

31.3.16

Den 4. erfaringskonferansen om kunnskapsbasert praksis i helse og omsorgstjenesten. Om erfaringskunnskap og implementering

Grunner til at klinisk skjønn ofte ikke blir bedre med erfaring

Geir Kirkebøen
Psykologisk institutt
UiO



De neste 45 minuttene 😊

- Om forskning på faglig skjønn og intuisjon i beslutningspsykologien.
- Grunner til at klinisk skjønn (ofte) ikke blir bedre med erfaring.
- Implikasjoner for praksis?

Klinisk skjønnsutøvelse

“Er mor eller far best egnet som forelder?”

“Vil han begå nye voldshandlinger?”

“Har løsemidlene påført henne en hjerneskade?”



Må typisk vurdere hva flere usikre indikatorer (tegn, opplysninger etc.) samlet indikerer.

Skjønnsmessig vurdering:

Vurdering basert «bare» på egen kunnskap og erfaring, uten bruk av hjelpemidler.

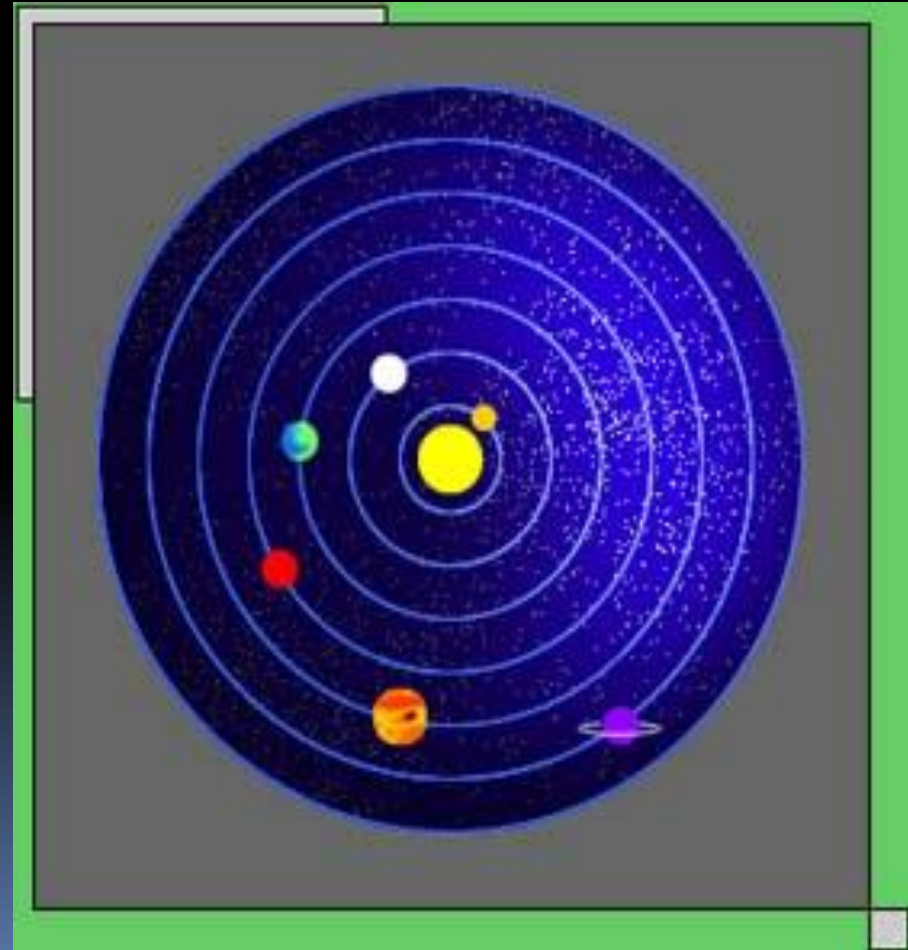
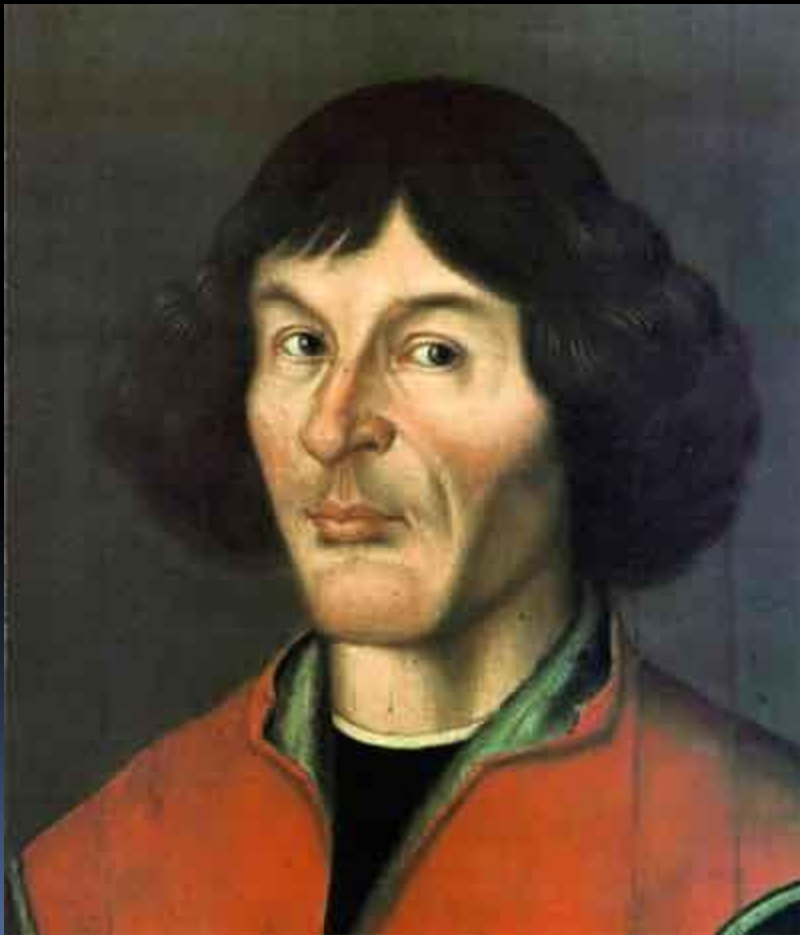
«DET FJERDE SLAGET»: FUNN FRA FORSKNINGEN PÅ FAGFOLKS SKJØNN



De tre første – i følge Freud - «slagene» mot menneskets selvfølelse ...

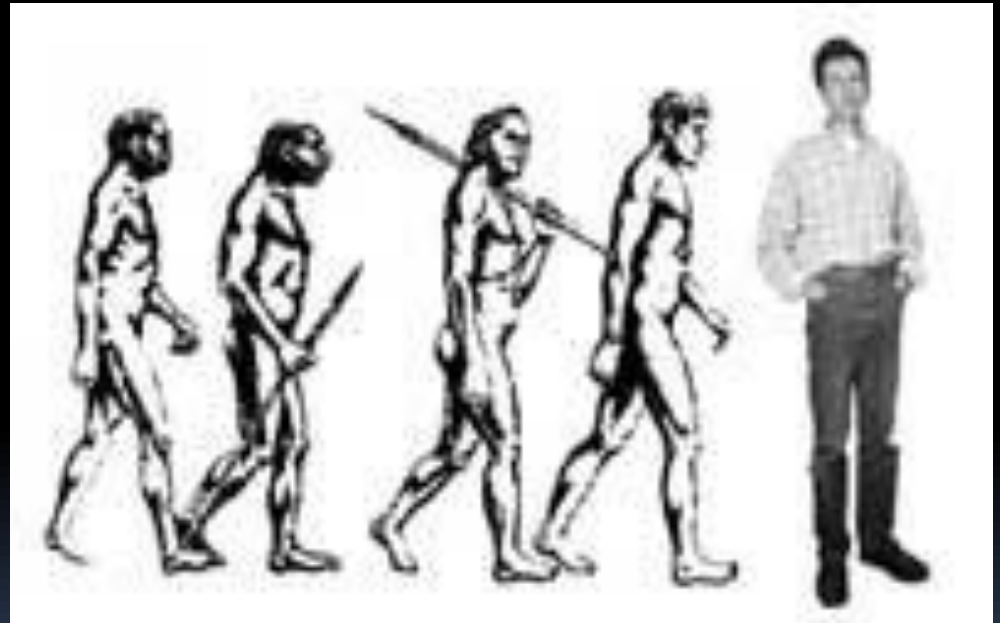
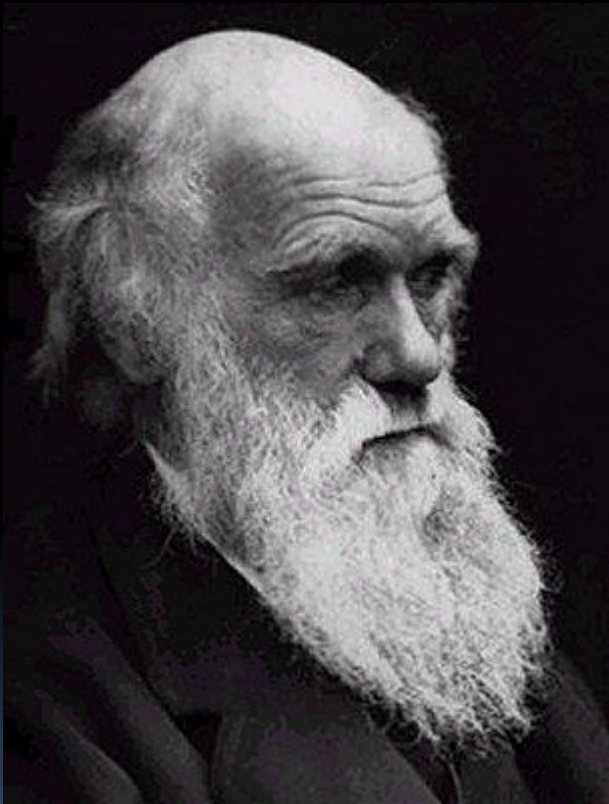
Det første slaget:

”Vi befinner oss ikke i universets sentrum”
(Nicolaus Copernicus, 1543)



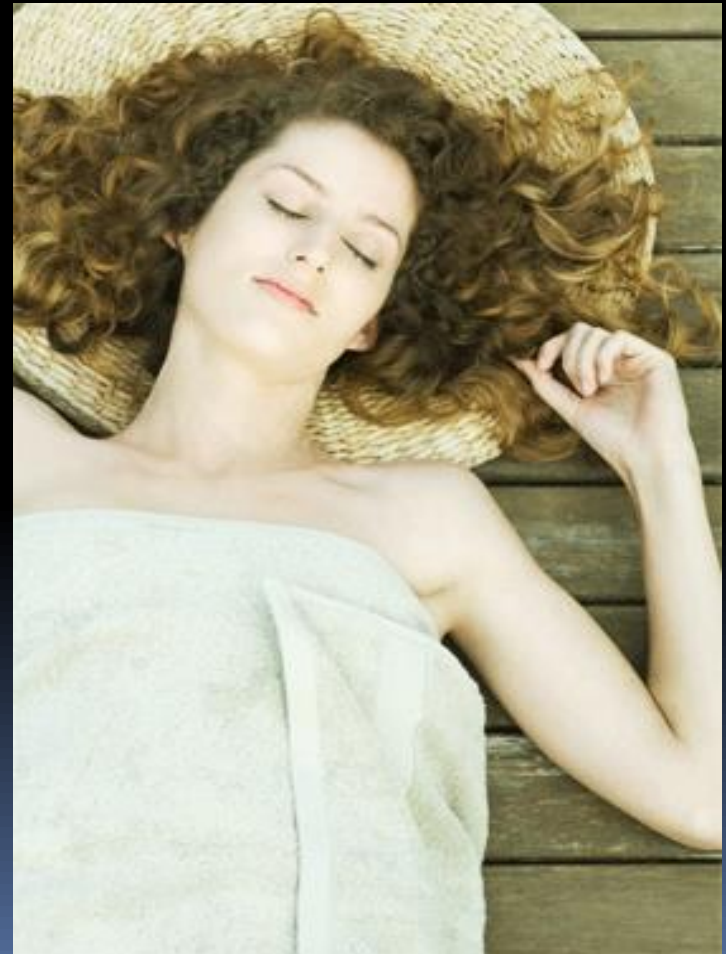
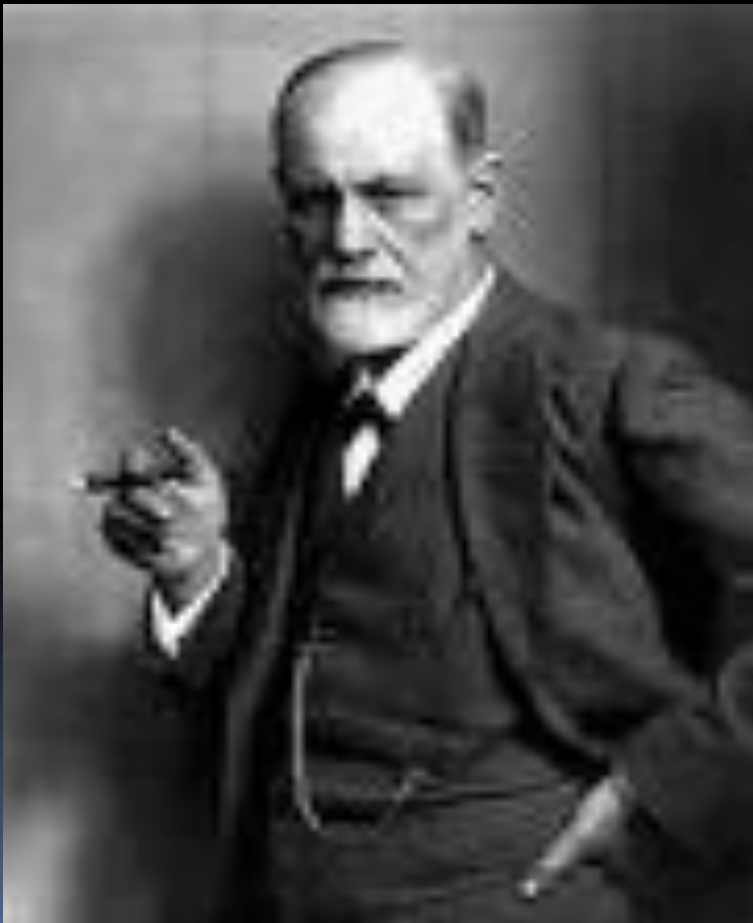
Det andre slaget:

”Det er intet skarpt skille mellom oss og dyr” (Charles R. Darwin, 1859)



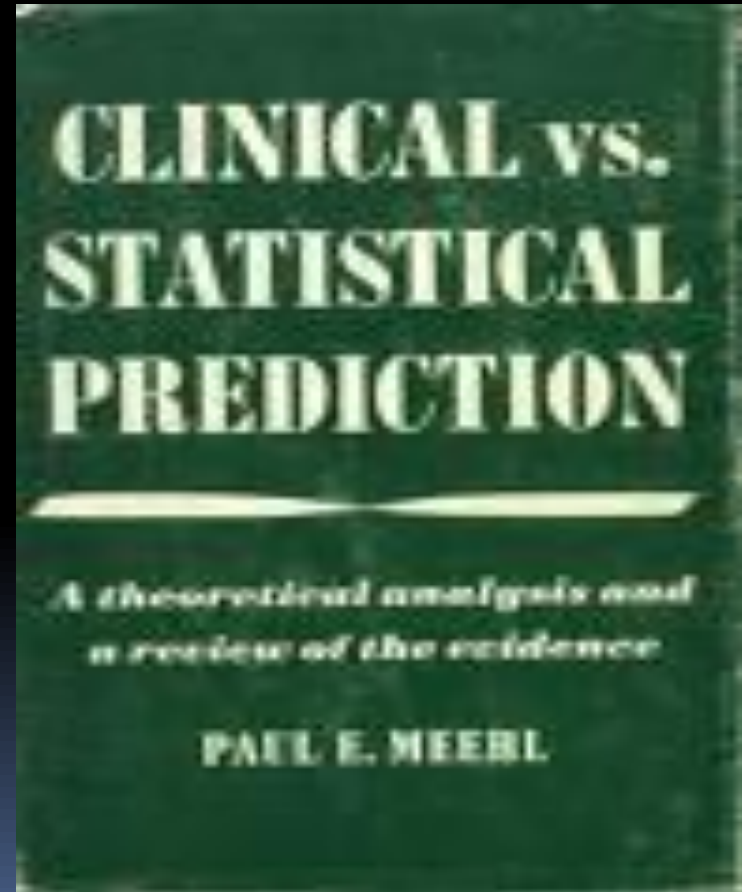
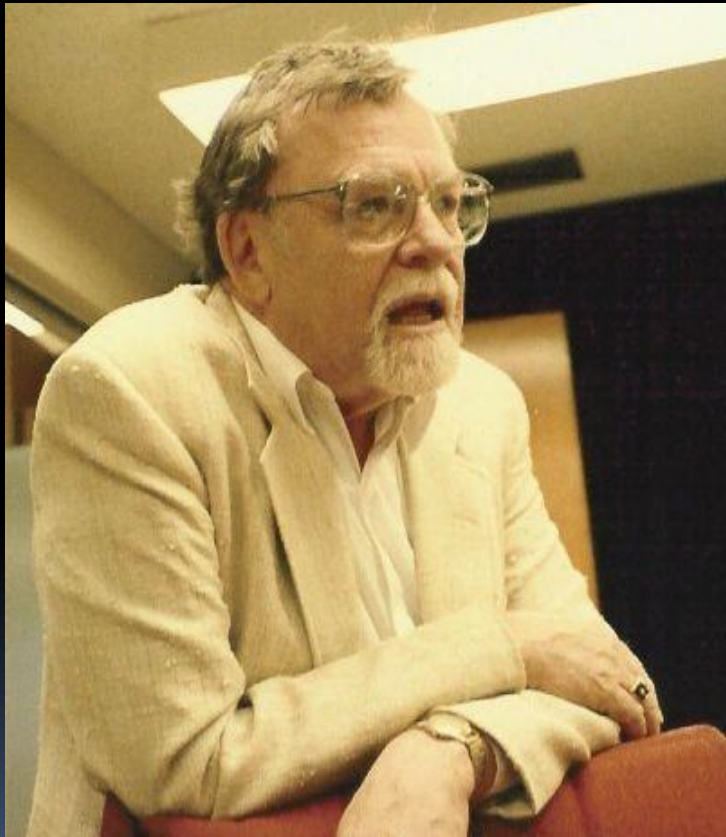
Det tredje slaget:

”Vi er ikke ’herrer i eget hus’, men styres av ubevisste krefter” (Sigmund Freud, 1900)



«Det fjerde slaget»:

Selv når vi ikke herjes med av ubevisste krefter, så er selv erfarne fagfolks skjønn sterkt begrenset.



Paul Everett Meehl (1920-2003)

Fagfolks skjønn - hovedfunn i Meehl- tradisjonen

Hovedfunn 1:

Erfarne profesjonsutøveres skjønnsmessige vurderinger overgås systematisk av helt enkle formler.

Hovedfunn 2:

Skjønnsmessige vurderingers nøyaktighet bedres (ofte) ikke med erfaring.

Funn 1 og 2 gjelder i særdeleshet «probabilistiske» kunnskapsområder som kliniske fag gjerne er.

SKJØNN VS. FORMLER: NOEN STUDIER

- Skjønn: Vurdering av data “bare” vha. hodet.
- Enkel formel: (f.eks.) $Y = k_0 + k_1 * x_1 + \dots + k_n * x_n$

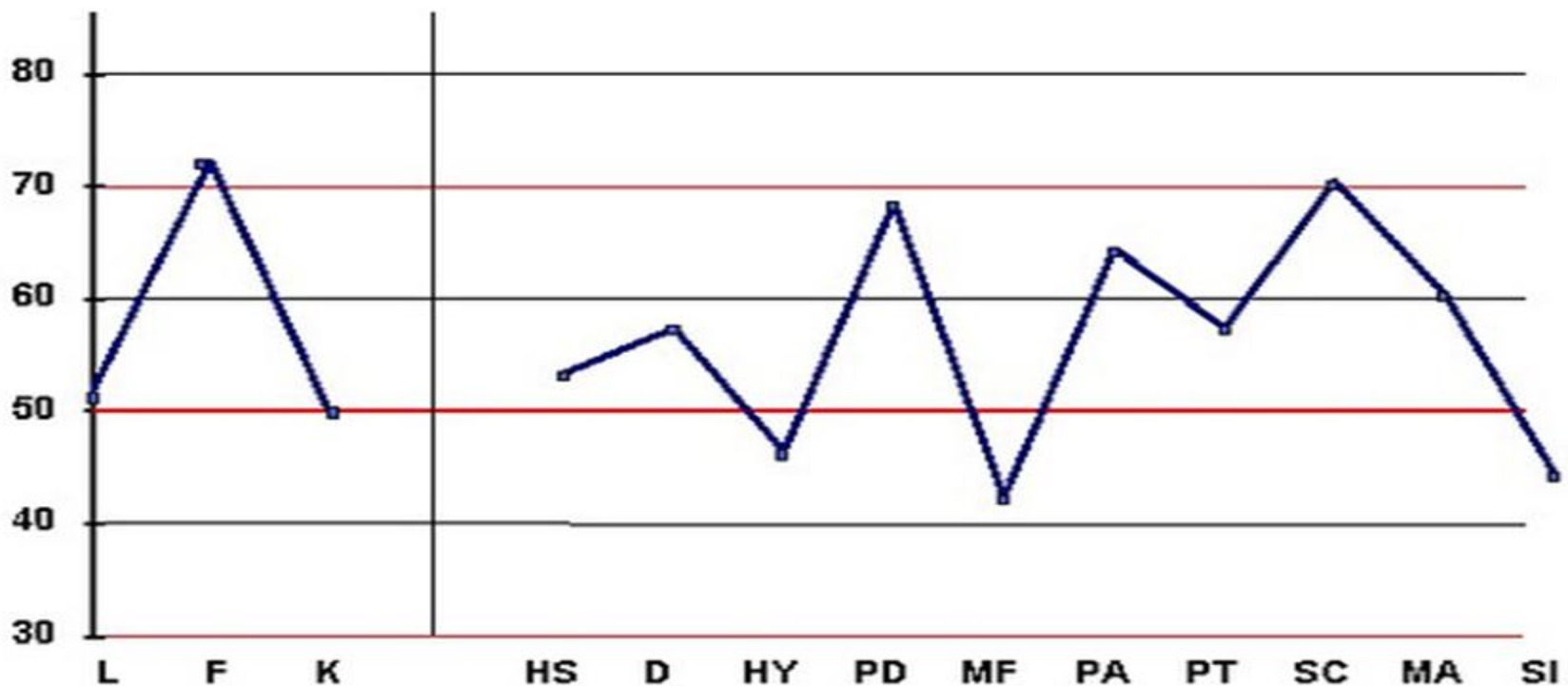
Hovedinnvending 1) mot Meehl (1954):

Fagfolks skjønn vil komme bedre ut om det kreves
«helhetlig» vurdering av data.

Eksempel 1:

Er denne pasienten nevrotisk eller psykotisk?
(Goldberg, 1965)

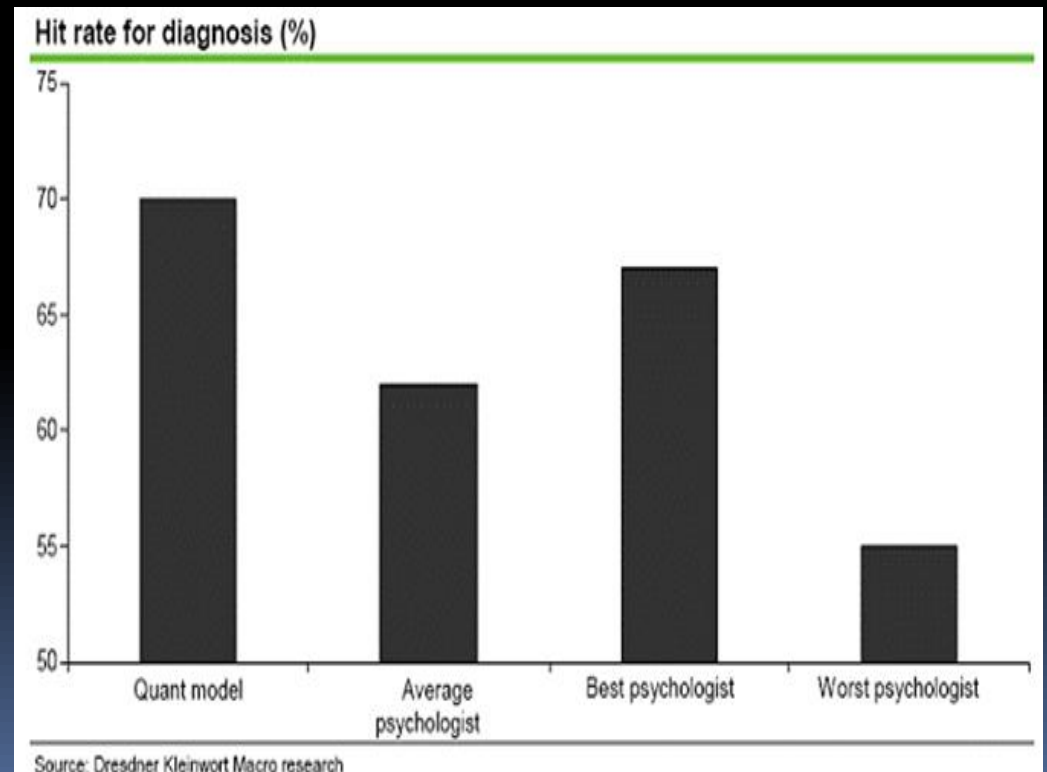
MMPI profil



Goldberg, 1965, forts

- Data: 861 MMPI-protokoller, dvs. 11 tallverdier
- Sammenligning: "Goldbergs regel" vs. 29 klinikere
- Resultat:

% korrekte vurderinger:
Snitt klinikere 62%
Beste kliniker: 67%
"Goldbergs regel": 70 %



Hvordan ser Goldbergs regel ut?

$$\begin{aligned} & \text{skala-1} + \text{skala-2} + \text{skala-3} \\ - & (\text{skala-4} + \text{skala-5}) \\ & \geq 45 \text{ ---> Psykotisk} \\ & < 45 \text{ ---> Nevrotisk} \end{aligned}$$

Så, klinisk skjønn kom dårligere ut enn en helt enkel formel selv når erfarne fagfolk hevdet at data måtte vurderes helhetlig.

Hovedinnvending 2) mot Meehl (1954):

Skjønnnet vil komme bedre ut enn formelen om fagfolk får tilgang til ”naturlig” informasjon.

Gjør tilgang på ”naturlig” informasjon
ekspertskjønnnet bedre?

Dawes, R. M. (1971). A case study of graduate admissions. *American Psychologist*, 26, 180-88.

Eksempel 2:

Hvem skal tas opp på college?

(Dawes, 1971)



Sammenligning:

Inntakskomiteer (m/ erfarne skolefolk) vs. enkle formler.

Data: Komiteen har mye mer info om elevene enn formlene og foretar bl.a. et lengre intervju med dem.

Dawes, 1971:

(1) $0.0006 \text{ GRE} + 0.76 \text{ GPA} + 0.2518 \text{ QI}$ forutså (**predikerte**) bedre enn komitéen hvordan studentene ville gjøre det på college.

(2) Bare GPA var tilstrekkelig for å gjøre det bedre enn inntakskomiteén.

(3) En enkel formel eliminerte 55% av søkerne som ble vurdert og avvist av komiteen, uten å eliminere en eneste søker som ble akseptert av komiteen. Ved å bruke formelen ville man spare 18 millioner dollar (i 1971) årlig i USA. (I 2005, anslagsvis 500 millioner dollar årlig, i følge Bazerman, 2006)

Eksempel 3:

Hvor god/dyr blir årets Bordeaux-vin etter ti års lagring?



Ashenfelter O, Ashmore, D. & LaLonde R. (1995)
Bordeaux wine vintage quality and the weather. *Chance*
8, 7 – 14

Vurderingene til de fremste vineksperter som tar
«alt» i betraktning: temperatur og nedbør, smak og
lukt etc.

... utkonkurreres av "Bordeaux-likningen":

$$\begin{aligned} & 0.616 && * \text{TempVekst} \\ - & 0.00386 && * \text{RegnInnhøst} \\ + & 0.00173 && * \text{RegnVinter} \\ + & 0.0238 && * \text{ÅrsLagring} \\ - & 12.145 && \\ = & \ln(\text{pris på vin}) && \end{aligned}$$

Vintips!

“Subject: Re: Bordeaux equation

From: Orley C Ashenfelter <c6789@Princeton.EDU>

To: geir.kirkeboen@iln.uio.no

Many thanks for the email.

(...)

Have fun! And if you can find some 1997 red Bordeaux around, it is both good and under-priced (critics speak poorly of it, but it sure tastes good, and the weather tells us it is too!)”

Eksempelstudie 4:

Hvor lenge vil denne pasienten med Hodgins' sykdom leve? (Einhorn, 1972)

Data (Predikatorer): Ni biopsi-trekk pekt ut av legene.

Funn:

- 1) Legene var ikke i stand til å predikere levetid basert på pasientens ni biopsi-trekk.
- 2) En regresjonslikning basert på de samme ni biopsi-data kunne (i en viss grad) predikere levetid.

Også «uekte» modeller:

- . «bootstrap»-modeller
- . modeller med enhetsvekter
- ...

overgår (systematisk) fagfolks helhetsvurderinger

Alt i alt:

“Eksperters skjønnsmessige vurderinger har vist seg å være dårligere enn selv de enkleste modeller på praktisk talt alle fagområder som er blitt studert”
(Camerer & Johnson, 1991, s. 203).

HVORFOR ...

- ... utkonkurreres fagfolks skjønn av helt enkle formler?
- ... blir (ofte) ikke det faglige skjønnnet bedre med erfaring?

«Professionals are not stupid, but they are human» (Bazerman, 2006)

- Det er i vår komplekse verden (ofte) både en fordel og en nødvendighet:

- 1) å forenkle

- 2) å la seg lett påvirke

Men:

- 1) og 2) leder også til **systematiske** feil i (fag)folks skjønnsmessige bedømminger og beslutninger.

Fagfolk forenkler: Klassisk studie

«Blant kvinner på 40 år har 1% brystkreft. Dersom en kvinne har brystkreft, så er sannsynligheten 80% for at hun vil teste positivt på mammografi. Dersom en kvinne ikke har brystkreft, så er sannsynligheten 10% for at hun likevel vil teste positivt. Hva er sannsynligheten for at en kvinne på 40 år som tester positivt, faktisk har brystkreft?» (Eddy, 1982)

ALTSÅ:

- 1% (av kvinner på 40 år) har brystkreft.
- 80% som har brystkreft tester positivt.
- 10% som ikke har brystkreft tester likevel positivt.

Hva er sannsynligheten for at en kvinne på 40 år som tester positivt, faktisk har brystkreft?

Legenes svar (95%): c. 70-80%

Korrekt svar: c. 7.5 %

Hvorfor tar velutdannede, intelligente og erfarne leger så fryktelig feil?

Skjønnsutøvelse krever resonnering med usikkerhet

- Normen: Sannsynlighetsregningens regler

- . Dels ukjente, selv for fagfolk

- . Komplekse å anvende skjønnsmessig



- (Fag)folk håndterer usikkerhet med noen enkle heuristikker (Tversky & Kahneman, 1974)

- Representativitetsheuristikk

- Tilgjengelighetsheuristikk

- Affektheuristikk

- Bruk av disse heuristikkene har felles at man (ubevisst) erstatter et komplisert spørsmål med et enklere, og svarer på dette siste i stedet.

Bruk av representativitetsheuristikk i vurderingen av mammografitester

Legene erstattet (ubevisst) det kompliserte spørsmålet:

«Hva er sannsynligheten for brystkreft ved positiv test?»

med det langt enklere:

«Hvor representativt (typisk, sannsynlig) er det at en som har brystkreft tester positivt?»

Legenes oppgave var å bestemme den **betingede sannsynligheten**:

$$P(\text{Brystkreft} | \text{Positiv test})$$

De burde brukt Bayes' formel:

$$P(\text{BK} | \text{T+}) = P(\text{T+} | \text{BK}) * \frac{P(\text{BK})}{P(\text{T+})}$$

Også fagfolk lar seg lett påvirke

i) Tyske dommere, alle med mer enn 15 års dommererfaring, fikk lese en beskrivelse av en kvinne tatt for butikktysterier.

To grupper: G1 & G2

ii) Hver dommer fikk se en terning bli slått på bordet foran seg. Terningen viste for dommerne i G1 og G2 hhv. 3 og 9.

iii) Spm 1: «Vil du gi kvinnen en fengselsstraff kortere eller lengre enn månedene gitt av tallet på terningen?»

iv) Spm. 2: «Hvor mange måneders straff vil du gi henne?»

Snittsvar (på spm. 2) i G1: 5 måneder .

Snittsvar (på spm. 2) i G2: 8 måneder.

(Englich, Mussweiler, & Strack, 2006)

Lineære modeller kommer bedre ut, fordi:

- Fagfolk 'ofre' for:
 - . Beslutningsskjevheter (forenkler, lar seg påvirke etc.)
 - . Omstendighetene (umotiverte, trøtte, uoppmerksomme etc.)
- Fagfolks 'helhetsvurderinger' er gjerne nettopp ikke det:
 - . Bruker sjelden mer enn 2-3 data
 - . Vurderer typisk ikke data helhetlig (i forhold til andre data)
- Lineære modeller (formler)
 - . «Helhetsvurderer» heller ikke data, men ...
 - . Gir hvert datum den vekt det (isolert sett) skal ha
 - . Konsistent integrering ('vurdering') av data

EKSEPSJONELT FAGLIG SKJØNN: EKSEMPLER

Intuisjon

En tankeprosess som gir et svar, en løsning, eller en idé, uten anstrengelse og uten bevissthet om prosessen bak.



Capablanca (1908): "Bare ett, det riktige"



Også erfarne brannfolk 'ser' (i blant) intuitivt hva som bør gjøres (Klein et al., 1986)

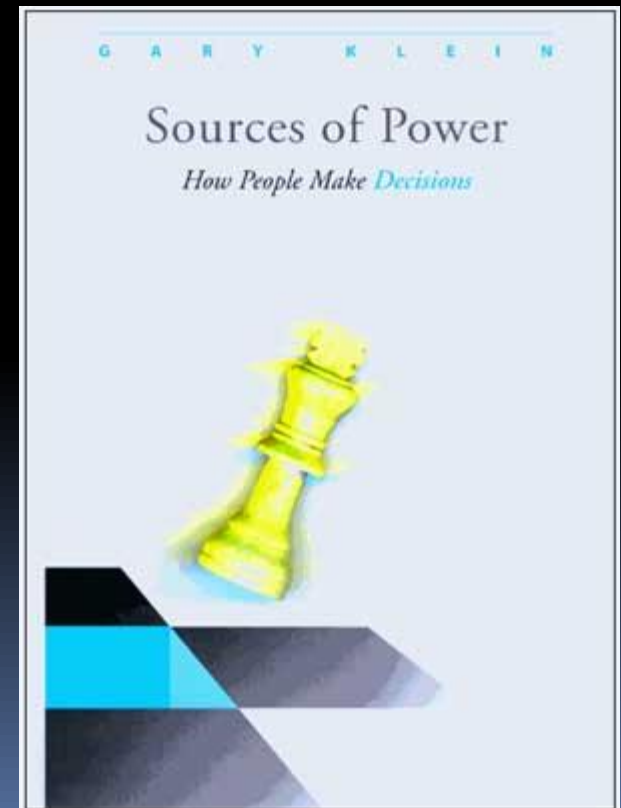
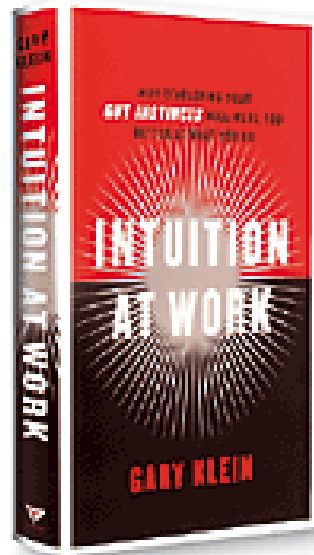


Sykepleiere kunne 'føle' at barn var i ferd med å utvikle livstruende infeksjoner.

(Crandall & Getchell-Reiter, 1993)



- De bare føler hva de bør gjøre, hva som er galt etc..
- De kan (ofte) ikke forklare sine intuisjoner



**FAGFOLK SOM IKKE BØR STOLE PÅ
SIN INTUISJON:
EKSEMPLER**



Aksjeanalytikere gjør det ikke bedre enn lekfolk
(f.eks. Törngren & Montgomery, 2004)

Politiske analytikere



"Vil det bryte ut åpen krig mellom Sør- og Nord-Korea i 2016?";

«Vil Eurosonen endres innen 2018?» etc.

284 analytikere

82361 intuitive forutsigelser,
over 20 år.

Politiske analytikere gjør det ikke bedre enn utrenede avislesere.

Fagfolks intuisjon

- Kvaliteten varierer voldsomt fra fag til fag.
- Erfaring og intelligens er ikke tilstrekkelig for å utvikle god intuisjon.
- Liten sammenheng mellom kunnskapsnivå og kvalitet på intuitive vurderinger/valg.



Når – under hvilke betingelser –
utvikler (fag)folk en intuisjon vi
kan stole på?

Når kan vi stole på fagfolks intuisjoner?

Ekspertintuisjon:

«Situasjonen gir ett tegn. Tegnet gir eksperten tilgang til informasjon lagret i hukommelsen, og denne informasjonen gir svaret. **Intuisjon er gjenkjenning, verken mer eller mindre.**” (Simon, 1992)



Det må på fagområdet **finnes** gode tegn å basere
intuisjonen på.

Intuisjoner basert på gode tegn

- Sjakk: «Sh4. Det er trekket!»
- Brannslukning: «Kom dere ut av kjøkkenet!»
- Sykepleie: «Det barnet har en alvorlig infeksjon!»



Intuisjoner basert på svært usikre tegn

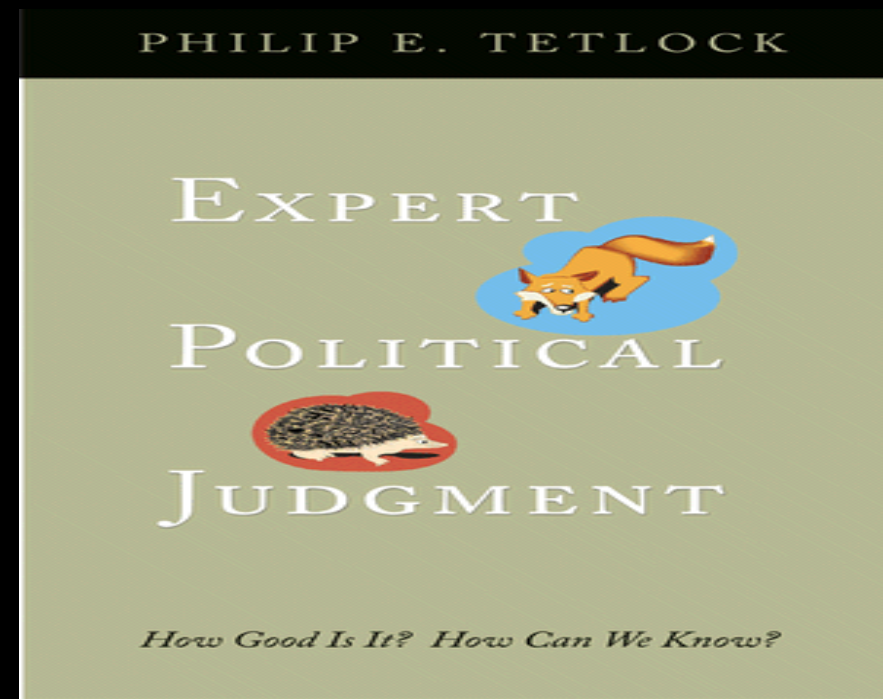
Politiske/økonomiske trender:

«Vil det bryte ut åpen krig mellom Sør- og Nord-Korea i 2016?»

«Vil Eurosonen endres innen 2018?»

Aksjeprediksjon:

«Vil den aksjen stige mer enn børsen de neste tre månedene?»



Karakteristika ved fagområdet er avgjørende

- Forutsetter stor grad av regularitet mellom det som skal vurderes og grunnlaget for vurderingene (tegn).
- Da vil det gjerne finnes gode - dvs. relativt sikre - tegn å basere intuisjonen på.
- Om det fins gode tegn, så har fagfolk en tendens til å finne fram til disse og utvikle god faglig intuisjon gjennom erfaring.

HVORFOR BEDRES OFTE IKKE NØYAKTIGHETEN AV PROFESJONSUTØVERES BEDØMMINGER OG BESLUTNINGER MED ERFARING?

AT det har seg slik:

“I praktisk talt alle studier av eksperter utført innen bedømmings- og beslutningspsykologien, så har erfaring vist seg å være urelatert til den empiriske nøyaktigheten av ekspertenes vurderinger.” (Hammond, 1996, *Human judgment and social policy*, s. 278)

”Klinisk erfaring [i psykodiagnostikk, atferdsprediksjon og nevropsykologi] har generelt vist seg ikke å være relatert til kliniske vurderingers gyldighet, verken når erfarne klinikere er blitt sammenlignet med uerfarne klinikere, eller når klinikere er blitt sammenlignet med studenter”
(Garb, 1998, *Studying the clinician*, s. 110).

Problemet: Usikkerhet

- Gode (intuitive) tegn er (relativt) sikre tegn (jf. ovenfor).
- Hovedproblemene usikkerhet medfører :
 1. Å bruke (resonnere med) usikre tegn (jf. ovenfor).
 2. Å lære fra erfaring hvilken (prediktiv) verdi enkelttegn har.
 3. Tilbakemeldingen man får på vurderinger og valg blir mangelfull og misvisende.

PROBLEMER MED Å LÆRE ENKEL SAMVARIASJON FRA ERFARING

	X	Ikke X
Y	A	B
Ikke Y	C	D

Hvilken samvariasjon er det mellom X og Y?
I hvilken grad indikerer X at Y er tilstede?

Må kjenne alle hyppighetene, A, B, C og D for å bestemme dette.

	Symptom	Ikke Symptom
Sykdom	A (37)	B (17)
Ikke sykdom	C (33)	D (13)

Smedslund, J. (1963). The concept of correlation in adults.
Scandinavian Journal of Psychology, 4, 165-173.

Spm.: «Er det en positiv relasjon mellom symptom og sykdom her: Indikerer symptomet sykdommen?»

JA : 86%

NEI: 7%

VET IKKE: 7%

Hvorfor svarer flertallet feil?

	Symptom	Ikke Symptom
Sykdom	A (37)	B (17)
Ikke sykdom	C (33)	D (13)

Problem 1:

Selv når man har tilgang til alle de fire hyppighetene (A – D) så formes våre oppfatninger om samvariasjon i all hovedsak av hyppigheter av type A, dvs. av hvor hyppig symptomet og sykdommen forekommer samtidig (Smedslund, 1963)

Erfaren 'voldsterapeut' i Dagbla':

“Basert på min lange erfaring kan jeg si at ingen menn som har slått/mishandlet sin samboer én gang, slutter med det uten profesjonell hjelp”

Hvorfor kan terapeuten vanskelig ha lært dette fra sin egen terapeutiske praksis?

	Terapi	Ikke terapi
Slutter å slå	A	B
Slutter ikke å slå	C	D

Og tilsvarende:

Kliniske psykologer erfarer at mange med psykiske problemer har hatt store barndomstraumer, men de erfarer ikke i sin praksis at det også er mange med store barndomstraumer som ikke har psykiske problemer (f.eks. Paris, 2000).

Problem 2:

I mange profesjonelle praksiser utelukker praksisen selv tilgang til alle de fire hyppighetene (A- D) som må kjennes for at sammenhengen mellom variablene skal kunne læres.

Problem 3:

Hukommelsen er selektiv. Hvordan vi husker (eller rekonstruerer) at to variabler har samvariert (altså basert på egen erfaring) er av mange grunner lite pålitelig. Vår hukommelse er (ofte) svært selektiv.

Problem 4:

Illusoriske korrelasjoner. Vi har en sterk tendens til å "se" i et erfaringsmateriale de sammenhenger som vi forventer å finne, ikke de sammenhenger som faktisk ligger i materialet.



Svært vanskelig å lære fra erfaring hvilken prediktiv verdi enkeltdata har.

Betingelser for effektiv erfaringslæring

1. Umiddelbar, utvetydig og konsistent tilbakemelding **når** man tar/gjør feil.
2. Tilbakemeldingen må gi en klar forståelse av **hva** som er feilen.
3. Feil må ha kostnader.

Hva blir vi (raskt) bedre til med erfaring ...

- 1) Sensomotoriske ferdigheter.
- 2) Ferdigheter vi utøver i "deterministiske" læringsomgivelser.

I **deterministiske** læringsomgivelser gir samme (type) valg/bedømming samme (type) tilbakemelding hver gang.

I **probabilistiske** læringsomgivelser derimot, kan samme type bedømming/valg gi **forskjellig** (type) tilbakemelding fra gang til gang.

Vanskeligheter med å lære av erfaring i probabilistiske læringsomgivelser.

Eksempel 1:

- Terapeut 1 (T1) velger å gi en pasient den type terapi som virker ofte (9 av 10 ganger) på denne type pasienter.

Utfall: Pasienten ble verre!

- Terapeut 2 (T2) velger en type terapi som sjelden (1 av 10 ganger) har effekt på denne type pasienter.

Utfall: Pasienten ble mye bedre!

Hvem tok det beste valget?

Igjen, hvem tok det beste valget?

Problem 1:

Kliniske vurderinger/valg tas ofte i probabilistiske læringsomgivelser karakterisert ved at *samme* type valg/vurdering kan gi ulike utfall.

«Utfalls-skjevhet» (*Outcome bias*)

Å vurdere kvaliteten på en beslutning ut fra hvilket utfall den (tilfeldigvis) fikk, og ikke ut fra hva beslutningstakeren visste da beslutningen ble tatt.

Eksempel 2:

Skal pasient X tvangsinnlegges eller ikke?

Problem 2:

Får ofte ikke tilbakemelding på bedømminger alene, men på resultatet av bedømmingen + alt det som gjøres i forlengelsen av den.

«Probabilistiske» kliniske fagområder: Sammenfatning

- Tegn (og mønstre av tegn) indikerer (ofte) det som skal vurderes med stor grad av usikkerhet.
- Faglig praksis utelukker (ofte) erfaringer som er nødvendige for læring av (usikre) sammenhenger (tegns prediktive verdi).
- Fagfolk, klinikere inkludert, blir ikke gjennom erfaring bedre til skjønnsmessig å vurdere hva flere usikre tegn samlet indikerer.
- Usikkerheten gjør at tilbakemeldingen på kliniske valg/vurderinger ofte ikke er informativ nok til å lære av den.

IMPLIKASJONER FOR PRAKSIS

Hvilke konsekvenser bør
forskningsfunnene i bedømmings-
og beslutningspsykologien få?

Krav til sakkyndige etc.

«... En konsekvens av for eksempel forskningen som viser at mange typer kliniske vurderinger knapt bedres gjennom erfaring, burde være en streng kvalitetskontroll av blant andre psykologer og psykiatere som opptrer som sakkyndige og vurderer enkeltmennesker i rettsapparatet (i barnefordelingssaker etc.). Disse bør for eksempel ikke slippe unna med å begrunne sine konklusjoner ved å vise til sin tittel og lange erfaring. Slike vitnemål bør generelt ikke tillegges vekt om de ikke kan belegges med relevant forskning.» (Kirkebøen, 2013, s. 125-6)

Faglig skjønn trenger støtte:

- Læring av relevante statistiske sammenhenger forutsetter systematisk registrering av erfaringer.
- Når man vet hvilke (usikre) data (indikatorer) som er særlig relevante for en type (helhets)vurdering man stadig gjør, så: Overlat (hvis mulig) «vurderingen» (integreringen) av hva de usikre indikatorene samlet indikerer til modeller.
- Bruk (i større grad) «sjekklister» etc. (Gawande, 2009) som støtte til klinisk skønnsutøvelse.

Hvorfor bruker ikke fagfolk i langt større grad beslutningsstøtte?

Mange ulike typer grunner:

1. Krevende å utvikle, integrere og evaluere beslutningshjelpemidler i praksis (f.eks. Yates, Veinott, & Patalano, 2003).
2. De beslutningene gjelder (pasienter etc.), vurderer gjerne **fagfolk som "trenger" beslutningsstøtte som mindre kompetente** enn fagfolk som ikke trenger slik hjelp! (f.eks. Arkes, Shaffer, & Medow, 2007).

3. Dersom beslutningen om for eksempel hvilken behandling man skal få, tas basert på en formel eller et dataprogram og ikke av legen alene, så opplever pasienter at **de ikke i samme grad har en person å dele ansvaret med** om utfallet av beslutningen skulle bli dårlig (f.eks. Promberger & Baron, 2006).

4. Mekaniske rutiner er langt fra feilfrie. Vi har problemer med å akseptere at vi skal "vurderes" av formler og maskiner som til og med ofte "tar" feil, selv om de er klart bedre enn fagfolks skjønn.

5. Om man erstatter skjønn med modellen, så gjør man eksplisitt og synliggjør den store faglige usikkerheten på området. Det liker verken beslutningstakeren selv eller dem beslutningene gjelder.

6. Trusselen mot eget selvbilde, og kanskje økonomiske profesjonsinteresser, bidrar muligens også til å sperre for en økt formalisering, automatisering og forbedring av faglige vurderinger?

Men – hovedgrunnen til at hjelpemidler ikke blir tatt i bruk i langt større grad, er nok den utbredte myten om erfarne fagfolks generelt gode skjønn.

KONKLUSJON

«Tiden er overmoden for at myndighetene pålegger profesjonsutøvere, særlig på "probabilistiske" kliniske fagområder, i langt større grad å utvikle og bruke beslutningshjelpemidler, sjekklistor etc., i stedet for å basere profesjonsutøvelsen på sin lange erfaring og sitt "gode" skjønn.» (Kirkebøen, 2013, s. 43)

Takk for oppmerksomheten!



Anbefalt sekundærlitteratur ☺ :

- Kirkebøen, G. (2013). Kan vi stole på fagfolks skjønn? I A. Molander & J.C. Smeby (red.) *Profesjonsstudier. Bok II*. Oslo: Universitetsforlaget
- Kirkebøen, G. (2012). *Hva er intuisjon*. Oslo: Universitetsforlaget.