

# Kapitel 10: Etik, redelighed og privacy

Henrik Kragh Sørensen      Mikkel Willum Johansen  
22. april 2022

<b>Indhold</b>	5
10.1 Deskriptiv etik, normativ etik og jura . . . . .	1
10.2 Første normative teori: Utilitarisme . . . . .	4
10.3 Anden normative teori: Kants pligtetik . . . . .	6
10.4 Nagels syntese . . . . .	8
10.5 Analyser i praksis . . . . .	10
10.6 Big data-etik og privacy . . . . .	12
10.7 Bias og algoritmisk transparens . . . . .	14
10.8 Systemisk etik . . . . .	17
10.9 IT-professionelles etiske ansvar . . . . .	18
10.10 Redelighed . . . . .	19
<b>Litteratur</b>	<b>23</b>

## 10.1 Deskriptiv etik, normativ etik og jura

På platformen Moral Machines (<http://moralmachine.mit.edu/>) har en gruppe forskere fra MIT i en årrække indsamlet information om, hvordan folk mener, at en selvkørende bil skal reagere i en række såkaldt no-win-scenarier, hvor bilen vil blive tvunget til at dræbe enten én gruppe personer eller en anden gruppe personer. Forskerne har løbende indsamlet og analyseret brugernes svar, hvilket der er kommet en række interessante artikler ud af. Fx har det vist sig, at der er store regionale forskelle i de valg, folk tager (Maxmen, 2018). Hvor man i Sydamerika har en præference for at redde kvinder, børn og folk af høj status, er man i den sydlige del af Asien mere optaget af at redde fodgængere og lovlydige, og foretrække desuden at undgå at handle. Undersøgelsen er et eksempel på det, man kan kalde deskriptiv etik, dvs. en undersøgelse af de etiske valg folk faktisk foretager (eller siger de vil foretage) i praksis. Som undersøgelsen antyder, kan der være kulturelle forskelle mellem de etiske præferencer, folk i praksis har.

Det, at folk handler på en bestemt måde, betyder naturligvis ikke, at det er den rigtige måde at handle på. Det ville være en klar fejlslutning (der i etikken kaldes *den naturalistiske fejlslutning*). Undersøgelsen af, hvordan folk bør handle for at opføre sig etiske korrekt kaldes normativ etik, og består til dels af teorier, der giver generelle bud på, hvordan man bør opføre sig i forskellige situationer. Som eksemplet med Moral Machines viser, er der ikke

35 enighed om, hvad man skal gøre i en given situation, og den uenighed søger de normative teorier at råde bod på.

Nu er der heller ikke enighed om, hvilken normative etik, der er den rigtige, og vi vil her se på to forskellige teorier, der repræsenterer meget forskellige etiske intuitioner.

Hvis man bliver overbevist af en af de normative teorier, vil man have (mere eller  
 40 mindre) klare handlingsanvisninger i alle de etiske valg, man fremover vil stå i. Som vi skal se, rummer de normative teorier dog selv en række problemer, men selv om man ikke accepterer teorierne i deres helhed er det stadig værdifuldt at overveje rækkevidden af de overordnede principper, teorierne repræsenterer. Fra det mere pragmatiske synspunkt på etikken, som vi vil tillade os at anlægge i denne bog kommer den mest grundlæggende etiske  
 45 fordring hverken fra den ene eller den anden etiske teori, men består i at man i det hele taget tager den etiske udfordring op og forsøger at nå frem til en form for reflekteret ligevægt (jf Rawls, 1972), hvor man er konsistent i den forstand, at der er sammenhæng mellem ens overordnede principper og de konkrete domme, man foretager, og at de argumenter man accepterer i en kontekst også bliver accepteret i andre. Vi vil senere i kapitlet give en  
 50 struktur for, hvordan den slags analyser kan foregå.

Udover deskriptiv og normativ etik er det i denne sammenhæng også værd at inddrage det juridiske område, dvs. spørgsmålet om, hvad der er lovligt og hvad der ikke er lovligt. Ideelt set burde der være et perfekt sammenfald mellem det etiske og det juridiske, sådan at lovlige handlinger altid er etisk korrekte og omvendt. I praksis er det dog ikke så simpelt,  
 55 og der opstår jævnligt situationer, hvor folk bliver nødt til at bryde love for at gøre, hvad de anser for etisk rigtigt. Et godt eksempel er whistleblowere som EDWARD SNOWDEN, (visse) ’white-hat’ hackere, der bryder loven for at gøre opmærksom på sårbarheder i kritiske IT-systemer, eller spørgsmålet om ophavsret og deling af vigtig information. Det sidste blev fx sat på spidsen af aktivisten AARON SWARTZ (1986–2013), der i det såkaldte *Guerilla Open*  
 60 *Access Manifesto* fra 2008 skrev:

It’s called stealing or piracy, as if sharing a wealth of knowledge were the moral equivalent of plundering a ship and murdering its crew. But sharing isn’t immoral — it’s a moral imperative. Only those blinded by greed would refuse to let a friend make a copy. [...] There is no justice in following unjust laws. It’s  
 65 time to come into the light and, in the grand tradition of civil disobedience, declare our opposition to this private theft of public culture. (Swartz, 2020)

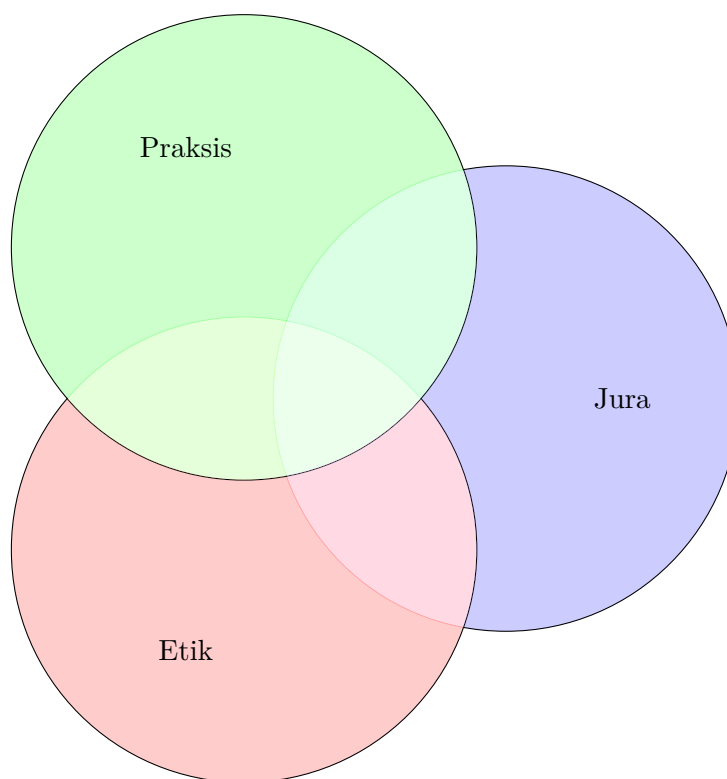
SWARTZ begik i 2013 selvmord mens han var under anklage for ulovligt at have downloaded et meget stort antal videnskabelige artikler til sin computer.

Det er et åbent spørgsmål, om SWARTZ havde ret i, at man har en moralsk forpligtelse  
 70 til at dele information, men der er her en klar brudflade mellem etik og jura, og diskussionen har været med til at flytte både juraen, vores praksis og hvad vi anser for etisk korrekt fx gennem opfindelsen af ny rettighedsformer som *creative commons* (som SWARTZ også var med til at udforme).

Brudfladen mellem etik og jura kan også udspille sig på den måde, at ting, vi anser for  
 75 etiske forkerte, faktisk er lovlige. Specielt på områder, hvor der er en hastig teknologiudvikling, vil der ofte være en form for juridisk vakuum, hvor teknologien muliggør nye praksiser, der forekommer åbenlyst uetiske, men som alligevel er lovlige i en årrække af den simple grund, at lovgiverne først skal opdage problemet og bagefter blive enige om en passende lovgivning for at ulovliggøre den uetiske praksis, og det kan tage meget lang tid. Kritikere som  
 80 SOSHANA ZUBOFF hævder, at det er præcis dette vakuum firmaer som Google og Facebook

bevidst har udnyttet til i en årrække at profitere af en forretningsmodel, der grundlæggende er uetisk (Zuboff, 2019). Det kan naturligvis diskuteres, om ZUBOFF og ligesindede har ret i deres kritik (vi vil vende tilbage til spørgsmålet senere), men der er til gengæld meget lidt tvivl om, at det juridiske vakuum hastig teknologiudvikling skaber, er en realitet.

Det tre områder, der udgøres af vores faktiske *praksis*, den *normative etik* og *juraen* er altså tæt forbundne, men ikke altid sammenfaldende. Af den grund er undskyldninger som „alle gør det, så er det etisk korrekt“, eller „det er ikke ulovligt, altså er det etisk korrekt“ åbenlyst fejlslutninger: Undskyldningerne ville kun virke, hvis der var perfekt overensstemmelse mellem hvad vi faktisk gør, hvad vi burde gøre og hvad der er lovligt; men som vi har set er dette ikke tilfældet, specielt i områder med hastig teknologiudvikling. De tre områder er dog tæt sammenflettede, og udviklingen på et område kan meget let påvirke udviklingen på de to andre. I 1987 fik vi således i Danmark *Det Ethiske Råd*, der skulle rådgive politikere om udviklingen inden for gen- og bioteknologi. Siden da har både praksis, lovgivningen og den normative vurdering af fx en teknologi som kunstig befrugtning udviklet sig dramatisk, men hele tiden som et tæt sammenspil mellem de tre områder, hvor fx ændrede etiske vurderinger har ført til ny lovgivning, der banede vejen til nye praksiser, der har ændret den normative vurdering osv. (se fx Adrian, 2016). Nu har vi fået et *Dataetisk Råd*, der bl.a. skal rådgive i spørgsmål om privatliv og brugen af kunstig intelligens. Der er derfor al mulig grund til, at man som datalog er både parat og klædt på til at gå ind i de etiske og juridiske debatter, der rejses af den teknologiudvikling, man er en del af.



Figur 1: De tre domæner.

Hvis man har en mere instrumentel tilgang til etikken, kan man sige følgende med reference til figur 1: Det er bedst at befinde sig, hvor der er overensstemmelse mellem etik, jura og praksis, men hvis man kommer uden for det område, gør man klogt i at være på vagt.

105 Hvis man befinder sig et sted, hvor den lokale praksis er lovlig, men uetisk risikerer man etisk fordømmelse (og samvittighedskvaler). Befinder man sig et sted, hvor praksis både er ulovlig og uetisk, gør man klogt meget hurtigt at ændre praksis, og er ens praksis etisk men ulovlig er det man gør nok etiske set prisværdigt, men det kan have store personlige omkostninger (som SNOWDEN og andre whistleblowere illustrerer).

## 110 10.2 Første normative teori: Utilitarisme

Den første normative etiske teori, vi skal se på, kaldes *nytteetik* eller *utilitarisme*. Utilitarismen er et eksempel på det, man lidt bredere kalder en *konsekvensetik*, dvs. en etik, hvor den etiske vurdering af en handling udelukkende inddrager konsekvenser af handlingen, og ikke det princip, handlingen repræsenterer. I sin reneste form siger utilitarismen ganske simpelt, 115 at den etiske set rigtige handling i enhver given situation er den handling, der maksimerer den samlede nytte (eller lykke) for samtlige involverede parter.

Utilitarismen har rødder tilbage til antikken, men i den moderne udgave blev den primært formuleret af den engelske filosof JEREMY BENTHAM (1748–1832) samt JAMES MILL (1773–1836) og hans søn JOHN STUART MILL (1806–1873) i den første halvdel af 1800-tallet. 120 Utilitarismen var (navnlig for J. S. MILL) et led i en bevægelse for social reform; utilitarismen tager ikke hensyn til rang eller stand, men stiller alles nytte lige. J. S. MILL var således en tidlig og stærk fortaler for kvinders stemmeret. Utilitarismen kan desuden ses som et forsøg på at videnskabeliggøre etikken. I stedet for at tage udgangspunkt i religiøse dogmer eller mere eller mindre obskure fornuftsprincipper tager utilitarismen udgangspunkt 125 i en enkelt kvantitet — mængden af nytte — og søger at maksimere denne.

Når nytteetikken søger at maksimere netop nytte skyldes det, at nytte for dem er et såkaldt *intrinsisk gode*, dvs. noget, der er godt i sig selv. Andre goder som penge, mad eller et sted at bo er instrumentelle goder, dvs. goder, vi gerne vil have, for at opnå noget andet. Man vil have en uddannelse, så man kan få et job, man vil have et job, så man kan tjene 130 penge, man vil have penge, så man kan købe ting, og man vil have de ting, man køber, fordi de gør en lykkelig. Så for alle andre goder kan man spørge: Hvorfor vil du have det gode? Men det giver ikke nogen mening at spørge: Hvorfor vil du være lykkelig? Lykke (eller nytte) er et gode i sig selv.

I udgangspunktet virker utilitarismen som en simpel og næste selvindlysende teori, men 135 når man kigger nærmere efter, er den forbavsende nok ingen af delene. Hvis vi ser på kravet om at maksimere nytten for alle involverede parter må man for det første spørge, hvad det egentlig er, vi skal maksimere. Vi har i teksten her skiftet mellem termene 'lykke' og 'nytte', og det er ikke en oversættelsesfejl; BENTHAM og J. S. MILL skifter selv mellem at bruge termene 'happiness' og 'utility', så allerede her er der en vis uklarhed, og diverse 140 udgaver af utilitarismen har sidenhen på forskellig vis forsøgt at præcisere begrebet, uden at der er opnået enighed.

Et andet godt spørgsmål er, hvem vi skal maksimere lykken for. Den oprindelige utilitarisme blev brugt til at argumenter for social reform, da alles lykke tæller lige, men den blev også brugt til at argumentere for dyrevelfærd. I et berømt citat siger BENTHAM: „The 145 question is not, Can they reason? nor, Can they talk? but, Can they suffer?“ (Bentham, 2000, kap. XVII §4). Med det mener han, at nyttemaksimeringen ikke skal begrænse sig til intelligente væsener (mennesker), men til alle dyr, der er i stan til at føle smerte og lykke. Det gør selvfølgelig ikke regnestykket simplere — hvor meget lykke kan en bædelorm føle, og har vi en moralsk forpligtelse til at maksimere den? Og hvordan skal vi forholde os

til robotter med syntetiske følelser? Tilsvarende kan man spørge hvilken vægt fremtidige 150  
generationer har i ligningen. I klimadebatten vægter hensynet til fremtidige generationers  
lykke højt, men hvor meget skal de vægte og hvor langt skal man se frem?

Alt i alt viser det sig, at den simple idé om at maksimere lykken for alle involverede 155  
parter er så vag, at den ikke udgør en egentlig normativ teori, men snarere en skitse til  
en teori. Den bliver først en operationaliserbar teori, når man har præciseret, hvad man  
mener med de forskellige begreber, og den præcisering kan foregå på flere forskellige måder.  
Så når man fremfører utilitaristiske argumenter eller indgår i en debat med en utilitarist,  
skal man være opmærksom på, hvordan teorien (eller rettere: teoriskitsen) konkret er blevet  
implementeret, og det er klart, at man ikke i en sammenhæng kan definere 'alle involverede  
parter' på en måde, hvis man i en anden sammenhæng har defineret det anderledes. 160

Utilitarismen har også andre og mere grundlæggende problemer. For det første er det  
ikke altid klart, hvilke konsekvenser en konkret handling har. Specielt i teknologiudvikling  
er dette et virkeligt og konkret problem; TIM BERNERS-LEE, som var hovedmanden bag  
World Wide Web, kunne således næppe have forudset de enorme konsekvenser internettet 165  
har haft for vores demokratiske debat og vores måde at knytte sociale relationer på. Uti-  
litarismen kan (i et vist omfang) håndtere den indvending ved udelukkende at inddrage  
sandsynligheden for forudsebare konsekvenser. Så hvis man arbejder på en forbedret algo-  
ritme til ansigtsgenkendelse, må man med andre ord overveje, hvor sandsynligt det er, at  
den bliver brugt til forskellige ting (at låse smartphones op, at undertrykke oppositionen i  
diktaturstater mv.) og derefter lave kalkulen. Det er et godt spørgsmål, hvor grundig af- 170  
dækningen af de mulige konsekvenser skal være. Er det nok, at jeg ser på de umiddelbart  
intenderede anvendelser af min teknologi, eller har jeg, som teknologifilosoffen HANS JONAS  
(1903–1993) hævdede, et mere grundlæggende ansvar for at være futerolog, dvs. for syste-  
matisk og grundigt at gennemtænke de mulige konsekvenser af den teknologi, jeg udvikler.  
I en datalogisammenhæng har navnlig udviklingen af autonome våbensystemer og kunstig 175  
intelligens fået forskere til at tage et aktivt ansvar for de langsigtede, mulige konsekvenser  
af deres forskning.

Det mest alvorlige problem for utilitarismen er imidlertid, at det entydige fokus på den  
samlede lykke betyder, at teorien ikke tager højde for, at vi som enkeltindivider kan have  
rettigheder, der beskytter os mod at blive udnyttet af flertallets behov. Hvis det fx var 180  
nyttmaksimerende at inficere guatemalanske børnehjemsbørn med kønssygdomme for at  
teste en ny medicin — som JOHN CHARLES CUTLER (1915–2003) rent faktisk gjorde det med  
støtte fra de amerikanske sundhedsmyndigheder i 1946–48 og uden børnenes samtykke —  
ville det fra et utilitaristisk synspunkt være en etisk set korrekt handling (Reverby, 2011).  
Tilsvarende kan utilitarismen blive brugt til at retfærdiggøre handlinger, der overskrider 185  
vores ret til privatliv, vores ret til autonomi og hensynet til retfærdighed.

Utilitarismen har to svar på den udfordring. Det ene er at sige „so what?“. Hvis vores  
intuition fortæller os, at det er forkert at bruge børnehjemsbørn til medicinske forsøg eller  
at give køb på vores privatliv for at fremme det fælles bedste, så er det ikke et problem for  
utilitarismen, men et problem for vores etiske intuition. Det viser os blot, at vores intuition 190  
ind i mellem er forkert, præcis som vores forestillinger på så mange andre områder var  
fejlagtige, før de blev guides af rationelle, systematiske undersøgelser.

Det andet mulige svar er at introducere den såkaldte regelutilitarisme. I den oprindelige  
utilitarisme tager man udgangspunkt i og søger at lykkemaksimere konkrete enkelthand-  
linger. Den version af utilitarismen kaldes *handlingsutilitarisme*: Hvis det i denne konkrete 195  
situation vil maksimere lykken i verden, at jeg bruger børn til medicinske forsøg uden deres  
tilladelse, så er det utilitaristiske set det rigtige at gøre. *Regelutilitarismen* ser derimod på

generelle handlingekategorier, og søger at opstille regler, der generelt maksimerer lykken, hvis de bliver fulgt. Her kan man opstille det argument, at hvis man generelt kan bruge  
 200 børn til medicinske forsøg uden deres tilladelse, vil det formentlig føre til et fald i lykke, da folk forståeligt nok vil være nervøse for, at deres børn vil blive brugt som forsøgskaniner. Regelutilitarismen bringer dermed til en vis grad den utilitaristiske tankegang bedre i overensstemmelse med vores moralske intuition, men på den løser næppe alle problemerne. Specielt er det ikke klart, hvad det er for en generel handlingskategori, jeg følger. Det er  
 205 måske rigtigt, at en regel om at bruge børn generelt til medicinske forsøg næppe vil være lykkemaksimerende, men det kan jo godt være, at en regel, der tillader brugen af en meget specifik gruppe børn — fx guatemalanske børnehjemsbørn — til medicinske forsøg faktisk vil være lykkemaksimerende. Og dermed er vi lige vidt. Det er dermed klart, at der skal en anden form for etisk teori i spil, hvis vi ønsker at redegøre for den intuitive fornemmelse af,  
 210 at individet på en eller anden måde skal beskyttes mod kravene fra det fælles bedste. Og netop et sådant synspunkt finder vi hos IMMANUEL KANT.

### 10.3 Anden normative teori: Kants pligtetik

Den anden normative etiske teori vi skal betragte her er en *pligtetik*, der blev udformet af den tyske filosof IMMANUEL KANT (1724–1804). I pligtetikken (eller *deontologien*, som  
 215 pligtetikken også kaldes, efter det græske ord for pligt 'deon') bliver en handlings moralske værdi bedømt på handlinge i sig selv og ikke på dens konsekvenser. Typisk vil man se på, om handlinge er i overensstemmelse med et sæt givne regler. Pligtetikken er dermed en direkte modsætning til konsekvensetikken.

Præcis som der findes mange forskellige former for konsekvensetik (hvor utilitarismen  
 220 kun er én udgave) findes der også mange forskellige former for pligtetik, og KANTS pligtetik er blot en af disse. Det er dog ikke helt tilfældigt at vi netop inddrager KANTS pligtetik, da den på mange måder har været med til at forme det pligtetiske modsvar til utilitarismens konsekvensetiske betoning af det fælles bedste.

KANT udviklede sin etik i slutningen af 1700-tallet primært i de to værker *Grundlegung zur Metaphysik der Sitten* (1785) og *Kritik der praktischen Vernunft* (1788). Vi vil ikke gå  
 225 i dybden med KANTS argumentation, men i korte træk påpeger KANT, at det er uholdbart at bruge en handlings konsekvenser til at vurdere dens moralske kvalitet, da konsekvenserne altid er usikre. Derfor er vi nødt til at se på handlingsprincippet, når vi afgøre om en handling er etisk korrekt eller ej. KANT hævder desuden, at menneskets status som moralske  
 230 subjekter primært skyldes, at vi er fornuftsvæsner, der er i stand til at træffe autonome, rationelle beslutninger. Det centrale spørgsmål for etikken er, hvordan vi forvalter denne frihed, dvs: hvad er den gode vilje? KANT afviser eksplicit, at mennesker skal stræbe efter lykke eller velfærd; hvis vi handler efter vane, lyst eller begær, gør vi os til slaver af vores dyriske natur, og optræder dermed ikke som frie fornuftsvæsner. Hvis vi vil hævde  
 235 vores frihed må vi lade fornuften styre og sætte sig sine egne mål uden hensyn til hverken konsekvenserne eller vores lyster og følelser. Og det, fornuften først og fremmest kræver, er konsistens. Vi skal derfor sørge for, at de handlingsprincipper vi følger, ikke kan føre til en interne modstrid.

På den (lidt snørklede) måde når KANT frem til et grundlæggende moralsk princip, kaldet det *kategoriske imperativ*. Som navnet antyder, skal princippet forstås som et absolut  
 240 påbud, man til enhver tid og uanset konsekvenserne skal følge, hvis man vil opføre sig etiske korrekt. I en af flere formuleringer, lyder princippet: „Handl kun ifølge den maksime ved

hvilken du samtidig kan ville, at den bliver en almen gyldig lov“ (Kant, 1999, s. 78). Det kategoriske imperativ skal forstås som en form for lakmusprøve, man kan bruge til at teste sine handlinger med. Hvis man vil udføre en handling, skal man først finde ud af hvilket handlingsprincip (eller ’maksime’) man handler efter, og derefter skal man overveje, om handlingsprincippet overholder det kategoriske princip. Det kategoriske imperativ fungerer med andre ord på den måde, at det udstikker en række forbud mod bestemte handlingstyper. 245

Som et eksempel, ser KANT på følgende handlinge: Du låner penge og lover at betale dem tilbage, selv om du ikke har tænkt sig at gøre det. Maksimen for din handling er her: At afgive et løfte, man ikke har tænkt sig at holde. Den maksime kan ikke klare at blive testet med det kategoriske imperativ, for hvis alle afgav løfter, selv om de ikke havde tænkt sig at holde dem, ville selve løfteinstitutionen bryde sammen, fordi ingen længere ville tro på folk, når de afgav løfter — løfter ville blive meningsløse. Af den grund kan man ikke ville, at maksimen bliver til almen lov. Bemærk, at det ikke er fordi, man ikke bryder sig om en verden, hvor der ikke findes løfter (hvilket vil være en konsekvensetisk tænkning), at handlingen er uetisk ifølge KANT. At afgive et falsk løfte er uetisk fordi det er selvmodsigende og dermed fornuftsstridigt at udføre en handling, der ville ophæve sig selv, hvis det blev almen lov at udføre den. Tilsvarende vil handlinger som at stjæle, at lyve og at slå folk ihjel være uetiske ifølge det kategoriske imperativ idet ingen af dem kan ophæves til almen lov uden at gøre sig selv meningsløse eller umulige. 250 255 260

Den version af det kategoriske imperativ, vi har set på her, er nok den, der forklarer KANTS tankegang bedst, men samtidig er den noget vanskelig at anvende. KANT formulerede dog også imperativet i en anden (og ifølge KANT ækvivalent) version, der er lettere at anvende i praksis, og derfor nogle gange går under navnet det *praktiske imperativ*. Dette lyder: „Handl således, at du altid tillige behandler menneskeheden, såvel i din egen person som i enhver andens person, som mål og aldrig blot som middel“ (Kant, 1999, s. 88). Denne version af det kategoriske imperativ har haft en kolossal indflydelse på den etiske debat, siden KANT formulerede den. Det praktiske imperativ giver individet en klar og tydeligt beskyttelse mod at blive udnyttet af andre, uanset om udnyttelsen er til fællesskabets bedste. Det fortæller os præcist, hvad der er galt i at inficere børnehjemsbørn med kønssygdomme for at teste en ny medicin; børnene bliver brugt udelukkende som et middel til at gøre andre mennesker raske, og det går ikke ifølge KANT. 265 270

Man skal selvfølgelig ikke forstå det praktiske imperativ på den måde, at man aldrig må bruge andre som et middel til et eller andet. Så længe man husker, at de også skal være et mål i sig selv, er det ok. Det kan man i praksis sikre på forskellig vis. I forbindelse med medicinske forsøg kan man fx sikre, at forsøgspersoner afgiver et informeret samtykke om at være med i forsøget. På den måde behandler man forsøgspersonerne som mål i sig selv idet man respekterer deres autonomi og ret til selvbestemmelse. Det er dog også klart, at et informeret samtykke let kan misbruges, hvis det ikke afgives frit og informeret. *Cookie-consents* og overdrevent lange og komplicerede slutbrugerftaler på software er gode eksempler, hvor den oprindelige intention om at et samtykke skulle udtrykke respekt for brugeren ikke altid er opfyldt. Så selv om den type samtykke fra et juridisk synspunkt er gyldige, er det fra et (kantiansk) etisk synspunkt tvivlsomt, hvilke reel værdi de har. Så her er nok et eksempel på, at jura og etik ikke altid er det samme. 275 280 285

Selv om KANTS etik har været meget indflydelsesrig, har den også sine problemer. Specielt betyder det ensidige fokus på princippet bag en handling (fremfor konsekvenserne), at teorien kommer i modstrid med almindelig moralsk intuition. I en diskussion af sin teori behandler KANT følgende eksempel (Kant, 1996): En ven, der er på flugt fra en morder, har søgt tilflugt i dit hus. Morderen kommer nu forbi og afkræver dig et svar på, om vennen 290

er i huset. Er det da etisk set korrekt at lyve for morderen? KANTS svar er et klart: Nej! Hvis du fortæller sandheden til morderen, har du opfyldt din moralske pligt, og hvad der derefter sker, er ikke dit ansvar. Man er kun forpligtet i forhold til sine egen frie handlinger, mens man, som KANT ser det, ikke har nogen etiske forpligtelser i forhold til hvordan andre vælger at udøve deres autonomi. Hvis du derimod lyver for morderen, vil du, hævder KANT, paradoksalt nok have pådraget dig et ansvar, hvis din løgn fører til, at morderen finde din ven (fx fordi venen har sneget sig ud af dit hus, uden at du vidste det). Det kategoriske imperativ udstikker med andre ord i KANTS øjne en radikal og undtagelsesløs pligt, men til gengæld har du udelukkende ansvar for dine egne handlinger.

Eksemplet med den udspørgende morder illustrerer også et andet og mere grundlæggende problem i KANTS etik. Specielt den første udgave af det kategoriske imperativ lader til at forudsætte, at man entydigt kan finde ud af, hvilket handlingsprincip (eller maksime) en konkret handling tilhører. Det er dog ikke indlysende, at det forholder sig sådan i praksis, for hvilket princip følger man, hvis man lyver for den udspørgende morder? Lyver man, eller forsøger man at redde et menneske, der er i fare? Og hvis det ikke er klart, hvilket handlingsprincip man følger i en konkret situation, forsvinder den absolutte sikkerhed, KANT ønskede i sin etiske tænkning.

## 10.4 Nagels syntese

Modsætningen mellem utilitarismens hensyn til almenvellets interesser og den kantianske deontologis beskyttelse af individet udgør en helt central og grundlæggende konflikt i den moderne etiske debat. Konflikten er konstant tilstede i den politiske og juridiske debat og udgør et væsentligt tema i populærkulturen.

Konflikten kan selvfølgelig løses ved at man tager et klart valg: Man kan vælge at være enten utilitarist eller deontolog, og så vil den normative teori, man har valgt, give (mere eller mindre) klare handlingsanvisninger i de etiske dilemmaer, man står i. Der har dog også været forsøg på at samtænke de to teorier. Vi vil her kort præsentere den amerikanske filosof THOMAS NAGELS forsøg på at gøre dette, baseret på Nagel (1979) og Nagel (1986).

Hvor utilitarismen og deontologien kan beskrives som en form for 'nede fra og op'-teorier, der tager udgangspunkt i hvert sit grundlæggende aksiom (hhv. lykkemaksimering og det kategoriske imperativ) og derfra udleder, hvad der er rigtigt og forkert i alle konkrete situationer, tager NAGEL udgangspunkt i det enkelte individs etiske virkelighed og søger at give vejledning i konkrete dilemmaer og etiske valgsituationer. NAGELS primære budskab er (i vores fortolkning), at vi ikke kan reducere etikken til et simpelt valg mellem utilitarismen og pligtetik, men at etiske valg typisk vil rumme både utilitaristiske og pligtetiske aspekter. Det etiske arbejde vil så bestå i at analysere og vægte disse aspekter i den konkrete situation.

I sin analyse af den etiske situation identificerer NAGEL to kvalitativt forskellige former for etiske værdier, nemlig *agentneutrale* og *agentrelative* værdier, og han undersøger de forskellige former for etisk relevante begrundelser, vi kan give for en handling, med udgangspunkt i disse værdier.

De agentneutrale værdier er de mest oplagte, så vi vil starte med dem. De agentneutrale værdier er værdier, der ikke er knyttet til et bestemt, personligt perspektiv. Folk deler dem generelt og kan derfor umiddelbart forstå, at de bør fremmes. Fravær af smerte er et eksempel på en agentneutral værdi: Hvis vi ser en person i smerte, kan vi alle identificere os med smerten, og vi kan forstå, at det vil være godt, hvis smerten hørte op. De agentneutrale værdier er dermed objektive i den forstand at de er universelle og søger at fremme almenvellet,



frem for enkeltindviders specifikke interesser. Det er klart, at det agentneutrale perspektiv er mere eller mindre sammenfaldende med utilitarismen, men hvor utilitarismen hævder, at de agentneutrale værdier er de eneste, der er etisk relevante, påpeger NAGEL, at der findes (mindst) tre andre former for etisk relevante begrundelser, der ikke kan reduceres til agentneutrale begrundelser, nemlig:

340

**Autonome begrundelser.** Dvs. begrundelser, hvor vi slet og ret tager udgangspunkt i vores egne autonome livsprojekter og de værdier, der følger af dem.

**Deontologiske begrundelser.** Dvs. begrundelser, der udspringer af andres ret til ikke blive behandlet dårligt af mig.

**Særlige forpligtelser.** Dvs. forpligtelser, man kan have fx over for familiemedlemmer eller i form af loyalitet rettet mod en arbejdsplads.

345

Vi vil her nøjes se nærmere på de to første af de tre typer agentrelative begrundelser.

Disse begrundelser knytter alle an til et bestemt individs unikke perspektiv og er derfor hvad NAGEL kalder *agentrelative*. De *autonome begrundelser* tager udgangspunkt i det forhold, at vi mennesker har en fri vilje og er i stand til at formulere vores egne livsprojekter og mål. Disse projekter kan spille en etisk relevant rolle, også selvom de muligvis kun giver mening fra vores eget individuelle perspektiv, og ikke nødvendigvis udspringer af eller fremmer agentneutrale værdier. Hvis du fx ønsker at løbe et maratonløb, er det næppe et projekt, der vil fremme det gode i verden, og andre mennesker vil ikke nødvendigvis kunne identificere sig med dit mål, på samme måde som de kan identificere sig med dit ønske om at blive fri for en hovedpine. Men det er et mål, du kan sætte dig, og det kan give mening set ud fra dit specifikke perspektiv. Autonome værdier og begrundelser er derfor agentrelative værdier, i den forstand at de er uløseligt knyttet til en bestemt agent, dvs. et bestemt autonomt handlende og rationelt væsen. De har ikke værdi i en objektiv forstand, men har kun værdi i forhold til det konkrete individ.

350

355

360

Man kan hævde (og det gør NAGEL indirekte), at det har en iboende etisk værdi, at mennesker kan forfølge deres egne autonome mål og livsprojekter, men de agentrelative, autonome værdier kan ikke på nogen simpel måde reduceres til agentneutrale utilitaristiske værdier (selvom forskellige versioner af utilitarismen med diverse krumspring har forsøgt). Så hvis man ønsker at anerkende autonome begrundelser og værdier som et gyldigt, etisk perspektiv, må man acceptere dem som en selvstændig klasse af etiske begrundelser og værdier. Det er også klart, at de autonome værdier og begrundelser indimellem vil være i modstrid med agentneutrale begrundelser og værdier; Du har lyst til at løbe en maraton, men verden ville være et bedre sted, hvis du i stedet brugte din energi på fx at hjælpe folk i nød. I det omfang man giver agentrelative, autonome begrundelser forrang fremfor agentneutrale begrundelser, kommer de autonome begrundelser til at sætte en grænse for, hvor meget hensynet til fællesskabets bedste kan kræve af dig som individ. Det er dog uklart, præcis hvor denne grænse går, og en væsentlig del af det etiske arbejde i en konkret situation vil bestå i at forhandle grænsen.

365

370

Skellet mellem agentneutrale og agentrelative værdier gør det dog klart, at der er en tydelig asymmetri: En agentneutral værdi kan bruges til at gøre et krav på andre individer, men det kan en agentrelativ ikke. Hvis du sulter eller er i smerte, kan du have en agentneutralt begrundet forventning om, at andre vil hjælpe dig, også selv om det måske vil gå ud over deres egne agentrelative livsprojekter, men du kan ikke på samme måde gøre krav på, at andre skal hjælpe dig med at løbe en maraton; Jeg kan etisk set være forpligtet på

380

at give dig mad, hvis du er sulten, men jeg er ikke på samme måde etiske set forpligtet til at købe løbesko til dig, så du kan træne din maraton — det er dit eget projekt, der kun har værdi relativt til dit perspektiv. Agentneutrale værdier kan derfor (i visse tilfælde) trumfe et individs autonome værdier, men det kan aldrig gå den anden vej. Så selv om vi stadig mangler at finde ud af, præcis hvornår agentneutrale værdier har forrang for de autonome, er vi i det mindste blevet lidt klogere.

Hvor de autonome værdier giver individets egne projekter en vis beskyttelse mod krav rejst ud fra hensynet til det fælles bedste, fungerer de *deontologiske værdier og begrundelser* som begrænsninger og forbud i forhold til både autonome og agentneutrale værdier. Hvis jeg skal respektere andres autonomi, er der grænser for, hvad jeg kan gøre imod dem (som KANT påpegede). Jeg kan ikke tillade mig at lyve for folk eller på anden vis behandle dem på måder, der går ud over deres integritet. De deontologiske værdier tager udgangspunkt i en konkret relation mellem mig og et andet menneske (eller en gruppe af andre mennesker), og ikke i et ønske om at gøre verden som helhed til et bedre eller lykkeligere sted. Derfor de deontologiske værdier også agentrelative, og de adskiller sig fra de agentneutrale (dvs. utilitaristiske) værdier og begrundelser. Som vi har set i diskussionen mellem utilitarisme og deontologi, er deontologiske værdier og begrundelser langt fra altid i overensstemmelse med agentneutrale, utilitaristiske værdier; der kan opstilles scenarier, hvor hensynet til det fælles bedste kræver, at vi på den ene eller anden måde overskrider andre menneskers autonomi. Deontologiske værdier kan altså ikke reduceres til agentneutrale, og de deontologiske værdier udgør dermed et selvstændigt tredje niveau af etiske værdier og begrundelser, der giver en begrænsning både i forhold til mine egne autonome værdier og i forhold til agentneutrale værdier, der søger at fremme almenvellet. Spørgsmålet er naturligvis, i hvilken grad den deontologiske begrænsning skal honoreres. NAGEL mener tilsyneladende, at begrænsningen skal forstås som absolut (dvs. at man aldrig må overskride deontologiske forpligtelser), men det kan naturligvis diskuteres.

Opsamlende kan man sige, at agentneutrale (utilitaristiske) værdier udspringer af et tredjepersonsperspektiv, hvor man ser på verden udefra som en neutral tilskuer. De agentrelative, deontologiske værdier udspringer af et andenpersonsperspektiv, der tager udgangspunkt i de relationer, mennesker kan have med hinanden, mens de agentrelative, autonome værdier tager udgangspunkt i et førstepersonsperspektiv, der er bundet til den enkeltes individuelle projekter. Alle de tre perspektiver er væsentlige i en nuanceret etisk vurdering af en situation, og ingen af dem kan reduceres til de andre. Første- og andenpersonsperspektiv gør det ikke muligt se almenvellets interesser tydeligt, men et rent tredjepersonsperspektiv, som det, utilitarismen indfører, overser også noget væsentligt: Der er forskel på, om noget sker for nogen, eller om det sker for mig, og der er forskel på, om noget bare sker, eller om det er mig, der får det til at ske. Alle tre perspektiver er dermed nødvendige og (hvis man accepterer tankegangen ovenfor) legitime.

## 10.5 Analyser i praksis

Både utilitarismen og KANTS deontologi kan give nogenlunde klare svar på, hvordan man (ifølge dem) skal handle i en given situation. NAGELS beskrivelse af de forskellige etiske værdier giver ikke den slags klare svar, og det er heller ikke pointen. Pointen er snarere, at man med udgangspunkt i en forståelse af de forskelligartede værdier, der er i spil i en given situation, kan give en mere nuanceret etisk analyse og vurdering. Der findes ikke nogen simple algoritme for, hvordan de etiske valg skal træffes, men man kan, med baggrund i

NAGEL, give en skabelon for, hvordan et etisk problem kan adresseres mere nuanceret. Vi vil her gengive en skabelon, der er bearbejdet og forkortet på grundlag af Christiansen (2015, s. 41):

- Identificér hvori det etiske dilemma består: Mellem hvilke grundlæggende etiske intuitioner er der konflikt? 430
- Hvilke etiske grundværdier er de etiske intuitioner, der er i konflikt med hinanden, baseret på (autonome værdier, deontologiske værdier, utilitaristiske værdier)? Siger det noget om, hvordan de forskellige intuitioner skal vægtes i forhold til hinanden?
- Hvilke interessenter er involveret i konflikten, og hvilke interesser har de hver især (1., 2. og 3. personsperspektiver)? 435
- Er der ikke-etiske perspektiver, fx konkret empirisk viden, der kan kaste lys over situationen og styrke den etiske analyse?
- Er der løsninger, hvor de etiske intuitioner, der er i konflikt med hinanden, kan respekteres? Og hvis ikke: Hvordan er det nødvendigt at modificere de etiske intuitioner du startede med? Kan du leve med de modifikationer i andre situationer? 440
- Beskriv dit løsningsforslag så konkret som muligt.

For at illustrere, hvordan en etisk analyse kan se ud i praksis, kan vi kort se på en sag fra 2015, hvor en datalogistuderende skrev speciale i samarbejde med Macom, firmaet bag læringsplatformen *Lectio*, som de fleste danske gymnasier bruger (Behrendt, 2014). Den studerende brugte data fra *Lectio* om godt 70.000 gymnasieelever og byggede ved hjælp af machine learning en model, der med 93% sikkerhed kunne forudsige, om en given elev ville droppe ud af gymnasiet inden for de næste tre måneder. Modellen blev kortvarigt implementeret som en funktion i *Lectio*, men blev fjernet efter en uge efter krav fra Danske Gymnasier (Møllerhøj, 2015). 445

Vi vil ikke lave en egentlig analyse af sagen, men blot illustrere, hvordan NAGELS teori kan bringes i spil i den etiske undersøgelse af en konkret sag. En etisk analyse af situationen kunne som deontologiske perspektiver fx inddrage det forhold, at de elever, hvis data modellen blev bygget på, ikke havde givet samtykke til, at deres data blev brugt til at lave en model af den slags. Man kunne også inddrage det forhold, at eleverne ikke fik del i den økonomiske gevinst, implementeringen af modellen kunne have givet Macom (hvis modellen ikke var blevet fjernet) og det, at eleverne ikke havde noget reelt valg i forhold til at benytte *Lectio*. Som et utilitaristisk perspektiv kunne en etisk analyse overveje modellens troværdighed og rækkevidde (ikke mindst i lyset af induktionsproblemet og de forskellige andre metodologiske problemer machine learning-modeller generelt står over for, se kapitel 5). Fra et utilitaristisk perspektiv ville det dog også være relevant at gøre sig nogle grundige overvejelser om de forudsebare konsekvenser indførelsen af en model vil have på elever og læreres adfærd. Her vil det muligvis være relevant at indhente ny viden fx gennem pilotstudier eller interviews med lærere og elever, så man ikke er nødt til apriori at antage, at en implementering af modellen vil fremme nytten for de involverede parter. Udover de 2.- og 3.-personsperspektiver, som deontologien og utilitarismen leverer, har de involverede parter desuden haft deres egne 1.-personsperspektiv på sagen, hvor udviklingen af modellen på forskellige vis har spillet en rolle i deres egne, individuelle livsprojekter. Det overordnede spørgsmål er nu, hvordan de forskellige perspektiver skal vægtes i forhold til hinanden, 465

og om det med passende justeringer (fx brug af informeret samtykke, bedre afdækning af modellens konsekvenser) vil være muligt i højere grad at tilgodese flere af de konkurrerende perspektiver.

## 10.6 Big data-etik og privacy

Som et andet eksempel på hvordan flere etiske perspektiver kan bringes i spil, kan vi se mere principielt på den etiske diskussion af brugen af big data. I et review identificerede de to etikere BRENT MITTELSTADT og LUCIANO FLORIDI følgende fem dominerende temaer i litteraturen om big data-etik (Mittelstadt og Floridi, 2016, 311ff):

1. *Informeret samtykke*, dvs. spørgsmålet om, hvorvidt folk giver reel tilladelse til, at deres data bliver brugt til de formål, de bliver brugt til.
2. *Privacy*, dvs. spørgsmålet om hvorvidt analyse af big data kan overskride retten til privatliv fx ved at udlede følsomme oplysninger fra ikke-følsomme.
3. *Ejerskab*, dvs. spørgsmålet om, hvem der ejer og har råderet over folks data (og hvorvidt folk bliver tilstrækkeligt kompenseret for den værdi, deres data besidder).
4. *Epistemologi*, dvs. spørgsmålet om troværdigheden af de konklusioner, man når frem til via en given big data-analyse.
5. „*Big data-skellet*“, dvs. den asymmetri i både magt og viden, der er omkring brugen af big data (et firma som Facebook har en langt bedre forståelse af, hvad data kan bruges til, end den typiske bruger og kan (til en vis grænse) diktere brugerbetingelserne, da man ikke kan sige nej til dem).

Af disse repræsenterer tema 1, 2, 3 og 5 klare deontologiske perspektiver, der i grove træk udtrykker bekymring for, om folks autonomi og ret til selvbestemmelse bliver respekteret (1 og 2) og for, om brugen af big data fører til, at man stjæler fra eller narre folk (3 og 5). Tema 4 adresserer derimod nytteværdien af big data-analyser og repræsenterer derfor et utilitaristisk (agent-neutralt) perspektiv. Vi vil her nøjes at gå i dybden med spørgsmålet om privacy, dvs. retten til privatliv.

Spørgsmålet om privatliv er ikke nyt, men det er i meget høj grad blevet aktualiseret af udviklingen af ny informationsteknologi, idet vi (både bevidst og ubevidst) efterlader os utallige digitale spor, der meget let kan samles, deles og samkøres, og fordi machine learning og andre former for statistisk analyse gør det muligt at prediktere data, der tilhører andre kategorier af personoplysninger end de data, der er tilgængelige. Det sidste er fx det der sker, hvis man bruger folks indkøbsvaner til at prediktere om de er gravide, eller folks Facebook-adfærd til at prediktere deres personlighedstype (Youyou, Kosinski og Stillwell, 2015).

Til at starte med kan vi overveje, hvorfor privatliv overhovedet er værd at beskytte. Et svar går på, at vi har et ønske om selv at kunne styre, hvem der ved hvad om os (det kaldes *kontekstuel integritet*). Når vi fortæller, hvad vi lavede i weekenden, vil vi måske fortælle nogle detaljer til vores mor, andre til vores arbejdskolleger eller studiekammerater og andre igen til vores kæreste. Vi opbygger forskellige narrativer om os selv i forskellige sociale kontekster, og retten til privatliv sikrer netop vores mulighed for at udfolde vores autonomi på det område. Problemstillingen bliver forstærket af, at visse typer af oplysninger

direkte kan skade én på forskellig vis, hvis de falder i de forkerte hænder; I en del lande er 510  
homoseksualitet fx belagt med dødsstraf og tilsvarende kan forskellige former for politiske  
aktiviteter og religiøse tilhørsforhold være en trussel mod ens liv og sikkerhed. Som et lidt  
mere fredsommeligt eksempel kan man også have gode grunde til ikke at fortælle sin chef,  
at man har været til jobsamtale i det konkurrerende firma, og man kan også have gode 515  
grunde til ikke at offentliggøre på Facebook hvilken julegave, man har købt til sin kone  
(som Facebook kom til at gøre det en overgang i det såkaldte *Beacon-program*, hvor indkøb  
fra visse butikker automatisk blev delt med alle ens facebookvenner, se Nakashima, 2007).  
Privacy er således væsentligt fra et 1.- og 2.-personsperspektiv, dvs. for at vi kan udøve  
vores autonomi og følge egne autonome projekter uden at blive reduceret til et middel for  
andres mål. Argumentet, at man ikke har brug for privatliv, hvis man ikke har noget at 520  
skjule, er dermed grundlæggende forfejlet: Alene det at kunne bestemme, hvad man deler  
hvornår og med hvem er en grundlæggende del af det at udøve sin autonomi.

Retten til privatliv er primært blevet udfordret af kommercielle hensyn og af hensynet til  
effektivitet og sikkerhed. I forhold til de kommercielle hensyn er det bemærkelsesværdigt, at  
folk i så høj grad har været villige til frivilligt at afgive persondata til private virksomheder 525  
som Google og Facebook. Spørgsmålet er selvfølgelig, hvor frivillig afgivelsen af data reelt  
har været, og specielt om folk forstår og er tilstrækkeligt oplyste om, hvilke informationer,  
de egentligt afgiver (jf. MITTELSTADT og FLORIDIS tema 1 og 2 ovenfor). Tilsvarende kan  
man hævde, at tjenester som Google Translate og Google Maps bliver mere effektive, jo  
flere data vi overgiver til dem — og hvem vil ikke gerne have adgang til gode oversættelser 530  
og trafikinformation om bilkøer mv. på ens typiske vej til arbejde. Så også effektivitet kan i  
visse tilfælde stå i modsætning til ønsket om privatliv. Spørgsmålet er selvfølgelig, om den  
type tjenester bliver mere effektive til at hjælpe dem, der afgiver data, eller om de private  
data blot er et middel til at opnå andre (kommercielle) mål. Fra et kantiansk synspunkt er  
der en afgørende forskel på de to situationer pga. det praktiske imperativ. 535

Det er dog nok i forhold til spørgsmålet om sikkerhed, at den etisk set mest betydnings-  
fulde og vanskeligste konflikt i forhold til privatliv opstår. Her er det typiske argument, at  
kriminelle har lettere spil, hvis de kan skjule sig i bag privatlivets fred, og derfor bliver vi  
alle — for vores egen sikkerheds skyld — nødt til at acceptere et vist niveau af overvågning.  
Hvis man fortolker sikkerhed til også at omfatte sundhed, ser man den samme konflikt: Hvis 540  
et system som Google FluTrends, der skulle forudsige influenza-epidemier (se Lazer m.fl.,  
2014), rent faktisk havde fungeret, ville der have været et klart argument for, at vi alle skul-  
le overgive vores søgedata (og måske også andre private data, som køn, alder, arbejdssted,  
sociale kontakter mv.) til Google, så modellen kunne fungere bedst muligt og dermed gøre  
verden til et sikrere og mere influenzafrit sted. 545

Der er dermed en række klare og vanskelige konflikter mellem privatliv og andre hensyn.

I forhold til den kommercielle udnyttelse af private oplysninger er det måske ikke så  
meget en konflikt, men mere et spørgsmål om, hvorvidt virksomhederne respekterer grund-  
læggende deontologiske forpligtelser. I forhold til spørgsmålet om sikkerhed, er konflikten  
mere reel og udspiller sig som nok et klart eksempel på den klassiske konflikt mellem deon- 550  
tologiske og utilitaristiske hensyn. Her antyder historien (uden at vi vil gå i dybden med  
det), at det er let at hylde de deontologiske principper, når alt går godt, men så snart der  
viser sig en trussel — fx terrorangreb eller en pandemi — har de utilitaristiske hensyn det  
med at få forrang.

Af samme grund er det, som etikeren DEBORAH JOHNSON påpeger, væsentligt at gøre 555  
sig klart, at den rent deontologisk funderede begrundelse, vi ovenfor gav for privatliv, fak-  
tisk ikke er fuldstændig (Johnson, 2009, 95ff). Privatliv er ikke blot et personligt gode, men

også et socialt gode, idet den grundlæggende autonomi og integritet, beskyttelse af privat-  
 livet muliggør, ifølge JOHNSON er en afgørende forudsætning for at have et demokrati. Et  
 560 demokrati kan kun fungere, hvis der er plads til at tænke anderledes og mulighed for at gøre  
 og afprøve nye ting, som de siddende magthaverne måske ikke billiger. Uden en privatsfære,  
 hvor man kan tage selvstændig stilling til og indgå med andre borgere i en kritisk dialog om  
 magthavernes beslutninger kan man ikke have et demokrati. Hvis det argument er korrekt,  
 er spørgsmålet om privatliv vs. sikkerhed ikke blot et spørgsmål om individuelle rettigheder  
 565 mod almenvellets velfærd, men også et spørgsmål om forskellige former for generel velfærd.  
 Sat på spidsen kan man sige, at valgte også står mellem den sikkerhed man kan få med  
 totalitær overvågning og de fordele, et oplyst, inddragende demokratisk styre giver. Man  
 kan i længden ikke have begge dele.

## 10.7 Bias og algoritmisk transparens

570 Brugen af algoritmer giver også en anden etisk konflikt, nemlig en konflikt mellem effek-  
 tivitet og retfærdighed. Det har i århundrede været helt almindeligt at bruge datadrevne  
 sandsynlighedsberegninger til at støtte beslutninger, der går ud på at vurdere en form for  
 risiko. Hvis man fx tegner en ulykkesforsikring, vil forsikringsselskabet bruge deres viden  
 om, hvor hyppigt folk, der på relevante træk ligner dig, kommer til skade, til at fastsætte  
 575 forsikringspræmien. Det er ret ukontroversielt. Det er straks mere kontroversielt, at man  
 bl.a. i det amerikanske retsvæsen siden starten af den 20. århundrede har brugt lignende  
 beregninger til at vurdere kriminelles risiko for at begå ny kriminalitet – og at den type  
 beregninger er blevet brugt i vurderingen strafudmåling, prøveløsladelse mv. (Carlson, A.  
 (2017). The Need for Transparency in the Age of Predictive Sentencing Algorithms. Iowa  
 580 Law Review, 103(1), 303–329.). I de sidste årtier har lettere adgang til store datamængder  
 og fremkomsten af diverse machine learning teknikker gjort det lettere at udvikle den type  
 systemer og diverse former for prædiktive algoritmer bruges i dag på en lang række områder  
 fra reklamer til sortering af jobansøgninger.

Det er dog ikke helt uproblematisk at bruge prædiktive algoritmer. Som vi så tidligere, er  
 585 der en række epistemiske problemer forbundet med machine learning. Der er typisk en meget  
 direkte sammenhæng mellem epistemiske og etiske overvejelser; hvis de etiske overvejelser  
 inddrager utilitaristiske komponenter er det i høj grad væsentligt at forstå, hvor godt en  
 given algoritme i praksis virker. Ville den etiske vurdering af algoritmen til forudsigelse af  
 frafald i gymnasiet fx falde anderledes ud, hvis systemet kun havde en træfsikkerhed på  
 590 67%? Der er imidlertid også et anden og vanskeligere etisk aspekt specielt hvis man bruger  
 algoritmiske forudsigelser til at træffe afgørelser, der har betydning for enkeltindviders  
 muligheder. Algoritmen til forudsigelse af frafald inddrog således direkte elevernes etnicitet  
 som en faktor i beregningen, men hvad nu hvis algoritmen havde en bias så den systematisk  
 gav elever med en etnicitet en højere risikovurdering end elever med en anden?

595 Spørgsmålet om diskrimination i algoritmer er navnlig blevet diskuteret i forbindelse  
 med den såkaldte COMPAS-algoritme, der bruges i det amerikanske retsvæsen til vurdering  
 af kriminelles tilbagefaldsrisiko. En gennemgang af algoritmen viste imidlertid en systema-  
 tisk skævhed i de fejl, algoritmen lavede. Således havde farvede, der ikke senere begik ny  
 kriminalitet, en langt højere chance for fejlagtigt at få en høj risikoscore end hvide, og hvide  
 600 havde omvendt en højere chance for fejlagtigt at få en lav score end farvede (Angwin m.fl.,  
 2016).

Hvis man imidlertid tager udgangspunkt i de to kategorier høj- og lav-risiko, og under-

søger, hvor god algoritmen er til at komme med korrekte forudsigelser, er der paradoksalt nok ingen forskel på farvede og hvide; blandt de, der blev bedømt som højrisiko, var procentdelen af farvede, der var blevet fejlplaceret, nogenlunde lige så stor som procentdelen af hvide, og tilsvarende for de, der blev vurderet som højrisiko. Det betyder med andre ord, at risikoscoren er nogenlunde lige pålidelig, uanset hvilken etnicitet den person, den bliver lavet for, har. Ydermere har det vist sig, at man ikke kan have lighed for begge mål samtidig. Enten vil algoritmen som nu være lige pålidelig i sine forudsigelser uanset etnicitet, men til gengæld have en systematisk skævhed i forhold til enkeltpersoners risiko for at blive fejlplaceret, eller også kan man gøre risikoen for at blive fejlplaceret lige stor uanset etnicitet, hvilket til gengæld betyder, at algoritmen vil være mindre pålidelig for den ene etniske gruppe end for den anden (for en stringent matematiske gennemgang, se Kleinberg, Mullainathan og Raghavan, 2017). Det er selvfølgelig en interessant etisk udfordring!

Når man skal diskutere hvorvidt brugen af prædiktive algoritmer kan forsvares etisk er der er en række ting, det kan være godt at få på plads. For det første er de etiske argumenter for at bruge prædiktive algoritmer typisk utilitaristiske; den grundlæggende påstand er, at en algoritme som COMPAS på den eller anden måde øger den samlede nytte i samfundet og det er på den baggrund algoritmen skal vurderes (der kan selvfølgelig også være ikke-etiske begrundelser som kommercielle interesser eller andet).

For andet skal være klar over, at prædiktive algoritmer altid diskriminerer, i den forstand at din fremtidige adfærd estimeres på baggrund af den fortidige adfærd af den gruppe, som du anses for at tilhøre. Hvis man fx vil estimere hvor længe en 25 årig, ikke-rygende, københavnsk studerende lever, vil man se på, hvor længe folk i en referenceklasse af personer, der minder om vores 25-årige studerende i gennemsnit har levet. Det er ganske simpelt sådan, statistiske forudsigelser fungerer (i det mindste i den frekventistisk tolkning af sandsynlighed, som er mest naturlig i denne sammenhæng). Spørgsmålet er altså ikke om algoritmerne må diskriminere, men i højere grad hvilke faktorer, vi vil tillade, at der indgår i referenceklassen. Hvilke træk ved dig, kan man det etisk set forsvare at diskriminere på baggrund af, når du skal have en livsforsikring, ansøger om et job eller søger ind på en uddannelse?

For at svare på det spørgsmål bliver vi for det tredje nødt til at inddrage begrebet retfærdighed. Begrebet retfærdighed befinder sig et sted imellem etik og politisk filosofi, idet det angår mere generelle egenskaber samfund, som fordeling af goder, adgang til politisk magt og det juridiske straffesystem. Præcis som det er tilfældet for etik findes der mange forskellige retninger indenfor politisk filosofi, og mange forskellige ideer om, hvad retfærdighed er. Det vil vi på ingen måde gå i dybden med her. Vi vil nøjes med at præsentere en enkelt og meget operationel forestilling om retfærdighed, til dels med udgangspunkt i den amerikanske filosof JOHN RAWLS (1921–2002). RAWLS hævder (Rawls, 1972), at hvis samfundet skal være retfærdigt, skal vi tage alle grundlæggende beslutninger om det indretning og institutioner, mens vi befinder os bag et *uvidenhedens slør*, der skjuler, hvilken position vi selv vil få i samfundet. Så man skal altså forestille sig, at man ikke ved, hvilket køn man har, at man ikke ved, om man er rig eller fattig, at man ikke kender sin etnicitet mv., og ud fra den position kan man beslutte, hvordan samfundet skal indrettes. I dette tilfælde: Hvilke former for diskrimination vil vi tillade i prædiktive algoritmer? Det er let at forestille sig (men det kan selvfølgelig diskuteres) at man i den situation vil vælge et samfund der er uden kønsdiskrimination og ude diskrimination mode fx seksuelle og etniske minoriteter. Så simpelt er det.

Og så alligevel ikke, for der kan være principielt forskellige former for diskrimination. I nogle tilfælde diskriminerer man ud fra en faktor, der slet ikke har nogen betydning for det udkomme, man er interesseret i at optimere. Fx ved man, at videnskabelige artikler har

letterer ved at bliver publiceret, hvis der bruges et mandligt forfatternavn end hvis der bruges et kvindeligt. Da det udelukkende er artiklens kvalitet og ikke forfatterens køn der skal afgøre, om en artikel er publicerbar eller ej, er der her tale om direkte (køns)diskrimination. I andre tilfælde er sagen imidlertid mindre klar. Fx lever mænd kortere end kvinder (en  
 655 tommelfingerregel siger, at det er lige så farligt at være mand som at være overvægtig). Så hvis man tegner en livsforsikring, burde mænds præmier være højere end kvinders. Hvis præmien er ens for begge køn kommer de i gennemsnit fornuftige kvinder til at betale for de i gennemsnit ufornuftige mænds dumheder. Man kan sige, at køn i forhold til livsforsikring er en informationsbærende dimension, så hvis man afskærer algoritmen fra at bruge din  
 660 dimension, vil den naturligvis blive mindre effektiv. Og i dette tilfælde er det kvinderne, der vil komme til at betale for den manglende effektivitet. Hvis man omvendt inddrager køn i beregningen af pensionspræmie, vil en mand, der opfører sig fornuftigt og tager sit helbred seriøst, komme til at betale en højere præmie, udelukkende fordi han har samme køn som en gruppe, der – gennemsnitligt – opfører sig ufornuftigt. Og hvad gør man så? Er  
 665 det mest retfærdige at kønsdiskriminere i dette tilfælde (og lade de fornuftige mænd betale prisen) eller er det mest retfærdigt at undlade at kønsdiskriminere (og lade kvinderne betale prisen)?

Det bringer os tilbage til COMPAS-algoritmen. I den population, hvor algoritmen er blevet testet, har farvede fanger reelt en større risiko for at begå ny kriminalitet. Når  
 670 algoritmen diskriminerer på baggrund af etnicitet er det derfor blot et udtryk for, at den har identificeret en informationsbærende dimension i datasættet. Og lige som ovenfor vil vi have valget imellem at forhindre algoritmen i at bruge informationen, hvorved den vil blive mindre effektiv i den forstand at dens overordnede fejlrate vil stige, eller at tillade den at bruge informationen, hvorved den vil diskriminere enkeltindivider, således at farvede, der  
 675 ikke vil begå ny kriminalitet har en meget større risiko for at få en høj kriminalitetsscore end ikke-farvede. Hvilket af de to scenarier forekommer dig mest retfærdigt, når du står bag uvidenhedens slør?

Der er dog også væsentlige forskelle på de to cases. Specielt kan man argumentere for, at den gennemsnitligt høje kriminalitetsrate blandt farvede til dels skyldes strukturelle forhold  
 680 i det amerikanske samfund; pga. en generel diskrimination mod farvede har de sværere ved at få uddannelser og job end ikke-farvede, og derfor vil de hyppigere ende i kriminalitet. En algoritme som COMPAS, der har en større tilbøjelighed til at holde farvede i fængsel end ikke-farvede, vil bidrage til fastholde denne strukturelle ulighed. En datadreven algoritme afspejler virkeligheden, som den er, men hvis virkeligheden er racistisk, vil algoritmen også  
 685 blive det, og hvis algoritmen bliver brugt uden forbehold, kan den bidrage til at fastholde historisk betingede diskriminerende strukturer. Der er derimod ikke noget, der tyder på, at mænds kortere levealder er et udtryk for strukturel diskrimination. Så med andre ord er det vigtigt, at se på den kontekst, en algoritme skal indgå i, når man overvejer, om den er retfærdig.

## 690 **Pandoras black box**

I en traditionel statistisk model vil man have nogenlunde styr på de variable, modellen inddrager. Det er dog værd at bemærke, at tingene selv med traditionelle algoritmer ikke altid er så klare; COMPAS-algoritmen inddrager således ikke direkte etnicitet som en variabel. Den inddrager til gengæld flere variable, der er tæt korrelerede med etnicitet, og  
 695 dermed indgår etnicitet indirekte i algoritmen. Det gør naturligvis tingene mere besværlige,



at man på den måde kan komme til at bruge variable, man ikke direkte har indbygget i sin algoritme.

Det problem bliver imidlertid kraftigt forstærket, hvis man træner en algoritme med machine learning. Her har vi styr på de data, algoritmen er trænet med, og vi kan opstille forskellige mål for, hvor godt den virker, men resten er en black box; vi aner reelt ikke hvorfor den virker, og som vi berørte ovenfor, kan machine learning algoritmer med lethed diskriminere langs variable, der er af en helt anden type end de data, vi har trænet algoritmen med. Hvis vi vil sikre, at de algoritmer, vi udvikler, ikke diskriminerer på en etisk kritisabel måde, bliver vi nødt til at gøre dem transparente. Det er imidlertid ikke altid let, at gøre en blackbox transparent!

## 10.8 Systemisk etik

Efter at have set på de normative teorier vil vi nu vende tilbage til den deskriptive etik og beskrive et interessant og i høj grad overset aspekt af etikken, nemlig det tilsyneladende empiriske faktum, at vores etiske beslutninger kan påvirkes af selv små ændringer i vores miljø. Som et enkelt eksempel testede JOHN M. DARLEY (1938–2018) og DANIEL BATSON teologistuderende i et eksperimentelt paradigme, hvor en forsøgsperson (typisk en studerende) bliver bedt om at forberede et kort oplæg om den barmhjertige samaritaner, som er en lignelse fra Det Nye Testamente, der i korte træk går ud på, at man ubetinget skal hjælpe andre i nød, også selv om det er en vildt fremmet af et andet folk (Darley og Batson, 1973). Forsøgspersonen skulle herefter gå hen til et andet lokale for at holde sit oplæg. På vejen kom forsøgspersonen forbi en skuespiller, der, forklædt som hjemløs, fingerede et ildebefindende. Eksperimentet gik nu ud på at undersøge, hvor mange af forsøgspersonerne, der faktisk stoppede op for at hjælpe den fremmede i nød. Resultatet var ikke så ringe — i den første setup forsøgte 63% af forsøgspersonerne at hjælpe den hjemløse. I den anden variant af forsøget lavede man imidlertid én lille ændring: Netop som forsøgspersonen gik ud af forberedelseslokalet kiggede forsøgslederen på sit ur og fortalte forsøgspersonen, at vedkommende havde travlt. Det fik antallet af forsøgspersoner, der standsede op, til at falde til 10%. Med andre ord viser eksperimentet, at selv små ændringer i vores miljø kan have en stor indflydelse på de moralske valg, vi tager, her om man skal hjælpe eller ikke-hjælpe en fremmed i nød.

Man skal absolut ikke lægge for meget i et enkelt adfærdspsykologisk eksperiment (med lav statistisk kraft) som det ovenstående. Eksperimenter er primært en illustration. Det slående er snarere, at effekten tilsyneladende er robust og kan gentages i forskellige varianter og eksperimentelle setups (se fx Pascual-Ezama m.fl., 2015).

Det er ikke helt oplagt, hvordan man skal fortolke den slags resultater fra et etisk synspunkt. Undersøgelser af, hvordan folk faktisk handler siger jo, som nævnt i indledningen, ikke noget om, hvordan vi bør handle. Og så alligevel. Fra en overfladisk betragtning kan man hævde, at kontekstafhængigheden af vores moralske valg mindsker vores personlige etiske ansvar: Hvis vores etiske beslutninger styres af konteksten, er det jo ikke rigtig vores skyld alligevel. Og dog. Man kan lige så vel sige, at den type resultater gør os opmærksom på et nyt og anderledes etisk ansvar, nemlig vores medansvar for, at den kontekst, vi befinder os i, tilskynder os til at opføre os etisk korrekt og ikke omvendt. Med andre ord kan man hævde, at man har et ansvar for at være opmærksom på, om de strukturer og den kultur, der er i den virksomhed eller det fagfællesskab, man befinder sig i, på den ene eller anden måde tilskynder til tvivlsom eller uetisk adfærd. Og hvis det er tilfældet, må man forsøge at ændre

kulturen eller flytte sig. For at vende tilbage til figur 1 ovenfor, må man med andre ord have en opmærksomhed på, om den lokale praksis, man befinder sig i, er i overensstemmelse med, hvad der er etisk korrekt.

I en forskningsetisk sammenhæng er det interessant, at man i den *code of conduct*, som alle danske universiteter har ratificeret, til dels har forladt det entydige fokus på individets ansvar (*Danish Code of Conduct for Research Integrity* 2014). Således betones det, at det ikke kun er individuelle forskere, der har et ansvar for at optræde ordentligt, men at institutionerne også har et ansvar for at skabe strukturer og arbejdsbetingelser, der gør det muligt for forskerne at optræde etisk og redeligt. Spørgsmålet er så selvfølgelig, i hvilket omfang de strukturer er på plads, hvilket peger hen på temaet om redelighed og forskningspolitik.

## 10.9 IT-professionelles etiske ansvar

Foreløbig har vi set på, hvordan man kan håndtere etiske valgsituationer, men vi har endnu ikke undersøgt, om IT-professionelle i det hele taget har et etisk ansvar eller hvor langt det ansvar egentlig går.

I en artikel, der diskuterer forskeres ansvar for militære anvendelser af deres forskning, behandlede den danske etiker JESPER RYBERG en række af de undskyldninger, forskere typisk kommer med, for at unddrage sig et etisk ansvar for deres forskning (Ryberg, 2003). Den første undskyldning går på, at staten håndterer moralen, idet staten gennem lovgivning sætter grænser for, hvad man må og hvad man ikke må. Som vi så ovenfor, holder denne undskyldning ikke, specielt for områder med hastig teknologiudvikling, da undskyldningen forudsætter, at lovgiverne både har perfekt moralsk og teknisk indsigt. En anden mulig undskyldning påpeger, at det ligger i videnskabens natur, at man hverken kender udfaldet af et eksperiment eller kan forudsige dets anvendelser. Det er korrekt, at man ikke altid ved, hvor man ender, når man bedriver videnskab, men i mange tilfælde har man faktisk klare mål og en klar forventning om, hvad der kommer ud af et givet forskningsprojekt. Og hvad angår anvendelserne, kan man som minimum gøre en fokuseret indsats for at overveje, hvad den teknologi, man udvikler, faktisk vil kunne bruges til. Så den undskyldning har også begrænset rækkevidde. En tredje undskyldning går på, at det jo ikke er os som teknologiudviklere, der bestemmer, hvad resultaterne af vores forskning skal bruges til. Vi gør det kun muligt. Den undskyldning fungerer heller ikke rigtigt. Hvis vi husker på casen med KANTS udspørgende morder, var problemet jo netop, at vi ikke bare kan hjælpe andre med at udføre en handling, vi ved er etisk forkert, uden selv at pådrage os et vist ansvar. Så med mindre man er parat til at blive en stålsat kantianer med alt, hvad deraf følger, virker 'vi gør det bare muligt'-undskyldningen ikke. Som en sidste undskyldning kan man hævde, at hvis ikke jeg havde gjort det, ville en anden bare have gjort det. Eller med andre ord, at man er *udskiftbar*. Den undskyldning fungerer dog heller ikke. Hvis man stjæler en pung, som nogen har glemt i bussen, vil man ikke nå langt ved at hævde, at hvis jeg ikke havde stjålet den, ville en anden have gjort det. Man tage ansvar for sin handling uanset hvad.

Der kan naturligvis være andre undskyldninger, men med mindre de fungerer bedre, end de fire, vi har set ovenfor, har man tilsyneladende som teknologiudvikler et ret udstrakt ansvar for konsekvenserne af den teknologi, man udvikler. Dominerende faglige foreninger som Institute of Electrical and Electronics Engineers (IEEE) og (specielt) Association for Computing Machinery (ACM) har da også meget klare etiske kodeks, der tydeligt tilskriver

foreningernes medlemmer et ganske udstrakt etisk ansvar. Ifølge IEEE's etiske kodeks accepterer medlemmerne således bl.a. „to hold paramount the safety, health, and welfare of the public, to strive to comply with ethical design and sustainable development practices, and to disclose promptly factors that might endanger the public or the environment“ (History of Computing, u.å.). Hos ACM finder man tilsvarende krav om, at IT-professionelle skal bruge deres viden og kunnen til samfundets bedste og til at fremme menneskerettigheder og individers autonomi (ACM, u.å.). Og tilsvarende punkter findes også i dansk sammenhæng implementeret i PROSA's etiske retningslinjer (PROSA, 2000). 790

Begge de amerikanske kodekser indeholder dog også paragraffer, der rækker ud over den enkeltes eget arbejde. ACM's kodeks rummer således et eksplicit krav om, at man skal 'blow the whistle' (ACM, u.å., sektion 1.2), hvis man ser tegn på, at et system kan gøre skade, og hos IEEE hedder det, at man skal „seek, accept, and offer honest criticism of technical work, to acknowledge and correct errors, and to credit properly the contributions of others“(History of Computing, u.å.). 795

Det er en interessant diskussion, hvor langt dette ansvar egentlig går. Man kan argumentere for, at man som IT-professionel sidder inde med en særlig viden, og at man af den grund har en særlig pligt til at reagere, hvis man bliver opmærksom på misbrug eller tvivlsom brug af computerteknologi, også selv om man ikke selv har været med til at designe de pågældende systemer. Med store kræfter kommer et stort ansvar! Det er dog også klart, at det kan have store omkostninger for den enkelte IT-professionelle at påtage sig det ansvar både i form af ekstraarbejde og repressalier, og derfor må den utilitaristisk funderede pligt til at handle afvejes i forhold til den enkeltes autonome projekter og mål. 800 805

## 10.10 Redelighed

Spørgsmålet om videnskabelig redelighed kom for alvor på dagsordenen i årene lige efter årtusindskiftet, hvor en række store skandaler viste, at videnskaben langt fra var så selv-korrigerende, som man hidtil havde troet, men at bevidst og grov uredelighed faktisk fandt sted. Det var skandaler, som sagen om den tyske faststoffysiker JAN HENDRIK SCHÖN, der en overgang publicerede omkring en artikel om ugen i de mest prestigefyldte tidsskrifter, indtil det viste sig, at mange af artiklerne var skrevet på baggrund af delvist konstruerede data. Eller sagen om den nederlandske psykolog DIEDERIK ALEXANDER STAPEL, der publicerede mere end 50 artikler på baggrund af frit opfunden data, og den koreanske læge HWANG WOO-SUK, der publicerede en række banebrydende artikler inden for stamcelleforskning, men også på baggrund af konstruerede data. I Danmark er den mest prominent sag til dato formentlig sagen om hjerneforskeren MILENA PENKOWA, der på linje med STAPEL og SCHÖN publicerede artikler på baggrund af konstrueret data fra forsøg, der aldrig havde fundet sted. 810 815 820

Nogenlunde samtidig med disse store skandaler blev man klar over, at direkte uredelighed måske slet ikke var det største problem i forhold til videnskabens integritet. En række store spørgeskemaundersøgelser af forskeres adfærd viste, at alvorlige overskridelser som plagiering og fabrikation af data kun forekommer sjældent (eller i hvert fald indrømmede kun få forskere i undersøgelserne at de havde gjort den slags), men til gengæld var mindre overskridelser af normerne forbavsende udbredte. I den mest berømte af undersøgelserne, der involverede mere end 3000 amerikanske forskere, indrømmede kun 0,3% af forskerne, at de havde forfalsket eller manipuleret med data, men 15,5% svarede, at de havde ændret på en undersøgelses metode eller resultat efter pres fra en sponsor, 15,3% svarede, at de 825 830

havde slettet afvigende data udelukkende ud fra en mavefornemmelse af, at datapunkterne var upræcise og 10,0% svarede, at de havde givet eller fået ufortjente forfatterskaber (Martinson, Anderson og Vries, 2005). Andre undersøgelser havde nogenlunde sammenlignelige resultater (Fanelli, 2009), og man begyndte at bruge betegnelsen *tvivlsom forskningspraksis* til at beskrive den type adfærd, der ligger i gråzonen mellem acceptabel forskningspraksis og direkte uredelighed.

I 1981 fortalte PHILIP HANDLER (1917–1981), der på det tidspunkt var præsident for det amerikanske National Academy of Sciences den amerikanske kongres, at videnskabelig snyd sjældent forekommer, og at det, hvis det forekommer, uundgåeligt bliver opdaget, fordi videnskaben er effektiv, demokratisk og selvkorregerende (Stroebe, Postmes og Spears, 2012, §677). 30 år efter stod det imidlertid klart, at den påstand ikke længere holdt vand. Videnskabelig uredelighed og tvivlsom forskningspraksis er virkelige og alvorlige fænomener, som videnskaben bliver nødt til at tage alvorligt, og som man kollektivt må finde metoder til at håndtere.

I Danmark nedsatte man allerede i 1998 tre stående udvalg, der kunne håndtere sager om uredelighed (BEK nr 933 af 15/12/1998). Udvalgene og de bekendtgørelser og lovtekster, der sætter rammerne for dem, har udviklet sig en del siden da, og udviklingen er et godt eksempel på, hvordan praksis, etik og jura gensidigt påvirker hinanden. Udvalgene blev i 2017 omdannet til et nævn, Nævnet for Videnskabelig Uredelighed (NVU), og definitionen af uredelighed blev ved samme lejlighed væsentligt revideret. I den nuværende lov beskriver man videnskabelig uredelighed som: 'Fabrikerings, forfalskning og plagiering, som er begået forsætligt eller groft uagtsomt ved planlægning, gennemførelse eller rapportering af forskning' (*Lov om videnskabelig uredelighed m.v.* 2017, §3.1), hvilket svarer til den typiske internationale definition, hvor uredelighed netop forstås som fabrikerings, forfalskning og plagiering (FFP). I den danske lov defineres de tre centrale begreber som:

Fabrikerings: Uoplyst konstruktion af data eller substitution med fiktive data.

Forfalskning: Manipulation af forskningsmateriale, udstyr eller processer samt ændring eller udeladelse af data eller resultater, hvorved forskning fremstår misvisende.

Plagiering: Tilegnelse af andres ideer, processer, resultater, tekst eller særlige begreber uden retmæssig kreditering (*Lov om videnskabelig uredelighed m.v.* 2017, §3.2–3.4).

Loven skelner desuden skarpt mellem uredelighed og tvivlsom forskningspraksis, hvor det sidste defineres som: 'Brud på alment anerkendte standarder for ansvarlig forskningspraksis, herunder standarderne i den danske kodeks for integritet i forskning og andre gældende institutionelle, nationale og internationale praksisser og retningslinjer for integritet i forskning.'

NVU tager sig udelukkende af sager om uredelighed, mens sager om tvivlsom forskningspraksis bliver håndteret af universiteternes praksisudvalg. Det kodeks for integritet, der henvises til i lovteksten, er kodekset *Danish Code of Conduct for Research Integrity*, der blev ratificeret af alle danske universiteter i 2014. Dette kodeks betoner specielt, hvordan man som forsker skal forsøge at leve op til de tre grundlæggende værdier *ærlighed*, *transparens* og *ansvarlighed*. Så hvor loven om videnskabelig uredelighed beskriver de ting, man som forsker absolut *ikke* må gøre, beskriver *Danish Code of Conduct for Research Integrity* et ideal, man så vidt muligt skal stræbe efter. De to tekster beskriver med andre ord hver sin ende af et spektrum for, hvordan man som forsker bør opføre sig.

Endelig skal det bemærkes, at NVU eksplicit *ikke* behandler spørgsmål om forskningskvaliteten af et videnskabeligt produkt (*Lov om videnskabelig uredelighed m.v.* 2017, §3, stk. 2). Spørgsmålet om kvalitet er uddelegeret til peer-reviewsystemet og de mekanismer til organiseret skepsis, det videnskabelige samfund i øvrigt har etableret. Der er med andre ord forskel på sjusk og fusk, og det er kun det sidste, NVU tager sig af. 880

I en fremsynet artikel beskrev den amerikanske sociolog STEPHEN HILGARTNER i 1990 fire forskellige mulige tiltag, man kunne gøre mod uredelighed, nemlig:

1. *Law enforcement*, dvs. afskrækkelse gennem hurtig og hård straf.
2. *Oversight*, dvs. mere kontrol med forskeren, fx gennem skærpede krav om opbevaring af forsøgsprotokoller og rådata. 885
3. Mere *oplysning og uddannelse* om god videnskabelig praksis, så folk ikke uforvarende kommer til begå uredelighed.
4. *Revision* af belønnings- og meriteringssystemerne i forskningsverdenen, så de i højere grad opfordrer til god praksis (Hilgartner, 1990, §2). 890

De første tre af disse strategier er tydeligt blevet bragt i spil. NVU er således et klart udtryk for *law enforcement*, både universiteter og tidsskrifter stiller i dag meget højere krav om publicering og opbevaring af rådata og forsøgsdesign end tidligere (*oversight*), og kurser i „Fagets Videnskabsteori“ er en styrkelse af uddannelsen i god videnskabelig praksis. Hvad angår den fjerde strategi er det til gengæld mere tvivlsomt om belønnings- og meriteringssystemet er blevet ændret, så de i højere grad opfordrer til god praksis. Specielt kan man diskutere, om overgangen til postakademiske forskning og det øgede fokus på forskningsproduktion hos de enkelte forskere fremmer god praksis eller det modsatte. 895

Vi er dermed tilbage ved spørgsmålet om systemisk etik og den rolle, de strukturer, vi befinder os i, spiller for vores moralske valg (afsnit 10.8). I hvilket omfang ligger ansvaret for uredelighed og tvivlsom forskningspraksis hos de individuelle forskere og i hvilken grad ligger ansvaret på de overordnede strukturer, man fra politisk og institutionelt hold har skabt til at bedrive forskning under? Mens loven om videnskabelig uredelighed udelukkende beskæftiger sig med forskernes individuelle ansvar, er det bemærkelsesværdigt, at *Danish Code of Conduct for Research Integrity* også tillægger forskningsinstitutionerne et vist (omend begrænset) ansvar. I forhold til publicering af forskningsresultater lyder det fx således (og der findes lignende paragraffer for andre aspekter af forskningsprocessen): 900

v. Institutions should promote and maintain an environment that supports honesty, transparency, and accuracy when disseminating research findings, e.g. through policies and training relating to publication and communication. 910

vi. Institutions should ensure that sponsors and other funders of research fully respect the duty of researchers to publish research and research results honestly, transparently, and accurately. (*Danish Code of Conduct for Research Integrity* 2014, §11)

Dermed er der åbnet for en diskussion, hvor hele ansvaret ikke pr. automatik placeres hos de individuelle forskere. 915

## God studiepraksis

Som studerende er man bl.a. underlagt regler om god studiepraksis, der i et vist omfang minder om de regler for uredelighed og god forskningspraksis, forskere er underlagt (ifølge  
 920 ordensreglerne på Københavns Universitet er man faktisk underlagt præcis de samme regler). Det er dog værd at bemærke, at der er nogle væsentlige forskelle mellem den situation, som hhv. forskere og studerende befinder sig i, specielt i forhold til plagiering og forfatterskab. Som bemærket ovenfor, er den officielle definition af plagiering faktisk ganske bred, og indeholder ikke bare tekstplagiering, men også plagiering af ideer, processer, resultater  
 925 og begreber. Selv om det kan være svært i praksis at bevise ideplagiering, er det bestemt noget, man tager alvorligt i forskningsverdenen; det er ikke velset at benytte andres ideer uden at kreditere dem.

For studerende ser situationen imidlertid lidt anderledes ud. En væsentlig del af meget undervisning består netop i at man tilegner sig og bruger ideer fra undervisere, undervisningsmateriale og medstuderende. Det ville ikke give mening, hvis man til eksamen skulle kreditere underviseren, hver gang man brugte en ide hun havde præsenteret til forelæsninger, eller hvis man skulle nævne sine medstuderende, hver gang man trak på en ide, de var kommet med i løbet af undervisningen. Overordnet set er de plagieringskrav, studerende er underlagt, dermed lempeligere end dem, der stilles til forskere. Der stilles dog højere  
 935 og højere krav til originalitet, jo længere man kommer på sin uddannelse, så dermed kan man sige, at universitetsstuderende er i en overgangsfase fra det simple plagieringsbegreb, men benytter i gymnasiet (hvor plagiering opfattes som synonym for tekstplagiering) til det mere omfattende plagieringsbegreb, man bruger i forskningsverdenen. Den slags overgange kan let skabe gråzoner og misforståelser, så derfor er det altid en god ide at spørge  
 940 sin underviser, hvis man er i tvivl om, hvordan plagieringsreglerne fortolkes på et specifikt kursus. Specielt skal det bemærkes, at der på nogle uddannelser er enkelte eksaminer, hvor de studerende under ingen omstændigheder må dele ideer med hinanden og hvor reglerne om ideplagiering håndhæves strengt.

Hvor plagieringsbegrebet generelt er mindre restriktivt for studerende end forskere, er kravene til forfatterskaber til gengæld strengere. Som vi var inde på i forbindelse med undersøgelsen af tvivlsom forskningspraksis (Martinson, Anderson og Vries, 2005), er det ikke ualmindeligt, selvom det er uærligt, at forskere angiver sig selv eller andre som forfattere på videnskabelige artikler, de egentlig ikke har bidraget noget videre til. Kravene til forfatterskab fremgår af den såkaldte Vancouver-protokol, hvor det bl.a. fremgår at man skal have  
 950 ydet et 'substantial contributions to 1) conception and design, or analysis and interpretation of data; and to 2) drafting the article or revising it critically for important intellectual content; and on 3) final approval of the version to be published' („Uniform Requirements for Manuscripts Submitted to Biomedical Journals“ 1997, §38). Forbavsende nok, betragtes brud på reglerne om forfatterskaber i almindelighed ikke som uredelighed, men kommer  
 955 under reglerne for tvivlsom forskningspraksis, hvor de dog tilsyneladende håndhæves særdeles lempeligt. Her er reglerne til de studerende til gengæld soleklare: Hvis man ikke har bidraget tilstrækkeligt til en gruppeopgave, må man ikke optræde som medforfatter. Det vil være eksamenssnyd.

Så der er altså visse forskelle på hvilke krav, der præcist bliver stillet til studerende og  
 960 forskere i forhold til redelighed. Det står dog også klart, at præcis som der er uklarheder og gråzoner for forskere, så er der også uklarheder og gråzoner for studerende, og derfor er det altid en god ide at snakke om reglerne og at spørge, hvis man er i tvivl.

Endelig skal det nævnes, at der kan opstå nye arbejdsformer, hvor ingen endnu har

besluttet, hvad standarderne for god praksis helt præcist er. Hvis jeg bruger en andens tekst er der nogle (nognelunde) klare standarder for, hvordan jeg skal citerer og henvise, men hvordan ser det ud, hvis jeg bruger en andens kode? I datalogien er det helt almindelig praksis, at folk deler og bruger hinandens kode via platforme som GitHub, men der har endnu ikke etableret sig klare og alment accepterede standarder for, hvor meget og hvordan man krediterer andres kode. Der er dermed en gråzone, hvor man let kan komme i klemme, hvor man let kan komme i klemme, hvis man ikke lever op til andres udtalte forventninger om, hvordan der krediteres - eller hvis andre bruger ens egen kode uden at kreditere sådan som man forventede det. Derfor er det vigtigt, at man har en åben dialog, hvor spillereglerne for brug af andres kode fastlægges klart.

## Litteratur

- ACM (u.å.). *ACM Code of Ethics and Professional Conduct*. URL: <https://www.acm.org/code-of-ethics>.
- Adrian, Stine Willum (2016). „Subversive Practices of Sperm Donation. Globalising Danish Sperm“. I: *Critical Kinship Studies*. Red. af C. Krolykke m.fl. Bd. 1. Rowman & Littlefield International, s. 185–202.
- Angwin, Julia m.fl. (maj 2016). *Machine Bias*. URL: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (bes. 04.04.2018).
- Behrendt, M. (jul. 2014). *Big Data-værktøj rykker ind i gymnasierne og fortæller hvem der dropper ud*. URL: <https://www.version2.dk/artikel/big-data-vaerktoej-forudsiger-gymnasie-dropouts-med-imponerende-traefsaekkerhed-59048> (bes. 11.01.2019).
- Bentham, Jeremy (2000). *An Introduction to the Principles of Morals and Legislation*. Batoche Books.
- Christiansen, F. V. (2015). „Rudimentær farmaceutisk etik“. *Danish Code of Conduct for Research Integrity* (nov. 2014). København.
- Darley, John M. og C. Daniel Batson (1973). „'From Jerusalem to Jericho'. A Study of Situational and Dispositional Variables in Helping Behavior“. *Journal of Personality and Social Psychology*, bd. 27, nr. 1, s. 100–108.
- Fanelli, Daniele (2009). „How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data (How Many Falsify Research?)“ *PLoS ONE*, bd. 4, nr. 5, e5738.
- Hilgartner, Stephen (1990). „Research Fraud, Misconduct, and the IRB“. *IRB: Ethics & Human Research*, bd. 12, nr. 1, s. 1–4. DOI: 10.2307/3563681.
- History of Computing, IEEE Annals of the (u.å.). *IEEE Code of Ethics*. URL: <https://www.ieee.org/about/corporate/governance/p7-8.html>.
- Johnson, Deborah G. (2009). *Computer Ethics. United States Edition*. 4. udg. Upper Saddle River (NJ): Pearson Education.
- Kant, Immanuel (1996). „On a Supposed Right to Lie from Philanthropy“. I: *Practical Philosophy*. Red. af Mary J. Gregor. Cambridge University Press, s. 605–615.
- (1999). *Grundlæggelse af sædernes metafysik*. København: Hans Reizels Forlag.
- Kleinberg, Jon, Sendhil Mullainathan og Manish Raghavan (2017). „Inherent Trade-Offs in the Fair Determination of Risk Scores“. I: *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Red. af Christos H. Papadimitriou. Leibniz International

- Proceedings in Informatics (LIPIcs) 67. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 43:1–43:23. DOI: 10.4230/LIPIcs.ITCS.2017.43.
- 1010 Lazer, D. m.fl. (2014). „The Parable of Google Flu. Traps in Big Data Analysis“. *Science*, bd. 343, nr. 6176, s. 1203–1205. DOI: 10.1126/science.1248506.
- Lov om videnskabelig uredelighed m.v.* (Apr. 2017).
- Martinson, Brian C., Melissa S. Anderson og Raymond de Vries (2005). „Scientists behaving badly“. *Nature*, bd. 435, s. 737–738. DOI: 10.1038/435737a.
- 1015 Maxmen, Amy (okt. 2018). „A moral map for AI cars“. *Nature*, bd. 562, s. 469–470. DOI: 10.1038/d41586-018-07135-0.
- Mittelstadt, Brent Daniel og Luciano Floridi (2016). „The Ethics of Big Data. Current and Foreseeable Issues in Biomedical Contexts“. *Science and Engineering Ethics*, bd. 22, nr. 2, s. 303–341. DOI: 10.1007/s11948-015-9652-2.
- 1020 Møllerhøj, Jakob (aug. 2015). *It-system til varsel af elevfrafald blev øjeblikkeligt standset af gymnasierne*. URL: <https://www.version2.dk/artikel/noejagtig-frafalds-algoritme-blankt-afvist-i-udbredt-gymnasie-it-320922> (bes. 09.07.2019).
- Nagel, Thomas (1979). *Limits of objectivity. Tanner Lectures on human values*. URL: [https://tannerlectures.utah.edu/\\_documents/a-to-z/n/nagel80.pdf](https://tannerlectures.utah.edu/_documents/a-to-z/n/nagel80.pdf).
- 1025 — (1986). *The view from nowhere*. Oxford: Oxford University Press.
- Nakashima, E. (nov. 2007). *Feeling Betrayed. Facebook Users Force Site to Honor Their Privacy*. URL: <https://www.washingtonpost.com/wp-dyn/content/article/2007/11/29/AR2007112902503.html>.
- Pascual-Ezama, David m.fl. (aug. 2015). „Context-dependent cheating. Experimental evidence from 16 countries“. *Journal of Economic Behavior & Organization*, bd. 116, s. 379–386. DOI: 10.1016/j.jebo.2015.04.020.
- 1030 PROSA (2000). *God Databehandlingsskik: En etisk guide for professionelle IT-folk*. København: PROSA.
- Rawls, John. (1972). *A theory of justice*. Oxford: Clarendon Press.
- 1035 Reverby, S. (2011). „Normal Exposure and Inoculation Syphilis. A PHS 'Tuskegee' Doctor in Guatemala, 1946–1948“. *Journal of Policy History*, bd. 23, nr. 1, s. 6–28.
- Ryberg, Jesper (2003). „Ethics and Military Research“. I: *Mathematics and War*. Red. af Bernhelm Booß-Bavnbek og Jens Høyrup. Basel, Boston og Berlin: Birkhäuser Verlag, s. 352–364.
- 1040 Stroebe, Wolfgang, Tom Postmes og Russell Spears (2012). „Scientific Misconduct and the Myth of Self-Correction in Science“. *Perspectives on Psychological Science*, bd. 7, nr. 6, s. 670–688. DOI: 10.1177/1745691612460687.
- Swartz, Aaron (2020). *Guerilla Open Access Manifesto*. URL: <https://archive.org/details/GuerillaOpenAccessManifesto> (bes. 26.05.2020).
- 1045 „Uniform Requirements for Manuscripts Submitted to Biomedical Journals“ (1997). *JAMA: The Journal of the American Medical Association*, bd. 277, nr. 11, s. 927–934.
- Youyou, W., M. Kosinski og D. Stillwell (2015). „Computer-based personality judgments are more accurate than those made by humans“. *Proceedings of the National Academy of Sciences of the United States of America*, bd. 112, nr. 4, s. 1036–1040.
- 1050 Zuboff, Shoshana (2019). *The age of surveillance capitalism. The fight for a human future at the new frontier of power*. London: Profile Books.