituation, da det netop er denne tanke, der faktisk ligger til grund for selve *retfærdiggørelsen* af de hårde straffe, som den gældende retspraksis i USA ofte baserer sig på, og som mange amerikanere tilsyneladende anser for at være helt igennem normal og legitim.

Som et alternativ til den retributivistisk orienterede strafferetspraksis i dagens USA forestiller Pereboom og Caruso sig, at der i tråd med *free will skepticism* kan benyttes dét, som Pereboom) kalder for "the quarantine model of punishment". Ifølge denne model bør vi se spørgsmålet om bekæmpelsen af kriminalitet analogt til spørgsmålet om bekæmpelsen af smitsomme sygdomme såsom ebola eller coronavirus. Sagt på en anden måde mener Pereboom, at ligesom vi normalt opfatter det som legitimt og acceptabelt, at visse personer smittet med farlige sygdomme sættes i karantæne for at forhindre videre smitte, også selvom de pågældende personer ikke nødvendigvis er moralsk ansvarlige for at være smittet af den pågældende sygdom, så kan vi også godt tillade os at sætte kriminelle i karantæne som en straf for diverse lovovertrædelser, hvis dette er med til at beskytte det omkringliggende samfund og dets beboere, også selvom disse kriminelle personer teknisk set ikke har det moralske ansvar for deres gerninger. Hermed gøres retfærdiggørelsen af straf ifølge Pereboom til noget, som er kompatibelt med tesen om, at kriminelle *ikke* har fri vilje og "*basic-desert moral responsibility*" for deres gerninger; og der kan dermed ifølge Pereboom i højere grad end aktuelt fokuseres på målsætninger såsom (1) "*to make the criminal safe for release*" og (2) "*to bring about moral change in an offender by non-punitively addressing conditions that underlie criminal behavior*" (Pereboom, 2001, s. 178).

Caruso tilslutter sig Perebooms tanker på dette punkt og kommer i denne forbindelse med følgende forsvar af dét, som han kalder for *the public health-quarantine model*:

Summarizing the public health-quarantine model, then, the core idea is that the right to self-defense and defense of others justifies incapacitating the criminally dangerous with the minimum harm required for adequate protection. The resulting account would not justify the sort of criminal punishment whose legitimacy is most dubious, such as death or confinement in the most common kinds of prisons in our society. The model also specifies attention to the well-being of criminals, which would change much of current policy. Furthermore, the public health component of the theory prioritizes prevention and social justice and aims at identifying and taking action on the [social determinants of health] and criminal behavior. This combined approach to dealing with criminal behavior, I maintain, is sufficient for dealing with dangerous criminals, leads to a more humane and effective social policy, and is actually preferable to the harsh and often excessive forms of punishment that typically come with retributivism. (Caruso, 2019, s. 65).

Ifølge Caruso vil en sådan *public-health quarantine model* bl.a. indebære et opgør med de såkaldte "supermax prisons" i dagens USA, og som i hans optik er "*cruel places, using a number of harsh forms of punishment, including extended solitary confinement.*" (Caruso, 2019, s. 51).

I *Determined* tilslutter Sapolsky sig Perebooms og Carusos ovenstående tanker om retfærdiggørelsen af straf og udtaler: "*I really like quarantine models for reconciling there's-no-free-will with protecting society from dangerous individuals*" (Sapolsky, 2023, s. 331). Dog mener Sapolsky, at der er visse praktiske udfordringer forbundet med denne tilgang – f.eks. hvor lang denne karantæne skal være, og om det vil være nok til at afskrække potentielle forbrydere. Overordnet set er Sapolsky dog overbevist om, at Perebooms og Carusos tilgang vil lede til et mere humant retsvæsen. Endvidere påpeger Sapolsky, at vores neurobiologiske støbning betyder, at det ofte er *dejligt* for os at se andre mennesker bliver straffet (hårdt og brutalt!) for diverse misgerninger og moralske forseelser – et faktum, der ifølge Sapolsky er ganske uheldigt, og som der findes tusindvis af rædsomme og uhyrlige eksempler på i vores kulturelle og civilisatoriske historie.³⁶ Og endelig bemærker Sapolsky i denne forbindelse, at troen på fri vilje desværre synes at være noget af dét, der gør at vi føler os *berettigede* til at straffe andre for deres forseelser. Følgelig mener Sapolsky alt i alt, at det er enormt vigtigt for os som

³⁶ Se kapitlet "The Joy of Punishment" – (Sapolsky, 2023, s. 341-384).

samfund at stille spørgsmålstegn ved tesen om menneskets frie vilje, da denne tese synes at understøtte og legitimere vores generelle tanke om, at andre mennesker i det hele taget *fortjener* at blive straffet for deres (stipulerede) ”frie” handlinger, og denne tro på fri vilje hermed er med til at underbygge og legitimere det meget ineffektive og på mange punkter noget sadistiske amerikanske retssystem i dets gældende form.

4.1.1. Kritik af Perebooms, Carusos og Sapolskys tanker om fri vilje og retfærdiggørelse af straf:

Spørgsmålet er da: Har Pereboom, Caruso og Sapolsky ret i, at troen på fri vilje står i vejen for progressive og humane fængselsreformer i et land som USA? Og er det virkelig rigtigt, at selve dét at tro på fri vilje gør os ekstra modtagelige over for idéen om, at gengældelsesorienterede straffeformer er moralsk legitime og retfærdige? Ja, går der virkelig en lige linje fra at fastholde en tro på menneskers frie vilje til en villighed til at straffe andre mennesker hårdt for deres ugerninger og kriminelle adfærd?

Jeg vil alt i alt mene, at Pereboom, Caruso og Sapolsky har nogle virkelig interessante tanker om koblingen mellem troen på fri vilje og vores tanker om retfærdiggørelsen af straf og betydningen af kriminel adfærd i det hele taget – og jeg vil desuden mene, at deres argumentation synes at være præget af den meget *sympatiske og næstekærlige tanke*, at vi bør lade være med at straffe andre mennesker unødigt hårdt og brutalt, og som grundlæggende ikke hjælper mennesker. Jeg er således enig med dem i, at det moderne amerikanske retsvæsen på mange måder synes at være et urimeligt og meget uhensigtsmæssigt system, og at der er alt for mange mennesker, som er *”warehoused in prisons”* i dagens USA, som Robert Kane formulerer dette i en forelæsning (Kane, 2010b).³⁷

Dog vil jeg samtidig mene, at der ikke synes at være nogen som helst evidens for, at troen på fri vilje skulle forhindre os i at agitere for en række vigtige og forskelligartede humanistiske og progressive reformer på retsområdet – og at dette synes at underminere en væsentlig del af Perebooms, Carusos og Sapolskys argumentation.

For at understøtte denne pointe vil jeg mene, at vi kan kigge på vores moderne historie i et land som Danmark, da denne historie netop synes at indikere, at vi har foretaget radikale ændringer af vores strafferetlige system, og at dette efter alt at dømme har fundet sted, *uden* at vi dermed

³⁷ Hør fra 5.20 og frem.

har forladt troen på vores vilje som menneskelige aktører. Eksempelvis kan man sige, at en stor del af vores strafferetlige system historisk synes at have baseret sig på ønsket om at se aktører blive straffet hårdt for dét, som vi gennem tiden har opfattet som klare brud på det moralske grundlag for vores samfundsorden og korrekt menneskelig opførsel – et eksempel på dette kunne være, at straffen for abort under Christian d. 5.'s Danske Lov fra 1683 var, at "*Letfærdige Qvindfolk, som deris Foster ombringe, skulle miste deris Hals, og deris Hovet sættis paa een Stage*"³⁸. Her kan man sige, at en sådan straf synes at udspringe af nogle klare retributivistiske tendenser koblet med nogle meget klare religiøse og moralske overbevisninger om livets hellighed og kvindens lod; og videre må vi antage, at man (eller kirken eller kongen) dengang har fundet den pågældende straf *retfærdig* eller *fair*, da man har antaget, at de pågældende kvinder måtte have udført disse aborter gennem brugen af deres egen frie vilje, og at disse kvinder hermed også havde det *moralske ansvar* for de pågældende gerninger, hvorfor det gav mening at straffe dem for gerningerne.

Man kan her sige, at vi i dagens Danmark er nået frem til et markant andet syn på abortspørgsmålet end i 1683, og at denne udvikling synes at have fundet sted i sammenhæng med en række større samfundsmæssige forandringer, der har præget danskernes syn på både straffens rolle i relation til det enkelte menneske og kvindens position i samfundet. Man kunne f.eks. argumentere for, at vi i dag har en mere oplyst, veluddannet befolkning med et generelt mere humanistisk præget og demokratisk baseret syn på menneskets status og kvinders rettigheder, og at vi i dag *generelt* ikke er villige til at hugge hovedet af folk pga. diverse lovovertrædelser, blot fordi vi finder, at de måtte have handlet moralsk ilde eller forkert. På denne baggrund vil jeg mene, at dét, at vi har ændret syn på abortspørgsmålet i Danmark i løbet af de seneste århundreder, i hvert fald ikke umiddelbart synes at have som helst at gøre med ændringer i vores opfattelse af, hvorvidt danske borgere generelt må siges at have fri vilje eller ej – i stedet skyldes disse ændringer i vort syn på abortspørgsmålet efter alt at dømme nogle bredere samfundsmæssige ændringer, som har at gøre med vores syn på menneskets rettigheder, kvinders status i samfundet og måske også religionens rolle i forhold til vort retsvæsen. Og på denne baggrund vil jeg mene, at der ikke synes at være noget klar evidens for, at netop troen på vores frie vilje skulle forhindre os i fremadrettet at advokere for nye humanistiske og progressive reformer på retsområdet. Vi har allerede udført en masse gode reformer på retsområdet i løbet af de seneste århundreder i et land som Danmark, og dette har givet mænd og kvinder i Danmark langt flere rettigheder end på Enevældens tid – og jeg kan

³⁸ Jf. (Wilken-Jensen, Rosenbeck, & Dübeck, 2024).

kort sagt ikke se, hvorfor det fremadrettet skulle være nødvendigt at underminere vores tro på fri vilje for at indføre flere humanistiske eller progressive reformer af vores retsstat, der f.eks. kunne give borgerne bedre rettigheder og sikre en endnu bedre behandling af vores indsatte i fængsler, etc.

Herudover vil jeg *for det andet* mene, at der mere generelt synes at være visse *moralske problemer* forbundet med *free will skepticism* som filosofisk position – for hvis vi fortæller folk, at de faktisk ikke har fri vilje og dermed ikke noget ultimativt moralsk ansvar for deres egne handlinger, så er spørgsmålet, hvordan vil de så reagere?

På dette punkt er jeg enig med Kane i, at *free will skepticism* muligvis kan have nogle uheldige følgevirkninger ifm. visse aktørers selvforhold og ansvarsfralæggelse – en frygt, som Kane opsummerer på denne måde i et interview fra filmen *Free Will? A Documentary* fra 2023:

In response to the free will skeptics, I must say that I am not as sanguine about it as some people like Pereboom and Caruso are[...]. I think there would be problems. Because, if we are interested in curtailing the evildoers of this world—whether they be dictators or Wall Street bankers who cheat people, or whatever—if we are ever going to be capable of curbing the evildoers of the world, it is not going to be by telling them that they are not ultimately responsible for being the way they are. Because if we tell them *that*, they are going to say something like, ‘Yeah, I work down here in Wall Street, I cheat people all the time, but *that’s the way I am!* I am not responsible!’ That’s what they’re going to say. My *real* problem with free will skepticism is precisely that one. (Walsh, 2023).³⁹

Som det fremgår af dette citat, er det altså Kanes overbevisning, at der potentielt er et stort *moralsk problem* forbundet med *free will skepticism* som filosofisk position, nemlig at denne position giver verdens forbrydere, skruppelløse og ondsindede aktører en slags moralsk *cart blanche* til at handle på lige den måde, de ønsker, uden at de behøver at stå til ansvar for deres gerninger, og hvem de er som personer. I tråd med denne pointe kommer Kane i en forelæsning fra 2010 også med følgende bemærkning:

Such skeptical views [about free will and ultimate responsibility] dangerously neglect, I think, the widespread human tendency to justify selfish and immoral behavior by rationalization, self-deception and denial. And rejection of free will and

³⁹ Se fra 1 time 40 min. inde i dokumentaren og frem.

moral responsibility is to my mind the ultimate incentive for rationalization, self-deception and denial, as the psychological experiments of Vohs, Schooler and others suggest. (Kane, 2010b).⁴⁰

Som det ses, mener Kane altså, at vi som mennesker er tilbøjelige til at benytte os af efterrationaliseringer eller bortforklaringer i alle mulige henseender; og disse psykologiske tilbøjeligheder vil ifølge Kane også have en betydning for, hvordan vi som aktører ofte har en hang til at retfærdiggøre selvisk og umoralsk adfærd og afvise at have ansvaret herfor både over for os selv og andre.

De adfærdseksperimenter, som Kane henviser til i dette citat, blev udført af de canadiske psykologer Jonathan W. Schooler og Kathleen D. Vohs, og viste bl.a., at en forsøgsperson er mere tilbøjelig til at snyde i en test, når denne person netop har læst en tekst, der sår tvivl om menneskers frie vilje (Vohs & Schooler, 2008). Ifølge Kane kan vi se dette som forskningsmæssig evidens for, at der i hvert fald *synes at være* et moralsk problem forbundet med dét at få folk til at tro, at vi *ikke* har fri vilje som aktører, da dette synes at gøre det meget sandsynligt, at de meget vel kan overbevise sig selv om, at de eksempelvis har ret til at snyde eller handle ilde i praksis, da de jo alligevel ikke er moralsk ansvarlige for deres handlinger. Kort sagt mener Kane altså, at *hvis* vi rent faktisk fortæller folk, at de ikke har fri vilje og det ultimative ansvar for deres valg og handlinger i livet, *og hvis de rent faktisk integrerer denne sandhed som en del af deres livssyn og levevis*, så vil dette alt andet lige føre til en række større moralske omvæltninger og udfordringer i vores nære relationer og samfund – ja, med filosofen Saul Smilansky kunne man her mere spekulativt hævde, at “[t]here are major dangers that a great decline in moral seriousness, leading to much cynicism and even nihilism,[...] would be the result” (Smilansky, 2024, s. 110).

Dette leder mig til min *tredje og sidste* indvending mod Perebooms, Carusos og Sapolskys position, nemlig at det ud fra karantænemodellen for retfærdiggørelse af straf ikke synes at være klart, hvordan (og hvorfor!) vi egentlig skulle skelne mellem moralsk og juridisk skyldige og uskyldige individer, og hvilken status selve denne distinktion egentlig ville have.

Som bl.a. filosofen Scott Sehon (2016) bemærker, så kan man sige, at et problem med *free will skepticism* som position er, at man ud fra denne position bliver nødt til at hævde, at intet menneske nogensinde kan siges at være *ultimativt* moralsk ansvarligt for dets handlinger, da intet menneske som bekendt nogensinde har handlet på basis af dets egen frie vilje – men hvis

⁴⁰ Hør fra 6.36 og frem.

dette er tilfældet, så er det svært at se, hvorfor det skulle være mere moralsk forsvarligt at putte kriminelle aktører i karantæne end at putte uskyldige aktører i karantæne, da *ingen* jo i sidste ende kan siges at have det ultimative ansvar for deres egne handlinger. I denne forbindelse kommenterer Sehon mere udførligt, at:

[A] problem for the quarantine model is that[...] there would no longer be any reason to restrict incarceration to those who have actually committed a crime. It is, after all, not only those with Ebola who have been quarantined, but also those who have simply been exposed to the disease. With diseases, the justification for isolating a person (even without their consent, if necessary) is based solely on the probability that they will infect others. We usually do not know the exact probability that a given person will spread a disease, and in some circumstances our best judgment concerning these probabilities will make no distinction between those who have the disease and those who were merely exposed to it. In such cases, the justification for quarantining those merely exposed to the disease is exactly as good as the justification for quarantining those known to have the disease. Following this reasoning, the entire justification for criminal incarceration will, on the quarantine model, lie in the likelihood that the person to be incarcerated would otherwise commit a crime in the future; that someone has committed a crime in the past is relevant only insofar as this is evidence that they will commit future crimes. So if we have other, equally good, evidence that someone will commit a crime in the future, then we would be equally justified in incarcerating them, regardless of whether they have actually done anything wrong. (Sehon, 2016, s. 11).

Kort sagt: Den eneste legitime årsag til at sætte folk i karantæne ud fra karantænemodellen er, at der er en vis sandsynlighed for, at folk kunne tænkes at begå kriminalitet i fremtiden; dette skyldes, at hele *skyldsspørgsmålet* i forbindelse med straf slet ikke er relevant for Pereboom & co., da de jo kun mener, at vi har ret til at straffe andre mennesker på basis af hensynet til den øvrige befolknings krav om at blive beskyttet mod kriminel adfærd i fremtiden, hvorfor en vis form for *karantæne* af farlige individer er acceptabelt – og dette også selvom disse individer jo ikke kan siges at have det ultimative ansvar for deres egen adfærd og være *moralsk ansvarlige i "the basic-desert sense"* for deres gerninger, uanset om sådanne gerninger så er kriminelle eller ej.

Eller mere simpelt sagt: *Hvis hele skyldsspørgsmålet er irrelevant ifm. udmålingen og legitimeringen af straf, hvorfor skulle det så være mere illegitimt at fængsle eller indespærre*

mennesker, som ikke har begået kriminalitet, end mennesker, der rent faktisk har begået kriminalitet? Det ville kort sagt være arbitrært at skelne moralsk mellem disse to grupper af mennesker; og som filosofen John Lemos (2016, s. 471) kommenterer, så kan man hermed sige, at vi effektivt set ville opføre os *diskriminerende* over de kriminelle aktører, hvis vi *kun* ønskede at indespærre disse!

Jeg vil mene, at disse overvejelser er yderst relevante ift. at vurdere de egentlige eller mere praktiske implikationer af at indføre karantænemodellen for straf, som Pereboom, Caruso og Sapolsky står på mål for – og jeg er enig med Lemos og Sehon i, at der synes at være nogle potentielle moralske problemer forbundet med denne karantæne-model. Mere konkret kunne vi ifølge Sehon snildt forestille os, at en organisation såsom The Chicago Police Department ville benytte sig af en masse data om befolkningen i Chicago til at finde ud af, hvem det ville være ”smart” eller ”behørigt” at putte i karantæne på et forebyggende grundlag – dette i det mindste indtil de pågældende borgere ikke var ”til fare” for deres omgivelser længere. Ifølge Sehon ville der på basis af karantænemodellen ikke være nogen meningsfuld eller logisk grund til *ikke* at bruge politiets data til at vurdere, om det kunne være nødvendigt at sætte visse uskyldige personer i karantæne, hvis det eksempelvis ud fra politiets beregninger kunne fastslås, at netop disse personer med vis sandsynlighed ville begå voldelig kriminalitet i den nærmere fremtid. Som Sehon i denne forbindelse påpeger, har The Chicago Police Department allerede computermodeller, der viser, at visse borgere i Chicago-området rent statistisk er “*500 times more likely than average to be involved in violence*” end andre borgere (Gorner, 2013). Vi ville ud fra karantænemodellen have præcist lige så megen ret til at sætte disse uskyldige (men statistisk set farlige!) mennesker i karantæne, som vi ville have ret til at sætte *faktiske* kriminelle aktører i karantæne, påpeger Sehon.

Omvendt kunne vi ud fra karantænemodellens logik også forestille os, at hvis nu en mand slår sin ekskone ihjel, og der er vitterligt tale om et drab, som *alene* var motiveret af denne mands personlige følelser ifm. sin ekskone (fx jalousi), så er der måske en høj sandsynlighed for, at denne mand *aldrig* nogensinde vil være i fare for at begå et drab igen, da han blot var ude på at myrde sin ny døde ekskone – og hvis dette er tilfældet, så har vi ifølge karantænemodellen faktisk *ingen ret* til at putte ham i karantæne, da han jo ikke længere er farlig for sine omgivelser! Med andre ord må vi altså ud fra karantænemodellen sige, at *hvis* vi kunne dokumentere at denne mand *slet ikke* ville være farlig for sine omgivelser fremadrettet, så har vi faktisk *ingen ret* til at putte ham i karantæne som en konsekvens af dette mord; og dette også selvom vi er helt sikre på, at det var ham, der stod bag mordet.

Jeg vil på baggrund af disse refleksioner mene, at der synes at være visse elementer af denne karantænemodel, som synes at være i strid med vores almindelige moralske dogmer og principper, fx (1) at *uskyldige mennesker ikke vilkårligt må fængsles, blot fordi vi vurderer, at de med en vis sandsynlighed kunne finde på at begå kriminalitet fremadrettet*; og (2) at *folk, der har begået et drab, dog fortjener én eller anden form for straf, også selvom de umiddelbart herefter måske ikke længere er til fare for deres omgivelser*. Og problemet synes at være, at hvis vi tilslutter os karantænemodellen, så bliver vi faktisk nødt til at skrotte disse moralske principper ikke blot i vores filosofiske tænkning, men som basis for hele vores retssamfund.

Opsummering af pointerne i det ovenstående:

Jeg vil alt i alt mene, at ovenstående refleksioner udgør et solidt forsvar af Kanes libertarianisme imod Perebooms, Caruso og Sapolskys tanker om koblingen mellem fri vilje og retfærdiggørelsen af straf. For det første (1) vil jeg mene, at Pereboom, Caruso og Sapolsky ikke på en fyldestgørende måde synes at demonstrere, at det skulle være nødvendigt at bortkaste tesen om vor frie vilje for at kunne agitere for forskellige former for progressive reformer på retsområdet i bl.a. det gældende amerikanske samfund. For det andet (2) kan man sige, at Pereboom, Caruso og Sapolsky synes at undervurdere, at det kunne vise sig at være moralsk problematisk at fortælle verdens skurke og andre egoistiske og ondsindede aktører, at de ikke har fri vilje og dermed heller ikke det ultimative ansvar for, hvordan de handler. Og endelig synes der for det tredje (3) at være nogle mere principielle problemer med karantænemodellen, som Pereboom, Caruso og Sapolsky går ind for, navnlig hvordan man inden for rammen af denne karantænemodel på et meningsfuldt grundlag rent faktisk kan skelne mellem (a) uskyldige og (b) skyldige personer og i det hele taget behandle disse to grupper af mennesker forskelligt.

Tilsammen vil her ud fra disse punkter mene, at der alt andet lige er god grund til at have store forbehold over for Perebooms, Carusos og Sapolskys filosofiske position baseret på dennes samfundsmæssige og moralske implikationer.

4.2. Om troen på fri vilje og vores ”moralske vrede” mod andre mennesker:

Dette leder mig til det andet punkt, som vi må komme ind på her, nemlig hvordan Pereboom, Caruso og Sapolsky anskuer koblingen mellem troen på fri vilje og vores vrede som personer.

Kan vi med andre ord sige, at der er noget ved selve troen på fri vilje, som er med til at afføde en særlig form for vrede mod andre mennesker? Og hvordan kan vi nærmere forstå koblingen mellem troen på fri vilje og så vores umiddelbare eller impulsive hang til at blive vrede på andre mennesker på grund af dét, som vi måtte mene, at de har gjort? Ja, er der noget ved selve troen på fri vilje, som så at sige blokerer for, at vi kan opnå et godt og sundt og opbyggeligt forhold til andre mennesker, simpelthen fordi vi fanges i nogle tankemønstre, der somme tider motiverer og begrundet vores vrede og forargelse imod disse mennesker?

For at nærme os nogle svar på disse spørgsmål er det nødvendigt først at tage et nærmere kig på, hvad Pereboom, Caruso og Sapolsky rent faktisk har at sige om koblingen mellem fri vilje, moralsk ansvarlighed og vrede. Nærmere bestemt er det Perebooms tese, at troen på fri vilje er koblet til en særlig form for vrede ("moralisk vrede"), der synes tæt forbundet med *forargelse* som fænomen, og som er en central del af vort almindelige etiske eller moralske liv. I *Living without Free Will* karakteriserer Pereboom mere konkret denne "moraliske vrede" som følger:

The type of anger at issue [in the free will debate] is the sort that is directed toward someone who is represented as having done wrong (sometimes the representation is accurate, sometimes not). Let us call this attitude *moral anger*. Perhaps moral anger, as I have characterized it, is closely related to indignation. Not all anger is moral anger. One type of non-moral anger is directed at someone because his abilities in some respect are scant or because he performs poorly in some particular circumstance. We are on occasion angry with inanimate machines for malfunctioning. Sometimes we are just angry without any target. However, by far most human anger is moral anger.

Moral anger forms an important part of the ethical life as it is ordinarily conceived. It motivates us to resist oppression, injustice, and abuse. But often expressions of moral anger have harmful effects. They may fail to contribute to the well-being of those to whom they are directed. Frequently, expressions of moral anger are intended to cause physical or emotional pain. Partly as a result of these problems, moral anger often has a tendency to damage or destroy relationships. In extreme cases, it can provide motivation to take very harmful and even lethal action against another. (Pereboom, 2001, s. 208).

Som det ses af dette citat, er det altså Perebooms holdning, at (1) moralsk vrede er den mest udbredte form for menneskelige vrede, og at (2) denne form for vrede synes at være nyttig i

forskellige sammenhænge, men at der også er tale om en potentielt enormt farlig reaktion på andres gerninger, som kan lede os til at handle på meget destruktive måder over for andre mennesker.

Nærmere bestemt mener Pereboom, at ”moralsk vrede” kan forstås som en såkaldt *reaktiv attitude*,⁴¹ dvs. én blandt mange forskellige ”*special emotion-laden responses to which human beings are naturally prone in their interactions with one another.*” (McGeer, 2012). Andre sådanne reaktive attituder er ifølge Pereboom forbitrelse, forargelse, taknemmelighed, hævnnergerrighed, stolthed, skyldfølelse, samvittighedsnag, kærlighed mv. – og af disse mener Pereboom altså, at moralsk vrede er meget tæt forbundet med vores *forargelse* over andres ugerninger. Problemet med moralsk vrede er i denne henseende ifølge Pereboom kort sagt, at denne type af vrede gør os særligt tilbøjelige til at ville skade eller straffe andre mennesker baseret på vores overbevisning om, at de netop *fortjener* at blive sårede eller lide på basis af, at de har gjort noget forkert, som har fået os til at mærke denne vrede. Således ser Pereboom altså den moralske vrede som en del af det psykologiske og biologiske grundlag for dét, som ofte manifesterer sig som vores retributivistiske trang til at straffe andre aktører for deres stipulerede forseelser.

For at karakterisere det destruktive ved den moralske vrede kommer Pereboom også mere bredt ind på vredens fænomen og dets betydning for os. Mere generelt mener Pereboom, at problemet med vrede som grundfænomen i vores liv er, at der her er tale om en følelse eller impuls, som ofte kan have en række uheldige bivirkninger eller effekter:

[Anger] increases tendencies to overlook mitigating details before attributing blame, to rely on stereotypes and to discount the role of uncontrollable factors in assigning blame. Anger makes us slower to associate positive traits than negative traits with an out-group, and activates an indiscriminate tendency to punish others in unrelated situations without regard for whether their actions were intentional. (Pereboom, 2021, s. 106).

Som det ses af dette citat, er det altså Perebooms standpunkt, at vrede ofte får os til at overse vigtige detaljer ifm. vores moralske domme og generelt gør os mere tilbøjelige til at dømme folk på basis af gruppetilhørsforhold og øvrige ydre parametre frem for deres faktiske gerninger og karakter. Videre betyder vrede ifølge Pereboom, at vi ofte dømmer folk på en måde, hvor vi ikke tager hensyn til, om dét, de gjorde, var intentionelt eller ej – hvilket i sig selv synes at være

⁴¹ Ophavsmanden til dette begreb er som nævnt P. F. Strawson (1962).

moralsk problematisk. På denne måde er det altså Perebooms holdning, at *vrede* som en grundfølelse i mennesket kan være med til at underminere vores dømmekraft og lede os til at handle på mere destruktive, ukontrollerede og irrationelle måder, som ofte vil forhindre mere harmoniske og stabile mellemmenneskelige relationer, både i vores private liv og i samfundslivet som helhed.

Det er netop på baggrund af disse overvejelser, at Pereboom mener, at *free will skepticism* har et potentiale til at gøre os mere forsonlige og mindre destruktive i vores adfærd over for hinanden, idet *free will skepticism* på et filosofisk plan viser os det irrationelle og frugtesløse ved selve den moralske vredes karakter. *Free will skepticism* lærer os, at andre mennesker aldrig er moralsk ansvarlige for deres adfærd i en mere ultimativ forstand, og dermed aldrig *fortjener* at blive lagt for had pga. dét, de har gjort, hvilket kan være med til at gøre det nemmere for os at styre eller regulere vores trang til at handle ud fra vredesmæssige impulser, uden at *free will skepticism* dermed helt kan fjerne vreden som et medfødt eller naturligt aspekt af vores menneskelige psykologi. Som Pereboom skriver:

Accepting [free will skepticism] is all by itself unlikely to alter our psychology so that moral anger is no longer a problem for us. Nevertheless, much of such anger feeds on the presupposition that its object deserves, in the basic sense, blame for a moral offense. Destructive anger in relationships is nourished by the belief that the other is in this sense blameworthy for having done wrong. The anger that fuels ethnic conflicts often results partly from the belief that an opposing group so deserves blame for some atrocity. [Free will skepticism] advocates retracting such beliefs because they are false, as a result of which the associated anger might be diminished, and its expressions reduced. (Pereboom, 2009, s. 172-173).

Som det ses af dette citat, mener Pereboom, at der er en del af vores vrede, der udspringer af processer i os, som det nok vil være svært at kontrollere fuldstændigt; og dog mener Pereboom, at vi godt kan forholde os mere reflektivt til vores vrede, og at *free will skepticism* mere generelt kan få os til at tænke over det forhold, at andre mennesker på et grundlæggende plan ikke er *skyldige* i de ting, som de gør som personer, og at det derfor alt andet lige ikke *giver mening* at rette vores moralske vrede imod dem. Dette kan altså ifølge Pereboom få os til at indtage en mere empatisk position over for andre mennesker, og generelt kan man sige at *free will skepticism* fortæller os, at vi ikke har *grund til* at udtrykke og føle moralsk vrede og forargelse over andres adfærd, fordi de ultimativt ikke selv er skyld i denne adfærd og moralsk ansvarlige for at have udført deres handlinger.

Videre mener Pereboom, at den moralske vrede i visse sammenhænge kan erstattes af øvrige reaktive attituder, der ikke har samme destruktive dimensioner eller aspekter – og i denne forbindelse kommer han med følgende kommentar:

[M]oral anger plays an important communicative role in personal and societal relationships, and one might object that if one were to strive to modify or eliminate this attitude, such relationships might well be damaged. However, when one is wronged in a relationship there are other emotions typically present that are not threatened by [free will skepticism], whose expression can also communicate the relevant information. These emotions include feeling hurt or shocked about what the other has done, and moral sorrow, sadness and concern for the other. Parents might feel intensely sad, and not angry, that their son has driven his car while intoxicated and hurt a pedestrian, and they might be concerned that he will continue to behave in this way. Ordinary human experience indicates that communicating such sadness and concern can be an effective way to motivate avoidance of future bad behavior. Often, communicating anger is at least not required to secure this effect, and frequently it is manifestly an inferior approach, given its deleterious consequences. (Pereboom, 2009, s. 173).

Som vi ser i dette citat, forestiller Pereboom sig altså, at dét at kommunikere *moralsk skuffelse, sorg og ærgrelse* over andre menneskers adfærd kan være med til at kommunikere til dem, at de bør ændre deres adfærd og måske arbejde med at forbedre deres moralske karakter – og at denne kommunikation af moralsk skuffelse og ærgrelse endvidere kan være med til at gøde jorden for forsoning og forbrødring. Til forskel herfra mener Pereboom, at den moralske vrede som fænomen som sådan ikke synes at have noget opbyggeligt ved sig, og at den videre er meget svær at tøjle og kontrollere og derfor ofte vil have skadelige sociale effekter. Med andre ord er det Perebooms samlede påstand, at hvis vi helt eller delvist kan erstatte kommunikationen af den moralske vrede med kommunikationen af moralsk skuffelse over andre aktørers adfærd, så vil dette i langt størstedelen af livets situationer vil være en *fordelagtig strategi*.

I denne sammenhæng kommenterer Pereboom også, at “[t]he virtue of compassion also has a crucial role [in a free will-skeptical outlook], and it undergirds the aims of reconciliation, rehabilitation, and reintegration” (Pereboom, 2021, s. 123).

Caruso er enig med Perebooms ovenstående overvejelser og mener alt i alt, at *free will skepticism* rummer et forsøg på at undgå en form for “*moral anger associated with the reactive*

attitudes of resentment and indignation [which] is often corrosive to our interpersonal relationships and to our social policies” (Caruso, 2018a, s. 212). Ligeledes mener Sapolsky, at noget af dét der står i vejen for, at vi som aktører kan “*shed hate and blame and desire for retribution*” (Sapolsky, 2023, s. 284), netop er tesen om fri vilje og den form for moralske ansvar, som tanken om fri vilje underbygger og legitimerer. Kort sagt kan man altså sige, at det netop er håbet om at (a) udfordre og (b) tøjle eller modarbejde vores moralske vrede som aktører, der gør, at Pereboom, Caruso og Sapolsky langt hen ad vejen er *optimistiske* omkring udsigten til en verden uden troen på menneskets frie vilje.

4.2.1. Kritik af Perebooms, Carusos og Sapolskys tanker om fri vilje og moralsk vrede:

Spørgsmålet er da: Hvordan skal vi med Kane forholde os til Perebooms, Carusos og Sapolskys tanker om vredens betydning og fænomen? Og giver det mening at sige, at vi bliver nødt til at forkaste tanken om vores frie vilje for at kunne bekæmpe vredens somme tider ganske destruktive betydning for vores fælles liv som mennesker?

Jeg vil alt i alt mene, at der er noget sympatisk ved Perebooms perspektiv,⁴² og at det meget vel kan være rigtigt, at vi ville skabe en bedre livssituation for os selv som mennesker, hvis vi faktisk lod os motivere af ønsket om forbrødring og forsoning frem for blot at insistere på vores ret til at føle vrede og harme over andres ugerninger (dvs. moralsk vrede). Min overbevisning er dog i denne forbindelse, at vi meget vel kan have et fornuftigt og afbalanceret forhold til vreden i os selv, også selvom vi fastholder, at vi er væsener med en fri vilje og ultimativt ansvar for vore egne handlinger – og jeg ser således ikke noget behov for at bortkaste tesen om vores frie vilje som mennesker for herved at kunne opnå et bedre forhold til vores egen vrede og forargelse og disses betydning i vore liv.

Nærmere bestemt mener jeg at vi ud fra Kanes libertarianistiske perspektiv må sige, at dét at have fri vilje samt at fastholde troen herpå faktisk synes at være en nødvendig betingelse for at kunne forsøge at regulere sit eget følelsesliv, herunder ens vredesmæssige impulser og tilbøjeligheder, på en nogenlunde hensigtsmæssig og moden og konstruktiv måde, som kan komme både én selv og andre mennesker til gode. Som bekendt definerer Kane fri vilje som “*the power to be the ultimate source and sustainer to some degree of one’s own ends or*

⁴² Som implicit også er Carusos og Sapolskys – Pereboom bruger bare mere tid på emnet.

purposes” (Kane, 2024, s. 10), og på basis af denne definition kan man sige, at vi alt andet lige bliver nødt til at tro på eksistensen af vores egen frie vilje for dermed faktisk at tro på, at vi har en *reel evne* til at definere vores egne målsætninger i livet på en konstruktiv og etisk opbyggelig facon. Kort sagt vil Kane altså mene, at selve min *autonomi*⁴³ eller min evne til at definere reglerne og lovene for min egen adfærd synes at afhænge helt og holdent af min egen frie vilje, og hvis ikke jeg var udstyret med en sådan fri vilje, så ville vi jeg heller ikke kunne *vælge* at forsøge at handle på basis af mit ønske om ikke at lade min egen vrede være bestemmende for min adfærd i en række af mit livs mangfoldighed af situationer.

I forlængelse heraf kunne vi ud fra Kanes teori om *self-forming actions*⁴⁴ konkret forestille os, at jeg står i en situation, hvor min gode ven Martin har stjålet penge fra mig, men jeg véd, at han er meget presset for tiden, fordi han er blevet hjemløs og ikke har nogen indtægtskilder, osv. Spørgsmålet er: Hvordan skal jeg da reagere? Der er måske en del af mig, der har lyst til at reagere meget vredt og bruge min vrede som afsæt for at udtrykke stor forargelse og forbitrelse og måske et vist had rettet mod Martin – det er meget muligt, at jeg mener at han er skyldig i en grov forseelse, og at jeg føler, at han har gjort noget, som udfordrer selve grundlaget for vores venskab. Men det ville også være muligt for mig at kommunikere en form for *moralsk skuffelse* eller *ærgrelse* i situationen, hvis jeg vurderer, at Martin faktisk på en måde ikke har *hele* skylden for dét, der er hændt, givet at han befinder sig i en så prekær situation, at jeg i hvert fald til dels kan forstå ham og sympatisere med ham. På baggrund af Kanes libertarianisme kan vi altså mere overordnet sige, at *valget er mit* i denne situation – der vil ud fra Kanes perspektiv ikke være noget, som *forhindrer mig* i at kommunikere moralsk skuffelse i denne situation; men jeg vil *også* være i stand til at handle på basis af min vrede og frustration, hvis jeg ønsker dette i situationen. Herudover kan man sige, at mit valg i denne situationen så at sige afhænger af, om jeg (1) gerne vil sætte grænser over for Martin ved at udtrykke min vrede eller (2) gerne vil fortælle ham, hvorfor skuffet jeg er, fordi han ikke har levet op til mine forventninger som en god ven – og vi kan også forestille os (3), at jeg ender med at kommunikere begge ting til ham (simultant eller sekventielt), fordi jeg netop er præget af *både* vrede og skuffelse, og at det derved i praksis ikke er muligt for mig blot at vælge at kommunikere én af delene til Martin.

Kort sagt vil jeg ud fra disse tanker mene, at vi ifølge Kanes teori om SFAs altid vil have valget om at *forsøge* at handle på basis af vrede eller skuffelse i en given situation, og at dette valg kan træffes på basis af vores egen frie vilje. Videre vil jeg ud fra denne case mene, at vi også kan sige,

⁴³ Se også del 5.

⁴⁴ Se del 1.

at der synes at være en væsentlig *forskel på*, hvad (a) vrede og (b) skuffelse udtrykker som ”reaktive attituder” – og man kan sige, at hvor vreden typisk udtrykker ønsket om at sætte visse grænser over for folk, dér kommunikerer skuffelsen i stedet, at den anden person ikke lever op til ens forventninger ift. en bestemt type adfærd. Dvs. ikke blot *udtrykket* eller *formen*, men *indholdet* af disse reaktive attituder er øjensynligt forskelligt.

I forlængelse heraf vil jeg mene, at Kane har en god og lavpraktisk pointe i, at det ikke giver mening at sige, at moralsk skuffelse *altid* vil være lige så effektivt og meningsfuldt et svar på ugerninger som vrede og forargelse. Som Kane skriver i *Tapestry*:

To be disappointed and sad in response to the behavior of cruel dictators, child abusers, ruthless seekers of profit, and many other cruel and exploitative individuals without supposing they are deserving of moral blame, any more than any other persons are, seems to undermine our views about morality as well as responsibility. To take an example (at the time I am writing), what are we to think of Vladimir Putin and other dictators of the present and past, such as Stalin, whom Putin is emulating? Not moral blame or anger, but simply disappointment and sadness? (Kane, 2024, s. 260).

Som det fremgår af dette citat, mener Kane altså, at det synes at være et alt for blidt eller behersket respons på forskellige former for misgerninger, at vi blot bliver skuffede eller kede af det – og at der i hvert fald i visse sammenhænge synes at være brug for et stærkere og mere markant følelsesmæssigt respons, der knytter an til vreden og forargelsen. Eksempelvis vil det ifølge Kane ikke være nok blot at udvise skuffelse over Putins beslutning om at invadere Ukraine, hvis vi rent faktisk ønsker at kommunikere til ham (og til verden), at han er ultimativt ansvarlig for sine handlinger, og at han faktisk må stå mål for dét, han har gjort, og de konsekvenser dette har for andre mennesker og disses skæbne. Således mener Kane altså, at det er *nødvendigt* somme tider at gøre brug af og udtrykke moralsk vrede for dermed også at sætte grænser over for andre mennesker og tilkendegive, at man finder disses adfærd etisk uacceptabel.

Ligeledes kan man sige, at hvis jeg gerne vil kommunikere til Martin, at han altså *skal stoppe med at stjæle mine penge*, så er det måske tilstrækkeligt og fint at udtrykke ”moralisk skuffelse” og ærgrelse over hans adfærd første eller anden gang, dette sker. Men hvis Martin nu stjæler fra mig fire, fem eller seks gange, og jeg bliver ved med blot at kommunikere moralsk skuffelse over hans adfærd til ham, og han ikke ændrer kurs og begynder at respektere mine ord, så kan man sige, at der måske kommer et tidspunkt, hvor dét at udtrykke en vis *forbitrelse, forargelse og*

vrede er på sin plads. Man kan sige, at jeg hermed udtrykker, at jeg ikke blot er ked af dét, Martin har gjort og fortsætter mig at gøre, men at jeg også er villig til at sætte visse grænser over for ham – og måske kan man sige, at jeg dermed også kan fortælle Martin, at han rent faktisk skal respektere mig som menneske, hvis han faktisk vil være min ven.

Tilsvarende kunne vi forestille os, at hvis min kæreste har været mig utro, så kan jeg måske i første omgang vælge ”blot” at kommunikere min *skuffelse* over hende til hende – men hvis jeg bliver ved med dette, og hun bliver ved med at være mig utro hver weekend, så vil det måske være berettiget med *et vredesudbrud* eller en kommunikation af *moralsk vrede og forargelse* på ét eller andet tidspunkt. Ja, man kan sige, at hvis end ikke jeg havde *mulighed* for at kommunikere min vrede og forargelse til Martin eller min kæreste, så synes der ikke at være nogen effektiv måde, hvorpå jeg rent faktisk kan fortælle dem, at de skal lade være med at fortsætte deres aktuelle adfærd, og at jeg ikke *finder mig i*, at de bliver ved med at tilsidesætte mine behov og min værdighed og behandle mig respektløst.

Hvad kan vi lære af dette? Jeg vil mene, at disse refleksioner illustrerer to store problemer ved Perebooms, Carusos og Sapolskys forhåbninger angående vores moralske vrede og vores forhold til denne som aktører. Nærmere bestemt kan man for det første sige, at vi har brug for både (a) skuffelsen og (b) vreden og forargelsen i vores katalog af ”reaktive attituder”; og disse reaktive attituder kan i hvert fald ikke til fulde erstatte eller afløse hinanden, simpelthen fordi de kommunikerer forskellige ting. På denne baggrund er jeg enig med filosofen Shaun Nichols i, at Pereboom og hans ligesindede synes at være ”*naively optimistic about the plasticity of human motivation*” (Nichols, 2007, s. 422), og at deres skeptiske filosofi baserer på nogle lidt urealistiske forestillinger om, hvordan vi vis skulle være i stand til at *udskifte* én reaktiv attitude med en anden på tværs af en række af livets situationer.

For det andet vil jeg også med Nichols mene, at cases som de ovenstående viser, at ”*there would be significant losses to human life if we made a concerted effort to eradicate moral anger from our lives*” (Nichols, 2007, s. 422), fordi den moralske vrede i hvert fald også nogle gange har en positiv *social rolle* i kraft af dens evne til at være en adfærdsregulerende instans, dvs. den moralske vrede tillader os at sige fra over for andre menneskers urimelige, selviske eller moralsk problematiske adfærd i forskellige sammenhænge, og vores liv som mennesker ville ikke være helt det samme uden denne mulighed. Og jeg vil på denne baggrund ikke mene, at Perebooms tanker om moralsk vrede helt generelt kan bruges som et normativt eller pragmatisk argument for, at vi burde tilslutte os *free will skepticism* som position.

Konklusion på del 4:

I denne del af specialet har vi kigget på, hvordan Pereboom, Caruso og Sapolsky mener, at der en ”mørk” side af troen på fri vilje, fordi denne tro på fri vilje (1) motiverer et retributivistisk syn på straf og (2) nærer og legitimerer vores moralske vrede som aktører. Jeg har dog forsøgt at vise, at deres argumenter på disse områder er behæftede med en række problemer, og alt i alt ser jeg ikke, at troen på fri vilje har den moralske skyggede eller slagside, som Pereboom, Caruso og Sapolsky hævder, at den skulle have.

Herudover vil jeg mene, at der syner sig at være nogle alvorlige problemer med Perebooms, Carusos og Sapolskys skeptiske position i relation til de ovenstående områder. Det mest alvorlige problem forbundet med deres skeptiske position synes nærmere bestemt at være, at vi ud fra *free will skepticism* ikke kan siges at have det ultimative moralske ansvar for vores egne handlinger som mennesker – og en *internalisering* af denne sandhed synes at lede til en situation, hvor det bliver meget nemt for os som aktører at undskylde forskellige former for moralsk problematisk adfærd med, at vi alligevel ikke har det ultimative ansvar for vores handlinger, og følgelig har andre ikke *ret til* at beskylde os for at have gjort noget forkert. Jeg vil mene, at det er meget nemt at se, hvordan dette kunne lede til en form for livssfære og samfundsorden, hvor vi giver et moralsk *carte blanche* til alle de egoistiske, usolidariske og ondsindede aktører i verden – og jeg vil give Kane ret i, at det på et mere grundlæggende plan derved ikke synes at være helt moralsk uproblematisk at fortælle folk, at de slet ikke har fri vilje og det ultimative ansvar for deres beslutninger og handlinger i livet.

Derudover vil jeg ud fra ovenstående refleksioner mene, at det synes at være tilfældet, at vi faktisk har brug for at kunne udvise og udtrykke *moralsk vrede* i mange af livets situationer for dermed at kunne sætte grænser over for andre aktører – og dette både i private relationer og på den internationale politiske scene, mv. Sagt på en anden måde lader det til at være tilfældet, at selv *hvis* vi kunne erstatte al moralsk vrede med moralsk skuffelse, så er det ingenlunde klart, at dette ville være fordelagtigt eller godt for os på et mellemmenneskeligt og samfundsmæssigt plan. Og på denne baggrund vil jeg samlet set hævde, at det ud fra drøftelserne i del 4 ikke kan siges, at det af praktiske eller moralske grunde skulle være mere fordelagtigt at tilslutte sig *free will skepticism* end f.eks. Kanes libertarianisme.

Del 5: Afrundende refleksioner: Hvad er de dybere moralske implikationer af troen på fri vilje i Kanes forstand?

I den ovenstående del af specialet har jeg kigget nærmere på Perebooms, Carusos og Sapolskys argumenter for, at der skulle være en ”dyster” eller ”mørk” side af troen på fri vilje; og dog jeg har forsøgt at vise, at der alt i alt synes at være grund til at holde fast i tanken om os selv som aktører med fri vilje og det ultimative ansvar for vores egne valg og handlinger i livet.

I denne, afrundende del af specialet vil jeg i forlængelse af ovenstående komme nogle lidt mere generelle refleksioner over spørgsmålet: *Hvad er de dybere moralske og eksistentielle implikationer af troen på fri vilje i Kanes forståelse heraf?*

Jeg vil mene at dette er et ret svært spørgsmål; men det synes at være muligt i det mindste at give nogle overordnede svar her på med basis i Kanes libertarianisme. Nærmere bestemt vil det være min påstand, at der synes at være to mere overordnede konsekvenser eller implikationer af at fastholde troen på fri vilje i Kanes forståelse heraf, dvs. fri vilje i betydningen *“the power to be the ultimate source and sustainer to some degree of one’s own ends or purposes”* (Kane, 2024, s. 10). Den første af disse implikationer har at gøre med vores selvforhold som aktører, og at vi ud fra Kanes teori (populært sagt) må siges at være forfatterne til vores egen karakter, og at vi hermed også motiveres til at undgå dét, Sartre kalder ”ond tro”. Og den anden af disse implikationer har at gøre med, hvordan vi kan og bør anskue vores iboende *værdighed* som aktører.

I det følgende vil jeg forsøge at sætte lidt flere ord på disse to implikationer. Min tese vil ikke være, at disse er de *eneste* implikationer af at tro på fri vilje, som Kane forstår dette – men min vurdering er, at de nok er de *vigtigste*.

5.1. Menneskelige aktører er forfatterne til deres egen karakter gennem deres SFAs:

Den første implikation af troen på fri vilje i Kanes forståelse heraf synes at være, at vi må betragtes som *forfatterne til vores egen karakter gennem vores SFAs i livet*, forstået på den måde at vi gennem SFAs altid er med til at danne og omdanne os selv og vor karakter som personer. Sagt på en anden måde kan man sige, at hvis det virkelig er rigtigt at vi har den form for fri vilje, som Kane tager afsæt i og hævder eksistensen af i sin libertarianistiske filosofi, så betyder dette også, at vi synes at have en særlig pligt eller motivation til at tænke dybere over,

hvordan vi gennem livets mangfoldighed af SFAs kan være med til at forme vores egen karakter og mentale egenskaber som individer.

I forlængelse heraf vil jeg mene, at det synes at være en slags implicit etisk fordring (om man vil), at vi bør tilgå spørgsmålet om vores egne SFAs som menneskelige aktører med en grad af alvor gennem livet – og at dette på en måde er den styrende eller grundlæggende normative antagelse i hele Kanes libertarianistiske position. Sagt på en anden måde vil jeg mene, at der synes at være noget *normativt på spil* i Kanes teori om fri vilje, idet dette ikke blot er en teori om, hvad vi eventuelt skal forstå ved fri vilje som fænomen, men endvidere en teori om, hvordan fri vilje er forbundet med vores ultimative ansvar for vores egne målsætninger og formål og værdier, og hvordan vi tilsyneladende også *bør* påtage os dette ultimative ansvar for, hvordan vi tilgår vores basale livsvalg og prioriteringer som aktører.

I regi heraf vil jeg også mene, at der synes at være en del af Kanes libertarianistiske position, som så at sige tvinger os eller motiverer os til at forholde os mere *kritisk, åben og eksperimenterende* til vores eget værdisæt som aktører samt til selve spørgsmålet om, hvem vi som mennesker egentlig ønsker og tilstræber at blive. Hvis vi eksempelvis tager udgangspunkt i casen med min ven Martin (jf. del 4), der desværre er blevet hjemløs og har stjålet nogle penge fra mig, så kan man sige, at jeg i denne situation stilles over for spørgsmålet om, hvem jeg egentlig at være som menneske, og ud fra hvilke værdier og moralske principper jeg egentlig ønsker at handle i denne situation. Jeg kan vælge at handle på en måde, som vil signalere gavmildhed og barmhjertighed og empati; eller jeg kan vælge at handle på en måde, som vil være udtryk for streghed, koldhed og den holdning, at jeg ”ikke finder mig i noget”, osv. Kort sagt er der forskellige handlemuligheder for mig; og gennem disse handlemuligheder kan jeg ifølge Kane være med til at definere ikke blot min egen livsbane som aktører, men endvidere min relation til Martin og til andre mennesker, som jeg er i kontakt med i mit liv. Jeg har det ultimative ansvar for, hvordan jeg reagerer i denne situation, og dermed kan man også sige, at både noget af mit eget liv og noget af Martins liv er i mine hænder – og at jeg må tage vare på både mit og hans liv så godt, jeg nu kan, og ud fra de principper, som jeg mener er de rigtige.

Endvidere vil jeg i denne forbindelse mene, at der synes at være et aspekt af Kanes libertarianisme, der læner sig op ad eller har noget tilfælles med Sartres eksistentialisme. Mere konkret synes det at være relevant at kigge på Sartres (1956) begreb om ”ond tro” i forbindelse med Kanes filosofi. Som nævnt indledningsvist i specialet, så er ”ond tro” (*mauvaise foi*) for Sartre en form for selvbedrag, hvor individer benægter deres frihed og ansvar for deres valg, ofte

ved at tilskrive deres handlinger til eksterne faktorer eller deterministiske kræfter. Dvs. at en person effektivt set siger noget i retning af følgende til sig selv:

1. "Jeg havde i situationen *ikke andet valg* end at gøre dette, og det *var ikke op til mig, hvad jeg gjorde!*"
2. "Jeg er *nødt til* at være på lige den her måde også fremadrettet, fordi min karakter og identitet jo allerede er etableret på en bestemt måde lige nu – det er bare sådan, jeg *ér!* Og jeg kan ikke (realistisk) lave dette om!"

Kort sagt er "ond tro" for Sartre en måde, hvorpå vi på paradoksal vis træffer en fri beslutning om at nægte os selv vores egen uundgåelige frihed i livet. Vi overbeviser os selv om, at vi *ikke* er frie, og dette gør vi bl.a. for at leve med den angst og usikkerhed, der er forbundet med at anerkende vores egen status som et radikalt frit individ, der har det ultimative ansvar for sine egne handlinger og valg i livet. Man kan således sige, at det er *forståeligt*, at vi somme tider flygter ind i denne "onde tro", men det er dog også ifølge Sartre *beklageligt og et udtryk for en inautentisk selvopfattelse*, når vi går dette.

Ifølge Sartre er denne form for "onde tro" altså farlig, fordi den er en måde at flygte fra den omstændighed, at vi reelt set er *dømt til frihed*, dvs. at vi faktisk ikke har noget andet valg end at træffe frie valg i livets mangfoldighed af situationer. Eksempelvis kunne vi i casen med Claudia (jf. del 1) sige, at hun ifølge Sartre er radikalt fri til at vælge mellem (1) det lukrative job i storbyen, langt fra familie og venner, og (2) det mindre lukrative job i den mindre by, der er tættere på hendes familier og venner. Sartres pointe ville her være, at der er *intet* ved denne situation, der *tvinger* Claudia til at vælge det ene eller det andet, og at det ville være udtryk for "ond tro" for hende at tro andet.

Hvordan kunne denne onde tro komme til udtryk? Eksempelvis kunne vi forestille os at Claudia siger til sig selv, at hun er en "landsbypige", og at hun derfor aldrig ville kunne klare storbylivets jag og det arbejdsmæssige pres forbundet med at være på et større advokatkontor, langt fra hendes rødder og omegn. Herved kan Claudia fortælle sig selv, at *hendes valg er truffet for hende*, idet hendes personlighed og natur som "landsbypige" ikke tillader hende at gå imod denne natur! Man kan herved sige, at hun opererer i "ond tro" ifølge Sartre, da hun ikke anerkender at valget om at gøre *A* eller *B* jo er hendes – det er op til hende selv, om hun vil lade sin identitet som "landsbypige" afgøre hendes frie valg i denne situation.

Tilsvarende kunne vi ifølge Sartre sige, at hvis nu Claudia havde en farmor – Anne – som var *enormt* ambitiøs på Claudias vegne, og som altid pressede hende til at prioritere sine

karrieremæssige interesser over alt andet, så ville dette måske også kunne præge Claudias beslutning. Lad os tænke os, at Claudia taler i telefon med Anne om de to jobmuligheder, og hvordan hun skal forholde sig til disse. Ifølge Annes vurdering bør Claudia selvklart vælge jobbet på det mere lukrative og interessante advokatkontor i storbyen. "Claudia har altid været en klog og dygtig pige", mener Anne, "og det er vigtigt at hun ikke spilder sit potentiale, og at hun faktisk udlever sine drømme om at blive en stor advokat!!"

På denne baggrund har Claudia lyst til at gøre sin farmor glad ved at leve op til de ambitioner, som farmor Anne har på hendes vegne. Ikke desto mindre kan man ud fra Sartre sige at der her er en fare for, at Claudia handler og agerer i "ond tro", fordi hun lader sit valg *determinere* af Annes forventninger til hende. Man kan sige, at hvis Claudia lader sin egen essens som person være defineret af, hvordan hendes farmor ser hende, og hvilke forventninger farmoren har til hendes liv, så lever hun reelt set ikke op til mantraet om, at hun er radikalt fri som person. Hun gør sig selv til en "ting" i verden, der er defineret af noget givet; hun flygter fra sin egen frihed. Sartre ville her sige, at uanset hvad må vi sige, at Claudia jo selv har valget om at *lade være med* at gøre dét, som farmor Anne ønsker, at hun skal gøre. Dvs. Claudia kan selv vælge om hun vil leve op til sin farmors forventninger eller ej. *Denne* frihed kan farmor Anne jo ikke tage fra hende – men det er så at sige denne sandhed, som Claudia glemmer eller flygter fra, når hun lever og handler i "ond tro".

Jeg vil mene at disse overvejelser er meget relevante ift. at vurdere implikationerne af Kanes libertarianisme for vort selvforhold som aktører. Kort sagt kan man sige, at Kane ville give Sartre ret i at vi som individer har en tendens til (1) at undskylde vores egen adfærd gennem henvisningen til alskens deterministiske kræfter i verden, og at vi videre (2) ofte er tilbøjelige til at glemme, at det jo er *os selv*, der har det ultimative ansvar for, hvordan vi handler i livets mangfoldighed af *self-forming actions*. Det er så at sige vores lod som mennesker, at denne usikkerhed og angst er en del af dét at træffe frie valg. Men det er ifølge både Sartre og Kane også nødvendigt, at vi faktisk behandler valgene som *vores*, hvis vi skal blive "herre i eget hus" og leve op til vores medfødte frihed. Eller som Kane skriver:

If we want to be independent sources of activity in the world, we must accept ambivalence, uncertainty, struggle, and conflict within ourselves — all of which are connected to the indeterminacy that is required for free will. (Kane, 1996, s. 214).

5.2. Troen på fri vilje underbygger vores iboende værdighed som menneskelige aktører:

Dette leder mig til dét, som synes at være en anden hovedimplikation af troen på fri vilje i Kanes forståelse heraf – nemlig at fastholdelsen af denne tro synes at være væsentligt for også at kunne begrunde vores værdighed som personer.

Nærmere bestemt kan man ifølge Kane sige, at vores idé om fri vilje er intimt forbundet med *“...a sense of personal worth that we associate with the ultimate sources of our actions and achievements”* (Kane, 1996, s. 87). Med andre ord mener Kane altså, at dét at have fri vilje synes at være en kilde til vores særlige status som væsner med et indre værd eller værdighed, og at dette endvidere kan siges at være koblet til vores forståelse af os selv som den ultimative årsag til vores egne valg og handlinger. I denne forbindelse kommer Kane også med følgende overvejelse baseret på Immanuel Kants filosofi:

The connection Kant envisaged between dignity and ultimacy can be explained as follows. To treat others as moral agents is to assume that, on some matters, the buck stops with them. They, and they alone, are ultimately accountable for their purposes and actions. (As James Felt has put it, the libertarian intuition is that “the buck-stops [here] because it started [here]”).[...] But there is another side to this coin. If I am to be held responsible for my purposes and actions because I am their ultimate source, then I likewise want those purposes and actions to be respected because I am their ultimate source, and not merely because what I do is useful to others. In other words, I want to be treated as an end in myself (with dignity), because I acknowledge ultimate responsibility for my own ends. (Kane, 1996, s. 87).

Kort sagt kan man altså ifølge Kane sige, at dét at vi er væsner med fri vilje indebærer, at vi er bevidste om vores ultimative ansvar for vores egne målsætninger og hensigter som personer, og at det er *dette*, der er en del af grundlaget for og kilden til vores værdighed som personer. Endvidere kan man sige, at Kane hermed også antyder, at det i hvert fald til dels er pga. vores frie vilje som personer, at vi *fortjener* og måske endda *kan kræve af andre* at blive behandlet og betragtet som værende *mål i os selv*, frem for blot som *midler* til andre aktørers gøremål og formål.⁴⁵

⁴⁵ Jf. Kants anden formulering af det kategoriske imperativ (Kant, 1999[1785]). Andre kilder til vores værdighed som personer kunne f.eks. være (1) at vi er mennesker, (2) at vi er fornuftsvæsener, (3) at vi har mulighed for at handle moralsk. Pointen er ikke, at disse *ikke* har betydning, men at dét at

I forlængelse heraf mener Kane, at vores status som væsener med fri vilje også er med til at underbygge det faktum eller forhold, at vi kan og bør se hinanden som *autonome* skabninger. Med en sådan "autonomi" forstås ifølge Kane netop vores evne til at udøve en vis *kontrol* over vores egne vilje og herigennem definere de maksimer, love og regler mv., som vi ønsker at bruge som basis for vores egen adfærd som aktører. Sagt på en anden måde er det altså Kanes tese, at alle mennesker har en vis grad af autonomi, for så vidt de faktisk kan siges at være mennesker, og at autonomi i en bredere forstand kan forstås som denne evne til at definere vores egne målsætninger og formål på en sådan måde, at vi i det mindste kan *forsøge* at handle ud fra disse i forskellige sammenhænge.

Hvad betyder alt dette for vores syn på hinanden som mennesker?

Man kan sige, at hvis Kane har ret i at der er denne indre kobling mellem fri vilje, autonomi og værdighed – en kobling, som vi altså også ser i Immanuel Kants filosofi – så synes det at være tilfældet, at vi er *etisk forpligtede til* at respektere hinandens evne til netop at gøre brug af vores frie vilje i praksis. Eller sagt på en anden måde: Lige meget hvem det er, jeg taler med og interagerer med, så har jeg i en vis forstand en pligt til at respektere og anerkende deres ret til at bruge deres egen frie vilje i praksis forstået som evnen til at definere de ultimative formål og målsætninger med deres egen eksistens. Og *kun* hvis jeg gør dette, respekterer jeg deres værdighed som personer.

En implikation heraf synes eksempelvis at være, at hvis jeg i et parforhold ikke accepterer og anerkender min kærestes ret til at bruge sin egen frie vilje til at definere dette parhold og evt. ønske noget andet i sit liv, så kan jeg på en måde ikke *elske* min kæreste – for så vidt kærlighed rummer dét at behandle min kæreste med værdighed. Dvs. hvis jeg ikke tillader, at min kæreste rent faktisk har rum til og mulighed for at *af- eller bekræfte* selve kærligheden i vores forhold af sin egen frie vilje, så kan denne kærlighed på en måde faktisk ikke eksistere i dens gældende form.

En anden implikation kunne være, at hvis en forælder ikke anerkender sin 2-årige datters eller søns ret til selv at forme en identitet på en sådan måde, at dette barn *en skønne dag* vil være i stand til at definere sine egne ultimative målsætninger og formål i livet, så behandler denne forælder ikke barnet som noget med en iboende værdighed og potentiel autonomi. Med andre ord er det altså etisk befordrende, at forælderen ikke *blot* ser barnet som et *middel* til sine egne

have fri vilje *også* har betydning for vores status som moralske aktører og dermed vores iboende værdighed.

livsplaner og formål og drømme, men *tillige* anskuer barnet som et mål i sig selv og som et væsen, der har *ret til* at definere sine egne formål og målsætninger i livet.

Jeg vil mene at disse to eksempler illustrerer, at der er en meget tæt kobling mellem (1) at anerkende menneskers frie vilje og (2) at være rationelt motiveret og etisk forpligtet til at behandle disse mennesker som *mål i sig selv*, dvs. som væsner med værdighed og en grad af autonomi. Kort sagt kan man sige at jeg ud fra Kanes libertarianisme må behandle alle mennesker som mål i sig selv, netop fordi de har *midlet* til selv at definere deres egne målsætninger og formålene med deres eksistens, og dette middel er netop deres *frie vilje*. Da jeg selv er et væsen med en fri vilje, så må jeg respektere andres ret til at gøre brug af *deres* frie vilje til at definere deres målsætninger og formål i livet.

Jeg vil ikke mene at disse overvejelser viser, at hvis vi tilslutter os Kanes libertarianisme, så *skal* vi tilslutte os Immanuel Kants moralfilosofiske principper og tanker i alle henseender.⁴⁶ Dog vil jeg mene, at vi ud fra Kanes libertarianisme kan sige, at vi alt andet lige er forpligtede til at skabe et samfund og en livssfære, hvor det dog er *muligt*, at vi alle kan gøre brug af vores egen frie vilje til at definere målsætninger og formål i vores liv. Thi et samfund og en livssfære, hvor dette *ikke* er muligt, vil fratage os en del af vores værdighed som personer og underminere et meget væsentligt aspekt af vores menneskelige eksistens og de forhåbninger, der knytter sig til denne. Og man kan sige, at hvis vi *ikke* fastholder troen på fri vilje, så har vi dermed heller ikke noget rationelt grundlag for at argumentere for det gode og nødvendige ved at opbygge et samfund og en livssfære, hvor vores frie vilje spiller en rolle, og hvor dens betydning anerkendes af os selv, institutionerne og magthaverne i samfundet og andre mennesker i almindelighed.

⁴⁶ Kane mener fx at vi (alt andet lige) godt må lyve til øksemorderen, der leder efter vores ven – for hans større analyse af forholdet mellem Kants moralfilosofi og hans egne tanker om moral og etik, se (Kane, 2010a, s. 38-59).

Konkluderende betragtninger:

I dette speciales dele har vi kigget på, hvordan vi nærmere kan forstå Kanes libertarianisme, og hvordan denne position kan forvares imod en række af indvendingerne præsenteret af skeptikerne Derk Pereboom, Gregg Caruso og Robert Sapolsky.

På et mere overordnet plan vil jeg mene, at vi kan konkludere to ting på basis af specialets dele:

Ad 1) For det første vil jeg mene, at selvom specialet har påvist, at Kanes libertarianisme godt kan forsvares over for bl.a. truslen fra moderne neuroscience og Galen Strawsons Basic Argument, så er dette kun to af de mange forskellige indvendinger imod Kanes position, som denne er blevet mødt med i løbet af de seneste årtiers debat herom. Der er mange andre indvendinger imod Kanes position, som det ville være nødvendigt at behandle for at kunne sige at hans position overordnet set er en *forsvarlig filosofisk position* – men jeg gjort mit bedste for at vise at det dog synes *plausibelt*, at hans position også kunne være mere forsvarlig på et mere generelt plan.

Videre vil jeg mene at selvom vi altså *ikke* kan bevise, at vi *ikke* har fri vilje, så betyder dette ikke, at vi omvendt kan påvise, at vi *har* fri vilje – og jeg er på dette grundlag enig i den følgende vurdering, som Kane kommer med i *Tapestry*:

...we cannot show with certainty that such a free will exists. But what we can do, *if we aspire to rationally believe in it*, is do our best to show what such a free will would have to look like if it did exist; [...] and to show that it is an intelligible and possible ideal, by arguing that the most common objections that have been raised to its possibility fail to show that it is an unintelligible or impossible, and by showing how it might be possible (though certainly not proven) given what is known in contemporary physics and other sciences, including the neurosciences. (Kane, 2024, s. 326).

Kort sagt: Det vil ifølge Kane måske altid være et "*open scientific question*", om vi på et empirisk grundlag kan siges at have den form for frie vilje, som hans teori baserer sig på – og denne *usikkerhed om eksistensen af fri vilje* må vi så at sige lære at leve med, hvis vi ønsker at være tilhængere af hans filosofiske perspektiv (eller øvrige perspektiver, der insisterer på at begrunde fri vilje på en naturalistisk basis). Jeg tror at Kane har ret i alt dette – og jeg ser på dette grundlag hans libertarianistiske position som et forsøg på at skabe en teori om fri vilje, som kan *motivere*

os til at tro på fri vilje på et rationelt grundlag, samtidig med at vi dog anerkender, at vi nok aldrig vil kunne *bevise*, at vi har det.

Ad 2) For det andet vil jeg mene, at vi kan benytte dette speciale som et indspark i debatten om, hvad der vil ske, hvis vi *ikke* tror på (eller tror på), at vi har fri vilje. Jeg vil mene at specialet har vist, at der er mange meget interessante tanker og perspektiver forbundet med *free will skepticism* som position; og man kan sige, at Pereboom, Caruso og Sapolsky bidrager med interessante bemærkninger om sammenhængen mellem fri vilje som fænomen og samfundsmæssige og eksistentielle problemstillinger, hvor jeg i dette speciale har behandlet dem der har at gøre med (1) ”moralsk vrede” og (2) retfærdiggørelsen af straf.

Dog er det min overbevisning, at der er så mange problemstillinger forbundet med *free will skepticism*, at vi alt i alt har gode *praktiske* grunde til ikke at gå ned ad denne sti. Og videre mener jeg – som jeg har forsøgt at vise i specialet – at der er gode etiske grunde til at fastholde troen på fri vilje i Kanes forståelse heraf, dvs. som *”the power to be the ultimate source and sustainer to some degree of one’s own ends or purposes.”* Hvorfor? Jo, fordi dette tillader os at se os selv som ultimativt ansvarlige for vores valg og handlinger og i stand til at gøre en forskel i verden – og jeg vil mene at dette er en gunstig måde at se vores eksistentielle situation på, hvis vi rent faktisk skal forsøge at være ordentlige og gode mennesker, der (1) kan tage vare på vores eget liv og (b) forsøge at være noget for andre mennesker i kraft af den måde, vi ønsker at leve livet på. Og herudover giver troen på fri vilje os en god grund til at behandle alle mennesker som skabninger med *værdighed*, for så vidt dét at behandle andre menneskers ret til at definere deres egne målsætninger og formål i livet synes at forudsætte, at vi anerkender, at de faktisk har en fri vilje.

Og netop *derfor* tror jeg, at det er vigtigt for os at bevare troen på fri vilje.

Bibliografi:

- Aristoteles. (2000). *Etikken*. (S. Porsborg, Ovs.) Frederiksberg: Det lille Forlag.
- Banks, W. P. & Isham, E. A., 2009. We infer rather than perceive the moment we decided to act. *Psychological Science* 20(1), s. 17-21.
- Bode, S., & Haynes, J. (2009). Decoding Sequential Stages of Task Preparation in the Human Brain. *Neuroimage* 45(2), s. 606-613.
- Caruso, G. D. (2012). *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lanham, MD: Lexington Books.
- Caruso, G. D. (2014). *The dark side of free will*. Foredrag ved TEDx Talks. Besøgt 26. juni 2025. Hjemmeside: <https://www.youtube.com/watch?v=rfOMqehl-ZA>.
- Caruso, G. D. (2018a). Origination, Moral Responsibility, and Life-Hopes: Ted Honderich on Determinism and Freedom. I: G. D. Caruso (Red.), *Ted Honderich on Consciousness, Determinism, and Humanity* (s. 195-216). Cham: Palgrave Macmillan.
- Caruso, G. D. (2018b). *Stanford Encyclopedia of Philosophy*. Se punkt 2.3: Impossibility of Ultimate Responsibility. Besøgt 9. juni 2025. Hjemmeside: <https://plato.stanford.edu/entries/skepticism-moral-responsibility/#ImpoUltiResp>.
- Caruso, G. D. (2019). Free Will Skepticism and Its Implications: An Argument for Optimism. I: *Free Will Skepticism in Law and Society* (s. 43-72). Cambridge: Cambridge University Press.
- Coyne, J. (2018). *You don't have free will*. Foredrag ved Science Media Center ved Indian Institute of Science, Education and Research, Pune. Besøgt 24. december 2024. Hjemmeside: <https://www.youtube.com/watch?v=hgCPQNeLplk>
- Cyr, T., Flummer, M., & Maoz, U. (2021). *The Free Will Show Episode 23: The Neurophilosophy of Free Will with Uri Maoz*. Besøgt 2. november 2024. Hjemmeside: <https://www.youtube.com/watch?v=UiBPo3YLILw>
- Dennett, D. C. (2003). *Freedom Evolves*. New York: Viking.
- Doyle, R. (2011). *Free Will: The Scandal in Philosophy*. Cambridge, Massachusetts: I-Phi Press.
- Frankfurt, H. (1969). Alternative Possibilities and Moral Responsibility. *Jornal of Philosophy* 66, s. 829-839.

- Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69(3), s. 548-562.
- Gorner, J. (21. april 2013). *Chicago Police Use 'Heat List' as Strategy to Prevent Violence*. Udgivet af Chicago Tribune. Besøgt 25. august 2025. Hjemmeside: http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list
- Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Science* 9(6), s. 290-295.
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research* 126(1), s. 128-133.
- Haggard, P., & Libet, B. (2001). Conscious intention and brain activity. *Journal of Consciousness Studies* 8(11), s. 47-63.
- Harari, Y. N.. (2018). *21 Lessons for the 21st Century*. New York: Spiegel & Grau.
- Hawking, S., & Mlodinow, L. (2012). *The Grand Design*. New York: Bantam Books.
- Kane, R. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Kane, R. (2007). Response to Fischer, Pereboom, and Vargas. I: R. K. John Martin Fischer, *Four Views on Free Will* (s. 166-183). Hoboken: Blackwell Publishing.
- Kane, R. (2010a). *Ethics and the Quest for Wisdom*. Cambridge: Cambridge University Press.
- Kane, R. (2010b). *Rethinking Free Will: New Perspectives on an Ancient Problem*. Foredrag ved College of the Holy Cross i 2010 som del af en konference om neuroscience, evolutionsteori og fri vilje. Besøgt 3. november 2024. Hjemmeside: <https://www.youtube.com/watch?v=o9E5q5w3Czk>
- Kane, R. (2023). Free Will, Complexity, Dynamical Systems, and All That Jazz. *The Harvard Review of Philosophy* 30, s. 1-22.

- Kane, R. (2024). *The Complex Tapestry of Free Will: A Philosophical Odyssey*. New York: Oxford University Press.
- Kant, I. [1785] (1999). *Grundlæggelse af sædernes metafysik*. (T. Bøgeskov, Ovs.) København: Hans Reitzels Forlag.
- Kant, I. [1788] (1996). *Critique of Practical Reason*. (M. Gregor, Ovs.) Cambridge: Cambridge University Press.
- Lemos, J. (2016). Moral Concerns About Responsibility Denial and the Quarantine of Violent Criminals. *Law and Philosophy* 35, s. 461-483.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain* 106, s. 623-642.
- Marchetti, C., & Sala, S. D. (1998). Disentangling the Alien and Anarchic Hand. *Cognitive Neuropsychiatry* 3(3), s. 197-207.
- McGeer, V. (2012). Co-reactive attitudes and the making of moral community. I: C. MacKenzie, & R. Langdon (Red.), *Emotions, Imagination, and Moral Reasoning* (s. 299-325). Psychology Press.
- Mele, A. (2014). *Free: Why Science Hasn't Disproved Free Will*. Oxford: Oxford University Press.
- Miller, E., & Cohen, J. (2001). An Integrative Theory of Pre-frontal Cortex Function. *Annual Review of Neuroscience* 24, s. 167-202.
- Mitchell, M. (2009). *Complexity: A Guided Tour*. New York: Oxford University Press.
- Nichols, S. (2007). After Incompatibilism: A Naturalistic Defense of the Reactive Attitudes. *Philosophical Perspectives* 21(1), s. 405-428.
- Obhi, S. S., & Haggard, P. (2004). Free will and free won't. Motor activity in the brain precedes our awareness of the intention to move, so how is it that we perceive control? *American Scientist* 92(4), s. 258-65.
- Pereboom, D. (2001). *Living without Free Will*. Cambridge: Cambridge University Press.

- Pereboom, D. (2007). Hard Incompatibilism. I: R. K. John Martin Fischer, *Four views on free will* (s. 85-125). Oxford: Blackwell Publishing.
- Pereboom, D. (2009). Free Will, Love, and Anger. *Ideas y Valores* 58(141), s. 169-189.
- Pereboom, D. (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Pereboom, D. (2016). *Free Will is Impossible. Interview with Derk Pereboom*. Udgivet af Moscow Center for Consciousness Studies. Besøgt 10. juni 2025. Hjemmeside: <https://www.youtube.com/watch?v=f5UblvofneY&t=746s>
- Pereboom, D. (2021). *Wrongdoing and the Moral Emotions*. Oxford: Oxford University Press.
- Pinker, S. (2018). *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. London: Penguin Books.
- Sapolsky, R. M. (2023). *Determined: The Science of Life without Free Will*. New York: Penguin Press.
- Sartre, J.-P. (1956). *Being and nothingness: An essay in phenomenological ontology*. (H. E. Barnes, Ovs.) New York: Philosophical Library. Oversættelse af *L'Être et le Néant: Essai d'ontologie phénoménologique*, oprindeligt udgivet af Édition Gallimards, Paris, i 1943.
- Schurger, A., Hu, P., Pak, J., & Roskies, A. L. (2021). What is the Readiness Potential? *Trends in Cognitive Sciences* 25(7), s. 558-570.
- Sehon, S. (2016). *Free Will and Action Explanation: A Non-Causal, Compatibilist Account*. Oxford: Oxford University Press.
- Smilansky, S. (2024). The Reality of Free Will. *Annals of Philosophy* 72(1), s. 99-114.
- Spinoza, B. [1677] (1992). *The Ethics and Selected Letters*. (S. Feldman, Red., & S. Shirley, Ovs.) Indianapolis: Hackett.
- Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 75(1), s. 5-24.
- Strawson, G. (2004). Free agents. *Philosophical Topics* 32, s. 371-403.
- Strawson, G. (2013). The Impossibility of Ultimate Responsibility? I: G. D. Caruso (Red.), *Exploring the Illusion of Free Will and Moral Responsibility* (s. 41-52). Lanham: Lexington Books.

- Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy* 48, s. 187-211.
- Usher, M. (2006). Control, Choice and the Convergence/Divergence Dynamics: A Compatibilistic Probabilistic Theory of Free Will. *Journal of Philosophy* 103(4), s. 188-213.
- van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Clarendon Press.
- van Inwagen, P. (2000). Free Will Remains a Mystery. *Philosophical Perspectives* 14, s. 1-19.
- Vohs, K. D., & Schooler, J. W. (2008). The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating. *Psychological Science* 19(1), s. 49-54.
- Walsh, M. (Filmskaber). (2023). *Free Will? A Documentary*. Dokumentarfilm om fri vilje. Besøgt 17. november 2024. Hjemmeside: <https://www.freewilldocumentary.com/>
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge: MIT Press.
- Wilken-Jensen, C., Rosenbeck, B., & Dübeck, I. (2. december 2024). *Abortens historie*. Danmarks Nationalleksion. Besøgt 28. august 2025. Hjemmeside: https://lex.dk/abortens_historie
- Wolf, S. (1990). *Freedom within Reason*. Oxford: Oxford University Press.
- Youngsteadt, E. (14. april 2008). *Case Closed for Free Will?* Besøgt 10. juni 2025. Udgivet af Science.org. Hjemmeside: <https://www.science.org/content/article/case-closed-free-will>