

**Put your money
where your mouth is**

Prediksjonsmarkeders
styrker og svakheter i lys
av forskningen på
gruppebeslutningstaking

Øyvind Rogne

(30 studiepoeng)

19. mai 2008



Sammendrag

Hvordan kan man ta så gode gruppebeslutninger som mulig? Noen metoder som det anbefales å bruke er gruppediskusjon, Delphi-metoden og statistisk gruppe. Ingen av dem gjør hva gode metoder egentlig bør: Stimulere kritisk tankegang, la de som er mer smarte eller som har viktige opplysninger få mer innflytelse, eller å tilby insentiver. 'Prediksjonsmarked', er en slags måte å skaffe prediksjoner på ved å få folk til å inngå veddemål. Det har egenskapene som andre metoder mangler, noe som også leder til bedre prediksjoner. Noen eksperter har uttalt at prediksjonsmarkeder er utsatt for manipulasjon. For eksempel fra de som måtte ønske å sabotere prediksjonene. Forskning tyder på at dette ikke stemmer. Det er i praksis vanskelig å manipulere markeder.

Innhold

1	Introduksjon	4
1.1	Fokuset i oppgaven	5
1.2	Innsamling av informasjon	6
I	Metoder for gruppebeslutningstaking	8
2	Beslutninger basert på gruppediskusjon	9
2.1	Svakheter ved gruppediskusjon	9
2.1.1	Når gruppediskusjon blir flertallstyranni	9
2.1.2	Når gruppediskusjon gjør 'vondt verre'	11
2.1.3	Feil personer får innflytelse	13
2.1.4	Folk hermer etter hverandre	14
2.1.5	Mindre kreativitet	15
2.2	Styrker ved gruppediskusjon	15
2.2.1	Demonstrerbare beslutninger	16
2.2.2	Når flertallet ofte har rett	18
2.2.3	Tid og fornøyde medlemmer	18
2.2.4	Når forenklet tankegang dominerer	18
2.2.5	Oppsummering: I hvilke situasjoner fungerer gruppe- diskusjon?	19
2.3	Oppsummering: Gruppediskusjon i beslutningstaking . . .	19

3	Statistisk gruppe og Delphi-metode	20
3.1	Statistisk gruppe	21
3.2	Delphi-metoden	22
3.3	Når metodene blir til 'Flertallstyrannier'	23
3.3.1	Hvorfor oppstår flertallstyrannier?	23
3.3.2	Delphi-metoden under flertallstyranni	24
3.3.3	Hva kan gjøres med flertallstyrannier?	25
3.3.4	Oppsummering: Flertallstyrannier	26
3.4	Oppsummering: Statistisk gruppe og Delphi	26
II	Prediksjonsmarked	27
4	Prediksjonsmarkeder	28
4.1	Historikk	28
4.1.1	1900-1950: Politiske veddemål og Hayek	28
4.1.2	1980-2008: Veddemål våkner til live igjen	29
4.2	Konsepter	31
4.2.1	Noen grunnleggende forklaringer	32
4.2.2	Hva består et marked av?	34
4.2.3	Oppsummering	38
4.3	Hypotesen om markedseffisiens	38
4.4	Oppsummering av kapitlet	40
III	Diskusjon	41
5	Hva mangler de andre metodene?	42
5.1	Stimulere kritikk	42
5.1.1	Kritisk tankegang i prediksjonsmarkeder	43
5.1.2	Oppfordrer andre metoder til kritisk tankegang?	45
5.1.3	Oppsummering	45

5.2	Insentiver	45
5.2.1	Insentiver og prediksjonsmarkeder	47
5.3	Slippe til mindretallet	51
5.3.1	Innflytelse på prediksjonsmarkeder	52
5.3.2	Innflytelse på andre metoder	54
5.4	Oppsummering	54
6	Svakheter ved prediksjonsmarkeder	56
6.1	Baksiden ved innflytelse: Manipulasjon	56
6.2	Kvaliteten på studier	60
7	Konklusjon	62

Kapittel 1

Introduksjon

Dette er en 'kort' form for masteroppgave på Institutt for Informatikk på Universitetet i Oslo. Den er på 30 studiepoeng. Problemstillingen er gitt av veileder Geir Kirkebøen:

'Drøft prediksjonsmarkeders styrker og svakheter i lys av forskningen på gruppebeslutningstaking'.

Jeg har valgt å svare på denne måten:

Hvordan kan en gruppe mennesker ta så gode beslutninger som mulig? Mange tror det fungerer å utveksle synspunkter og prediksjoner ved å diskutere. Jeg stikker hull på oppfatningen av gruppediskusjon som en egnet metode i kapittel 2.

Det er i svært begrensede situasjoner hvor diskusjon faktisk viser seg å fungere. For eksempel når det er snakk om løsninger som lett kan demonstreres for andre. Vi behøver et bedre alternativ enn gruppediskusjon.

Opgaven utforsker derfor alternative metoder som eventuelt forbedrer beslutninger (kapittel 3 og 4), eller botemidler som kan gjøre at man unngår vanlige feil (kapittel 5). For eksempel å tilby insentiver, å få folk til å bli mer kritiske, eller å rett og slett unngå all diskusjon. En mer radikal metode er å sette opp et veddemål, og la de utfallene som folk setter mest penger på, være avgjørende for gruppebeslutningene (del II). Dette kalles 'Prediksjonsmarked' (Prediction market).

For eksempel hvis en gruppe forskere skal velge ut det riktige eksperimentet å investere tid og ressurser i, kan de inngå et veddemål om hvilke

hypoteser som vil vise seg å stemme, og eventuelt et veddemål om hvilke eksperimenter som vil vekke oppmerksomhet fra andre.

På prediksjonsmarkeder kan du kjøpe og selge veddemål akkurat som du kjøper og selger aksjer på aksjemarkeder. En nærmere forklaring av hvordan prediksjonsmarkedet fungerer, skal jeg gå inn på i et senere kapittel. Prediksjonsmarkeder blir til slutt sett i forhold til andre metoder i del III.

1.1 Fokuset i oppgaven

När vi i inköpsarbetet (vad gäller externa produkter, alltså inte egenproduktion) bestämmer oss för vilka jeans vi tror ska bli de största volymprodukterna så är det en blandning av känsla, trend, statistik och diskussioner mellan leverantören och oss som köpare

Jeg sendte mail til flere bukseleverandører og spurte dem hvordan de tok beslutninger. Sitatet er fra en direktør i et buksefirma. Det ser ut til å være representativt for liknende firmaer: Å holde diskusjon om eventuelle prediksjoner, basert på for eksempel tidligere statistikk og så ta beslutninger på grunnlag av dette (Winklhofer et al., 1996).

Opgaven fokuserer på situasjoner som småbedrifter kan befinne seg i når de skal ta gruppebeslutninger. For eksempel å komme på nye forslag til beslutninger eller å predikere salgstall som de kan basere beslutninger på. Dette kan ses i motsetning til sære situasjoner som er populære blant psykologer, for eksempel krasjlandinger på grunn av NASA eller invaderingen av grisebukta. Formålet med å fokusere på en konkret situasjon, er at jeg lettere kan finne ut hva slags forskning jeg kan vektlegge i konklusjonen min. I blant, så bruker jeg et tenkt eksempel med ledelsen i firmaet Gundas bukser A/S. De kan for eksempel være interessert i å finne ut leveringsdatoene på bukser, hvor store salg det blir, og sannsynligheten for at ulike buksemerker lykkes.

Jeg har valgt å fokusere på beslutningstaking i grupper på små grupper (cirka 3-30 personer), fordi jeg kan ende opp med forskjellige svar hvis det er snakk om større grupper.

Jeg fokuserer på *nøyaktighet*. En god prediksjon treffer godt. Nøyaktighet er sett på som det viktigste ved prediksjon av mange firmaer. Det er blant annet sett på som viktigere enn kostnader og hvor enkelt prediksjonene kan utføres (Winklhofer et al., 1996).

Jeg fokuserer også på hvordan man kan komme på gode idéer til beslutninger. Mange norske bedrifter behøver å forbedre idéene sine. 80 % av dem investerer i ressurser for å komme på nye forretningsidéer til produkter og tjenester, men bare 7 % sier de får gode resultater ut av investeringene¹.

Kort sagt, er fokuset i oppgaven på gruppebeslutningstaking for små grupper i bedrifter, hvor interessen er å gjøre nøyaktige prediksjoner eller å komme på gode forslag til beslutninger.

1.2 Innsamling av informasjon

Jeg samler inn artikler som fokuserer på empiriske observasjoner og eksperimenter, og overser de matematiske, for å unngå diskusjoner jeg ikke har tilstrekkelig kunnskap om. Det kan føre til at jeg overser noen eventuelle svakheter ved prediksjonsmarkeder (for eksempel kritikken som er tatt opp i Manski (2006)).

Jeg har brukt ISI Web (som inkluderer Social Science Citation Index) og Google Scholar da jeg søkte etter informasjon. Like søk er gjentatt i løpet av våren 2008.

For å forsikre at jeg hadde relevant nok stoff, har jeg sett på referanser fra viktige 'review-artikler' (artikler som oppsummerer forskning). Jeg har ikke fått tak i alle originalartiklene som demonstrerer fenomenet 'konjunksjonsfeilen', fordi en av dem er publisert i et japansk tidsskrift. Derfor bruker jeg sekundærkilder når denne feilen omtales. For å forsikre meg om at primærkilden er gjengitt på riktig måte, har jeg sett på flere av sekundærkildene som omtaler den. Jeg har bare måtte overse noen få artikler fordi universitetet ikke hadde betalt for tidsskriftene.

Jeg har kontaktet hovedforfatterne av viktige artikler de gangene jeg behøvde å presisere eventuelle uklarheter eller for å oppklare funn som ikke samsvarte med andre studier. Dette gjelder studiene Sunstein (2005), Pennock and Sami (2007), Robin Hanson og Scott Armstrong. Noen er kontaktet for å skaffe artikler som jeg ikke fikk tak i (for eksempel Hastie (1986)).

Jeg har også lagt ut problemstillinger for oppgaven på Google Groups diskusjonsforum for å høre synspunkter fra viktige personer som David Pennock, Robin Hanson, Cass Sunstein, Jed Christiansen, John Delaney (direktør for intrade.com), Chris Hibbert, Emile Servan-Schreiber og G. Tzi-

¹I følge en undersøkelse rapportert i bladet Grunder Januar 2008

ralis. De fleste er viktige navn innenfor prediksjonsmarked-feltet, og du vil se mange av dem dukker opp i oppgaven.

Del I

Metoder for gruppebeslutningstaking

Kapittel 2

Beslutninger basert på gruppediskusjon

Think of what a meeting is. It's some fat, obnoxious guy who talks for three minutes despite of the fact that he knows nothing. In the meantime, there's a woman who sits in the back and says nothing because she may feel her opinion isn't taken into account. And then there's the brown-noser, who wants to be senior VP and says anything the boss wants to hear

- Justin Wolfers, midasoracle.org (25/7-07)

Under gruppediskusjon, så utveksler man argumenter for og mot ulike beslutninger. Deretter, så kommer man til enighet om hva som bør gjøres. For eksempel en bedrift skal velge om de bør kjøpe inn flere olabukser. Et argument mot å gjøre det kan peke på at salgshallene gikk ned forrige gang det ble gjort. Argumenter for, kan peke på at en spørreundersøkelse viser at kunder liker å kjøpe olabukser.

2.1 Svakheter ved gruppediskusjon

2.1.1 Når gruppediskusjon blir flertallstyranni

Et 'flertallstyranni' oppstår når flertallet tar dårlige beslutninger på vegne av alle i gruppa.

I utgangspunktet er det flertallet som bestemmer utfallet av gruppediskusjoner. For eksempel forskning på juryer har vist at det er i 85 % av alle

tilfeller, hva flertallet i utgangspunktet hadde bestemt seg for, som avgjør hva slags beslutning grupper til slutt kommer frem til (Nemeth, 1986).

Vi vet også fra en veldig stor mengde med forskningslitteratur at når flertallet i en gruppe har kommet frem til feil svar, så lar mindretallet seg presse til å si seg enige med dem. Dette kan du se i oppsummeringen til Hogg and Vaughan (2005). For eksempel i en studie som er gjentatt over hundre ganger, blir gruppe-medlemmene bedt om å finne ut lengden på pinner. Hvis flertallet er falske forsøkspersoner som blir bedt om å oppgi feil lengde, så gjør ofte mindretallet akkurat det samme. ca. 36 % av tilfellene.

Innflytelsen fra flertallet er derimot mindre hvis gruppe-medlemmene ikke diskuterer med hverandre. Hvis folk i de samme pinne-eksperimentene ikke gjorde vurderinger i gruppe, begikk de samme feil bare 1 % av gangene. Og hvis de fikk lov til å være anonyme, eller oppga sine estimater uten at andre så på, ble prosentandelen med feil redusert dramatisk (Hogg and Vaughan, 2005). Det samme skjedde hvis medlemmer kommuniserte over datamaskin istedenfor å være på møte (Cinnirella and Green, 2006).

Når flertallet (eller mange personer) deler de samme opplysningene, så blir disse opplysningene vektlagt mer i de endelige beslutningene, enn hvis det bare er et lite mindretall som kjenner til dem. For eksempel hvis mange kjenner til en viss markedsrapport for salg av bukser, så blir den sett på som mer viktig enn hvis det er få som kjenner til den samme rapporten. Dette fenomenet går under navnet '*Common knowledge effect*' (CKE-effekt) (Sunstein, 2005; Gigone and Hastie, 1993).

Gigone and Hastie (1993) satte smågrupper til å diskutere seg frem til prediksjoner av karakterene på oppdiktete studenter. De fikk utdelt opplysninger som snittkarakter, oppmøte på skolen og oppmøte på forelesning. Hvis Gigone og Hastie gjorde slik at flere gruppe-medlemmer delte en opplysning, ble den også mer vektlagt i gruppeprediksjonen enn hvis det bare var noen få medlemmer som visste om den.

Stasser and Titus (1985) demonstrerer på en glimrende måte hva som kan være konsekvensene av at opplysninger som mange kjenner til blir vektlagt mer. De gir smågrupper oppgaven å velge mellom ulike jobb-kandidater. De tre kandidatene kalles Mr. Best, Mr. Ohum og Mr. Okay. Mr. Best har alltid flere positive egenskaper enn Mr. Ohum og Mr. Okay. Likevel, klarer Stasser og Titus å manipulere diskusjonsgruppene på en slik måte at Mr. Best ikke blir valgt. De gjør slik at negative opplysninger om Mr. Best blir delt, mens de positive opplysningene om Best er spredd på ulike personer. For eksempel prøvde Stasser & Titus å gjøre slik at ingen hadde

felles positive opplysninger om Mr. Best, og ingen felles negative om Mr. Okay. Da ble Okay valgt av de fleste. Bare 12 % av diskusjonsgruppene valgte Mr. Best, og 21 % valgte ham hvis de ikke diskuterte.

Grunnen til at kjente opplysninger er vektlagt mer, kan ha opphav i at slike opplysninger også har større sjanse for å bli nevnt av de som kjenner dem, enn de opplysningene som få personer kjenner til.

Kort sagt, hvis flertallet kjenner til en viss opplysning, så vektlegges den mer enn hvis mindretallet gjør det. Og hvis flertallet på forhånd har bestemt seg for å ta en viss beslutning, så er det stor sjanse for at gruppa også ender opp med å ta denne beslutningen. Det er derimot lite som tyder på at flertallet lar seg overtale av noen få, enn det er at noen få lar seg overtale av flertallet. Gruppediskusjon fører altså lett til et 'flertallstyranni'.

2.1.2 Når gruppediskusjon gjør 'vondt verre'

Hvis flertallet i en gruppe har sjanse for å begå en viss type feil, så kan en gruppediskusjon gjøre denne feilen enda verre.

Eksempler: Overoptimisme, sannsynlighetsvurderinger og økonomiske beslutninger

Feil i sannsynlighetsvurderinger kan bli enda verre av diskusjon. Et eksempel på dette er konjunksjonsfeilen. Det går ut på følgende: At to utfall A og B skjer samtidig, 'konjunksjonen' av A og B, vil alltid ha mindre sannsynlighet enn det A eller B skjer hver for seg. Mange tror likevel at A & B er mest sannsynlig. En måte å demonstrere konjunksjonsfeilen på, er å gi folk en personlighetsbeskrivelse av en hypotetisk person som heter Linda. Sørg for å ta med egenskaper som er sett på som typiske for feminister. Linda har for eksempel hovedfag i filosofi og er 31 år. Som student var hun opptatt av sosial rettferdighet og diskrimineringstemaer. Til slutt, ber du forsøkspersonene om å vurdere sannsynligheten på to påstander: *Linda er ansatt i bank, og Linda er feminist og ansatt i bank*. Den siste påstanden blir ofte sett på som mest sannsynlig, selv om det egentlig er den første påstanden som, i følge sannsynlighetsteori, er mest sannsynlig. Konjunksjonsfeilen har også blitt demonstrert med andre former for oppgaver. For eksempel trodde mange at tennisspilleren Bjørn Borg i 1980 hadde større sjanse for å tape første sett og vinne matchen, enn sjansen han hadde for å

bare tape første sett. Folk som blir bedt om å diskutere seg frem til løsninger i gruppe, er mer utsatt for konjunksjonsfeilen enn folk som blir bedt om å løse problemet alene. (Plous, 1993; Sunstein, 2005; Crupi, 2008).

Gruppediskusjon kan også føre til dårligere økonomiske beslutninger. En *'sunk cost'* er en kostnad som ikke lar seg gjøre om. For eksempel hvis du har kjøpt en kinobillett, og du ikke har rett til å få refundert den, så er betalingen en sunk cost. Hvis du da lar fremtidige avgjørelser (for eksempel om du bør se kinofilmen eller ikke) avhenge av pengene du først brukte på kinobillett, så gjør du en sunk cost feil (*'sunk cost fallacy'*). Hvis det for eksempel dukket opp noe mer interessant etter du hadde kjøpt billetten, så kunne du ha følt deg presset til å likevel se kinofilmen fordi du i utgangspunktet brukte penger på den. Men disse pengene er egentlig irrelevante, fordi de lar seg ikke refundere uansett. Grupper gjør flere sunk cost feil enn individer (Sunstein, 2005; Plous, 1993). Se f.eks. Smith et al. (1998).

Det er vanlig i psykologi å finne at folk er overoptimistiske. Dette gjelder for eksempel tidsbruk (vi oppgir ofte mindre tid enn hva vi egentlig kommer til å bruke) og selvinnsikt (vi er ofte mer sikre på hvor godt vi kan gjøre ting enn hva vi egentlig kan) (se f.eks. Kahneman et al. (1982).

Gruppediskusjon kan øke folks overoptimistiske tidsestimater. Buehler et al. (2005) satte studenter til å estimere hvor mye tid de kom til å bruke på ulike prosjekter. Det ene prosjektet var innlevering av en oppgave i et økonomifag, og det andre var å løse et puslespill. Noen ble bedt om å skrive estimatet på en lapp, mens noen ble bedt om å diskutere estimatet med prosjektgruppa. Både estimatene som folk kom frem til i grupper, var mer overoptimistiske enn estimatene til folk som ikke hadde deltatt i diskusjon. For eksempel for innleveringsprosjektene, så bommet de som hadde diskutert med 6 til 8 dager mer enn de som ikke hadde diskutert.

Gruppediskusjon fører også til større selvsikkerhet. Plous (1995) ga folk quiz-spørsmål som krevde estimer av tall, for eksempel året Isac Newton døde, og antall innbyggere i Australia. Et konfidensintervall, er et intervall som du tror at svaret kommer til å ligge innenfor en viss prosentandel av tilfellene. Hvis du oppgir et for kort intervall, så betyr det at du er altfor sikker på deg selv. Folk ble gitt konfidensintervaller hvor svaret skulle ligge i 90 % av tilfellene, og enten bedt om å vurdere svarene i gruppe, eller alene. Det viste seg at grupper oppga altfor korte konfidensintervaller i forhold til hva folk gjorde alene. For eksempel i et av eksperimentene til Plous, så hadde de som oppga intervaller alene, 7 til 9 antall riktige svar 84 % av tilfellene, mens de som oppga intervaller i gruppe hadde 7 til 9

riktige bare 8 % av tilfellene!

For flere eksempler på hvordan gruppediskusjon fører til økt overoptimisme, se også Sniezek (1992).

Svakheten ved eksperimentene vi har omtalt her, er at de fleste av dem bruker studenter som forsøkspersoner. Dette gjør at vi ikke vet så mye om hva som skjer når ekte eksperter eller kompetente personer diskuterer sammen.

Det vi har sett nå, er altså at gruppediskusjon kan føre til at folk gjør flere feilvurderinger enn hva de gjør uten diskusjon. For en rekke handlinger fra å ta feil i sannsynlighetsvurderinger til investeringsfeil, så kan det å la folk diskutere med andre bare 'gjøre vondt verre'.

Men hvorfor gjør gruppediskusjon ting verre?

'Gruppepolarisering' (group polarization) betyr at hvis folk vanligvis har et visst syn, så kan dette synet bli enda mer ekstremt etter at de også har sett hva andre mener. For eksempel hvis de fleste er i mot våpenlover, blir de enda mer mot våpenlover hvis de diskuterer dette med hverandre. Hvis de færreste er rasister, så vil det også bli enda færre rasister av diskusjon. Hvis de fleste i Gundas Bukser A/S tror at olabukser vil slå an, så kan de ende opp med å tro at olabukser vil slå an enda mer etter å ha sett hva andre mener om samme sak. Mange studier har demonstrert gruppepolarisering (Isenberg, 1986; Plous, 1993; Sunstein, 2005, 2002).

Gruppepolarisering har større sjanse for å oppstå etter diskusjon. Grunnen til dette, er at hvis det er mange argumenter for eller mot et tema, så lar folk seg også lettere overtale av dem (Isenberg, 1986). For eksempel vil de fleste på Gundas bukser argumentere mot å selge olabukser. Da vil mange lett la seg overtale, selv om argumentene kanskje ikke er så gode.

Hvis folk i utgangspunktet gjør en viss feil, så kan dette forsterkes av gruppepolarisering. Da vil folk også lett komme med argumenter som er feil når de diskuterer. For eksempel tenk deg at mange gjør konjunksjonsfeilen (beskrevet over). Da vil det også fort være mange argumenter for å tro at Linda er bankansatt og feminist, og færre for å tro at hun bare er bankansatt.

Kort sagt, så kan gruppepolarisering være en årsak til at gruppebeslutninger noen ganger forverres av gruppediskusjon.

2.1.3 Feil personer får innflytelse

En person som ikke bidrar med noe nyttig for gruppebeslutningene, kan fort få mer innflytelse enn andre i en gruppediskusjon. De som prater mer, og som kommer med flere retoriske argumenter, påvirker også gruppebeslutningene mer. De er til tross for dette, ikke nødvendigvis flinke til å foreslå beslutninger (Mendelberg, 2002).

Også de som har høy status, for eksempel i form av utdanning eller sosio-økonomiske forhold, får mer innflytelse enn andre. Det er konklusjonen til Hastie (1986) på grunnlag av forskningen han har gjort på gruppebeslutningstaking i juryer. Hastie kommer i tillegg frem til at det ikke eksisterer noen sammenheng mellom status, og hvor flink man er til å ta beslutninger.

Når det er feil person som påvirker gruppebeslutningene, så er det fort gjort at gruppa også blir 'dratt ned' av denne personen. For eksempel i Sprenger et al. (2007) så skulle grupper på 4 diskutere hvor populære ulike mobiltelefoner kom til å bli. Noen ble sett på av gruppedeltakerne som mer innflytelsesrike enn andre. Samtidig, så hadde nøyaktigheten til de med innflytelse, en sterk sammenheng med nøyaktigheten til gruppene de deltok i. Grupper som presterte dårlig ble altså dratt ned av å ha innflytelsesrike personer som også presterte dårlig.

2.1.4 Folk hermer etter hverandre

Rekkefølgen på hvordan folk snakker påvirker også hva de deler av informasjon med hverandre. Det er ikke alltid likegyldig at Per eller Unni snakker først. Per kan begynne å herme etter Unni, og la være å bidra med det han vet selv.

Et problem oppstår når mange hermer etter hverandre istedenfor å stole på egen kunnskap. Dette kalles en '*informasjonskaskade*' ('Informational cascade') (Sunstein, 2005). Eksperimentene på informasjonskaskader demonstrerer at flertallet i grupper ofte *kunne* ha kommet frem til riktige beslutninger hvis de bare hadde delt informasjonen sin mer med hverandre. Men ofte hermer for mange etter andre til at dette er mulig. For eksempel, Per har utført en trendanalyse som viser at hullete bukser selger bra. Kari og Ola får snakke først på møtet. De sier at de tror hullete bukser vil bli en flopp, og baserer konklusjonen på synsing. Andre gruppemedlemmer sluker hva Kari og Ola sier blindt. Per får snakke sist. Han tenker da at det ikke er noe poeng i å dele trendanalysen sin, fordi de andre kan ha god

grunn for å si det de sier. Dermed kan han la være å rapportere analysen til andre.

I typiske eksperimenter på informasjonskaskader (f.eks. Hung and Plott, 2001), trekker forsøkspersonene røde og hvite baller fra en urne. Oppgaven de får går ut på å gjette om det er mest av de hvite eller de røde ballene. Urna kan for eksempel ha en fordeling på 30 % hvite og 70 % røde. Hver person trekker ball på tur, og må oppgi til de andre fargen han eller hun tror er den riktige. Men det er ikke lov å oppgi fargen på ballen man har trukket selv. De fleste stoler mer på hva andre gruppemedlemmer har gjettet, enn på fargene de selv trekker. For eksempel hvis fire stykker sier 'rød ball', kan Ole bli lurt til å tro at den hvite ballen han har trukket ikke spiller noen rolle, og dermed også si 'rød ball'. Hvis det er mange som tenker som Ole, kan det føre til at de fleste i gruppa gjetter feil.

Slike eksperimenter kan ikke uten videre generaliseres til virkeligheten. I forretnings situasjoner blir man sjelden bedt om å gjette om det er flest av hvite eller røde baller i en urne, men vi blir kanskje bedt om å gjette salgstall og leveringsdatoer. Disse situasjonene krever for eksempel forskjellige former for kunnskap (f.eks. om hvordan salg og leveringsdatoer har vært tidligere).

Også andre studier finner at folk har lett for å herme etter hverandre. For eksempel studier på gruppepress finner ofte at folk har lett for å herme etter 'falske' forsøkspersoner som er satt til å gjøre vurderinger som ikke er riktige (Hogg and Vaughan, 2005). Det holder ofte at de falske forsøkspersonene er i flertall for å få folk til å herme etter dem.

Kort sagt, kan vi forvente at folk hermer etter hverandre i diskusjonsgrupper. Dette kan fort føre til dårligere beslutninger hvis de det hermes etter egentlig tar feil.

2.1.5 Mindre kreativitet

De som kommer på idéer under gruppediskusjon, kommer på færre idéer enn de som kommer på idéer hver for seg. Flere studier finner ut at det er best at gruppas medlemmer kommer på idéer uavhengig av hverandre, for deretter å kombinere dem. Likevel, viser det seg at deltakerne på diskusjonsgrupper blir mer fornøyde med idéene de kommer på enn hva folk blir når de kommer på idéer alene (Hill, 1982; Nijstad et al., 2006).

2.2 Styrker ved gruppediskusjon

I tillegg til de sårbarhetene jeg har trukket frem i dette kapittelet, går det an å finne flere fra psykologisk litteratur. Forskere på feltet har derfor kommet med ganske pessimistiske konklusjoner om det å holde diskusjoner ansikt til ansikt. Professor Scott Armstrong har publisert en artikkel med tittelen *How to make better forecasts and decisions: Avoid face-to-face meetings* (Armstrong, 2006). Der argumenterer Armstrong at man bør unngå å holde 'normale' møter med diskusjon. Sunstein (2005) skriver at

a central question is whether groups avoid the errors of the individuals who compose them. There is no clear evidence that they do, and there is considerable evidence that they do not (p. 30)

Armstrong og Sunstein sine konklusjoner holder ikke alltid. Det finnes unntakssituasjoner. Jeg skal hente frem evidens som støtter at man noen ganger bør holde gruppediskusjoner. De kan i noen situasjoner forbedre beslutningsprosessen. I tillegg tyder noe evidens på at diskusjon har positive bieffekter. Det går for eksempel an å diskutere for å gjøre hurtige beslutninger.

Under følgende betingelser, så fungerer det fint å holde gruppediskusjon:

1. Hvis det er lett å demonstrere løsningene på beslutningene som skal tas (med f.eks. logikk eller matematikk)
2. Hvis flertallet i gruppa ofte har rett
3. Hvis tidsbruk er viktig
4. Hvis det er viktig at folk har følelsen av å ha tatt de riktige beslutningene eller kommet på gode idéer
5. Når gruppedlemmer overforenkler tankegangen sin

Jeg skal kort gå inn på hver av dem.

2.2.1 Demonstrerbare beslutninger

En gruppe står ovenfor en 'demonstrerbar' beslutning når det går an å komme frem til klare svar. Man kan resonnerer seg frem til en demonstrerbar løsning med logikk, puslespill eller noe liknende. Den demonstrerbare

beslutningen har løsninger som alltid er like. Beslutninger under usikkerhet gir derimot ikke entydige svar, selv når du har all informasjon som er nødvendig for å ta dem. For eksempel når detektiver skal finne de skyldige, når statistikere skal predikere klimaet, eller når gamblere skal vedde på hester. Hvis man står ovenfor en usikker beslutning, så bør man helst unngå å diskutere. Men hvis man står ovenfor en demonstrerbar beslutning, bør man gjøre det.

Gigone and Hastie (1993) og Hastie (1986) gjennomgår studier som der små grupper predikerer eller tar beslutninger. De finner ingen forskjell på nøyaktigheten på det statistiske gjennomsnittet av prediksjonene folk gjør individuelt, og nøyaktigheten på prediksjoner de kommer frem til med diskusjon. De rapporterer også to tidligere review-artikler som finner det samme.

De finner derimot at diskusjonsgrupper fungerer bedre hvis det dreier seg om typiske hjernetrim-problemer, eller problemer som krever svar man logisk kommer frem til (altså problemer med 'demonstrerbare' løsninger), eller for typiske spørsmål du finner i quiz på puber, som har med fakta om verden å gjøre (f.eks. størrelsen på Australia).

Men studiene som er gjennomgått av Hastie og Gigone har to alvorlige svakheter. For det første, er det få av dem som er gjort på 'ekte' eksperter. De fleste bruker studenter, fordi det er lett for forskere å skaffe dem til eksperimenter (professoren som utfører eksperimentet trenger ofte bare å bevege seg fem meter ut av kontoret sitt). For det andre, er det få studier som ser på situasjoner som er relevante for min problemstilling - prediksjoner av salgstall, eller kvaliteten på produkter. Og i noen av de få studiene som ser på dette, gjør diskusjonsgrupper det bedre enn gjennomsnittet av individuelle prediksjoner (f.eks. Sniezek sin 1989-studie).

Baltes et al. (2002) gjør en meta-analyse av mange studier, og konkluderer med at grupper som kommuniserer ansikt til ansikt, tar beslutninger mer effektivt enn grupper som ikke gjør det. Men Baltes et al. (2002) er vage på hva de mener med 'effectiveness'. En nærmere titt på hvordan de har utført analysen viser at bare 3 studier som måler effektivitet, fokuserer på beslutningstaking under usikkerhet, mens 13 av dem fokuserer på beslutningstaking med lett demonstrerbare løsninger¹

Når det ikke er klare fasitsvar på beslutningen, så er folk mer utsatt for gruppepress enn hva de ellers ville ha vært. I de eksperimentene som for

¹Armstrong har generelt sett, uavhengig av om det er demonstrerbar eller usikker beslutning, kommet frem til at man bør unngå gruppediskusjon. Jeg har kontaktet han angående Baltes et al. (2002), og han er enig i at han har oversett studien

eksempel har prøvd å få medlemmer til å estimere lengden på pinner, så er det større sjanse for at gruppepress skjer med de som diskuterer *etter* at de har sett pinnene (ikke mens de ser pinnene). Grunnen til dette er at det da blir mer uklart hva folk har sett, og de har ikke lenger noe konkret å vise til (Hogg and Vaughan, 2005).

Kort sagt, så kan man bruke diskusjonsgrupper til beslutninger som har demonstrerbare løsninger. Hvis det ikke er snakk om demonstrerbarhet, så er det større fare for både at diskusjon ikke fungerer og for gruppepress.

2.2.2 Når flertallet ofte har rett

Tidligere, var oppgaven inne på det at i 85 % av alle tilfeller når juryer skal diskutere beslutninger, så er det hva folk før diskusjonen har bestemt seg for som fører til den endelige beslutningen (Nemeth, 1986). Da sier det seg selv at man bør være sikker på at de fleste har rett før man gir seg ut på gruppediskusjon.

2.2.3 Tid og fornøyde medlemmer

Hvor viktig er det at medlemmene blir fornøyde med beslutningene? Hvor viktig er tidsbruk?

Folk blir mer fornøyde med idéene hvis de får diskutere mens de kommer på idéer, enn hvis de må komme på idéer hver for seg (Nijstad et al., 2006). Dette skjer selv om folk er mer kreative når de arbeider alene.

Flere studier kan tyde på at diskusjon gjør at man tar beslutninger raskere enn hva man gjør uten (se f.eks. Baltés et al. (2002)).

2.2.4 Når forenklet tankegang dominerer

'Tilgjengelighetsheuristikken' (availability heuristic) er en slags tommelfingerregel hvor vi lar idéer som det er lett å tenke på avgjøre vurderinger av hyppigheter (f.eks. hvor ofte får man fallende flydeler i hodet, og hvor ofte blir man drept av hai?). Jeg har funnet noen studier som tyder på at diskusjon reduserer feil som oppstår med tilgjengelighetsheuristikken². Du kan

²Jeg har ikke fått tilgang på de opprinnelige studiene, så det er en sekundærkilde fra Benbasat and Lim (2000)

demonstrere heuristikken ved å spørre noen hvor frekvente ulike bokstaver er som første og tredje posisjon på ord. Den første posisjonen er lettere å tenke på. Mange bokstaver som du kan finne oftere på tredje posisjon, er sett på som mer sannsynlige å dukke opp på første posisjon. For eksempel mange tror at R oppstår mer som den første bokstaven i ord enn den gjør som den tredje (men det motsatte er egentlig tilfellet) (Plous, 1993).

Stasson et al. (1988) rapportert i Benbasat and Lim (2000) fant at grupper på 4 hadde mindre sjanse for å begå denne feilen enn individer. Slovic et al. (1978) også rapportert i Benbasat and Lim (2000) ba deltakere om å estimere frekvensen på ulike dødsårsaker. Grupper var ofte mer nøyaktige enn individuelle bedømmelser, fordi de vurderte dødsårsaker som det var lett å tenke på som mindre sannsynlig enn hva individene gjorde.

2.2.5 Oppsummering: I hvilke situasjoner fungerer gruppediskusjon?

Jeg har satt opp en oversiktlig tabell som oppsummerer det som er skrevet.

Situasjon	Bruke gruppediskusjon?
Idéfasen	Nei
Demonstrerbare Løsninger	Ja
Usikre løsninger	Nei
Tid viktig	Ja
Vanskelige beslutninger	Nei
Lette beslutninger	Ja
Fornøyde gruppemedlemmer viktig	Ja
' Forenklet tankegang	Ja

2.3 Oppsummering: Gruppediskusjon i beslutningstaking

Gruppediskusjoner kan fort ende opp som 'flertallstyrannier'. Det vil si at flertallet i gruppene avgjør beslutningene, selv når det ikke er noe som skulle tilsi at de burde gjøre det.

Hvis de fleste i utgangspunktet gjør en viss feil, så kan en diskusjon mellom dem, bare forverre feilen (med noen unntak). F.eks. hvis de fleste i utgangspunktet er overoptimistiske, kan de bli enda mer overoptimistiske etter å ha hatt en diskusjon med andre.

Men det finnes selvfølgelig også situasjoner hvor diskusjon er effektivt, for eksempel hvis man vil få unnagjort beslutningene hurtig.

Kapittel 3

Statistisk gruppe og Delphi-metode

You know how dumb the average guy is? Well, by definition, half of them are even dumber than *that*

J. R. 'Bob' Dobbs

For å forutsi fremtiden, kan vi spørre hver enkelt person i en gruppe om hva vedkommende tror. Så kan vi ta gjennomsnittet av hva alle har sagt. For eksempel Tor sier 3, Per sier 4 og Gunda 8. Gjennomsnittet av dette er $3 + 4 + 8/3 = 5$. Dette kalles '*statistisk gruppe*'.

En videreutviklet versjon av statistisk gruppe, går ut på å gi alle tilbakemelding etter å ha samlet inn deres prediksjoner. Tilbakemeldingen kan inneholde for eksempel gjennomsnittet og begrunnelser for andres prediksjoner. Vanligvis sørger vi også for at tilbakemeldingen holder folk anonyme. Etter tilbakemeldingen, spør vi gruppemedlemmene om prediksjoner på nytt. Så gir vi en ny tilbakemelding, og så videre. Vi fortsetter på denne måten til prediksjonene til en viss grad har blitt like hverandre. Dette går under navnet '*Delphi-metoden*' (Jones and Hunter, 1995). For eksempel hvis tre personer hver for seg har sagt 1, 2 og 3, så vil de først få tilbakemeldingen at gjennomsnittet er 2. Så blir de spurt på nytt. Denne gangen svarer de kanskje 2, 2 og 3 (med andre ord har de blitt mer enige).

3.1 Statistisk gruppe

En statistisk gruppe, kan du se på som en form for spørreundersøkelse. I en *spørreundersøkelse*, så stiller man mange personer de samme spørsmålene. De kan for eksempel stilles over internett, telefon, eller på papir. De som deltar i undersøkelsen er dessuten oppsøkt uavhengig av hverandre. F.eks. ved at tilfeldige personer plukkes ut fra telefonkatalogen. Gode eksempler på spørreundersøkelser er de du ofte ser i media like før politiske valg, og de du ser etter politiske debatter. Deltakerne blir her spurt om hvilke politikere de kommer til å stemme på. Svarene kan brukes for å predikere hvilke politikere som med størst sjanse vil bli valgt.

Statistisk gruppe er en form for spørreundersøkelse hvor man blir bedt om å gjøre kvantitative vurderinger av tall. F.eks. Ole mener at Per kommer til å spise 10 bønner, men Katrine mener 20 bønner. Vi kan da sette sammen enkeltprediksjoner til en gruppeprediksjon. Gjennomsnittet av Ole og Katrine sine prediksjoner er $(10 + 20)/2 = 15$. Det går selvfølgelig også an å sette sammen prediksjonene på andre måter (f.eks. med median), men jeg skal ikke gå inn på dem her.

Det statistiske gjennomsnittet av mange prediksjoner er ofte mer nøyaktig enn hver enkelt av dem. Det betyr at hvis du har mange prediksjoner, og rangerer dem etter hvor nøyaktige de har vært - på førsteplass, andreplass, og så videre, så vil ofte prediksjonen som er basert på det statistiske gjennomsnittet rangeres høyt opp. Dette resultatet kan du ofte få uansett hva det er du tar gjennomsnittet av. Basert på flere studier, finner Armstrong (2001) ut at feilene i gjennomsnitt reduseres med 12.5 %¹ når man tar gjennomsnittet av mange metoder (ekspertvurderinger, tall fra undersøkelser og statistikk) istedenfor å bruke metodene hver for seg.

Det samme funnet gjelder når man tar statistiske gjennomsnitt av medlemmene i grupper. Statistiske gjennomsnitt av mange individers prediksjoner, er ofte bedre enn hvert enkeltindivid sine prediksjoner. Hill (1982), Giguere and Hastie (1997) og Hastie (1986) rapporterer flere studier som finner dette.

Condorcet's jury-teorem er et matematisk bevis som også sier at flertallet i en gruppe, gjør mer nøyaktige prediksjoner enn enkeltpersoner eller flertallet i grupper på mindre størrelse. Teoremet stiller noen krav. For det første, kan ikke folks prediksjoner avhenge av hverandre. Du må for eksempel unngå at de har snakket med hverandre. For det andre, kan ikke

¹målt i Mean Absolute Percentage error. Det vil si gjennomsnittlig prosentandel bom på det riktige utfallet

folk bare gjøre prediksjoner i blinde. Hvis det er snakk om N utfall å forutse (f.eks. for fotballkamper er $N = 3$ hvor de tre utfallene er hjemme, borte og uavgjort), så må hvert gruppemedlem ha rett over $1/N$ av gangene de blir spurt (Sunstein, 2005; Boland, 1989).

I flere situasjoner, kan statistisk gruppe være like bra, og noen ganger bedre enn gruppediskusjon (Gigone and Hastie, 1997). For eksempel der vi mistenker at flertallet kommer til å gjøre en spesifikk feilvurdering (som overoptimisme). Eller der vi mistenker at de opplysningene mange sitter med er vektlagt mer i en gruppes beslutninger enn de opplysningene som noen få kjenner til (se kapitlet om gruppediskusjon). Da kan det nemlig hende at det bare gjør vondt verre å la folk diskutere med hverandre, og dermed bedre å bare ta gjennomsnittet av hva de predikerer. Flere studier tyder på at statistiske grupper er like bra eller bedre enn gruppediskusjon på å predikere fremtiden (Hastie, 1986; Gigone and Hastie, 1997). I noen av studiene Gigone and Hastie (1997) bruker som eksempler, forbedres nøyaktigheten med ca. 20 %.

Kort sagt, så får du ofte mer nøyaktighet ut av en statistisk gruppe enn hva du får ut av enkeltindivider eller andre statistiske grupper som er mindre i størrelse. Statistisk gruppe ser også ut til å være et lovende alternativ der gruppediskusjon svikter som effektiv metode.

3.2 Delphi-metoden

Når du bruker statistisk gruppe, mister du muligheten for at gruppemedlemmer kan utveksle informasjon med hverandre. Du bare spør folk hva de synes uten å la dem høre med andres synspunkter.

Det kan tenkes at for eksempel Gunn gjør en prediksjon som har sterkere støtte enn Per, fordi hun baserer seg på en markedsrapport. Per baserer seg derimot bare på historier som han har hørt fra sine kolleger i matpausen. Han kunne kanskje ha endret mening hvis han visste at Gunn hadde mer håndfast grunnlag enn han.

Delphi-metoden lar deltakerne bli bevisste på andres synspunkter i tillegg til å først, som med normal statistisk gruppe, ta gjennomsnittet av alle prediksjoner. Etter man har tatt gjennomsnittet, så utdeles tilbakemeldinger til deltakerne på hva dette gjennomsnittet er. Deretter, så holdes en ny 'runde' hvor alle gir nye prediksjoner. Hovedmålet med å bruke flere runder, er at deltakerne skal bli så enige med hverandre som mulig. Prediksjoner som er basert på gjennomsnitt blir ofte mer nøyaktige av det å

bruke runder (Rowe and Wright, 1999). I følge boka 'Principles Of Forecasting', bør hele prosessen foregå over cirka 2-3 runder (Armstrong, 2001). Det er mange varianter av Delphi-metoden. Den jeg holder meg her, er er beskrevet i Jones and Hunter (1995).

De fleste studier tyder på at Delphi fungerer bedre enn andre prediksjonsmetoder for grupper. Rowe and Wright (1999) går gjennom en rekke studier, og finner ut at Delphi kommer bedre ut enn statistisk gruppe i 12 av 14 studier. Delphi kommer dessuten bedre ut enn gruppediskusjon i 5 av 6.

Rowe og Wright advarer mot å gjøre for raske generaliseringer fra disse studiene til den 'virkelige verden'. Forsøkspersonene har ofte hatt mindre ekspertise og lavere personlig engasjement enn hva man har i virkelige situasjoner. Hvis du dessuten ser nærmere på studiene, så er det bare fire av dem som ser på prediksjoner av salgstall og økonomiske forhold som er relevante til mitt eksempel (Gundas bukser A/S). Resten av dem ser på hvordan folk svarer på generelle spørsmål, som hvor store ulike land er, eller spørsmål som er typiske for quiz på puber. Og av de 4 som har størst relevanse, så er det 2 hvor diskusjonsgrupper slår Delphi-metoden.

Kort sagt, så får gruppemedlemmene muligheten til å utveksle informasjon med hverandre med Delphi-metoden. Med Delphi kan man på lik linje med statistisk gruppe, holdes både anonym og separert fra hverandre. De fleste studier tyder på at Delphi slår gruppediskusjon og statistiske grupper.

3.3 Når metodene blir til 'Flertallstyrannier'

Et 'flertallstyranni' er når flertallet i en gruppe bestemmer beslutningene på vegne av alle de andre, selv når disse beslutningene ikke er gode. De få som kjenner til de beste beslutningene, kan bli overkjørt av de mange som tar feil.

3.3.1 Hvorfor oppstår flertallstyrannier?

Hvordan har det seg slik at flertallet på en gruppe tar feil på spørsmålene som blir stilt?

De kan ta feil av mangel på kunnskap. For eksempel Hanson (2003) trekker frem undersøkelser hvor flertallet av amerikanske deltakere svarer ja

på spørsmål som 'astrology has some scientific truth', 'US government is hiding that it knows of the existence of Aliens' og 'I believe in angels'.

Flertallet kan også tenke på måter som ikke er i tråd med teoriene som forteller hvordan man egentlig bør tenke ('normative' teorier). For eksempel i forhold til matematiske beregninger eller sannsynlighetsteori. Konjunksjonsfeilen som jeg har nevnt tidligere er et eksempel på dette. Et generelt funn fra psykologi, er at folk ofte ikke tenker i tråd med normative teorier (f.eks. Plous (1993) og Hastie and Dawes (2001)).

Flertallet kan dessuten ta feil på områder hvor det er komplisert å forutse fremtiden. For eksempel fotball (rykker LSK ned?) og politikk (Øker bruttonasjonalproduktet til Zimbabwe etter å ha byttet ut regjeringen?). Disse områdene har mange indikatorer på hva som kommer til å skje. Fotball har for eksempel indikatorer som skader før kamp, hjemmebane/bortebane, trener, strategi, tidligere spillerhistorikk, innbyrdes lagstatistikk og tabellstatistikk. Dermed blir det utrolig komplisert å forutse fremtiden. Noen ganger er det mulig å kaste mynt og krone eller terning og likevel gjøre det bedre eller like bra som ekspertene på slike områder. Philip Tetlock har helt siden 80-tallet registrert rundt 80 000 prediksjoner fra eksperter på politikk.. De har stort sett ikke vært bedre enn hva det hadde vært å kaste mynt og krone (Tetlock, 2005). Andersson et al. (2005) undersøkte hvordan eksperter (som hardbarka fans og sportsjournalister) predikerte utfall i fotball-VM 2002. De fant at ekspertene ble slått av enkle statistikker som plassering på tabellen.

Kort sagt, så kan flertallstyrannier oppstå hvis det er komplisert å spå fremtiden og hvis man rett og slett mangler kunnskap.

3.3.2 Delphi-metoden under flertallstyranni

Som sagt, så er Delphi-metoden en slags statistisk gruppe med flere 'runder'. Runder kan forbedre prediksjonsnøyaktigheten i følge studiene jeg diskuterte tidligere. På en annen side, så kan de gjøre den dårligere hvis det finnes et flertallstyranni. Hvis det er slik at mange mennesker tar feil, så kan de få som egentlig har rett, føle seg presset til å si det samme som alle andre.

For eksempel kan det tenkes en situasjon hvor man vanligvis gjør overoptimistiske prediksjoner av tidsbruk. Anta at riktig antall timer tidsbruk til slutt viser seg å være 1000. 80-90 % av alle som skal predikere den oppgir rundt 20 timer, og resten oppgir 800. De få som har oppgitt 800, oppdager etter tilbakemeldingen at alle andre har oppgitt noe som er annerledes. De

føler seg dermed presset til å gjøre prediksjonene sine mer like hva alle andre har sagt. Dette kan videre føre til at gjennomsnittet av alles prediksjoner blir dårligere etter flere runder enn hva det hadde blitt uten runder. For eksempel hvis mindretallet forandrer prediksjonene sine fra 800 til cirka 600 timer, så vil gjennomsnittet av alles prediksjoner bli 390 etter første runde, og 310 etter andre.

Teorien om at runder fører til gruppepress har støtte fra to studier. Rowe et al. (2005) stilte grupper spørsmål som krevde ja/nei svar. For eksempel 'et ran kommer til å skje i England', 'revolusjon på Filippinene', 'jordskjelv med over 100 døde'. Alle spørsmålene hadde betingelsen 'innen 2 måneder'. Hvis det var snakk om spørsmål hvor man vanligvis hadde rett cirka. 80 % av tilfellene, hadde flertallet også ofte rett. Runder forbedrer i de situasjonene nøyaktigheten med omtrent 5 %. Runder gjorde derimot prediksjonene omtrent 5 % dårligere på spørsmål hvor et lite mindretall ofte hadde rett og flertallet tok feil (nøyaktighet på under 30 %). Studien til Sniezek (1990) viser liknende resultater. Sniezek studerte prediksjoner innenfor økonomi, for eksempel prediksjon av rentenivå. Runder gjorde Delphi-metoden dårligere enn statistiske grupper da det var snakk om spørsmål som var betegnet som vanskelige.

Kort sagt, kan Delphi-metoden fungere dårligere enn en statistisk gruppe hvis man befinner seg i situasjoner hvor gruppeflertallet har lett for å ta feil.

3.3.3 Hva kan gjøres med flertallstyrannier?

For å unngå flertallstyrannier, så kan vi la de få som er flinkere enn andre eller som har viktigere informasjon også få mer innflytelse. For eksempel anta at Ola sier tallet 5 og at Kari sier 8. Du kan se på det som at begge disse tallene er automatisk multiplisert med 1 med Delphi-metoden eller statistiske grupper. Men hvis Ola er flinkere enn Kari, så bør kanskje hans femtall heller multipliseres med 2 eller 3? Å utføre dette er lettere sagt enn gjort. I praksis, er det vanskelig å få til.

Det fungerer ikke å la folk selv bestemme innflytelsen de bør ha på gruppebeslutningene. Dette er konklusjonen til flere forskere. For eksempel at Ola selv får bestemme at prediksjonen hans bør multipliseres med 2. Det fungerer heller ikke å prøve å vektlegge de personene som er flinkere enn andre, fordi det er vanskelig å skaffe håndfaste bevis på hvem som virkelig er flinkere enn andre i det lange løp (Se Armstrong (2001) og Hastie (1986)). For eksempel et prinsipp i boka 'Principles of Forecasting' er at alle

bør vektlegges likt. Boka begrunner dette med flere studier som finner det vanskelig å identifisere personene som er flinkere enn andre (Armstrong, 2001).

3.3.4 Oppsummering: Flertallstyrannier

Et 'Flertallstyranni' hvor flertallet bestemmer hva alle andre skal gjøre, selv når det fører til dårligere beslutninger, kan oppstå på kompliserte områder med mange indikatorer som fotball og politikk, og spesielt hvis det er vanlig å ta feil eller gjøre dårlige prediksjoner.

Det er ikke lurt å bruke Delphi-metoden hvis du mistenker at flertallstyrannier kan oppstå. Den fører til at de få som har rett føler seg presset til å si det samme som alle andre. Det idéelle, ville ha vært å gi de flinke gruppemedlemmene mer innflytelse enn andre, både i statistiske grupper og i Delphi-metoden. Det er uheldigvis vanskelig i praksis å finne ut hvem dette er.

3.4 Oppsummering: Statistisk gruppe og Delphi

Det finnes to metoder som kan brukes på beslutningsstøtte som holder gruppemedlemmer separerte fra hverandre og anonyme. Begge av dem virker lovende i forhold til andre alternativer (eksperter og gruppediskusjoner), men de kan svikte i situasjoner hvor det er lett å ta feil. Da kan flertallet få uberettiget makt, og de få i gruppa få som vet hva de driver med, får ikke innflytelse på prediksjonene. Delphi-metoden er i slike situasjoner, et dårligere alternativ enn en statistisk gruppe.

Del II

Prediksjonsmarked

Kapittel 4

Prediksjonsmarkeder

Fundamentally, in a system in which the knowledge of the relevant facts is dispersed among many people, prices can act to co-ordinate the separate actions of different people in the same way as subjective values help the individual to co-ordinate the parts of his plan (Hayek, 1945)

I løpet av årene som kommer, er det 15 % sikkert at USA kommer til å angripe Iran og 14 % sikkert at en astrologisk metode viser seg å fungere. Obama kommer dessuten til å bli presidenten i 2008, og i 2040 kommer verden til å få første mann på Mars. Alle disse prediksjonene kommer fra noe som kalles *Prediksjonsmarked*. Det er en måte å spå fremtiden på ved å få folk til å inngå veddemål med hverandre. Formålet med dette kapittelet, er å gi en generell innføring i Prediksjonsmarkeder.

4.1 Historikk

4.1.1 1900-1950: Politiske veddemål og Hayek

For 100 år siden, var det veldig vanlig å spå politiske valg basert på hvordan folk investerte i veddemål. Pressen elsket å rapportere fra veddemål på politiske kandidater. Antropologen Andrew Carnegie uttalte til New York Times på en pressekonferanse så tidlig som i 1904:

From what I see of the betting, ... I do not think that Mr. Roosevelt will need my vote. I am sure of his election... (Rhode and Strumpf, 2004)

Men så kom 1940-tallet, og interessen for å spå med veddemål ble borte på grunn av 'vitenskapelige' spørreundersøkelser. De kunne reklamere med å ha stikkprøver av befolkningen som var representative for de fleste folkeslag (mørkhudede, blonde, bønder, arbeidere, osv..) (Rhode and Strumpf, 2004).

I 1945, kom professor Hayek med artikkelen '*The use of knowledge in society*'. Denne artikkelen ble utrolig populær, og har i etterkant blitt referert av 2885 artikler og bøker¹.

Hayek (1945) argumenterer at en markedspris er et slags middel som mange mennesker kan bruke for å kommunisere kvaliteten på varer med hverandre. Dette skal være bedre enn at noen få sitter og synser. For eksempel er den vanlige prisen på en elektrisk tannbørste akkurat nå, cirka. 240 kroner. Hvis regjeringen hadde satt ut et ekspertpanel som fant ut at prisen burde være under 100 kroner, eller at den burde være over 1000 kroner, så kunne konsekvensen fort vært at enten kjøperne ikke hadde klart å betale, eller at selgerne ikke hadde fått god nok fortjeneste på å selge. På en annen side, så har du den prisen selgere og kjøpere av elektriske tannbørster har avtalt på egen hånd mellom hverandre. Hayek mener at denne prisen er mer 'treffsikker' når det gjelder hvor godt kjøpere og selgere kommer ut av handelen.

Hayek argumenterer også at ulike personer sitter på veldig spesifikke biter av den kunnskapen som må til for å forutse fremtiden. Bare markedspriser kan i følge han gi et helhetlig bilde av disse bitene. En nærmere forklaring av hvordan en markedspris dannes får du se senere.

Hayek sin 1945-artikkel har to gode poenger. Det første er at markedsprisen kan brukes som kommunikasjon mellom mange mennesker. Det andre er at disse menneskene sitter på så spesifikke deler av kunnskapen om det de setter en pris på at det er meget vanskelig å på egen hånd eller gjennom planlegging, finne ut hva denne prisen bør være.

4.1.2 1980-2008: Veddemål våkner til live igjen

På slutten av 80-tallet, skjedde en gjenfødelse det å se på hvordan folk investerte i veddemål for å forutse fremtiden. Denne gangen var det sterkt inspirert av Hayek sine idéer (Forsythe et al., 1992, 1999; Ortner, 1998).

Det begynte med at økonomi-undervisningen på universitetet i Iowa tok i bruk elektroniske markeder. Deltakerne kunne kjøpe og selge kontrakter

¹I følge Google Scholar

på politiske valg, hvor kontraktene ga utbetalinger avhengig av valgenes utfall. For eksempel, en kontrakt kunne betale \$1 hvis Bush ble president, og en kontrakt kunne betale \$1 hvis han ikke ble president. Hvis prisen på Bush var \$0.6, ble han president med 60 % sjanse.

PM'ene fra Iowa slo ofte andre metoder som var til for å predikere valg. De fungerte så godt, at de ofte gjorde en bedre jobb enn både undersøkelser (som Gallup), eksperter og statsvitenskapelige modeller. For eksempel Williams (2005) sammenlignet PM'et fra Iowa med modeller, ekspertpanel og spørreundersøkelser på prediksjoner av presidentvalget i 2004. Markedet spådde at Bush endte opp med 50.1 % av stemmene. Han fikk i virkeligheten 50.9 %. Gjennomsnittet av 14 spørreundersøkelser gjorde det dårligere fordi de spådde 48.9 %. Bare 3 av 8 statistiske modeller var nærmere det riktige utfallet enn PM'et.

PM'er for politiske valg har siden 80-tallet predikert godt selv om de ikke tar representative stikkprøver av deltakerne (som spørreundersøkelser gjør). På Iowa-markedet er det ofte velutdannede hvite menn som deltar. Dette forverrer ikke markedets prediksjoner i forhold til undersøkelser som er representative (dvs. ikke bare har velutdannede hvite menn). Iowa Electronic Markets har siden 1988 slått nesten tusen politiske spørreundersøkelser 74 % av gangene (Berg et al., 2008).

På 'Project Xanadu' foregikk verdens første 'hypertekst'-prosjekt (mekanismen bak World Wide Web), startet av Ted Nelson. Dette var også det tidligste forsøket på å bruke et Prediksjonsmarked i bedrift² Ansatte håpet på at produktet de laget på Xanadu skulle hjelpe demokratiet i Kina før generalsekretær Deng døde. Derfor prøvde markedet å finne ut lanseringstidspunktet. Markedet spådde 7 % sjanse for at produktet ville bli lansert før Deng døde, og han døde før produktet ble lansert. I etterkant, har flere store firmaer blitt inspirert til å prøve PM'er, blant annet Siemens (Ortner, 1998), Google (Cowgill et al., 2008), Microsoft og Hewlett-Packard. På Hewlett-Packard fant de at markedet slo bedriftens interne metoder 6 av 8 ganger (Chen and Plott, 2002).³

I 2003 ble også Pentagon nysgjerrige på hvordan prediksjonsmarkeder fungerte. De satte de opp et marked for å spå økonomisk og militær aktivitet i Midtøsten. For eksempel økonomisk vekst, mulige angrep mot

²Professor Robin Hanson skriver dette på bloggen sin: www.overcomingbias.com/2006/11/first_known_bus.html

³Chen og Plott sier ikke noe om hva PM'et ble sammenlignet med, men jeg har kontaktet forfatterne på mail, og de har svart at det var en blanding av ekspertvurderinger og statistikk.

Israel og forandringer i oljeprisen (Hanson, 2007b). Så kom de til skade for å bruke prediksjonsmarkeder på situasjoner som det ble oppfattet som umoralsk å gamble med. For eksempel, foreslo Pentagon en markedskontrakt på at Arafat skulle dø. Dette reagerte media negativt på. Også norske aviser kritiserte prosjektet. Noen eksempler er at Per Egil Hegge i Aftenposten 2/8-03 kaller idéen for *'usmakelig terrortipping'*. I Dagbladet 27/7-03, bruker Helge Øgrim og Mads Larsen uttrykket *'drapsspill'*. Per Olav Ødegård skriver i VG 3/8-03 at *'Massedød kunne gitt jackpot i dette skrekens lotto'*, og at *'Det er i beste fall naivt å tro at markedet kan forutse neste terroranslag'*. Han skriver også at :

Det amerikanske forsvarsdepartementets kontor for avanserte forskningsprosjekter skal bestå av de beste og skarpeste hjerner. De lyse hoder har fremlagt årets mest utrolige og ubrukelige forslag.

Etter den enorme mediekritikken, ble Pentagon sitt marked lagt ned. Om-talen Pentagon fikk kom nesten samtidig med lanseringen av den populære boka *'Wisdom of crowds'* (Surowiecki, 2004) som forteller om flere suksesshistorier med prediksjonsmarkeder. Derfor økte antallet forskningsartikler som handlet om PM'er. Tziralis and Tatsiopoulos (2007) tyder på at fra å bli publisert ca. 0-15 artikler om PM per år fra begynnelsen av 90-tallet, har det fra og med 2004 (samme år som Pentagons påfunn og lansering av boka) vært en økning til 20-40 artikler.

Kort sagt: Veddemål før 1940 ble ofte brukt for å spå fremtidige utfall, men interessen døde brått ut. Veddemål ble utkonkurrert av vitenskapelige spørreundersøkelser. Den samme interessen fikk en gjenfødelse på slutten av 80-tallet. For det første, begynte folk å bruke prediksjonsmarkeder internt i bedrifter. For det andre, gjorde man det på politiske valg. Den *'bevegelsen'* som oppstod var inspirert av idéene i Hayek (1945). De siste årene, har oppmerksomheten rundt PM'er eksplodert på grunn av en populær bok som omtaler PM'er, og Pentagon sitt mislykkede forsøk på å bruke PM'er.

4.2 Konsepter

Formålet med den seksjonen her er å gi innsikt i hva som foregår *'bak kulissene'* i et prediksjonsmarked. For eksempel brukes begrepene *'kjøp'* og *'salg'* ofte. Det ligger en litt komplisert forklaring av hva som skjer bak dem.

For det andre, fokuserer denne oppgaven på beslutningstaking i små grupper på 3-30 personer. Noen tror da at det ikke noe poeng i å bruke prediksjonsmarkeder, fordi få deltakere gjør det vanskelig å få folk til å selge og kjøpe mellom hverandre⁴ Denne innvendingen gjelder (som oppgaven snart viser) bare for en spesifikk form for markedsmekanisme.

4.2.1 Noen grunnleggende forklaringer

Hvordan bevege på prediksjonene?

For å få det tallet markedet predikerer til å gå oppover, er du nødt til å kjøpe andeler av utfallet tallet gjelder for. For eksempel hvis Kåre kjøper kontrakter på at buksesalget vil øke, øker markedet også prediksjonen av antall bukser som blir solgt.

For å få tallet nedover, er du nødt til å selge. Hvordan kan man selge noe uten å først eie det? Den ene måten kalles 'short-sale'. Den andre er å presse prisen opp på kontrakter på motsatte utfall enn utfallet som skal presses ned. For eksempel det motsatte utfallet av 'vi kommer til å selge mer enn 10 000 bukser' er 'vi kommer til å selge mindre enn eller like mye som 10 000 bukser'.

Short-sale er å låne noe og selge det videre med avtale om å betale tilbake til markedsprisen som gjelder på et senere tidspunkt. Du har tjent penger på short-sale hvis prisen på det du låner går ned. For eksempel Kari låner en sykkel av Knut, og selger den på nettauksjonen qxl.no for 400 kroner. Hun er utspekulert, og vet at prisen vil gå ned til 100 kroner om et år. Når året endelig har gått, kjøper hun sykkelen til 100 kroner, og gir den tilbake til Knut. Da har Kari tjent 300 kroner.

Et alternativ til short-sale på prediksjonsmarkeder, er å kjøpe kontrakter av de utfallene som er motsatt av de du vil presse prisene ned på. For eksempel hvis du vil presse ned prisen på kontrakten 'brune bukser vil lykkes', kan du kjøpe kontrakter av 'brune bukser vil ikke lykkes'.

Kort sagt, er det to handlinger en person kan gjøre for å bevege på markedsprisen: Kjøpe eller selge. Kjøp beveger prisen oppover, og salg beveger den nedover. Hva som skjer når en person selger, har en mer komplisert forklaring enn hva som skjer når noen kjøper. Grunnen til dette, er at

⁴Green et al. (2007) prøver seg for eksempel på denne kritikken når de skriver at '*Delphi requires only 5 to 20 experts who have agreed to participate and should therefore be superior to thin markets (those with few participants) where the incentive to trade, and thereby reveal information, is weak*'

man bør gi muligheten for at en person kan bevege prisen nedover (selge) uten å egentlig ha noe å selge til andre.

Hvordan hente ut en prediksjon basert på markedsprisene?

Det er to ting du må gjøre når du skal hente en prediksjon fra prediksjonsmarkedet. For det første, så må du hente ut en *markedspris*. Det er den prisen som folk vanligvis selger og kjøper for på markedet. Du kan finne markedsprisen ved å ta gjennomsnittsprisen av alle kjøp og salg over en viss tidsperiode, og så vekte den etter antallet det er kjøpt og solgt for. Dette kalles *volumvektet gjennomsnittspris* (Volume Weighted Average Price [VWAP]).

Tallet du får basert på VWAP, kan brukes for å gjøre en prediksjon. For eksempel at sjansen for å selge en bukse er 40 %, eller at firmaet kommer til å selge 10 000 bukser. For å tolke VWAP som en sannsynlighet, kan du ta VWAP som gjelder for alle kontraktene på markedet, summere disse og så dele dette på VWAP for kontrakten du vil ha sannsynligheten på.

Hvis du vil ha en variabel basert på VWAP, behøver du bare å gjøre det samme som når du får en sannsynlighet, kall dette tallet p , men denne gangen tar du intervallet variabelen ligger innenfor, og beregner:

$$Min + p(Max - Min) \tag{4.1}$$

Hvor 'Min' og 'Max' er minimum og maksimumsverdiene på intervallet.

F.eks. tenk deg at markedet har kontrakter som predikerer antall solgte bukser i løpet av måneden. Du forventer at minimum antall solgte er 5000 og at maksimum er 15 000.

Du har funnet ut at VWAP på kontrakten som sier at buksesalget vil gå bra (den betaler mer desto mer vellykket buksesalget blir) er på 12 000 og VWAP for kontrakten som sier det vil gå dårlig er på 7 000. Markedet predikerer da at antall solgte bukser blir følgende: $N = 12000 / (7000 + 12000) = 0.63$, $5000 + (15000 - 5000)0.63 = 11300$ ⁵

Kort sagt, er det to ting som må gjøres for å kunne gjøre en prediksjon basert på markedsprisene. Først bør du beregne den gjennomsnittlige prisen som folk har kjøpt og solgt kontrakter for. Så kan denne prisen konverteres om til en variabel innenfor et visst intervall.

⁵For kontrakten på at salget skulle gå dårlig (som er på 7000) så hadde du fått akkurat det samme tallet, fordi da ville du ha måtte multiplisere med (N-1)

4.2.2 Hva består et marked av?

Markedsmekanisme

En markedsmekanisme avgjør hvordan priser på kontrakter forandrer seg. Den mekanismen som vanligvis blir brukt, er en slags auksjon hvor kontrakter blir kjøpt og solgt mellom deltakerne (forklart nærmere nedenfor). Dette kalles '*Continuous Double Auction*' (CDA). I de fleste tilfeller hvor CDA ikke er passende (som når det er få brukere på markedet), kan det brukes en '*Automated Market Maker*' (AMM) istedenfor. AMM kan ses på som en slags simulering av hva som foregår på CDA med mange mennesker. Forskjellene på ulike AMM-utgaver er hvordan de beveger på markedsprisene etter kjøp og salg.

The Market

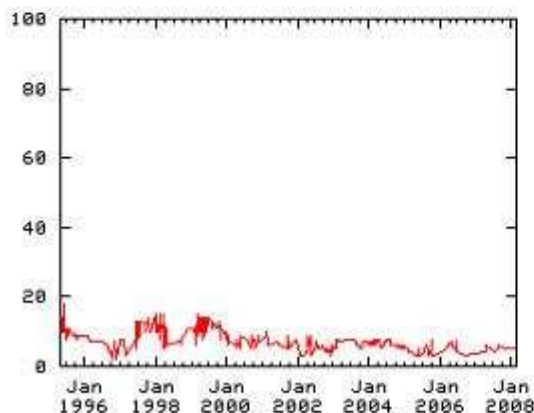
Your Buy YES Orders Players Buying YES Coupons

1 @ 1	haggis (2155)
1 @ 1	nazone (573)
8 @ 1	LotC (4057)
20 @ 1	evand (4465)
28 @ 1	DavidSJ (4176)
48 @ 1	Lavender Green (124)
1 @ 2	gustrot (4624)
1 @ 2	Varulv (3126)
7 @ 2	lucifer (395)
10 @ 2	evand (4465)
14 @ 2	DavidSJ (4176)
20 @ 2	eidolon (355)
25 @ 2	RobinZimm (9316)
30 @ 2	MOXman (4219)
50 @ 2	MaxiBide (7739)
500 @ 2	aof (617)
1 @ 3	aaronwinborn (4821)
1 @ 3	gustrot (4624)
1 @ 3	Varulv (3126)
1 @ 3	Glenn Alcoe (2562)
5 @ 3	evand (4465)
6 @ 3	jakala (3294)
9 @ 3	DavidSJ (4176)
11 @ 3	madmooneyuk (4243)
20 @ 3	wilian (4844)
1 @ 4	Varulv (3126)
4 @ 4	wilian (4844)
1 @ 5	caprandom (6829)
1 @ 5	Varulv (3126)
47 @ 5	wilian (4844)

Erase Quantity Price

<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
<input type="checkbox"/>	<input type="text"/>	<input type="text"/>

Price Plot for life of Astr



Last trade price: 5
 Current ask price: 6
 Current bid price: 5
 Pairs outstanding: 4341
 Players participating: 103
[View the ticker for this Claim.](#)

Your cash balance: 1996.00
 Your holdings in this Claim: 0
 Your UID:
 Your password:

Your Sell YES Orders Players Selling YES Coupons

Erase Quantity Price

<input type="checkbox"/>	<input type="text"/>	<input type="text"/>
<input type="checkbox"/>	<input type="text"/>	<input type="text"/>

1 @ 6	Georgi (8895)
20 @ 7	wilian (4844)
5 @ 7	gustrot (4624)
1 @ 7	wedgingt (171)
1 @ 7	Varulv (3126)

36

Figur 4.1: Slik ser en bid/ask kø ut i CDA (ideosphere.com). Bruker 'wilian' ligger øverst på buy-køen fordi tilbyr det beste kjøpet på 5 lekepenger.- Bruker 'georgi' er øverst på sell-køen med beste salg. På grafen ser du hvordan markedsprisen har forandret seg over tid basert på hvor stor pris kjøpere og selgere har blitt enige om.

Continuous Double Auction (CDA) består av to køer. En for forespørsler om salg ('ask'), og en for forespørsler om kjøp ('bid'). Køene er sortert etter de forespørslene som har best pris. 'Beste pris' vil si enten det billigste salget, eller det dyreste kjøpet. Systemet må følge en ganske enkel regel: Hvis beste salgspris er mindre enn eller lik beste kjøpspris, så gjennomføres en handel mellom de respektive forespørslene.

For å tjene så bra som mulig, bør en person prøve på å kjøpe så billig som mulig, og selge så dyrt som mulig (derav det kjente uttrykket 'buy low, sell high'). CDA fungerer på en slik måte at deltakerne selv slipper å følge denne strategien. På grunn av køsystemet, har de 'beste' salgene har alltid størst sjanse for å få gjennomført en handel med de 'beste' kjøpene.

For eksempel ønsker Ole 10 kontrakter kjøpt for 50 kroner stykket. Gunda selger 20 for 60 kroner stykket, mens Per selger 5 for 10 kroner. I dette tilfellet, får Ole og Per en match med hverandre på markedet. Ole kjøper da fem av kontraktene han ønsker fra Per for for ($5 * 10$) kroner.

Det kan oppstå flere problemer med CDA-mekanismen. For eksempel noen ganger, så er det ikke folk til stede som er villige til å motta andres forespørsler om salg eller kjøp. Hvis Per ikke hadde vært til stede i forrige eksempel, så hadde det ikke skjedd noen handel, fordi det bare hadde vært selgere til stede.

Noen har utviklet mekanismer som kan takle problemet CDA-mekanismen når det er for deltakere. Disse mekanismene kalles for 'Automated market makers' (AMM'er). AMM kan ses på som form for CDA hvor det også er en robot som deltar. Roboten sørger alltid for å legge ordre øverst på køene slik at ikke folk står igjen uten å finne egnede kjøpere eller selgere (Pennock and Sami, 2007). Deltakerne på AMM får illusjonen av å delta på markeder hvor det hele tiden er andre som er rede til å ta i mot forespørsler om kjøp og salg. Hver gang noen prøver å kjøpe, skyver AMM prisen opp, og hver gang noen selger, så skyves prisen ned. For eksempel hvis Ole ønsker å kjøpe 10 kontrakter, og det ikke er noen selgere til stede, så betaler AMM'en for kontraktene istedenfor, og gjør en beregning som skyver markedsprisene oppover i henhold til en viss formel. De mest kjente AMM'er er Pennock's Dynamic Pari-mutuel (DPM) og Hanson's Logarithmic Market Scoring Rule (LMSR) (Pennock, 2004; Hanson, 2007a).

Et eksempel på en studie hvor AMM har blitt brukt i praksis, er Bruggen et al. (2006). De finner at PM'er med bare 6 deltakere er bedre enn eller like bra som statistiske grupper på å predikere utfall som oljeprisen, aksjekurser og fotballkamper. På det høyt kritiserte prediksjonsmarked-prosjektet til Pentagon, så brukte de dessuten en AMM som er svært lik LMSR (Han-

son, 2007b).

Forskningen på AMM'er er på et veldig tidlig stadié. DPM har blitt forandret flere ganger, og man bør være forsiktig med å stole på at den fungerer helt som den skal. For noen år siden, fant for eksempel noen tenåringer ut at en matematisk feil med DPM gjorde at de kunne utnytte Yahoo sitt prediksjonsmarked for penger ved å selge og kjøpe på en spesiell måte (Mangold et al., 2005). Heldigvis, ble det bare brukt lekepenger, så de klarer ikke å ødelegge for Yahoo. LMSR er den mest utbredte AMM'en akkurat nå. Den brukes for eksempel på intrade.com, og de har rundt 40 000 brukere. Men det er ingen selvfølgelighet at LMSR derfor er overlegen DPM. En datasimulering som er gjort, tyder på at prisene du får med DPM blir omtrent like prisene du får med LMSR, men de varierer mindre. Dette kan føre til at DPM gir større prediksjonsnøyaktighet enn LMSR (Peters et al., 2007).

Jeg har gått nærmere inn på noen av de mekanismene som prediksjonsmarkeder bruker. Den største forskjellen er mellom CDA og AMM-mekanismene. Beslutningen om å velge en mekanisme, avhenger av hvor mange det er som skal delta på markedet.

Markedskontrakter

Markedet besvarer ulike spørsmål. For eksempel: Hvilket parti blir valgt? Vil renta gå opp? Hvor mange rømmevafler kommer Moss Idrettslag til å selge etter å ha malt pølsebøden sin rød? Vil Arbeiderpartiet føre til økte renter? For hvert spørsmål, kan det skje flere mulige *utfall*. For eksempel 'Arbeiderpartiet' og 'JA' er mulige utfall for de to første spørsmålene ovenfor. Utfallene kan også være tall innenfor spesifikke intervaller. Moss IL kan for eksempel selge fra 0 til 1000 rømmevafler.

På prediksjonsmarkedene, formuleres utfallene som '*kontrakter*'. De er på formen 'Jeg er en kontrakt som betaler X hvis Y skjer'. F.eks. 'Jeg er en kontrakt som betaler 100 kroner hvis vi selger over 50 rømmevafler'. Det å eie en kontrakt er noe likt det å eie en aksje i et aksjeselskap. Du kan både kjøpe og selge dem. Når du gjør dette, inngår du et veddemål med den du kjøper fra eller selger til.

Hvis utfallet er en sannsynlighet, har det en verdi mellom 0 og 1. Markeder som predikerer sannsynligheter har ofte 'gjensidig utelukkende' kontrakter på hvert utfall. Det vil si at ingen av utfallene det er kontrakter på kan skje på likt, og at sannsynligheten for hvert utfall til sammen har summen

1 (dvs. 100 %). For eksempel 'betaler 100 hvis vi selger over 50 rømmevafler' er gjensidig utelukkende med 'betaler 100 hvis vi selger like mye som eller under 50 rømmevafler'. Hvis Gundas bukser A/S skal spå antall bukser solgt, så kan det være to kontrakter. En kontrakt sier 'Vi selger flere bukser enn før', og en sier 'Vi selger færre eller like mange bukser som før'.

Kontrakter kan også gjøres mer kompliserte. En type kalles *kontingente kontrakter* ('contingent contracts'). De betaler avhengig av to variabler: Den ene er om beslutningen blir tatt, og den andre er konsekvensen av å ta beslutningen (Hanson, 1999). For eksempel kan en kontrakt si at 'Vi får økt salget hvis vi selger mørke bukser'. Jeg skal ikke gå inn på en teknisk beskrivelse av kontingente kontrakter i denne oppgaven.

Kort sagt er fremtidige utfall på på prediksjonsmarkedene formulert som 'kontrakter' som betaler en viss pris avhengig av hvordan utfallene utvikler seg.

4.2.3 Oppsummering

I denne seksjonen har jeg diskutert nærmere hva som ligger bak prediksjonsmarkedet. Jeg har forklart hvordan noen forandre på prediksjonene, hva som avgjør hvordan prisen forandrer seg (CDA og Automated Market Maker), og hvordan prisen til slutt kan tolkes som en prediksjon.

4.3 Hypotesen om markedseffisiens

For dem som drømmer om å tjene raske penger på aksjemarkedet, er effisiensteorien deprimerende lesning - Grøtte (2006)

Hypotesen om '*markedseffisiens*' (Market efficiency) sier at en markedspris reflekterer all informasjon som er tilgjengelig om varen den er ment for. For eksempel hvis en tannbørste er priset med 60 kroner, så er det god grunn for at den ikke for eksempel koster 50 eller 70 kroner.

Poenget med prediksjonsmarkeder, er å gå inn for å bevisst bruke markedseffisiens for å spå fremtiden. Hvis du setter markedspriser på ulike hendelser som kan forekomme, så bør disse prisene i følge effisienshypotesen, reflektere all den informasjonen som er tilgjengelig om hendelsene. For eksempel hvis kontrakten 'betaler 100 hvis regnbukser selger godt' har

en markedspris på 60 kroner, kan det tolkes som at all tilgjengelig informasjon om at regnbukser selger godt sier at det har $60/100 = 60\%$ sjanse for å skje. Hvis effisienshypotesen stemmer, kan for eksempel ikke Astrid plutselig komme og foreslå at sannsynligheten bør være 70 % istedenfor uten å tape penger på dette.

Effisienshypotesen har støtte fra studier av finansmarkeder. Hvis prisen på en aksje ligger på 60 kroner, er det ofte god grunn for at den ikke ligger på, si 50 eller 70 kroner istedenfor. De siste tiårene, har det vært vanlig å tjene penger på aksjeinvesteringer fordi aksjekursen generelt har gått oppover. Det er veldig få som klarer å tjene mer enn denne kursstigningen. Dette betyr at få har informasjon som sier at aksjekursen vil bevege seg annerledes enn hva den pleier å gjøre.

Likevel prøver både meglere og aksjefond å skaffe seg bedre avkastning enn normalt. De fleste av dem lykkes ikke. Gjennomsnittsprisen av alle aksjer på markedet (indeksen) slo i perioden 1961-1991 90 % av alle aksjefond på å forutse fremtidige aksjekurser (Malkiel, 2003). Williams (2005) konkluderer i en gjennomgang av et dusin studier fra finansmarkeder at det er svært få tilfeller hvor man med ekspertvurderinger og investeringsstrategier kan gjette at prisen som allerede er satt bør bli annerledes. For eksempel 20 analytikere som skulle spå kursen på utenlandsk valuta, hadde rett 50 % av gangene. I en fire ukers periode, var til og med spådommene deres dårligere enn 50 %. Det betyr altså at ekspertene hadde blitt slått hvis du hadde latt mynt og krone avgjøre dine investeringer i valutakurser.

Noen ganger, finner du argumenter som strider mot effisienshypotesen. De trekker frem investeringsstrategier som har vist seg å være mer lønnsomme enn normalt. Men Williams konkluderer med at disse strategiene ofte ikke gir så mye mer fortjeneste enn hva som er vanlig (Williams, 2005). Slike strategier blir også fort 'borte' etter offentliggjøring, f.eks. etter publiseringer i akademisk litteratur. Det er tilstrekkelig at noen få på markedene kjenner til strategiene og tjener penger på dem, for at ingen andre skal få gjort det samme. For eksempel Schwert (2002) går gjennom en rekke 'effekter' som investorer kan utnytte for å tjene penger på aksjehandel. Schwert undersøker om det har gått an å tjene penger på dem etter at de har blitt offentlig kjente. De fleste effektene 'forsvinner' fra markedene nesten med en gang. For eksempel har det eksistert en Weekend-effekt hvor aksjekursen pleide å dale ved ukeslutt. Men den forsvant fort fra markedene etter å ha fått omtale i akademisk litteratur.

Jeg har nå kort diskutert hypotesen om markedseffisiens - at en markedspris reflekterer all den informasjonen som er tilgjengelig om hva den kommer til å bli i fremtiden. Studier fra finansmarkeder tyder på at få klarer å finne ny informasjon om markedsprisen. Men mange prediksjonsmarkeder er ulike finansmarkeder. For eksempel kan de inneholde færre personer, og bruke forskjellige markedsmekanismer⁶. Så det er uklart om observasjoner fra finans kan overføres direkte til hva som skjer på prediksjonsmarkeder.

4.4 Oppsummering av kapitlet

Interessen for prediksjonsmarkedene har eksplodert etter at Pentagon prøvde å bruke dem, og etter lanseringen av den populære boka 'Wisdom of crowds'.

Hypotesen om markedseffisiens, sier at markedspriser inneholder all kunnskap som er tilgjengelig om det de brukes på. Hypotesen har støtte fra for eksempel finansmarkeder hvor de gjennomsnittlige markedsprisene ofte slår eksperter og aksjemeglere på å forutse fremtiden. Tanken bak Prediksjonsmarkeder er å bruke markedseffisiens bevisst for å spå fremtidige utfall.

⁶Eks. en 'Automated Market Maker' som jeg skal redegjøre for senere

Del III

Diskusjon

Kapittel 5

Hva mangler de andre metodene?

Tidligere diskuterte jeg metoder som gruppediskusjon, Delphi-metoden og statistiske grupper, Det er noen egenskaper prediksjonsmarkeder har som disse metodene mangler: Å stimulere kritikk, gi viktige personer økt innflytelse og tilby insentiver.

5.1 Stimulere kritikk

- At the risk of oversimplification, and without intending to reduce all debiasing strategies to a common theme, there does seem to be an element of continuity: Many of the most effective debiasing techniques involve the consideration of alternative perspectives - (Plous, 1993) (s. 256)

Å tenke kritisk er å tenke gjennom alternative forklaringer på hva som kommer til å skje, selv om de alternative forklaringene kanskje er upopulære.

Noen ganger kan vi låse oss inn i ensidig tankegang ved å overfokusere på ting vi ønsker skal skje, og ignorere de tingene vi ikke ønsker skal skje. Konsekvensen er en risiko for å ta dårligere beslutninger enn hva vi ellers ville ha gjort. For eksempel: Per ønsker å tenke mer på rapportene som viser at salget går bra. Det er mer behagelig for Per å tenke sånn, fordi lønnen hans avhenger av hvordan salget går. Hvis han bare tar beslutninger basert på positive rapporter, og ignorerer negative, så kan konsekvensene være at han gjør for optimistiske investeringer, og dermed går med tap.

En *'Pollyanna'* ser noe som er positivt, i selv de situasjonene som er ille. Begrepet er inspirert av en bok som er nesten 100 år gammel som et lite barn som heter Pollyanna. Flere forskere har i etterkant av boka, interessert seg for Polyannas oppførsel, og det har ført til flere studier av fenomenet. Etterhvert, så har man kommet frem til et prinsipp som kalles Pollyanna-prinsippet: Mennesket har en naturlig trang til å søke positiv informasjon, og ignorere negativ (Matlin, 2004).

Man bør påminne folk til å tenke kritisk for å unngå at Polyanna-prinsippet skal ødelegge beslutningstaking. Connolly et al. (1990) fikk forsøkspersonene sine til å prøve å komme på løsninger vedrørende et parkeringsproblem. I hver gruppe, ble det plassert falske forsøkspersoner. I en gruppe, var den falske personen satt til å hele tiden si seg enig med hva de andre sa. Han brukte uttrykk som 'god idé' og 'helt enig', da han omtalte de andres idéer. I den andre gruppa, var det derimot en som hele tiden var kritisk innstilt. Han brukte uttrykk som 'dårlig idé', og 'det var et forferdelig forslag' da han omtalte andres idéer. Så ble et dommerpanel gitt oppgaven å evaluere løsningene fra hver gruppe. Det viste seg at gruppene som hadde kritiske personer, kom på flere og mer originale idéer (men det var ingen forskjell på idéenes kvalitet).

Påminnelser om å tenke kritisk fungerte også i Valacich and Schwenk (1995). De ga grupper oppgaven å fordele salgsterritorier til handelsmenn. Hver gruppe fikk en rapport, men noen grupper fikk et vedlegg som var kritisk til hovedrapporten. De som fikk se det kritiske vedlegget, kom på både flere idéer og hadde større sjanse for å komme med idéer som solgte godt.

Når man kritiserer hverandre, så spiller egentlig ikke type kommunikasjonsmedium så stor rolle. Kritikk som foregikk elektronisk (anonymt) var enten like bra eller bedre enn kritikk som folk ble gitt ansikt til ansikt i Valacich and Schwenk (1995) og Connolly et al. (1990).

I denne seksjonen, har jeg vist at det ikke faller naturlig for folk å tenke kritisk. Når folk først blir oppfordret til å tenke kritisk, viser det seg at de kommer på flere og bedre idéer, enn hvis de ikke oppfordres til å tenke kritisk.

5.1.1 Kritisk tankegang i prediksjonsmarkeder

Hvis du er flink til å være kritisk til andre, så blir du godt betalt i prediksjonsmarkeder (du må selvfølgelig også ha en god grunn for å være det). De utfallene som ingen andre er enige i, har lave priser på et marked. For

eksempel kontrakten 'betaler 100 hvis vitenskapsmenn påviser at det eksisterer engler' har sannsynligvis lav pris på markedet (ca 1 krone). Det kan tenkes at du har en grunn for å tro at vitenskapen i nær fremtid kommer til å bevise at engler eksisterer. Da gjør du et røverkjøp hvis du kjøper 10 kontrakter for 10 kroner stykket, som senere viser seg å betale ut 1000 kroner.

For det andre, så gir markedet deg stor fleksibilitet til å 'bake inn' funksjonalitet som belønner folk for å være kritiske. David Pennock, har gjort nettopp dette på prediksjonsmarkedet 'Yahoo Buzz Game'. Funksjonen kalles *Dynamic Pari-mutuel Market Maker* (DPM).

Utbetalingen per kontrakt med Share-ratio Dynamic Pari-mutuel marked som beskrevet av Pennock and Sami (2007) er:

$$\frac{\sqrt{(\sum_i^n q_i^2)}}{q_i} \quad (5.1)$$

Hvor n er antall forskjellige kontrakter på markedet og q_i er hvor mye en deltaker eier av kontrakt i . Et eksempel på dette, er hvis det er 3 deltakere på markedet, Per, Ove og Gunda. Per eier 1 stk av kontrakten 'Tredje verdenskrig skjer før 2010', mens Ove og Gunda eier 5 hver av kontrakten 'Tredje verdenskrig skjer ikke før 2010', så er $n = 2$. I dette tilfellet har Per valgt å vedde på utfallet som de andre ikke tror på. Hvis Per har rett, vil han få utbetalt:

$$\frac{\sqrt{(1^2 + 10^2)}}{1} \simeq 101 \quad (5.2)$$

Hvis Ove og Gunda har rett, vil hver av dem bare få utbetalt:

$$\left(\frac{\sqrt{(1^2 + 10^2)}}{10}\right) * 5 \simeq 50.5 \quad (5.3)$$

Kort sagt, så lønner det seg mest vedde på upopulære utfall.

To hovedpoenger kom frem i denne seksjonen. For det første, så lønner det seg å være kritisk på et prediksjonsmarked. For det andre, så kan du til og med sette opp prediksjonsmarkedet slik at det belønner kritiske personer mer enn de ukritiske.

5.1.2 Oppfordrer andre metoder til kritisk tankegang?

Kritiske prediksjoner har ikke så lett for å slippe frem under gruppediskussjoner. Desto mer noens prediksjoner avviker fra det som er sett på som vanlig, desto mindre blir de lagt vekt på av andre gruppemedlemmer (Yaniv, 2004).

Med Delphi-metoden, så er man faktisk oppfordret til å fortsette å be folk om prediksjoner inntil de blir mer enige med hverandre (Jones and Hunter, 1995). Dette kan være veldig uheldig, fordi folk kan ha gode grunner for å i utgangspunktet være uenige (kritiske til hverandre).

I statistiske grupper, så er det ikke mulig å være kritisk, fordi folk får uansett ikke vite hva andre synes.

Kort oppsummert, så er det ingen av de andre metodene som forholder seg til det å tenke på en kritisk måte.

5.1.3 Oppsummering

I motsetning til alternative metoder, belønner altså prediksjonsmarkeder kritisk tankegang.

5.2 Incentiver

Et incentiv er noe som skal øke motivet for å gjøre en handling. Incentivet kan være alt fra pengebetaling til et klapp på skulderen. Hvis Knut får incentiv for å finne ut hvilke bukser som selger best, økes motivet hans for å gjøre dette.

Incentiver fungerer for å komme på idéer til beslutninger. De øker kvaliteten på idéer, og øker antallet idéer. Folk er også ofte mer fornøyd med idéene sine etter å ha fått belønning for dem (DeRosa et al., 2007)

Det fungerer å gi folk incentiver for å unngå at de hermer etter hverandre. Vi har sett at dette kan være et problem spesielt på grunn av gruppediskussjoner. Tidligere så vi et eksperiment hvor folk skal gjette hvilke baller det er flest av i en urne. Det kan for eksempel være 60 % røde og 40 % hvite. Hver person trekker en ball som vedkommende ikke får vise til de andre. Så annonserer denne personen til gruppa hva han eller hun tror er det riktige valget. Hung and Plott (2001) brukte dette eksperimentet og betalte folk enten etter hvor godt flertallet i gruppa gjettet, hvor godt de gjettet

selv, eller hvor enige de var med andre. Å betale folk for å være enige med flertallet førte også til at flertallet gjettet dårligere. I 96.7 % av tilfellene begynte folk å herme etter hverandre, og flertallet hadde 74 % av gjetningene korrekt. Å betale etter hvor godt man gjettet selv, eller etter hvor godt gruppa gjettet, viste seg å være gode måter å unngå at folk hermet etter hverandre på. Da var det betydelig færre som hermet, og flertallet hadde mellom 80 og 90 % svar korrekt. Et liknende resultat finnes i Anderson (2001): Folk hermer mindre etter hverandre hvis de blir betalt for å gjette riktig, og de har også oftere riktige svar.

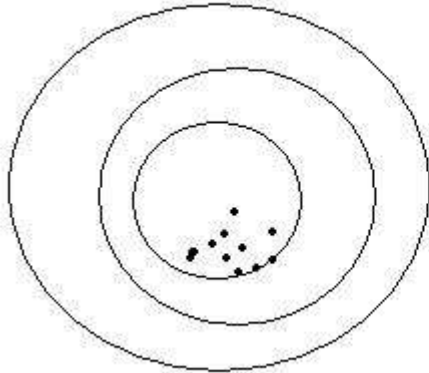
Insentiver forbedrer ikke hvordan folk predikerer (ved første øyekast). Camerer and Hogarth (1999) rapporterer bare 3 av 5 studier hvor insentiver forbedrer sannsynlighetsvurderinger. Og i 2 av 6 studier forbedrer insentiver prediksjoner, men forverrer dem i 4 andre! Bonner et al. (2000) finner at insentiver for god bedømmelse har mindre effekt enn insentiver for å gjøre andre ting, som f.eks. fysisk arbeid. Insentivene fungerer i så lite som under halvparten av studiene de gjennomgår.

Det er likevel grunner for å tro at insentiver vil kunne forbedre prediksjoner. Den ene grunnen er at folks prediksjoner blir mindre varierte av insentiver. Stor variasjon er når noen i gjennomsnitt gjør veldig forskjellige prediksjoner. Liten variasjon, er å gjøre veldig like prediksjoner. Camerer and Hogarth (1999) konkluderer med at insentiver gjør folks oppførsel mindre variert. Den samme konklusjonen kommer Smith and Walker (1993) til når de gjennomgår eksperimenter fra psykologi og økonomi.

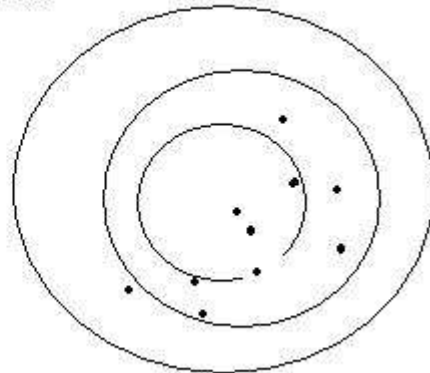
Mindre variasjon kan bidra til å øke *presisjonen* på hvordan folk predikerer. Her kommer en nærmere forklaring. Presisjon kan forklares ved å tenke seg to skyttere som konkurrerer om å skyte på en blink. Begge treffer i gjennomsnitt på samme sted. Den ene av dem treffer mye mer variert enn den andre, noe som fører til at han får dårligere poengsum i det lange løp.

På samme måte, vil en person som predikerer mindre variert enn en annen, men som i gjennomsnitt predikerer likt, i gjennomsnitt treffe flere ganger. Siden mange studier finner ut at folk predikerer mindre variert og likevel at de i gjennomsnitt predikerer likt, kan altså dette bidra til å øke prediksjonsnøyaktigheten i lengden.

Presis



Upresis



For det andre, hvis det er snakk om en metode som setter sammen ('aggregerer') individuelle prediksjoner, så kan effekten av individuell variasjon ramme hele gruppa (ikke bare de få individene som har mer variert oppførsel enn andre). For eksempel Per, Pål og Espen predikerer at de skal selge henholdsvis 1, 2 og 13 epler. Det riktige antallet viser seg å være 3. På grunn av Espens ekstreme prediksjon, ødelegges også den gjennomsnittlige prediksjonen. Individuelt, er det bare Espen som rammes av den dårlige prediksjonen, men som gruppe, er det alle sammen som rammes. Camerer and Hogarth (1999)[side 32] foreslår at markeder kan være sensitive for variasjon på denne måten, fordi markedsprisene er avhengig av hvordan de fleste enkeltindividene oppfører seg.

Altså er insentiver et godt virkemiddel for å forbedre idégenerering og prediksjoner. De kan også føre til at folk hermer mindre etter hverandre.

5.2.1 Insentiver og prediksjonsmarkeder

I prediksjonsmarkeder brukes penger som insentiver for å få folk til å gjøre gode prediksjoner.

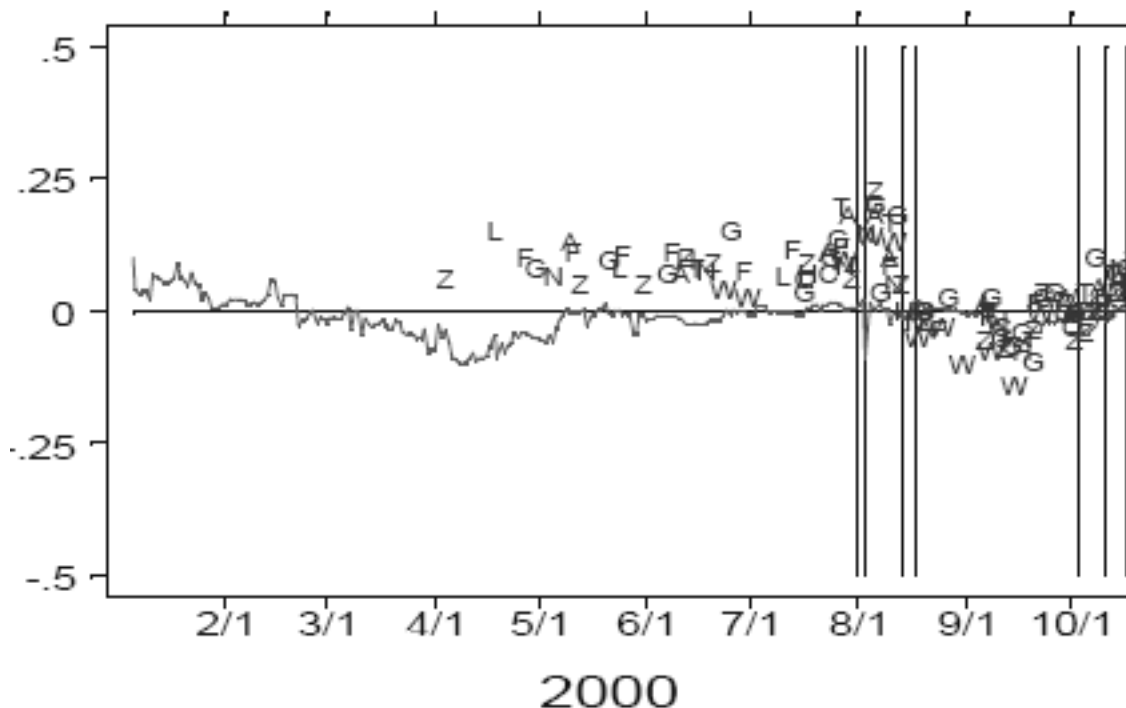
Økt presisjon

Jeg argumenterte i forrige seksjon at insentiver kan gjøre prediksjoner mer presise på grunn av redusert variasjon. I denne seksjonen viser jeg at studier tyder på at dette er tilfellet på prediksjonsmarkeder. Flere studier tyder på at prediksjonsmarkedene gjør like prediksjoner som andre metoder,

men at det slår dem fordi prediksjonene som gjøres er mindre varierte (og dermed mer presise).

Wolfers and Zitzewitz (2004) og Gurkaynak and Wolfers (2006) rapporterer fra prediksjonsmarkeder for økonomiske utfall som lønningsnivå og arbeidsløshet. Prediksjonsmarkedene slår en statistisk gruppe bestående av eksperter. Forskjellen på prediksjonsmarkedet og den statistiske gruppa, er ikke stor. Grunnen til at prediksjonsmarkedet gjør det bedre, er at prediksjonene som gjøres er mindre varierte.

Berg et al. (2003) rapporterer liknende resultater når de sammenligner de kjente markedene fra Iowa med spørreundersøkelser. Spørreundersøkelsene sier nesten det samme som prediksjonene til markedene, men markedene varierer mye mindre enn dem. Dette gjelder nesten for alle spørreundersøkelser. Det er trolig derfor prediksjonsmarkeder fra Iowa har slått nesten 1000 spørreundersøkelser i 74 % av tilfellene siden 1988. (Berg et al., 2008). Se også eksempelet på figuren nedenfor.



Figur 5.1: Bokstavene står for ulike spørreundersøkelser (f.eks. G er Gallup). Den lange ujevne streken er prediksjonene fra Iowa sitt marked. En prediksjon som har 0 på Y-aksen er helt perfekt. Som du kan se, varierer prediksjonene i PM'et mye mindre enn prediksjoner som er basert på spørreundersøkelsene. Det betyr at prediksjonsmarkedet blir mer presist i det lange løp.

Gruca et al. (2003) sammenligner prediksjoner av vellykkede kinofilmer. De finner heller ingen forskjell på en statistisk gruppe med eksperter og prediksjonsmarkeder, men prediksjonsmarkedene har mindre variasjon.

Problemet med studiene ovenfor, er at de har sett på prediksjonsmarkeder som har mange deltakere. Min problemstilling har fokus på markeder med få deltakere. Dessuten, vet vi ikke om vi kan få det samme resultatet ved å velge ut vilkårlige forsøkspersoner. I studiene jeg nettopp har sett på, så valgte folk å delta på markedene selv.

Heldigvis, så eksisterer det en studie som både velger ut forsøkspersoner, og som prøver markeder med få deltakere. Sprenger et al. (2007) får grupper på 4 til å prøve på å forutse populære mobiltelefoner. Prediksjons-

markeder blir da sammenlignet med diskusjonsgrupper. Det viser seg at prediksjonsmarkedene, som i de fleste andre studier, har mindre variasjon. Dette fører til at de får bedre presisjon, og predikerer de bedre enn diskusjonsgruppene.

Kort sagt, så tyder flere studier på at prediksjonsmarkedet ikke varierer så mye som andre metoder. Dette fører til økt presisjon slik at markedene blir mer nøyaktige på å forutse fremtiden.

Gir andre metoder noe form for insentiv?

Med verken Delphi-metoden eller en statistiske grupper, blir deltakerne stilt til ansvar for hva de gjør. De har for eksempel ikke noe å tape hvis de oppgir gale svar. Det nærmeste du kommer ansvar med Delphi-metoden, er at deltakerne ofte blir bedt om å begrunne prediksjonene sine. Prediksjonene viser seg faktisk å bli mer nøyaktige av dette (Rowe et al., 2005).

Hvordan folk vil belønne hverandre i diskusjonsgrupper er litt uklart. Jeg har hatt vansker for å finne studier som svarer på dette, men det er grunn til å tro at folk frykter konsekvensene av å være uenige med andre. I eksperimenter hvor folk har blitt lurt til å angi feil svar (f.eks. lengden på pinner) fordi flertallet med falske forsøkspersoner også har angitt feil svar, har det vist seg at mange ikke blir påvirket av flertallet hvis de får lov til å være anonyme (Hogg and Vaughan, 2005). I en gruppediskusjon, kan det altså være insentiver for å *ikke* si hva man mener.

Ikke bare på grunn av insentiver

Det er ikke bare på grunn av insentivene at prediksjonsmarkeder kan gjøre bedre enn andre metoder. Servan-Schreiber et al. (2004) sammenlignet prediksjonsmarkeder med en webside hvor folk ble belønnet for å gjøre riktige prediksjoner. Servan-Schreiber et al. meldte seg på denne websiden, og brukte prediksjoner som var basert på prediksjonsmarkeder. Av cirka 2000 deltakere, kom deltakerne som brukte prediksjonsmarkedene på 6. og 8. plass. Med andre ord, så må det være noe mer enn insentiver som gjør prediksjonsmarkedene nøyaktige.

Oppsummering

På prediksjonsmarkeder, så får du betalt hvis du er flink til å predikere. Andre metoder har ikke det tilsvarende insentivet. Dette kan føre til at

prediksjonsmarkedet får en rekke fordeler som ikke andre metoder har. Jeg har vært inne på noen av fordelene i et tidligere kapittel, for eksempel at folk hermer mindre etter hverandre og at de blir mer kreative.

Dessuten har jeg skrevet at siden insentiver fører til at folk oppfører seg mindre variert, så fører det også til at prediksjonene de gjør sammen kan bli mer presise. Flere studier jeg diskuterte her støtter den teorien.

5.3 Slippe til mindretallet

Sagen er den, ser I, at den sterkeste mand i verden, det er han, som står mest alene. - Henrik Ibsen ('En folkefiende')

Hva om det eksisterer en liten gruppe med noen få som ikke deler meninger med de andre? F.eks. enten ved å være smartere eller ved å kjenne til viktig informasjon. Disse personene vil ofte ikke få slippe til hvis det brukes statistiske grupper, Delphi-metode (kapittel 3) eller diskusjonsgrupper (kapittel 2). Disse personene er det dessuten vanskelig å identifisere i praksis (kapittel 3).

Studiene gir grunn til å tro at slike personer ofte eksisterer. I de fleste studier, kommer ofte noen få gruppemedlemmer ut bedre enn hva resten av gruppa gjør. Denne eller disse personene har også mer nøyaktige prediksjoner enn nøyaktigheten på gjennomsnittet av alles prediksjoner. Hill (1982) finner at det beste medlemmet slår alle andre både når hun ser på studier hvor man prøver å komme på nye idéer og når de skal forutse fremtiden. De fleste studiene rapportert i Hastie (1986) og Gigone and Hastie (1997) tyder på det samme. Det beste medlemmet slår også ofte prediksjonene gruppa kommer frem til etter diskusjon.

Det kan også oppstå situasjoner hvor mindretallet i gruppa har opplysninger som ikke andre har. Da har de mindre sjanse for å bli vektlagt i de endelige beslutningene. Dette kan skje selv når det er snakk om viktige opplysninger (se kapittel 2). For eksempel hvis mange kjenner til en trendanalyse som viser at olabukser selger godt, så vil den vektlegges mer i beslutningene enn hvis bare noen få kjenner til samme analyse.

Kort sagt: Enten så kan det være noen som kjenner til viktige opplysninger som ikke andre har, eller så kan det være noen som er flinkere enn andre til å predikere. Disse personene bør få mer innflytelse på gruppebeslutningene, men det er i praksis vanskelig å identifisere dem.

5.3.1 Innflytelse på prediksjonsmarkeder

På prediksjonsmarkeder, får noen mer innflytelse enn andre. Hvis man føler at man har mer viktig informasjon enn hva andre har, så kan man gjøre en større investering enn hva som er vanlig. Hva kjennetegner deltakerne som gir seg selv større innflytelse på markeder?

Marginal Trader hypotesen

Høy inntekt betyr at man er flink til å spå fremtiden. For eksempel Kåre kjøper 10 kontrakter av 'betaler 100 hvis olabukser kommer på moten'. Hvis dette skjer i virkeligheten, får han altså 1000 kroner. Sprenger et al. (2007) finner at i grupper på 4 som predikerer populære mobiltelefoner, så har inntekten til deltakerne korrelasjon med hvor gode de er på å spå fremtiden (mellom lekepengene en person tjener på markedet og hvor mye vedkommende i gjennomsnitt bommer på å spå fremtiden¹ (Sprenger et al., 2007)).

De med høy inntekt, påvirker ofte markedprisene mer enn andre. Altså bidrar de mer til å forandre prediksjonene. Dette er en liten gruppe med smarte deltakere kalt '*Marginal traders*'² For eksempel Oliven and Rietz (2004) og Forsythe et al. (1999) rapporterer fra studier på politiske valg at en liten gruppe som utgjør omtrent 15 % av deltakerne ofte tjener mer enn normalt. Denne gruppa er ofte mer aktiv enn normalt, og bidrar dermed mer til å forandre prediksjoner (Oliven and Rietz, 2004).

Marginal Trader fenomenet har ikke bare dukket opp på markeder for politiske valgutfall. Andre studier rapporterer også at markedene har noen få med bedre inntekt som påvirker markedsprisene mer enn andre. For eksempel Christiansen (2007) undersøkte prediksjonsmarkeder på roing. Det var 12 (6 %) veldig aktive personer, og fem av dem lå på topp 10 i inntekt. Halvparten av deltakerne på markedet veldig passive, og gjorde få investeringer selv om det skjedde ting som gjorde at de burde reagere. De lot være å blant annet selge kontrakter da det kom klar informasjon om at de burde (fordi f.eks. roere skadet seg). Likevel ga prisene nøyaktige prediksjoner, trolig på grunn av de få som var mer aktive enn andre. Kirchler and Maciejovsky (2002) rapporterer at personlig inntekt korrelerte positivt med antall kontrakter folk kjøpte og solgte. F.eks. hvis en person kjøpte 100 kontrakter og det var vanlig å kjøpe 10 kontrakter, så ville den personen

¹Mean Absolute Percentage Error

²Også kalt 'the smart few' av noen studier (f.eks. (Camerer, 1992))

også mest sannsynlig være en deltaker som også påvirket markedsprisene mer enn alle andre.

Men er det ikke en selvfølge at de som tjener best også påvirker markedene mer enn andre? Tvert i mot. Noen forskere har foreslått at det motsatte er tilfelle: At de med størst tap påvirker markedsprisene mest. For eksempel personer som er mer sikre i sin sak enn de egentlig burde være ('overkonfidente' personer), kan også gjøre flere investeringer enn andre. Thaler (anerkjent innenfor økonomi og beslutningspsykologi) har hevdet at overkonfidens påvirker markedsprisene (Bondt and Thaler, 1994). På en annen side, så tyder de fleste studier på at overkonfidens ikke har noen sammenheng med påvirkning på markedspriser (Kirchler and Maciejovsky, 2002; Glaser and Weber, 2007).

Camerer et al. (1989) demonstrerer at Marginal trader fenomenet reduserer feilvurderinger på markedsprediksjonene. Fenomenet som kalles '*kunnskapsforbannelsen*' (curse of knowledge), går ut på at når vi skal predikere hva andre tror, så tror vi at det de tror er mer basert på hva vi selv vet enn hva det egentlig er. En gruppe på femti personer predikerte profitten på ulike firmaer. Et marked ble brukt for å predikere hva denne gruppa predikerte. Markedsdeltakerne fikk vite hva det var firmaene egentlig fikk i profitt. Den gjennomsnittlige deltaker gjorde prediksjoner som lå nærmere kunnskapen om firmaene enn gruppa sine egentlige prediksjoner (altså forekom en kunnskapsforbannelse). Markedets prediksjoner av gruppa viste 50 % mindre kunnskapsforbannelse enn den gjennomsnittlige markedsdeltaker sine prediksjoner. Grunnen til dette er at det var noen som påvirket markedet mer enn andre, som også ble rammet mindre av forbannelsen.

Andre studier har liknende funn. Feilvurderinger reduseres på grunn av muligheten til å få større innflytelse (ved å handle i et antall kontrakter). For eksempel i Cowgill et al. (2008), så viste det seg at gjennomsnittlige kjøp eller salg både var overoptimistiske og feilvurderte sannsynligheter. Likevel, så førte antallet kjøp og salg som var gjort til at graden på feilvurderinger ble redusert. De kjøp/salg som hadde større antall, var også kjøp/salg som hadde færre feilvurderinger, og dermed ble det mindre grad av feilvurderinger i markedsprisene.

Samtidig som flere studier demonstrerer MT-fenomenet (noen få deltakere som er mer aktive, har bedre inntekt enn andre), er det vanskelig å finne studier hvor noen få deltakere som er mer aktive, har dårligere inntekt enn andre.

Kort sagt: Det kan eksistere en liten gruppe som har mer å si enn andre.

Enten fordi de har viktige opplysninger, eller fordi de rett og slett er smartere enn andre. Disse slipper ikke nødvendigvis til på diskusjonsgrupper, med Delphi-metoden eller i en statistisk gruppe. De får derimot slippe til på prediksjonsmarkeder. Dette støttes av en rekke studier, når du ser på alt fra kapping til prediksjon av politiske valg. Marginal Traders kan føre til at graden av feilvurderinger (f.eks. overoptimistiske prediksjoner) blir redusert når man skal gjøre prediksjoner basert på markeder.

5.3.2 Innflytelse på andre metoder

I kapittelet om Delphi-metoden og statistiske grupper, var jeg inne på at det viste seg å være vanskelig i praksis å gi folk mer innflytelse fordi det blant annet var vanskelig å identifisere hvem det gjaldt (f.eks. Armstrong (2001)).

Under gruppediskusjoner, så viste det seg at flertallet ofte bestemte beslutningene. Altså vil det være vanskelig å slippe til. På en annen side, i de situasjonene vi vet at folk klarer å slippe til på gruppediskusjoner, så er dette ofte folk med høyere status (f.eks. akademisk eller sosioøkonomisk) eller folk som er pratsomme. Det er sjelden noen sammenheng mellom innflytelsen som disse personene får, og hvor flinke de egentlig er på å styre beslutningsprosessen.

Altså er det ingen grunn til å tro at de andre metodene gir folk innflytelse. På gruppediskusjoner vet vi at de som i blant får slippe til ikke nødvendigvis er de rette personene som bør få slippe til.

5.4 Oppsummering

Den statistiske gruppa og Delphi-metoden gir alle deltakere like stor påvirkning på prediksjonene. For eksempel hvis tre personer sier tallene 2,4 og 6, så regnes gjennomsnittet ut av dette, uten å gjøre noe videre ut av tallene. På normale møter, bestemmer enten flertallet, eller så har feil folk mer innflytelse enn andre (noen med høy status eller som er flinke til å prate for seg). Et prediksjonsmarked gir derimot alle gruppas medlemmer den innflytelsen de selv ønsker. For eksempel Knut kan velge om han skal kjøpe 10, 20 eller 50 kontrakter som sier 'betaler 100 hvis olabukser selger godt'. Desto flere kjøp han gjør, desto høyere opp går prisen på kontrakten (og dermed går prediksjonen også høyere opp). Jeg kommer frem til tre hovedpoenger:

For det første, har prediksjonsmarkeder et fenomen som heter '*the smart few*' eller '*Marginal Traders*'. Fenomenet oppstår når de få som gjør færre feil enn andre, også bruker mulighetene de har til å påvirke markedsprisen mer enn andre. Konsekvensen av dette er at selv om flertallet gjør feilvurderinger eller viser ignoranse, så kan markedsprisene likevel brukes for å gjøre nøyaktige prediksjoner. Folk kan bruke sin kunnskap om psykologi for å 'reparere' feilene andre gjør. Markedet sørger for at de også får betalt for å gjøre det. For eksempel akkurat nå tror ideosphere.com at all kreft i USA vil bli helbredet før 2010 med 14 % sannsynlighet. Min kunnskap om prospekt-teorien i psykologi sier meg at dette er en vurdering som er altfor høy. Jeg kan altså tjene penger på å gå inn på markedet og kjøpe kontrakter som presser prediksjonen på 14 % ned til omtrent 0 %.

Prediksjonsmarkeder har også to andre gode egenskaper som ikke andre alternativer har. Det ene er at deltakere blir tilbudt insentiver (som premier og penger) for å bidra til å gjøre gruppas prediksjoner mer nøyaktige. Verken et naturlig møte, Delphi-metoden eller en statistisk gruppe tilbyr insentiver. Tvert i mot, så gir møtet i blant insentiver for at folk skal la være å si det de vet, f.eks. fordi man kan ha frykt for å bli utstøtt. Det andre er at man blir oppfordret til å tenke kritisk til det andre sier og gjør. Verken en diskusjonsgruppe, Delphi-metoden eller en statistisk gruppe har 'innebygde' oppfordringer til å tenke kritisk. Tvert i mot, oppfordres folk til å bli mer enige med hverandre med Delphi-metoden.

Kapittel 6

Svakheter ved prediksjonsmarkeder

Dette kapitlet ser på noen mulige svakheter ved prediksjonsmarkeder og studiene som undersøker dem.

6.1 Baksiden ved innflytelse: Manipulasjon

it could be subject to manipulation, particularly if the market has few participants - providing a false sense of security or an equally dangerous false sense of alarm.

- Joseph Stiglitz, Nobelprisvinner i økonomi. Om prediksjonsmarked-prosjektet på Pentagon (leserinnelegg til Los Angeles Times, 31. Juli 2003)

Prediction markets are vulnerable to speculative attacks mounted in order to manipulate the results. (Green et al., 2007)

Siden det er så lett å påvirke markedspriser, så kan det tenkes at noen bevisst går inn for å manipulere prediksjonene. For eksempel Per er egentlig ansatt i Pers bukser A/S og han vil sabotere Gundas bukser A/S sine beslutninger. Han later som at han er en fredelig medarbeider, og registrerer seg så som deltaker på markedet. Deretter inngår Per veddemål på at det

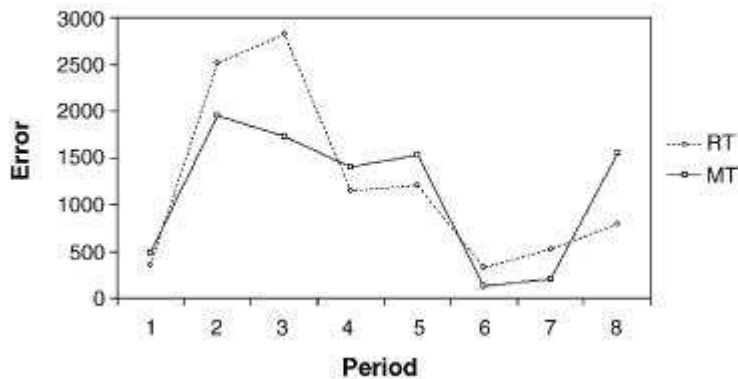
vil gå bra med bukser han vet det kommer til å gå dårlig med. Bedriftens ledelse blir derfor lurt til å gjøre investeringer i bukser som fører til tap.

Mange har antydnet at prediksjonsmarkeder er mer utsatt for manipulasjon enn hva andre metoder er. Nobelprisvinner Joseph Stiglitz antyder i sitatet ovenfor at manipulasjon spesielt kan skje på markeder som har få deltakere. Jeg skal gå inn på forskningen som undersøker temaet nærmere.

Forskning på manipulasjon

Hanson et al. (2006) på George Mason universitetet replikerte et tidligere eksperiment hvor grupper på 12 ble satt til å predikere tall. Prediksjonsmarkedet de brukte hadde en CDA-mekanisme (se forklaring av CDA i 4.2.2). Tallet på det riktige utfallet kunne enten være 0, 40 eller 100. Halvparten av deltakerne ble informert om et tall som ikke kom til å bli det riktige utfallet. Den andre halvparten ble fortalt om et annet tall som ikke var riktig. For eksempel hvis det riktige tallet egentlig var 40, fikk en halvpart vite at 0 ikke var riktig, og en annen fikk vite tallet 100. I eksperimentgruppa, fikk halvparten en bonus desto mer markedsprisen økte. Med andre ord, så fikk de insentiv for å manipulere prisen opp. Alle ble dessuten gjort oppmerksomme på at halvparten fikk bonus for å manipulere markedet.

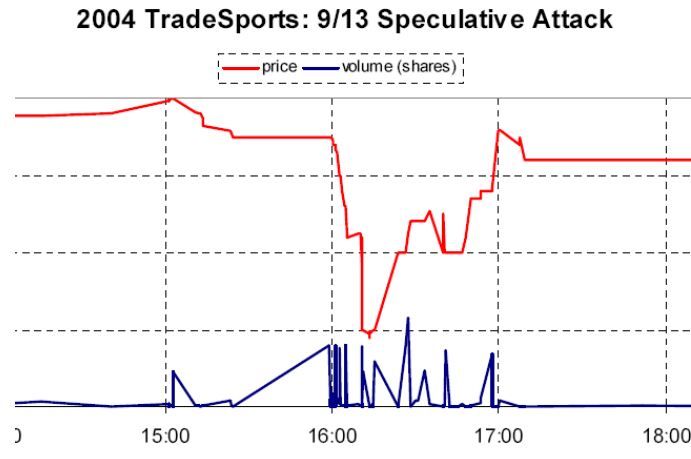
Det viste seg at insentiv for manipulasjon ikke hadde noen påvirkning på prisen i forhold til kontrollgruppa. Hanson et al. tror grunnen er at folk ikke var så villige til å akseptere en høy pris fra andre da de visste at noen prøvde å manipulere (fordi statistiske analyser tyder på dette).



Figur 6.1: Hvordan markedet bommet på det riktige utfallet. Desto høyere error-verdi, desto dårligere. MT er gruppa hvor manipulasjon foregikk, og RT er kontrollgruppa. Prisen var helt uventet, mer nøyaktig med manipulasjon enn hva den var uten manipulasjon.

Svakheten ved dette eksperimentet er at det vanskelig lar seg generalisere til virkelige situasjoner. For eksempel hvis man vil sabotere det interne markedet til Gundas bukser A/S, blir man ikke på forhånd informert om at det kan være andre deltakere som har motiv for å sabotere prisen. Dessuten kan størrelsen på insentivene som går til manipulasjon være langt høyere enn i eksperimentet. Kanskje snakk om flere millioner, hvis konkurrenten får en stor fordel på buksemarkedet av å bruke manipulasjon.

Undersøkelser fra åpne markeder (markeder som alle utenfra kan besøke) og veddemål har liknende funn som Hanson et al. Det er vanligvis nytteløst å manipulere markedsprisene. Rhode et al. (2006) oppsummerer en rekke mistenkelige forsøk på å manipulere åpne markeder på politiske utfall med mange deltakere, inkludert forsøk på manipulasjon av veddemål mellom 1880 og 1940. I de fleste av disse tilfellene, forandres prisen raskt tilbake til opprinnelig nivå igjen. Rhode et al. (2006) prøver også på å manipulere markeder selv ved å investere i tilfeldig utvalgte veddemål på Iowa Electronic Markets. De observerer akkurat det samme: Prisen går hurtig tilbake på opprinnelig nivå etter de har investert store summer for å få den til å bli annerledes.



Figur 6.2: Et mulig forsøk på manipulasjon av sjansen for at Bush vinner valget i 2004. Prisen kommer raskt tilbake på opprinnelig nivå igjen.

En generell svakhet ved disse studiene, er at de fleste av dem bare har sett på markeder som bruker CDA-mekanismen. Som du kanskje husker, så krever CDA at de som deltar må bli enige med hverandre om markedsprisene for at en handel skal skje. Men hvis vi derimot snakker om AMM-mekanismene, så er det mye lettere å flytte på prisene, fordi banken tar blindt i mot kjøp og salg fra deltakerne, og så beveger på prisene etter dette.

Oppsummering: Manipulasjon

Hvis noen har ønsker om å manipulere prediksjonsmarkedet, så er det en rekke studier som tyder på at dette ikke er så lett (til tross for ekspertuttalelser for at det motsatte er tilfellet).

Det er derimot flere grunner til å heller tro at andre metoder (Delphi, statistisk gruppe, diskusjonsgrupper) er mer utsatt for manipulasjon enn hva prediksjonsmarkedet er. For det første, stiller ikke noen av metodene folk til ansvar for å ha gjort dårlige prediksjoner. For det andre, så gir ikke de andre metodene motivasjon til andre gruppemedlemmer å rette på de som bidrar med feil informasjon. Det er ingen form for belønning for å gjøre dette. Hvis Bjørn ser at prisen på rosa bukser er helt på jordet (markedet spår at rosa bukser blir vellykkede med 90 % sjanse), så kan han tjene gode penger på å presse ned prisen på dem til det nivået han tror det bør være.

I eksperimentet av Hanson et al. beskrevet tidligere, var halvparten av 12 deltakere satt til å manipulere prediksjonene. Likevel, klarte de ikke oppgaven. Det hadde vært interessant å se hvordan deltakere med Delphi-metoden hadde taklet at halvparten av deltakerne ble satt til å gi feil prediksjoner (eller halvparten av en statistisk gruppe). Det samme med et normalt møte, som vi vet er ekstra sårbare for gruppepress.

Kort sagt, selv om flere personer har kommet med påstander om at prediksjonsmarkeder kan manipuleres, så har ikke disse påstandene empirisk støtte. Og det er heller ingen garanti for at alternativ beslutningsstøtte ikke kan utsettes for like mye (eller mer) manipulasjon.

6.2 Kvaliteten på studier

Denne oppgaven spør hva som skjer hvis PM'er brukes på en bedrift hvor ansatte ikke selv har valgt å delta. Det er få studier som undersøker nettopp ufrivillig deltakelse. Få forskere velger på egen hånd ut forsøkspersonene sine. De fleste av dem undersøker markeder hvor folk over internett (som ikke er tilfeldig valgt ut) på egen hånd går og prøver lykken. Bare 13 av ca. 150 studier laget inntil 2006, er definert som 'eksperimenter'. De fleste andre er 'observasjonsstudier' hvor forsøkspersonene ikke er valgt ut bevisst (Tziralis and Tatsiopoulos, 2007). Det kan være noe med de deltakerne som selv velger å delta som man ikke finner igjen hos andre deltakere. For eksempel Forsythe et al. (1999) rapporterer at prisene flere ganger blir 'feil' når de setter opp eksperimenter på PM'er i laboratoriet enn hva som skjer på prediksjonsmarkeder hvor forsøkspersonene ikke er valgt ut bevisst.

For det andre, utelater mange å sammenligne prediksjonsmarkeder med aktuelle alternativer som Delphi-metoden eller gruppediskusjon. Noen lar til og med være å rapportere hva de sammenligner prediksjonsmarkedet med. F.eks. Ortner (1998) rapporterer et eksperiment fra Siemens, men lar være å sammenligne prediksjonsmarkedet med andre metoder. Chen and Plott (2002) har ikke nevnt hva de sammenligner PM'et med når de skriver at PM'et slo 'internal forecasts' 6 av 8 ganger. Jeg måtte kontakte dem på mail. Da svarte de at det var en blanding av statistikk og ekspertvurderinger.

Hvis du introduserer noe nytt i en organisasjon, så økes ofte innsatsen til de ansatte uansett hva det nye er (f.eks. ny belysning, nye møbler). Dette kalles en 'Hawthorne-effekt'. Vi kan ikke vite sikkert om Prediksjonsmar-

kedet er mer enn bare en Hawthorne-effekt når det er mange studier som er dårlige på å sammenligne det med alternative metoder.

Denne oppgavens problemstilling er interessert gruppebeslutningstaking i organisasjoner. PM-feltet kan ha en større sjanse for å publisere positive enn negative resultater (kalt 'publikasjonsbias') fra organisasjoner. Bruk av PM'er i kjente organisasjoner siteres ofte av forskere. F.eks. Studien fra Hewlett-Packard som jeg har gått inn på tidligere. Andre er Microsoft, Dell og Google. Men det er uklart hva som hadde skjedd hvis slike organisasjoner hadde funnet negative resultater. Vil organisasjonen det rapporteres fra gå med på å publisere en mislykket historie?

Andre har Prediksjonsmarkedet som levebrød, og holder taler på dyre konferanser. For eksempel selger Jed Christiansen seg som en 'prediction market consultant' via sin hjemmeside. Han er aktiv på PM-konferanser og har publisert en studie (Christiansen, 2007). Hva om Christiansen fikk et negativt resultat? Ville han ha våget å rapportere det? For eksempel Christiansen (2007) finner ikke igjen det fenomenet mange andre studier finner som er kalt 'favorite-longshot bias'.

Kort sagt, så bør man være varsom med å trekke konklusjoner basert på studier fra prediksjonsmarkeder. I denne oppgaven, har jeg vært forsiktig med å gjøre det. For det første fordi få studier er 'ekte' eksperimenter, for det andre fordi studier bruker dårlige sammenligninger og for det tredje fordi positive funn har større sjanse for å bli publisert enn negative.

Kapittel 7

Konklusjon

There are many methods for predicting the future. For example, you can read horoscopes, tea leaves, tarot cards, or crystal balls. Collectively, these methods are known as 'nutty methods'.

- Scott Adams

For å predikere konsekvensene av beslutninger, er det foreslått flere forskjellige metoder: Delphi-metoden, statistiske grupper, diskusjonsgrupper og prediksjonsmarkeder.

Alternative metoder til prediksjonsmarkeder, fører til 'flertallstyrannier'. Under flertallstyranni, avgjør flertallet beslutningene, selv når det har negative konsekvenser. Prediksjonsmarkeder har på en annen side fenomenet 'Marginal traders'. Det vil si at de få som gjør færre feil enn andre også har makt til å påvirke prediksjonene mer enn andre.

Verken Delphi, statistiske grupper eller diskusjonsgrupper bruker insentiver. Siden prediksjonsmarkeder gjør dette, fører det til økt prediksjonsnøyaktighet, færre som hermer etter hverandre, og større kreativitet.

Pollyanna-prinsippet sier at vi helt naturlig søker positiv informasjon, og ignorerer negativ. Når vi overser viktig informasjon, så fører det også til at vi tar dårligere beslutninger. Jeg var inne på at kritisk tankegang kan fungere som et botemiddel mot pollyanna-prinsippet. Ingen av de andre metodene stimulerer kritisk tankegang. Prediksjonsmarkeder derimot, gjør dette ved å belønne deltakerne bedre for å investere pengene sine i veddemål som ikke er populære.

Egenskap	Delphi	Stat. gruppe	Diskusjon	Prediksjonsmarked
Innflytelse	-	-	+	+
Kritikk	-	-	-	+
Insentiver	-	-	?	+

Bibliografi

- L.R. Anderson. Payoff Effects in Information Cascade Experiments, 2001.
- P. Andersson, J. Edman, and M. Ekman. Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21(3):565–576, 2005.
- J.S. Armstrong. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, 2001.
- J.S. Armstrong. How to Make Better Forecasts and Decisions: Avoid Face-to-Face Meetings. *Foresight*, 5:3–8, 2006.
- B.B. Baltes, M.W. Dickson, M.P. Sherman, C.C. Bauer, and J.S. LaGanke. Computer-Mediated Communication and Group Decision Making: A Meta-Analysis. *Organizational Behavior and Human Decision Processes*, 87(1):156–179, 2002.
- I. Benbasat and J. Lim. Information Technology Support For Debiasing Group Judgments: An Empirical Evaluation. *Organizational Behavior and Human Decision Processes*, 83(1):167–183, 2000.
- J. Berg, F. Nelson, and T. Rietz. Accuracy and Forecast Standard Error of Prediction Markets. *Tippie College of Business Administration, University of Iowa*, 2003.
- J. Berg, F. Nelson, T. Rietz, and W. Draft. Prediction Market Accuracy in the Long Run. *Working paper*, 2008.
- P.J. Boland. Majority Systems and the Condorcet Jury Theorem. *The Statistician*, 38(3):181–189, 1989.
- W.F.M.D. Bondt and R.H. Thaler. Financial Decision-Making in Markets and Firms: A Behavioral Perspective, 1994.

- S.E. Bonner, R. Hastie, G.B. Sprinkle, and S.M. Young. A review of the effects of financial incentives on performance in laboratory tasks: Implications for management accounting. *Journal of Management Accounting Research*, 12:19–64, 2000.
- GH Bruggen, M. Spann, GL Lilien, and B. Skiera. Institutional Forecasting: The Performance of Thin Virtual Stock Markets. *draft*, 2006.
- R. Buehler, D. Messervey, and D. Griffin. Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation? *Organizational Behavior and Human Decision Processes*, 97(1):47–63, 2005.
- C.F. Camerer and R.M. Hogarth. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19(1):7–42, 1999.
- C. Camerer, G. Loewenstein, and M. Weber. The Curse of Knowledge in Economic Settings: An Experimental Analysis. *The Journal of Political Economy*, 97(5):1232–1254, 1989.
- C. Camerer. The rationality of prices and volume in experimental markets. *Organizational behavior and human decision processes(Print)*, 51(2):237–272, 1992.
- K.Y. Chen and C.R. Plott. Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem. *Lee Center Workshop*, 2002.
- J.D. Christiansen. Prediction Markets: Practical Experiments in Small Markets and Behaviours Observed. *Journal of Prediction Markets*, 2007.
- M. Cinnirella and B. Green. Does cyber-conformity vary cross-culturally? Exploring the effect of culture and communication medium on social conformity. *Computers in Human Behavior*, 2006.
- T. Connolly, L.M. Jessup, and J.S. Valacich. Effects of Anonymity and Evaluative Tone on Idea Generation in Computer-Mediated Groups. *Management Science*, 36(6):689–703, 1990.
- B. Cowgill, J. Wolfers, U.P. Wharton, and E. Zitzewitz. Using Prediction Markets to Track Information Flows: Evidence from Google. *Google*, 2008.

- V. Crupi. Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, 99999(1):1–1, 2008.
- D.M. DeRosa, C.L. Smith, and D.A. Hantula. The medium matters: Mining the long-promised merit of group interaction in creative idea generation tasks in a meta-analysis of the electronic group brainstorming literature. *Computers in Human Behavior*, 23(3):1549–1581, 2007.
- R. Forsythe, F. Nelson, G.R. Neumann, and J. Wright. Anatomy of an Experimental Political Stock Market. *The American Economic Review*, 82(5): 1142–1161, 1992.
- R. Forsythe, T.A. Rietz, and T.W. Ross. Wishes, expectations and actions: a survey on price formation in election stock markets. *Journal of Economic Behavior and Organization*, 39(1):83–110, 1999.
- D. Gigone and R. Hastie. The common knowledge effect: information sharing and group judgment. *Journal of personality and social psychology*, 65 (5):959–974, 1993.
- D. Gigone and R. Hastie. Proper analysis of the accuracy of group judgments. *Psychological bulletin*, 121(1):149–167, 1997.
- M. Glaser and M. Weber. Overconfidence and trading volume. *The GE-NEVA Risk and Insurance Review*, 32(1):1–36, 2007.
- K.C. Green, J.S. Armstrong, and A. Graefe. Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. *draft*, 2007.
- T.S. Gruca, J. Berg, and M. Cipriano. The Effect of Electronic Markets on Forecasts of New Product Success. *Information Systems Frontiers*, 5(1): 95–105, 2003.
- O. Grøtte. *Aksjekjøp og Daytrading - metode, psykologi, risiko og strategier*. Hegnar media, 2006.
- R.S. Gurkaynak and J. Wolfers. *Macroeconomic Derivatives: An Initial Analysis of Market-based Macro Forecasts, Uncertainty, and Risk*. National Bureau of Economic Research Cambridge, Mass., USA, 2006.
- Robin Hanson. Decision markets. *IEEE Intelligent Systems*, pages 16–19, 1999.

- R. Hanson, R. Oprea, and D. Porter. Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior and Organization*, 60(4):449–459, 2006.
- R. Hanson. Shall We Vote on Values, But Bet on Beliefs? *Draft*, 2003.
- R. Hanson. Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation. *Journal of Prediction Markets*, 1(1):3–15, 2007a.
- R. Hanson. The Policy Analysis Market. *Innovations*, 2007b.
- R. Hastie and R.M. Dawes. *Rational Choice in an Uncertain World: The Psychology of Judgement and Decision Making*. Sage, 2001.
- R. Hastie. Experimental evidence on group accuracy. *Decision Research*, 2: 129–157, 1986.
- FA Hayek. v. 1945. The use of knowledge in society. *American Economic Review*, 35(4):519–530, 1945.
- G.W. Hill. Group versus individual performance: Are N+ 1 heads better than one. *Psychological Bulletin*, 91(3):517–539, 1982.
- M. Hogg and V. Vaughan. *Social Psychology, 4th edition*. Prentice Hall, 2005.
- A.A. Hung and C.R. Plott. Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions. *The American Economic Review*, 91(5):1508–1520, 2001.
- D.J. Isenberg. Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50(6):1141–1151, 1986.
- J. Jones and D. Hunter. Qualitative Research: Consensus methods for medical and health services research. *British Medical Journal*, 311(7001): 376, 1995.
- D. Kahneman, P. Slovic, and A. Tversky. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
- E. Kirchler and B. Maciejovsky. Simultaneous Over- and Underconfidence: Evidence from Experimental Asset Markets. *Journal of Risk and Uncertainty*, 25(1):65–85, 2002.
- B.G. Malkiel. *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*. WW Norton & Company, 2003.

- B. Mangold, M. Dooley, GW Flake, H. Hoffman, T. Kasturi, DM Pennock, and R. Dornfest. The Tech Buzz Game [stock market prediction]. *Computer*, 38(7):94–97, 2005.
- C.F. Manski. Interpreting the predictions of prediction markets. *Economics Letters*, 91(3):425–429, 2006.
- M.W. Matlin. *Cognitive Illusions: A Handbook On Fallacies And Biases In Thinking, Judgement And Memory*. Psychology Press, 2004.
- T. Mendelberg. The Deliberative Citizen: Theory and Evidence. *Political Decision Making, Deliberation and Participation*, 6:151–93, 2002.
- CJ Nemeth. Differential contributions of majority and minority influence. *Psychological review*, 93(1):23–32, 1986.
- B.A. Nijstad, W. Stroebe, and H.F.M. Lodewijkx. The illusion of group productivity: A reduction of failures explanation. *European journal of social psychology*, 36(1):31–48, 2006.
- K. Oliven and T.A. Rietz. Suckers Are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market. *Management Science*, 50(3):336–351, 2004.
- G. Ortner. Forecasting Markets—An Industrial Application. *Working pa*, 1998.
- D.M. Pennock. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. *Proceedings of the 5th ACM conference on Electronic commerce*, pages 170–179, 2004.
- D. Pennock and R. Sami. Computational aspects of prediction markets, 2007.
- M. Peters, AM So, and Y. Ye. Pari-Mutuel Markets: Mechanisms and Performance. *LECTURE NOTES IN COMPUTER SCIENCE*, 4858:82, 2007.
- S. Plous. *The psychology of judgment and decision making*. McGraw-Hill New York, 1993.
- S. Plous. A comparison of strategies for reducing interval overconfidence in group judgments. *Journal of applied psychology*, 80(4):443–454, 1995.
- P.W. Rhode and K.S. Strumpf. Historical Presidential Betting Markets. *The Journal of Economic Perspectives*, 18(2):127–141, 2004.

- P.W. Rhode, K.S. Strumpf, U.N.C.C. Hill, and N.U.N.C.C. Hill. Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data. *University of North Carolina at Chapel Hill Working Paper*, 2006.
- G. Rowe and G. Wright. The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, 15(4):353–375, 1999.
- G. Rowe, G. Wright, and A. McColl. Judgment change during Delphi-like procedures: The role of majority influence, expertise, and confidence. *Technological Forecasting & Social Change*, 72(4):377–399, 2005.
- G.W. Schwert. *Anomalies and Market Efficiency*, 2002.
- E. Servan-Schreiber, J. Wolfers, D.M. Pennock, and B. Galebach. Prediction Markets: Does Money Matter? *Electronic Markets*, 14(3):243–251, 2004.
- C.M. Smith, R.S. Tindale, and L. Steiner. Investment Decisions by Individuals and Groups in Sunk Cost Situations: The Potential Impact of Shared Representations. *Group Processes & Intergroup Relations*, 1(2):175, 1998.
- V.L. Smith and J.M. Walker. Monetary Rewards and Decision Cost in Experimental Economics. *Economic Inquiry*, 31(2):245–61, 1993.
- J.A. Sniezek. A Comparison of Techniques for Judgmental Forecasting by Groups with Common Information. *Group & Organization Management*, 15(1):5, 1990.
- JA Sniezek. Groups under uncertainty: an examination of confidence in group decision making. *Organizational behavior and human decision processes(Print)*, 52(1):124–155, 1992.
- T. Sprenger, P. Bolster, and A. Venkateswaran. Conditional Prediction Markets as Corporate Decision Support Systems An Experimental Comparison with Group Deliberations. *Journal Of Prediction Markets*, 2007.
- G. Stasser and W. Titus. Pooling of unshared information in group decision making: biased information sampling during discussion. *Journal of personality and social psychology*, 48(6):1467–1478, 1985.
- C.R. Sunstein. The Law of Group Polarization. *Journal of Political Philosophy*, 10(2):175–195, 2002.

- C.R. Sunstein. Group Judgments: Statistical Means, Deliberation, and Information Markets. *New York University Law Review*, 80(3):962–1049, 2005.
- J. Surowiecki. *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday, 2004.
- P.E. Tetlock. *Expert Political Judgment: How Good is It? How Can We Know?* Princeton University Press, 2005.
- G. Tziralis and I. Tsiopoulos. Prediction Markets: An Extended Literature Review. *Journal of Prediction Markets*, 2007.
- J.S. Valacich and C. Schwenk. Devil' s Advocacy and Dialectical Inquiry Effects on Face-to-Face and Computer-Mediated Group Decision Making. *Organizational Behavior and Human Decision Processes*, 63(2):158–173, 1995.
- L.V. Williams. 7 Models, markets, polls and pundits: a case study of information efficiency. *Information Efficiency in Financial and Betting Markets*, 2005.
- H. Winklhofer, A. Diamantopoulos, and S.F. Witt. Forecasting practice: A review of the empirical literature and an agenda for future research. *International Journal of Forecasting*, 12(2):193–221, 1996.
- J. Wolfers and E. Zitzewitz. Prediction Markets. *The Journal of Economic Perspectives*, 18(2):107–126, 2004.
- I. Yaniv. The Benefit of Additional Opinions. *Current Directions in Psychological Science*, 13(2):75–78, 2004.