



FFI-RAPPORT

18/00108

FFIs prediksjonsturnering

idé- og metodebeskrivelse

—
Alexander W. Beadle

FFIs prediksjonsturnering idé- og metodebeskrivelse

Alexander W. Beadle

Emneord

Prediksjon
Fremtidsstudier
Forsvarspolitik
Sikkerhetspolitikk
Forsvarsplanlegging

FFI-rapport

FFI-RAPPORT 18/00108

Prosjektnummer

1421

ISBN

P: 978-82-464-3014-0

E: 978-82-464-3015-7

Godkjent av

Sigurd Glærum, *forskningsleder*

Espen Berg-Knutsen, *forsknings sjef*

Espen Skjelland, *forskningsdirektør*

Sammen drag

Så lenge fremtiden er preget av usikkerhet, vil det være nødvendig å legge noen antagelser om den fremtidige utviklingen til grunn i alle forsvars- og sikkerhetspolitiske studier, ekspertutredninger, fagmilitære råd og langtidsplaner for Forsvaret. Å «predikere» handler om å gjøre beregninger av fremtidige forhold. Flere stridsspørsmål i den norske forsvarsdebatten kan f.eks. spores tilbake til ulike «prediksjoner» om hvordan et fremtidig angrep på Norge vil se ut, hvor sikre vi kan være på alliert støtte og hva en angriper kan tenkes å gjøre. Konklusjoner om hva vi anser som sannsynlige utviklinger i våre strategiske omgivelser, kan også legge føringer for hvilke løsninger vi velger å utrede videre og beregne de økonomiske kostnadene av senere.

Prediksjon er derfor en uunngåelig del av forsvarsplanleggingen. Problemet er at *alle* eksperter, forskere og offiserer kan ikke ha like rett om hva som vil skje i fremtiden. Gitt de store konsekvensene av å ta feil, er det heller ikke likegyldig hvem beslutningstagerne hører på. Treffsikkerheten til det norske forsvars- og sikkerhetspolitiske miljøet har imidlertid aldri blitt målt.

De neste tre årene vil FFI derfor arrangere en «prediksjonsturnering» for alle forsvars- og sikkerhetspolitisk interesserte i Norge. Denne rapporten beskriver idéen og metoden bak. Turneringen vil gjennomføres som en spørreskjemabasert konkurranse, der alle deltagere får en mail med fem spørsmål i begynnelsen av hver måned. Spørsmålene vil handle om alt fra valgresultat, krig og konflikt til økonomisk utvikling og teknologi. Tidsperspektivet til spørsmålene vil variere fra noen uker til flere år. Turneringen vil være åpen for alle interesserte, både i og utenfor forsvarssektoren, med og uten forsvars- og sikkerhetspolitisk erfaring. Hensikten er å måle hvor gode eller dårlige vi er til å forutsi hendelser og utviklinger av relevans for det norske forsvaret.

Treffsikkerheten vil bli målt ved hjelp av Brier-score. Her måles deltagerne ut fra hvor sikre de er i sine prediksjoner. Deltagerne må derfor anslå hvor *sannsynlig* (i antall prosent) de tror hvert utfall er. Prediksjonsevne handler i denne turneringen om å oppgi høy sannsynlighet til de utfallene som *faktisk* skjer, og lav til de som *ikke* gjør det. Alle deltagere vil få tilsendt resultater med sin score og plassering etter hvert spørsmål som blir avgjort. De fem beste deltagerne navngis, og én gang i året kåres det vinnere innenfor ulike kategorier. Deltagernes treffsikkerhet vil også bli sammenlignet med enkle algoritmer som «tilfeldig gjetning» og «prediker ingen endring».

Resultatene fra turneringen vil for første gang gjøre det mulig å *tallfeste* treffsikkerheten til det norske forsvars- og sikkerhetspolitiske miljøet på spørsmål av relevans for Forsvaret i fremtiden. Resultatene vil også si noe om hvilke *temaer* vi klarer å predikere mer nøyaktig enn andre, om det er store forskjeller mellom ulike *grupper* deltagere og om det finnes noen *enkeltpersoner* i Norge som har en systematisk og målbar evne til å forutsi fremtiden mer nøyaktig enn resten.

Det endelige målet er at ved å få mer kunnskap om hva vi kan og ikke kan forutsi om fremtiden, vil vi også kunne bidra til å øke treffsikkerheten i beslutningsgrunnlaget for fremtidens forsvar.

Summary

In all defence planning, it is necessary to make some assumptions about the future strategic environment. However, given that it is impossible to be certain about what the future may entail, defence and security analyses, expert opinions, threat assessments and scenarios will always involve *some* degree of forecasting. To forecast, or to predict, is to make estimates about future events, e.g. about the scope of threats, changes in defence spending or the conduct of war.

Yet, the forecasting accuracy of the Norwegian defence and security community has never been measured. Furthermore, *all* researchers, experts and officers cannot be equally right about what may happen in the future. Given the potentially catastrophic consequences of preparing for the wrong threats, the degree to which the predictions we base our planning on are correct, will be decisive for how well prepared the Norwegian Armed Forces will be in face of future challenges.

Over the next three years (2017–2020), FFI will therefore be conducting a forecasting tournament for everyone who is interested in or works on defence and security policy-related topics in Norway. The main objective is to measure the Norwegian defence and security community's ability to forecast developments and events of relevance to the Norwegian Armed Forces. The tournament will be conducted through a monthly survey with five questions on topics ranging from international politics and the economy to armed conflicts, technology and future warfare. The tournament will be open to everyone, both from within and outside the Norwegian defence sector, regardless of formal background or previous experience. Thus, it will be possible to compare the forecasting ability of amateurs with that of professionals who work on the questions posed on a daily basis, including military planners and intelligence analysts.

Forecasting accuracy will be measured using Brier score, which is a commonly used proper score function that measures the accuracy of probabilistic predictions. Thus, all participants will be asked to estimate how likely (in percentages) they believe each different outcome is. The goal is to assign high probabilities to outcomes that *do* occur, whilst assigning low probabilities to outcomes that *do not*. All participants will receive an individual score and rank on every question. The five best participants will be named and winners will be announced once a year. All scores will also be measured against various algorithms and groups of other participants.

The tournament's results will provide the first quantitative measure of the Norwegian defence and security policy community's forecasting accuracy. It will also provide insights on whether there are some *topics* that are harder to predict than others, whether some *groups* of people are better than others and whether there are some *individuals* in Norway who are particularly good.

Hopefully, the knowledge gained through this forecasting tournament will help increase the accuracy of analyses in support of decisions on the future of the Norwegian Armed Forces.

Innhold

Sammendrag	3
Summary	4
Forord	6
1 Innledning	7
2 Idéen	8
2.1 Prediksjon i forsvarssammenheng	8
2.2 Tidligere forskning	10
2.3 Hensikt	13
2.4 Opplegg	13
3 Deltagerne	16
3.1 Målgrupper	16
3.2 Registrering	17
4 Resultater	29
4.1 Hva er prediksjonsevne?	29
4.2 Hva slags spørsmål?	31
4.3 Hvordan måles treffsikkerheten?	34
4.4 Eksempel på en turnering	43
4.5 Begrensninger	46
5 Analyser	47
5.1 Det norske forsvars- og sikkerhetspolitiske miljøets prediksjonsevne	47
5.2 Prediksjonsevnen til ulike grupper	47
5.3 De beste prognosemakerne	48
A Brier-score	49
Referanser	54

Forord

Denne rapporten beskriver hensikten og gjennomføringen av FFIs treårige prediksjonsturnering.

Turneringen finansieres av FFI og er initiert som en del av FFI-prosjektet «Globale trender og militære operasjoner II» (2016–2019), som blant annet skal bidra til kompetansebygging på militære fremtidsstudier. Bakgrunnen for turneringen er tidligere studier av hvordan beslutningsgrunnlaget for fremtidens forsvar kan forbedres, der én av anbefalingene har vært å arrangere en prediksjonsturnering for det forsvars- og sikkerhetspolitiske miljøet i Norge.

Turneringen ble startet opp i august 2017. Første spørsmålsrunde ble sendt ut i begynnelsen av september 2017. I januar 2018 har turneringen over 800 deltagere. Mer informasjon og kontinuerlig oppdaterte resultater finnes på turneringens nettside: <https://prediksjonsturnering.ffi.no>

Forfatteren ønsker å takke forskningsleder Sigurd Glærum for støtte til gjennomføringen av turneringen, tidligere kollega Tobias Lillekvelland for hjelp til utviklingen av spørreskjemaer og verktøy for resultatberegning, kollegaer på FFI for innspill til opplegget (Sverre Diesen, Maria Fleischer Fauske, Jostein Grønflaten, Tore Nyhamar, Jan Pay, Kari Røren Strand og Narve Toverød) og medlemmer av projektrådet (Gaute Indseth fra Hærstaben (HST), Per Martin Norheim-Martinsen fra Forsvarets høgskole (FHS), Erlend Kibsgaard Vestad fra Forsvarsdepartementet (FD II) og spesielt Håvard Mogleiv Nygård fra Institutt for fredsforskning (PRIO)) for nyttige tilbakemeldinger på testrunder av turneringen og tidligere utkast av denne rapporten.

Alexander W. Beadle
Kjeller, 11. januar 2018

1 Innledning

All forsvars- og sikkerhetspolitisk planlegging baserer seg på en idé om hvordan trusselbildet, de forsvarsøkonomiske rammene, militærteknologien og våre politiske omgivelser vil utvikle seg i fremtiden. Alle trusselvurderinger, langtidsplaner og innlegg i forsvarsdebatten vil derfor måtte legge noen antagelser om den videre utviklingen til grunn når spørsmålet om hva slags forsvar Norge trenger, skal besvares. Slike antagelser kan handle om alt fra spenningsnivået i nordområdene til innenrikspolitiske utfordringer i Russland, Norges forhold til allierte og endringer i måten krig føres på. Men hvor gode er vi egentlig til å forutsi slike utviklinger?

I dag finnes det ingen studier av vår evne til å forutsi («predikere») den forsvars- og sikkerhetspolitiske utviklingen i Norges omgivelser. De neste tre årene vil Forsvarets forskningsinstitutt (FFI) derfor arrangere en prediksjonsturnering for hele det norske forsvars- og sikkerhetspolitiske miljøet. Målet er å tallfeste hvor gode eller dårlige vi er til å forutsi relevante utviklinger og hendelser og å kunne forbedre treffsikkerheten i beslutningsgrunnlagene for fremtidens forsvar.

Denne rapporten beskriver *idéen* og *metoden* bak FFIs prediksjonsturnering. Kapittel 2 forklarer hvorfor turneringen arrangeres, tidligere forskning på prediksjon av politiske hendelser, hensikten med FFIs turnering og opplegget for gjennomføringen. Kapittel 3 beskriver turneringens målgrupper, hvilken informasjon som samles inn under registreringen og hvordan deltagerne deles inn i ulike grupper. Kapittel 4 diskuterer hva prediksjonsevne handler om i forsvars- og sikkerhetspolitisk sammenheng, hva slags spørsmål turneringen vil inneholde og hvordan treffsikkerheten til deltagerne vil bli målt. Den matematiske formelen som brukes til å måle treffsikkerhet i FFIs turnering er nærmere beskrevet i vedlegg A. Til slutt diskuteres det i kapittel 5 hva slags forskningsspørsmål resultatene fra turneringen forhåpentligvis kan gi svar på.

Hensikten med denne rapporten er å dokumentere hvordan turneringen skal gjennomføres i praksis og å gi innsyn i hvordan prediksjonsevnen og resultatene beregnes.

Nøkkelbegreper	Beskrivelse
Prediksjon	Generelt; beregning av et fremtidig forhold. Her brukes det mer spesifikt om sannsynlighetsberegninger av ulike utfall.
Prediksjonsevne	Evne til å kunne systematisk oppgi <i>høy</i> sannsynlighet til de hendelsene som <i>faktisk</i> skjer, og <i>lav</i> til de som <i>ikke</i> gjør det.
Brier-score	Et mål på treffsikkerhet på en skala fra 0 til 2.
Accuracy-score	Et mål på treffsikkerhet sammenlignet med andre.
Binomiske spørsmål	Spørsmål med ett riktig ett galt svar, f.eks. «ja» eller «nei».
Kategoriske spørsmål	Spørsmål med ett riktig og flere gale alternativer.
Ordinale spørsmål	Spørsmål med ett riktig og flere gale alternativer, men der noen gale alternativer ligger «nærmere» riktig svar enn andre.

Tabell 1.1 Definisjoner av nøkkelbegreper.

2 Idéen

Idéen om å arrangere en prediksjonsturnering kommer fra tidligere studier av det norske forsvars- og sikkerhetspolitiske miljøets evne til å tenke langsiktig. I en FFI-rapport fra 2016 ble det identifisert en rekke fallgruver i alt fra langtidsplanlegging til scenarioutvikling og krigsspill som kan få potensielt katastrofale konsekvenser for beslutninger om fremtidens forsvar.¹

Felles for alle fallgruvene er at de handler om hvordan vi innhenter og anvender informasjon om fremtiden. Et oppløftende funn var at det *er* fullt mulig å forbedre vår treffsikkerhet på spørsmål om internasjonal politikk, krig og konflikt 3–5 år frem i tid. For å kunne bli bedre, forutsettes det imidlertid at vi *vet* når vi tar feil og at vi får muligheten til å *øve* på å predikere.

Én av rapportens anbefalinger var derfor å arrangere en «prediksjonsturnering», der forskere, offiserer og eksperter blir bedt om å forutsi utviklinger av relevans for det norske forsvaret på kort- og mellomlang sikt (0–5 år). Bakgrunnen var et behov for mer kunnskap om hva vi kan og ikke kan forutsi om fremtiden og et ønske om å skape større bevissthet om fallgruvene som gjør at vi tenker feil. Hvor treffsikre er vi egentlig? Er det noen deler av forsvarssektoren som er bedre enn andre? Er det store forskjeller mellom ulike grupper, som kvinner og menn, forskere og offiserer, erfarne og uerfarne? Finnes det enkeltpersoner som er spesielt gode?

I dette kapittelet forklares det hvordan vi (ofte ubevisst) baserer oss på prediksjoner i forsvarssammenheng, hva dagens forskning sier om hva som er mulig og hvem som er best, hva som er målsettingen med akkurat FFIs prediksjonsturnering og hvordan den vil gjennomføres i praksis.

2.1 Prediksjon i forsvarssammenheng

Ingen forsvars- og sikkerhetspolitiske spørsmål kan forutsies med sikkerhet. I så fall hadde Forsvarets langtidsplanlegging vært en ren optimaliseringsprosess av strukturvurderinger framkøbet usikkerheten ved fremtiden generelt og krigføring spesielt. Selv på kort sikt kan det være vanskelig å forutsi utfallet av hendelser. Det er likevel vanskelig å unngå å gjøre *noen* antagelser om hvor verden går, så lenge vi må ta beslutninger i dag om hva slags forsvar vi skal ha i fremtiden.

Å *predikere*, eller forutsi, handler om å beskrive en fremtidig utvikling eller hendelse, gjerne ved hjelp av sannsynlighetsvurderinger. Selv om de aller fleste forsvars- og sikkerhetspolitiske studier innledes med beskrivelser av fremtiden som «usikker» eller «uforutsigbar», baserer anbefalingene seg ofte på antagelser og «prediksjoner» av hva som «sannsynligvis» vil skje.²

I Etterretningstjenestens åpne trusselvurderinger (Fokus) formuleres det f.eks. eksplisitte prediksjoner for hvert år. I 2017 vurderte tjenesten det bl.a. som sannsynlig at Nord-Korea ville

¹ Beadle, Alexander W. (2016), 'Å forske på Forsvaret i fremtiden – muligheter, begrensninger og kognitive fallgruver', *FFI-rapport 2016/01810* (Kjeller: Forsvarets forskningsinstitutt).

² Se f.eks. FFIs og denne turneringens ansvarlige prosjekt sin egen fremtidsstudie, Beadle, Alexander W. og Diesen, Sverre (2015), 'Globale trender mot 2040 – implikasjoner for Forsvarets rolle og relevans', *FFI-rapport 2015/01452* (Kjeller: Forsvarets forskningsinstitutt). Her brukes varianter av ordet «sannsynlig» 75 ganger på 150 sider.

holde frem med atom- og missiltester, at Russland ville skjerme sine strategiske kjernefysiske kapasiteter fra kutt i forsvarsbudsjettet, at regjeringen i Afghanistan ville bli sittende og at det *ikke* ville bryte ut åpen krig mellom den irakiske regjeringen og kurdiske selvstyresmakter.³

I forsvarsanalyser med flere års perspektiv er det ikke vanlig med så eksplisitte prediksjoner, men de baseres likevel ofte på en rekke, mer implisitte antagelser om hva som vil skje i fremtiden. I Forsvarets siste langtidsplan beskrives f.eks. en rekke sikkerhetspolitiske «rammer», inkludert betydningen av nordområdene, fortsatt globalisering, russisk opprustning og spredning av masseødeleggelsesvåpen.⁴ Denne typen premisser er både vanlig og nødvendig for å kunne komme med anbefalinger i alle langtidsplaner, fagmilitære råd og ekspertgruppers rapporter.⁵

Et av FFIs viktigste bidrag til langtidsplanleggingen er «scenarioklassene» som ble identifisert i 2006 og som brukes til å analysere mulige gap i fremtidens forsvarsstruktur.⁶ Hensikten med scenarioklassene er nettopp *ikke* å måtte forutsi én bestemt fremtid, men å kartlegge *spekteret* av mulige trusler Forsvaret må kunne håndtere de neste 10–25 årene. Målet er altså ikke å «treffe» på en bestemt situasjonsbeskrivelse, men ikke å utelukke trusler som *likevel* viser seg å skje.

FFIs scenarioklasser omtales likevel som «delvis sannsynlighetsbaserte», fordi *utvelgelsen* av dem bygger på subjektive vurderinger av hva som kan regnes som «mulig» i fremtiden.⁷ Ekskluderingen av invasjon som den dimensjonerende trusselen i FFIs scenarioklasser bygger eksempelvis på en vurdering av at «Russland ikke disponerer styrker – hverken i dag eller innenfor denne analysens tidshorisont – med kapasitet til å gjennomføre en invasjon av Norge».⁸

De seks klassene som i dag utgjør grunnlaget for FFIs scenarioportefølje, spenner fra «det svært usannsynlige til det svært sannsynlige».⁹ De minst alvorlige scenarioklassene (Fredstidsoperasjoner, Kriminalitet, Terrorangrep og Tvangsdiplomati) anses som «utfordringer som kan oppstå til enhver tid». Noen av dem anses likevel for å være mer sannsynlige enn andre. De to mest alvorlige scenarioklassene (Strategisk overfall og Begrenset angrep) ble derimot vurdert som «svært usannsynlige innenfor rammen av dagens sikkerhetspolitiske situasjon». De er ikke «umulige», men det må skje noe som gjør «militære styrker mer relevante enn hva tilfellet er i dag som virkemiddel for konfliktløsning» for at truslene skal kunne bli realiteter.

Scenarioklassenes *grader* av sannsynlighet kan derfor variere med utviklingen i Norges forsvars- og sikkerhetspolitiske omgivelser, selv om scenarioklassene i seg selv forblir de samme. Et sentralt funn i FFIs gjennomgang av scenariogrunnlaget i 2014 var nettopp det at Russlands intervensjon i Ukraina samme år, hadde vist at landet har utviklet en større evne og vilje til å

³ Etterretningstjenesten (2017), *Fokus 2017*. I oktober 2017 brøt det ut kamper mellom irakiske og kurdiske styrker.

⁴ Se Forsvarsdepartementet (2016), Proposisjon til Stortinget (forslag til stortingsvedtak), Prop. 151 S (2015–2016), *Kampkraft og bærekraft: Langtidsplan for forsvarssektoren*.

⁵ Se f.eks. Ekspertgruppen for Forsvaret av Norge (2015), *Et felles løft*.

⁶ Johansen, Iver (2006), 'Scenarioklasser i Forsvarsstudie 2007: En morfologisk analyse av sikkerhetspolitiske utfordringer mot Norge', *FFI-rapport 2006/02664* (Kjeller: Forsvarets forskningsinstitutt).

⁷ Bukkvoll, Tor; Glærum, Sigurd; Johansen, Iver; Diesen, Sverre og Lia, Brynjar (2014), 'En gjennomgang av FFIs scenariogrunnlag for Forsvarets langtidsplanlegging', *FFI-rapport 2014/01154* (Kjeller: Forsvarets forskningsinstitutt). Begrenset.

⁸ Johansen (2006), 'Scenarioklasser i Forsvarsstudie 2007', s. 28.

⁹ Alle sitater i dette avsnittet er hentet fra Johansen (2006), 'Scenarioklasser i Forsvarsstudie 2007', ss. 37–38.

anvende militærmakt på den måten som beskrives i FFIs mest alvorlige scenarioklasse «Strategisk overfall».¹⁰ Gjennomgangen konkluderte videre med at det var flere trender, bl.a. teknologiske, som vil kunne påvirke muligheter og handlemåter for militær maktbruk i de konkrete scenariene innenfor hver scenarioklasse. Utvalget av utviklinger som diskuteres her, er igjen et resultat av subjektive vurderinger av hvilke trender som antas å kunne bli de mest avgjørende.

Militære planer og strukturutviklingen i Forsvaret kan imidlertid aldri baseres på prediksjoner. Til det er usikkerheten ved fremtiden for stor, og det vil være for farlig å redusere fremtidig krig til én enkelt situasjon som vi må utvikle én konseptuell løsning på. Målet i forsvarsplanleggingen er ikke å forutsi fremtiden, men å bomme så lite at det er mulig å tilpasse seg når krigens realiteter avdekkes. Vi må likevel erkjenne at mye av *grunnlaget* som planene, krigsspillene og anbefalingene bygger på – som de sikkerhetspolitiske rammene, scenarioklassene og utvalget av bestemte teknologiske trender – er i realitetene basert på prediksjoner av hva vi tror vil skje.

Det er derfor ikke mulig å trekke et klart skille mellom «prediksjon» og «kartlegging» av trusler i forsvarssammenheng. Studier har nemlig vist at de fleste av oss har en tendens til å applisere sin forståelse av dagens situasjon *direkte* på spørsmål om fremtidige hendelser, uten å ta hensyn til hvor langt frem i tid vi blir spurt om å vurdere.¹¹ Det vi oppfatter som «mulig» i et 15–25 års perspektiv vil derfor alltid påvirkes av hendelsene, forskningsrapportene og ekspertene vi har i dag. Debatten om store forsvarsinvesteringer, som kampfly og ubåter, domineres f.eks. ofte av kortvarige stemningsskifter i lys av nylige hendelser, som Ukraina-konflikten i 2014, selv om dette er beslutninger som har konsekvenser i mange tiår fremover, der én konflikt ikke nødvendigvis vil være en god indikasjon på hverken det langsiktige trusselbildet eller hva det vil kreve.

Det vi tror er (u)sannsynlig på kort sikt kan derfor påvirke hva vi tror er (u)mulig på lang sikt.¹²

2.2 Tidligere forskning

Felles for nesten alle «prediksjoner» som gjøres i forsvars- og sikkerhetspolitisk sammenheng, er imidlertid at de sjeldent tallfestes eller formuleres på en måte som gjør at de kan etterprøves.

Det finnes ingen studier av hvor godt tidligere langtidsplaners sikkerhetspolitiske beskrivelser har stemt overens med den virkelige utviklingen i verden. Ingen av de mest alvorlige scenarioklassene i FFIs scenari portefølje har (heldigvis) materialisert seg heller. Det betyr imidlertid at vi vet lite om hvor presise de prediksjonene og antagelsene vi baseres oss på faktisk er. Evnen vår til å forutsi internasjonal politisk utvikling har imidlertid blitt forsket på tidligere av andre.¹³

¹⁰ Bukkvoll et al. (2014), 'En gjennomgang av FFIs scenariogrunnlag for Forsvarets langtidsplanlegging'.

¹¹ Mellers, Barbara m.fl. (2015), 'Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions', *Perspectives on Psychological Science*, Vol. 10, No. 3, ss. 267–281, ss. 273–275.

¹² Kommentar fra forskningsleder Sigurd Glærum på et tidligere utkast av denne rapporten.

¹³ Se spesielt Tetlock, Philip (2005), *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton: Princeton University Press), og Tetlock, Philip og Gardner, Dan (2015), *Superforecasting: The Art and Science of Prediction* (London: Random House Books).

Fra 1980-tallet og frem til 2003 undersøkte professor Philip E. Tetlock hvor gode eksperter var til å predikere politiske hendelser og utviklinger 2, 5, 10 og 20 år frem i tid. Resultatene var nedslående: Generelt viste det seg at selv de bedre ekspertene oppnådde bare en treffprosent på rundt 60 %, og de ble dårligere dess lenger frem i tid de forsøkte å predikere. De dårligste slet med å slå tilfeldig gjetning, der lik sannsynlighet ble gitt til begge svaralternativene (50/50). Dette gav opphav til utsagnet som studien senere ble mest kjent for, nemlig at politiske eksperter er like dårlige til å predikere som en pilkastende ape med bind for øynene, der det er helt tilfeldig hvor godt man treffer.¹⁴ Samtidig var ikke *alle* ekspertene like dårlige.

Tetlock fant nemlig at det var mulig å skille mellom to typer eksperter – pinnsvin og rever – på bakgrunn av deres kognitive egenskaper, altså *hvordan* de tenkte:¹⁵

- *Pinnsvin*-ekspertene kjennetegnes av at de kan én eller to store sentrale emner eller teorier, f.eks. maktbalanseprinsippet eller sivilisasjonskonflikt, som de appliserer på alt. De plasserer komplekse problemer inn i årsak-virkningsforhold som de kjenner fra før, og behandler det som ikke passer inn som irrelevant. Pinnsvin er også svært selvsikre i sine prediksjoner og har lettere for å avvise motsigende synspunkter. De bruker gjerne ord som «dessuten», «og så videre» og «i tillegg til» for å trekke inn ytterligere grunner til at de har rett, og skyr ikke ord som «umulig» eller «sikkert» i sine omtaler av fremtiden. Gale prediksjoner bortforklares ved at de «bommet litt på tidspunktet», var «nesten riktige» eller at de ble avsporet av uforutsigbare hendelser.
- *Reve*-ekspertene kjennetegnes derimot av at de kan mange forskjellige, men ikke så store ting. De har ikke tro på at verden kan forstås gjennom én overordnet idé, men bruker ulike analytiske verktøy avhengig av problemet som skal løses. De samler så mye informasjon som mulig, fra så mange kilder som mulig. I språket sitt bruker de oftere ord som «men», «imidlertid», «selv om» og «på den annen side». De snakker også om muligheter og sannsynligheter, ikke sikkerheter – og har lettere for å innrømme feil.

Hovedfunnet fra Tetlocks studie var at *reve*-ekspertene var bedre til å predikere enn *pinnsvin*-ekspertene.¹⁶ Ekspertene traff bedre dess kortere tidsperspektivet var, men treffsikkerheten nærmet seg evnen til den pilkastende apen, med lik tilfeldighet for alle utfall, på spørsmål med 3–5 års perspektiv.¹⁷ Pinnsvinene gjorde det faktisk ofte *dårligere* enn apen. Revene gjorde det bedre, men ikke med en stor margin. Revene klarte bare så vidt å slå enkle algoritmer som predikerte «ingen endring» eller «dagens endrings-tempo».

Fra 2011 til 2015 arrangerte det amerikanske *Intelligence Advanced Research Projects Activity* (IARPA) en prediksjonsturnering som hadde til hensikt å identifisere metoder for å unngå den type feil som ble gjort i forkant av terrorangrepene 11. september 2001 og invasjonen av Irak i

¹⁴ På engelsk: *dart-throwing chimpanzee*. For en diskusjon av metaforen, se forordet og s. 68 i Tetlock og Gardner (2015), *Superforecasting*.

¹⁵ Se kapittel 3–6 i Tetlock (2005), *Expert Political Judgment*, for mer om de to typene eksperter, og Tetlock og Gardner (2015), *Superforecasting*, ss. 68–73 for en kort oppsummering av de samme funnene.

¹⁶ Tetlock og Gardner (2015), *Superforecasting*, s. 68ff.

¹⁷ Tetlock og Gardner (2015), *Superforecasting*, s. 5 og s. 244.

2003.¹⁸ Turneringen stilte rundt 500 spørsmål over en periode på fire år, som f.eks. hvor mange som vil flykte fra Syria det neste året, hvor raskt Kinas økonomi vil vokse det neste kvartalet, og om Russland ville offisielt annektere ytterligere ukrainsk territorium de neste tre månedene.

Fem miljøer fra industri og academia ble invitert til å delta, som igjen rekrutterte tusenvis av deltagere til hvert sitt lag. Disse lagene skulle oppnå så høy treffsikkerhet som mulig, mens de konkurrerte både mot hverandre og en referansegruppe som bestod av aktive etterretningsanalytikere som svarte på de samme spørsmålene. Målet var å slå referansegruppen med 20 % det første året og med 50 % det fjerde året.¹⁹

I forbindelse med denne turneringen etablerte Tetlock og hans kollegaer et lag som het *The Good Judgment Project* (GJP).²⁰ Til forskjell fra den forrige studien var ikke målsettingen her å undersøke hvor gode eller dårlige eksperter var til å predikere, men å finne ut hva som kjenner tegner de menneskene som er *best*. Til dette laget rekrutterte de derfor ikke bare eksperter, men også fra akademiske miljøer, alumniorganisasjoner, vitenskapelige blogger og andre som registrerte seg på internett. Minimumskravet var en bachelorgrad og at de gjennomførte en test av kognitive egenskaper og kunnskap om politiske forhold.

Til sammen samlet GJP over én million prediksjoner fra rundt 2 800 deltagere.²¹ Allerede det første året slo GJP referansegruppen med 50 %, som var målet for det fjerde året.²² Ved hjelp av algoritmer, trening og teknikker klarte GJP-laget å forbedre treffsikkerheten i sine prediksjoner mer enn forbedringene til alle de fire andre lagene til sammen. De andre lagene ble lagt ned etter det andre året, og GJP gjorde det til slutt 70 % bedre enn referansegruppen.²³

Hovedfunnet fra GJP-prosjektet var at det *er* mulig å predikere. Det finnes personer som har en reell, målbar evne til å forutsi hvordan hendelser vil utvikle seg tre måneder, seks måneder, ett år, og ett og et halvt år i forveien.²⁴ De var det Tetlock kaller *superforecastere*, som bestod av de 2 % beste deltagerne. Det var heller ikke flaks, fordi 70 % av dem som var *superforecastere* ett år, forble det året etter.²⁵ Som i den opprinnelige studien av eksperter, handlet det også i denne sammenhengen mest om *hvordan* deltagerne tenkte. *Superforecasterne* kom utelukkende fra kategorien rever og scoret bedre på en rekke kognitive evner som kan motvirke mentale snarveier som gjør at vi alle tenker feil om fremtiden. Alle disse evnene kan trenes og læres.²⁶

¹⁸ For mer informasjon om turneringen, se [‘The Aggregative Contingent Estimation Program: Predicting Global Events Through Crowdsourcing’](#), *Federal Citizen Science and Crowdsourcing Toolkit*, 23. februar 2016.

¹⁹ Joseph, Regina (2014), ‘Keeping Score: Better Policy Through Improved Predictive Analysis’, *5th International Conference on Future-Oriented Technology Analysis (FTA)*, Brussel, 27-28. november 2014.

²⁰ For mer informasjon om GJP-prosjektet, se Tetlock og Gardner (2015), *Superforecasting*, ss. 16–20 og ss. 87–96. Se også Campbell, Mary Pat (2015), ‘What I’ve Learned from the Good Judgment Project’, *Forecasting and Futurism*, July 2015, No. 11, ss. 17–22.

²¹ Tetlock og Gardner (2015), *Superforecasting*, ss. 16–17.

²² Mellers, Barbara m.fl. (2014), ‘Psychological Strategies for Winning a Geopolitical Forecasting Tournament’, *Psychological Science*, Vol. 25, Issue 5, ss. 1106–1115.

²³ [‘The Aggregative Contingent Estimation Program: Predicting Global Events Through Crowdsourcing’](#).

²⁴ Tetlock og Gardner (2015), *Superforecasting*, s. 18.

²⁵ Tetlock og Gardner (2015), *Superforecasting*, s. 104.

²⁶ For en oppsummering av evnene og teknikkene, se kapittel 3 i Beadle (2016), ‘Å forske på Forsvaret i fremtiden’.

2.3 Hensikt

Med utgangspunkt i eksisterende forskning om hva som kan forutsies, er hensikten med FFIs turnering å kartlegge prediksjonsevnen til det *norske* forsvars- og sikkerhetspolitiske miljøet.

Målsettingen er å avdekke kunnskap om vår treffsikkerhet i dag for å kunne oppnå en høyere presisjon i fremtiden, på samme måte som vi kan kalibrere værmeldingene ut fra kunnskap om hvor mange ganger vi har truffet tidligere og hvilke fenomener vi treffer bedre på enn andre.

Ved å måle treffsikkerheten til personer som til daglig arbeider med forsvars- og sikkerhetspolitiske spørsmål, vil det være mulig å *tallfeste* hvor godt vi klarer å forutsi relevante utviklinger for Forsvaret og avdekke hvilke typer spørsmål eller tematiske områder vi kan forutsi med høyere presisjon enn andre. I tillegg er det en målsetting å skape økt *bevissthet* om usikkerheten ved analyser av fremtidens forsvar og hvordan den kan reduseres. Alle kan ikke være like gode til å forutsi hva som vil skje, men det er mulig for alle å bli bedre. Turneringen kan derfor skape større interesse for teknikker og metoder som øker treffsikkerheten i våre analyser. Institusjonelt kan treffsikkerheten økes ved å basere seg i større grad på personer eller grupper som har vist seg å ha en målbar og bedre evne enn andre til å forutsi fremtiden i denne turneringen.

2.4 Opplegg

Det ble utforsket flere alternative løsninger for gjennomføringen av FFIs prediksjonsturnering.

Et første alternativ var å benytte seg av Good Judgment Inc., som er en kommersiell etterfølger av GJP-prosjektet. Dette selskapet tilbyr løsninger for bedrifter som ønsker å forbedre de ansattes prediksjonsevne ved hjelp av skreddersydde turneringssystemer og kurs i regi av *superforecastere* fra GJP. En alternativ løsning ble tilbudt av Cultivate Labs, som tilbyr en turneringsløsning der deltagerne kan logge seg på, predikere spørsmålene de ønsker og oppdatere sine prediksjoner helt til spørsmålet avgjøres. Begge disse alternativene ble vurdert som for kostnadskrevenende for FFI. Disse løsningene forutsetter også svært motiverte deltagere, som vil logge seg inn for å oppdatere vurderingene sine flere ganger i lys av ny informasjon. Dette ble vurdert som for urealistisk gitt målgruppen FFI ønsket å rekruttere fra.

FFIs prediksjonsturnering bygger i stedet på en enklere spørreskjemabasert løsning, der alle deltagere får en mail med spørsmål i begynnelsen av hver måned over en periode på tre år.

For deltagerne består turneringen av tre steg:

- 1) I begynnelsen av hver måned får alle registrerte deltagere tilsendt fem spørsmål. Her blir de bedt om å oppgi hvor sannsynlig (i antall prosent) de tror ulike utfall er. Deltagerne må ikke svare på alle spørsmålene og blir ikke diskvalifisert om de ikke svarer på noen. Spørsmålsrundene er åpne i fem dager (normalt fra mandag til fredag). Etter dette er det ikke mulig for deltagerne å svare på noen av spørsmålene. Dette tidsvinduet er

nødvendig for å sikre at alle deltagerne har et likest mulig informasjonsgrunnlag. Hvis utfallet det spørres om skjer mens runden fortsatt er åpen, utgår spørsmålet.

- 2) Etter at en spørsmålsrunde er avsluttet, får alle deltagerne tilsendt en «prediksjonskupong» med egne prediksjoner og gjennomsnittsprediksjonen til alle deltagerne som svarte på samme spørsmål. Hensikten med kupongen er at deltagerne kan huske hva de (og andre) predikerte etter hvert som fremtiden avslører svarene.
- 3) Etter at et spørsmål har blitt avgjort, får deltagerne tilsendt en ny mail med score og plassering, både på det aktuelle spørsmålet og sammenlagt så langt i turneringen.

Målet i FFIs turnering er å rekruttere minst 500 deltagere og at de svarer på minst 30 spørsmål i løpet av turneringen. Det vil i så fall gi 15 000 spørsmålsbesvarelser. Hvert svar vil inneholde mellom to og rundt seks prediksjoner fra hver person, avhengig av hvor mange svaralternativer spørsmålet har. Med 500 deltagere som svarer på 30 spørsmål vil det bety mellom 30 000 og 90 000 prediksjoner, der én prediksjon tilsvarer én sannsynlighetsvurdering. I tillegg kan det bli aktuelt med ekstra spørreunder i forbindelse med spesielt viktige hendelser som oppstår. Tre måneder inn i turneringen er det per januar 2018 over 800 deltagere registrert, men det er for tidlig å trekke noen konklusjoner om hvor mange spørsmål de vil svare på til sammen.

Både registreringsskjemaet og spørsmålsrundene lages og sendes ut gjennom *Confermit*, som er et kommersielt verktøy for å lage spørreundersøkelser på internett. All informasjon om deltagerne og prediksjonene deres lagres i turneringens datasett hos FFI, der også scorene beregnes. Det betyr at beregningen av deltagerens scores bare gjennomføres av FFI. Deltagere fra skjermede miljøer i Forsvaret, som spesialstyrker og etterretningsanalytikere, deltar over gradert nett. Det er utviklet egne registrerings- og spørsmålsskjemaer til dette formålet. Prediksjonene deres importeres inn i det samme datasettet, men alle deltagerne fra graderte nett er anonymiserte.

Turneringen ble i april 2017 meldt inn til og formelt godkjent av FFIs personvernombud, som er et eget oppnevnt personvernombud godkjent av Datatilsynet. Turneringen innhenter ikke sensitive opplysninger (som etnisitet, politisk oppfatning, religion og helseopplysninger) og all deltakelse er frivillig. Identiteten til deltagerne og resultatene fra turneringen behandles konfidensielt, og det er utarbeidet retningslinjer for håndteringen av personopplysninger, gjennomføring av nye spørsmålsrunder, importering av data fra graderte nett og utsendelse resultater.

Behandlingen av resultatene og produksjonen av analyser, rangeringer og figurer gjøres i *R*, som er et programmeringsspråk og system for statistiske beregninger og grafikk. For denne turneringens formål er det utviklet script som bl.a. lager prediksjonskupongene, analyserer hvem som predikerer hva og som sammenligner treffsikkerheten til ulike grupper, forsvars- og fagmiljøer.

Det er også laget en egen nettside for turneringen: <https://prediksjonsturnering.ffi.no>. Her legges det ut all nødvendig informasjon om registrering, spørsmålene, foreløpige resultater og topplister med de fem beste deltagerne. I tillegg deles alle nye spørsmålsrunder og oppdateringer på turneringens Facebook-side: <https://www.facebook.com/ffiprediksjonsturnering/>.

FFIs prediksjonsturnering

– en konkurranse for alle forsvars- og sikkerhetspolitisk interesserte

HEM OM TURNERINGEN REGISTRERING RESULTATER AVMELDING SPØRSMÅL OG SVAR



Ny spørsmålsrunde: Desember 2017

🕒 4. desember 2017 📄 Spørsmål og svar

En ny spørsmålsrunde er ute. I denne runden spores det om russisk krenking av norsk luftrom, terrorfaren for nordmenn i utlandet, Bitcoin-kursen, NATO og forholdet mellom USA og Storbritannia. Her kan du lese alle spørsmålene, og deltagere kan stille eventuelle oppklarende spørsmål i kommentarfeltet.



Alle som registrerer seg etter 4. desember kl. 09, vil få tilsendt spørsmålene påfølgende morgen til og med fredag 8. desember. Hvis du har registrert deg før dette, men ikke har fått tilsendt spørsmålene, er sannsynligvis eposten din registrert feil. Ta i så fall kontakt med alexander.beadle@ffi.no.

[Fortsett å lese →](#)

[Se alle 2 kommentarer](#)

Om FFI's prediksjonsturnering

På denne siden kan du følge og delta i Forsvarets forskningsinstitutt (FFI) sin prediksjonsturnering.

Alle som vil kan delta i turneringen. Du kan melde deg på når som helst.

Registrer deg

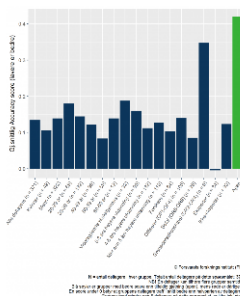
Neste runde begynner om

24 dager **21** timer **41** minutter **14** sekunder

Nytt svar: Nord-Koreas missiltesting

🕒 1. desember 2017 📄 Spørsmål og svar

Et nytt spørsmål er avgjort: Vil Nord-Korea gjennomføre en ny test av et interkontinentalt ballistisk missil i løpet av de neste tre månedene?



Oppsummert viser resultatene fra dette spørsmålet at:

- Deltagerne predikerte i snitt at hendelsen var 76,1 % sannsynlig og traff dermed bedre enn om de bare hadde gjettet (50/50 %).
- Hele 62 av 371 deltagere kom på delt 1. plass ved å oppgi 100 % sannsynlighet for at svaret ville bli ja.
- Av alle deltagergruppene var det ekspertene som traff klart best. Deretter følger personer i aldersgruppen 50-59 år og befal (ORS-OR9).
- Forsvarets operative hovedkvarter (FOH), Forsvarets fellestjenester (FFT) og Hæren hadde i snitt de beste deltagerne fra forsvars- og fagmiljøene.

[Fortsett å lese →](#)

[Legg igjen en kommentar](#)

Har du et spennende spørsmål!?

Her kan du sende inn forslag til spørsmål eller tema til turneringen.

Her kan du se tidligere spørsmål.

Hvem leder sammenlagt?

NAVN	SCORE
1 Anonym	-1,2042
2 Lasse Lærum	-1,0962
3 Kristian Johansen	-1,0258
4 Anonym	-1,0075
5 Eirik Glimsdal	-0,9858

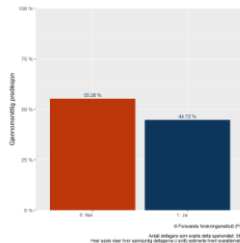
Antall spørsmål avgjort: 4.

Sist oppdatert: 1. desember 2017.

Prediksjoner: November 2017

🕒 13. november 2017 📄 Spørsmål og svar

45 % sannsynlig med en dødelig hendelse som involverer nordkoreanske og sørkoreanske/amerikanske styrker det neste året. 25 % sannsynlig med et vedtak om å etablere en FN-operasjon i Ukraina i løpet av de neste fem månedene. Dette er to av prediksjonene til de 598 deltagerne i november-runden av turneringen.

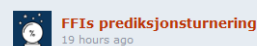


Alle som deltok har fått tilsendt sin individuelle "prediksjonskupong", der alle egne prediksjoner er listet og sammenlignet med snittet til alle deltagerne.

[Fortsett å lese →](#)

[Legg igjen en kommentar](#)

Siste Facebook-oppdateringer



Putin har nå erklært at han stiller i det russiske presidentvalget i mars 2018. Dermed er følgende spørsmål fra oktober-runden med videre i turneringen: Gitt at Putin stiller i presidentvalget 2018, hvor mange prosent av stemmene vil han få?

Svaralternativene var: Mindre enn 60 %, 60 %-69,9 %, 70 %-79,9 %, 80 % eller mer. Ifølge meningsmålinger har Putin i dag mer enn 80 % oppslutning.

Nettsiden til FFI's prediksjonsturnering (prediksjonsturnering.ffi.no), 7. desember 2017.

3 Deltagerne

Dette kapittelet beskriver hvem turneringen retter seg mot og hvordan registreringen foregår. Informasjonen som innhentes her er utgangspunktet for inndelingen av deltagerne i grupper, basert på bl.a. kjønn, alder, utdanningsnivå, institusjonell tilhørighet, erfaring, interesse for og tro på egen evne til å forutsi forsvars- og sikkerhetspolitiske spørsmål.

3.1 Målgrupper

Den første gruppen deltagere som det er viktig å rekruttere til prediksjonsturneringen, er *ansatte i den norske forsvarssektoren i dag*. Dette er personer som, gjennom forskning, rådgivning og utredningsarbeider, er direkte eller indirekte involvert i Forvarets langtidspanlegging og dermed kan ha en reell innflytelse på hvilke oppfatninger som legges til grunn om fremtidens forsvar. Dette gjelder spesielt Forsvarsdepartement (FD), Forsvarets forskningsinstitutt (FFI) og deler av Forsvaret, spesielt grenledelsene, Forsvarets høyskole (FHS), inkludert krigsskolene og Institutt for forsvarsstudier (IFS). Det kan derfor være interessant å sammenligne prediksjonsnivoen til de ulike delene av forsvarssektoren, inkludert forskjellene mellom avdelinger hvor forsvars- og sikkerhetspolitiske spørsmål er en del av oppgaveporteføljen og ikke. For å rekruttere deltagere fra denne målgruppen ble det i august 2017 laget nyhetsartikler med informasjon om turneringen på FFIs og Forsvarets interne nett. I tillegg er opplegget presentert for ulike deler av forsvarssektoren i ulike fora, og det er laget plakater til krigsskolene.

Den andre målgruppen er *andre som tilhører det norske forsvars- og sikkerhetspolitiske miljøet*, men som ikke er ansatt i forsvarssektoren. Dette er f.eks. akademikere på universitet og forskningsinstitutter, risikoanalytikere og konsulenter som på ulike måter arbeider med og uttaler seg om relevante spørsmål. Omtrent 40 % av de mest siterte forsvarseksperter i dag, er ikke ansatt i forsvarssektoren.²⁷ De største sikkerhetspolitiske forskningsmiljøene i Norge, som NUPI, PRIO og universitetene, befinner seg også utenfor forsvarssektoren. For å nå ut til denne målgruppen ble det skrevet en kronikk på forskning.no i september 2017 og sendt ut personlige invitasjoner til et utvalg eksperter innenfor relevante fagområder.²⁸

Den siste målgruppen for turneringen er personer som *ikke* har noe arbeidserfaring med, men er interessert i, forsvars- og sikkerhetspolitiske spørsmål. Til forskjell fra IARPA-turneringen, vil FFIs turnering derfor være åpen alle. Hensikten er å kunne sammenligne profesjonelle miljøer og deltageres prediksjonsevne med «amatører» i fagfeltet. For å nå ut til den forsvars- og sikkerhetspolitiske allmennheten er turneringen promotert på forsvarsrelaterte Facebook-grupper, og det er hengt opp plakater på universiteter for å nå ut til studenter. En offentlig turnering kan også bidra til å skape større oppmerksomhet til hvilke krav vi stiller til hvem vi hører på i spørsmål om fremtidens forsvar, som også er en del av formålet med turneringen.

²⁷ Beadle (2016), 'Å forske på Forsvaret i fremtiden', s. 45.

²⁸ [Beadle, Alexander \(2017\). 'Er du bedre til å forutsi fremtiden enn ekspertene?'. *forskning.no*, 30. september 2017.](#)

3.2 Registrering

For å kunne kategorisere deltagerne på en hensiktsmessig måte, blir alle som ønsker å delta bedt om å oppgi en rekke opplysninger om seg selv, hvor de jobber og deres kjennskap til forsvars- og sikkerhetspolitiske spørsmål fra før. I dette kapitlet følger en beskrivelse av registreringsprosessen alle deltagerne må gjennom og hvilken informasjon som innhentes om dem.

3.2.1 Informasjon og samtykke

Det første steget i registreringen gir deltagerne viktig informasjon om turneringen.

Velkommen til FFIs prediksjonsturnering!

For å kunne delta i denne turneringen, må vi først registrere noen opplysninger om deg. Dette er nødvendig for å kunne sende deg spørsmålene på epost og for å kunne identifisere hvilke personer og grupper som er best til å predikere. Etter at registreringen er fullført, vil du motta et spørreskjema med fem spørsmål i begynnelsen av hver måned. Når et spørsmål som du har svart på er blitt avgjort, vil du motta en ny epost med informasjon om hvilken plass du havnet på og hvilken score du fikk.

Ingen navn vil bli offentliggjort, med unntak av de fem beste deltagerne på et bestemt spørsmål og de fem beste sammenlagt gjennom turneringen. Hensikten er å hedre de som gjør det bra og motivere andre til å bli bedre, men du kan også reservere deg mot å bli navngitt på disse topp-listene under denne registreringen. Alle andre vil forbli anonyme og vi vil bare publisere grupperesultater der enkeltpersoner ikke kan identifiseres, f.eks. kvinner vs. menn, offiserer vs. forskere, amatører vs. eksperter.

Formål med prosjektet. Hensikten med turneringen er å kartlegge hvor gode eller dårlige vi alle er til å predikere forsvars- og sikkerhetspolitiske utviklinger av relevans for det norske forsvaret, f.eks. valg, krig og konflikt, økonomisk utvikling og viktige hendelser i ulike deler av verden. Hvor treffsikre er forsvarsplanleggere og forskere? Kan forsvarsinteresserte enkeltpersoner slå eksperter som brukes i media? Er vi alle bedre enn en pilkastende ape med bind for øynene, der det er helt tilfeldig hvor godt han treffer?

Bli med og hjelp oss med å finne svar på disse og mange andre spørsmål. Lykke til!

Deretter får deltagerne en erklæring, der de må samtykke til bruken av personopplysninger og resultatene til forskningsformål. Ved avmelding blir navn og epostadresse fjernet fra datasettet.

Samtykkeerklæring

Samtykke. Ved å registrere deg her samtykker du i at Forsvarets forskningsinstitutt (FFI) kan bruke personopplysningene dine og resultatene fra denne prediksjonsturneringen til forskningsformål. Formålet med å innhente personopplysninger er å kunne sammenligne prediksjonsevnen til ulike etater, avdelinger og grupper deltagerne. Med unntak av de fem beste deltagerne, vil ikke navnene til deltagerne bli publisert. Deltagerne kan også reservere seg mot å bli navngitt på topplistene. Alle deltagerne må være over 18 år ved registrering.

Behandling av opplysninger. Du vil ikke bli bedt om å oppgi noen sensitive opplysninger under registreringen eller senere i turneringen. Alle personopplysninger (navn, epost, alder, kjønn, utdanningsnivå, bakgrunn) vil bli behandlet konfidensielt. Personopplysninger vil bare være tilgjengelig for forskerne som er ansvarlig for turneringen. Hverken personopplysninger eller resultater fra prediksjonsturneringen vil bli overlevert eller behandlet av andre enn FFI. Fordi noen av spørsmålene i turneringen vil handle om hendelser 20 år frem i tid og fordi de endelige resultatene ikke vil foreligge før alle spørsmål er avgjort, er det ikke lagt planer for å slette datamaterialet ved endt turnering.

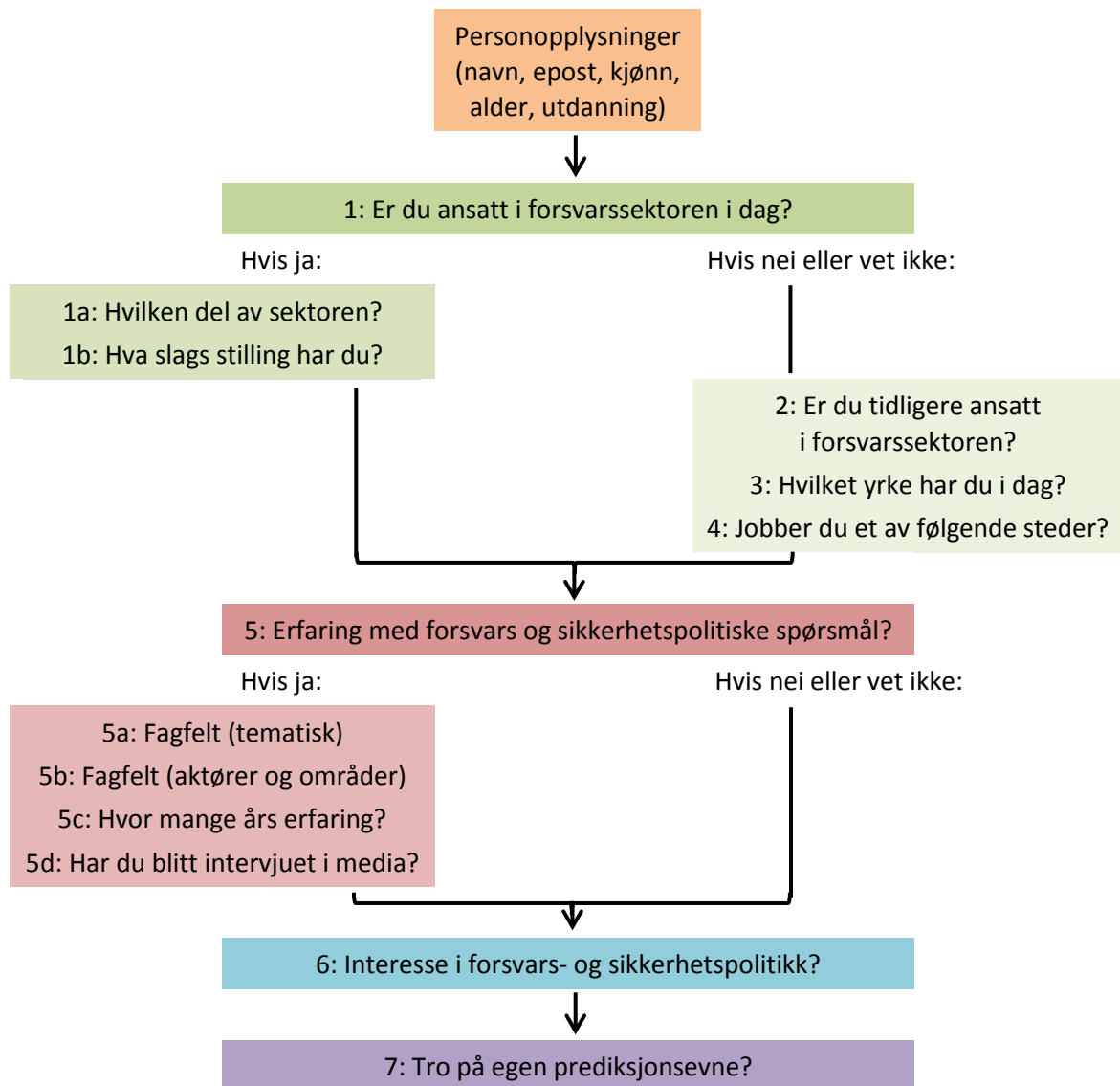
Frivillig deltakelse. Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil opplysningene dine bli anonymisert.

Spørsmål. Dersom du har spørsmål til studien, ta kontakt med ansvarlig forsker ved Forsvarets forskningsinstitutt (FFI), Alexander W. Beadle (alexander.beadle@ffi.no).

Etter samtykkeerklæringen bes deltagerne om å fylle inn følgende personopplysninger: *navn, epostadresse, kjønn, fødselsår og høyeste fullførte utdanning* (grunnskole, videregående skole/gymnas, 0–3 års høyere utdanning, 4–5 års høyere utdanning, mer enn 5 års høyere utdanning). Navn er nødvendig for å ha oversikt over hvem deltagerne er og identifisere hvem som er best. Deltagerne har likevel mulighet til å reservere seg mot å bli navngitt på offentlige «topp fem»-lister under registreringen. Epostadresse er nødvendig for å kunne sende ut skjemaene med spørsmålene som skal predikeres. Informasjon om kjønn, alder og utdanningsnivå gjør det mulig å analysere variasjoner i prediksjonsevnen ut fra disse variablene.

Etter personopplysningene følger en rekke spørsmål som brukes til å kategorisere respondentene. Spørsmålene er basert tidligere forskning om hva som kan ha betydning for prediksjonsevnen og hvilke sammenligninger som kan være særlig relevante å analysere i norsk sammenheng.

Her innhentes det informasjon om fire hovedaspekter: yrke i dag (spørsmål 1-3), erfaring med forsvars- og sikkerhetspolitiske spørsmål (spørsmål 4), interesse for denne typen spørsmål (spørsmål 5) og tro på egen evne til å forutsi dem (spørsmål 6). Avhengig av respondentenes svar, kan de få ytterligere oppfølgingsspørsmål på ulike steg (se figur 3.1).



Figur 3.1 Registreringsprosess i FFIs prediksjonsturnering.

Resten av dette kapitlet gir en beskrivelse av hvert spørsmål i registreringsprosessen og hvilke svaralternativer deltagerne får. Under følger også en diskusjon av relevante forskningsspørsmål og avveininger som ble gjort underveis i utformingen av registreringskjemaet.

3.2.2 Yrke

Prediksjonsturningens primære målgruppe er det norske forsvarsmiljøet. Det første spørsmålet undersøker derfor om respondenten er *ansatt* i forsvarssektoren i dag eller ikke:

Spørsmål (1)	Er du i dag ansatt innenfor forsvarssektoren?
Instruksjoner	Forsvarssektoren inkluderer Forsvaret, Forsvarsdepartementet (FD), Forsvarets forskningsinstitutt (FFI), Forsvarsbygg, Nasjonal sikkerhetsmyndighet (NSM) og Forsvarsmateriell (FMA). Dette inkluderer <u>ikke</u> pågående førstegangstjeneste, innkalling til repetisjonstjeneste, vanlig heimevernstjeneste eller tjeneste i Heimevernets innsatsstyrker. Hvis du er ansatt både innenfor og utenfor forsvarssektoren, svar ut ifra hvor du jobber mest.
Svaralternativer	Ja, Nei, Vet ikke

Hvis respondentene svarer «ja» på dette spørsmålet, får de et oppfølgingsspørsmål (1a), der de blir bedt om å spesifisere *hvor* i forsvarssektoren de er ansatt (ved hjelp av en nedtrekksmeny):

Spørsmål (1a)	I hvilken del av forsvarssektoren er du for øyeblikket ansatt?
Instruksjoner	
Svaralternativer	<ul style="list-style-type: none">• Forsvarsdepartementet (FD)<ul style="list-style-type: none">○ FD I - Avdeling for utvikling, administrasjon og forebyggende sikkerhet○ FD II - Avdeling for sikkerhetspolitikk og operasjoner○ FD III - Avdeling for økonomi, styring og investeringer○ FD IV - Avdeling for forsvarspolitik og langtidspanlegging○ FD V - Avdeling for personellpolitikk, kompetanse og felles juridiske tjenester○ Annet• Forsvarets forskningsinstitutt (FFI)*<ul style="list-style-type: none">○ Avdeling Analyse○ Avdeling Cybersystemer og elektronisk krigføring○ Avdeling Landsystemer○ Avdeling Maritime systemer○ Avdeling Luft- og romsystemer○ Avdeling Beskyttelse og samfunnsikkerhet○ Annet (inkl. staben, strategi og plan, næringsstrategi og andre)• Forsvarsstab (FST)• Forsvarets høgskole (FHS), inkl. krigsskolene og Ingeniørhøgskolen• Forsvarets operative hovedkvarter (FOH)• Hæren

-
- Sjøforsvaret
 - Luftforsvaret
 - Cyberforsvaret
 - Heimevernet
 - Forsvarsmateriell (FMA)
 - Forsvarets logistikk (FLO)
 - Forsvarsbygg
 - Forsvarets fellestjenester, inkl. bl.a. FSA, FRA, FMS og FPK
 - Forsvarets sanitet (FSAN)
 - Forsvarets personell- og vernepliktssenter
 - LOS-programmet
 - Etterretningstjenesten**
 - Forsvarets spesialstyrker**
 - Nasjonal sikkerhetsmyndighet (NSM)**
 - Annet
-

* FFI vil fra 1. januar 2018 ha en ny avdelingsstruktur. Alle deltagere fra FFI vil derfor grupperes på nytt etter at den nye strukturen og prosjektfordelingen er endelig bestemt.

** Deltagere fra Nasjonal sikkerhetsmyndighet, Etterretningstjenesten og Forsvarets spesialstyrker deltar på gradert nett, fordi ansatte i disse etatene ikke vil kunne oppgi navn og institusjonell tilhørighet over internett. Prediksjoner fra disse skjermede delene av forsvarssektoren samles inn separat.

Hensikten med spørsmål 1 og 1a er å kunne undersøke prediksjonsevnen til forsvarssektoren sammenlignet med andre sektorer, samt å kunne sammenligne de ulike etatene innad i forsvarssektoren. Deltagere fra FD og FFI blir også bedt om å oppgi hvilken avdeling de er ansatt i. Dette gjøres fordi disse etatene har avdelinger som det vil være særlig interessante å se på den relative prediksjonsevnen til, fordi de har større miljøer med forsvars- og sikkerhetspolitikk som fagområde. I tillegg er sannsynligheten større for at ansatte i FD og FFI fortsatt vil være ansatt i de samme etatene tre år ut i turneringen enn for ansatte i Forsvaret.

Det samme vil ikke være mulig for avdelinger i Forsvaret på grunn av en høyere stillingsrotasjon, offiserer som er i Garden i dag, kan være på Krigsskolen i morgen, og i FST i overmorgen.

Faktaboks: Stillingsrotasjon i Forsvaret

En FFI-studie fra 2014 fant at tidsbegrensede stillinger er «normen» for militært personell i Forsvaret og at «stillingsbytter for befal i Forsvaret inntreffer oftere enn intensjonen som er hvert 3. år».²⁹ Det betyr at militært ansatte i Forsvaret trolig vil endre stilling i løpet av de tre årene turneringen arrangeres. Rundt 60 % av stillingsbyttene i Forsvaret i 2012 skjedde uten at man byttet bevilgningssted (f.eks. en bestemt avdeling), men rotasjonshyppigheten var lavere for sivilt enn militært personell.

²⁹ Lillekvelland, Tobias og Strand, Kari Røren (2014), 'En analyse av sluttrater og stillingsrotasjon i Forsvaret', *FFI-rapport 2014/00343* (Kjeller: Forsvarets forskningsinstitutt), s. 3.

For vervede militært ansatte skjer rotasjonene like ofte ut av et bevilgningssted som innad.³⁰ Generelt roterer yrkesbefal fra Hæren og Sjøforsvaret hyppigere enn de fra Luftforsvaret og FLO, og yrkesbefal med operativ bransje roterer noe hyppigere enn øvrige bransjer.³¹

Inndelingen av deltagere som er ansatt i Forsvaret begrenses til grenene og overordnede etater som FST, FOH og FLO, fordi sannsynligheten er noe mindre for at de militært ansatte roteres helt ut av disse enhetene enn til en annen avdeling i samme gren eller etat. Denne avveiningen gjøres for å kunne rapportere den relative prediksjonsevnen til ulike deler av Forsvaret underveis i turneringen, uten at den blir for misvisende. For å kunne rapportere den relative prediksjonsevnen til alle avdelinger i Forsvarets, også innad i grenene og etatene, er det mulig å gjennomføre en årlig spørreundersøkelse der respondentene blir bedt om å oppgi nøyaktig hvor i Forsvaret de for øyeblikket er ansatt, f.eks. i Brigade Nord eller Garden. Dette vil gi grunnlag for å lage en samlet score for deltagere fra enkeltavdelinger i Forsvaret.

Alle respondentene som svarer «ja» på at de jobber i forsvarssektoren får også spørsmål 1b:

Spørsmål (1b)	Har du i dag noen av de følgende stillingene?
Instruksjoner	Her kan du krysse av for flere alternativer. Du må ikke krysse av noen, hvis ingen passer.
Svaralternativer	Forsker, Offiser (OF1–OF4), Offiser (OF5–OF9), Befal (OR5–OR9), Grenader/konstabel (OR2–OR4)

Hensikten med spørsmål 1b er å sammenligne prediksjonsevnen til militære og sivile ansatte, militære ansatte på ulike nivåer og mellom forskere og andre sivile ansatte.

Hvis deltagerne derimot svarer «nei» eller «vet ikke» på spørsmål 1 om de er ansatt i forsvarssektoren i dag, får de i stedet et spørsmål om de *tidligere* har vært ansatt i forsvarssektoren:

Spørsmål (2)	Har du tidligere vært ansatt i forsvarssektoren?
Instruksjoner	Forsvarssektoren inkluderer Forsvaret, Forsvarsdepartementet (FD), Forsvarets forskningsinstitutt (FFI), Forsvarsbygg, Nasjonal sikkerhetsmyndighet (NSM) og Forsvarsmateriell (FMA). Dette inkluderer <u>ikke</u> pågående førstegangstjeneste, innkalling til repetisjonstjeneste, vanlig heimevernstjeneste eller tjeneste i Heimevernets innsatsstyrker.
Svaralternativer	Ja, Nei, Vet ikke

³⁰ Lillekvelland og Strand (2014), 'En analyse av sluttrater og stillingsrotasjon i Forsvaret, ss. 39–40.

³¹ Lillekvelland og Strand (2014), 'En analyse av sluttrater og stillingsrotasjon i Forsvaret, s. 42.

Hensikten med spørsmål 2 er å undersøke om prediksjonsevnen til de som samlet sett har bakgrunn fra forsvarssektoren skiller seg fra de som ikke har *noe* arbeidserfaring fra sektoren.

For å undersøke hva slags yrke deltagerne som ikke er ansatt i forsvarssektoren i dag tilhører, vil alle som svarte «nei» på det innledende spørsmålet (1) også bli spurt om hva slags næringskategori de arbeider innenfor. Svaralternativene her baserer seg på SSBs standard for næringsgruppering (SN).³² For å tydeliggjøre hva som ligger i de ulike kategoriene er det i registreringsskjemaet inkludert noen eksempler på de vanligste yrkene innenfor noen av kategoriene. I tillegg vil respondentene få mer informasjon om hver kategori ved å holde musepekeren over den.

Spørsmål (3)	Innenfor hvilken næringskategori arbeider du i dag?
Instruksjoner	Hvis du er usikker, hold musen over alternativene for mer informasjon.
Svaralternativer	<ul style="list-style-type: none">• A – Jordbruk, skogbruk og fiske• B – Bergverksdrift og utvinning (f.eks. kull, malm, olje og gass)• C – Industri (f.eks. anlegg, fabrikker, bakeri)• D – Elektrisitet-, gass-, damp- og varmtvannsforsyning• E – Vannforsyning, avløps- og renovasjonsvirksomhet• F – Bygge- og anleggsvirksomhet• G – Varehandel (f.eks. engros-, butikk- og detaljhandel), reparasjon av motorvogner• H – Transport og lagring• I – Overnattings- og serveringsvirksomhet• J – Informasjon og kommunikasjon (f.eks. forlag, film, TV, radio, telekommunikasjon, IT-tjenester, nyhetsbyrå)• K – Finansierings- og forsikringsvirksomhet• L – Omsetning og drift av fast eiendom• M – Faglig, vitenskapelig og teknisk tjenesteyting (f.eks. forskning og utvikling, juridiske tjenester, regnskap, kommunikasjon, arkitekt, ingeniør, design, fotografi, veterinær og ulike typer faglig rådgivning)• N – Forretningsmessig tjenesteyting (f.eks. utleie av varer, reisebyrå, arrangør-, vakt- og kontortjenester)• O – Offentlig administrasjon og forsvar, og trygdeordninger underlagt offentlig forvaltning• P – Undervisning (f.eks. førskole, grunnskole, videregående skole, høyere utdanning og ulike typer kurs)• Q – Helse- og sosialtjenester

³² <http://www.ssb.no/klasse/#!/klassifikasjoner/6>

- R – Kulturell virksomhet, underholdning og fritidsaktiviteter (f.eks. kunst, freelancejournalistikk, kultur og idrett)
- S – Annen tjenesteyting (f.eks. medlemsorganisasjoner, reparasjon av personlige varer, vaskeri og kroppspleie)
- T – Lønnet arbeid i private husholdninger
- U – Internasjonale organisasjoner og organer
- 0 – Ikke yrkesaktiv/pensjonert
- Vet ikke

Hensikten med spørsmål 3 er å undersøke hvorvidt det er noen yrkeskategorier som skiller seg ut. Et potensielt interessant spørsmål er f.eks. hvor gode ansatte i informasjons- og kommunikasjonsnæringen, som gjerne bruker «eksperters» vurderinger i sitt virke, er til å predikere selv sammenlignet med intervjuobjektene sine.

Til slutt får alle som ikke jobber i forsvarssektoren i dag følgende spørsmål:

Spørsmål (4) Er du i dag ansatt ved et av de følgende forsknings- og fagmiljøene?

Instruksjoner Hvis ingen av alternativene passer, trykk neste.
Hvis du er ansatt flere steder, velg det stedet du jobber mest.

- Svaralternativer**
- Chr. Michelsens Institutt (CMI)
 - Fafo
 - Fridtjof Nansens Institutt (FNI)
 - Institutt for fredsforskning (PRIO)
 - International Law and Policy Institute (ILPI)
 - Norsk utenrikspolitisk institutt (NUPI)
 - Nord universitet
 - Norges miljø- og biovitenskapelige universitet (NMBU)
 - Norges teknisk-naturvitenskapelige universitet (NTNU)
 - Universitetet i Agder (UiA)
 - Universitetet i Bergen (UiB)
 - Universitetet i Oslo (UiO)
 - Universitetet i Stavanger (UiS)
 - Universitetet i Tromsø (UiT)
 - Utenriksdepartementet (UD)

Hensikten med spørsmål 4 er å kunne skille ut deltagere fra spesielt relevante miljøer utenfor forsvarssektoren. Disse kan sammenlignes både seg imellom og med andre i forsvarssektoren.

3.2.3 Forsvars- og sikkerhetspolitisk erfaring

Det andre aspektet som undersøkes i turneringen, er hvorvidt det er forskjeller mellom de som har *arbeidserfaring* med forsvars- og sikkerhetspolitiske spørsmål og de som ikke har det. Alle respondenter får derfor spørsmål om dette, uavhengig av hvilket yrke de har i dag:

Spørsmål (5)	Arbeider du eller har du arbeidet med forsvars- eller sikkerhetspolitiske spørsmål som en del av jobben din?
Instruksjoner	Forsvars- eller sikkerhetspolitiske spørsmål inkluderer bl.a. spørsmål om internasjonal politikk, strategi, militærteori, økonomi, innenrikspolitikk, terrorisme, krig og konflikt, NATO, militære operasjoner, teknologi, Forsvaret, Russland, USA og en rekke andre aktører og områder som kan være relevant for Norges sikkerhet. Med «arbeider» menes det her personer som til daglig jobber eller har jobbet med denne typen spørsmål på et noe overordnet nivå, f.eks. akademisk, politisk, analytisk, strategisk eller operasjonelt, men ikke taktisk.
Svaralternativer	Ja, Nei, Vet ikke

Hvis respondentene har svart «ja» på dette spørsmålet (5), får de også fire oppfølgingsspørsmål:

Spørsmål (5a)	Har du særlig kompetanse innenfor noen av de følgende tematiske områdene?
Instruksjoner	Her kan du krysse av flere alternativer. Du må ikke krysse av noen alternativer, hvis ingen av de nevnte passer.
Svaralternativer	<input type="checkbox"/> Internasjonal politikk generelt (f.eks. stormaktpolitikk, FNs sikkerhetsråd) <input type="checkbox"/> Krig, konflikt og militære operasjoner <input type="checkbox"/> Teknologi <input type="checkbox"/> Økonomi (inkl. forsvarsøkonomi) <input type="checkbox"/> Terrorismen <input type="checkbox"/> Norsk innenrikspolitikk <input type="checkbox"/> Norsk utenrikspolitikk <input type="checkbox"/> Det norske forsvaret

Spørsmål (5b)	Har du særlig kompetanse om noen av disse aktørene eller geografiske områdene?
----------------------	--------------------------------------------------------------------------------

Instruksjoner	Her kan du krysse av flere alternativer. Du må ikke krysse av noen alternativer, hvis ingen av de nevnte passer.
----------------------	------------------------------------------------------------------------------------------------------------------

Svaralternativer	<ul style="list-style-type: none"><input type="checkbox"/> Russland/nordområdene<input type="checkbox"/> Norden<input type="checkbox"/> Europa (inkl. EU)<input type="checkbox"/> NATO<input type="checkbox"/> USA og Nord-Amerika<input type="checkbox"/> Mellom og Sør-Amerika<input type="checkbox"/> Midtøsten og Nord-Afrika<input type="checkbox"/> Afrika sør for Sahara<input type="checkbox"/> Øst-Asia (inkl. Kina, Nord- og Sør-Korea)<input type="checkbox"/> Sentral-Asia (de fem tidl. sovjetrepublikkene)<input type="checkbox"/> Sør- og Sørøst-Asia (inkl. Afghanistan, Pakistan, India og Indonesia)
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Hensikten med denne kartleggingen er å undersøke hvorvidt deltagerne er bedre, dårligere eller like gode som andre til å predikere spørsmål innenfor sitt eget fagfelt. Her er det også mulig å undersøke hvorvidt noen fagområder synes å være vanskeligere å predikere innenfor enn andre.

Spørsmål (5c)	I hvor mange år til sammen har du arbeidet med forsvars- og sikkerhetspolitiske spørsmål?
----------------------	-------------------------------------------------------------------------------------------

Instruksjoner	Forsvars- eller sikkerhetspolitiske spørsmål inkluderer bl.a. spørsmål om internasjonal politikk, strategi, militærteori, økonomi, innenrikspolitikk, terrorisme, krig og konflikt, NATO, militære operasjoner, teknologi, Forsvaret, Russland, USA og en rekke andre aktører og områder som kan være relevant for Norges sikkerhet.
----------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Svaralternativer	Et tall mellom 0 og 100
-------------------------	-------------------------

Spørsmålet 5c gjør det mulig å undersøke om det er noen sammenheng mellom hvor lang erfaring respondentene har med forsvars- og sikkerhetspolitiske spørsmål og evnen til å predikere. Et av funnene i Tetlocks studie var nemlig at det *ikke* var noen sammenheng mellom ekspertenes formelle bakgrunn, som utdanningsnivå og arbeidserfaring, og prediksjonsevne.³³

³³ For en gjennomgang av alle disse faktorene, se kapittel 3 i Tetlock (2005), *Expert Political Judgment*.

Spørsmål (5d)	Har du noen gang blitt spurt eller intervjuet av media som en fagperson ("ekspert") om forsvars- og sikkerhetspolitiske spørsmål?
Instruksjoner	Media inkluderer f.eks. aviser, radio, TV, nettaviser, både nasjonalt og internasjonalt. Med forsvars- og sikkerhetspolitiske spørsmål menes det alt media kan tenkes å spørre eksperter – fra kommentarer om konkrete hendelser i verden, utviklingen i bestemte regioner og land til generelle fenomener som terrorisme, krig og konflikt eller forsvarsøkonomisk utvikling.
Svaralternativer	Ja, Nei, Vet ikke

Spørsmål 5d skal skille ut forsvars- og sikkerhetspolitiske «eksperter» i en egen gruppe, tilsvarende den kategorien Tetlocks første studie målte prediksjonsevnen til. Alle som svarer «ja» på dette spørsmålet vil derfor havne i ekspertgruppen. Denne inndelingen kan imidlertid kvalifiseres ved hjelp av andre mål, f.eks. basert på antall års erfaring fra spørsmålet over eller antall siteringer i den digitale arkivtjenesten ATEKST, som inneholder både internettkilder og artikler fra de største norske papiravisene, fagbladene og magasinene.

3.2.4 Interesse for forsvars- og sikkerhetspolitikk

Det tredje aspektet som skal undersøkes er mulige sammenhenger mellom prediksjonsevnen og deltagerens *interesse* for forsvars- eller sikkerhetspolitiske spørsmål, uavhengig av om de har erfaring fra å jobbe med dem eller ikke:

Spørsmål (6)	På en skala fra 1–7, hvordan vil du vurdere din egen interesse for forsvars- eller sikkerhetspolitiske spørsmål?
Instruksjoner	Forsvars- eller sikkerhetspolitiske spørsmål inkluderer bl.a. spørsmål om internasjonal politikk, strategi, militærteori, økonomi, innenrikspolitikk, terrorisme, krig og konflikt, NATO, militære operasjoner, teknologi, Forsvaret, Russland, USA og en rekke andre aktører og områder som kan være relevant for Norges sikkerhet.
Svaralternativer	<ol style="list-style-type: none"> 1. Svært liten 2. Ganske liten 3. Litt liten 4. Hverken liten eller stor 5. Litt stor 6. Ganske stor 7. Svært stor

Tetlock og andres studier fant blant annet at de som var best til å predikere geopolitiske spørsmål hadde et høyere kunnskapsnivå om internasjonal politikk.³⁴ I FFIs turnering gjennomføres det ikke tester av deltagernes kunnskap om politiske forhold på forhånd, fordi det ville innebære en betydelig lengre registreringsprosess og høynet terskelen for å delta. Spørsmål 5 er likevel med fordi det kan være interessant å undersøke en mulig sammenheng mellom interessen deltagerne selv oppgir at de har og deres reelle prediksjonsevne.

3.2.5 Tro på egen evne til å forutsi forsvars- og sikkerhetspolitikk

Det fjerde og siste aspektet som vil bli undersøkt er sammenhengen mellom prediksjonsevne og tro på egen evne til å forutsi forsvars- eller sikkerhetspolitiske spørsmål:

Spørsmål (7)	På en skala fra 1–7, hvordan vil du vurdere din egen evne til å forutsi forsvars- eller sikkerhetspolitiske utviklinger?
Instruksjoner	Forsvars- eller sikkerhetspolitiske spørsmål inkluderer bl.a. spørsmål om internasjonal politikk, strategi, militærteori, økonomi, innenrikspolitikk, terrorisme, krig og konflikt, NATO, militære operasjoner, teknologi, Forsvaret, Russland, USA og en rekke andre aktører og områder som kan være relevant for Norges sikkerhet.
Svaralternativer	<ol style="list-style-type: none">1. Svært dårlig2. Ganske dårlig3. Litt dårlig4. Hverken dårlig eller god5. Litt god6. Ganske god7. Svært god

Et annet funn i Tetlocks forskning var nemlig at ekspertene i studien hans ofte hadde en overdreven tro på deres egen evne til å predikere, spesielt pinnsvin-ekspertene.³⁵ Spørsmål 7 er derfor med for å kunne undersøke en mulig sammenheng mellom troen på egen evne til å forutsi fremtiden før turneringen starter og den reelle prediksjonsevnen det viser seg at de har. Denne troen kan også sammenlignes med svaret på samme spørsmål etter turneringen er over.

³⁴ Se kapittel 9 og ss. 107–110 i Tetlock og Gardner (2015), *Superforecasting*. For en oversikt over egenskapene som betyr mest for prediksjonsevnen, se Mellers m.fl. (2015), 'The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics'.

³⁵ Tetlock (2005), *Expert Political Judgment*, s. 138; Tetlock og Gardner (2015), *Superforecasting*, s. 184.

4 Resultater

Dette kapittelet beskriver hva som menes med prediksjonsevne i FFIs turnering, hva slags spørsmål deltagerne vil få og hvordan prediksjonsevnen måles i praksis, med et eksempel på en turnering. Til slutt følger en diskusjon av hva resultatene fra turneringen *ikke* kan si noe om.

4.1 Hva er prediksjonsevne?

I forsvarssammenheng er det, som vist i kapittel 2.1, vanskelig å unngå å gjøre noen antagelser, eller «prediksjoner», om den fremtidige utviklingen i våre omgivelser. Hvilke trusler vi ser for oss, hvilke trender vi anser som viktige og hvilke scenarier vi vurderer som «mulige», vil alltid avhenge av hvem vi er og hva vi tror om fremtiden. Alle kan imidlertid ikke ha like rett om hva som vil skje. Det er alltid en fare for at det vi har lagt til grunn ikke stemmer likevel.

Prediksjonsevne handler nettopp om denne graden av samsvar mellom det vi *tror* vil skje og det som *faktisk* skjer. En årsak til at FFI arrangerer en turnering, er at denne evnen sjeldent måles og at vi derfor ikke vet hvor godt vi treffer. For å kunne måle treffsikkerheten er det imidlertid nødvendig å spesifisere *hva* prediksjonsevne betyr i forsvars- og sikkerhetspolitisk sammenheng.

For det første må prediksjonsevnen som skal måles være *relevant* i forsvars- og sikkerhetspolitisk sammenheng. Ofte baseres prediksjoner om fremtiden på hvor ofte noe vanligvis skjer.

Værmeldingene varsler f.eks. at det er 90 % sannsynlighet for regn når tidligere erfaring har vist at det har regnet 90 % av gangene hvor forholdene var like. Med forsvars- og sikkerhetspolitiske spørsmål er dette vanskeligere. Få hendelser er svært like, og selv like hendelser, som f.eks. valg, skjer ikke veldig ofte. Stortingsvalget i 2017 var f.eks. det 18. valget i Norge siden andre verdenskrig. Andre typer hendelser, som terrorangrep, antall konflikter og andelen av BNP som brukes på forsvar, finnes det gode tall på utviklingen over tid, men det er også hendelser som i større grad enn været avhenger av hva vi som mennesker velger å gjøre selv. Noen hendelser skjer også kanskje bare én gang. De underliggende premissene for forsvars- og sikkerhetspolitiske utviklinger er derfor langt mer foranderlige enn naturlovene bak været.

Den typen prediksjon hvor vi baserer oss på hvor sannsynlig noe er statistisk sett, er derfor ikke det vi mener med god prediksjonsevne i forsvars- og sikkerhetspolitisk sammenheng. Det kan være et godt utgangspunkt, men de virkelig nyttige prediksjonene er de som kan fortelle oss at «dette vil skje» og «dette vil ikke skje», uavhengig av hvor ofte det skjer og hvor stor usikkerheten opprinnelig var. Personer som har en systematisk og målbar evne til å forutsi faktiske og konkrete utfall over tid, kan derfor sies å ha en bedre «prediksjonsevne» enn andre i forsvars- og sikkerhetspolitisk sammenheng.

For det andre må det være *mulig* å predikere forsvars- og sikkerhetspolitiske utfall, for at evnen til å gjøre det i det hele tatt skal kunne måles. Forsvars- og sikkerhetspolitikk omtales imidlertid

som et område preget av «*deep uncertainty*».³⁶ Dette begrepet brukes til å beskrive beslutningsområder der usikkerheten er så stor at det ikke er *mulig* (hverken konseptuelt eller i praksis) å beregne sannsynligheter, der en ikke klarer å enes om årsakssammenhenger og der ekspertenes estimater fordeler seg over hele spennet fra 0 % til 100 % sannsynlig.³⁷ I tillegg kommer litteraturen som hevder at det uansett er de store, overraskende hendelsene – såkalte «sorte svaner» – som egentlig avgjør verdens gang.³⁸ Resten er bare støy. Fremtiden anses derfor som grunnleggende uforutsigbar. Da er det også liten vits i å forsøke å predikere den.

Dette bildet må imidlertid nyanseres noe. Selv om det alltid er en fare for plutselig oppdøkkende sorte svaner, «kravler» også historien gjennom inkrementelle og langsomme endringer. Tetlocks forskning har også vist at det *er* mulig å predikere *konkrete* forsvars- og sikkerhetspolitiske spørsmål med høyere presisjon enn tilfeldig gjetning opp til 3–5 år frem i tid.³⁹ Utover dette tidsperspektivet er det imidlertid bare overordnede og til dels åpenbare utviklingstrekk, som globale trender, som kan forutsies og analyseres med en viss grad av sikkerhet. Prediksjonsevnen som skal måles i denne turneringen, begrenser seg derfor til evnen til å forutsi konkrete forsvars- og sikkerhetspolitiske hendelser og utviklinger opp til og med 5 år frem i tid.

Et tredje og siste kriterium for at forsvars- og sikkerhetspolitisk prediksjonsevne skal kunne måles, er *presisjonsnivået*. Selv om noen eksperter har stor nok tro på egen evne til å predikere fremtidige hendelser at de omtaler dem som «helt sikre» eller «helt umulig», vil de fleste personer nyansere sine vurderinger ved å bruke ord som «kanskje», «muligens» og «trolig». Denne kvalitative måten å beskrive fremtidig utvikling på, er normen i de fleste forsvars- og sikkerhetspolitiske studier, ekspertuttalelser og utredninger. Slike språklige vurderinger har imidlertid begrensninger som kan få konsekvenser for kvaliteten på beslutningsgrunnlaget de skal støtte.⁴⁰

Faren med språklige vurderinger er at ord som «kan» og «muligens» kan bety alt fra 1 % til 60 % sannsynlighet. For beslutningstagere som må ta avgjørelser basert på slike språklige vurderinger, kan en så stor kvantitativ forskjell være utslagsgivende for hva de velger å gjøre. Så lenge sannsynlighetsvurderingene ikke tallfestes, er det derfor vanskelig for beslutningstageren å vite hva analytikeren *egentlig* mener.

Et av de mest kjente eksemplene på dette er det mislykkede forsøket på å invadere Cuba i 1961.⁴¹ Bare tre måneder etter at John F. Kennedy ble innsatt som president, iverksatte han en operasjon som skulle styrte Castro-regimet ved å landsette en liten styrke med eksilcubanere i Grisebukta. Det øverste amerikanske lederskapet hadde i forkant sagt at planen hadde en «*fair chance*» for å lykkes. Mannen bak ordene «*fair chance*» har senere uttalt at han med denne for-

³⁶ Se f.eks. Walker, Warren E.; Lempert, Robert J. og Kwakkel, Jan H. (2013), 'Deep Uncertainty', i Gass, S. og Fu, M., red., *Encyclopedia of Operations Research and Management Science* (New York: Springer).

³⁷ For definisjoner, se Lempert, R. J.; Popper, S. W. og Bankes, S. C. (2003), 'Shaping the Next One Hundred Years: New Methods for Quantitative Long-term Strategy Analysis', *MR-1626-RPC* (California: The RAND Pardee Center) og professor Simon French sin oppsummering: <https://www.expertsinuncertainty.net/Portals/60/onepagerdeep.pdf>.

³⁸ Taleb, Nassim Nicholas (2010), *The Black Swan* (NY: Random House).

³⁹ Tetlock og Gardner (2015), *Superforecasting*, s. 5.

⁴⁰ For mer om mulighetene og farene ved å tallfeste sannsynlighetsvurderinger, se spesielt kapittel 3 i Tetlock og Gardner (2015), *Superforecasting*.

⁴¹ For mer om dette eksempelet, se Tetlock og Gardner (2015), *Superforecasting*, ss. 55–56, 193–195 og s. 200.

muleringen mente at det var omtrent 1/3 sjans for å lykkes. President Kennedy ble imidlertid aldri fortalt hva som ble ment med «*fair chance*» – og tolket formuleringen som en høyere sannsynlighet for å lykkes enn den var ment som. Vi kan aldri vite om Kennedys beslutning ville ha vært en annen om han hadde fått tallfestet sannsynligheten å forholde seg til, men det er mulig han ville ha tenkt seg mer om før han iverksatte en operasjon som viste seg å bli en katastrofe.

En viktigere årsak til at vi bør forsøke å tallfeste sannsynligheter i forbindelse med støtte til beslutninger om fremtidens forsvar, er at hvis vi *ikke* gjør det, er det en fare for at vi fortsetter å predikere galt, uten at vi selv er klar over det. Uten å tallfeste sannsynligheter er det nemlig alltid mulig å strekke språklige formuleringer som «kan» og «muligens» til riktig side av svaret. Konsekvensen er at man alltid får «rett», men fortsetter å legge gale antagelser til grunn uten at dette får konsekvenser for fremgangsmåten. Vi har alle en tendens til å underspile egne feilprediksjoner og å finne forklaringer på hvorfor vi likevel hadde rett, uansett hva vi trodde før. I verste fall kan dette gjøre at vi senere tar sjansen på risikable valg med store, og potensielt katastrofale, konsekvenser i forbindelse med store forsvarsinvesteringer og bruk av militærmakt.

Alle deltagerne i FFIs prediksjonsturneringen vil derfor bli bedt om å tallfeste hvor sannsynlig (i antall prosent) de tror ulike utfall er. For å gjøre det bra må deltagerne oppgi *høye* sannsynligheter til hendelser som *skjer*, og *lave* sannsynligheter til de som *ikke* gjør det. I forkant av det amerikanske presidentvalget kunne du f.eks. ha blitt spurt: «Vil Donald Trump vinne det amerikanske presidentvalget i 2016?». La oss si at du svarte at det var 90 % sannsynlig at Trump kom til å vinne. Trump vant som kjent valget. Du ville da ha fått en bedre score enn en annen som svarte at det var 60 % sannsynlig, selv om dere begge hadde «riktig» fordi dere begge trodde han kom til å vinne. Hvis Trump derimot *ikke* vant, ville du ha blitt straffet svært hardt.

Det er disse sannsynlighetsberegningene som vil danne grunnlaget for å vurdere prediksjonevnen til individer og grupper av deltagere i FFIs forsvars- og sikkerhetspolitiske turnering. På den ene siden tvinger tallfestingen deltagerne til å være mer presise enn de mener de har grunnlag for å være. På den annen side er denne tallfestingen nødvendig for å kunne måle prediksjonevnen over tid. Resultatene fra turneringen vil også kunne gi svar på nettopp hvilke temaer vi har grunn til å stole på våre egne tallfestede prediksjoner og hvilke vi ikke har det.

4.2 Hva slags spørsmål?

For at spørsmålene i turneringen skal kunne brukes til å kartlegge det norske forsvars- og sikkerhetspolitiske miljøets evne til å forutsi fremtidige utviklinger *av relevans for Forsvaret*, baserer de fleste spørsmålene seg på trender, aktører og områder som ble identifisert som potensielt relevante for Norge i FFIs fremtidsstudie fra 2015.⁴² Her ble det identifisert over 200 hypoteser om utviklinger som kan få implikasjoner for det norske forsvaret. Andre kilder til spørsmål er Forsvarets langtidsplaner, Etterretningstjenestens trusselvurderinger, andre studier, utredninger og aktuelle saker i nyhetsbildet eller i den norske forsvarsdebatten, fordi det er i disse dokumentene og diskursene premissene for fremtidens forsvar ofte etableres.

⁴² Beadle og Diesen (2015), 'Globale trender mot 2040 – implikasjoner for Forsvarets rolle og relevans'.

Noen av spørsmålene hentes også fra GJ Open sin geopolitiske turnering (www.gjopen.com), fordi mange av temaene som inkluderes her også vil være relevant i en norsk sammenheng. Fordelen med å bruke spørsmål fra GJ Open er at disse allerede har vært gjennom en kvalitets-sikringsprosess og utfallet avgjøres av deres eksperter. Dette vil derfor kreve lite arbeid fra prosjektets side. De fleste spørsmål, som de som berører norske forhold spesielt, vil prosjektet generere og avgjøre selv. Her kan prosjektet dra nytte av andre forskere på FFI som kan hjelpe til med å kvalitetssikre spørsmålene og svarene. Fordi utvalget av spørsmål i turneringen naturligvis vil kunne ha noe å si for hvem som gjør det godt, er det også mulig for alle å foreslå sine egne spørsmål på nettsiden.

Alle spørsmålene må imidlertid være *falsifiserbare*, dvs. at det må være mulig å slå fast om hendelsen faktisk har skjedd eller ikke. Et eksempel på denne typen spørsmål er «Vil Putin stille som kandidat i Russlands presidentvalg i 2018?». Det betyr samtidig at det vil være spørsmål som er høyst relevante, men ikke mulig å ha med i turneringen fordi de ikke er falsifiserbare. Et eksempel på dette er: «Vil Russland utgjøre en økt militær trussel mot Norge i fremtiden?».

For å ivareta dette benyttes en teknikk som heter «*question clustering*».⁴³ Idéen er at det «store» spørsmålet vi ønsker svar på («Vil Russland utgjøre en økt militær trussel mot Norge i fremtiden?») er for stort, mens de «små» spørsmålene er for små alene. I stedet kan det lages klynger av mindre, falsifiserbare spørsmål, der svarene samlet sett kan gi en indikasjon på det store spørsmålet vi ønsker å kunne forutsi med høyere presisjon.

Et eksempel på en slik klynge vises i under:



Hvis svarene på de små spørsmålene i eksempelet over er mer «ja» enn «nei», kan det peke i retningen av en økt militær trussel mot Norge. Andre «store» spørsmål som vil være relevant å undersøke i FFIs turnering, er f.eks. den videre utviklingen i NATO, faren for væpnet konflikt i Europa, terrortrusselen, den forsvarsøkonomiske utviklingen og trender i krigføring.

Alle spørsmålene vil i utgangspunktet bli avgjort på bakgrunn av rapportering i åpne kilder, f.eks. internasjonale nyhetsbyråer som Reuters og AP samt etablerte medier som BBC, Aften-

⁴³ For mer om denne teknikken, se Tetlock og Gardner (2015), *Superforecasting*, ss. 261–266.

posten, VG og Dagbladet. Alternativt er det mulig å basere seg på fageksperter, der det er nødvendig med mer skjønnsmessige vurderinger. Om ikke kriteriene for hvordan spørsmålet vil bli avgjort er selvsagte, vil det bli opplyst hva som må til for at hendelsen regnes for å ha skjedd, f.eks. at minst én person må bli drept for at et terrorangrep kan regnes som «gjennomført».

Tidsperspektivet på spørsmålene i FFIs turnering vil i hovedsak variere fra et par uker til noen få år, gitt begrensningen på hvor langt det har vist seg å være mulig å predikere. Turneringen vil imidlertid ha noen spørsmål om hendelser av særlig stor betydning med et lengre tidsperspektiv (10–20 år). Hensikten er både å etterprøve tidligere funn om at det ikke går an å predikere så langt frem innenfor en forsvarsspesifikk kontekst og å kartlegge hvor sannsynlig miljøet faktisk anser slike hendelser. I et av turneringens første spørsmål vurderte f.eks. deltagerne det som 18 % sannsynlig at en stat eller allianse av stater vil bruke ett eller flere kjernevåpen i forbindelse med en væpnet konflikt i løpet av de neste ti årene.

Alle spørsmålene vil også kategoriseres i henhold til de samme kompetanseområdene som deltagerne ble bedt om å oppgi under registreringen (spørsmål 5a og 5b). Slik vil det være mulig å undersøke om personer med «Russland» eller «økonomi» som kompetanseområder er bedre til å predikere spørsmål innenfor disse kategoriene.

Turneringen vil benytte seg av tre ulike typer spørsmål:

- **Binomiske spørsmål.** Her finnes det bare ett riktig og ett galt svar, som «ja» og «nei». Eksempel: «Vil Putin stille som kandidat i Russlands presidentvalg i 2018?». Hvis du svarer at det er 80 % sannsynlig for at Putin stiller, betyr det samtidig at de tror det er 20 % sannsynlighet for at han *ikke* gjør det.
- **Kategoriske spørsmål.** Her finnes det bare ett riktig, men flere gale svar, der ingen av dem er «riktigere» enn andre. Eksempel: «Fra hvilket parti (R, SV, Ap, MDG, Sp, V, KrF, H, FrP) vil forsvarsministeren komme fra etter Stortingsvalget i 2017?». Her må du oppgi estimater for alle alternativene. Summen av alle alternativene må bli 100 %.
- **Ordinale spørsmål.** Her finnes det bare ett riktig svar, men noen gale svar er «nærmere» enn andre. Eksempel: «Hvor mange vil bli drept i islamistiske terrorangrep i Europa i 2017?». Du får følgende alternativer: A: 0–49, B: 50–99, C: 100–149, D: 150–199 og E: 200+. Som ved kategoriske spørsmål må du oppgi estimater for *alle* alternativene, men hvis det riktige svaret ble 65 (B), vil 0–49 (A) og 100–149 (C) være riktigere enn 150–199 (D) og 200+ (E). Denne vektingen må derfor tas hensyn til.

I tillegg kan noen av spørsmålene være betingede, der man blir bedt om å predikere et utfall *gitt en bestemt begivenhet*, f.eks. om hvorvidt Norge vil delta i en operasjon *hvis* våre allierte bestemmer seg for å gjennomføre den. Disse spørsmålene kan være av alle de tre overnevnte typene. Hvis den forutsatte begivenheten ikke skjer, utgår spørsmålet fra turneringen.

Fordi både spørsmålene og deltagerens ekspertiseområder vil variere betydelig, vil de fleste spørsmål ha en kort beskrivelse av eksisterende kunnskap på temaet, f.eks. fjorårets verdier eller

tidligere utfall. Her er det fare for *ankring*, en psykologisk mekanisme der vi tar utgangspunkt i et tall vi får presentert (ankeret) og justerer beregningene våre deretter.⁴⁴ Et dårlig anker kan bety at vi ikke justerer nok, og det er overraskende lett å begynne med et dårlig anker. Samtidig vil et felles utgangspunkt gjøre at forskjellene ikke skyldes mangel på grunnleggende kunnskap, men heller ulike *vurderinger* av sannsynligheten. I tillegg øker det sjansen for at personer som ikke er reelle fagpersoner på temaet forsøker å predikere spørsmålet. Dette kan også være en mer realistisk setting for mange ansatte i forsvarssektoren, der de gjerne har en mulighet til å sette seg litt, men ikke fullstendig, inn i forsvars- og sikkerhetspolitiske spørsmål.

4.3 Hvordan måles treffsikkerheten?

Et siste viktig spørsmål er *hvordan* prediksjonsevne kan måles.⁴⁵ FFIs prediksjonsturnering vil basere seg på en forenklet modell av måten prediksjonsevne ble målt i IARPA-konkurransen og måten det måles på i GJ Open-turneringen nevnt tidligere. Her måles prediksjonsevnen ved hjelp av et system som heter *Brier-score*, der deltagerne ikke bare måles ut fra om de *treffer*, men også hvor *sikre* (i antall prosent) de var i sin prediksjon – derav begrepet «treffsikkerhet».

Med utgangspunkt i Brier-score, beregnes det også en Median-score og en Accuracy-score.

4.3.1 Brier-score

Brier-score er basert på Glenn W. Briers studie fra 1950 for å måle treffsikkerheten til værmeldinger.⁴⁶ Brier-score måler avstanden mellom det du predikerte og det som virkelig skjedde, basert på den tallfestede sannsynligheten du gir til hvert mulig utfall på spørsmålet.

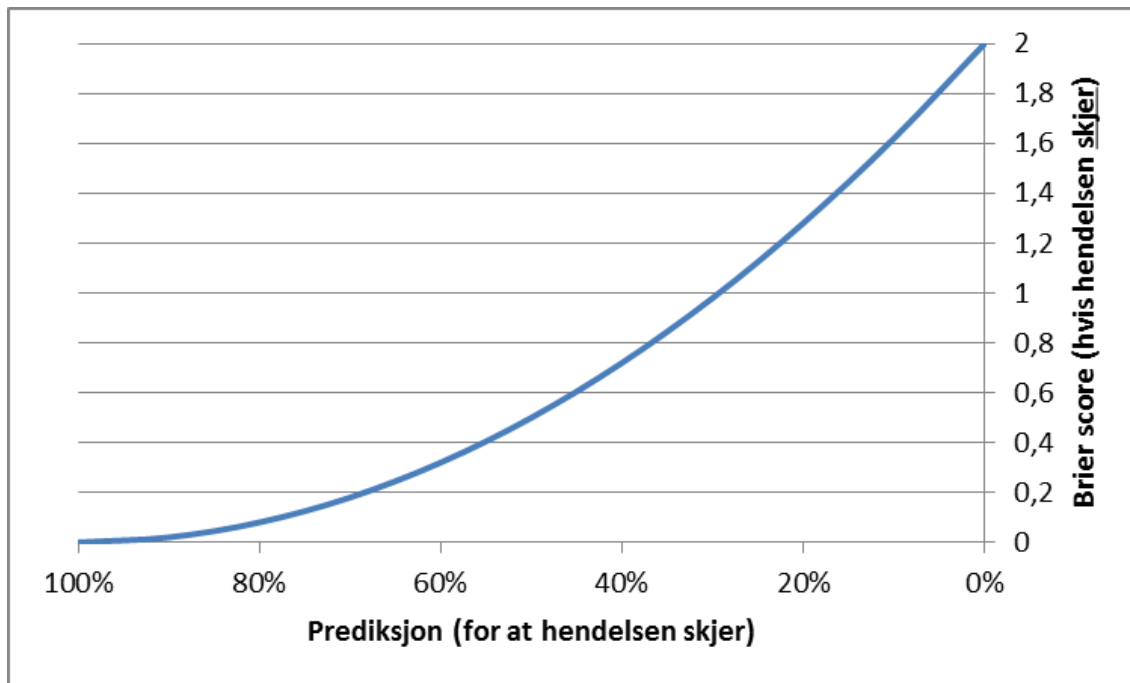
Brier-score-skalaen går fra 0 til 2, der *lavere* score betyr høyere treffsikkerhet. Du får en Brier-score på 0 hvis du predikerer «helt riktig», dvs. at du sier at en hendelse er 100 % sannsynlig og den skjer. Du får en Brier-score på 2 hvis du predikerer «helt feil», dvs. at du sier at en hendelse er 100 % sannsynlig, men den *ikke* skjer. Hvis du predikerer 50 % sannsynlighet, der du gir lik sannsynlighet til begge alternativene som den pilkastende apen, får du en Brier-score på 0,5 uansett utfall. Figurene 4.1 og 4.2 illustrerer hvordan Brier-scoren varierer ved ulike estimater.

⁴⁴ For mer om ankring, se kapittel 11 i Kahneman (2013), *Tenke, fort og langsomt* (Oslo: Pax Forlag). For den opprinnelige forskningen, se Tversky, Amos og Kahneman, Daniel (1974), 'Judgment under Uncertainty: Heuristics and Biases', *Science*, Vol. 185, No. 4157, ss. 1124–1131.

⁴⁵ Følgende delkapittel og figurer er basert på Tetlock og Gardner (2015), *Superforecasting*, ss. 59–64.

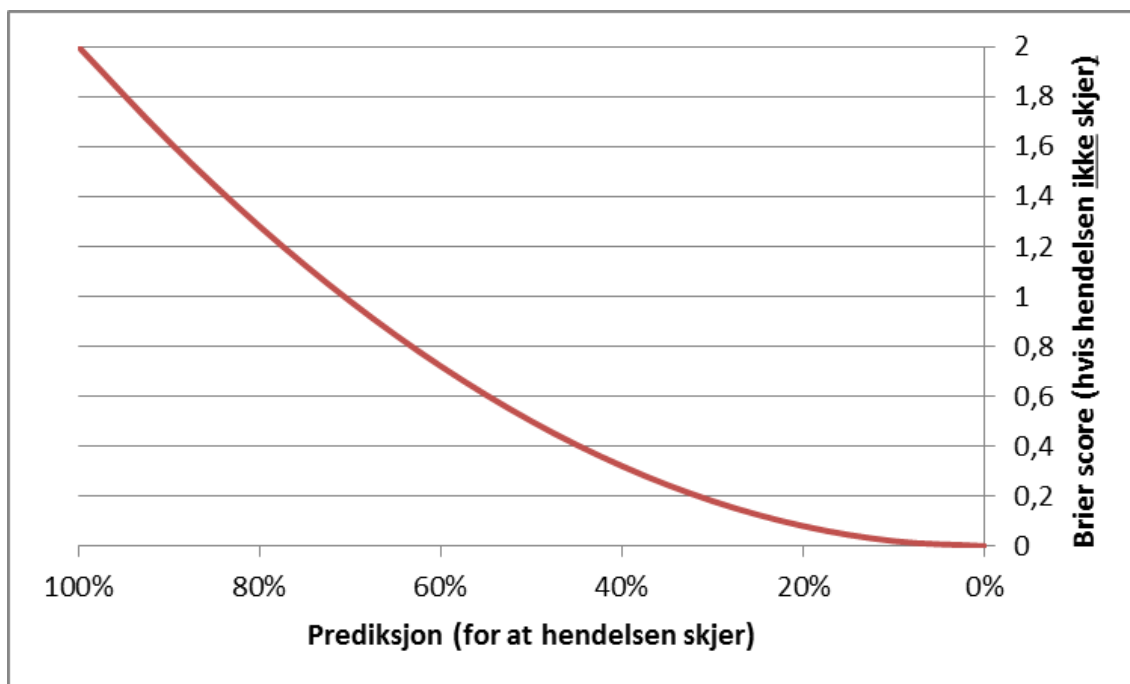
⁴⁶ [Brier, Glenn W. \(1950\), 'Verification of Forecasts Expressed in Terms of Probability', *Monthly Weather Review*, Vol. 78, No. 1.](#)

Hvis du ble bedt om å predikere sannsynligheten for at en hendelse skjer, og hendelsen *skjedde*, ville du ved ulike sannsynlighetsestimater få følgende Brier-scores:



Figur 4.1 Brier-score ved ulike estimater, gitt at hendelsen skjedde.

Hvis hendelsen derimot *ikke* skjedde, ville du få følgende Brier-scores ved samme estimater:



Figur 4.2 Brier-score ved ulike estimater, gitt at hendelsen ikke skjedde.

Som det fremgår av figurene, straffes du hardt for å havne på «feil side» av 50/50. Estimerer på 80 % og 90 % for hendelser som *ikke* skjer, staffes hardere enn gevinsten du hadde oppnådd ved å justere fra 80 % til 90 % sannsynlighet hvis du hadde predikert riktig. Det betyr at hvis du tar sjansen på bastante prediksjoner og er heldig nok til å treffe flere ganger, men bommer av og til, vil du over tid gjøre det dårligere enn om du hadde vært mer forsiktig, men traff mindre spektakulært. Samtidig vil du slite med å slå den pilkastende apen, hvis du «helgarderer» deg ved å fordele sannsynligheten omtrent likt på alle svaralternativer. Brier-score er derfor et system som måler din systematiske evne til å gi høye sannsynligheter til hendelser som *faktisk* skjer, og lave sannsynligheter til de som *ikke* gjør det, uavhengig av hvor «vanlige» de er statistisk sett.

I det videre beskrives det hvordan Brier-scoren blir beregnet ved binomiske, kategoriske og ordinale typer spørsmål. Den matematiske formelen er nærmere beskrevet i vedlegg A.

4.3.1.1 Binomiske spørsmål

Den enkleste utregningen av Brier-score skjer ved et binomisk ja/nei-spørsmål. La oss ta spørsmålet om hvorvidt Putin vil stille til valg i Russland i 2018. Du anslår at det er 80 % sannsynlig at han vil det. Det betyr samtidig at du mener det er 20 % sannsynlig at han *ikke* vil det.

Brier-scoren beregnes da på følgende måte:

1. Sannsynlighetsberegningene deles på 100, slik at prediksjonene blir et tall mellom 0 og 1. Det vil si at prediksjonen for at Putin stiller til valg blir: $80/100 = 0,8$, mens prediksjonen for at han *ikke* stiller til valg blir: $20/100 = 0,2$.
2. Deretter avgjøres det om hendelsen det ble spurt om har skjedd eller ikke skjedd. Hvis hendelsen skjedde, gis den verdien 1. Hvis den *ikke* skjedde, gis den verdien 0. La oss si at Putin stilte til valg. Da får dette utfallet (at Putin stiller til valg) verdien 1 (markert med rød farge under), mens det gale utfallet (at Putin ikke stiller til valg) får verdien 0.
3. Brier-scoren finnes ved å beregne forskjellen mellom prediksjonene og disse utfallene (1 og 0), multiplisere disse forskjellene med seg selv, og legge dem sammen.

Gitt at Putin stilte til valg, ville du fått følgende Brier-score:

Ja		Nei	
$(0,8 - 1)^2$	+	$(0,2 - 0)^2$	=
0,04	+	0,04	= 0,08

Hvis Putin *ikke* stilte til valg, ville du med samme prediksjoner fått en langt høyere Brier-score:

Ja		Nei	
$(0,8 - 0)^2$	+	$(0,2 - 1)^2$	=
0,64	+	0,64	= 1,28

Legg merke til at prediksjonene er de samme (0,8 og 0,2), mens det er hendelseskodene (1 og 0) som har byttet plass, fordi utfallet har endret seg fra «skjedde» (1) til «skjedde ikke» (0).

4.3.1.2 Kategoriske spørsmål

Fremgangsmåten er den samme for kategoriske spørsmål med flere alternativer, men utregningen utvides for hvert alternativ som er med. La oss f.eks. si at du ble bedt om å predikere hvem av fem nominerte som ville vinne kategorien «rock» under Spellemannsprisen 2017.

Du velger å estimere sannsynligheten for de fem ulike alternativene slik:

Alternativer	Sannsynlighetsvurdering
A: «Pineapple Pizza» – Death By Unga Bunga	40 %
B: «Vannmann86» – Hjerteslag	20 %
C: «Nattesferd» – Kvelertak	10 %
D: «Ført bak lyset» – Tasmørke	30 %
E: «The Gospel – Årabrot	0 %
Totalt	100 %

Hvis alternativ A var riktig, ville du derfor fått følgende Brier-score:

A	B	C	D	E					
$(0,4 - 1)^2$	+	$(0,2 - 0)^2$	+	$(0,1 - 0)^2$	+	$(0,3 - 0)^2$	+	$(0 - 0)^2$	=
0,36	+	0,04	+	0,01	+	0,09	+	0,0	= 0,5

Hvis alternativ C var riktig, ville du fått følgende Brier-score:

A	B	C	D	E					
$(0,4 - 0)^2$	+	$(0,2 - 0)^2$	+	$(0,1 - 1)^2$	+	$(0,3 - 0)^2$	+	$(0 - 0)^2$	=
0,16	+	0,04	+	0,81	+	0,09	+	0,0	= 1,1

Hvis alternativ E var riktig, ville du fått følgende Brier-score:

A	B	C	D	E					
$(0,4 - 0)^2$	+	$(0,2 - 0)^2$	+	$(0,1 - 0)^2$	+	$(0,3 - 0)^2$	+	$(0 - 1)^2$	=
0,16	+	0,04	+	0,01	+	0,09	+	1	= 1,3

Selv om man i det siste eksempelet mente at det var 0 % sannsynlighet for det riktige alternativet og dermed tok «helt feil», får man likevel ikke 2 i Brier-score hvis de resterende prosentene fordeles på flere alternativer. Hvis man derimot satser alt (100 %) på ett alternativ som viser seg å være feil, vil man også i spørsmål med flere alternativer kunne få en Brier-score på 2.

Til sammenligning vil den pilkastende apen, der sannsynligheten fordeles likt på alle alternativene, få følgende Brier-score ved forskjellige antall alternativer: Ved tre alternativer, der sannsynligheten for hvert alternativ er 33,3 %, vil apen få en Brier-score på 0,6667. Ved fire alternativer, der sannsynligheten for hvert alternativ er 25 %, vil Brier-scoren bli 0,75. Ved fem alternativer, der sannsynligheten for hvert alternativ er 20 %, vil apen få en Brier-score på 0,8.

4.3.1.3 Ordinale spørsmål

Ordinale spørsmål har også flere svaralternativer, men her er noen alternativer «riktigere» enn andre. Da må også prediksjonsevnen vektas annerledes. La oss si at Ole og Per blir bedt om å predikere hvor mange som blir drept av islamistisk terror i Europa i 2017.

De får følgende alternativer, og estimerer sannsynligheten for hvert utfall slik:

Alternativer	Ole	Per
A: 0-49 drepte	25 %	0 %
B: 50-99 drepte	25 %	25 %
C: 100-149 drepte	40 %	30 %
D: 150-199 drepte	10 %	35 %
E: 200+ drepte	0 %	10 %
Totalt	100 %	100 %

La oss si 65 ble drept av islamistisk terror i 2017. Da var B det riktige svaralternativet. Som vi ser anslo begge det som 25 % sannsynlig at B var riktig, men Ole vektla de nærmeste alternativene A og C mer enn Per, som avfeide A som helt usannsynlig og la mer vekt på C og D.

Ved vanlig Brier-utregning ville Ole fått følgende score:

A (0-49)	B (50-99)	C (100-149)	D (150-199)	E (200+)	
$(0,25 - 0)^2$	$+ (0,25 - 1)^2$	$+ (0,4 - 0)^2$	$+ (0,1 - 0)^2$	$+ (0 - 0)^2$	=
0,0625	+ 0,5625	+ 0,16	+ 0,01	+ 0	= 0,795

Ved vanlig Brier-utregning ville Per fått følgende score:

A	B	C	D	E	
$(0 - 0)^2$	$+ (0,25 - 1)^2$	$+ (0,3 - 0)^2$	$+ (0,35 - 0)^2$	$+ (0,1 - 0)^2$	=
0	+ 0,5625	+ 0,09	+ 0,1225	+ 0,01	= 0,785

Her ser vi at begge Brier-score blir omtrent 0,79, men Per får faktisk en *lavere* (og dermed bedre) Brier-score enn Ole. Forklaringen er at Ole blir straffet mer for å ha bommet med et høyere anslag på 40 % på alternativ C enn det Per gjorde, som på det meste bommet med 35 % på D. Dette gir imidlertid et galt bilde av prediksjonsevnen siden Ole egentlig var nærmere det riktige svaret ved at han vektla A, B og C langt høyere enn Per, som la mest vekt på B, C, D og litt E. Brier-scorene deres bør derfor reflektere denne relativt bedre prediksjonsevnen til Ole.

For å regne ut denne typen spørsmål vil FFI benytte samme fremgangsmåte som GJ Open:⁴⁷

1. Først deles alle svaralternativene opp i par. I eksempelet over deles Ole og Per sine svaralternativer inn i fire par: A vs. BCDE, AB vs. CDE, ABC vs. DE og ABCD vs. E.
2. Deretter beregnes Brier-scoren til hvert par på samme måte som ved et ja/nei-spørsmål, men hendelseskodingen for at det skjedde (1) vil alltid trekkes fra det paret som inneholder det riktige svaralternativet, i dette tilfellet det paret som inneholder alternativ B.

For Ole vil alle parene utregnes slik:

A		BCDE	
$(0,25 - 0)^2$	+	$((0,25 + 0,4 + 0,1 + 0) - 1)^2$	=
0,0625	+	0,125	= 0,125

AB		CDE	
$((0,25 + 0,25) - 1)^2$	+	$((0,4 + 0,1 + 0) - 0)^2$	=
0,25	+	0,25	= 0,5

ABC		DE	
$((0,25 + 0,25 + 0,4) - 1)^2$	+	$((0,1 + 0) - 0)^2$	=
0,01	+	0,01	= 0,02

ABCD		E	
$((0,25 + 0,25 + 0,4 + 0,1) - 1)^2$	+	$(0 - 0)^2$	=
0	+	0	= 0

⁴⁷ For en beskrivelse av metoden, se http://training.goodjudgment.com/Ordered_Categorical_Scoring_Rule.pdf.

3. Til slutt regnes gjennomsnittet av alle fire parene.

For Ole blir dette regnestykket som følger:

A vs. BCDE	AB vs. CDE	ABC vs. DE	ABCD vs. E				
0,125	+	0,5	+	0,02	+	0	= 0,645
						0,645/4	= 0,16125

Med denne vektete fremgangsmåten får Ole en Brier-score på 0,16125.

Den samme fremgangsmåten må gjøres for alle deltagerne. For Per blir utregningen som følger:

A		BCDE	
$(0 - 0)^2$	+	$((0,25 + 0,3 + 0,35 + 0,1) - 1)^2$	=
0	+	0	= 0

AB		CDE	
$((0 + 0,25) - 1)^2$	+	$((0,3 + 0,35 + 0,1) - 0)^2$	=
0,5265	+	0,5265	= 1,125

ABC		DE	
$((0 + 0,25 + 0,3) - 1)^2$	+	$((0,35 + 0,1) - 0)^2$	=
0,2025	+	0,2025	= 0,405

ABCD		E	
$((0 + 0,25 + 0,3 + 0,35) - 1)^2$	+	$(0,1 - 0)^2$	=
0,01	+	0,01	= 0,02

Deretter beregnes snittet av alle parene:

A vs. BCDE		AB vs. CDE		ABC vs. DE		ABCD vs. E	
0	+	1,125	+	0,405	+	0,02	= 1,55
						1,55/4	= 0,3875

Nå viser det seg at Per får en vektet Brier-score på 0,3875, og dermed et dårligere resultat enn Ole, som fikk 0,16125 med samme utregningsmåte. Denne fremgangsmåten er den samme uavhengig av hvor mange flere alternativer enn to svaret har. Til sammenligning vil den pilkastende apen få en Brier-score 0,3 ved å oppgi 20 % sannsynlighet for hvert av de fem alternativene.

En deltagers gjennomsnittlige Brier-score gir en indikasjon på hvor god eller dårlig han eller hun er til å predikere generelt. Samtidig vil scoren kunne variere mye fra spørsmål til spørsmål, fordi usikkerheten kan være svært ulik. Større usikkerhet vil føre til høyere Brier-score. Det betyr blant annet at personer som predikerer spørsmål innenfor fagfelt preget av mye usikkerhet, vil generelt få en høyere Brier-score enn personer som predikerer innenfor fagfelt preget av mindre usikkerhet. Brier-scorene til deltagere på ulike spørsmål kan derfor ikke sammenlignes uten videre. Det er imidlertid mulig å sammenligne Brier-scorene til deltagere som har predikert samme spørsmål eller spørsmål innenfor tilnærmet like omgivelser, f.eks. innenfor samme fagfelt. Det går også an å bruke Brier-scorene til å si noe om hvor «lett» eller «vanskelig» det i etterkant viser seg å være å predikere spørsmål innenfor ulike fagfelt. Endelig er det viktig å være klar over at ordinale spørsmål vil resultere i bedre Brier-scores enn kategoriske spørsmål. Med den kategoriske utregningsmetoden fikk Ole og Per en Brier-score rundt 0,79, mens med den vektete fikk de mellom 0,16125 og 0,3875. Det må derfor skilles mellom Brier-scores på ulike typer spørsmål, spesielt ordinale.

4.3.2 Median-score

Det andre målet på treffsikkerheten er Median-score. Median-score er ganske enkelt *medianen* til alle deltageres Brier-score på et spørsmål. Denne brukes bare til å beregne Accuracy-score.

4.3.3 Accuracy-score

Accuracy-score er det tredje målet på treffsikkerhet og måler prediksjonsevnen din *sammenlignet med de andre deltagerne*. Accuracy-score beregnes ved å trekke Median-score fra Brier-score på det aktuelle spørsmålet:

$$\text{Accuracy score} = \text{Brier score} - \text{Median score}$$

En *negativ* Accuracy-score betyr at deltageren er *bedre* til å predikere enn halvparten av deltagerne (medianen). En *positiv* Accuracy-score vil si at deltageren er *dårligere*.

Fordelen med Accuracy-scoren er at den utligner forskjellen mellom lette og vanskelige spørsmål. Så lenge Median-scoren trekkes fra, er det deltagerens *relative* prediksjonsevne som måles, uavhengig av om spørsmålet er «lett» eller «vanskelig». Dermed kan det også lønne seg å predikere spørsmål hvor usikkerheten er stor og anslaget vil ligge i nærheten av 50 %, så lenge man treffer bedre enn de andre deltagerne. På spørsmål hvor deltagerne ikke svarer, får de ingen Accuracy-score, som betyr at de hverken straffes eller premieres for å hoppe over et spørsmål.

Den største fordelen med Accuracy-score er at den kan brukes til å beregne sammenlagte resultater på basis av mange og ulike typer spørsmål. F.eks. vil *summen* av alle Accuracy-scorene brukes til å identifisere de sammenlagt beste deltagerne gjennom hele turneringen. For å undersøke deltagergruppers prediksjonsevne kan man bruke *snittet* av Accuracy-score, f.eks. ved å sammenligne snittet til kvinner og menn eller til ansatte i FD og FFI. Det er derfor deltagerens Accuracy-score som i hovedsak vil bli benyttet til å analysere treffsikkerheten til deltagerne i FFIs prediksjonsturnering.

4.4 Eksempel på en turnering

For å illustrere hvordan prediksjonsevnen kan måles ved hjelp av Brier-, Median- og Accuracy-scores, følger det her eksempler på beregninger av ulike typer spørsmål. La oss si at det arrangeres en turnering med fire deltagere, der hver av dem får totalt tre spørsmål og ett av hver type.

4.4.1 Binomiske spørsmål

Spørsmål	Vil norske F-16-fly bli involvert i militære operasjoner i Syria innen utgangen av 2018?
Spørsmålstype	Binomisk (ja/nei)
Instruksjoner	Dette inkluderer alle typer militære operasjoner, som ikke er trening, med eller uten bruk av våpen, inkl. alt fra tilstedeværelse til angrep.
Svaralternativer	Oppgi sannsynligheten for «ja».

La oss si at norske fly *ikke* blir involvert i militære operasjoner i Syria innen utgangen av 2018.

Følgende personer hadde oppgitt disse sannsynlighetsvurderingene på spørsmålet:

Deltagere	Ja	Nei	Brier-score	Median-score	Accuracy-score
Ole	10 %	90 %	0,02	0,0125	0,0075
Per	25 %	75 %	0,125	0,0125	0,1125
Ingrid	2 %	98 %	0,0008	0,0125	-0,0117
Stine	5 %	95 %	0,005	0,0125	-0,0075

Her er det Ingrid som treffer best med sitt anslag på 2 %. Hun får en Brier-score på 0,0008. Deretter følger Stine og Ole med 0,005 og 0,02 Brier-score for sine anslag på hhv. 5 % og 10 %. Per får en langt høyere Brier-score på 0,125 og kommer klart dårligst ut. Dette reflekteres også i Accuracy-scoren hans, som er 0,105 dårligere enn nest siste plass, mens forskjellen mellom tredje og førsteplass til sammenligning bare er på 0,0192 (mellom 0,0075 og -0,0117).⁴⁸

4.4.2 Kategoriske spørsmål

Spørsmål	Hvem vil vinne kategorien «rock» under Spellemansprisen 2017?
Spørsmålstype	Kategorisk, der alle gale alternativer er like feil
Instruksjoner	Vinneren publiseres 28. januar 2017.
Svaralternativer	A: «Pineapple Pizza» – Death By Unga Bunga B: «Vannmann86» – Hjerteslag C: «Nattesferd» – Kvelertak D: «Ført bak lyset» – Tusmørke E: The Gospel» – Årabrot

28. januar 2017 ble det annonsert at vinneren var C: «Nattesferd» – Kvelertak.

Følgende personer hadde oppgitt disse sannsynlighetsvurderingene på spørsmålet:

Deltagere	A	B	C	D	E	Brier-score	Median-score	Accuracy-score
Ole	60 %	20 %	10 %	5 %	5 %	1,215	0,94	0,275
Per	40 %	30 %	10 %	10 %	10 %	1,08	0,94	0,14
Ingrid	35%	5 %	50 %	0 %	10 %	0,385	0,94	-0,555
Stine	20 %	20 %	20 %	20 %	20 %	0,8	0,94	-0,14

Her er det Ingrid som gjør det klart best med en Brier-score på 0,385 og en Accuracy-score som er 0,415 bedre enn nestemann på listen. Det er en like stor forskjell (0,415) som mellom andre- og fjerdeplassen.

⁴⁸ Merk at Median-scoren her er gjennomsnittet av de to Median-scorene i midten siden antallet deltagere er et partall.

4.4.3 Ordinale spørsmål

Spørsmål	Hvor stor andel vil mene at Trump gjør en god jobb 28. april 2017?
Spørsmålstype	Ordinal, der noen gale alternativer er riktigere enn andre
Instruksjoner	
Svaralternativer	A: Under 35 %, B: 35–39 %, C: 40–44 %, D: 45 % eller over

28. april 2017 var det 43 % som mente Trump gjorde en god jobb. Dermed ble riktig svaralternativ C. Følgende personer hadde oppgitt disse sannsynlighetsvurderingene på spørsmålet:

Deltagere	A: Under 35 %	B: 35–39 %	C: 40–44 %	D: 45 % eller over	Brier-score	Median-score	Accuracy-score
Ole	25 %	25 %	25 %	25 %	0,25	0,215	0,035
Per	40 %	30 %	30 %	0 %	0,433	0,215	0,2183
Ingrid	0 %	30 %	30 %	40 %	0,167	0,215	-0,0483
Stine	10 %	40 %	40 %	10 %	0,18	0,215	-0,035

Som vi ser er det Ingrid og Stine som får best Brier- og Accuracy-scores. Ingrid får noe bedre score enn Stine, til tross for at Stine veide alternativ C tyngst, fordi Ingrid også vektla alternativene rundt (B og D) mer enn Stine. Ole, som gav alle svaralternativene lik sannsynlighet gjorde det, gjorde det bedre enn Per, som til tross for at også han anslo sannsynligheten for C for å være 30 %, vektla de lavere alternativene enda mer til sammen og av slo D som helt umulig.

4.4.4 Sammenlagte resultater

Til slutt kan deltagerens relative treffsikkerhet beregnes ut fra deres samlede Accuracy-score, mens deres objektive treffsikkerhet kan beregnes ut fra deres gjennomsnittlige Brier-score.

Rangering	Deltagere	Samlet Accuracy-score	Gj.snitt Brier-score
1. plass	Ingrid	-0,615	0,1843
2. plass	Stine	-0,1825	0,3283
3. plass	Ole	0,3175	0,495
4. plass	Per	0,4708	0,546

Ingrid skiller seg her ut som den som den klart mest treffsikre på de tre spørsmålene, fordi Accuracy-scoren hennes er mye bedre enn resten. Den samme reflekteres i den gj.snittlige Brier-scoren. Til sammenligning ville den pilkastende åpen, med lik sannsynlighet til alle alternativer på akkurat disse tre spørsmålene, fått en gj.snittlig Brier-score på 0,5167 $((0,5 + 0,8 + 0,25) / 3)$.

4.5 Begrensninger

Avslutningsvis er det viktig å minne om hva FFI prediksjonsturnering *ikke* kan si noe om.⁴⁹

Treffsikkerheten, slik den måles i denne turneringen, sier bare noe om evnen til å forutsi de forsvars- og sikkerhetspolitiske spørsmålene som blir stilt. Brier- og Accuracy-scorene til deltagerne sier altså ikke noe om evnen til å predikere spørsmål som *ikke* er med i turneringen eller spørsmål som ikke *kan* inkluderes, fordi de ikke oppfyller kravene til relevans, tidsspesifisering, tidsperspektiv og falsifiserbarhet.

Prediksjonsevnen er heller ikke et mål på generelt kunnskapsnivå. Det går an å predikere godt, uten å kunne forklare *hvorfor* det skjedde. Det å predikere enkelthendelser på kort sikt er altså ikke nødvendigvis det samme som å forstå det overordnede strategiske bildet best. Det vil også være andre typer vurderingsevner som kan være like verdifulle i forsvars- og sikkerhetspolitisk sammenheng. Spesielt gjelder dette evnen til å identifisere hendelser som ingen andre kommer på å spørre om, men som kan få store konsekvenser hvis de skulle skje. Evnen til å *stille* de gode spørsmålene og evnen til å *forutsi* svarene på dem, er derfor ikke den samme.

Kunnskap om et fagfelt, en aktør eller region har imidlertid begrenset nytteverdi for beslutningstagere, hvis den ikke også kan anvendes til å si noe om hva som vil skje i fremtiden. Det er her en målbar evne til å kunne forutsi relevante hendelser og utviklinger, over tid og med høyere presisjon enn andre, kan være betydningsfullt. Det å vite hva vi er gode på og ikke, hvem som er bedre enn andre og hva vi rett og slett ikke klarer å forutsi, er uansett et bedre utgangspunkt enn *ikke* å vite svaret på disse spørsmålene, selv om funnene skulle vise seg å være ubehagelige.

Brier-systemet har også noen begrensninger. Det er ifølge Tetlock et «*work in progress*».⁵⁰ Den største svakheten er at det ikke skilles mellom *false alarms*, der man hevder at noe vil skje, men det ikke gjør det, og *misses*, der man hevder at noe ikke vil skje, men så gjør det likevel.⁵¹ Mange vil hevde at det er mye farligere å *ikke* forutsi et militært anslag eller terroristangrep som faktisk skjer (*miss*), enn det er å hevde at det vil skje, uten at det gjør det (*false alarm*). Denne svakheten kan tas høyde for ved å straffe *misses* hardere enn *false alarms* – og informere deltagerne om dette, slik at de kan ta dette med i beregningen.⁵² Dette vil imidlertid ikke gjøres i FFIs turnering, fordi de fleste spørsmålene ikke vil handle om hendelser med så katastrofale konsekvenser at det vil være viktigere å varsle om dem, selv om de ikke skjer. Dette er også noe FFIs scenarier er spesielt utviklet for å gjøre i norsk forsvarsplanleggingssammenheng.

Selv med sine svakheter kan Brier-systemet derfor være et viktig forskningsbasert *supplement* til andre måter eksperter og forskere blir målt og vurdert i dag, som f.eks. antall publikasjoner, grad av selvtillit, evne til å fortelle en god historie, antall bøker solgt og siteringer i media.⁵³

⁴⁹ For mer prediksjonsevners begrensninger, se Tetlock (2005), *Expert Political Judgment*, ss. 14–15.

⁵⁰ Tetlock og Gardner (2015), *Superforecasting*, s. 261. For mer om måling av prediksjonsevne, se kapittel 12.

⁵¹ For mer om dette, se også Tetlock (2005), *Expert Political Judgment*, ss. 11–12 og Technical Appendix.

⁵² Tetlock og Gardner (2015), *Superforecasting*, s. 261.

⁵³ Tetlock og Gardner (2015), *Superforecasting*, s. 261.

5 Analyser

Etter tre år vil de fleste spørsmålene i turneringen være avgjort. Det vil da være mulig å studere resultatene mer systematisk. Her skisseres det tre forskningsspørsmål som vil være aktuelle å analysere. Målet er å bidra til å forbedre treffsikkerheten i beslutningsgrunnlaget for Forsvaret.

5.1 Det norske forsvars- og sikkerhetspolitiske miljøets prediksjonsevne

Et første spørsmål som turneringens resultater kan gi svar på er: *Hvor nøyaktig klarer det norske forsvars- og sikkerhetspolitiske miljøet å forutsi utviklinger av relevans for Forsvaret?*

Til forskjell fra Tetlocks studier, der deltagerne hovedsakelig var eksperter, akademikere og andre spesielt interesserte som var villige til å gjennomføre flere timer med tester i forkant, er minst halvparten av deltagerne i FFIs turneringen ansatt i forsvarssektoren, og omtrent en tredjedel arbeider med forsvars- og sikkerhetspolitiske spørsmål til daglig.⁵⁴

Ved å analysere Brier-scorene til disse deltagerne vil det være mulig å *tallfeste* treffsikkerheten til forsvars- og sikkerhetspolitiske miljøer – både generelt og innenfor bestemte temaer, aktører og geografiske områder av særlig relevans for Norge. Finnes det noen temaer eller typer utfall vi treffer bedre eller dårligere på enn andre? Hvilke temaer klarer vi *ikke* å forutsi? Svar på spørsmål som dette kan gi innsikt i hvilke typer premisser vi har grunn til å stole mer på enn andre.

Her kan det også være relevant å sammenligne treffsikkerheten til deltagerne med andre måter å forutsi fremtiden. Hvor mye mer eller mindre nøyaktig klarer deltagerne å predikere sammenlignet med tilfeldig gjetning eller ingen endring på alle spørsmål? Gitt at deltagerne treffer bedre enn slike algoritmer, kan det også være relevant å utforske andre måter å vekte prediksjonene på som kvalifiserer nytteverdien i en forsvarssammenheng, f.eks. å straffe hardere feilslåtte prediksjoner der konsekvenser er spesielt store (som et militært angrep i motsetning til et valgresultat).

5.2 Prediksjonsevnen til ulike grupper

Et annet spørsmål som vil være interessant å studere er: *Hvilke grupper deltagere treffer best?*

Tetlocks studier fant ingen klar sammenheng mellom prediksjonsevne og utdanningsnivå eller tidligere erfaring. Dette er funn som kan etterprøves i en norsk kontekst og med denne turneringens utvalg av deltagere. I tillegg vil det være mulig å sammenligne treffsikkerheten til andre grupper enn de Tetlock studerte, f.eks. kvinner og menn, unge og eldre, offiserer og forskere, og militære med ulik rang. Her kan det gjøres statistiske analyser av potensielt signifikante sammenhenger mellom hvem deltagerne er og treffsikkerheten deres. Slike funn kan f.eks. si noe om hvilke grupper sine prediksjoner som bør vektlegges mer enn andre i studier der vurderinger av fremtidige hendelser og utviklinger er en del av arbeidet. Fordi deltagerens navn og epostad-

⁵⁴ Per 14. desember 2017. Det forventes ikke store endringer i disse forholdene.

resser vil være registrert, vil det også være mulig å gjennomføre ytterligere spørreundersøkelser etter endt turnering, for å undersøke nye hypoteser som kan dukke opp underveis om hvem deltagerne er og treffsikkerheten deres.

Turneringen vil også ha informasjon om prediksjonsevnen til deltagere fra ulike deler av forsvarssektoren. Disse resultatene kan imidlertid bare si noe om treffsikkerheten til *deltagerne* fra de respektive miljøene, ikke om treffsikkerheten til miljøene *generelt*. Til det er utvalget per i dag ikke representativt eller stort nok. Når det likevel vises sammenligninger av treffsikkerheten til ulike forsvarsmiljøer, handler dette først og fremst om å stimulere til konkurranse mellom grener og avdelinger. Dette kan føre til økt deltagelsen, men er også et mål i seg selv, fordi konkurranseinstinkt var en av egenskapene som kjennetegnet *superforecasterne* til Tetlock.

5.3 De beste prognosemakerne

Det tredje forskningsspørsmålet som kan ha verdi for treffsikkerheten i fremtidige analyser av Norges forsvars- og sikkerhetspolitiske omgivelser er: *Hvem i Norge er de beste til å forutsi utviklinger av relevans for Forsvaret, og hva kjennetegner dem?*

Basert på turneringens endelige resultater kan det bli mulig å identifisere enkeltpersoner som har en systematisk høyere evne til å forutsi fremtiden enn andre deltagere og algoritmer. Disse personene kan f.eks. være de som havner blant de 10 % beste alle årene. Dersom det viser seg å være mulig å identifisere en gruppe «norske *superforecastere*», kan disse f.eks. brukes eller vektes tyngre i forbindelse med nye fremtidsstudier for Forsvaret.

En viktig forutsetning er imidlertid å ha nok data over prediksjonsevnen til individer over tid, slik at det ikke er tilfeldigheter som avgjør hvem som havner i toppen. Flere av lagene i IARPA-turneringen hadde vanskeligheter med å rekruttere og beholde antallet deltagere de trengte, fordi prediksjon var en tidkrevende aktivitet som tok rundt én time per uke.⁵⁵ FFIs prediksjonsturnering vil bli noe mindre tidkrevende enn i IARPA-turneringen, fordi det bare er snakk om én mail med fem spørsmål per måned. Det kan likevel bli en utfordring å opprettholde engasjementet gjennom turneringens varighet. For å øke motivasjonen til å fortsette å predikere vil det kåres årlige vinnere innenfor kategorier som «Beste militær», «Beste fagperson» og «Beste ekspert».

Et siste viktig poeng er at denne turneringen ikke er et kontrollert eksperiment. Det vil være ukjente variabler ved deltagerne som ikke avdekkes her, f.eks. grad av forhåndskunnskap om spesifikke spørsmål, tilgang på informasjon og hvor mye tid de bruker på å vurdere hvert spørsmål. Resultatene fra turneringen vil bare ha informasjon om hvor lenge deltagerne har undersøkelsen åpen i nettleseren og hvilke hovedkategorier av forsvars- og sikkerhetspolitiske områder de har kompetanse innenfor. En mulig oppfølgingsaktivitet etter at turnering er over, kan derfor være å studere nærmere hva som *kjennetegner* de beste deltagerne, f.eks. gjennom intervjuer med de norske *superforecasterne* om hvordan de innhenter og anvender informasjon til å predikere. En slik studie kan potensielt avdekke nye idéer om hvordan vi kan bli bedre.

⁵⁵ Se [‘The Aggregative Contingent Estimation Program: Predicting Global Events Through Crowdsourcing’](#).

A Brier-score

Brier-score er et scoringssystem som brukes til å måle hvor nøyaktig probabilistiske prediksjoner er. Glenn W. Brier utviklet dette systemet for å kunne måle hvor presise værmeldinger er.⁵⁶ Systemet kan bare brukes til å måle sannsynlighetsvurderinger med binære utfall; det vil si prediksjoner der noe enten skjer eller ikke skjer, f.eks. «regn» eller «ikke regn».

Her oppsummeres matematikken bak Brier-score-systemet. Først beskrives den vanligste måten å beregne Brier-score i dag, der skalaen går fra 0 til 1. Deretter beskrives Briers opprinnelige formel, som brukes i forsvars- og sikkerhetspolitiske prediksjonsturneringer som GJ Open. Her går skalaen fra 0 til 2. Til slutt forklares det hvordan denne formelen er tilpasset FFIs turnering.

A.1 Den vanligste formelen

Den vanligste matematiske formelen for Brier-score i dag, er:⁵⁷

$$\text{Brier score} = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

Her er N antall prediksjoner totalt (f.eks. antall dager et spørsmål er åpent), t er tidspunktet for prediksjonen (f.eks. dagen prediksjonen gjøres på), f_t er sannsynlighetsvurderingen på tidspunkt t , og o_t er utfallet på tidspunkt t , som er 1, hvis utfallet som predikeres skjedde, eller 0, hvis utfallet som predikeres *ikke* skjedde.

La oss si at du 13. mai ble spurt om predikere om det vil regne i Oslo de neste fire dagene. Det betyr at spørsmålet er åpent i fire dager ($N = 4$). Du predikerer følgende for hver av dagene (t):

Dag (t)	Prediksjon for «regn» (f_t)
14. mai ($t = 1$)	30 %
15. mai ($t = 2$)	40 %
16. mai ($t = 3$)	60 %
17. mai ($t = 4$)	90 %

La oss si at det bare regnet 16. og 17. mai. Da får utfallet (o_t) verdien 1 («regn») ved ($t = 3$) og ($t = 4$), mens verdien blir 0 («ikke regn») for ($t = 1$) og ($t = 2$).

⁵⁶ [Brier, Glenn W. \(1950\), 'Verification of Forecasts Expressed in Terms of Probability', *Monthly Weather Review*, Vol. 78, No. 1.](#)

⁵⁷ Se https://en.wikipedia.org/wiki/Brier_score og <http://www.statisticshowto.com/brier-score/>.

Ved bruk av formelen over summeres (\sum) alle scorene for hver dag (t), før denne summen multipliseres med ($\frac{1}{N}$), som gir den endelige Brier-scoren på dette spørsmålet:

Dag (t)	Utrekning ($(f_t - o_t)^2$)	Brier-poeng
14. mai ($t = 1$)	$(0,3 - 0)^2$	0,09
15. mai ($t = 2$)	$(0,4 - 0)^2$	0,16
16. mai ($t = 3$)	$(0,6 - 1)^2$	0,16
17. mai ($t = 4$)	$(0,9 - 1)^2$	0,01
Brier-score ($\frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$)		$(\frac{1}{4}) * 0,42 = 0,105$

Denne matematiske formelen over gir en Brier-score mellom 0 og 1. Dersom du predikerte helt riktig alle dagene, dvs. anslo 100 % sannsynlighet for regn de dagene det regnet og 0 % sannsynlighet for de dagene det ikke regnet, ville du fått en perfekt Brier-score på 0. Motsatt ville du fått en Brier-score på 1, hvis du bommet fullstendig på alle dagene. I eksempelet over fikk vi en Brier-score på 0,105. Til sammenligning er dette dobbelt så treffsikkert som ved tilfeldig gjettning, der det anslås 50 % for regn alle dagene, som ville gitt en Brier-score på $(0,5 - 1)^2 = 0,25$.

A.2 Den opprinnelige formelen

Til forskjell fra formelen i eksempelet over, tar FFI og GJ Open utgangspunkt i formelen Brier opprinnelig utviklet i 1950. Her er utregningen er noe annerledes og litt mer komplisert:

$$\text{Brier score} = \frac{1}{N} \sum_{t=1}^N \sum_{i=1}^R (f_{ti} - o_{ti})^2$$

Her er N antall prediksjoner totalt (f.eks. antall dager et spørsmål er åpent), t er tidspunktet for prediksjonen (f.eks. dagen prediksjonen gjøres på), R er antall kategorier eller mulige svaralternativer (som er 2 ved binomiske spørsmål og mer enn 2 for kategoriske og ordinale spørsmål), i er et svaralternativ (f.eks. «regn», «ikke regn», A, B eller C), f_{ti} er en sannsynlighetsvurdering for svaralternativ i på tidspunkt t (f.eks. 30 % for alternativ A, 40 % for B og 30 % for C på den første dagen), og o_{ti} er utfallet av svaralternativ i på tidspunkt t (som er 1, hvis svaralternativet som predikeres er riktig, eller 0, hvis svaralternativet er feil).

Forskjellen mellom den første og andre formelen er at i den siste formelen summeres differansen mellom prediksjonene (f_{ti}) og utfallet (o_{ti}) for *alle* svaralternativene i , ikke bare mellom prediksjonen for at det skjer (f_t) og det endelige utfallet (o_t).

I spørsmålet om hvorvidt det vil regne, er det derfor to mulige svaralternativer ($R = 2$): «regn» ($i = 1$) og «ikke regn» ($i = 2$). Ved bruk av den opprinnelige formelen gjøres det nå en sannsynlighetsberegning for *begge* svaralternativene for hver dag (t):

Dag (t)	Prediksjon for «regn» (f_{t1})	Prediksjon for «ikke regn» (f_{t2})
14. mai ($t = 1$)	30 %	70 %
15. mai ($t = 2$)	40 %	60 %
16. mai ($t = 3$)	60 %	40 %
17. mai ($t = 4$)	90 %	10 %

Med likt utfall (regn 16. og 17. mai) beregnes Brier-scoren med den opprinnelige formelen slik:

Dag (t)	Utrekning ($(f_{ti} - o_{ti})^2$)	Brier-poeng
14. mai ($t = 1$)	$(0,3 - 0)^2 + (0,7 - 1)^2$	0,18
15. mai ($t = 2$)	$(0,4 - 0)^2 + (0,6 - 1)^2$	0,32
16. mai ($t = 3$)	$(0,6 - 1)^2 + (0,4 - 0)^2$	0,32
17. mai ($t = 4$)	$(0,9 - 1)^2 + (0,1 - 0)^2$	0,02
Brier-score ($\frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$)		$(\frac{1}{4}) * 0,84 = 0,210$

Med denne formelen får vi en dobbelt så høy Brier-score som med formelen beskrevet først. Her går Brier-score-skalaen mellom 0 og 2, ikke 0 og 1. Tilfeldig gjetning (50 % alle dager) ville her gitt en score på $((0,5 - 1)^2 + (0,5 - 0)^2) = 0,5$, i motsetning til 0,25 som med den første formelen.

Mens den første formelen bare beregner forskjellen mellom prediksjonen for ett av alternativene og utfallet, beregnes her også forskjellen mellom alle prediksjonene og utfallene. Dette er et nødvendig steg hvis det bare er snakk om to mulige utfall, men er et nødvendig steg for å kunne beregne Brier-score på spørsmålet med flere alternativer, fordi sannsynlighetsberegningene vil fordeles på mer enn to mulige utfall.

Eksempel: Ole og Per får et spørsmål med tre alternativer (A, B og C). Ole predikerer A: 30 %, B: 70 % og C: 0 %, mens Per predikerer A: 30 %, B: 35 % og C: 35 %. La oss si at riktig svar ble A. Med den første, vanligste Brier-formelen vil begge få lik Brier-score, fordi begge trodde det var 30 % sannsynlig at svaret ble A. Med den andre, opprinnelige Brier-formelen ville imidlertid Ole få en dårligere Brier-score enn Ole, fordi scoren blir negativt påvirket av at Per var i overkant skråsikker på et galt svar når han anså det som 70 % sannsynlig at svaret ville bli B.

Dette ekstra steget i beregningen i Briers opprinnelige formel er også en forutsetning for å kunne regne ut Brier-scores ved ordinale spørsmål, der noen svar ligger «nærmere» det riktige enn andre (se kapittel 4.3). I likhet med den amerikanske prediksjonsturneringen GJ Open, tar derfor FFI utgangspunkt i Briers opprinnelige formel og skalaen mellom 0 og 2.

Som illustrert i eksemplene over var hensikten med Briers formel å kunne vurdere treffsikkerheten til værmeldingene, der utfallet (o_{ti}) kan variere fra dag til dag avhengig av om det f.eks. regner eller ikke. I prediksjonsturneringer som GJ Open og FFIs er det en annen type prediksjonsevne som skal måles. Her handler prediksjonsevne om å predikere én bestemt fremtidig hendelse, der utfallet (o_{ti}) av alle prediksjonene avgjøres av om denne hendelsen skjedde eller ikke.⁵⁸ Utfallet som prediksjonene vurderes opp mot, vil derfor ikke variere fra dag til dag.

I eksempelet over vil ikke spørsmålet den 13. mai være om det vil regne hver dag frem til og med 17. mai, men om det kommer til å regne *akkurat* den 17. mai. Denne typen spørsmål kan derfor også være åpent flere dager i forveien (N) og det kan gjøres nye prediksjoner hver dag (t), men utfallet avgjøres av om det regner på 17. mai, ikke om det regner de andre dagene. I GJ Open-turneringen gjøres dette ved at deltagerne kan logge seg på og oppdatere sine prediksjoner i lys av ny informasjon helt til spørsmålet avgjøres.

Med de samme prediksjonene som ble benyttet i eksempelet over vil Brier-scoren beregnes slik:

Dag (t)	Utregning ($\sum_{i=1}^R (f_{ti} - o_{ti})^2$)	Brier-poeng
14. mai ($t = 1$)	$(0,3 - 1)^2 + (0,7 - 0)^2$	0,98
15. mai ($t = 2$)	$(0,4 - 1)^2 + (0,6 - 0)^2$	0,72
16. mai ($t = 3$)	$(0,6 - 1)^2 + (0,4 - 0)^2$	0,32
17. mai ($t = 4$)	$(0,9 - 1)^2 + (0,1 - 0)^2$	0,02
Brier-score ($\frac{1}{N} \sum_{t=1}^N \sum_{i=1}^R (f_{ti} - o_{ti})^2$)		$(\frac{1}{4}) * 2,04 = 0,510$

Denne måten å beregne Brier-score på er nærmere beskrevet på GJ Opens [hjemmesider](#), i en [instruksjonsvideo](#) og på hjemmesidene til [Cultivate Labs](#), som leverer turneringsplattformen.⁵⁹

A.3 Formelen i FFIs prediksjonsturnering

FFIs prediksjonsturnering benytter en forenklet versjon av GJ Opens fremgangsmåte, fordi deltagerne i FFIs turnering bare vil gjøre én prediksjon per spørsmål ved bruk av spørreskjemaet og ikke ha mulighet til å oppdatere sine beregninger etter det. (N), som er antallet prediksjoner som gjøres totalt, vil derfor alltid være lik 1 i FFIs turnering. Derfor kan ($\frac{1}{N}$) fjernes, sammen med (t), som er indeksen til (N). (R) vil variere, avhengig av antallet svaralternativer.

⁵⁸ Se også [Campbell, Mary Pat \(2015\), 'What I've Learned from the Good Judgment Project', *Forecasting and Futurism*, July 2015, No. 11, ss. 17–22.](#)

⁵⁹ For en nærmere forklaring av hvordan Brier-score utregnes i GJ Open-turneringen, se [videoen](#) og [FAQen på turneringens hjemmeside](#), samt forklaringen på [Cultivate Labs' hjemmeside](#).

Formelen som brukes til å beregne Brier-score i FFIs turnering kan derfor forenkles til:

$$Brier\ score = \sum_{i=1}^R (f_i - o_i)^2$$

Her er R antall kategorier eller mulige svaralternativer (som er 2 ved binomiske spørsmål og mer enn 2 for kategoriske og ordinale spørsmål), i er et svaralternativ (f.eks. regn, ikke regn, A, B eller C), f_i er en sannsynlighetsvurdering for svaralternativ i (f.eks. 30 % for alternativ A, 40 % for B og 30 % for C), og o_i er utfallet av svaralternativ i (som er 1, hvis svaralternativet som predikeres er riktig, eller 0, hvis svaralternativet er feil).

17. mai-eksempelet vil derfor begrense seg til én sannsynlighetsberegning, f.eks. gjort 15. mai:

Dag	Prediksjon for «regn» (f_1)	Prediksjon for «ikke regn» (f_2)
15. mai	40 %	60 %

Fordi det regnet 17. mai blir utfallet (o_2) verdien 1, mens utfallet «ikke regn» (o_1) får verdien 0. Med den forenklede formelen vil man i FFIs turnering få følgende Brier-score på spørsmålet:

Dag	Utrekning ($f_i - o_i$) ²	Brier-poeng
15. mai	$(0,4 - 1)^2 + (0,6 - 0)^2$	0,72
Brier-score ($\sum_{i=1}^R (f_i - o_i)^2$)		0,72

Eksempler på hvordan Brier-scoren brukes til å beregne treffsikkerhet er beskrevet i kapittel 4.3.

Referanser

‘The Aggregative Contingent Estimation Program: Predicting Global Events Through Crowdsourcing’, *Federal Citizen Science and Crowdsourcing Toolkit*, 23. februar 2016, <https://crowdsourcing-toolkit.sites.usa.gov/ace-forecasting/>.

Baruch Fischhoff. Se Fischhoff, Baruch (1975), ‘Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty’, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 1, No.3, ss. 288–299.

Beadle, Alexander W. (2016), ‘Å forske på Forsvaret i fremtiden – muligheter, begrensninger og kognitive fallgruver’, *FFI-rapport 2016/01810* (Kjeller: Forsvarets forskningsinstitutt).

Beadle, Alexander W. (2017), ‘Er du bedre til å forutsi fremtiden enn ekspertene?’, *forskning.no*, 30. september 2017, <https://forskning.no/meninger/kronikk/2017/09/er-du-bedre-til-forutsi-fremtiden-enn-ekspertene>.

Beadle, Alexander W. og Diesen, Sverre (2015), ‘Globale trender mot 2040 – implikasjoner for Forsvarets rolle og relevans’, *FFI-rapport 2015/01452* (Kjeller: Forsvarets forskningsinstitutt).

Brier, Glenn W. (1950), ‘Verification of Forecasts Expressed in Terms of Probability’, *Monthly Weather Review*, Vol. 78, No. 1.

Bukkvoll, Tor; Glærum, Sigurd; Johansen, Iver; Diesen, Sverre og Lia, Brynjar (2014), ‘En gjennomgang av FFIs scenariogrunnlag for Forsvarets langtidspanlegging’, *FFI-rapport 2014/01154* (Kjeller: Forsvarets forskningsinstitutt). Begrenset.

Campbell, Mary Pat (2015), ‘What I’ve Learned from the Good Judgment Project’, *Forecasting and Futurism*, July 2015, No. 11, ss. 17–22.

Ekspertgruppen for Forsvaret av Norge (2015), *Et felles løft*.

Etterretningstjenesten (2017), *Fokus 2017*, https://forsvaret.no/fakta/_ForsvaretDocuments/Fokus2017-utskriftvennlig.pdf.

Fischhoff, Baruch (2007), ‘An Early History of Hindsight Research’, *Social Cognition*, Vol. 25, No. 1, ss. 10–13.

Fischhoff, Baruch og Beyth, Ruth (1975), ‘I Knew It Would Happen: Remembered Probabilities of Once – Future Things’, *Organizational Behavior and Human Performance*, Vol. 13, No. 1, ss. 1–16.

Forsvarsdepartementet (2016), Proposisjon til Stortinget (forslag til stortingsvedtak), Prop. 151 S (2015–2016), *Kampkraft og bærekraft: Langtidsplan for forsvarssektoren*.

-
-
- Johansen, Iver (2006), 'Scenarioklasser i Forsvarsstudie 2007: En morfologisk analyse av sikkerhetspolitiske utfordringer mot Norge', *FFI-rapport 2006/02664* (Kjeller: Forsvarets forskningsinstitutt).
- Joseph, Regina (2014), 'Keeping Score: Better Policy Through Improved Predictive Analysis', *5th International Conference on Future-Oriented Technology Analysis (FTA)*, Brussel, 27-28. november 2014, https://ec.europa.eu/jrc/sites/jrcsh/files/fta2014-t2practice_88.pdf.
- Kahneman, Daniel (2013), *Tenke, fort og langsomt* (Oslo: Pax Forlag).
- Kunda, Ziva (1999), *Social Cognition: Making Sense of People* (Cambridge, MA: Massachusetts Institute of Technology Press).
- Kuosa, Tuomo (2012), *The Evolution of Strategic Foresight* (Surrey: Gower).
- Lillekvelland, Tobias og Strand, Kari Røren (2014), 'En analyse av sluttrater og stillingsrotasjon i Forsvaret', *FFI-rapport 2014/00343* (Kjeller: Forsvarets forskningsinstitutt).
- Mellers, Barbara, Stone, Eric, Murray, Terry, Minster, Angela, Rohrbaugh, Nick, Bishop, Michael, Chen, Eva, Baker, Joshua, Hou, Yuan, Horowitz, Michael, Ungar, Lyle, og Tetlock, Philip (2015), 'Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions', *Perspectives on Psychological Science*, Vol. 10, No. 3, ss. 267–281.
- Mellers, Barbara; Ungar, lyle; Baron, Jonathan; Ramos, Jaime; Gurcay, Burcu; Fincher, Katrina; Scott, Sydney E.; Moore, Don; Atanasov, Pavel; Swift, Samuel A.; Murray, Terry; Stone, Eric og Tetlock, Philip E. (2014), 'Psychological Strategies for Winning a Geopolitical Forecasting Tournament', *Psychological Science*, Vol. 25, Issue 5, ss. 1106–1115.
- Moore, Don A. og Healy, Paul J. (2007), 'The Trouble with Overconfidence', *Psychological Review*, Vol. 115, No. 2, ss. 502–517.
- Roese, Neal J. og Vohs, Kathleen D. (2012), 'Hindsight Bias', *Perspectives on Psychological Science*, Vol. 4, No. 5, ss. 411–426.
- Taleb, Nassim Nicholas (2010), *The Black Swan* (NY: Random House).
- Tetlock, Philip (2005), *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton: Princeton University Press).
- Tetlock, Philip og Gardner, Dan (2015), *Superforecasting: The Art and Science of Prediction* (London: Random House Books).
- Tversky, Amos og Kahneman, Daniel (1974), 'Judgment under Uncertainty: Heuristics and Biases', *Science*, Vol. 185, No. 4157, ss. 1124–1131.

About FFI

The Norwegian Defence Research Establishment (FFI) was founded 11th of April 1946. It is organised as an administrative agency subordinate to the Ministry of Defence.

FFI's MISSION

FFI is the prime institution responsible for defence related research in Norway. Its principal mission is to carry out research and development to meet the requirements of the Armed Forces. FFI has the role of chief adviser to the political and military leadership. In particular, the institute shall focus on aspects of the development in science and technology that can influence our security policy or defence planning.

FFI's VISION

FFI turns knowledge and ideas into an efficient defence.

FFI's CHARACTERISTICS

Creative, daring, broad-minded and responsible.

Om FFI

Forsvarets forskningsinstitutt ble etablert 11. april 1946. Instituttet er organisert som et forvaltningsorgan med særskilte fullmakter underlagt Forsvarsdepartementet.

FFIs FORMÅL

Forsvarets forskningsinstitutt er Forsvarets sentrale forskningsinstitusjon og har som formål å drive forskning og utvikling for Forsvarets behov. Videre er FFI rådgiver overfor Forsvarets strategiske ledelse. Spesielt skal instituttet følge opp trekk ved vitenskapelig og militærteknisk utvikling som kan påvirke forutsetningene for sikkerhetspolitikken eller forsvarsplanleggingen.

FFIs VISJON

FFI gjør kunnskap og ideer til et effektivt forsvar.

FFIs VERDIER

Skapende, drivende, vidsynt og ansvarlig.

FFI's organisation



Forsvarets forskningsinstitutt
Postboks 25
2027 Kjeller

Besøksadresse:
Instituttveien 20
2007 Kjeller

Telefon: 63 80 70 00
Telefaks: 63 80 71 15
Epost: ffi@ffi.no

Norwegian Defence Research Establishment (FFI)
P.O. Box 25
NO-2027 Kjeller

Office address:
Instituttveien 20
N-2007 Kjeller

Telephone: +47 63 80 70 00
Telefax: +47 63 80 71 15
Email: ffi@ffi.no