



Det helsevitenskapelige fakultet – Institutt for Psykologi

Moralpsykologi – Hvordan kan kunnskap om psykologi bidra til moralteori?

//

Moral Psychology – How can psychological knowledge contribute to moral theory?

–

Bendik Cheetham Overrein

*Hovedoppgave PSY-2901 ved profesjonsstudiet i psykologi – Mai 2020, Tromsø
Veileder: Joar Vittersø*

::

Hvordan kan kunnskap om psykologi bidra til moralteori?
How can psychological knowledge contribute to moral theory?

Bendik Cheetham Overrein

Veileder: Joar Vittersø

PSY-2901

Hovedoppgave for graden Cand. Psychol.

Det Helsevitenskapelige Fakultet

Institutt for Psykologi

Universitetet i Tromsø – Norges Arktiske Universitet

Mai 2020



UiT / NORGES ARKTISKE
UNIVERSITET

Forord

I denne oppgaven har jeg studert moral ved tradisjonell moralfilosofi og empirisk tilegnet kunnskap om moralpsykologi. I prosessen rundt valg av tema var jeg hovedsakelig opptatt av å velge veileder og håpe at vedkommende kunne gi meg inspirasjon til å skrive om noe innenfor hans/hennes kompetanseområde. Det viste seg at Joar Vittersø gjorde nettopp det. Jeg opplever at vi har hatt et åpent, ærlig og interessant samarbeid. I løpet av perioden har vi hatt flere berikende faglige og ikke-faglige samtaler. Han har pekt meg i retning av relevant og stimulerende litteratur, og underveis gitt meg tilbakemelding på innhold, struktur og språk.

Jeg ønsker å takke Joar for at jeg fikk muligheten til å lære av hans kunnskap og visdom. Jeg vil også takke min kjæreste og samboer, Martine Utland Glasø, for støtte og tålmodighet i perioder der selvkritikk og pessimisme har vært en utfordring. Til slutt vil jeg takke mine foreldre, brødre, svigerforeldre og venner som – til tross for ikke å i særlig grad ha vært involvert i oppgaven – legger stabile og trygge rammer som muliggjør at jeg kan bruke tid på å fordype meg i mine interesser.

I oppgaven er tekst i kursiv mine oversettelser av sitater fra ulike kilder. Lengre sitater er oppgitt med sidetall dersom det finnes i originalteksten (mitt elektroniske eksemplar av Michael Tomasellos bok «A Natural History of Human Morality» inneholder ikke sidetall). Enkelte ord og begreper er oversatt, det opprinnelige begrepet står da i parentes. Ved mangel på en god oversettelse står begrepet i «anførselstegn» på engelsk. Begrepene lykke og livskvalitet brukes om hverandre og kan i denne oppgaven forstås som et uttrykk for både subjektiv tilfredshet med livet og objektive parametre for velferd.

Sammendrag

Hvordan kan kunnskap om moralpsykologi bidra til moralteori? Kan moral reduseres til én universell, allmenngyldig sannhet, eksempelvis grad av lykke, eller kreves det en pluralistisk tilnærming der moral forstås som bestående av flere uavhengige fundamenter som sammen forklarer ulike menneskers moralske valg? I denne oppgaven diskuteres disse problemstillingene med utgangspunkt i normativ moralfilosofisk teori – Aristotelisk dydsetikk, Kantiansk pliktetikk og Mills og Benthams konsekvensetikk – og deskriptiv kunnskap basert på empiri fra sosialpsykologi, utviklingspsykologi, nevrovitenskap og nevrobiologi. Av betydning for drøftingen av begge problemstillingene er moralfilosofiens kjente innvending om at en ikke kan trekke slutninger fra er til bør, fra fakta til verdier. Slik er det hevdet at vitenskapelig metode ikke kan tilføre noe til spørsmål om hva som er moralsk sant eller riktig. I oppgavens konklusjon vektlegges innsikten i at kunnskap om hjernen og dens fungering kan øke vår bevissthet rundt komplekse sammenhenger i moralske kontroverser. Eksempelvis bakgrunnen for ulike individers stridende interesser, i tillegg til generell kunnskap om forhold som påvirker våre beslutninger. Dette da våre verdier og deres materielle eksistens (nevrale korrelater) er fakta som gir grunnlag for å avveie hva som betyr noe for oss i moralske spørsmål.

Introduksjon

Menneskets utvikling av moral kan spores tilbake til dannelsen av gjensidig avhengighet som grunnlag for samarbeid mellom individer for flere titalls tusen år siden (Tomasello, 2016). Det i større grad målrettede forsøket på å etablere en forståelse av moral har sitt utspring fra antikkens filosofer; Platon, Sokrates, Aristoteles, stoikerne, epikurerne, skeptikerne med flere. Disse, blant andre tenkere og filosofer, eksempelvis Østens Confucius og Mencius, har lagt grunnlaget for en stadig aktuell diskurs om hvordan menneskets moral har blitt utviklet og hvilken moralsk filosofi som er best egnet til å begrunne valg i komplekse samfunnsproblemer. Parallelt har samtlige av verdens religioner formulert normative retningslinjer for hvordan en bør leve som et rettskaffent og godt menneske. Disse er avledet fra én, eller flere guders, vilje, med bare en av de store verdensreligionene som unntak; Buddhismen (Churchland, 2013). Siden opplysningstiden har moral vært gjenstand for interesse blant vestlige skoler innenfor filosofi, psykologi, sosiologi, antropologi og biologi. Hovedtyngden av denne oppgaven vil bestå av aktuell empiri fra disiplinene nevrovitenskap, nevrobiologi, sosialpsykologi og utviklingspsykologi. Som et bakteppe for denne kunnskapen innledes oppgaven med en gjennomgang av de mest innflytelsesrike moralfilosofiske teoriene og tilhørende foregangsfigurer med utgangspunkt i antikken og opplysningstiden. Dette vil forhåpentligvis bidra til en forståelse av hvordan moralfilosofi og moralpsykologi kan sees i sammenheng for bedre å forstå moral.

Antikk moralsk teori

Antikk moralsk teori vektlegger den moralske agenten ved forsøk på å forklare moral. Hvilke verdier og holdninger kjennetegner en moralsk karakter (Parry, 2014)? Aristoteles vektla balanse av avgjørende betydning for å være et moralsk menneske. Han argumenterte for at det finnes en menneskelig funksjon, og at denne er å finne i den menneskelige sjels typiske aktivitet, nemlig fornuft. Han forsto dyder, og etterlevelse av disse, forankret i at sjelens

aktivitet er fullendt, at en lever ut sin menneskelige funksjon. En slik utmerket utøvelse av funksjon knyttet han til utøvelse av praktisk visdom – perfektjonisme i hverdagslig tenkning, valg og atferd. Å ha en følelse av sinne, avhengig av tidspunkt, situasjon, årsak, hvem sinnet er rettet mot, og på hvilken måte en handler som følge av sinnet, verken ekstremt eller mildt, men moderat og balansert, var i følge Aristoteles definisjonen av moralsk dyd i praksis. En passende mengde og intensitet av følelser – gjennomsnittet mellom ekstremer – medfører passende handling. Om forholdet mellom lykke og moral mente han at lykke i seg selv er en dyd for sjelen, mens en moralsk dyd utføres dersom det moderate, balanserte eller gjennomsnittlige uttrykkes i følelser og utøves i handlinger.

Aristoteles antok at dersom en visste hva den menneskelige funksjon var, kunne en si noe om på hvilken måte den menneskelige funksjon kunne utøves utmerket. Denne antagelsen er kritisert fordi den avleder fra et premiss om hva et menneske *er*, til hva det *bør* være (*er/bør*-distinksjonen utdypes senere). Altså å argumentere for at det som er godt, i betydningen av at det fungerer godt, også er godt i moralsk forstand (Parry, 2014). I motsetning til sin mentor Platon, tilbød ikke Aristoteles en formel for hvordan å være moralsk god. Han fremla heller ikke et sett av prinsipper å følge for å oppnå balanse, men fremhevet praktisk visdom, tilegnet gjennom erfaringer, refleksjon og over tid etablert som vaner, som forutsetninger for en balansert atferd (Churchland, 2011).

Moderne filosofiske teorier

Der antikkens filosofi primært forsøker å definere en moralsk agent, er moderne moralsk teori opptatt av å avgjøre hva som kjennetegner moralske handlinger utfra hvilke betingelser og omgivelser handlingene finner sted i. Moderne teoretikere kan deles inn i to overordnede grupper; de som bedømmer en handlings moral ut fra kjente eller forventede konsekvenser, konsekvensialister – konsekvensetikk, og de som vektlegger en handlings overensstemmelse med vedtatte normer, lover eller regler, deontologer – pliktetikk (Parry, 2014).

Den tyske filosofen Immanuel Kant (1724 - 1804) utarbeidet en regelbundet måte å forholde seg til moral på. Han erkjente at rettferdighet var et viktig aspekt ved moral. Sentralt sto tanken om fornuften og menneskets evne til å resonnerer seg frem til løsninger på moralske spørsmål. Snarere enn å formulere én, eller flere, konkrete regler for hva en burde gjøre, utformet han en abstrakt metode for på et vis å kvalitetssikre at ens handlinger var konsistente på tvers av situasjoner. Hans kategoriske imperativ spesifiserte en unntaksfri, ubetinget regel for moralsk atferd basert på rettferdighet og plikt: *Handle bare i tråd med en maksime som du på samme tid kan ønske som en universell lov*. Imperativet var utformet for å filtrere ut hva som var umoralsk, rasjonalisert ved at regelen måtte ivareta ens egen livskvalitet. Idéen var at et menneske ikke konsistent kan bekrefte en regel som går på bekostning av ens egen livskvalitet da etterlevelse av en slik regel vil være irrasjonelt (Churchland, 2011).

Filosofen Patricia Churchland (2011, p. 174) formulerer en tydelig test på hvorvidt imperativet faktisk fungerer som et filter for moralske/umoralske handlinger slik Kant hevder: *kan vi karakterisere en universelt bindende regel som åpenbart er en umoralsk regel, men som kan etterleves uten kontradiksjon eller irrasjonalitet?* Svaret er ja, illustrert ved følgende hypotetiske regel: Alle hvite menn over 50 år med en formue på over 100 millioner kroner bør, dersom de ikke betaler minst 50 % skatt, fengsles. La oss si at dette er et forslag til en universell regel og at vi er enige om at den er umoralsk. Fra et kantiansk perspektiv ville jeg, dersom jeg passet beskrivelsen, ikke kunne godta en slik regel fordi det i verste fall ville medført fengselsstraff, tilsynelatende er regelen da inkonsistent og irrasjonell. Premisset står imidlertid for fall da jeg kunne tenkes å rasjonelt være enig i at jeg, gitt en stor formue, burde blitt beskattet strengt, eller fengslet hvis ikke. Altså kan de mest umoralske handlinger forsvares med det kategoriske imperativ så lenge den (u)moralske agenten erkjenner at regelen som følges også gjelder en selv, til og med dersom utfallet er døden. Eksempelvis ved å håndheve en regel som postulerer at alle immigranter bør tortureres som straff for å utvanne et lands kulturelle identitet,

begrunnet i at en ville anerkjent lignende konsekvenser dersom en selv emigrerte til et annet land. Problemet demonstrert her antyder at et svar på hva som er en moralsk rett handling vanskelig kan avledes fra en utelukkende rasjonell og konsistent fremgangsmåte (Flanagan, 2007).

Konsekvensialisme bedømmer en handling som moralsk rett kun ut fra konsekvenser som er direkte forårsaket av handlingen. Omgivelsene handlingen finner sted i, og foranledningen til handlingen, anses som irrelevant. Innunder konsekvensialistisk filosofi finner vi utilitarismen, utviklet av de britiske filosofene Jeremy Bentham (1789) og John Stuart Mill (1861), begge innflytelsesrike på 1800-tallet. Filosofien vekter konsekvensene av ens handlinger som overordnet hvert enkelt menneskes rettigheter og prinsipper utledet fra resonnering (Kants imperativ), samt religiøse normative tekster. Forenklet fremholder den en moral der målet er å maksimere total lykke for flest mulig mennesker. Det kan tenkes at en handling medfører en økning i lykke for et høyt antall mennesker, men samtidig fører til en forholdsvis større reduksjon i lykke for en liten gruppe mennesker. En slik handling vil ikke være i tråd med utilitarismen da den store gruppens økning i lykke i dette tilfellet ikke oppveier den mindre gruppens nedgang i lykke (Sinnott-Armstrong, 2019). Rent matematisk vil det imidlertid finnes tilfeller der en stor gruppes lykkegevinst kan oppveie en liten gruppes lykketap, eller motsatt, der en liten gruppes lykkegevinst oppveier en stor gruppes lykketap. Et av utilitarismens fortrinn er at den tilbyr en mulighet til å vekte betydningen av ulike verdier: Oppveier verdien om ytringsfrihet verdien om personvern og privatliv, eller forholder det seg motsatt? Utilitarismens svar er at ingen verdi har absolutt forrang, men at verdier må balanseres slik at den normativt sett mest moralsk riktige balansen er den som gir de totalt sett beste konsekvenser, målt i lykke (Greene, 2013). Et forsøk på å balansere, vurdere og predikere konsekvenser av ulike valg er ikke enkelt å gjennomføre i praksis, og gir opphav til en større diskusjon (Sinnott-Armstrong, 2019).

Bentham betydret at vi i våre valg generelt burde håndheve et mål om å maksimere lykke generelt sett. Mill forsvarte en noe mindre radikal versjon av dette. For Mill var handlinger som medførte skade på andre (vold og overgrep) galt og straffbart. Eksempelvis bør en ikke punktere en annen persons bildekk, fordi dette potensielt vil kunne medføre fysisk skade på vedkommende (se for argumentasjonens skyld bort fra innvendingen om at punkteringen av bildekket er en form for skade på vedkommende uavhengig av om det resulterer i trafikkuhell og gir mer alvorlige helsemessige konsekvenser). Mill mente at handlinger som ikke ga slike konsekvenser ikke burde bedømmes som gale. En er ikke nødt til å avstå fra klatring som fritidsaktivitet da dette som regel ikke skader noen. Dette på tross av at en kunne valgt å bruke tiden på andre aktiviteter som med større grad av sannsynlighet ville resultert i totalt sett høyere lykke. Mills kjente eksempel illustrerer dette: En mann ønsker å gå over en bro og spør en kjentmann om brua er farlig å krysse. Kjentmannen sier at ja, brua er farlig å gå over. På tross av dette er mannen i sin fulle rett til å forsøke å krysse brua da den eneste han eventuelt skader ved å falle ned er seg selv. Oppsummert: Etter Mills syn er en handling som direkte forvolder andres skade eller ulempe (reduksjon i livskvalitet) moralsk gal, mens en handling som ikke skader andre, men som prioriteres over en handling som kan medføre totalt sett høyere livskvalitet, er moralsk akseptabel. En maksimerende konsekvensialist kunne hevde at utøvelse av fritidsaktiviteter som personlig gir økt livskvalitet burde vike plass for en aktivitet som ga større total økning i lykke, eksempelvis arbeid som leksehjelp i Røde Kors på fritiden. For Mill er imidlertid dette en absurd utledning fra tanken om å unnlate å utøve skadelig oppførsel (Churchland, 2011).

Problemet med Mills forståelse er at det til en viss grad alltid er slik at vi forvolder andre en form for skade uansett hva vi gjør. Det å leve innebærer for eksempel et behov for energi, en energi det er konkurranse om. Den enes død er på et grunnleggende vis den andres brød. Planter konkurrerer for eksempel om lys og næring med hverandre; et grantre eller en granskog

vil derfor ved sin blotte eksistens skade annet liv, fordi den utnytter lys som ellers kunne tilfalt andre planter og organismer. I eksempelet over er det slik at mannen ikke bare skader seg selv dersom han faller ned fra brua. Han vil også påføre andre lidelse; kjentmannen, egen familie og vennekrets. En kan ikke utsette seg selv for fare uten å påføre andre lidelse (gitt at de andre har et normalt følelsesliv og er klar over at en utsetter seg for fare). Følgelig er spørsmålet aldri *om* handlinger medfører skade på andre, men *hvor mye* skade som påføres, og hvor mye skade som kan aksepteres, både for en selv og de rundt.

Moralpsykologi

I innledningen har jeg gitt en kort innføring i ulike normative moralfilosofiske teorier. Normative i den forstand at de forsøker å argumentere for hva en til enhver tid bør vektlegge for å opptre moralsk, enten ved å henvise til karaktertrekk hos det handlende individet eller til kjennetegn ved selve handlingen og dens omgivelser. Mot dette filosofiske grunnlaget kan vi anvende funn fra tverrdisiplinær forskning for bedre å forstå hva moral er rent deskriptivt. Dette anliggendet er tilknyttet den første problemstillingen:

1. Hvordan kan kunnskap om psykologi bidra til moralteori? Med andre ord; på hvilken måte kan kunnskap om menneskers adferd og kognisjon bidra til teorier om moral?

Med kjennskap til normativ moralfilosofisk teori og deskriptiv moralpsykologisk kunnskap kan vi deretter drøfte hva som kan være den rette moral i praksis. Denne diskusjonen er tilknyttet den andre problemstillingen:

2. Kan moral reduseres til én universell, allmenngyldig sannhet, eksempelvis grad av lykke, eller kreves det en pluralistisk tilnærming der moral forstås som bestående av flere uavhengige fundamenters som sammen forklarer ulike menneskers moralske valg?

Det videre innholdet i oppgaven er forankret i lesning av bøker forfattet av følgende psykologer/filosofar: Patricia Churchland, Joshua Greene, Jonathan Haidt, Sam Harris, Michael Tomasello og «The Atlas of Moral Psychology». Jeg vil og referere til relevante enkeltstudier.

I denne moralpsykologiske delen vil jeg beskrive evolusjonær teori om moral, nevrobiologiske mekanismer på hormon-nivå som grunnlag for det faktum at vi bryr oss om andre mennesker, og beslutningstaking i moralske valgsituasjoner.

Evolusjonær basis for moral

Greene (2013, p. 22) formulerer følgende vitenskapelige definisjon av moral: *Moral er et sett av psykologiske tilpasninger som tillater ellers egoistiske individer å høste fordelene av samarbeid.* Gitt at dette er en presis definisjon; hvordan har det blitt slik at angivelig egoistiske individer har utviklet evnen til å samarbeide med hverandre?

På gen-nivå er vår forståelse av menneskelig evolusjon opplyst av «the Selfish Gene» (Dawkins, 1976), forenklet forklart en bok som beskriver evolusjon ut fra mekanismene knyttet til gens overlevelse og reproduksjon. I noen grad kan bokens budskap ha bidratt til en forståelse av at seleksjon på andre nivåer enn gener, altså individ-og gruppenivå, ikke forekommer (Haidt, 2012). Det er teoretisert at seleksjonspress på gen-nivå gjør et menneskelig individ tilbøyelig til å opptre oppofrende på vegne av sine slektninger da de bærer felles genmateriale («kin selection»); ens brødre og søstres reproduksjonssuksess er til en viss grad lik ens egen (Hamilton, 1964). Hamiltons regel (utarbeidet med utgangspunkt i eusosiale insekter) beskriver at altruistisk atferd utvikles dersom $br - c > 0$, der c uttrykker det altruistiske individets kostnad, b er mottakerens fordel og r er koeffisienten for genetisk likhet som spenner fra 0 til 1 (identiske tvillinger). Dersom r er 0,5 er en kostnad lik 1 verdt å ofre dersom mottakerens fordel er minst 2 ($2 \times 0,5 - 1 = 0$).

Regelen var i utgangspunktet et alternativ til teorien om gruppeseleksjon. Hamilton innså senere at en økning av antallet kopier av et altruistisk gen ikke var det samme som en relativ økning i levedyktighet hos altruistiske individer i samme gruppe som ikke-altruistiske individer. I samarbeid med George Price fant de at altruisme ikke selekteres innenfor en gruppe selv om gruppen består av slektninger; altruisme utvikles i den totale populasjonen bare ved

mellomgruppe-seleksjon. Likningen predikerte fortsatt altruisme i den totale populasjonen, men i stedet for å utelukke gruppeseleksjon var den nå avhengig av slik seleksjon for å gi mening. Koeffisienten for genetisk likhet, tidligere forstått som sannsynligheten for arvbart likt genetisk materiale, var nå en indikator for genetisk variasjon mellom grupper. Når $r = 0$ impliserer det at grupper dannes tilfeldig. Dersom $r = 1$ er individer i hver gruppe genetisk identiske. Naturlig seleksjon forekommer da på gruppenivå. Med andre ord forekommer seleksjon mellom gener i et individ, seleksjon mellom individer innenfor en gruppe, og seleksjon mellom grupper i en populasjon (Wilson & Wilson, 2008). Menneskelig evolusjon kan da forstås som multinivå-seleksjon: *seleksjon av slektinger opererer på gen-nivå, gruppeseleksjon opererer på sosial gruppe-nivå; og mutualisme og resiprositet opererer på en individuell organisms nivå* (Tomasello, 2016). En konseptualisering av evolusjonære mekanismer som forekommende på ulike nivåer – gener, individer og grupper – er av moralpsykologisk betydning da den åpner for forståelsen av mennesket som kapabel til å midlertidig utsette selvsentrerte behov til fordel for samarbeid med andre individer. Dette i stedet for å utelukkende operere som et egoistisk, asosialt vesen motivert for å spre sine gener videre i konkurranse med andre individer (Tomasello, 2016).

Michael Tomasello's definisjon av moral avviker noe fra Greene's forståelse av individer som i utgangspunktet egoistiske: *Fra et evolusjonsperspektiv er moral en form for samarbeid. Samarbeid forutsetter at individer enten undertrykker egeninteresse eller likestiller denne med andres* (M. Tomasello & Vaish, 2013). Tomasello (2016) foreslår at mennesker har gjennomgått en biologisk utvikling som fremmer investering av energi med mål om å øke andre menneskers livskvalitet. I et første steg utviklet tidlige mennesker seg fra et stadium av individuell egoisme til annen-persons moral («second-personal morality»). I et andre steg, som følge av folketallsøkning, delte de seg opp i mindre grupper og samarbeidet i konkurranse mot andre grupper om ressurser, hvorpå «objektiv» moral bestående av normer for rett og galt

vokste frem. Fortellingen hviler på hypotesen om tidlige menneskers gjensidige avhengighet av hverandre under gitte sosioøkologiske betingelser. Gjensidig avhengighet førte til kognitive adaptasjoner, eksempelvis den psykologiske kapasiteten delt intensjonalitet («shared intentionality»). Dette var i seg selv ikke et moralsk fenomen, men samspillet med annenpersons målorientering («second-personal agency»), en sosial-interaktiv funksjon basert på gjensidig respekt mellom to individer, og felles forpliktelse («joint commitment»), sosial selvregulering basert på samarbeidende rasjonalitet, resulterte i individer med gjensidig respekt og verdighet for hverandre. Disse individene opplevde det som en oppgave å etterleve sosiale avtaler.

Tomasello (2016) beskriver det slik: *Tidlige menneskers fornemmelse av gjensidig respekt og rettferdighet mot partnere kom hovedsakelig fra en ny type samarbeidende rasjonalitet der det ga mening å gjenkjenne ens avhengighet av en partner for samarbeid, til et punkt der en oppga noe av kontrollen over ens handlinger til det selv-regulerende «vi» skapt av en felles forpliktelse. Dette «vi» var en moralsk kraft fordi begge partnere anså det som legitimt, basert på at de hadde skapt det selv spesifikt for selvregulerende hensikter, og at de begge så deres partner som genuint verdig deres samarbeid.* Slik utviklet mennesket seg fra en tilstand av strategisk samarbeid – eksempelvis koordinert jakting på større byttedyr – til moral. Tomasello presiserer at hypotesen om gjensidig avhengighet beskriver proksimale (i et evolusjonsperspektiv nylig utviklede) psykologiske mekanismer. Mekanismene er vist ved tidlige og moderne menneskers økende samarbeid med hverandre hvis evolusjonære utgangspunkt var andre primaters (eksempelvis sjimpansens) sosiale natur.

Angående Trivers (1971) teori om resiprok altruisme som forklaring på samarbeid mellom mennesker som ikke er beslektet, gjennom at en part utfører en tjeneste, som deretter blir gjengjeldt av en annen, påpeker Tomasello at det uten en sosial kontrakt mellom de to partene er vanskelig å forstå hvordan slik altruisme psykologisk sett fungerer. Det første

problemet er at teorien ikke forklarer den første handlingen av altruisme, som i så tilfelle må være et resultat av naivitet, optimisme eller tilfeldigheter. Videre vil parten som er mottaker av den altruistiske handlingen ha et sterkt insentiv til ikke å gjengjelde tjenesten. Uten noen form for enighet står resiprositet uten noen rasjonell eller emosjonell kraft til å motivere altruistisk atferd på egen hånd (Tomasello, 2016).

Gjensidig avhengighet tar, i motsetning til resiprok altruisme, høyde for betydningen av sosiale relasjoner for samarbeid. Roberts (2005) «stakeholder model» uttrykker at individer har en innsats («stake») i spesifikke andre individers livskvalitet, eksempelvis deres partnere. For en individuell altruistisk handling gjelder: $sB > C$. Denne er sammenlignbar med Hamiltons regel da B representerer den reproduktive fordelen for aktøren, som må være større enn aktørens kostnad (C) for å utføre handlingen. B er betinget av aktørens innsats (s) i mottakeren av handlingen, lignende Hamiltons koeffisient for genetisk likhet. Slik tilsvarer s aktørens egeninteresse av at mottakeren er i live og i stand til fremtidig samarbeid. Et viktig poeng er at modellen fungerer asymmetrisk, det er aktørens vurdering av mottakerens betydning for en selv som gjelder. Hvorvidt mottakeren har en innsats (s) i aktøren er irrelevant. Slik vil aktøren være avhengig av et annet individ som varsler om fare. Det vil da være i aktørens interesse å hjelpe varsleren, et tilfelle av gjensidig avhengighet.

Ulikt klassisk resiprositet er ikke den altruistiske aktørens atferd betinget av en forventet eller faktisk respons fra mottakeren. Mottakeren fortsetter å handle slik som ellers – som varsler, jaktpartner, eller i par for reproduksjon – fordi det er i hans/hennes interesse. Biproduktet er at det belønner individet som opptrer altruistisk. Altruisme er slik ikke et usannsynlig tilfelle i kamp mot krefter som fremmer individualisme på bakgrunn av naturlig seleksjon. Det er en integrert del av det sosiale livet til alle organismer som lever i gjensidig avhengighet med andre – opp til et matematisk punkt. Resiprositet er ikke egnet til å forklare variasjonen i menneskelig moralpsykologi; overholdelse av løfter, håndhevelse av sosiale

normer, følelsen av harme og regulering av ens handlinger gjennom en opplevelse av ansvar, forpliktelse og skyld. Gjensidig avhengighet som evolusjonært legger til rette for en opplevelse av et «vi» (delt intensjonalitet spesielt i form av felles forpliktelse) og upartiskhet i form av selv-andre ekvivalens synes å bedre forklare denne variasjonen (Tomasello, 2016).

Tomasello (2016) fremhever også skillet mellom proksimale psykologiske mekanismer og evolusjon med evidens som viser at små barn er sympatiske uten noen strategisk kalkulering for adaptasjon på et psykologisk nivå. *Ved å sidestille psykologisk og evolusjonær terminologi kan vi kalle slike handlinger moralsk-adaptive, på den måten at individet hjelper basert på en genuin følelse av sympatisk bekymring, i uvisshet om at det for en selv er adaptivt fordelaktig (...) en del av forklaringen på det evolusjonære grunnlaget for menneskelig moralpsykologi kan derfor ansees å være en versjon av det noen kaller en «feil», i betydningen at individet tror det handler ut fra en genuin bekymring for en annen, selv om han/hun på det evolusjonære nivået selv drar nytte av det.* Han trekker en parallell mellom ontogenetisk og fylogenetisk utvikling ved argumentasjon om at barns utvikling følger den samme kvalitative endringen i to steg som menneskeheten, beskrevet over. Barn fra 1 til 3 år er indre motivert til å hjelpe andre barn spontant. Ved samarbeid deler de ressurser rettferdig i tillegg til å forvente at andre overholder felles forpliktelser. Dette indikerer at de innehar en naturlig annen-persons moral som kommer til uttrykk ved interaksjon med andre barn i løpet av de første tre leveår (Tomasello, 2016).

Først etter tre år synes barn å forstå sosiale normer som delte forventninger innenfor deres egen sosiale gruppe. Da begynner de å håndheve hvorvidt andre følger gruppens normer. Denne adferden lar seg ikke enkelt forklare av egoistiske motiver, men kan sees som uttrykk for barnets forståelse av å være en del av en kulturell identitet, et «vi», hvor alle medlemmer gjør ting på lik måte (Schmidt & Tomasello, 2012). Fra tre-fire års alder begynner barn å praktisere upartiskhet, generaliserbarhet og autonomi fra autoriteter ved beslutninger som omhandler rettferdighet, likhet og rettigheter (Gray & Graham, 2018). Påstanden om at barn

gjennomgår steg av moralsk utvikling med tidsavgrensede kvalitative sprang, er i samsvar med Jean Piaget (1948), og de forstår interaksjon med jevnaldrende, jevnbyrdige individer med lik makt og innflytelse hvor en er avhengig av argumentasjon og forhandling som avgjørende for utvikling av moralske konsepter og holdninger. Overholdelse av normer kun basert på voksnes autoritet er ikke tilstrekkelig, understøttet av at en autoritativ foreldrestil fører til at barn i mindre grad internaliserer ønskede verdier og utviser mer strategisk normativ adferd. En induktiv foreldrestil med vekt på å gi barn årsaker bak handlinger, fører til mer internalisering av verdier, og derav moralsk selv-regulering (Hoffman, 2000).

Sammenfattet gir de moralpsykologiske hypotesene i denne seksjonen opphav til deskriptiv moralteori som foreslår følgende tre ting: at samarbeid forstås som resultat av seleksjon på ulike nivåer – multinivå-seleksjon; at gjensidig avhengighet er av grunnleggende betydning for utvikling av bestemte psykologiske mekanismer som kom til å utgjøre moral; og at de samme psykologiske mekanismene som ligger til grunn for menneskets moralske utvikling også gjør seg gjeldende på individnivå observert ved barns tidlige utvikling.

Nevrobiologi

Hvilke nevrobiologiske prosesser kan tenkes å ligge til grunn for de psykologiske mekanismene som driver samarbeid og der igjennom moral? Pattedyr utviklet evnen for omsorg til å inkludere, på bakgrunn av økologiske og overlevelsesrelaterte årsaker, slektninger, venner og i noen tilfeller fremmede (Churchland, 2011, 2013, 2019). Pattedyrs store behov for energi resulterer i en reproduksjonsstrategi ulik den hos kaldblodige dyr; beskyttelse og omsorg til få avkom i kontrast til eksempelvis skilpaddens masseproduksjon av egg hvis avkom senere overlates til seg selv. Churchland (2011) foreslår at menneskets behov for tilknytning forsterket av smerten ved separasjon og velbehaget ved sosial tilhørighet, og drevet av komplekse nevrologiske forbindelser og nevrobiologiske prosesser, utgjør grunnlaget for moral. Det finnes en betydelig mengde forskning som antyder at neuropeptidene Oxytocin (Oxt) og Vasopressin

(Avp) i tillegg til neurotransmitteren Dopamin spiller en viktig rolle ved omsorg og tilknytning hos andre pattedyr (Walum & Young, 2018). Tilknytningen innebærer et ønske om å være sammen med og ta hånd om sitt avkom, samt opplevelsen av separasjon fra sine barn som stressutløsende. Omlag 3 % av pattedyr viser tilknytning til sin make (Porges & Carter, 2011).

Interessante funn fra forskning på de beslektede «prairie vole» (PV, latin: *Microtus Ochrogaster*) og «mountain vole» (MV, latin: *Microtus Montanus*) foreligger. Da disse to dyrene er nært beslektet, men utviser ulik sosial atferd er det stilt spørsmål ved hva som er forskjellen i deres respektive hjerner? PV utviser flere former for adferd til grunn for et monogamt levesett som ikke er observert hos MV. Dette gjelder partnerpreferanse (parring med partner heller enn fremmed), selektiv aggresjon etter parring (rettet mot fremmede fra samme art), foreldresamarbeid om omsorg for avkom (fedres omsorg for barn er uvanlig blant pattedyr), og tilfellet av at juvenile individer tar vare på yngre søsken. Hos PV og den monogame slektingen «pine vole» (PiV, latin: *Microtus Pinetorium*), er det påvist høyere reseptortetthet for Oxt i flere deler av hjernens belønningssystem – nucleus accumbens (NAcc), medial prefrontal korteks (mPFC) og «bed nucleus av stria terminalis» (BNST) – sammenlignet med de promiskuøse «meadow vole» (MwV, latin: *Microtus Pennsylvanicus*) og MV (Young, Gobrogge, Liu, & Wang, 2011). Hos PV og MV er det generelt høyere mengde Oxt hos hunner enn hos hanner, motsatt for Avp. Det synes som om ulik tetthet av reseptorer for disse to hormonene tydeligst forklarer forskjellen i lang tids tilknytning mellom hann og hun, samt samarbeid om oppfostring av avkom. Dersom PVs reseptorer blokkeres eksperimentelt slik at Oxt og Avp ikke kan binde seg, medfører det at makene ikke knytter seg til hverandre etter første parring, i tillegg til at de ikke utviser sosial atferd typisk sett hos PV (Churchland, 2011, 2013, 2019).

Oxt og Avp er imidlertid ikke tilstrekkelig. Makenes gjenkjenning av hverandre forutsetter læring mediert av dopamin. Blokkering av dopaminreseptorer resulterer i at hannen

og hunnen ikke gjenkjenner hvem de paret med. Resultatet er det samme – tilknytning uteblir (Churchland, 2013). Tetthet av reseptorer for Oxt (Oxt-r) er videre presist predikert av genetisk variasjon (Walum & Young, 2018). Dersom nyfødte individer av PV med lav tetthet av Oxt-r i NAcc utsettes for omsorgssvikt er resultatet manglende utvikling av preferanse for prososial atferd med partner sammenlignet med en fremmed. Individer med tilsvarende reseptortetthet som ikke er utsatt for neglekt utvikler derimot sterk partnerpreferanse. Grad av omsorgssvikt påvirker ikke partnerpreferanse hos PV-individer som har høy tetthet av Oxt-r (Walum & Young, 2018).

Flere studier gjort på ulike primater styrker antagelsen om Oxt's rolle for sosial tilhørighet. Det er funnet positiv sammenheng mellom mengde Oxt i urin og ivaretagende fysisk kontakt («grooming») mellom sjimpanser uavhengig av genetiske bånd og seksuell interesse. Forfatterne foreslår Oxt som en underliggende forklaring på hvordan nære relasjoner er utviklet mellom individer som ikke deler slektskap. Følgelig antyder resultatene at vedvarende relasjoner preget av samarbeid mellom individer som ikke er beslektet medieres av hormonelle og ikke bare kognitive mekanismer (Crockford et al., 2013). Videre er intranasalt administrert Oxt på rhesusaper vist å dempe sosial vaksomhet (Ebitz, Karli, & Michael, 2013), samt å modulere hjerneaktivering representert ved fMRI ved selektivt å redusere respons til bilder av både aggressive og engstelige ansikter i hjerneområder assosiert med prosessering av ansikter. Aktivering ved visning av nøytrale og beroligende ansikter forble uendret etter Oxt-administrasjon (Liu et al., 2015). Hos mennesker er det funnet indikasjoner på sammenheng mellom Oxt og regulering av sosial adferd ved å se på i hvilken grad genetiske variasjoner av Oxt-r genet er assosiert med monogami og sosiale relasjoner (Walum et al., 2012). Det er også funnet at Oxt øker positiv kommunikasjon, reduserer kortisolnivået ved konflikter (Ditzen et al., 2009) og er en nevrobiologisk årsak bak tillit hos mennesker (Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005).

Basert på funn av typen nevnt over, er det foreslått at intranasal injeksjon av Oxt (Dal Monte, Noble, Turchi, Cummins, & Averbeck, 2014) kan endre menneskelig adferd, eksempelvis som behandling hos individer diagnostisert med autismespektrumlidelse. Flere studier, både med mennesker og primater, rapporterer lovende resultater (Anagnostou et al., 2014; Guastella et al., 2010; Parr et al., 2016). En gjennomgang av litteraturen konkluderer imidlertid med manglende statistisk styrke, bias og at det meste av rapporterte resultater sannsynligvis ikke er sanne effekter. Forfatterne maner til sunn skepsis inntil flere studier som tilfredsstillende påpekte mangler er gjort. Studier på bruk av intranasal Oxt med formål å fremme sosial adferd er i tidlig utvikling og behøver metodologisk og teknologisk utbedring (Walum, Waldman, & Young, 2016).

Oppsummert synes det å være bred enighet om at Oxt og Avp i komplekse interaksjoner med ulike nevrotransmittere, deriblant Dopamin, medierer sosial kontakt. Moral synes å være dypt biologisk fundert på en slik måte at vi ønsker å være sammen med andre og at vi finner tilknytning belønnende (Churchland, 2011). Slik dokumenterer denne seksjonen det biologiske grunnlaget for et individs tilbøyelighet til medfølelse, bekymring og velvilje for et annet individ, sammenfallende med moralfilosofiens oppmerksomhet knyttet til *det gode* – en moral for sympati.

Psykologiske teorier om moralsk beslutningstaking

Greene (2013) argumenterer for at vår moralske beslutningstaking styres av to ulike prosesser. Han beskriver de metaforisk som to ulike funksjoner på et kamera; automatisk funksjon – rask, intuitiv, følelsesstyrt, lite fleksibel; og manuell funksjon – gjennomtenkt, rasjonell, langsom men fleksibel («dual-process theory»). Teorien er sammenlignbar med Kahneman's (2012) anerkjente teori om system 1 og 2-tenkning. Greene har drevet eksperimentell utprøving av ulike versjoner av det klassiske tankeeksperimentet (heretter

benevnt «gangbru») der en jernbanevogn er på kollisjonskurs med fem arbeidere (Thomson, 1985). Noe modifisert lyder problemet (Greene, 2013, p. 113):

En vogn er i bevegelse med retning mot fem jernbanearbeidere som vil dø dersom vognen fortsetter på sin nåværende kurs. Du står på en gangbru over togsjennene mellom den kommende vognen og de fem arbeiderne. Ved siden av deg er en jernbanearbeider med en stor ryggsekk. Den eneste måten du kan redde de fem arbeiderne på er å dytte denne mannen av bruene og ned på skinnene under. Mannen vil dø som resultat, men hans kropp og ryggsekk vil forhindre at vognen når de andre arbeiderne. (Du kan ikke selv hoppe fordi du alene ikke er stor nok til å stoppe vognen uten ryggsekk, og du har ingen tid til å ta på en ryggsekk). Er det moralsk akseptabelt å redde de fem jernbanearbeiderne ved å dytte den ene arbeideren i hans død?

I tråd med resultatene rapportert nedenfor vil de fleste av oss umiddelbart tenke at det føles feil å dytte en person i døden. Fra et utilitaristisk perspektiv medfører derimot denne handlingen det totalt sett beste resultatet, da fem liv spares. Den mest utbredte forklaringen på hvorfor dette likevel er moralsk galt er at det bryter med individets rettigheter. Kants formulering om humanitet i det kategoriske imperativ gjør seg gjeldende: *Handle mot menneskeheten, deg selv eller noen annen, alltid som et mål og aldri som et middel* (Cureton & Johnson, 2019). I en annen versjon («spak») kan man redde de fem arbeiderne ved å dytte en spak slik at vognen bytter til et nytt spor, men som følge tar livet av en arbeider som står på sidesporet. I «gangbru» føles det som et brudd på individets rettigheter og derfor feil å ofre ett liv i bytte mot fem. I «spak» er det mulig å forsvare det samme, til tross for at det ikke er ideelt. Dette er Kantiansk filosofi satt opp mot Mills filosofi, deontologisme versus utilitarisme (Greene, 2013). Greene (2013, p. 116) påpeker at dilemmaene adresserer noe som er relevant i nær sagt alle viktige moralske spørsmål: *Når, og hvorfor, veier individets rettigheter tyngre enn det felles beste?*

Hypotesen hans for hvordan noen mennesker vil foreta en utilitaristisk beslutning var todelt: et slikt valg følger fra anvendelsen av en eksplisitt utilitaristisk regel («gjør hva enn som produserer mest godt»); og – i situasjoner som er emosjonelt aktiverende kreves anvendelse av kognitiv kontroll – å si ja i «gangbru» kan sammenlignes med å lese ordet «rød» dersom det er skrevet med grønne bokstaver på Stroop-testen for nevropsykologisk undersøkelse av eksekutive funksjoner. Med hensyn til kamera-metaforen kan dette forstås som en overstyring av den emosjonelle aktiveringen representert ved den automatiske funksjonen. Dette gjennom påkobling av den i større grad følelseskalde, kognitive, manuelle funksjonen. Underbygget av Antonio Damasio's observasjoner av pasienter med skade i nedre del av medial prefrontal korteks (MPC), ventromedial prefrontal korteks (VMPFC) – pasientene presterer godt på kognitive tester, men foretar valg i hverdagslivet som gir store negative konsekvenser grunnet problemer med å integrere emosjonell forståelse ved beslutninger – administrerte Greene disse to og andre lignende dilemmaer mens deltagerens hjerneaktivitet ble undersøkt ved bruk av fMRI. Slik forventet aktiverte de «personlige» dilemmaene («gangbru» og lignende) MPC inkludert VMPFC, hjerneområdene assosiert med emosjonell prosessering som var skadet hos Damasio 's pasienter. I tillegg fant han økt aktivitet i Amygdala da deltakerne vurderte «personlige» eksempler, kontra «upersonlige» eksempler. Motsatt, de «upersonlige» dilemmaene («spak» og lignende) der deltakerne i signifikant større grad responderte at én kunne ofres for fem, ga økt aktivitet i dorsolateral prefrontal korteks (DLPFC), et område i hjernen assosiert med kognitiv kontroll. De fant og økt aktivering i en nærliggende region av DLPFC i de tilfelle deltakere ga utilitaristiske svar på «personlige» dilemmaer, dytt én for å redde fem. Samsvarende med dette vil pasienter med skade i VMPFC med større sannsynlighet godkjenne at vognen treffer familiemedlemmer med hensikt å redde et høyere antall fremmede (Greene, 2013).

Pasienter med frontallappsdemens (FTD), en degenerativ sykdom som blant annet rammer VMPFC og kjennetegnes av mangel på empati og dempet emosjonell respons, Alzheimer og friske kontroller ble presentert for «gangbru» og «spak». Alle tre gruppene viste lik respons på «spak» der minst 80 % av respondentene vurderte det som akseptabelt å endre vognens kurs for å redde fem på bekostning av én. Omtrent 20 % av pasientene med Alzheimer og friske kontroller godkjente å dytte mannen i døden for å redde fem i «gangbru», sammenlignet med 60 % i gruppen som hadde FTD, altså tre ganger høyere (Mendez, Anderson, & Shapira, 2005). Det er og vist at psykopater og mennesker diagnostisert med alexitymi, en psykiatrisk lidelse som beskriver manglende oppmerksomhet/forståelse for ens egne emosjonelle tilstander, i større grad foretar utilitaristiske beslutninger. Motsatt, deltakere som blir mer fysiologisk aktivert som respons på stress, samt de som rapporterer at de følger magefølelsen, gjør færre utilitaristiske valg (Greene, 2013).

Var det en kausal sammenheng mellom respons på ulike varianter av dilemmaene og aktivering i ulike hjerneområder assosiert med forskjellige sinnstilstander? Greene og medarbeidere ga deltakere kognitive anstrengelser ved respons til dilemmaene nevnt over og lignende dilemmaer, en form for simulering av effekten en lesjon i VMPFC har på emosjonell aktivering. I betingelsen med kognitiv anstrengelse passerte en serie nummer på skjermen under dilemmateksten og i betenkingsperioden før deltakerne skulle gjøre et valg. Økt kognitiv belastning ga lengre reaksjonstid for utilitaristiske, men ikke for ikke-utilitaristiske beslutninger. Funnet støtter hypotesen om at utilitaristiske beslutninger drives av kontrollerte kognitive prosesser, mens ikke-utilitaristiske beslutninger drives av automatiske prosesser. Kognitiv anstrengelse reduserte imidlertid ikke antallet utilitaristiske beslutninger (Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008). En sterkere påvirkning på deltakernes beslutninger er funnet med tidspress – mer tidspress førte til færre utilitaristiske valg – og motsatt. Samlet er det en rekke ulike typer studier (fMRI, nevrologiske pasienter med

emosjonelle avvik, fysiologiske mål på emosjonell aktivering, induksjon av emosjoner, manipuleringer som vanskeliggjør manuell-modus tenkning, farmakologiske intervensjoner) som viser at mennesker sier nei til å dytte arbeideren av gangbrua (og lignende «personlige» utilitaristiske handlinger som volder skade) som følge av emosjonelle responser i VMPFC og Amygdala (Greene, 2013).

Greene stiller følgende psykologiske spørsmål: Hvorfor finner 87 % av deltakerne det akseptabelt å bytte ett liv mot fem i «spak», mot 31 % av deltakerne i «gangbru»? I begge eksempler engasjeres vi i utilitaristisk, kost-nytte tenkning. Det er kun i «gangbru» at den emosjonelle responsen vanligvis trumfer vår utilitaristiske (manuelle) tenkning. For å undersøke dette har Greene og kolleger gjort eksperimenter med ulike versjoner av de to nevnte dilemmaene. To faktorer ble studert; personlig nærhet – avstanden mellom protagonisten og arbeideren som ofres når fem personer reddes – og personlig kraft – hvorvidt det påføres kraft for å forårsake arbeiderens død.

Felles for alle dilemmaene er at en vogn er i bevegelse på et spor i retning fem personer og at protagonisten heter Per. I «gangbru-avstand» står Per ved en spak et stykke unna sporet. Over sporet står en arbeider med ryggsekk på en gangbru. Dersom Per vender spaken vil arbeideren falle gjennom en lem i gangbrua og dø som følge av at han blokkerer vogna. Fem personer overlever. I «gangbru-spak» står Per ved en spak på en gangbru over sporet. Hvis han vender spaken vil resultatet bli det samme som i «gangbru-avstand», forskjellen mellom de to er graden av personlig nærhet mellom han og arbeideren. «Gangbru-kjepp» likner «gangbru», men her dytter Per arbeideren over ende og ned på sporet med en kjepp og ikke med hendene. Til forskjell fra «gangbru-avstand» er Per nær arbeideren, i tillegg til å bruke personlig kraft. For det første antyder resultatene at grad av personlig nærhet sannsynligvis ikke er den utslagsgivende faktoren da 59 % av deltagerne samtykker til å ofre én for fem i «gangbru-spak» (ikke signifikant forskjellig fra «gangbru-avstand» 63 %). Personlig kraft synes å være mer

avgjørende for hvorvidt det utløses en automatisk emosjonell respons: I «gangbru-kjepp» svarer 33 % av deltakerne at det ikke er akseptabelt å ofre én for å redde fem, ikke signifikant forskjellig fra «gangbru» (31 %). Dette er interessant da resultatet (én dør, fem overlever) er det samme uavhengig av om en berører personen som dør eller om en gjør et valg som medfører denne personens død uten direkte fysisk kontakt. Personlig kraft synes altså å ha psykologisk betydning selv om det ikke er moralsk relevant (Greene, 2013).

Videre undersøkte de betydningen av om skaden er påført den ene arbeideren som et middel for å oppnå et mål (redde fem) eller som en bivirkning, kjent som doktrinen av dobbel effekt beskrevet av Thomas Aquinas. Her følger tre nye dilemma. I «hindring-kollisjon» står Per på en smal, høy gangbru over et sidespor. Vogna er på et hovedspor i bevegelse med retning mot fem personer. Mellom Per og spaken på enden av gangbrua står det en arbeider. For å redde fem mennesker må Per springe mot spaken så fort som mulig og vende denne i visshet om at han kolliderer med arbeideren som da faller av gangbruen og dør. Slik som i «gangbru» og «gangbru-kjepp» er den skadelige handlingen personlig (nærhet og kraft). I dette tilfellet dør imidlertid arbeideren som en bivirkning og ikke som et middel. «Sløyfe» er likt «spak» bortsett fra at avsporingen svinger tilbake igjen på hovedsporet. Altså fungerer den ene personen stående på sidesporet som et middel for å stoppe vognen fra å rulle tilbake på hovedsporet og treffe de fem. Til slutt «kollisjonsalarm»: Her beveger to vogner seg på to parallelle spor. Vogn én treffer fem personer dersom ingenting skjer. Dersom Per vender en spak skifter vogn to spor og treffer en person hvorpå en alarm utløses. Når alarmen utløses kuttes elektrisiteten til hele systemet og stopper som følge vognen som er på vei mot de fem. Igjen vil den ene arbeideren dø som et middel for å redde fem.

Tabell 1*Oversikt over tilstedeværelse av fire faktorer i de ulike vogn-dilemmaene*

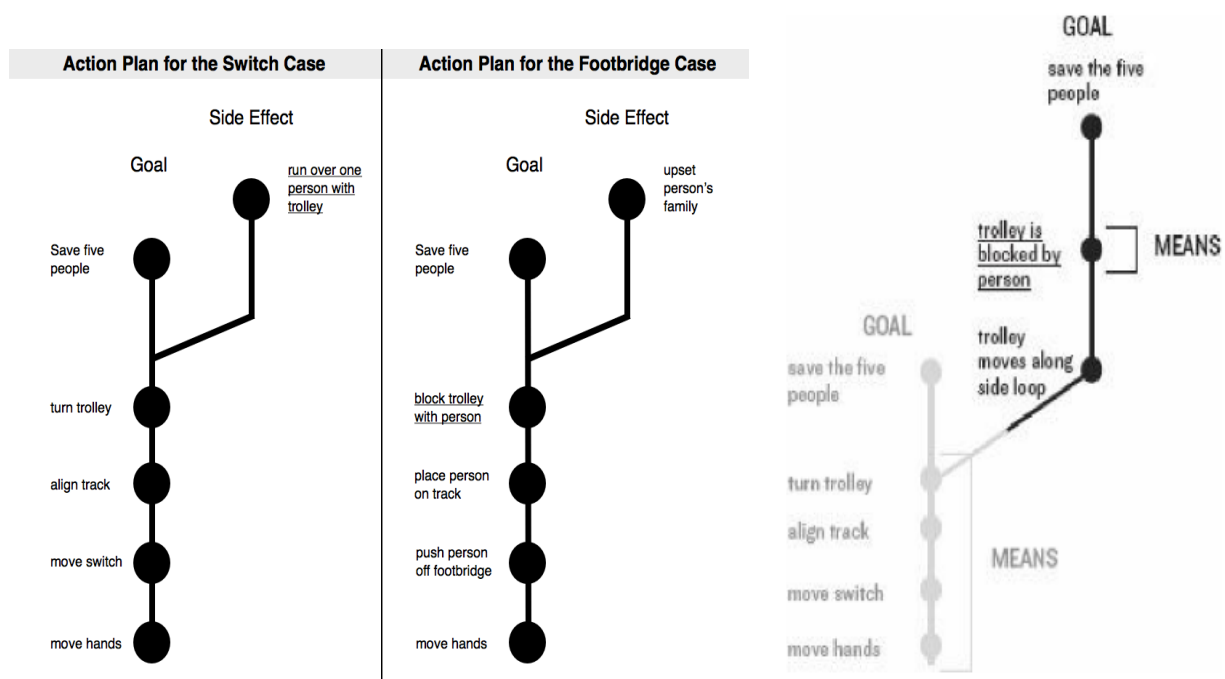
	«Gangbru» (31%)	«G-K» (33%)	«G-S» (59%)	«G-A» (63%)	«H-K» (81%)	«sløyfe» (81%)	«KA» (86%)	«Spake» (87%)
PK	ja	ja	nei	nei	ja	nei	nei	nei
PN	ja	ja	ja	nei	ja	ja	nei	ja
Middel	ja	ja	ja	ja	nei	ja	ja	nei
Bivirkning	nei	nei	nei	nei	ja	nei	nei	ja

Note. % = andel deltakere som besluttet å ofre én for fem. PK = Personlig kraft, PN = Personlig nærhet. «G-K» = Gangbru-kjepp, «G-S» = Gangbru-spake, «G-A» = Gangbru-avstand, «H-K» = Hindring-kollisjon, «K-A» = kollisjonsalarm.

I «gangbru» svarte 31 % av deltakerne at de var villige til å ofre den ene for fem. I «hindring-kollisjon» svarte 81 % at de godkjente at Per forårsaket arbeiderens død som en forventet bivirkning av å redde fem. Dette er ikke signifikant forskjellig fra det originale «spake» (87 %). Svarene antyder at våre automatiske innstillinger er sensitive til middel/bivirkning, et skille som i filosofien er anerkjent som moralsk relevant. Vi sier hovedsakelig nei dersom arbeideren brukes som et middel for å redde fem («gangbru», «gangbru-kjepp»), og ja dersom arbeideren er skadelidende som en bivirkning («spake», «hindring-kollisjon»). Et avvik fra mønsteret er at de fleste finner det akseptabelt å bruke arbeideren som et middel for å redde fem i «gangbru-avstand», «sløyfe», og i «kollisjonsalarm». Resultatene viser en varierende virkning av to faktorer, personlig kraft og middel/bivirkning. Personlig kraft synes å være av betydning i «gangbru-kjepp», sammenlignet med «gangbru-spake». Det synes derimot ikke å gjelde ved «hindring-kollisjon» sammenlignet med «spake». Middel/bivirkning synes å gjelde ved sammenligning mellom «hindring-kollisjon» (bivirkning) og «gangbru» (middel), men ikke i «spake» (bivirkning) kontra «sløyfe» (middel) og «kollisjonsalarm» (middel). Skade forårsaket av personlig kraft, men som en bivirkning, virker ikke så ille («hindring-kollisjon»). Skade forårsaket som et middel, men uten bruk av personlig kraft fremstår heller ikke ille («sløyfe», «kollisjonsalarm»). Derimot er

interaksjonen mellom personlig kraft, og bruk av arbeideren som et middel, utslagsgivende. Da virker det galt for de fleste å ofre én for fem («gangbru, «gangbru-kjepp»). Våre automatiske innstillinger aktivert i moralske situasjoner synes altså å være påvirkbar både av den vidt anerkjente distinksjonen mellom middel/bivirkning og den moralsk irrelevante faktoren personlig kraft. Greene hevder dessuten at forskjellen av moralsk betydning mellom middel/bivirkning ikke er begrunnet. Våre intuisjoner bidrar til at vi opplever det som en meningsfull, moralsk forskjell. Doktrinen av dobbel effekt er derfor en ufullstendig oppsummering av våre intuitive bedømmelser.

På bakgrunn av resultatene gjennomgått over foreslår Greene (2013) hypotesen om Modulær Myopi (automatiske innstillinger). Han beskriver modulen som et kognitivt delsystem som kontrollerer planer for atferd og utløser en alarm dersom vi vurderer å skade andre mennesker. Myopisk (nærsynt) fordi systemet bare kan se skade som er planlagt som et middel for å oppnå et mål og ikke skade som bivirkninger av disse planlagte handlingsrekkene. Modulær fordi systemets operasjoner foregår utenfor de delene av hjernen som muliggjør bevisst tenkning. Den er emosjonelt, men ikke kognitivt blind og fungerer slik at den forhindrer oss fra å være voldelig da vi utgangspunktet har evnen til planlagt vold, slik vist i vogn-dilemmaene. Hypotesen gir opphav til tre prediksjoner: (1) den aktiveres av ekte vold, (2) den aktiveres av typiske triggere for ekte vold, (3) dens funksjon er å forhindre en selv fra å utøve vold. Hvorfor er den blind for skade som skjer som en forutsett bivirkning? Her implementerer Greene en teori fra Mikhail (Levine, Leslie, & Mikhail, 2016), nemlig at handlingsrekkene i de ulike dilemmaene representeres mentalt som en asymmetrisk forgreinet aksjonsplan. Dersom forskjellen mellom middel/bivirkning betyr noe for beslutningene vi tar i de ulike dilemmaene må det være ved å påvirke systemet som utløser de emosjonelle responsene slik at vi reagerer kraftigere ved dilemma der skaden er forårsaket som et middel for å nå et mål. Hvis så, hvorfor skjer det ikke i «sløyfe»? Teorien om aksjonsplaner og dobbel-prosessering kan kombineres.



Figur 1. Representasjon av kausale handlingsrekker i «spak», «gangbru» og «sløyfe».

Kanskje er det slik at «sløyfe» ikke utløser en emosjonell respons fordi den er et komplisert tilfelle av et dilemma der arbeideren ofres som et middel for å redde de fem. Den forutsetter at vi forholder oss til flere samtidige kausale handlingsrekker, i motsetning til én kausal handlingsrekke i «gangbru» og «spake». Kan det være at vårt automatiske system (den myopiske modulen) inspiserer aksjonsplaner og responderer på muligheten for skade med en emosjonell alarm? Den er imidlertid blind for skadelige bivirkninger fordi den bare registrerer hendelser som er kausalt nødvendige for oppnåelsen av målet. Med andre ord følger den opp hendelsene som utgjør aksjonsplanens primære kausale kjede. Derfor presenteres «sløyfe» slik: beveg hånd → vend spake → skift spor → sving vogn → redd fem mennesker (se figur 1). Det som skjer på den sekundære handlingsrekken (uthevet i figuren) – vognen går over på sidesporet → vognen blokkeres av en arbeider (middel) → fem mennesker reddes – registreres ikke av den myopiske modulen. Den maksimerende manuelle modus (dobbel-prosess teori) veier tyngst dersom ingen emosjonell alarm utløses. Dette er årsaken til at vi gir utilitaristiske svar ved respons på «spake» og «sløyfe», men ikke «gangbru». Slik integreres den myopiske

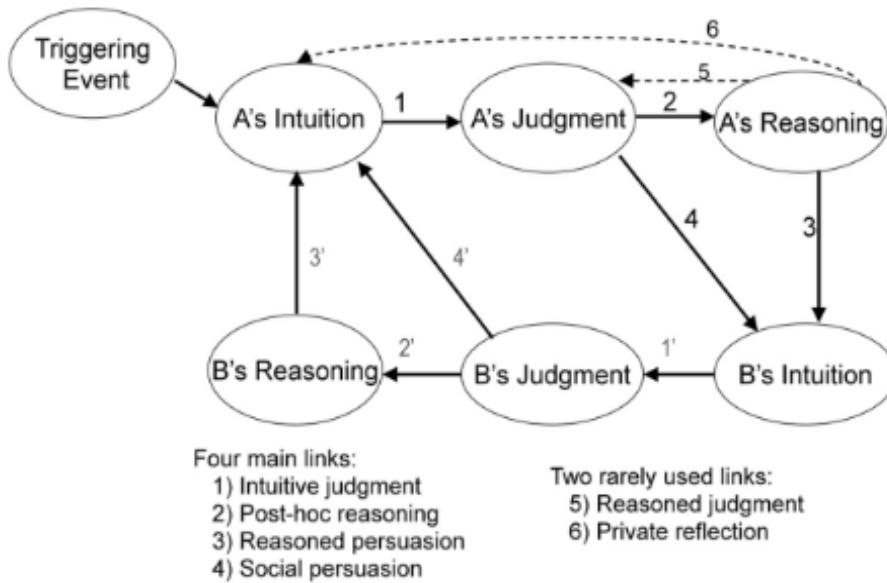
modulen, Mikhails teori om aksjonsplaner og teorien om dobbel-prosessering for moralsk beslutningstaking, til å forklare hvorfor vi bryr oss mindre om skade som oppstår som bivirkninger (Greene, 2013).

Churchland (2011) forholder seg skeptisk til generaliserbarheten av disse funnene til moralske valg i det virkelige liv da scenarioene, av metodiske hensyn for å begrense konfunderende variabler, ikke oppgir informasjon om detaljer ved individene involvert, deres historie og konteksten. Detalj mangelen medfører at hjernens behov for moralsk relevant fakta tiltrengt for å løse problemer undergraves. Hun foreslår følgende som en mer sannsynlig forklaring: *alle bruker situasjons-basert resonnering, men gitt våre individuelle historier og temperament, bruker vi ulike hendelser for å veilede beslutninger i valget vi står ovenfor. Én student kan respondere negativt til «ofre en for mange» etter å ha lært om unødvendige offer i Russland under kommunismen. En annen student kan respondere positivt basert på en assosiasjon til en film med en torpedert ubåt der alle ville druknet med mindre døren til maskinrommet var avstengt og som følge druknet kun én person, maskiningeniøren (...)* Dersom, slik jeg mistenker, de moralske dilemmaene vi konfronteres med i det virkelige liv typisk løses ved begrenset behovstilfredstillelse, så er situasjons-baserte analogier, emosjoner, hukommelse og oppfinnsomhet nesten alltid involvert (Churchland, 2011, p. 183).

I visshet om at resultatene i klassiske vogn-dilemma ikke uten videre kan generaliseres til moralske valg i det virkelige liv, rekrutterte Greene og medarbeidere (Ranshoff et al. sitert i Greene, 2013) leger og profesjonelle innenfor folkehelse, yrkesgrupper som tar beslutninger i spørsmål om liv og død. De presenterte begge grupper for «gangbru» og lignende dilemma, i tillegg til mer realistiske dilemma: rasjoner av medisiner distribuert for å redde flere liv, med resultatet at enkelte ikke får medisinen; hvorvidt en skal prioritere billig medisin til mange mennesker, eller dyr medisin til noen få som allerede er syke; eller et dilemma der én pasient med infeksjon isoleres for å beskytte andre pasienter. I begge grupper fant de en sterk

korrelasjon mellom individuelle deltakeres svar på klassiske hypotetiske vogn-dilemma og mer realistiske helsedilemma. Med andre ord vil en lege/helseprofesjonell som velger å dytte en person i døden for å redde fem, med større grad av sannsynlighet støtte rasjonering av medisiner osv. De fant i tillegg at helseprofesjonelle ga mer utilitaristiske svar på begge typer dilemma, sammenlignet med legene. De konkluderte med at dette enten kan skyldes at mennesker som er mer utilitaristisk orientert søker seg til arbeid i folkehelse, eller at mennesker som jobber innenfor folkehelse inntar et mer utilitaristisk standpunkt som følge av utdanning og jobb. Eventuelt begge deler. De hevder at funnene viser at en avveining mellom individets rettigheter og det felles beste i utilitaristisk forstand er fenomener som forekommer i den virkelige verden og som løses av de to moralpsykologiske prosessene, analog og manuell funksjon.

Liknende Greene's teori om dobbel-prosessering argumenterer Haidt (2012) for sitt første prinsipp om moralpsykologi – intuisjoner kommer først, strategisk resonnering deretter – med en todelt modell. Den sosiale intuisjonsmodellen («social intuitionist model») for moralsk bedømming beskriver at moralske beslutninger styres av raske, intuitive, emosjonelle prosesser. Først etter at vi har gjort et valg mener Haidt at vi rasjonelt overveier på hvilket grunnlag vi tok valget med hensikt å overtale andre mennesker. Han bruker metaforen om en rytter på en elefant, der rytteren illustrerer kontrollerte prosesser og elefanten gjenspeiler automatiske prosesser. Dette er en noe upresis metafor da Haidt argumenterer for at automatiske prosesser (elefanten) hovedsakelig styrer moralske beslutninger. I virkeligheten får rytteren nesten alltid elefanten til å gjøre det rytteren vil. Skjematisk ser modellen slik ut:



Figur 2. Haidt's sosiale intuisjonsmodell for moralsk bedømming.

B representerer vår sosiale omgangskrets som kan utfordre våre synspunkter og produsere motargumenter slik at vi kan endre våre intuisjoner. Han utelukker ikke at vi selv kan endre våre intuisjoner gjennom link 6 privat refleksjon, men konstaterer at dette forekommer langt sjeldnere enn ved sosial påvirkning. Begrunnelsen for hvorfor politiske og moralske diskusjoner ofte er frustrerende illustrerer han i eksempelet navngitt den emosjonelle hunden og dens rasjonelle hale («the emotional dog and it's rational tail»). *Vårt moralske liv er forpestet av to illusjoner. Den første illusjonen kan kalles «wag-the-dog» illusjonen: Vi er av den oppfatning at vår egen moralske dømmekraft (hunden) er drevet av vår egen moralske resonnering (hundens hale). Den andre illusjonen kan kalles «wag-the-other-dogs-tail» illusjonen: I en moralsk diskusjon forventer vi at vårt vellykkede motbevis av vår motstanders argument skal forandre vår motstanders oppfatning om saken som diskuteres. En slik overbevisning er analog til å tro at å tvinge en hunds hale til å bevege seg vil gjøre hunden glad (Haidt, 2001). Haidt bemerker at mennesker ofte blir moralsk målløse («morally dumbfounded») dersom de bes begrunne sine responser til spesifikke moralske, og pseudomoralske, dilemma.*

Harris (2010) kritiserer Haidt's forklaring på moralske beslutninger og generelle skepsis til om det er mulig å realistisk definere galt og rett, ved å vise til Monty Hall-problemet som et tilfelle der mange mennesker opplever å bli logisk målløse: Forestill deg at du er med i et TV-show der du må velge mellom tre lukkede dører. Bak en av dørene er en ny bil, de to andre skjuler geiter. La oss si at du har valgt dør #1. Deretter åpner programlederen dør #2 som skjuler en geit. Du får nå en sjanse til å bytte fra dør #1 til den gjenstående dør #3. Bør du bytte? Det rette svaret er ja. Det oppleves for de fleste ulogisk da det bryter med antagelsen om at det må være 50/50 sjanse at bilen befinner seg bak en av de to stadig uåpnede dørene (#1 og #3). Hvis du holder deg til ditt første valg (#1) er sjansen for å vinne bilen 1/3, dersom du bytter til dør #3 er den 2/3. Folk klarer ikke å gi slipp på sin intuitive vurdering selv om de konseptuelt forstår at det rette er å bytte dør. Denne feilen i vår evne til å resonnerer betyr ikke at det ikke finnes et riktig svar på problemet. På samme måte er ikke vår manglende evne til å begrunne våre moralske valg ensbetydende med at det ikke finnes mer eller mindre rette/gale handlinger. *Haidt har rett når han påpeker at hjernens emosjonelle nettverk ofte styrer våre moralske intuisjoner, og måten følelser innvirker på beslutninger er helt klart verdt å forske på. Men det følger ikke at det ikke er noen rette eller gale svar på spørsmål om moral. Akkurat slik mennesker ofte kan være mindre rasjonelle når de hevder å være rasjonelle, kan de være mindre moralske når de hevder å være moralske* (Harris, 2010, p. 119). Argumentet viser forskjellen mellom Haidt's deskriptive moralteori og Harris normative moralteori, ytterligere diskutert i siste del av denne oppgaven.

På linje med Harris kritikk av Haidt, er det ikke bred enighet om betydningen av raske, emosjonelle prosesser som i større grad styrende for vår moralske beslutningstaking enn rasjonalitet. Det er utbredt innenfor psykologiske teorier om moral å søke basale forklaringer på komplekse kognitive fenomener. Ved studier på moralsk beslutningstaking tillegges aversjon spesielt stor vekt for hvorvidt noe er rett eller galt, eksempelvis slik beskrevet i

Greene's hypotese om Modulær Myopi som en evolusjonær alarmreaksjon nevralt representert i hjerneområder assosiert med emosjonell aktivering, Blairs redegjørelse for skillet mellom moral og konvensjoner der moral innbefatter ubehaget ved å se andre lide, eller Cushman (2013) og Crocketts (2013) forklaring på ikke-utilitaristiske beslutninger som drevet av læringsteori. Slike forklaringer er tiltalende da de bygger på prosesser som udiskutabelt er en del av menneskets psykologi. De fleste finner det aversivt å se mennesker lide, og det er godt dokumentert at mennesker gjennom habituell læring kan etablere assosiasjonen mellom enkelte typer handlinger og aversive utfall (Gray & Graham, 2018).

Eksempelvis er det funnet aversjon mot å utføre en pseudovoldelig handling og vurdering av moralsk fordømmelse, ved å be deltakerne gjøre handlinger som var harmløse, men aversive (knivstikke en annen deltaker med en falsk scenekniv). Moralsk fordømmelse var mer presist predikert av deltakernes aversjon mot å utøve simulerte voldelige handlinger, enn affektive responser ved å simulere den forulempedes smerte (Miller, Hannikainen, & Cushman, 2014). Videre er det funnet fysiologisk aversjon målt ved perifer vasokonstriksjon ved tanken på å utføre pseudovoldelige handlinger slik som å dunke hodet til en babydukke mot et bord eller å slå en realistisk kunstig fot med en hammer (F. Cushman, Gray, Gaffey, & Mendes, 2012). Dette antyder at voldelige handlinger som vanligvis medfører skade kan tillegges en negativ verdi tilstrekkelig til å utløse en aversiv respons selv uten at vanligvis forekommende skadelige konsekvenser inntreffer.

Shaun Nichols (Gray & Graham, 2018) illustrerer at selv om noe er dårlig (aversivt) er det ikke nødvendigvis normativt galt. En opplever tømning av søppel som aversivt, men ikke galt, motsatt kan en oppleve spising av kjøtt som galt men ikke aversivt. Adferd styres ofte av aversjoner og attraksjoner; unngå søppel, spis kjøtt. Følgelig hevder Nichols at en behøver mer enn aversjon for å forklare hvorfor noe er galt. Ved bruk av Stalins kjente sitat om at ett dødsfall er en tragedie, mens en million er en statistikk, viser han at vi kan oppleve det mer aversivt å

forestille oss en persons død enn en million, men vi bedømmer en million dødsfall som mer galt. Fenomenet beskrevet av Stalins sitat kalles den identifiserbare offereffekten («the identifiable victim effect»). Vi reagerer følelsesmessig sterkere på lidelsen til et bestemt offer enn statistiske ofre. Nichols mener derfor bedømmelsen av i hvor stor grad handlingen er moralsk gal kan synes å være begrunnet i den abstrakte kategorien mord, og ikke i aversjonen. Forklaringsmodeller for moralsk beslutningstaking begrunnet i aversjon har da vanskelig for å fange opp nyansene ved ulike moralske spørsmål, i tillegg til at moralske valg oftest handler om begrensninger for hva en kan gjøre. Følgelig mener han at moralske beslutninger er avhengig av strukturerte regler som kan fange opp variasjonen i uenigheter på moralske spørsmål.

Moralfilosofi begrunnet i moralpsykologi

Til nå har jeg gjennomgått normative moralfilosofiske teorier (dydsetikk, konsekvensetikk og pliktetikk) som på ulike måter argumenterer for hvordan å handle moralsk, samt deskriptiv teori, basert på empiriske studier, om hvordan moral kan ha blitt utviklet, hvilke hjernemekanismer som fremmer tilknytning og omsorg for andre (hormoner og nevralt aktivitet i ulike hjerneområder), samt hvordan vi treffer beslutninger i konstruerte moralske situasjoner. Med en utvidet forståelse av moral basert på moralpsykologisk empiri gjenstår en normativ diskusjon rundt den gjenstående problemstillingen: Kan moral reduseres til én universell, allmenngyldig sannhet, eksempelvis grad av lykke, eller kreves det en pluralistisk tilnærming der moral forstås som bestående av flere uavhengige fundamentet som sammen forklarer ulike menneskers moralske valg? En slik diskusjon vil nødvendigvis inneholde spørsmålet om hvorvidt vi kan bruke vår stadig økende deskriptive kunnskap om moralpsykologiens utvikling, nevrobiologiske korrelater og moralpsykologiske mekanismer til å besvare normative moralske spørsmål.

Finnes det flere uavhengige moralske fundamententer?

Haidt (2012) presiserer at psykologiens bidrag til forståelsen av moral er deskriptiv da hensikten er å undersøke hvordan vår moralske tenkning faktisk fungerer, ikke hvordan den bør fungere. Dette kan bare gjøres ved empatisk observasjon, i motsetning til filosofiens prioritering av resonnering og logikk. Et av hans moralpsykologiske prinsipper lyder; *det er mer ved moral enn skade og rettferdighet*. Om moralsk monisme uttrykker han: (...) *vi tror at moralsk monisme – forsøket på å basere moral på et enkelt prinsipp (vanligvis en variant av maksimering av velferd) – leder til samfunn som ikke tilfredsstillende sine menneskers behov og som risikerer å bli inhumane fordi de ignorerer flere moralske prinsipper* (Haidt, 2012, p. 113). Med andre ord hevder han at spørsmål om rett og galt ikke kan favnes av én enkelt moralsk dimensjon, én moralsk sannhet.

Ved å lete etter sammenhenger mellom etterlevelse av ulike dyder rundt om i verden og veletablerte evolusjonære teorier, har Haidt utarbeidet en gjennomgripende teori om menneskets moralske fundament («Moral Foundations Theory»). Den beskriver moral som uttrykk for ulike adaptasjoner på evolusjonært sett langvarige trusler og muligheter i det sosiale liv. Fundamentene er funnet ved å identifisere adaptive utfordringer ved sosialt liv mye omtalt av evolusjonspsykologer og deretter knytte disse til dyder som finnes i en eller annen form i mange kulturer. Han trekker frem fem adaptive utfordringer som korresponderer med fem bipolare fundamententer: omsorg for sårbare barn – omsorg/skade («care/harm»), dannelsen av parforhold med ikke-slekt for å høste fruktene av resiprositet – rettferdighet/bedrageri («fairness/cheating»), koalisjoner dannet for å konkurrere mot andre koalisjoner – lojalitet/forræderi («loyalty/betrayal»), mekling i hierarkier av status – autoritet/underdanighet («authority/subversion»), og forsøk på å holde en selv og ens slekt fri for parasitter og patogener – sakralitet/nedbrytning («sanctity/degradation»). Da våre forfedre møtte disse adaptive utfordringene i flere hundre tusen år ville naturlig seleksjon favorisere de individene med

hjerner som inneholdt medfødte kognitive moduler som løste utfordringene på rett måte – raskt og intuitivt – sammenlignet med individer som var avhengig av sin generelle intelligens (rytteren i rytter/elefant-analogien) for å løse tilbakevendende problemer.

Hvert fundament har et tilhørende sett av moralske emosjoner. Eksempelvis er omsorg/skade assosiert med sympati. Verdien sakralitet er forbundet med en følelse av ærefrykt for det som er hellig, og avsky for det som er fordervet eller nedbrutt (det motsatte av sakralt). Den senere tillagte frihet/undertrykkelse («liberty/oppression») er angivelig basert på folks følsomhet for, og forakt mot, å bli forsøkt dominert. Eksempelvis uttrykkes den hos den amerikanske politiske venstresiden gjennom viktigheten av å opprettholde et antiautoritært, egalitært samfunn, på høyresiden som sinne mot staten og sosialiseringspolitik. Haidt sammenligner våre moralske fundamenters reseptorer med tungens smaksreseptorer. Et lidende barn aktiverer det moralske sinnets reseptor for omsorg/skade, som medfører sympati og medfølelse. Ulike kulturelle grupper har ulike moralske smaksløker, sensitive for ulike moralske «smaker». Ifølge Haidt har den amerikanske liberale venstresiden lite utviklede moralske smaksløker kun sensitive for omsorg/skade, rettferdighet/bedrageri og frihet/undertrykkelse, i motsetning til den konservative, i stor grad religiøse, høyresidens moral fundert på alle seks.

Teoretisk inndeling av moral i ulike fundamenters/faktorer ut fra dyder er tidligere gjort med ulik vektlegging. Sokrates argumenterte for følgende fem: visdom, mot, moderasjon, rettferdighet og gudfryktighet, men utelot senere sistnevnte. Buddhismen vektlegger de fire noble sannheter: «loving-kindness», «compassion», «appreciative joy» og «equanimity». Slik nevnt anbefalte Aristoteles moderasjon og balanse mellom ekstremer. Haidt (2012) går et skritt videre da han argumenterer for en evolusjonær forklaring på hvorfor hans moralske intuisjoner er fundamentale og medfødte. Han presiserer at dette ikke er et ikke-verifiserbart etiologisk narrativ («just-so story»), men viser ikke til evidens fra molekylærbiologi, nevrovitenskap eller evolusjonsbiologi. Hans liste av kryss-kulturelle dyder kan forklares på ulike måter

(Churchland, 2011). Noe av begrunnelsen hans for hvorfor listen utgjøres av akkurat disse seks fundamentale faktorene er hvorvidt de er enkle å lære. Slik avgrenser han hvilke ferdigheter hjernen er naturlig forberedt for, fra ferdigheter hjernen er mindre forberedt for. Churchland (2011, p. 116) innvender: *Men hvordan forsvarer du, uten å ty til ad hoc løsninger, hvorfor noen ferdigheter som er enkle å lære er medfødte mens andre utelates, slik som å sykle, knyte en båtmannsknop, ikle seg sko, eller fiske etter ørret med en mark på en krok? Omvendt, å lære seg selvkontroll, ganske bestemt noe hjernen sannsynligvis er «forberedt på», er ofte ikke enkelt. Grad av enkelhet for å lære en ferdighet er konsistent med at det er medfødt, men impliserer det ikke.*

Harris (2010) foreslår at de seks distinkte moralske fundamentene simpelthen kan være ulike fasetter av en mer generell bekymring for omsorg/skade som både konservative og liberale deler, men som kommer til uttrykk på en noe ulik måte. *Hva er problemet med å skjende et eksemplar av Koranen? Det ville ikke vært noe problem med mindre noen mennesker var av den oppfatning at Koranen er en guddommelig forfattet tekst. Slike mennesker tror ganske sikkert at en eller annen form for skade ville tilfalle dem eller deres gruppe som et resultat av slik skjending – hvis ikke i denne verden, så i den neste* (Harris, 2010, p. 119).

Greene (2013) påpeker at konservative amerikanere ikke generelt sett respekterer autoriteter (eksempelvis Barack Obama), kun de som anerkjennes av deres egen kulturelle gruppe. Majoriteten av republikanere har uttalt at en ikke bør ha tillit til en muslimsk-amerikaner i en autoritetsposisjon. Følgelig er og deres bekymring for sakralitet dypt forankret i gruppetilhørighet (dersom du holder profeten Muhammed hellig, bør du ikke ha makt). Han påpeker at også deres lojalitet er begrenset til deres egen gruppe da konservative ikke ønsker at alle mennesker skal være lojale mot sitt eget land, eksemplifisert ved at Iranske protester mot regjeringen ville mottatt støtte fra konservativt hold. Samtidig motsier Haidt (2012, p. 272) sin egen uttalelse om moralsk monisme og avstand fra normative påstander når han senere i boken

skriver følgende: *Jeg vet ikke hva som er den beste normative etiske teorien for individer i deres private liv. Men når vi snakker om å lage lover og implementere offentlig politikk i vestlige demokratier bestående av noen grad av etnisk og moralsk mangfold, tenker jeg at det ikke finnes noe tiltalende alternativ til utilitarisme.* På tross av sin uttalte deskriptive tilnærming til moralske spørsmål gir Haidt (2012) en generell normativ rettesnor. For at vi mennesker skal komme bedre overens, bør vi alle være mindre selvrettferdige. Vi bør erkjenne at de fleste av oss er gode mennesker og at våre uenigheter har sitt opphav i at vi tilhører ulike kulturelle grupper med ulike moralske intuisjoner. Vi bør bli flinkere til å se våre egne moralske rasjonaliseringer, slik vi ser andres. På den måten kan vi forsøke å forstå hverandre, være mindre hyklerske og mer åpne for kompromiss.

Hva med en moralsk sannhet?

Spørsmålet om hvorvidt det finnes en moralsk sannhet – kan forstås som moralsk monisme – er historisk sett forsøkt besvart ut fra hovedsakelig tre ulike fremgangsmåter: religiøse tekster, matematisk logikk og naturvitenskapelig evidens. Haidt's (2012) redegjørelse kan sees som et forsøk på en vitenskapelig beskrivelse av moral uten å argumentere for en moralsk sannhet. Platon påpekte en logisk svikt ved forholdet mellom moralsk autoritet og en guddommelig vilje ved å spørre noe slikt som: Er noe galt fordi Gud fordømmer det, eller er det slik at Gud fordømmer noe fordi det er galt? Drap og vold er moralsk galt, noe de fleste mennesker og Gud er enige om. Men har Gud makt til å, dersom han ønsker, gjøre drap moralsk rett? Kan Gud, i kraft av å være den øverste moralske autoriteten, beslutte å gjøre en umoralsk handling, moralsk? Hvis en, slik som Platon, tenker at svaret på dette er nei, betyr det at noen viktige moralske regler er uavhengige av Guds vilje, og at Gud følgelig ikke er en ubestridt moralsk autoritet. Dersom svaret er ja, betyr det at det i praksis er tilfeldig hva som til en hver tid er å regne som rett og galt, altså utvannes betydningen av at noe er moralsk. Platons argument hentyder at menneskeheten har behov for sekulært funderte begrunnelser for hva som

er rett og galt (Greene, 2013). Dersom vi forutsetter at Guds vilje kan bestemme hva som er moralsk rett og galt oppstår et problem da vi må velge mellom ulike religioners stridende moralske bestemmelser som alle angivelig er basert på Guds vilje (Harris, 2004). Greene (2013) konkluderer med at religiøs tro i de fleste tilfeller lykkes med å sette gruppen foran individet, men at den snarere enn å tilby løsninger på, øker og polariserer klimaet mellom grupper, «oss» og «dem», på bakgrunn av motstridende verdier. Dette er i overensstemmelse med Haidt's (2012) tredje prinsipp om at moral binder og blinder. En gruppe bindes sammen av felles verdier og blindes samtidig for andre gruppers avvikende verdsett.

Den rasjonelle overbevisningen vurderer moral på linje med matematikk ved å hevde at det må finnes én eller flere grunnleggende aksiomer, matematiske påstander som er selvsagt sanne uten bevisbyrde, og at moralske påstander da må utgå fra slike aksiomer. Denne overbevisningen har søkt en logisk allmenngyldig moral som eksisterer uavhengig av menneskets dynamiske natur, sammenlignbart med matematikkens lover. Fornuft er tuftet på premisser. Dersom Stines premisser imidlertid ikke er selvsagt sanne, og Jon misliker konklusjonen som følger av Stines premisser, står Jon fri til å avvise Stines premisser og derav konklusjon. Denne dynamikken ser en til stadighet i norsk politikk, både i moralske og konvensjonelle debatter mellom politikere, der de underliggende antagelsene er mange, motstridende og gir mening for den enkelte debattant, i kontrast til matematikken, som er begrunnet i teoremer avledet fra aksiomer ubestridelig for enhver. Det er til nå i moralfilosofiens historie ikke funnet slike selvbeviselige aksiomer egnet til å løse utfordringer i den virkelige verden. Fornuft kan, på samme måte som religion men av andre årsaker, ikke alene balansere mellom ulike gruppers/individens inkongruente verdsett og rettigheter (Greene, 2013).

Hva med vitenskap? Her aktualiseres problemet med å gå fra *er* til *bør*, å slutte fra det som er naturlig til det som er rett, fra fakta til verdier, deskriptivitet til normativitet, kjent som

«Hume's giljotin», opprinnelig aktualisert av den skotske filosofen David Hume, men ofte tilskrevet George Edward Moore i det 20ende århundre (Harris, 2010). Det er moral, og refleksjon rundt moralske spørsmål, som er egnet til å fortelle oss hva vi *bør* gjøre, hvordan vi *bør* handle. Biologi kan kun fortelle oss noe om hva som *er* tilfellet. Altså har ikke biologi og naturvitenskapelig tenkning noe å tilføre mulige svar på moralske spørsmål (Churchland, 2011). Dersom biologi skulle ha betydning forutsetter det at vi i møte med moralske spørsmål, eller situasjoner som er av moralsk karakter (i seg selv vanskelig å definere), tar valg basert på deduksjon. Deduksjon er logisk stringente utledninger fra konkrete, eksplisitte premisser om det vi allerede vet. Eksempelvis syllogismen; (1) Alle dyr puster, (2) alle hunder er dyr, derfor (3) puster alle hunder. Eksempelen er logisk koherent fordi det åpner med to premisser om noe som *er*, noe som vi har allmenn kunnskap om, og konkluderer med et faktum som utledes fra begge er-premissene. Det er en faktabasert deduksjon uten moralsk relevans.

Problemet tydeliggjøres dersom vi bruker et moralsk ladet eksempel; (1) Kvinner og menn har ulike fysiske forutsetninger, (2) ulike fysiske forutsetninger mellom kjønn er naturlig og forekommer hos flere dyrearter, derfor (3) bør menn og kvinner behandles ulikt. I dette tilfellet fremkommer det tydelig at selv om enkelte påstander beviselig forholder seg slik, kan de gi grunnlag for å trekke en normativ konklusjon som av mange i dagens samfunn ansees som umoralsk. Altså er ikke definisjoner og «sannheter» om hva som er, relevant for hva som bør være. Er/bør-innsigelsen har i moderne vestlig akademisk filosofi i hovedsak blitt brukt som et uomtvistelig argument mot biologiens, psykologiens, og andre vitenskapelige disipliners forsøk på å besvare moralske spørsmål. Hume's opprinnelige hensikt med å belyse denne problemstillingen var imidlertid å blottlegge umoralske praksiser som ble begrunnet med slik deduksjon, samt å kritisere overbevisningen om at fornuft alene, uten innblanding fra følelser er tilstrekkelig for å forstå moral: «reason is and ought only to be the slave of the passions» (Hume, Nidditch, & Selby-Bigge, 1978). Han mente derimot ikke at kunnskap om menneskets

biologi, emosjoner, tanker, hukommelse, motivasjon og planer ikke kunne påvirke vår formening om menneskets moral på et generelt grunnlag (Churchland, 2011).

Historien har vist at en overbevisning om at evolusjonær utvikling representerer det moralsk rette (sosial-darwinisme) kan bære galt av sted, eksemplifisert ved fascisme og nazisme. Begrunnelsen for hva som er moralsk rett basert på evolusjonær funksjon kan forstås som et forsøk på å definere et moralsk aksiom: Det som er rett er det som i størst grad er i tråd med en evolusjonær utvikling av moral. Påstanden er derimot ikke et aksiom da den ikke er selvsagt sann (Greene, 2013). Evolusjonen har i seg selv ingen retning, verdi eller hensikt. Husk Greene's (2013, p. 22) definisjon: *Moral er et sett av psykologiske tilpasninger som tillater ellers egoistiske individer å høste fruktene av samarbeid.* Dersom moralens funksjon er å fremme samarbeid, hvorfor kan vi da ikke hevde at den moralske sannhet er den som best gjør nettopp det? Evolusjonen medfører ikke samarbeid med mindre resultatet er en videreføring av genene til individene i gruppen som samarbeider. Altså vil evolusjonen kunne videreføre en gruppes gener, som til tross for samarbeid internt, søker å utslette alle mennesker som ikke slutter seg til gruppen. Følgelig kan resultatet av den evolusjonære prosessen være umoralsk. Vi kan derfor ikke bruke vår forståelse for utviklingen av moralens naturlige funksjon til å finne en moralsk sannhet, vi kan ikke gå fra er til bør. Altså er, i følge Greene (2013), verken religion, matematisk begrunnet logikk eller vitenskapen i stand til å tilby en moralsk sannhet. *Betyr dette at det ikke finnes noe slikt som en moralsk sannhet? Jeg forholder meg agnostisk. En gang i tiden anså jeg dette som det store spørsmålet, men jeg har siden endret mening. Det som virkelig betyr noe er om vi har en direkte, pålitelig tilgang til den moralske sannhet – en tydelig sti gjennom moralske kontroverser – ikke hvorvidt en moralsk sannhet eksisterer* (Greene, 2013, p. 188).

Det er viktig å understreke at Greene (2013) utelukkende omtaler kunnskap om den evolusjonære bakgrunnen for moral i diskusjonen om vitenskapens bidrag til søken etter en

moralsk sannhet. Sam Harris (2010) opponerer mot det han beskriver som en brannmur mellom fakta og verdier da forskere innenfor psykologi og nevrovitenskap sjelden trekker konklusjoner om hvordan mennesker bør tenke eller handle. Han lister opp tre poenger, som han senere argumenterer for, med hensikt å vise at skillet mellom fakta og verdier er en illusjon: *(1) Hva som enn kan vites om maksimering av bevisste skapningers livskvalitet – som er det eneste vi med fornuft kan finne verdifullt – må på et punkt være relatert til fakta om hjernen og dens interaksjon med verden; (2) selve idéen om «objektiv» kunnskap (kunnskap ervervet gjennom ærlig observasjon og resonnering) har innebygde verdier, da enhver anstrengelse vi gjør for å diskutere fakta avhenger av prinsipper som vi først må verdsette (logisk konsistens, evidensbasert, stringens); (3) oppfatninger om fakta og oppfatninger om verdier synes å ha sitt utspring fra liknende prosesser i hjernen: det synes som om vi har et felles system for å bedømme sannhet og usannhet i begge domener* (Harris, 2010, pp. 23-24). Det siste punktet har han funnet indikasjoner for i sin egen nevrovitenskapelige forskning på tro/overbevisning («belief») ved bruk av hjerneavbildning. Da ulike bevissthetstilstander er fenomener underlagt naturlovene, hevder Harris at det må finnes rette og gale svar på moralske spørsmål som kun kan besvares ved bruk av vitenskapelig metode. Slik vil spørsmålet om hva som er viktig i livet besvares mer eller mindre rett av ulike mennesker og kulturer.

Da Harris ønsker å argumentere for at vitenskapen kan og bør brukes til å etablere en universell forståelse av hva som er moralsk sant og rett, gjør han et forsøk på å avvise umuligheten ved å slutte fra er til bør. Harris spesifiserer at en vitenskapelig redegjørelse for menneskelige verdier (forståelse av verdier som tilknyttet korrelasjonen mellom hendelser i verden og hendelser i den menneskelige hjerne) ikke er det samme som en evolusjonær redegjørelse da mesteparten av det som i dag utgjør menneskelig livskvalitet omfatter mer enn hva som kan forklares evolusjonært. Han eksemplifiserer det med at en mann, fra et evolusjonært perspektiv, kun vil være opptatt med å spre sine gener ved å få så mange barn som

mulig. Disse må i sin tur være i stand til å spre sine gener videre. Det kan innvendes at dette i ytterste konsekvens avhenger av at ens barn har gode oppvekstvilkår og derigjennom omsorgsevne – faktorer som utvilsomt er viktig for livskvalitet.

Harris begrunner at maksimering av lykke er den eneste moralske sannhet vi med overbevisning kan ha i vår nåværende eksistens som bevisste vesener, illustrert med et hypotetisk eksempel der vi har valget mellom en verden der alle mennesker opplever den verst tenkelige lidelsen kontra en verden der alle mennesker opplever den høyest tenkelige lykke. Han forfekter at enhver handling, på individ-og samfunnsnivå, begrunnet i et forsøk på å bevege seg vekk fra førstnevnte tilstand, og i retning sistnevnte, representerer det moralsk rette. Innvendingen om at vitenskapen til nå ikke har funnet svar på hvordan vi best veier ulike scenarioer opp mot hverandre ved å treffe beslutninger som medfører størst økning i lykke tilbakevises: det at vi hittil ikke har løst problemet med å balansere ulike hensyn mot hverandre for at konsekvensene skal bli høyest mulig lykke for flest mulig levende vesener, betyr ikke at det ikke eksisterer et svar på dette spørsmålet. På samme måte er det vanskelig å finne ut av hvor mange biler som til enhver tid er i bevegelse på kloden, selv om svaret finnes.

Harris overbevisning om vitenskapens potensiale til å tette gapet mellom er og bør har møtt seriøs kritikk i hovedsakelig tre former: *At det ikke finnes noen vitenskapelig basis for å argumentere for at vi bør verdsette vår egen eller noen andres livskvalitet; følgelig, dersom noen ikke bryr seg om livskvalitet, eller kun bryr seg om sin egen livskvalitet og ingen andres, så finnes det ingen måte å argumentere mot dette standpunktet på et vitenskapelig grunnlag; Selv om vi ble enige om å prioritere livskvalitet øverst i enhver moralsk diskusjon, er det vanskelig eller umulig å definere hva livskvalitet er. Det er følgelig umulig å måle livskvalitet vitenskapelig. Derfor kan det ikke finnes noen moralvitenskap* (Harris, 2011). Sentralt i Harris motsvar er at vitenskapen i seg selv er basert på verdier som ikke i utgangspunktet er vitenskapelig begrunnet – slik nevnt over. På samme måte som vitenskapen forutsetter verdien

av bevisførsel og logisk konsistens, kan moralvitenskapen basere seg på verdien av livskvalitet, i den forstand at en forsøker å unngå den verst tenkelige lidelsen for alle mennesker. Dette uten at utgangspunktet for hvorfor en prioriterer et slikt mål må være vitenskapelig dokumentert. Vår formaning om hva som bør gjøres er ifølge Harris slik sett innebygd i vitenskapens fundament. Harris sammenligner begrepet livskvalitet med det vanskelig definerbare begrepet helse. Det finnes ingen vitenskapelig basis for å begrunne at vi prioriterer å bedre menneskers helse, men dette betyr ikke at det er i strid med vitenskapen å søke å etterstrebe bedre helse for flest mulig mennesker. Det betyr heller ikke at det ikke finnes bedre eller dårligere svar på hva som fremmer god helse, i så fall ville medisinen som vitenskapelig disiplin være retningsløs. Kjernen i Harris argumentasjon er at det må finnes potensiale for utvikling i vår forståelse av livskvalitet – og at enkelte interesser og preferanser gjør det vanskelig for oss utnytte dette potensialet. I prinsippet anser han det som sikkert at enkelte mennesker og kulturer ikke vet sitt eget beste hva angår å leve et moralsk riktig liv (Harris, 2011).

Greene (2013) erkjenner og at våre moralske valg må ta utgangspunkt i at alle mennesker opplever negative og positive følelser, glede og lidelse. Både han og Harris støtter en utilitaristisk begrunnet filosofi med formål å maksimere lykke. Likevel, der Sam Harris argumenterer for at anstrengelser med formål å øke lykke utgjør den moralske sannheten, presiserer Greene at han ikke hevder at utilitarisme er den moralske sannhet, ei heller at vitenskapen beviser at utilitarisme er den moralske sannhet. Det han derimot hevder er at utilitarisme er den foretrukne filosofi for å avgjøre moralske kontroverser i det øyeblikket vår moralske tenkning er objektivt forbedret av en vitenskapelig forståelse av moral. Greene påpeker at moral ikke er avgrenset til et sett av frittstående abstrakte sannheter som vi ikke kan få tilgang til grunnet våre begrensede menneskelige hjerner. Moralpsykologi er ikke noe som tidvis berører moralfilosofiens abstrakte landskap. Moralfilosofien er en manifestasjon av

moralpsykologien, den intellektuelle toppen av et langt større og dypere psykologisk og biologisk isfjell.

Illustrativt forstår han stridende moralske filosofier som predikerbare produkter av våre dobbelt-prosesserende hjerner. Han knytter de tre mest sentrale skolene i vestlig filosofi; utilitarisme/konsekvensialisme (Bentham og Mill), deontologi (Kant) og dydsetikk (Aristoteles), til måter vår manuelle modus kan gi mening til de automatiske innstillingene den er innebygd i (husk analogien til funksjoner i et kamera). Manuell-modus kan brukes til eksplisitt å beskrive våre automatiske innstillinger (Aristoteles), til å rettferdiggjøre våre automatiske innstillinger (Kant), eller til å overstige begrensningene i våre automatiske innstillinger (Bentham og Mill). Aristoteles dyder er et produkt av vår evne til å beskrive våre intuitive, emosjonelle reaksjoner i moralske situasjoner, mens Kants pliktetikk er et forsøk på å innramme disse på en regelbundet, rasjonell måte. Bentham og Mills utilitarisme ser han derimot som den eneste mulige fremgangsmåten for å fleksibelt kunne variere mellom avgjørelser drevet av automatiske innstillinger (raske, intuitive), styrt av emosjoner (empati, skam, sinne), i situasjoner hvor det er fordelaktig, kontra overstyring av i større grad kalde, rasjonelle beslutningsstrategier i andre situasjoner hvor dette kreves. Distinksjonen mellom automatisk/manuell er ikke abstrakt men representert i hjerneaktivitet i ulike områder av prefrontal korteks, slik beskrevet i forrige seksjon. Alle forstår vi poenget med å øke livskvalitet for flest mulig mennesker da vår alles hjerne innehar den samme manuelle modusen. Samtidig vekker det utilitaristiske argumentet en motforestilling da vi alle har ikke-utilitaristiske automatiske innstillinger, intuitive følelsesdrevne reaksjoner i situasjoner der lykkemaksimering totalt sett går på bekostning av enkelte individers rettigheter (Greene, 2013).

Dyp pragmatisme – foretrukket som begrep fordi utilitarisme blant annet er misforstått som maksimering av velstand i form av penger og materielle verdier – er Greene's (2013) forslag til en verdensomspennende meta-moral, en felles moralsk valuta. Den er begrunnet i

hjernens dobbelt-prosesserende moralpsykologi, da manuell modus er egnet til å oppnå fremtidige mål-orienterte konsekvenser. *Dersom jeg har rett, er utilitarisme spesielt, og Bentham og Mill gjorde noe hittil ukjent i intellektuell historie. De tvang moralfilosofi vekk fra våre automatiske innstillinger, vekk fra begrensningene i våre biologiske og kulturelle historier, og gjorde det håndterlig for hjernens system for generell problemløsning. Den manuelle modus gir ikke opphav til en moralfilosofi, men den kan skape en dersom den næres av to universelt tilgjengelige moralske verdier: lykke og upartiskhet* (Greene, 2013, p. 203). Jeg forstår Greene slik at utilitarismen i moralske kontroverser – situasjoner der ulike grupper er grunnleggende uenige og henviser til sine respektive rettigheter – tilbyr en mulighet til å tilsidesette våre biologiske (automatisk funksjon) og kulturelle begrensninger (gruppetilhørighet rundt felles verdier) til fordel for upartisk utarbeidede kompromisser mellom grupper, begrunnet i lykkemaksimering.

Konklusjon – refleksjoner

Hva er implikasjonene av kunnskapen om at moral er en evolusjonært utviklet menneskelig kapasitet for samarbeid mellom individer, og at denne kapasiteten drives av grunnleggende nevrobiologiske interaksjoner der hormonene Oxytocin og Vasopressin, sammen med transmittere som Dopamin og Serotonin, er involvert? Er det slik at kunnskap om prosessene som utviklet oss og komplekse forbindelser i hjernen gjør oss bedre i stand til å, hvis noe slikt finnes, agere moralsk rett? Utfordret med konkrete moralske spørsmål i dagens samfunn, hevder Churchland (2018) at svaret er nei. Det bringer oss sannsynligvis ikke nærmere svar på spørsmål om abort, genredigering og prioriteringer i helsevesenet. I et bredere perspektiv er svaret muligvis ja, da en forståelse av at vår biologiske natur er et resultat av naturlig seleksjon og at våre kulturelle praksiser dypest sett følger av disse fakta utfordrer filosofiens søken etter fundamentale moralske prinsipper som kan anvendes i enhver moralsk situasjon universelt og til alle tider.

Churchland (2011) vektlegger Aristoteles innsikt om balanse i hverdagslig liv og tenkning opparbeidet gjennom vanedanning og utvikling av dyder, sammen med forståelsen av at moralske valg alltid innebærer komplekse vurderinger der en, om enn aldri, sjelden kan treffe beslutninger som imøtegår alle hensyn. Da er/bør-innvendingen er begrenset til strengt deduktive slutninger kan den tilsidesettes for et dypere psykologisk og nevrobiologisk perspektiv på hva resonnering og problemløsning er, hvordan å håndtere det sosiale liv, og hvordan vurderinger og beslutninger gjengis nevralt. En kan forsøke å finne ut hva som bør gjøres ved å anvende ens kunnskap, persepsjon, emosjoner og rasjonalitet på en måte der ulike forhold balanseres opp mot hverandre. I moralsk ladde situasjoner i det virkelige liv tar dette form som en prosess av begrenset behovstilfredstillelse («constraint satisfaction»), ikke som deduksjon. Hun vektlegger en naturlig dynamisk moral informert av vitenskapelig tilegnet kunnskap. Navigering i den sosiale verden avhenger av de samme nevralt mekanismer som kreves for å gjenkjenne og kategorisere kausale prosesser; motivasjon og driv, belønning og prediksjon, persepsjon og hukommelse.

Dersom Churchland skal tillegges et svar på hva som er den moralske sannhet, uttrykkes den ikke i form av et fundamentalt moralsk prinsipp utledet fra resonnering, men som en erkjennelse av at våre verdier er forankret i hjernens nettverk organisert for omsorg og omtanke for vår egen, våre barns, venners og slektingers lykke. Dette former hvordan vi tenker om alt fra konfliktløsning, distribusjon av ressurser og flere andre komplekse tema i samfunnet. Samtidig som at verdiene og deres materielle eksistens (nevralt korrelater) innehar begrensninger for måten vi typisk løser moralske kontroverser på, er de fakta som gir grunnlag for å avveie hva som betyr noe for oss i slike spørsmål, slik som at vi bryr oss om våre barn, ønsker at deres livskvalitet skal være god, og at vår gruppe av familie og venner er betydningsfulle for oss. Disse verdiene, begrunnet med fakta, gir grunnlag for løsninger på sosiale problemer som er bedre enn andre (Churchland, 2011).

Vi kan sammenfatte denne innsikten med etablert sosialpsykologisk forskning om at mennesker finner fellesskap, tilhørighet og trygghet i grupper, men at dette samtidig kan innebære favorisering av ens egen gruppe over andre grupper (Balliet, Wu, & De Dreu, 2014). Det er da av praktisk betydning å forsøke å etablere enighet om en verdi, eller noen få verdier, som er viktige for de fleste mennesker og som er direkte knyttet til menneskers opplevelse uavhengig av gruppetilhørighet. Greene (2013) beskriver at mennesker må løse moralske problemer på to ulike nivåer; avveininger mellom mine og dine interesser («the tragedy of the commons») og avveininger mellom våre og deres interesser («the tragedy of commonsense morality»). Aristotelisk dydsfilosofi kan gi lærdom om hvordan å være et godt menneske på et generelt grunnlag. Den er anvendbar ved spørsmål om mine og dine interesser, men kommer ifølge Greene til kort som filosofi da den ikke er av praktisk relevans hva angår samtidens moralske kontroverser: abort, donasjoner til hjelpetrequende fremmede mennesker i u-land og strukturering av helsesystem. Dette da oppfatninger av hva som er moralsk rett i slike spørsmål besvares ulikt avhengig av hvilken gruppe («tribe») en tilhører – i situasjoner der det står mellom våre og deres interesser. Én gruppes dyder anses av en annen gruppe å være umoralsk (Greene, 2013). Det kan derfor argumenteres for å bruke vår dobbelt-prosesserende moralpsykologi på rett problem til rett tid. Vår moralske intuisjon (samvittighet) gjør oss som hovedregel i stand til å samarbeide og løse uenigheter innad i grupper. Ved konflikter mellom grupper med ulike verdsett er vi derimot avhengig av å bruke manuell modus for å navigere oss frem til valg basert på de universelle verdiene upartiskhet og lykke (Greene, 2013).

En dyp pragmatisk moralfilosofi – begrunnet i en teori om hjernens dobbelt-prosesserende måte å håndtere moralske tema på – med siktemål å ta beslutninger som gir lykkemaksimerende konsekvenser, møter på innvendinger. For det første er det vanskelig å predikere konsekvenser av en beslutning i innviklede moralske spørsmål. For det andre kan maksimering av livskvalitet måles enten i gjennomsnitt eller sum lykke. Dersom gjennomsnitt

lykke er målet, kan en i teorien se for seg et samfunn bestående av få, men veldig lykkelige mennesker. Derimot vil sum lykke i teorien kunne bety et samfunn bestående av mange mennesker der hvert individ bare er moderat lykkelig, men der den totale lykken er høyere enn i det første scenarioet. En utilitarist kan derimot stille det betimelige spørsmålet; hvorfor kan vi ikke få begge? Hvis en forplikter seg til idéen om at noe skal maksimeres aksepterer en samtidig at det finnes en endimensjonal, kvantifiserbar skala, hvor det som skal maksimeres kan plasseres. Hvis en da maksimerer gjennomsnittet vil ikke summen bli maksimal, og omvendt. Riktignok kan det matematisk være mulig å maksimere både sum og gjennomsnitt lykke, men i virkeligheten vil det alltid være en viss variasjon i enhver populasjon. Så selv om vi beveger oss fra x mengde global lykke til høyere global lykke representert ved y , er det vanskelig å forestille seg at alle individer i praksis vil være eksakt like lykkelige slik at sum og gjennomsnitt er maksimert. Dersom en holder maksimering av gjennomsnittlig lykke som den moralske sannhet, vil en være etisk forpliktet til å eliminere individene som trekker ned gjennomsnittet. Dersom kun summen gjelder vil andre praksiser med nesten like absurde konsekvenser måtte etterleves.

Videre er sinnelagsetikkens (deontologiens) appell til rettigheter vanskelig å omgå. Hvis livet har verdi i seg selv, uavhengig av dets kvalitet, i tråd med menneskerettighetserklæringen, så kan vel ikke livskvalitet totalt sett være den eneste verdien vårt moralske kompass skal orientere etter? Enhver debatt om moral vil bygge på argumenter, men argumentasjonen foregår mellom mennesker med sine emosjonelle og kognitive begrensninger. Utilitarisme fjerner ikke disse begrensningene. Det at våre følelser påvirker moralske beslutninger er ikke i seg selv et argument for at måten vi oppfører oss på i bestemte emosjonelle tilstander er moralsk galt (eller riktig). Et viktig argument mot utilitarismen er at vi ikke skal bruke mennesker som midler til å nå våre mål. Det er vanskelig å se at dette argumentet blir tilbakevist av å påvise at det i noen sammenhenger er en korrelasjon mellom visse følelser og visse sinnelagsetiske handlinger.

Grunnen til at siviliserte samfunn fordømmer slavehandel, tortur etc. har ikke referanse til følelseslivet vårt, men til den abstrakte ideen om at alle mennesker har det samme iboende menneskeverdet.

Etter en gjennomgang av utilitarisme utøvd til det ekstreme der alles livskvalitet (inkludert ens egen) vektlegges likt og der økning av den totale livskvalitet veier tyngre enn alle andre moralske prinsipper, konkluderer Churchland (2011, p. 179) med følgende: *Etter min mening er konsekvensialisme mest anvendbar dersom den aksepteres som et sett med eksempler av moralske prototyper, problemstillinger der vi kan enes om at kalkulering av prototypiske gode konsekvenser (de som omhandler livskvalitet) tjener oss vel, snarere enn å forholde oss til det som en absolutt regel. På den måten, slik som ved ikke-moralske kategorier, kan vi fra prototypene utvide til nye problemstillinger ved analogi, sensitive til multiple begrensninger... ikke desto mindre, slik Mill korrekt påpekte, er vektlegging av konsekvenser for menneskelig livskvalitet, på tross av vanskeligheten med å kalkulere presist, alltid relevant.*

Hva kan vi så konkludere med? Det synes som at kunnskap om moralpsykologi kan gi oss verdifull innsikt i moralens utvikling, nevrobiologiske grunnlag, og moralsk beslutningstaking. Kunnskapen om disse tre temaene er deskriptiv – beskrivende – og gir oss verdifull moralteori som foreslår hvordan ting har blitt eller er. Men kan denne kunnskapen også komplementere normativ moralfilosofisk teori? Innbakt i begge problemstillingene ligger nemlig spørsmålet om vi kan trekke slutninger fra det som er til det som bør være.

Det foreligger gode argumenter mot å kunne bruke kunnskap om moralens evolusjon til å treffe normative beslutninger. Det samme gjelder for strengt deduktive avledninger basert på avgrensede empiriske fakta. Derimot bruker Greene (2013) forskningsbasert kunnskap om moralpsykologisk beslutningstaking til å begrunne sin dypt pragmatiske moralfilosofi. Også Churchland (2011) ser forbi er/bør – distinksjonen ved å hevde at psykologi, biologi og generelt sett vitenskapelig metode, kan føre oss nærmere gode svar på normative moralske spørsmål.

For det første fordi måten vi gjør beslutninger, og løser moralske problemer, ikke arter seg som logisk stringente deduksjoner. For en syllogisme vil det alltid være slik at konklusjonen må være sann dersom premissene er sanne, noe som ikke gjelder for forholdet mellom er og bør. Den syllogistiske skråsikkerheten er vanskelig, om ikke umulig, å oppnå i moralske spørsmål. Denne begrensningen har historisk sett ikke hindret vitenskapelige fremskritt. Altså kan vi kanskje utvikle en fruktbar moralteori ved bruk av faktisk kunnskap uten å måtte forholde oss til syllogistisk logikk. For det andre da vitenskapelig metode er bedre egnet enn religion og moralfilosofisk resonnering til å lære mer om hjernen og dens navigering i miljøet vi omgir oss med. En hjerne som på et grunnleggende vis er utgangspunktet for alt vi verdsetter og etterstreber i vårt liv som mennesker.

Vi har i dag ikke et definitivt svar på hvorvidt det finnes én eller flere moralske sannheter, eventuelt én faktor av moralsk betydning eller flere. Angående dette anser jeg Harris (2010) observasjon som fornuftig; det foreløpige fraværet av et svar impliserer ikke at et svar ikke finnes. Jeg tenker at denne oppgaven har vist at moralpsykologi kan bidra til deskriptive moralteorier. Jeg er og av den oppfatning at moralpsykologi – og andre vitenskapelige disipliner som beskjeftiger seg med moral – har potensialet til å veilede oss i normative spørsmål. Først og fremst gjennom at kunnskap om hjernen og dens fungering kan øke vår bevissthet rundt komplekse sammenhenger i moralske spørsmål. Eksempelvis bakgrunnen for ulike individers stridende interesser, i tillegg til generell kunnskap om forhold som påvirker våre beslutninger.

Jeg ønsker å avslutte med et sitat som vektlegger vår generasjons ansvar for å ta beslutninger i vår samtids moralske kontroverser. En samtid der jeg er av den oppfatning at innsikt i tradisjonell moralfilosofi, både dydsetikk, pliktetikk og utilitarisme – utfylt av vitenskapelig tilegnet kunnskap om moral – gjør oss best skikket til å foreta avveininger i komplekse moralske spørsmål. *Moral er hvordan vi finner løsninger sammen med andre uten bruk av makt og autoritet, og, selv om vi kan få hint fra tidligere generasjoner, er det vi i den*

nåværende generasjon som må forhandle og finne ut av våre egne moralske relasjoner
(Tomasello, 2016)

Referanser

- Anagnostou, E., Soorya, L., Brian, J., Dupuis, A., Mankad, D., Smile, S., & Jacob, S. (2014). Intranasal oxytocin in the treatment of autism spectrum disorders: A review of literature and early safety and efficacy data in youth. *Brain Research, 1580*, 188-198. doi:<https://doi.org/10.1016/j.brainres.2014.01.049>
- Balliet, D., Wu, J., & De Dreu, C. K. W. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin, 140*(6), 1556-1581. doi:10.1037/a0037737
- Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650.
- Churchland, P. S. (2011). *Braintrust : what neuroscience tells us about morality*. Princeton, N.J: Princeton University Press.
- Churchland, P. S. (2013). *Touching a nerve : the self as brain*. New York: W.W. Norton & Co.
- Churchland, P. S. (2018). *Preface to the Princeton Science Library Edition, Braintrust*. New York: Princeton University Press.
- Churchland, P. S. (2019). *Conscience: the Origins of Moral Intuition*. New York: W.W Norton & Co.
- Crockett, M. J. (2013). Models of morality. *Trends in cognitive sciences, 17*(8), 363-366.
- Crockford, C., Wittig, R. M., Langergraber, K., Ziegler, T. E., Zuberbuhler, K., & Deschner, T. (2013). Urinary oxytocin and social bonding in related and unrelated wild chimpanzees. *Proc Biol Sci, 280*(1755), 20122765. doi:10.1098/rspb.2012.2765
- Cureton, A., & Johnson, R. (2019). Kant's Moral Philosophy. *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*. Retrieved from <https://plato.stanford.edu/archives/spr2019/entries/kant-moral/>.

- Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and social psychology review, 17*(3), 273-292.
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: the aversion to harmful action. *Emotion, 12*(1), 2-7. doi:10.1037/a0025071
- Dal Monte, O., Noble, P. L., Turchi, J., Cummins, A., & Averbeck, B. B. (2014). CSF and Blood Oxytocin Concentration Changes following Intranasal Delivery in Macaque.(Research Article). *PLoS ONE, 9*(8).
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Ditzen, B., Schaer, M., Gabriel, B., Bodenmann, G., Ehlert, U., & Heinrichs, M. (2009). Intranasal oxytocin increases positive communication and reduces cortisol levels during couple conflict. *Biol Psychiatry, 65*(9), 728-731.
- Ebitz, R. B., Karli, K. W., & Michael, L. P. (2013). Oxytocin blunts social vigilance in the rhesus macaque. *Proceedings of the National Academy of Sciences, 110*(28), 11630. doi:10.1073/pnas.1305230110
- Flanagan, O. J. (2007). *The Really Hard Problem : Meaning in a Material World*. Cambridge, UNITED STATES: MIT Press.
- Gray, K., & Graham, J. (2018). *Atlas of Moral Psychology*. New York, UNITED STATES: Guilford Publications.
- Greene, J. D. (2013). *Moral tribes : emotion, reason, and the gap between us and them*. New York: Penguin Press.
- Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*(3), 1144-1154. doi:10.1016/j.cognition.2007.11.004
- Guastella, A. J., Einfeld, S. L., Gray, K. M., Rinehart, N. J., Tonge, B. J., Lambert, T. J., & Hickie, I. B. (2010). Intranasal Oxytocin Improves Emotion Recognition for Youth with

- Autism Spectrum Disorders. *Biol Psychiatry*, 67(7), 692-694.
doi:<https://doi.org/10.1016/j.biopsych.2009.09.020>
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev*, 108(4), 814-834. doi:10.1037/0033-295x.108.4.814
- Haidt, J. (2012). *The righteous mind : why good people are divided by politics and religion*. London: Allen Lane.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7(1), 1-16. doi:[https://doi.org/10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4)
- Harris, S. (2004). *The End of Faith: Religion, Terror and the Future of Reason*. New York: W. W. Norton & Co.
- Harris, S. (2010). *The Moral Landscape : How Science Can Determine Human Values*. Great Britain: Bantam Press.
- Harris, S. (2011, 25.05). A Response to Critics. *HuffPost*. Retrieved from https://www.huffpost.com/entry/a-response-to-critics_b_815742?guccounter=1
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York, NY, US: Cambridge University Press.
- Hume, D., Nidditch, P. H., & Selby-Bigge, L. A. (1978). *A treatise of human nature* (2nd ed. with text revised and variant readings by P. H. Nidditch. ed.). Oxford: Clarendon Press.
- Kahneman, D. (2012). *Thinking, fast and slow*. London: Penguin Books.
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, 435(7042), 673-676. doi:10.1038/nature03701
- Levine, S., Leslie, A., & Mikhail, J. (2016). The Mental Representation of Human Action. *Cognitive Science*, 42. doi:10.1111/cogs.12608

- Liu, N., Hadj-Bouziane, F., Jones, K. B., Turchi, J. N., Averbeck, B. B., & Ungerleider, L. G. (2015). Oxytocin modulates fMRI responses to facial expression in macaques. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), E3123. doi:10.1073/pnas.1508097112
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cogn Behav Neurol*, *18*(4), 193-197. doi:10.1097/01.wnn.0000191292.17964.bb
- Miller, R., Hannikainen, I., & Cushman, F. (2014). Bad Actions or Bad Outcomes? Differentiating Affective Contributions to the Moral Condemnation of Harm. *Emotion (Washington, D.C.)*, *14*. doi:10.1037/a0035361
- Parr, L. A., Brooks, J. M., Jonesteller, T., Moss, S., Jordano, J. O., & Heitz, T. R. (2016). Effects of chronic oxytocin on attention to dynamic facial expressions in infant macaques. *Psychoneuroendocrinology*, *74*, 149-157. doi:<https://doi.org/10.1016/j.psyneuen.2016.08.028>
- Parry, R. (2014). «Ancient Ethical Theory». *The Stanford Encyclopedia of Philosophy*(Fall 2014 edition).
- Piaget, J. (1948). *The moral judgment of the child*. New York, NY, US: Free Press.
- Porges, S., & Carter, C. (2011). Neurobiology and evolution: Mechanisms, mediators, and adaptive consequences of caregiving. *Self Interest and Beyond: Toward a New Understanding of Human Caregiving*, 53-71.
- Roberts, G. (2005). Cooperation through interdependence. *Animal Behaviour*, *70*(4), 901-908. doi:10.1016/j.anbehav.2005.02.006
- Schmidt, M. F. H., & Tomasello, M. (2012). Young Children Enforce Social Norms. *Current Directions in Psychological Science*, *21*(4), 232-236. doi:10.1177/0963721412448659

- Sinnott-Armstrong, W. (2019). «Consequentialism». *The Stanford Encyclopedia of Philosophy, Summer 2019 Edition*
- Thomson, J. J. (1985). The Trolley Problem. *Yale Law Journal, 94*, 1395-1855.
- Tomasello, M. (2016). *A Natural History of Human Morality*. London: Harvard University Press.
- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annu Rev Psychol, 64*, 231-255. doi:10.1146/annurev-psych-113011-143812
- Trivers, R. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol., 46*. doi:10.2307/2822435
- Walum, H., Lichtenstein, P., Neiderhiser, J., Reiss, D., Ganiban, J., Spotts, E., . . . Westberg, L. (2012). Variation in the Oxytocin Receptor Gene Is Associated with Pair-Bonding and Social Behavior. *Biol Psychiatry, 71*, 419-426. doi:10.1016/j.biopsych.2011.09.002
- Walum, H., Waldman, I. D., & Young, L. J. (2016). Statistical and Methodological Considerations for the Interpretation of Intranasal Oxytocin Studies. *Biol Psychiatry, 79*(3), 251-257. doi:10.1016/j.biopsych.2015.06.016
- Walum, H., & Young, L. J. (2018). The neural mechanisms and circuitry of the pair bond. *Nature reviews. Neuroscience, 19*(11), 643-654. doi:10.1038/s41583-018-0072-6
- Wilson, S., David, & Wilson, O. E. (2008). Evolution «for the Good of the Group». *American Scientist, 96*.
- Young, K. A., Gobrogge, K. L., Liu, Y., & Wang, Z. (2011). The neurobiology of pair bonding: insights from a socially monogamous rodent. *Front Neuroendocrinol, 32*(1), 53-69. doi:10.1016/j.yfrne.2010.07.006

