

Å vite hva man tror

Selverkjennelse og førstepersonsautoritet hos Davidson, Hacker og Wittgenstein

Steinar Bøyum

Spørsmålet om førstepersonsautoritet i tilskrivelse av psykologiske tilstander er blitt et sentralt filosofisk spørsmål. Denne artikkelen diskuterer Wittgenstein-kommentatoren Peter Hacker sin kritikk av Donald Davidsons klassiske artikkel om førstepersonsautoritet.

“Hvorfor gjorde du det?” – vi forventes å kunne svare på dette spørsmålet. En handling som vi ikke kan angi noen grunn for, er irrasjonell. Derimot forventes vi normalt ikke å kunne svare på spørsmålet “Hvorfor gjorde han det?” Å ikke vite hvorfor noen utførte en viss handling er ikke et tilfelle av irrasjonalitet, men uvitenhet. Det er noe vi ikke vet, og for å få vite det, spør vi først den som gjorde det. Og vi forventer at han kan svare. Men forventningen trenger ikke bli innfridd. Personen kan spørre seg selv: “Hvorfor gjorde jeg det?” Han kan vite det, og si det. Han kan vite det, men ikke ville si det. Eller han kan lete etter forklaringen hele livet.

Dette viser noen av forskjellene mellom førsteperson og tredjeperson med hensyn til spørsmål etter grunner. Det viser også noe av kompleksiteten i dagligdags sosial interaksjon som det vitenskapsfilosofiske problemet om handlingsforklaringer oppstår fra. I dette essayet skal jeg derimot ikke eksplisitt behandle grunner for handling, men snarere noen av de tingene som handlingsgrunner refererer til: de oppfatninger (beliefs), ønsker (desires) og hensikter (intentions) man gjerne tenker seg bestemmer en handling. Og med hensyn til disse skal jeg bare behandle ett spørsmål, om enn et sentralt: Hvem sitter til syvende og sist med sannheten om hva X tror, ønsker og intenderer? Noe som videre betyr: Hvem vet til syvende og sist (best) hvorfor X utfører en handling? Et naturlig svar er: X selv. En slik oppfatning kan man grovt sagt kalle for tesen om førstepersonsautoritet, og under dette navnet har temaet blitt en hel liten filosofisk industri de siste par tiårene. Av blant annet denne grunnen skal jeg heller ikke behandle temaet førstepersonsautoritet i sin fulle bredde i det følgende. Jeg skal bare behandle en enkelt vending i denne debatten, en bevegelse som gjerne forbindes med Wittgenstein, nemlig spørsmålet om vi i det hele tatt kan sies å ha *kunnskap* om egne tanker, ønsker og følelser.

I stedet for å la dette bli et essay i Wittgenstein-filologi vil jeg ta utgangspunkt i en klassiker på feltet, Donald Davidson. Jeg vil deretter se på Wittgenstein-kommentatoren Peter Hacker sin kritikk av Davidsons syn på førstepersonsautoritet. Resten av artikkelen vil så bestå av en kritisk undersøkelse av Hackers sentrale antagelser, med særlig vekt på hans måte å lese Wittgensteins tekster på.

1) Davidson om førstepersonsautoritet

Førstepersonsautoritet innebærer en tanke om asymmetri mellom det jeg sier om mine egne tanker og følelser og det andre sier om mine tanker og følelser. I sin artikkel “First Person Authority” åpner Donald Davidson med å formulere asymmetrien på denne måten: “When a speaker avers that he has a belief, hope, desire or intention, there is a presumption that he is not mistaken, a presumption that does not attach to his ascriptions of similar mental states to others”.¹ Han stiller videre det som for han er det sentrale spørsmålet: “What accounts for the authority accorded first person present tense claims of this sort, and denied second or third person claims?”²

Davidson skynder seg med å kvalifisere denne førstepersonsautoriteten. Den gjelder kun utsagn i *nåtid*. Den gjelder i varierende *grad* og på ulike *måter* for ulike “propositional attitudes”. Og den utelukker ikke at man kan ta feil om eller tvile på hva man selv tror, tenker og føler. Davidson benekter at vi har en særegen ufeilbarlig evne til introspeksjon – det er nettopp ved å forkaste ideen om en slik introspeksjon og den tilhørende “myten om det subjektive” at førstepersonsautoritet blir problematisk. Davidson formulerer derfor førstepersonsautoritet adskillig svakere enn det som er vanlig. Han nøyer seg med å si at førsteperson “speaks with a special weight”.³ Men også dette kaller på forklaring. Det finnes nemlig visse forhold som man skulle tro impliserte at førsteperson *ikke* var privilegert. I lys av disse forholdene kan det til og med fremstå som ekstremt *underlig* at man privilegerer førsteperson.

I følge Davidson har den filosofiske utviklingen resultert i at kunnskap om eget selv fremstår som mer problematisk enn kunnskap om andres selv. Som han sier i innledningen til “Knowing One’s Own Mind”: “There is no secret about the nature of the evidence we use to decide what other people think”.⁴ Vi får vite hva andre tenker gjennom å se hva de gjør, høre hva de sier, vurdere hva de har opplevd osv. Dette gjelder følgelig også andre sin kunnskap om hva *jeg* tenker. Men de fleste er samtidig enige i at jeg vet hva jeg tenker uten å observere min egen adferd. Hvordan er det mulig? Og hvorfor tar da de andre vanligvis mitt ord for hva jeg sier om egne tanker? Påstander med belegg veier normalt sett tyngre enn påstander som

mangler belegg. Likevel veier min påstand om at jeg tror p tyngre enn en annen sin påstand om at jeg tror p , selv om han har observert mine handlinger og ytringer aldri så grundig, mens jeg kanskje bare sier det uten å ha tenkt særlig mye på det tidligere.

Og ikke nok med det. Aksept av slik førstepersonsautoritet synes å gjøre begreper for mentale tilstander radikalt ambivalente. Siden slike begreper anvendes på en bestemt måte om oss selv og på en fullstendig annen måte om andre, må de ikke da rett og slett ha ulik *betydning* alt etter om de står i forbindelse med førsteperson eller tredjeperson? Altså at det du sier om meg når du sier “Han tror p ” betyr noe *helt annet* enn når jeg sier “Jeg tror p ”? Dette synes å lede til en ny form for skeptisisme med hensyn til det mentale, men nå med rot i det problematiske ved mitt *eget* sinn. Følgene er imidlertid identiske med den tradisjonelle skeptisismen om andres sinn: Det åpner seg en avgrunn mellom meg og andre – og ingen form for kunnskap eller påstander kan brolegge gapet. Vi er havnet i et typisk filosofisk dilemma. Det *må* være slik, men det *kan* ikke være slik.

Den vesentlige mangelen ved tidligere behandlinger av dette temaet er i følge Davidson at man verken har maktet eller forsøkt å *forklare* førstepersonsautoritet, dvs. man har ikke forklart *hvorfor* førsteperson veier tyngre enn tredjeperson. Han nevner påstanden om at førsteperson ikke er basert på noen form for belegg, en påstand som han assosierer med Wittgenstein, og hevder at dette, selv om det i og for seg er riktig, ikke kan forklare førstepersonsautoritet fordi påstander som ikke er basert på belegg generelt sett ikke er mer autoritative enn de som er det. Snarere tvert imot. Ut av dette kan vi slutte oss til at Davidson ønsker en forklaring som i en eller annen forstand kan *rettferdiggjøre* førstepersonsautoritet. Hvorfor skal vi eller må vi privilegere førstepersonsutsagn? En slik begrunnelse må følgelig ikke være sirkulær; den må tilbakeføres til noe som ikke forutsetter en slik asymmetri.

Davidsons egen forklaring er lettest å forstå ut fra en situasjon av “radical interpretation”. Hvis A i en slik situasjon forsøker å få B til å forstå hva han sier, må A gjøre seg forståelig eller, som Davidson sier, interpretabel. Det gjør han blant annet ved å ytre gjenkjennelige lydsekvenser på en konsistent måte, dvs. ytre dem i relasjon til (omtrent) de samme objekter og situasjoner hver gang. På den andre siden må B forutsette visse ting for at interpretasjon skal være mulig. I andre skrifter har Davidson argumentert for at B må forutsette både at A sine ytringer har karakter av påstander (“assertions”), dvs. at A med sine ytringer “holder noe for sant” (et uttrykk som må forstås i sin aller videste og minimale betydning), og at disse “påstandene” i stor grad er sanne, dvs. stemmer overens med hva B holder for sant. Dette er bakgrunnen for Davidsons argumentasjon i “First Person Authority” og “Knowing One’s Own Mind”.

Davidsons poeng er at i en slik radikal fortolkningssituasjon så må B også forutsette at *A vet hva han mener med sine ytringer*. Det gir ingen mening å undre på om A kanskje misbruker språket han snakker – ingen andre kan korrigere ham, siden de, *ex hypothesi*, er alene. Hvordan ville det være for B, gitt den forståelsen av radikal fortolkning som Davidson tegner opp, å forkaste denne antagelsen og for eksempel undre på om A kanskje forsøker å villedde han? Det ville innebære at når A sier *p* først i en kontekst og så i en annen, så vet ikke B om han med det mente det samme i de to tilfellene eller om han kanskje heller mente *q* (eller *r*) første gang og *r* (eller *p*) andre gang. Det er lett å se at interpretasjon dermed ville bli umulig.

Denne måten å forklare det på gir imidlertid ikke et helt riktig bilde av hvilken status denne forutsetningen har. For en slik forutsetning er ikke bare nødvendig for at interpretasjon skal være praktisk mulig. Poenget til Davidson er sterkere enn som så. Å ikke gjøre en slik forutsetning vil være ensbetydende med å ikke behandle A som et rasjonelt vesen i det hele tatt. A vil ikke være interpretabel. Men dermed kan hans ytringer heller ikke interpreteres som uttrykk for intensjoner og oppfatninger – i en slik radikal fortolkningssituasjon vet vi i begynnelsen ikke en gang om A har slike (dette er lettere å forstå dersom vi tenker på A som en marsboer som ikke ser ut som et menneske). Følgelig er det inkohærent å “lure på om A kanskje forsøker å villedde”, siden dette forutsetter at A er rasjonell og interpretabel. Også det å forandre meningen med det en sier fra gang til gang forutsetter rasjonalitet. Forutsetningen om at A vet hva han mener med sine ytringer er derfor en nødvendig forutsetning for interpretasjon, og videre for språk og rasjonalitet overhodet.

Hvordan forklarer dette førstepersonsautoritet med hensyn til mentale tilstander? La oss ta utgangspunkt i det begrepet som Davidson baserer sin redegjørelse på, nemlig “belief”. La oss tenke oss en person A som ytrer *p*. Hvis A vet hva han mener med *p*, og også vet at han holder *p* for sant, så følger det at han også vet hva han tror. Men begge disse to fundamentale forutsetningene må i følge Davidson antas i en radikal fortolkningssituasjon. Og siden en slik situasjon bare setter på spissen det som gjelder for all kommunikasjon, må førstepersonsautoritet med hensyn til “beliefs” antas som nødvendig for all interpretasjon, og dermed for all forståelse og rasjonalitet.

2) Hackers kritikk av Davidson

Den dominerende Wittgenstein-fortolkeren Peter Hacker har siden 1990-tallet lagt ut på et korstog mot Davidson sin filosofi. Dette har blant annet resultert i artikkelen “Davidson on First-Person Authority” som kritiserer Davidsons redegjørelse for førstepersonsautoritet på de

fleste viktige og mindre viktige punkter. Noen av disse innvendingene går opp i røyk ved nærmere undersøkelse, da det gjerne viser seg at Hacker tross alt er enig med Davidson. Men på i hvert fall et punkt kan det se ut som om Hacker formulerer en relevant og fundamental kritikk av Davidsons forklaring.

I følge Hacker blir Davidson hengende fast i myten om det subjektive fordi han ikke stiller spørsmål ved en grunnleggende forutsetning i det tradisjonelle, empiristiske bildet av det mentale. Denne forutsetningen er hva Hacker kaller “the cognitive assumption” hos Davidson, dvs. antagelsen om at en person normalt sett *vet* hva han tror. I følge Hacker er dette galt. Det er nemlig meningsløst for en person som tror p å si, “Jeg vet at jeg tror p ”. Dette betyr ikke at personen *ikke* vet hva han tror. Hacker forsvarer en form for førstepersonsautoritet, som vi skal se senere. Poenget er snarere at når jeg tror at p , så verken vet jeg eller ikke vet jeg at jeg tror p , fordi det er nonsens å si både “Jeg vet at jeg tror p ” og “Jeg vet ikke om jeg tror p ”. Dermed kommer ikke Davidsons resonnement av bakken, siden det allerede i problemstillingen baserer seg på at slike formuleringer er meningsfulle.

Det argumentet som Hacker serverer oss her, er en velkjent strategi i Wittgenstein-tradisjonen for å underminere en rekke problemer omkring selvet og sinnet. Men dersom man et øyeblikk forsøker å se bort fra denne tradisjonen og heller se argumentet med nye øyne, virker det svært underlig. For det første fordi det med et pennestrøk fjerner hele den dimensjonen av våre liv som vi fanger i begreper som selverkjennelse, selvrefleksjon og selvbevissthet. For det andre fordi det rett og slett er feilaktig.

Hacker utvider argumentet sitt til å gjelde de fleste påstandsholdninger: Det er for eksempel også nonsens å si “Jeg vet at jeg ønsker p ”. Dessuten utvider han påstanden til å gjelde hele det epistemiske vokabularet, dvs. ikke bare “å vite”, men også “å tvile”, “være sikker på”, “å tro” osv. Eller for å være mer presis, det er nettopp fordi dette epistemiske vokabularet ikke kan anvendes i forbindelse med psykologiske begreper i førsteperson at det ikke gir mening å si “Jeg vet at jeg tror x ”. Når man bruker ordet “å vite” forutsetter man samtidig at det gir mening å anvende relaterte uttrykk som for eksempel “å tvile” og “å ha belegg”, og siden det ikke gir mening å tvile på hva man tror, og heller ikke å anføre belegg for at man tror det man tror, gir det ikke mening å si “Jeg vet at jeg tror p ”. Ulempen med å generalisere poenget slik er imidlertid at faren for moteksempler øker.

Hacker vedgår at det tilsynelatende finnes moteksempler. Han sier at et utsagn som “Jeg vet ikke om jeg tror at p ” normalt er nonsens, men at det også kan være “a slightly odd confession of *indecision*”.⁵ Det finnes også forekomster av manglende selvinnsikt, slik som uoverensstemmelser mellom ord og handling. I slike tilfeller er vi tilbøyelige til å si at taleren

tror at han tror det han sier han tror, men at han egentlig ikke tror det. Hacker regner likevel ikke slike formuleringer som moteksempler på sin generelle påstand om at “the cognitive assumption” er gal. De er nemlig “anomalous” eller “abnormal”.⁶ Og han synes å mene at de er unormale fordi de ikke er epistemiske. Men hva betyr det?

Hvorfor er ikke tilfeller av manglende selvinnsikt genuine moteksempler på Hackers generelle påstand? Hackers svar er at den som ikke har tilstrekkelig selvinnsikt ikke tar feil om hva han selv tror på samme måte som man kan ta feil om hva andre tror, fordi man ikke observerer ens eget indre slik som man observerer hva andre sier og gjør. Og det er den eneste grunnen han gir. Dette er merkelig. Det rammer ikke Davidson, siden han for det første forkaster den myten om det subjektive som begrepet om introspeksjon er en del av, og for det andre trolig vil mene at asymmetrien mellom førsteperson og tredjeperson nettopp gjør at det ikke er samme type feil i de to tilfellene. Hackers forklaring gir heller ingen grunn for å si at slike bruksmåter er unormale på noen måter, siden Hacker mener at ingen andre førstepersonutsagn heller hviler på en indre sans analog med den ytre sans. At disse spesifikke førstepersonsutsagnene ikke gjør det, gir derfor ingen grunn for å kalle dem “unormale”.

Heller ikke tilfeller av “indecision” representerer genuine moteksempler, i følge Hacker. For hvis man sier “Jeg vet ikke om jeg tror p ”, uttrykker man usikkerhet om p , ikke usikkerhet om man tror p . Dette er også merkelig. Det virker naturlig å si at man i et slikt tilfelle er usikker *både* om p og om man tror p , og så lenge vi ikke får servert tungtveiende grunner for å si noe annet, bør vi holde fast ved det. Hacker vil nok også argumentere med at dersom jeg sier “Jeg vet ikke om jeg tror p ”, så betyr ikke det at det finnes et bestemt “trosobjekt” inne i meg som jeg ikke vet hvordan jeg skal beskrive. Men selv om det ofte ville vært *kunstig* å hevde noe slikt, noe Davidson dessuten ikke gjør, så finnes det situasjoner der det ikke er direkte galt eller meningsløst å si det. I en diskusjon kan jeg for eksempel føle at jeg har en helt bestemt oppfatning, men jeg klarer ikke å formulere den helt, og jeg er usikker på om den formuleringen du tilbyr meg, “ p ”, er dekkende. Dette er enda tydeligere ved andre påstandsholdninger: Jeg kan føle at det er noe helt bestemt jeg vil, men jeg vet ikke helt hva. Slike situasjoner er fullstendig “normale”, selv om de kanskje ikke forekommer så *ofte*.

Hacker virker å ignorere i hvilken grad problemer med selverkjennelse og selvinnsikt er sentrale for oss. Man trenger ikke være freudianer for å være skeptisk til tesen om at vi normalt sett vet hva vi tror og føler. Denne feilen gjør også Davidson – han virker å tro at en rask avvisning av relevansen av Freuds begrep om det ubevisste er tilstrekkelig for å rydde bort all tvil om førstepersonsautoritetens generelle gyldighet (en avvisning som dessuten er basert på en kontroversiell lesning av Freuds syn på drømmeinterpretasjon).⁷ Men selv om

slik usikkerhet om egne oppfatninger, følelser, motiver og ønsker har økt i og med moderniteten, og til en viss grad som resultat av Freuds innflytelse, så er slik tvil et kjennetegn ved menneskesinnet heller enn et aparte unntak. Davidson kommer derfor nærmere førstepersonsautoritetens kjerne med sin svakere formulering om at førstepersonsutsagn blir tillagt en annen *vekt* enn de korresponderende tredjepersonsutsagnene.

Forsømmelsen av problemet med selverkjennelse tyder på en ensidig eksempeldiett. I skjemaet “Jeg tror *p*” velger man gjerne for *p* en type helt enkle forhold, fortrinnsvis om visuelt nærværende objekter, som nettopp gjør det vanskelig å tenke seg at personen ikke vet hva han tror om dem, til dels fordi det er så opplagt hva de *er*. Men er slike eksempler representative? Debatten om førstepersonsautoritet ville sett annerledes ut hvis man tok utgangspunkt i følgende eksempler: “Jeg tror at Gud passer på meg”, “Jeg tror at hun vil komme tilbake til meg” eller “Jeg tror at jeg er den personen som jeg tror jeg er”. Og her havarerer Hackers tese om at det er meningsløst å bruke epistemiske operatører på førstepersonsutsagn. I forbindelse med alle disse tre eksemplene kan man tenke seg en som repliserer: “Tror du egentlig det?” Og på det spørsmålet kan svaret være “Ja, jeg vet at jeg tror det” eller “Nei, jeg er ikke sikker – jeg vet ikke hva jeg tror”. Dette blir åpenbart når vi vurderer andre påstandsholdninger. “Jeg tror jeg vil at du skal komme” eller “Jeg vet ikke om jeg egentlig ønsker at hun skal komme tilbake til meg” er typiske eksempler på den refleksjonen, og stundom tvilen, som vi kan ha over våre egne oppfatninger, følelser og ønsker. Og er de ikke snarere mer karakteristiske for menneskesinnet enn å si “Jeg tror at dette er en gul bok” når den gule boken ligger foran en?

I fraværet av noen egentlig begrunnelse står Hacker igjen med gjentatte insisteringer på at tilsynelatende moteksempler er “unormale” bruksmåter. Men når han møter ethvert moteksempel med å si at de bare er et unntak fra regelen, så kan man mistenke ham for rett og slett å være dogmatisk, at henvisningene til hva som er normalt er *ad hoc*. For hva betyr det å klassifisere uttrykk i normale og unormale? Hacker mener det nok ikke i empirisk forstand – han har neppe telt forekomster og funnet ut at vi sjelden sier “Jeg vet at jeg tror *p*” og lignende. Han kunne ha ment at unormale bruksmåter anvender sentrale ord *i en annen betydning* enn de normale. Men denne strategien er ikke brukbar i det foreliggende tilfellene. Ingenting tilsier at man må bruke “vet” i en *spesiell* betydning i uttrykk som “Jeg vet at jeg tror *p*” eller “Jeg vet ikke om jeg tror *p*” – “spesiell” virker her like *ad hoc* som “unormal”. Å snakke om “en annen betydning” virker ikke å bety noe annet enn at det er *forskjeller* mellom førstepersonsutsagn og tredjepersonsutsagn, noe som alle forsvarere av førstepersonsautoritet,

deriblant Davidson, er enige om. Så langt er det derfor ingenting som tilsier at “Jeg vet at jeg tror p ”, “Jeg er usikker på om jeg tror p ” eller “Jeg vet ikke om jeg ønsker p ” er nonsens.

Det er imidlertid en kjerne av sannhet i Hackers argumentasjon: Det er snodig å sette epistemiske operatører foran “Jeg tror at p ” dersom p er et helt opplagt forhold, for eksempel “Det der er en gul bok”. Men dette kan ikke brukes til noen allmenn tese om hva som gir mening og hva som ikke gir mening. For det første fordi det overser hele det feltet vi kaller selverkjennelse. For det andre fordi det er uklart hvorvidt et slikt utsagn er meningsløst eller om det bare er kunstig. Jeg vakler mellom å ville si at “Jeg vet at jeg tror at dette er en gul bok”, sagt i en vanlig situasjon hvor det ikke er spørsmål om noe annet, er kunstig og å ville si at det er meningsløst. Betyr det at jeg, som filosof, må se til å bestemme meg for hva jeg mener eller betyr det at det betyr omtrent det samme? For man skal ikke ta det for gitt at en term som “meningsløs” har en enhetlig betydning, skarpt avgrenset fra beslektede begreper som “kunstig”, “uinformativt”, “irrasjonelt” og “idiotisk”.

Likevel, dette bringer oss på sporet av et dilemma. For i kontekster der det ikke oppstår noe virkelig spørsmål om hva som er tilfelle og hva rasjonelle vesener tror er det ikke bare kunstig med epistemiske operatører på begrepet ’å tro’ – det er kunstig å si “tro” i det hele tatt. I slike situasjoner er kanskje tesen om førstepersonsautoritet korrekt, men til gjengjeld unaturlig, dersom den forstås som en tese om at vi normalt sett vet hva vi tror. Hvis man derimot tar utgangspunkt i påstandshodninger som er grunnet i en kompleks livssituasjon, slik som “Jeg tror at hun kommer tilbake til meg” eller “Jeg vil at hun kommer tilbake til meg”, så virker det, i motsetning til det Hacker sier, fullstendig normalt, ikke bare å si “tro”, men også å bruke epistemiske operatører på egne påstandsholdninger. Til gjengjeld blir førstepersonsautoritet mer tvilsomt. I slike situasjoner, hvor utsagnet er forbundet med en komplisert vev av følelser, tanker, oppfatninger, fortid og fremtid, er det ikke uten videre slik at vi normalt sett vet hva vi tror, ønsker og føler – her mister “normalt sett” sin berettigelse. Med en slik oppfattelse av førstepersonsautoritet synes vi altså å være tvunget til å velge mellom kunstighet og uriktighet. (Nå er det ikke sikkert at Davidson, med sitt filosofisyn, ville regnet dette dilemmaet som særlig betydningsfullt, for det forutsetter at han regner “kunstig” som en filosofisk relevant kritikk. Og det er det ikke sikkert at han gjør, selv om han insisterer på at en forklaring av førstepersonsautoritet må være “naturlig”⁸).

Finnes det en formulering av førstepersonsautoritet som unngår dette dilemmaet? Nå skal vi ikke ta for gitt at det finnes noen slik formulering, for vi skal ikke ta det for gitt at det finnes noe slikt som førstepersonsautoritet, dvs. en rimelig enhetlig egenskap ved psykologiske førstepersonsutsagn generelt sett. Det virker mer fruktbart å begynne med å

karakterisere nærmere den asymmetrien mellom førstepersonsutsagn og tredjepersonsutsagn som førstepersonsautoritet hviler på. For det er et faktum at vi ikke forholder oss til våre egne tanker og følelser slik vi forholder oss til andre sine, og at vi ikke forholder oss til A sine utsagn om A på samme måte som til B sine utsagn om A.

Vi kan begynne med å bemerke det enkle, men vesentlige forhold at det har andre følger dersom førstepersonsutsagn er falske enn det har hvis tredjepersonsutsagn er det. Hvis A tar feil med hensyn til hva A selv tror og føler, så *krever det en annen forklaring* enn hvis B tar feil om A. Ofte består forskjellen simpelthen i at vi krever en forklaring, noe vi normalt sett ikke gjør når andre tar feil om A. Det er dessuten noe annet å tro på hva A sier om seg selv enn å tro på hva B sier om A. Når jeg velger å ikke tro på hva B sier om A sine tanker og følelser, kan det naturligvis være fordi jeg ikke har tillit til B som person, men det kan også simpelthen være fordi jeg ikke fortolker belegget slik han gjør – vi kan rett og slett være uenige. Når jeg derimot ikke tror på hva A sier om sine egne tanker og følelser, blir begrep som tillit, respekt, selvrespekt og autonomi uvegerlig brakt på bane. Dette blir tydelig når vi tenker oss selv i rollen til den som ikke blir trodd. Jeg kan akseptere at andre ikke tror det jeg sier om andre, men når de ikke tror det jeg sier om meg selv, rører det ved min person på en annen og dypere måte – her kan jeg ikke slå meg til ro med å si “Vi er rett og slett uenige”. Dette leder ut i et felt som verken Davidson eller Hacker utforsker nevneverdig, et felt hvor det epistemiske slår over i det etiske og hvor sannhet og kunnskap blir forbundet med tillit og mistillit, ærlighet og uærlighet.

Å slik artikulere de begrepsmessige forbindelsene som danner seg rundt førstepersonsutsagn til forskjell fra tredjepersonsutsagn, som jeg her har antydnet en begynnelse på, er kanskje ikke å si alt som er å si om asymmetrien mellom dem, men det er å si noe som altfor sjelden blir sagt i forbindelse med førstepersonsautoritet. Og det å bemerke slike forbindelser er ikke bare *forberedelser* til den egentlig filosofiske undersøkelsen av det som ligger til grunn for disse forbindelsene – det er disse forbindelsene selv som er kjernen i asymmetrien mellom førsteperson og tredjeperson.

3) Hacker og Wittgenstein

I del II av Wittgensteins *Filosofiske Undersøkelser* finner vi følgende: ““Jeg vet hva jeg vil, ønsker, tror, føler...” (osv. gjennom alle de psykologiske verbene) er enten filosof-sludder, eller i hvert fall *ikke* en apriorisk dom.”” Denne bemerkningen er hyppig sitert i Wittgenstein-litteraturen, som oftest til støtte for et syn som ligner Hackers: Det er sludder/nonsens/meningsløst å si “Jeg vet hva jeg *t*” der *t* er et psykologisk verb. Som sådan

har den tidvis vært brukt til intet mindre enn å løse problemet med skeptisisme om andres sinn. Men som vi har sett, er dette rett og slett feilaktig. Nå kan man spørre seg hva det er med Wittgensteins stil ellers filosofi som gjør at en slik feilaktighet har levd så lenge, men det skal ikke bry oss her. I stedet vil jeg forsøke å vise at denne lesningen er gal.

Det er typisk for Hacker sin argumentasjon, ikke bare i denne artikkelen, men også i hans øvrige Wittgenstein-arbeid, å si at en viss type språklig utsagn er meningsløst, dvs. ikke blir sagt eller kan sies “i dagliglivet”, og at filosofen *derfor* ikke kan bruke lignende type utsagn. Med andre ord, han skiller ikke mellom nonsens i dagligspråket og nonsens i filosofien – dersom noe er nonsens i det første er det det automatisk også i det andre. Og med denne modellen for øye vil man få ut av ovenstående paragraf akkurat det Hacker får ut av den. Men det er nettopp denne modellen som bærer galt av sted. For legg merke til at de to alternativene Wittgenstein gir i paragrafen over begge er av filosofisk art. Wittgenstein sier her ikke noe om hva som gir mening og hva som ikke gjør det i dagligspråket, dvs. slik vi bruker ord og uttrykk i spesifikke kontekster. Det han derimot sier noe om, er hvordan visse formuleringer blir nonsens *i filosofien*. Og det er ingenting som tilsier at slike formuleringer er meningsløse i “dagligspråket”, dvs. *i spesifikke situasjoner*. Det avhenger av konteksten de brukes i, og om vi er i stand til å forstå dem i den konteksten. Og som vi har sett, er det ikke vanskelig å tenke seg sammenhenger hvor de gir mening. Det filosofen stundom forsøker, er imidlertid å bruke dem løsrevet fra enhver spesifikk kontekst – Wittgensteins replikk impliserer at en slik kontekstløshet resulterer i noe annet enn det en ønsker å si. Slik utgjør bemerkningen en kommentar til den personen som i løpet av en filosofisk diskusjon slår seg på brystet og utbryter “Men jeg vet nå i alle fall hva jeg føler selv *nå!*” – en kommentar hvis poeng er at et slikt utbrudd bokser hull i luften, eller i hvert fall ikke makter å uttrykke en apriori dom.¹⁰

En annen nedtegning hos Wittgenstein som ofte siteres som støtte for et slikt argument, og som Hacker siterer i den aktuelle artikkelen, er ved første øyekast vanskeligere å befri fra en tradisjonell fortolkning. For her uttrykker Wittgenstein seg på en mer bombastisk og mindre subtil måte enn vanlig:

Jeg kan vite hva den annen tenker, ikke hva jeg tenker.

Det er riktig å si “Jeg vet hva du tenker” og galt å si: “Jeg vet hva jeg tenker.”

(En hel sky av filosofi kondensert til en liten dråpe grammatikk)¹¹

Nå må det først nevnes at selv om denne bemerkningen fortolkes i den retningen som Hacker vil, så gir den likevel ingen grunn for noen generell påstand. Det vil være helt i tråd med

Wittgensteins vekt på det partikulære å ikke uten videre slutte fra tilfellet med “tenke” til alle psykologiske begreper. For det andre, dersom man tar seriøst Wittgensteins insistering på at han ikke ønsker å komme med bestemte filosofiske påstander eller teser, men at han oppfatter filosofi som en klargjørende aktivitet som virker ved hjelp av grammatiske bemerkninger, så er det heller ingen grunn til å fortolke dette som en tese om språkbruk. Med det mener jeg at en slik bemerkning kan utøve sin *funksjon*, nemlig å fri oss fra visse bilder som holder oss fanget, selv om de ikke er korrekte som generelle beskrivelser av hva vi sier og ikke sier. Begge disse trekkene ved Wittgensteins filosofiske metode synes Hacker å glemme når han transformerer slike bemerkninger til allmenne påstander om hva som gir mening og ikke.

Men det er likevel ikke slike forglemmelser jeg vil fokusere på her. For dersom man ser denne bemerkningen i den sammenhengen den står i, er den langt mer subtil, kompleks og tvetydig enn det bombastiske inntrykket den først gir. Paragrafen den avslutter, er verdt å sitere i sin helhet:

“Et nyfødt barn har ikke tenner.” – “En gås har ikke tenner.” – “En rose har ikke tenner.” – Dette siste – ville man være fristet til å si – er jo innlysende sant! Faktisk sikrere enn at en gås ikke har det. – Og likevel er det ikke så opplagt. For hvor skulle en rose ha tennene sine? Gåsen har ingen i kjevene sine. Og den har naturligvis heller ingen i vingene sine, men det er det ingen som mener når de sier at den ikke har tenner. – Ja, men hva om man sier: Kua tygger føret sitt og gjødsler deretter rosen med det, altså har rosen tenner i munnen på et kreatur. Dette ville ikke være absurd ettersom man på forhånd slett ikke vet hvor på rosen man skulle se etter tenner. ((Sammenheng med ’smerter i den annens kropp’.)¹²

Denne paragrafen står videre som del av en lengre diskusjon om nettopp hva Hacker kaller “the cognitive assumption”. Den foregående paragrafen går slik:

Med dette henger den kjensgjerningen sammen at f.eks. setningen “Jorden har eksistert i millioner av år” har en klarere mening enn denne: “Jorden har eksistert de siste fem minuttene.” For jeg ville spørre den som hevdet dette siste: “Hvilke iakttagelser henviser setningen til; og hvilke ville tale imot den?” – mens jeg vet hvilken tankekrets og hvilke iakttagelser den første setningen er en del av.¹³

Wittgensteins poeng er nok en gang rettet mot den *filosofiske* diskursen. Hans poeng er at et utsagn som “Jeg vet hva jeg tenker (tror, føler osv.)” er beslektet med utsagn som “En rose har ikke tenner” og “Jorden har eksistert de siste fem minuttene”. De er alle av typen man reagerer på med å si først “Naturligvis!”, men deretter “Nonsens!” – fra først å se dem som selvinnlýsende oppdager man at man egentlig ikke vet hva de sier, fordi man ikke vet hvilken kontekst man skal sette dem inn i. Man går med andre ord fra å oppfatte dem som dype, apriori utsagn til å oppfatte dem som filosofisk nonsens. Men Wittgenstein sier ikke at de *per se* er nonsens. Dersom man gir dem en passende kontekst kan de være så meningsfulle som bare det. Ingen *setninger* er i og for seg nonsens, men i visse kontekster gir de mening og i

andre ikke. Problemet som gjerne oppstår i filosofien er at man bruker utsagn som trenger å få angitt en kontekst for å være meningsfull uten å angi en slik kontekst – *dermed* blir resultatet nonsens. Visse utsagn gir mening uten å få indikert en kontekst, ikke fordi de gir mening uavhengig av kontekst, men fordi de liksom bærer sin kontekst med seg, dvs. vi “vet hvilken tankekrets og hvilke iakttagelser” de er en del av (en formulering som viser at konteksten naturligvis ikke trenger å være noen praktisk kontekst). Dette gjelder utsagnene “En gås har ikke tenner”, “Jorden har eksistert i millioner av år” og, vil jeg hevde, “Jeg vet hva du tror”, i motsetning til “En rose har ikke tenner”, “Jorden har eksistert i de siste 5 minuttene” og “Jeg vet hva jeg tror”, som trenger å få spesifisert en sammenheng for å gi mening. I dette perspektivet forstår vi hva Wittgenstein mente med å si hans mål i filosofien er å gå fra en forkledd til en åpenbar meningsløshet.¹⁴ Man går fra å oppfatte et utsagn i filosofien som uttrykk for et dypt, apriori forhold til å se det som noe som i kraft av sin kontekstløshet går på tomgang. Wittgensteins bemerkninger er nettopp ment som replikker som skal ta oss den veien.

Slike bemerkninger forsøker å komme inn *forut for* det filosofiske problemet, ved å ta fra en trangen til å stille spørsmålet, dvs. ved å ta bort problemet. Et slik prosjekt er dypt utilfredsstillende for mange. For som de vil si, problemet er jo der selv om vi ikke reiser det: førsteperson og tredjeperson er der og eksisterer i et eller annet forhold selv om vi ikke plages med det. Med andre ord, i stedet for å feie problemet under teppet bør man forsøke å løse det, dvs. finne ut hva forholdet *egentlig er* mellom førsteperson og tredjeperson, det indre og det ytre. De impliserer altså at det er et eller annet vesentlig vi ikke vet om dette forholdet. Men Wittgensteins poeng, tror jeg, er at vi allerede vet alt som er å vite om dette forholdet. Det er bare å se på det vi sier og gjør (som ikke er så lett som det høres ut som).

Man kunne rette Hacker sin kritikk av det kartesianske bildet av det indre mot hans eget bilde av grammatikken: Han snakker om en grammatisk struktur som om den var et visuelt foreliggende objekt, om enn et veldig komplisert objekt, som man kan beskrive, og der beskrivelsene er sanne eller falske alt etter om de samsvarer med denne objektivt gitte strukturen. Men Wittgensteins grammatiske bemerkninger er ikke bemerkninger *om* “grammatikken”, de er ikke om en struktur som ligger i eller under språket og utgjør rammene for hva som er mulig å si og sikrer betydningen til det vi sier. De grammatiske bemerkningene er bemerkninger om hva vi sier og gjør som ikke skal vise at visse oppfatninger gale, men å hindre at de problemene som utløser disse oppfatningene oppstår, ved å vise at alt vi trenger allerede ligger der foran oss, det vi sier og gjør.

I dette perspektivet trer likhetene mellom Hacker og Davidson tydeligere frem. De er enige om at “det kartesianske bildet” eller “myten om det subjektive” må forkastes, uten at man dermed skal omfavne noen enkel form for behaviorisme eller materialisme. De er dessuten enige om alle fakta omkring førstepersonsautoritet: at personen selv normalt sett er en autoritet med hensyn til hva han tror, tenker og føler, at det finnes unntak fra denne regelen, at tredjepersonsutsagn er basert på hva personen sier og gjør, at førstepersonsutsagn ikke er basert på belegg osv. Men de er også enige om at dette ikke bare er fakta, at førstepersonsautoritet på en eller annen måte er et nødvendig trekk ved vårt språk og forståelse, om enn de er uenige om den filosofiske forklaringen av denne nødvendigheten. Slik sett deler de antagelsen om at det virkelig filosofiske arbeidet begynner *etter* at man har samlet sammen de relevante fakta. For Hacker ved å vise at de uttrykker logisk-grammatiske regler. For Davidson ved å vise at de er nødvendige forutsetninger for interpretasjon. Dermed står de begge langt fra Wittgenstein, tror jeg. Det virkelig filosofiske arbeidet er snarere å vinne en innsikt i hvordan arbeidet er over når innsamlingen av fakta er over. Og å kalle dem “fakta” er én måte å uttrykke en slik innsikt på.

Abstract: First person authority has become a main philosophical topic the last decades. The current paper examines Peter Hacker’s criticism of two classic articles by Donald Davidson on first person authority. It also discusses Hacker’s use of some remarks from Wittgenstein in the course of this criticism. It is argued, first, that both Hacker and Davidson disregard how questions of self-knowledge are both central to and pervasive in human life. Second, Hacker’s position leads to a fairly typical philosophical quandary: we run into a position where we are forced to choose between the false and the weird (moreover, it just this hovering between falsity and weirdness that creates the illusion of substantial truth in Hacker’s view). Third, Hacker and Davidson have more in common than Hacker acknowledges, but the features that unite them separate them from Wittgenstein’s view of philosophy.

Key words: first person authority, self-knowledge, Hacker, Davidson, Wittgenstein

Steinar Bøyum holds a doctorate in philosophy on the concept of philosophical education. Recent publications include “Philosophical Experience” (*Metaphilosophy* 3/2008) and “Words from Nowhere – Limits of Criticism” (*Philosophical Investigations* 2/2008).

¹ Davidson, Donald: "First Person Authority", *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*, Oxford: Clarendon Press 2001, s. 3.

² *ibid.*

³ *ibid.*, s. 5.

⁴ Davidson, Donald: "Knowing One's Own Mind", *Subjective, Intersubjective, Objective: Philosophical Essays Volume 3*, Oxford: Clarendon Press 2001, s. 15.

⁵ Hacker, P. M. S.: "Davidson on First-Person Authority," *The Philosophical Quarterly* 47, no. 188 (July 1997), s. 291.

⁶ *ibid.* s. 291, s. 293.

⁷ Davidson 2001, s. 7.

⁸ Davidson 2001, s. 9.

⁹ Wittgenstein, Ludwig: *Filosofske Undersøkelser*. Oversatt av Mikkel B. Tin. Oslo: Pax Forlag 1997, s. 251. Jf. §246.

¹⁰ Jf. *ibid.*, §253.

¹¹ *Ibid.*, s. 251-252.

¹² *Ibid.* s. 251.

¹³ *ibid.*

¹⁴ *ibid.*, §464.

PREPRINT