

Nudging, hva er det og bør vi benytte oss av det?

Fredrik Andersen og Anders Dechsling
Høgskolen i Østfold

Nudging omtales som valgarkitektur og har til hensikt å «dulte» individer og grupper mot ønsket atferd. Vi introduserer bakgrunnen for nudging og tar utgangspunkt i en gjennomarbeidet definisjon av nudging som vi kategoriserer innenfor atferdsanalyse. Vi argumenterer for at nudging er stimuluskontroll og prompting-prosedyrer. Vansker med å kategorisere nudging innenfor atferdsanalyse har også medført manglende etiske betraktninger. Ved å kategorisere og plassere nudging innenfor subsystemer i atferdsanalyse kan vi peke på hvilke etiske retningslinjer som bør følges. Vi argumenterer for at «libertariansk paternalisme» ikke skiller seg fra paternalisme. Videre påstår vi at nudging ikke står i noen særstilling, men at det ved nudging trengs uttrykt og informert samtykke på lik linje som ved andre intervensjonsformer. Der hvor nudgingen ikke er utarbeidet i samhandling med den nudgede, er nudging klassisk paternalisme og må begrunnes prinsipielt.

Nøkkelord: Nudging, dulting, paternalisme, stimuluskontroll, prompting, etikk

Nudging – what is it and should we employ it?

Nudging is often labeled as choice architecture with the intention of nudging individuals and groups towards planned behaviors. We introduce nudging as an element of behavior economics, and categorize nudging within behavior analysis, where we argue that nudging is best understood as stimulus control and prompting procedures. After categorizing nudging under subsystems within behavior analysis, we present a brief ethical discussion and argue that “libertarian paternalism” does not differ considerably from classical paternalism. As such, nudging is not isolated from other paternalistic procedures, and needs expressed and informed consent as with other form of intervention. In cases where nudging procedures are introduced without the nudgee’s informed consent, nudging is classical paternalism and in need of principled justification.

Key words: Nudging, paternalism, stimulus control, prompting, ethics

Nudging¹ er metoder for endring av atferd og blir hyppig brukt som virkemiddel opp mot samfunnsrelevante tema som personlig økonomi, folkehelse og hygiene. Prosedyrens nærhet til atferdsanalyse gjør at stadig flere atferdsanalytikere ønsker å forstå, kategorisere og evaluere nudging på ulike nivåer. Som atferdsanalytikere er vi spesielt opptatt av hvordan nudging skal

forstås i et atferdsanalytisk rammeverk og hvilke etiske evalueringer man kan gjøre av nudgingiltak både i helse- og omsorgstjenestene og for den øvrige befolkningen. Vi vil se på hva nudging er, hvor det kommer fra, hvordan nudging kan kategoriseres innenfor atferdsanalyse, og hvilke etiske konklusjoner vi kan trekke på et overordnet nivå.

¹Nudging er et engelsk begrep, men brukes også i norsk dagligtale. Nudging kan oversettes til «dulting» på norsk. I denne artikkelen bruker vi flere engelske betegnelser med norske endinger.

Det er ingen konflikter mellom forfatterne med hensyn til dette manuskriptet. Takk til Yngve Herikstad og Line Børtveit for nyttige kommentarer. Korrespondanse vedrørende manuskriptet kan sendes til fredrik.andersen@hiiof.no

Begrepet «nudging» eller «dulting» fikk fullt innpass i atferdsøkonomien etter Thaler og Sunnsteins (2008) bok: *Nudge: improving decisions about health, wealth, and happiness*. Thaler vant den Svenske riksbankens pris i økonomisk vitenskap i Alfred Nobels minne i 2017. Det atferdsøkonomiske grunnlaget Thaler bygger på var allerede etablert som vitenskap da Kahnemann vant samme pris i 2002. For å forstå nudging er det nyttig med en innføring i det faghistoriske perspektivet teorien og prosedyrene springer ut av.

Nyklassisk økonomisk teori² har tradisjonelt tatt utgangspunkt i et begrep om rasjonalitet som omhandler evnen til å effektivt jobbe mot preferansetilfredsstillelse (Sen, 1977). Forutsetninger for slik rasjonalitet er at man kan ta inn over seg nødvendig informasjon, prioritere preferanser og handle effektivt i retning av disse preferansene. Deretter har man postulert fiktive rasjonelle aktører eller *homo economicus* i modellkonstruksjon. Det postuleres med andre ord aktører som har tilgang på informasjon, evner å forstå denne informasjonen, og evner å handle effektivt med grunnlag i den. Mennesker er ikke rasjonelle aktører i denne forstanden. Når vi snakker om mennesker, må vi snakke om begrenset rasjonalitet og rasjonelle forventninger. «Begrenset rasjonalitet» viser til at vi tar beslutninger som vi ikke forventer at skal være optimale, men heller gode nok. Til dette benytter vi oss blant annet av tommelfingerregler som forenkler kompleks informasjon. Med rasjonelle forventninger menes det at til tross for at det ikke finnes fullt ut rasjonelle aktører, er det effektivt å anta deres eksistens når vi skal analysere på gruppenivå. Dette er spesielt relevant når man skal modellere hvordan grupper av økonomiske aktører endrer atferd i lys av ny informasjon. Tanken er at folks atferd viker fra rasjonalitet på ulike måter og retninger og at den overordnede trenden vil tendere mot en rasjonell

aktørs atferd (Sargent, 2021). Nyklassisk økonomisk teori modellerer altså som om rasjonelle aktører eksisterer, selv om de vet at dette er en idealisering. Det er en lang tradisjon innenfor økonomifaget å kritisere modellering ved bruk av fiktive elementer på denne måten (se for eksempel Sen, 1977).

Atferdsøkonomisk teori oppstod på mange måter som en opposisjon mot nyklassisk modellering og bygger på en tanke om at modellene må reflektere faktisk atferd og faktisk eksisterende analyseenheter. Teorien bygger blant annet på psykologisk teori og det er i denne tradisjonen vi finner teoriene til Daniel Kahnemann og Amos N. Tversky, og Richard H. Thaler og Cass R. Sunnstein. Kahneman og Tversky har avdekket og kategorisert en rekke systematisk irrasjonell atferd. For eksempel at vi oppfører oss annerledes enn det en rasjonell aktør ville gjort blant annet fordi vi forenkler informasjonen vi tar til oss (Kahnemann, 2011). Et eksempel på slik forenkling er den vanlige journalistiske fortolkningen av Pfizers Covid-19 vaksine som 95% effektiv. Vår første tanke er gjerne at dersom vaksinen er 95% effektiv så vil 19 av 20 personer med vaksinen være trygge, mens den resterende personen (5%) befinner seg i samme situasjon som en person som ikke har fått vaksine. Altså tenker vi at med en vaksine med 95% effektivitet er hver 20. vaksinerte person ubeskyttet mot viruset. De faktiske forhold er at kun 0,04% av de vaksinerte i Pfizers studie fikk Covid-19, og de fleste av disse fikk mildere symptomer enn hva man ville forvente uten vaksine (Polack et al., 2020). De 95% refererer til relasjonen mellom sannsynligheten for at du får Covid-19 med eller uten vaksine. Med vaksinen har du 95% mindre sjanse for å få Covid-19-symptomer. Dette er et eksempel på hvordan vi systematisk henter opp heuristikker ($95\% = 1/20$) som hjelper oss å finne en umiddelbar fortolkning, men hvor heuristikken er villedende og fortolkningen dermed blir feil. Vi har altså forenklet oss frem til en rask, men feil løsning på problemet. Kirkebøen

²I vårt henseende kan vi tenke om nyklassisk økonomisk teori som en samling av de økonomiske teorier som tar som utgangspunkt at individer har utfallspreferanser og at individer handler rasjonelt etter disse preferansene. Se for eksempel Weintraub (2002).

(2011) gir en god oppsummering av ulike heuristikker av denne typen.

I tillegg til at vi systematisk henter opp heuristikker som leder til feilfortolkning vet vi at folk tar valg som går mot deres egne langtidsprefranser selv i de tilfellene hvor de har tolket informasjonen korrekt. Systematikken i dette kan blant annet vises gjennom litteraturen på temporal diskontering (se for eksempel Critchfield & Kollins, 2001; Story et al., 2014).

Nudging er en samlebetegnelse for atferdsendrende tiltak som aktivt drar nytte av disse systematiske feilene i menneskelig valgatferd. Det er flere grunner til at atferdsanalytikere bør sette seg inn i nudging-litteraturen. Nudging og atferdsanalyse har en rekke fellestrekk (se for eksempel da Rocha & Hunziker, 2020; Simon & Tagliabue, 2018). I tillegg ser nudging ut til å være en prosedyre som i liten grad benytter seg av tvang, og nudging-litteraturen åpner inn mot Skinners beskrivelse av utopien Walden Two hvor protagonisten Frazier hevder at det vil være umoralsk ikke å kontrollere påvirkningen av valg ettersom valg uansett påvirkes (Skinner, 2005/1948).

Hva er nudging?

Nudging er en prosedyre for atferdsendring som drar nytte av våre systematiske og forutsigbare tankefeil, og som retter seg spesifikt mot valgatferd. Det som gjør nudging til noe særegent er mantraet om at disse atferdsendrende prosedyrene ikke faller under kategorien klassisk paternalisme. Klassisk paternalisme er enhver inngripen i en persons liv, mot dens vilje, som begrunnes med hensynet til personens eget beste (for eksempel påbud om bruk av bilbelte eller forbud mot bruk av heroin). Fraværet av klassisk paternalisme i nudging begrunnes i de begrensinger som settes på hvordan strategiene skal gjennomføres. Det foreligger for eksempel et forbud mot insentivstyring av valgene. Dersom du belønner eller straffer en gitt atferd, driver du ikke med nudging (Hansen, 2016).

Begrensninger av denne typen — du skal ikke fjerne valgmuligheter, du skal ikke sanksjonere mot uønsket atferd, du skal ikke belønne ønsket atferd — har ledet til begrepet “libertariansk paternalisme” (Hansen, 2016). Tanken er at nudging er paternalisme ettersom det er et atferdsendrende tiltak med et intendert og forutsigbart resultat, men libertariansk fordi man hverken fjerner valgmuligheter eller sanksjonerer mot uønsket atferd. Tilstedeværelsen av valgmuligheter og fraværet av sanksjoner stemmer godt overens med den overordnede libertarianske tanken om at man skal unngå tvungen inngripen i individets frie utfoldelse.

Nudging er et begrep med utstrakt anvendelse, hvor definisjonene ofte er mer folkelige enn tekniske. Dette leder til en potensiell konseptuell uklarhet som kan tåkelegge diskusjonen om hva nudging er. Hansen (2016) tar for seg Thaler og Sunnsteins stipulative definisjoner av nudging-begrepet og forbedrer disse med fokus på klarhet og konsistens. I så måte er Hansens (2016) mye brukte definisjon et godt utgangspunkt for å forstå nudging som teknisk begrep. Hansen definerer nudging på følgende måte:

“A nudge is a function of (1) any attempt at influencing people’s judgment, choice or behaviour in a predictable way, that is (1) made possible because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally in their own self-declared interests, and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts. Thus a nudge amongst other things works independently of: (i) forbidding or adding any rationally relevant choice options, (ii) hanging incentives, whether regarded in terms of time, trouble, social sanctions, economic and so forth, or (iii) the provision of factual information and rational argumentation.” (Hansen, 2016, s. 174)

En nudge, eller en dult, er altså et forsøk på å påvirke valg eller atferd på en forutsigbar måte. Dette faller naturlig inn under vårt begrep om atferdsendring og er i så måte

ikke noe nytt. Videre skal atferdsendringen *være mulig på grunn av* kognitive begrensninger, bias, rutiner, og vaner i individuelle og sosiale beslutninger som begrenser folks evne til å ta rasjonelle valg som fremmer egne preferanser. Det vil si at dersom et atferdsendrende tiltak skal kunne defineres som nudging, må tiltaket bygge på allerede eksisterende mangler i våre evner til å ta rasjonelle valg. Avslutningsvis gjøres det eksplisitt at en nudging-prosedyre ikke kan fjerne valg, innføre insentiver og sanksjoner, eller tilføre faktaopplysninger. Nudging er altså atferdsendrende tiltak som ikke fjerner valg, tilfører regler/faktaopplysninger, eller kontrollerer forsterkning og straff. Et naturlig spørsmål er derfor hvordan vi skal kategorisere nudging innenfor atferdsanalytisk teori.

Kategorisering av nudging innenfor atferdsanalytisk teori

Den tradisjonelle forståelsen av pragmatisme innenfor behaviorismen setter prediksjon og kontroll som vitenskapelige hovedmål (Hayes & Brownstein, 1986). De fleste atferdsøkonomiske beskrivelsene av nudging vektlegger prediksjon og toner ned kontrollaspektet. Dersom vi skal kunne kategorisere nudging innenfor atferdsanalyse er vi derimot avhengig av også å kunne gjøre rede for kontrollaspektet. Prediksjon og kontroll innebærer at de fleste atferdsbeskrivelsene vi gir innenfor atferdsanalyse er beskrivelser av prosesser, men utledet av spesifikke prosedyrer som gjør at man kan kontrollere og bedømme hvorvidt beskrivelsen var korrekt. For eksempel sier vi at forsterkning både er prosedyre og prosess (Catania, 2013).

Rapporter fra nudging-prosedyrer innebærer ofte en beskrivelse av utfordringer knyttet til temporal diskontering. Temporal diskontering kan beskrive hvordan umiddelbare forsterkere foretrekkes på bekostning av utsatte konsekvenser som personen i utgangspunktet vurderer som mer verdifulle. Nudging vil i dette perspektivet kunne handle om å arrangere foranledninger som

sannsynliggjør valg med utsatte konsekvenser. Utfordringen går ut på at eventuelle forsterkere ligger langt frem i tid og ofte fjernt fra målatferden. Hvordan skal man da kunne kontrollere atferd via forsterkere? Denne utfordringen er også velkjent innenfor økonomi og spillteori. Der hvor man ikke kan gi vel-timedede forsterkere overlates dette til tilfeldigheter eller naturlige forsterkere utenfor vår kontroll. Avstanden mellom atferden og et senere produkt (e.g., bedre helse, bedre økonomi osv.) er en av grunnene til at det har vist seg vanskelig å gi en tilfredsstillende beskrivelse av nudging som prosess. Simon og Tagliabue (2018) påpeker at *terms* beskrivelser av nudging er problematiske, noe vi støtter opp under, men hvordan skal man da kategorisere nudging? Vi går her gjennom mulige forklaringer hva angår både prosess og prosedyre.

Atferdsprosesser kan forklares med utgangspunkt i genetikk, læringshistorie, og de til enhver tid rådende kontingenser. I nudging-litteraturen kan prosessene strekkes til også å inkludere langtidskonsekvenser. Motivasjonen for nudging er ofte nudgerens ønsker om å oppnå langtidskonsekvenser slik som redusert matsvinn, bedre resirkulering og økt pensjonssparing. Dette er ønskede langtidseffekter hvor forventningen om effekten ikke i seg selv er sterk nok til å endre atferd i øyeblikket. Dersom kunnskap om langtidseffektene var sterke nok motivatorer ville vi ikke trenge andre tiltak enn tilførsel av informasjon (i.e., regelstyring). Langtidseffektene i seg selv kan dermed ikke regnes inn som kausale ledd i selve nudging-prosessen.

Nudget atferd har naturligvis direkte konsekvenser på samme måte som all annen atferd. Det er likevel problematisk å regne inn de direkte konsekvensene av atferden som en del av nudging-prosessen. Som påpekt av Donahoe et al. (1993) må vi skille mellom tilfeldige og pålitelige ledd i en forsterkningsprosess, og i Hansen (2016) legges det vekt på at nudgen må fungere *uavhengig* av de direkte konsekvensene. Ettersom vi ikke kan kontrollere de umiddelbare conse-

kvensene og atferden ikke produserer predikerbare direkte konsekvenser, vil de direkte konsekvensene av atferden være tilfeldige. Tilfeldige konsekvenser er ikke effekter av nudgen og kan dermed ikke innlemmes i den kausale forklaringen av hva som skjer i en nudging-prosess.

Et eksempel på hvordan nudging fungerer er finanstilsynets regulering av norske bankers nudging av kredittkortkunder. I sin opprinnelige nudge sendte norske banker ferdig-utfylte, elektroniske kredittkortregninger til sine kunder. Disse regningene kom med et forhåndsutfylt alternativ som var å betale kun gebyr (250 kr.). Dersom kunden klikket seg kjøpt gjennom regningen ville kun dette gebyret bli betalt, men ikke selve gjelden. Etter finanstilsynets regulering av banksektoren sendes regningene ut ferdig utfylt med det totale gjeldsbeløpet. Dersom kunden klikker seg kjøpt gjennom regningen betaler hen dermed hele gjeldsbeløpet. Et sentralt poeng her er at den direkte og umiddelbare konsekvensen kunden opplever -ferdig med å betale regninger- forblir uendret. Den indirekte konsekvensen, gjeld eller ikke gjeld, endres riktignok, men enhver kontrafaktisk analyse forteller oss at det ikke er den indirekte konsekvensen som styrer atferden. Om dette var tilfelle ville kundene trykke en gang ekstra på knappen og betale hele beløpet/kun gebyret uavhengig av hvilket valg som var forhåndsutfylt.

Poenget med kredittilsynets regulering var at kundene lot seg styre av de umiddelbare, direkte konsekvensene og ikke av de indirekte konsekvensene og langtidskonsekvensene. Dette kan igjen begrunnes i rasjonalitetsbegrensninger. Ved å spille på rasjonalitetsbegrensningene kunne kredittilsynet, via bankene, nudge kundene til å betale gjelden sin ved å endre på valgarkitekturen. Ettersom endrede indirekte konsekvenser av atferden ikke kan gi noen kausal forklaring på atferden og de endrede konsekvensene forblir uoppdaget kan vi ikke forklare nudging-prosessen ved hjelp av atferdsseleksjon styrt av rådende kontingenser. Dersom

vi kunne presentert forsterkere som pålitelige, umiddelbare konsekvenser ville saken stilt seg annerledes, men da ville vi ikke lenger snakke om nudging. Dersom vi skal anvende oss av tre-terms kontingente beskrivelser av nudging må vi flytte diskusjonen over i den læringshistorie som ligger til grunn for at valgarkitektur påvirker atferd. Vårt forslag er at nudging som prosess forstås best gjennom stimuluskontroll.

Stimuluskontroll brukes som begrep når klasser av forutgående stimuli kontrollerer samme responsklasse (Cooper et al., 2020). Stimuluskontroll innebærer at en atferds frekvens endres i nærvær av spesifikke stimuli som følge av læringshistorie med lignende stimuli. Et kjent fenomen kalt stimulusgeneralisering beskriver prosesser som leder til klasser av stimuli - ofte uten at relasjonen mellom stimuliene er direkte trent (Gå på grønn mann generaliseres til grønn sone, grønt lys, grønn pil etc.). Alle de ulike stimuliene kontrollerer samme responsklasse — gå på grønt.

Nudging vil ofte være under stimuluskontroll som følge av stimulusgeneralisering. Altså at en forsterkningshistorie knyttet til spesifikke stimuli generaliseres til stimuli med lignende egenskaper (Green & Saunders, 1998). Vi agerer ofte automatisk med utgangspunkt i heuristikker basert i stimulusgeneralisering. Heuristikkene her er de samme tommelfingerreglene det refereres til forbindelse med begrepet «begrenset rasjonalitet» innenfor atferdsøkonomi. Automatisert atferd kan utnyttes i nudging-tiltak ved at man presenterer kjent stimulus, men nå med «nytt» innhold. Dette er hva som foregår i «default option» tiltak slik som i bankeksempelet over. Slike tiltak betegnes som de mest effektive nudging-tiltakene (se Hummel & Maedche, 2019).

En fremstilling av nudging-prosedyrer basert på tretermskontingens vil ikke være i tråd med definisjonen av en nudge fordi det vil være en endring av insentiver dersom man endrer på konsekvensene av atferden. Et videre problem i så måte er valgarkitek-

turets forbud mot å fjerne alternativer. Vi kan altså ikke påvirke atferden ved å endre foranledningen i radikal grad fordi det i definisjonen kun er tillatt med små arkitektoniske endringer, og videre vil det heller ikke være anledning til å forandre foranledningen ved å tilføre informasjon slik man ser ved regelstyrt atferd. Dette betyr selvsagt ikke at det er umulig å påvirke foranledninger og konsekvenser ved hjelp av informasjon eller fjerning av alternativer, men da vil prosedyrene ikke havne under definisjonen av nudging.

Nudging innebærer å gjøre arkitektoniske endringer i miljøet for å sannsynliggjøre noen valg framfor andre. Man er derfor nødt til å manipulere foranledninger som sannsynliggjør slike valg. Vi foreslår derfor at nudging-prosedyrer bør kategoriseres som stimuluskontroll- og at valgarkitekturen i prinsippet er det samme som prompting-prosedyrer.

Prompts er forutgående stimuli, eller hjelpebetingelser, som skal bidra til sannsynliggjøringen av korrekte responser i nærvær av relevante stimuli (Cooper et al., 2020). Nudging-tiltak i form av valgarkitektur faller naturlig inn under kategorien prompting ettersom prompts brukes til å modifisere karakteristika i miljøet som sannsynliggjør ønsket atferd. Til tider konstrueres nudging-tiltaket slik at det skal opprettes ny, eller brytes med allerede etablert stimuluskontroll. En nudging-prosedyre kan med andre ord forstås som prompts som leder individet til å handle i tråd med eller til å bryte tidligere atferdsmønstre.

Det er også andre forutgående stimuli som kan være med å påvirke effekten av nudging-tiltak. Motivasjonelle operasjoner er stimuli som igangsetter atferd eller påvirker forsterkeres verdi (Catania, 2013; Laraway et al., 2003). Et ofte brukt eksempel er at deprivasjon på vann både påvirker vanns verdi som forsterker i tillegg til at atferd som tidligere har ledet til vann igangsettes. Motivasjonelle operasjoner bidrar til forklaringer på hvorfor individer responderer ulikt på nudging-prosedyrer. Motivasjonelle opera-

sjoner som dårlig tid, deprivasjonstilstander, og habituering påvirker hvor sensitive vi er for prompts i miljøet.

Nudging er mest brukt som prosedyre rettet mot grupper fremfor individer. Det faktum at individer responderer ulikt på nudging styrker vår tese om at nudging er prompting, som igjen er sensitivitetsstyrt av motivasjonelle operasjoner. Det vil være mulig å predikere at en viss andel av befolkningen vil handle i tråd med en nudge, men umulig å kontrollere for hvilke individer som utgjør denne andelen. Vi møter dermed velkjente problemer knyttet til slutninger som trekkes fra kunnskap om grupper til kunnskap om individer. I tillegg møter vi problemer knyttet til overføringsverdiene av kunnskap fra en gruppe til andre grupper (se f.eks. Cartwright & Munroe, 2010). Enhver sosial gruppe har en mengde felles atferd, læringshistorie og stimulusklasser som muliggjør predikerbar atferdskontroll. Et eksempel er at de fleste nordmenn vil reise seg dersom nasjonalsangen spilles av i en idrettshall. Dette er en predikerbar atferd blant nordmenn som følger av læringshistorien vi har til felles. Vi vil ikke forvente samme respons om vi spiller: "Ja, vi elsker", under et japansk idrettsarrangement.

Slik vi ser det må prosessene i nudging forklares med en prediksjon på individens læringshistorie (e.g., stimulusgeneralisering) og eventuelle motivasjonelle faktorer. En nudge virker dersom alle disse faktorene er til stede. Nudging som prosedyre er således stimuluskontroll og prompts, men man kan ikke beskrive prosessene nudging i en tretermkontingens fordi de umiddelbare konsekvensene ikke lar seg kontrollere.

Kategorisering og etikk

Innenfor atferdsanalyse har man tidligere sett eksempler på at argumenter for gjennomføring av intervensjoner har vært basert på rent paternalistiske forestillinger om andres ve og vel, kombinert med en utilitaristisk etikk hvor målet i stor grad helliger middelet. Som følge av en rekke hendelser hvor

atferdsanalytiske prinsipper og verktøy har blitt brukt på etisk uforsvarlig vis, har etiske retningslinjer blitt viet stor oppmerksomhet innen atferdsanalyse (Bailey & Burch, 2016). I vår tid vies etiske vurderinger og hensyn stor plass i både atferdsanalytisk forskning og praksis. Ved å kategorisere nudging som stimuluskontroll og prompting åpner det seg muligheter for å anvende dette etiske rammeverket på nudgingprosedyrer i samme grad som alle andre atferdsanalytiske prosedyrer.

Atferdsanalyse kan deles i fire domener (Løkke et al., 2012) med tilhørende etiske retningslinjer: (1) *Eksperimentell atferdsanalyse* er underlagt forskningsetiske retningslinjer og for eksempel Helsinki deklarasjonen ved gjennomføring av eksperimenter. (2) *Konseptuell eller teoretisk atferdsanalyse*, som innebærer raffinering av teorier og begreper, vil i stor grad være uten regulering utover allmennvitenskapelige etiske retningslinjer om språkbruk og omtale av personer og grupper (3) *Anvendt atferdsanalyse* er underlagt etiske retningslinjer knyttet til behandling og avklarte og definerte retningslinjer for bruk av atferdsanalyse (Bailey & Burch, 2016). (4) *Tjenesteyting basert på atferdsanalyse*, som innebærer en bevissthet knyttet til prinsipper i atferdsanalyse i daglig tjenesteyting, vil være bundet av yrkesetiske retningslinjer og lovverk knyttet til profesjonsutøvelsen. Løkke og Salthes (2012) sjekkliste har i tillegg foreslått viktige normative premisser som bør ligge til grunn i systematisk miljøarbeid. En fellesnevner for alle disse delene er en forventning om respekt for individets frihet og menneskeverd. Det er viktig å merke seg at det ikke er nudging (som alle prosedyrer) i seg selv som kategoriseres i de ulike domene, men at domenet nudgeren opererer i påvirker hvilke formelle etiske retningslinjer som er rådende.

Nudging, prompting og etikk

Skinners Walden Two føyer seg inn i rekken av klassisk paternalistiske politiske tekster hvor tanken er at dersom “noen”

kan ta tilstrekkelig kontroll over viktige valg i menneskenes liv så vil den totale lykken økes, og samfunnet såvel som individene kan befris fra en mengde onder. I Platons “Staten” foreslås det, i tillegg til eugeniske prosjekter gjennom falske formeringslotterier, sensur av deler av religion, teater og musikk (Platon, 2015). I “manufacturing consent” argumenterte den amerikanske propaganda- og PR-guruen Edward Bernays at det ville være mer effektivt å styre opinionen gjennom tiltak som folket i liten grad legger merke til (Bernays, 1928). Det disse har til felles er tankene om at samfunnet må styres, det må styres i en gitt retning, og retningen må pekes ut av noen andre enn majoritetsbefolkningen. Det finnes naturligvis mindre totalitære måter å være paternalistisk på, og det er rimelig å tenke at Skinners utopi ikke speiler Skinners reelle politiske ideologi. Allikevel er det slik at Skinners, Platons, Bernays og en rekke andres paternalistiske teksters uttrykk for kravet om kontroll har ført til skepsis blant liberalistiske og libertarianske røsters møte med enhver paternalistisk praksis.

Nudging-litteraturen møter denne skepsisen ved å hevde at nudging riktignok er paternalistisk, men at det er en særegen form for paternalisme som går klar av den klassiske paternalismens største problemer. Thaler og Sunstein (2009) innførte begrepet “libertariansk paternalisme” for å fange denne særegenheten og det diskuteres flittig i atferdsøkonomiske tekster hvorvidt nudging er manipulasjon eller ikke, og hvis det er manipulasjon, hvorvidt det er valg som manipuleres (Thaler & Sunstein, 2009). Det er verdt å merke seg at Thaler og Sunsteins opprinnelige beskrivelse av nudging som “libertariansk paternalisme” har i seg en forutsetning om at vi nudger til fordel for den som nudges. Denne forutsetningen er ikke en del av Hansens (2016) definisjon. Flere av tiltakene vi har nevnt her nudger i en retning som er nøytral eller direkte ufordelaktig for den nudgede, og allerede her kan vi se at paternalismebegrepet ikke alltid er anvendelig i nudging. Det er ikke noe formelt krav

om at man må bedrive “*nudging for good*”, og dermed ingen garanti for at tiltaket har sosial validitet. Tilfeller hvor man nudger i strid mot den nudgedes interesser og til fordel for statlige eller økonomiske interesser er forhåpentlig ikke en del av atferdsanalytikerens intenderte bruk av nudging som strategi. Vi velger dermed ikke å diskutere dette i detalj.

Vi skal omgå de dype filosofiske diskusjonene og ganske enkelt spørre hvorvidt nudging som prompting og stimuluskontroll trenger etisk evaluering. For å holde diskusjonen innenfor tema, vil vi kun forholde oss til prompting-prosedyrer hvor ingen alternativer er fjernet og hvor vi ikke kontrollerer de umiddelbare forsterkerne, altså nudgingprosedyrer. Vi vil også begrense oss til anvendt atferdsanalyse og tjenesteyting basert på atferdsanalyse. Dersom vi prompter med tjenestemottakers reelle samtykke og innsikt i prompting-prosedyren, kan vi ikke se større etiske utfordringer. Dersom tjenestemottaker enten ikke forstår eller ikke er gjort oppmerksom på prosedyren stiller saken seg annerledes.

Her må vi i første hånd spørre hvorfor vi prompter. Svaret på dette er at vi prompter for å sannsynliggjøre en gitt atferd. Dette er klassisk atferdsendring hvor enten den ønskede atferden, eller prosedyren som fremskaffer atferden er bestemt av noe ytre (oss) og ikke av tjenestemottakeren selv. Dersom ønsket atferd er tenkt å være for tjenestemottakers eget beste, er prosedyren en klassisk paternalistisk prosedyre. I da Rocha og Hunziker (2020) tar forfatterne for seg hvordan etikken rundt nudging skal forstås innenfor et perspektiv hvor autonomi fremstilles som illusorisk. Vi samstemmer med Løkke og Salthe (2012) at alle paternalistiske tiltak må begrunnes prinsipielt, og at autonomi må respekteres (Bailey & Burch, 2016). Vår posisjon er dermed at alle nudging-prosedyrer som gjennomføres uten den nudgedes viten og uttrykte vilje, må begrunnes prinsipielt. Nudging står ikke i noen særstilling.

En viktig grunn til kravet om tjeneste-

mottakers uttrykte aksept av både ønsket atferd og prosedyre er at enhver person må ha mulighet til å motsi seg atferdsendrende tiltak fra staten og hjelpeapparatet. Nudgingens særegenhet er at dette er en form for prosedyre som vi gjerne ikke legger merke til og dermed er det spesielt viktig at vi innhenter reelt og tydelig samtykke. Hvordan samtykket skal innhentes på nivåer over det individuelle nivået er en etisk utfordring for de som benytter seg av nudging på gruppenivå. Dersom tjenestemottaker eksplisitt samtykker til både prosedyre og ønsket atferd, anser vi nudging for å være en velkommen del av effektive og lite inngripende atferdsanalytiske prosedyrer.

Referanser

- Bailey, J. S. & Burch, M. R. (2016) *Ethics for behavior analysts* (3rd ed). Lawrence Erlbaum Associates.
- Bernays, E. (1928). *Propaganda*. Horace Liveright.
- Cartwright, N., & Eileen, M. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16(2), 260–266. <https://doi.org/10.1111/j.1365-2753.2010.01382.x>
- Catania, C. (2013). *Learning* (5th ed). Sloan.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied Behavior Analysis* (3rd ed.). Pearson Education.
- Critchfield, T. S., & Kollins, S. H. (2001). Temporal discounting: basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis*, 34, 101–122. <https://doi.org/10.1901/jaba.2001.34-101>
- da Rocha, C. A. A., & Hunziker, M. H. L. (2020). A behavior-analytic view on nudges: Individual, technique, and ethics. *Behavior and Social Issues*, 29(1), 138-161. <https://doi.org/10.1007/s42822-020-00037-9>
- Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). A selectionist approach

- to reinforcement. *Journal of the Experimental Analysis of Behavior*, 60(1), 17–40. <https://doi.org/10.1901/jeab.1993.60-17>
- Green, G., & Saunders, R. R. (1998). Stimulus equivalence. I K. A. Lattal, & M. Perone (Red.), *Handbook of research methods in human operant behavior* (s. 229–262). Plenum.
- Hansen, P. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, 7(1), 155–174. <https://doi.org/10.1017/S1867299X00005468>
- Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *The Behavior Analyst*, 9(2), 175–190. <https://doi.org/10.1007/BF03391944>
- Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, 80, 47–58. <https://doi.org/10.1016/j.socec.2019.03.005>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Strauss and Giroux.
- Kirkebøen, G. (2011). Kan vi stole på fagfolks skjønn? I Molander, A. & Smeby, J. C. (red). *Profesjonsstudier II* (27–43). Universitetsforlaget.
- Laraway, S., Snyerski, S., Michael, J., & Poling, A. (2003). Motivating operations and terms to describe them: Some further refinements. *Journal of Applied Behavior Analysis*, 36(3), 407–414. <https://doi.org/10.1901/jaba.2003.36-407>
- Løkke, J. A., Arntzen, E., & Løkke, G. E. H. (2012). Om begreper: Inndeling av atferdsanalysen i subsystemer eller arbeidsområder. *Norsk Tidsskrift for Atferdsanalyse*, 39(1), 63–71.
- Løkke, J. A., & Salthe, G. (2012). Sjekklister for målrettet tiltaksarbeid: fra normative og deskriptive premisser til tiltak og evaluering. *Norsk Tidsskrift for Atferdsanalyse*, 39(1), 17–32.
- Plato, & Jowett, B. (2015). *The republic / by Plato; translated and introduction by Benjamin Jowett*. First Avenue Editions.
- Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Frenck, R. W. ... Gruber, W. C. (2020). Safety and efficacy of the BNR162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine*, 383, 2603–2615. <https://doi.org/10.1056/NEJMoa2034577>
- Sargent, T. J., (2021, mars). Rational expectations. *The Library of Economics and Liberty*. Hentet fra: <https://www.econlib.org/library/Enc/RationalExpectations.html>
- Sen, A. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 6(4), 317–344. <https://www.jstor.org/stable/2264946>
- Simon, C., & Tagliabue, M. (2018). Feeding the behavioral revolution: Contributions of behavior analysis to nudging and vice versa. *Journal of Behavioral Economics for Policy*, 2(1), 91–97. <http://sabeconomics.org/wordpress/wp-content/uploads/JBEP-2-1-13.pdf>
- Skinner, B. F., (2005/1948). *Walden two*. Macmillian.
- Story, G. W., Vlaev, I., Seymour, B., Darzi, A., & Dolan, R. J. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in Behavioral Neuroscience*, 8. <https://doi.org/10.3389/fnbeh.2014.00076>
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth and happiness*. Yale University Press.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: improving decisions about health, wealth and happiness*. SD Books.
- Weintraub, E. R. (2002). Neoclassical

Economics. *The Library of Economics and Liberty*. Hentet fra: <https://www.econlib.org/library/Enc/NeoclassicalEconomics.html>

<https://www.econlib.org/library/Enc/NeoclassicalEconomics.html>
