Masteroppgave om misoppfatninger innenfor sannsynlighet

En vurdering av et spill som diagnostisk verktøy

Arve E. Handeland

Lektorprogrammet

30 studiepoeng

UiO

Det utdanningsvitenskapelige fakultet

15.06.2021

Sammendrag

Denne kvantitative studien har som formål å undersøke et spills evne til å diagnostisere misoppfatninger innenfor sannsynlighet. Fagfornyelsen løfter frem viktigheten av å tenke kritisk for å fatte gode valg og ta stillig til viktig spørsmål i eget liv og samfunnet (Utdanningsdirektoratet, 2019a, s. 2). Sannsynlighetsundervisningen anses som sentral for oppøvingen av den kritiske tenkningen og beslutningstakingen, og forstås som høyst relevant for å bidra til den teknologiske utviklingen (Utdanningsdirektoratet, 2019b; Kahneman, 2013). Sannsynlighet er et fagområde hvor mange har en svak intuitiv forståelse, og misoppfatninger hindrer videre læring (Shaughnessy, 1977). For å overkomme disse misoppfatningene trengs det egnede ressurser, samt mer lærerkompetanse (Brekke, 2002). Forskningsspørsmålet mitt lyder: Hvordan kan bruken av spillet Borel være en ressurs i matematikkundervisning for å gi informasjon om elevers misoppfatninger innenfor sannsynlighet?
For å undersøke dette ble det utviklet et undervisningsopplegg, sentrert rundt spillet. Gjennom undervisningsopplegget skulle forekomsten misoppfatninger innenfor sannsynlighet kartlegges. Det var snakk om et utvalg misoppfatninger som henger tett sammen med temaets kompetansemål. Undervisningsopplegget ble gjennomført på en ungdomskole, og 107 elever på 10. trinn deltok i undersøkelsen. Metoden anvendte en survey for det meste bestående av dikotome variabler. Surveyen var innarbeidet som en naturlig del av spillet. Funnene peker på at undervisningsopplegget har styrker for didaktisk bruk og kan anvendes som en ressurs som gir informasjon om misoppfatninger. Undersøkelsen viser videre at undervisningsopplegget ikke frembringer éntydige data, og derfor har svakheter som konkluderende diagnostisk verktøy.

Forord

Det har vært en fryd og et mareritt å lage denne masteren. Fryden har knyttet seg til å overkomme hindringer, kjenne mestringsfølelse etter hodestanging, og huke av punkter på gjøremålslista. Og marerittet har vært å ha en bekmørk sky hengende over meg i halvannet år; en sky som sjeldent har latt meg glemme hvor nærliggende den neste fristen er og hvor dårlig jeg har ligget an.

Takk til all støtte som har vært å hente.

Takk til veileder Helmer Aslaksen for dyptskuende tolkninger, artige anektoder og stadig kreative innspill.

Takk til de fine studiekameratene som har lokket frem smil hver eneste dag. Dere skal virkelig krediteres for at dette studieløpet har gått sin gang, og at selve reisen virkelig har vært en gledelig affære. Takk til Gautefar for at du alltid baner vei med din effektivitet og er forbilledlig med din ærlighet. Og takk for at du har vært en bror i troen i fem fete og magre år. Takk til GlowNikolai for sang og skuespill, evinellige erindringer, pinlig nøyaktighet, og abstrakt humor. Takk til EvnerikeAhmed som har inspirert med dype tanker, lat livsførsel, og visjonen om å ha visjoner. Og takk til alle andre i krinker og lektorkroker som har brakt relasjonelle lyspunkt inn i hverdagen.

Takk til mamma, pappa og storebror, og alle venner som i like stor grad anses som nær familie, som stadig spør, lytter og tilbyr utstrakte hender.

Takk til min kjære, tilkommende fru, Ingrid, som nå har nilest og rettet denne oppgaven. Takk for at du alltid støtter og bekrefter meg i min galskap og barnlighet. Takk for at du inspirerer meg med godhet, kjærlighet og tro. Takk for all mat du lager, og – akkurat nå – alle kaker du *ikke* bager.

Takk til det norske utdanningstilbudet, som lar meg ta en utdanning hvor jeg får bruke meg selv og mine gaver.

Ikke minst: Takk til Jesus, for talenter, nåde, fred og håp!

Og takk og pris for at dette nå er ferdig.

Oslo, juni 2021
Arve Esteban Handeland

Innhold

1 Introduksjon

1.1 Bakgrunn for valg av oppgaven

Sannsynlighet har lenge stått som et diffust og utfordrende fagområder for meg. I møte med det har jeg vært en antikkens sjømann; stadig innhyllet i forfjamselsens tåke har jeg blitt dratt inn i ugunstige farvann av sirener som lokker med lettvinte løsninger. Og med en strandet forståelse og en arbeidslyst knust mot forvirringens feilskjær har jeg funnet det fornuftig å la sannsynlighetens tematikk ligge i fred. Når nå studiet skulle fullendes meldte en boblende følelse seg. Skulle vi ta en siste dans, sannsynligheten og jeg? Så på tross av dette anstrengte forholdet – eller kanskje på grunn av – ble det altså til at dansegulvet skulle inntas.
Sakte men sikkert ble veien klarere. En delt spillinteresse med veileder brakte en umiskjennelig mulighet på banen, og jeg bestemte meg for å bruke et bestemt spill til å studere elevers misforståelser innenfor sannsynlighet. Således var min didaktisk harpun funnet, og jeg var klar til å møte med min hvite hval nok en gang.

1.2 Rasjonale

Det blir sagt at skolen «har et dobbelt samfunnsoppdrag. Den skal både danne og utdanne» (Kunnskapsdepartementet, 2017). Mellin-Olsen (1981, ss. 356-360) adresserer denne dualismen med det han kaller skolens *S-rasjonale*, den sosiale viktigheten av skolelæring og læreplanmål, og *I-rasjonale*, skolen som instrument for at eleven skal få en god fremtid og videreutvikle samfunnet. Utdanningen av elever er klart knyttet til I-rasjonalet, hvor kompetanser utvikles for at elevene skal kunne ta plass i yrkeslivet, og bidra til reproduksjon og utvikling av samfunnet. Danningen av elever skjer gjennom alle de erfaringene og utfordringene de møter i skolehverdagen, og har «enkeltmenneskets frihet, selvstendighet, ansvarlighet og medmenneskelighet som mål» (Kunnskapsdepartementet, 2017). Dette aspektet ved opplæringen er en dynamisk prosess mellom individet og kulturen, og kan slik forstås som en del av S-rasjonalet (Ulvik & Sæverot, 2017, s. 35). Dette dobbelte samfunnsoppdraget kan settes i en spesifikk matematikkfaglig kontekst. Den danske professoren Niss (1996, ss. 13-14) fremmer tre fundamentale grunner for å undervise matematikk, som matematikkfaget må bidra til dersom det skal være et *viktig* fag. Det må bidra til: 1) den teknologiske og sosioøkonomiske utviklingen av samfunnet, 2) bevaring og videreutvikling av samfunnets politikk, og ideologi, og 3) å gi enkeltindivid forutsetninger for å takle livets ulike sfærer (Niss, 1996, s. 13). Slik kan vi forstå hvordan matematikkfaget må svare på rasjonalene for å forsvare sin sentrale posisjon i skolen. I Fagfornyelsen blir matematikkfagets relevans og sentrale verdier beskrevet, og man ser her hvordan den norske skolen forstår og implementerer Niss sine punkter. Utdanningsdirektoratet (2019a, s. 2) sier:

> Matematikk er et sentralt fag for å kunne forstå mønstre og sammenhenger i samfunnet (...) Matematikk skal bidra til at elevene utvikler et presist språk for resonnering, kritisk tenkning og kommunikasjon gjennom abstraksjon og generalisering. Matematikk skal forberede elevene på et samfunn og arbeidsliv i utvikling ved å gi dem kompetanse i utforsking og problemløsing. (...) Kritisk tenkning i matematikk omfatter kritisk vurdering av resonnementer og argumenter og kan ruste elevene til å gjøre egne valg og ta stilling til viktige spørsmål i sitt eget liv og i samfunnet.

Utdraget knytter matematikkens mål klart sammen med skolens dobbelte samfunnsoppdrag. Det legges vekt på at elevene skal kunne tre inn i et yrkesliv og samfunn i utvikling, som henspiller på I-rasjonalet. Samtidig har Fagfornyelsen hatt et forøket fokus på å oppøve elevenes evne til å lære og tenke selvstendig, som kan forstås som S-rasjonalet (Regjeringen, 2019). Dette gjenspeiles i kompetansenes ordlyd i utdraget ovenfor. Elevene skal kunne: *forstå* og *utforske*; *løse problemer* og *resonnere*; *tenke kritisk* og *ta valg*. Her står elevenes metakognisjon mer sentralt og prominent enn det har gjort i tidligere læreplaner. Således svarer matematikkkfaget på det Niss (1996) etterlyser, og er både med på å danne og utdanne. Det skal likevel nevnes at flere kompetansemål har blitt *mindre* eksplisitte med Fagfornyelsen (Utdanningsdirektoratet, 2019a, ss. 13-14). Niss (1996) sier at for at slike hensikter og mål skal ha slagkraft er det viktig at de er så eksplisitte som mulig. Bare på den måten vil det kunne sikres godt samsvar mellom den intenderte læreplanen og den oppfattede læreplanen. Dette viser noe av det spennet skolen står i; den ønsker åpne mål med rom for personlig danning, samtidig som læreplanene må være konkrete og konsekvente for å kunne implementeres og utdanne til arbeidslivet.

Med Fagfornyelsen har fagområdene *statistikk* og *sannsynlighet* blitt reduseret for å gi rom til dybdelæreing (Utdanningsdirektoratet, 2019b). På ungdomstrinnet er fagområdet bevart i kompetansemålene for 9. trinn, og på videregåedene videreføres det i valgfagene Matematikk S1 og S2 (Utdanningsdirektoratet, 2019a; Utdanningsdirektoratet, 2020). Reduksjonen har ført til at disse fagområdene som konkrete kompetansemål har blitt fjernet fra 10. klasse og alle de øvrige valgfagene på videregående (Utdanningsdirektoratet, 2013; Utdanningsdirektoratet, 2019a). Emnets omfangsreduksjon ble kommentert under Fagfornyelsens høringsrunder, og det fremkom at flere instanser mener at statistikk og sannsynlighet «er et kunnskapsområde som vil bli stadig mer viktig med fremveksten av teknologier som maskinlæring, store data, o.l. I tillegg fremheves statistikk som særlig relevant knyttet til demokrati og medborgerskap og bærekraftig utvikling, herunder forståelse for samfunnet og politiske og personlige valg» (Utdanningsdirektoratet, 2019b). Å kunne henge med på den teknologiske utviklingen har tradisjonelt sett vært et av de viktigste argumentene, ut fra I-rasjonalet, for å undervise matematikk (Niss, 1996, s. 14; Mellin-Olsen, 1981, s. 358). Statistikk og sannsynlighet fremmes her som både relevante og viktige for dette formålet. Høringsinnspillet fremhever videre statistikkens og sannsynlighetens rolle i de nyinnførte tverrfaglige temaene. De tverrfaglige temaene sier at matematikkfaget skal utvikle forståelse som hjelper å ta ansvarlige livsvalg, og en metakognisjon tilknyttet beslutninger rundt eget liv og i samfunnet (Kunnskapsdepartementet, 2017). Disse målene peker direkte på opplæringslovens formålsparagraf; elevene skal kunne mestre livene sine og delta i samfunnet (Regjeringen, 2021, §1 − 1). Som samfunnsaktør er det mange valg å ta stilling til, om det er relativt trivielle ting som værmeldinger, forsikringsavtaler og lottotipping, eller mer fagspesialiserte ting som entropiøkning eller DNA-bevis i rettsaker (Hirsch & O'Donnell, 2001, 1. avsnitt). I skrivende stund faller det også naturlig å nevne valg tilknyttet koronapandemien, både på nasjonale og personlige plan. Det er i møte med slike valg og usikre utfall at statistikken og sannsynligheten er verktøy som hjelper en å velge fornuftig. Slik blir statistikken og sannsynligheten relevant for enkeltmennesket i sin egne sosiale virkelighet – det hjelper mennesket å fatte gode valg i egne liv. Med bakgrunn i høringsinnspillene ble det fremmet et forslag om å styrke statistikkens og sannsynlighetens omfang i den nye læreplanen, noe som ikke skjedde (Utdanningsdirektoratet, 2019b). Fagutvalget i Norsk Lektorlag (2018) konstaterer at «elever som bare tar matematikk som fellesfag fra 1. til 11. trinn og 2P i videregående, blir veldig lette å slå i brettspill». Dersom dette er den eneste konsekvensen fra en slik redusksjon tyder det kanskje på at man i norsk skole ikke har vært helt klare på hva emnets hensikter og mål er, og hvordan disse knytter seg til skolens oppdrag.

For norsk utdanning aktualiseres dette ytterligere ved å se på norske elevers prestasjoner innenfor statistikk og sannsynlighet på internasjonale testresultater. En slik undersøkelse som man kan se til er TIMSS-undersøkelsen, som, blant annet, måler elevers kompetanse i matematikk på 5. og 9. trinn (ILS, 2020). Formålet med undersøkelsen er å innhente data for »å bidra til å styrke læring og undervisning i realfagene» (ILS, 2020). Når man gjennom læreplanen har sikret at dette er nyttig for elevers danning og utdanning, er det viktig at man sikrer elevene best mulig forutsetninger for å lære. TIMSS-undersøkelsen viser en signifikant nedgang hos norske elever innen området statistikk og sannsynlighet fra 2015 til 2019 (Kaarstein, Radišić, Lehre, Nilsen & Bergem, 2019, s. 18). Hva som er skyld i denne nedgangen er ikke lett å konkludere med, og skyldes trolig ikke bare én ting. TIMSS-undersøkelsen viser blant annet at det er en klar sammenheng mellom hvordan elevene opplever undervisningen og hvordan de presterer (Kaarstein et al., 2019, s. 38). Forskning sier at *ingen* annen enkeltfaktor har så mye å si for elevenes læring som læreren og hennes undervisning (Klette, 2013). Lærerens kompetanser har mye å si for å skape gode rammer hvor elevene motiveres (Manger, 2013). Noe av det TIMSS-undersøkelsen viser er en nedgang i andelen elever som rapporterer om høy faglig selvtillit og motivasjon (Kaarstein et al., 2019). Tyder dette på at lærerpraksisen bør utvikles på dette fagfeltet? Undervisningen er avhengig av at læreren klarer å gi god undervisningsmessig støtte som brer seg over de ulike undervisningssituasjonene (Klette, 2013, ss. 180-190). Dette er komplekse begrep som rommer mye, men noe av det de klarest taler imot er monotoni og oppgaver som ikke er kognitivt utfordrende.

En annen ting TIMSS viser er at norske lærere er mye mer ensformige i vurderingspraksisen sin en andre lærere på internasjonalt nivå (Kaarstein et al., 2019, 45). Spesielt innenfor matematikk vektlegger lærere lite annet enn lange, skriftlige prøve (Kaarstein et al., 2019, 45). For å utfordre slike satte mønstre passer det kanskje fint at underveisvurderinger har fått en stor rolle i Fagfornyelsen (Kunnskapsdepartementet, 2017). Kunnskapsdepartementet (2017) sier at underveisvurdering, all vurdering som skjer før avsluttningen av opplæringen, «(...) skal brukes til å fremme læring, tilpasse opplæringen og øke kompetansen i fag.» Dette gir diagnostiske oppgaver og arbeid en særegen plass i den nye læreplanen. God vurderingspraksis er viktig å kunne ha en god tilbakemeldings-praksis, hvor elevene gis rettledning som gjør at de kan utvikle seg (Hattie & Timperley, 2007). Å bruke vurderinger på en slik måte er essensielt for å tilpasse opplæringen til enkeltelevens behov og forutsetninger, slik det kreves i opplæringsloven (Regjeringen, 2021, §1 – 3; Engh, 2011).

Denne tematikken skal likevel tillegges et lite tankekors. Man må se resultatene for det de er verdt, og ikke tilegne hensikten deres mer enn det den opprinnelig skulle ha (Hopfenbeck, 2016). Er man ikke tydelig på dette vil testresultater kunne bli for styrende for undervisning og skolepolitikk, slik at skolen viker fra dens formål (Biesta, 2009; Hopfenbeck, 2016, s. 140). Er man denne problematikken bevisst vil slike undersøkelser kunne være gode ressurser og bidra med viktige pekepinner om hvilke retninger undervisningen bør ta.

Ut fra de foregående avsnittene fremstår det tydelig at sannsynlighets- og statistikkundervisning knytter seg tydelig til skolens S- og I-rasjonal, og fremtrer som høyst relevant i dagens skole. Fra TIMSS kan det virke som om den norske undervisningen har rom for endring om den sannsynlighetsundervisningen skal kunne nå sin hensikt og bidra til elevers danning og utdanning. En av de største utfordringene man møter når man skal lære bort om sannsynlighet, er at emnets egenart gjør at man møter mange misoppfatninger hos elevene (Batanero & Sanchez, 2005; Konold, 1991). Denne tematikken har ikke blitt drøftet enda, men som det skal vises i avsnitt 2.1 spiller slike oppfatninger en viktig rolle i undervisning. Slike misoppfatninger står ofte i veien for den videre læringen, og det er således viktig for den videre undervisningen at disse blir oppklart (Sjøberg, 2014).

> [I]f you don't understand probability, then you'll probably find yourself making some bad decisions. (…) Sometimes you'll act when you shouldn't, and other times you'll fail to act when you should. (Rowbottom, 2015, s. 1)

1.3 Forskningsspørsmål

Denne oppgaven skal se på muligheter og utfordringer som bruken av seriøse spill som vurderingsverktøy i undervisningen medbringer, og stiller spørsmålet:

Hvordan kan bruken av spillet Borel være en ressurs i matematikkundervisning for å gi informasjon om elevers misoppfatninger innenfor sannsynlighet?

For å besvare dette spørsmålet ble det utviklet et undervisningsopplegg viss formål var å avdekke misoppfatninger i sannsynlighet. Undervisningsopplegget er en didaktisk videreutvikling av et kommersielt spill kalt Borel. Dette kommer til uttrykk gjennom undervisningsoppleggets rammer og oppgaveformuleringer. Brorparten av opplegget går ut på at elevene skal anslå hva som er det mest sannsynlige utfallet i ulike oppgaver, og så ta valg knyttet til dette. Svarene utgjør dataene i denne oppgaven, og skal gjennom analyse belyse forskningsspørsmålet. Oppgavene ble laget med den hensikt å adressere misoppfatninger som gjør seg spesielt gjeldende i sannsynlighet på skolenivå. I tillegg til disse egenutviklede oppgavene ble det tatt i bruk oppgaver som har gitt tidligere forskningsprosjekter interessante funn, og disse ble tilpasset spillets format.

En pilotering ble gjennomført i flere klasser på en videregående skole, men på grunn av restriksjoner lot ikke innhentingen av data seg gjennomføre på videregående skoler. Selve datainnsamlingen ble da gjort på 10. trinn på en ungdomsskole. Det var i alt 107 elever som deltok i undersøkelsen. Hver elev skulle svare på ti spørsmål.

Undervisningsopplegget er ikke tiltenkt å måtte gjennomføres på en rigid måte, slik det ble gjennomført i datainnsamlingen. Min hypotese er at undervisningsopplegget vil kunne gi flere spennende pedagogiske angrepspunkter inn til sannsynlighetsundervisning. Jeg tror at undervisningsoppleggets utforming er så fleksibelt av natur kunne brukes på forskjellige måter og i ulike sammenhenger, og at det, i sin annerledeshet, vil kunne åpne noen nye pedagogiske dører. Oppgaven vil også belyse hvordan bruken av spill vil kunne fremme spennende nytenkning rundt vurderingspraksisen til lærere, om det skulle vise seg å være et godt verktøy for formativ vurdering. Vurderingen av disse antagelsene vil være viktige nøkkelpunkter å dvele ved for å kunne besvare problemstillingen.

1.4 Oppbygning

Denne masteroppgaven består av seks kapitler, medregnet det innledende. Dette første kapittelet forklarer valg av oppgaven, og begrunner dens relevans sett i kontekst av dagens skole og samfunn. Det andre kapittelet er en presentasjon av tidligere forskning på dette fagfeltet og ser denne oppgaven i rekken av en rik forskningsarv. Kapittel tre beskriver og drøfter relevant teori og litteratur, og danner et bakteppe for videre analyse og diskusjon. De mest aktuelle misoppfatningene vil bli gjennomgått, og det vil bli redegjort for bruk av spill og diagnostiske oppgaver undervisning. I kapittel fire beskrives metoden som ble brukt, og disputerer forskningens troverdighet. Det femte kapittelet favner forskningens funn, og består av to deler. Den første delen er en gjennomgang av de innsamlede dataene, mens den andre delen vil analysere av de gjeldende dataene. Her vil funn bli knyttet opp mot hvorvidt dataene indikerer at undervisningsopplegget er et godt diagnostisk verktøy. I det siste kapittelet, kapittel seks, vil analysen diskuteres ytterligere i lys av de foregående delene. Hovedmomentene blir samlet og danner en konklusjon for masteroppgaven.

2 Tidligere forskning

Dette kapittelet vil se på hvordan tidligere forskning på misoppfatninger og diagnostiske oppgaver relaterer seg til sannsynlighet. Den presenterte forskningen er med på å fremheve hvilket rom forskningsspørsmålet i denne masteravhandligen er med på å fylle.

2.1 Misoppfatninger

En *misoppfatning* er en konsekvent ufullstendig tanke knyttet til et begrep (Brekke, 2002, s. 10). Brekke (2002, s. 10) sier at misoppfatninger ofte er et «resultat av en overgeneralisering av tidligere kunnskaper til nye områder der disse kunnskapene ikke gjelder fullt ut». Forskning viser at det nesten ikke til å unngå at misoppfatninger danner seg under oppveksten, da generalisering er en del av naturlig oppvekst (Brekke, 2002, s. 11). Problemet fremgår når grunnlaget for å definere gyldighetsområdet er for svakt. Denne prosessen kan forstås ut fra kognitiv læringsteori. Et komplekst konsept vil kunne ha forskjellige sider, der noen stemmer overens med elevens tidligere erfaringer, mens andre motsier dem. Om eleven bare ser de sidene som sammenfaller med hennes erfaringer vil hun *assimilere* kunnskapen; ingen store endringer skjer i tankesettet (Säljö, 2013). Dette generaliseres til å gjelde hele konseptet, og den nødvendige *akkomodasjonen*, hvor store kognitive endringer skapes, uteblir (Säljö, 2013). Säljö (2013, s. 65) sier at kognitiv konflikt er viktig for menneskets utvikling, om man skal oppdage nye aspekter, men det avhenger av at man oppdager at det nye erfaringen ikke passer inn i de alleredeeksisterende skjemaene. Dersom disse misoppfatningene ikke blir bearbeidet vil de kunne skape problemer for den videre forståelsen og læringen (Brekke, 2002, s. 11; Sjøberg, 2014, s. 338). Med grunnlag i empiri peker Nygaard og Zernichow (2006) på læreren som en av de vanligste kildene til at elever får misoppfatninger i matematikk. De sier at læreres upresise begrepsbruk gjør at man tilegner matematikken egenskaper den aldri var tiltenkt å ha. For eksempel vil fraser som «så stryker vi oppe og nede» og «så flytter vi over og bytter fortegn» være med å mystifisere temaet *ligninger*, fordi de undergraver de faktiske matematiske operasjonene som foregår (Nygaard & Zernichow, 2006, s. 4). På den måten vil læreren kunne ende opp med å gjøre elevene en real bjørnetjenste om de ikke er sin egen undervisning og egne kunnskaper bevisst. Også lærebøker og læreplaner får noe av skylden (Nygaard & Zernichow, 2006). I matematikk er det mange slike misoppfatninger som går igjen, f.eks. at det lengste tallet er det med størst tallverdi og multiplikasjon og divisjon gjør alltid tallet henholdsvis større og mindre (Nygaard & Zernichow, 2006).

Kognitiv konflikt kan være en god ressurs som hjelper eleven å begrense et konsepts gyldighetsområde. For å skape kognitiv konflikt må en være sin egen tenkning bevisst, og se hvordan noen biter ikke passer inn i puslespillet. Evnen til å tenke over egen tenkning kalles *metakognisjon*, og er en viktig ferdighet å oppøve om man skal bli matematisk kompetent (National Research Counsil, 2001). Som vist ovenfor har denne kompetansen også fått en mer sentral plass i Fagfornyelsen, og er en viktig del av matematikkens hensikt i skolen. Shoenfeld (1987) delte metakognisjonen opp i tre områder, hvor det siste området omfattet nettopp elevens oppfatninger og ideer, og hvorvidt eleven tenker over hvordan disse former det matematiske arbeidet. Shoenfeld (1987) sin forskning viser at mye av nøkkelen til å utvikle elevenes metakognisjon ligger i at elevene skal samarbeide, utforske og rettferdiggjøre ulike tanke- og fremgangsmåter. Det er også viktig at denne praksisen skjer i sammenheng med at elevene utvikler sin forståelse av matematiske konsepter, og ikke er en separert læring (Lesh & Zawojewski, 2007, s. 771). I så henseende settes misoppfatninger i et lys av sosiokulturelle læringsteorier. For å overkomme misoppfatninger må eleven gis støtte i den proskimale utviklingssonen av andre kapable partnere, slik at eleven kan appropriere de medierende kompetansene som trengs for å få utvikle kognisjonen (Säljö, 2013).

Tidligere forskning viser at misoppfatninger er spesielt motstandsdyktige mot vanlige undervisningsmetoder som at en prøver å ignorere misoppfatningene, eller prøver å unngå at de kommer i det hele tatt ved å være presis i innføringen (Angell et al., 2018, s. 150; Brekke, 2002, s. 11). Det kan ha en sammenheng med at lærere har svak kompetanse på misoppfatninger (Öçal, 2018). Det kan bety at læreren ikke gjenkjenner misoppfatningene, eller ikke vet hvordan de bør håndteres. I fraværet av dette vil ikke det sosiokulturelle stilaset kunne bygges rundt eleven, og hun vil ikke få den støtten som trengs. Med fare for å sette læreren for sentralt etter den kognitive tradisjonen: en annen måte å se på det er at læreren ikke vil klare å tilrettelegge for at elevene skal akkomodere. Om dette er tilfellet faller det naturlig å spørre hvordan denne undervisningskompetansen kan økes. Det vil være lett å peke på lærerutdanningen og kreve mer undervisning opp mot misoppfatninger. En utfordring med en slik tilnærimg vil være gjennomførbarhet. En rent pedagogisk tilnæring vil være for generell, mens fra et didaktisk perspektiv vil være veldig mange misoppfatninger å ta hensyn til. Jones og Thornton (2005, s. 82) mener at forskningen på identifiserende verktøy allerede har fått stor plass i forskningen, og etterlyser kognitive rammeverk innenfor de ulike matematiske domenene, slik at lærere får større innsikt i *hvordan* den matematiske kunnskapen utvikles. Behovet for mer kunnskap og mer forståelse rundt utviklingen av læringsstøttende verktøy i møte med misoppfatninger var med å bane vei for denne oppgaven.

2.1.1 Misoppfatninger i sannsynlighet

Fra tidligere forskning vet man at mennesker, voksne som barn, har en svak intuitiv forståelse av sannsynlighet (Shaughnessy, 1977; Green, 1986; Kahneman, 2013). Resonneringen innenfor sannsynlighet skiller seg fra annen logisk resonnering (Batanero & Sanchez, 2005, s. 241). Allerede hos elever i tidlig skolealder nivå finner man kontraintuitive resultater, som er tett knyttet sammen med misoppfatninger (Batanero & Sanchez, 2005, s. 241). Sammenlignet med andre fagområder er dette uvanlig; slike resultater henger vanligvis sammen med høyt abstraksjonsnivå, og pleier ikke å forekomme før man er på et høyere faglig nivå (Batanero & Sanchez, 2005, s. 241). Disse resultatene kan altså ikke bare skyldes på lærere eller andre skoleforhold; det må være en annen faktor i spill. Denne faktoren kan forstås å være tett tilknyttet emnets egenart (Batanero & Sanchez, 2005, s. 241; Konold, 1991, s. 139). Sannsynlighet fortelle om hyppigheten en hendelse inntreffer med. Det kan for eksempel være 20 % sjanse for at hendelse skal inntreffe, men hva betyr det? Et utfall kan være lite sannsynlig og likevel inntreffe, og et annet kan være svært sannsynlig og fremdeles utebli. Å gjøre en tolkning av en slik tallfestet sannsynlighet er noe av det mange elever sliter med (Kahneman, 2013). For elevene fremstår dette som en motsats til de ellers faste matematiske rammene, hvor $2 \cdot 2 = 4$ uansett hva, gitt at man operere innenfor vanlig aritmetikk (Ang & Shahrill, 2014, s. 23). Denne egenskapen gjør at når man skal bruke sannsynlighet som et objektivt verktøy, så vil elevene se på fagområdet som «flyktigere» enn andre fagområder. Dette gjør at elever ofte overgeneraliserer semantikk og fremgangsmåter, og misoppfatninger oppstår (Batanero & Sanchez, 2005). At en tolkning av et korrekt svar også inneholder usikkerhet, gjør at elever i mindre grad opplever den kognitive konflikten som er viktig for å oppdage misoppfatninger (Shaughnessy, 1977, s. 295; Batanero & Sanchez, 2005).

Elevenes erfaringer spiller en viktig rolle for deres forståelse av sannsynlighet (Konold, 1991, s. 140). Gjennom dagligtalen blir elever kjent med mange sannsynlighetsbegrep, som at noe er *sannsynlig*, *uavhengig*, *tilfeldig* etc. (Konold, 1991, s. 144). Den hverdagslige bruken sammenfaller sjelden med den rigorøse og faglige, noe som ofte skaper forvirring og misoppfatninger når disse to verdenene møtes (Konold, 1991, s. 145). Læreren må derfor være bevisst på hvilke forkunnskaper og oppfatninger eleven tar med seg inn i klasserommet (Konold, 1991, s. 145; Batanero & Sanchez, 2005, s. 262). Slike oppfatninger trenger ikke å være semantiske, men kan være av praktisk art (Kahneman, 2013). Har man spilt Yatzy kan man ha opplevd å få Yatzy, 5-like, opptil flere ganger når man har samlet på 6-ere, men aldri med 2-ere. En antakelse basert på tidligere erfaringer kan fort medføre at man tenker at det er mer sannsynlig å vinne om man satser på 6-erne. Det fins mange misoppfatninger knyttet til *heuristikker* (Kahneman, 2013, s. 143). Disse vil bli sett mer på i avsnitt 3.2.2. Rowbottom (2015, ss. 8-9) sier at et viktig poeng, for å hjelpe elever distingvere mellom sannsynlighetens funksjonelle fasetter, kan være å omformulere spørsmålet «hva er *den* riktige tolkningen av sannsynlighet» til «hva er den riktige tolkningen utfra den-eller-den konteksten». Selv etter formell opplæring i sannsynlighet er det ikke uvanlig å finne misoppfatninger hos elever (Shaughnessy, 1977, s. 295; Öçal, 2018). Igjen peker dette på behovet for – ikke bare mer kompetanse på misoppfatninger – men mer kompetanse på sannsynlighets-misoppfatninger. Dette stemmer overens med det Jones og Thornton (2005, s. 82) sier: sannsynlighet har blitt innført i skolen, uten at lærere har fått tilstrekkelig med opplæring.

2.2 Diagnostiske oppgaver

2.2.1 Vurdering for læring

Tradisjonelt sett har vurderinger i skolen kun blitt brukt til å bedømme elevers avsluttende prestasjon og evne til å reprodusere faktakunnskaper (Brevik & Blikstad-Balas, 2014). Med nyere skoleforsknings inntog har perspektivet på vurderinger endret seg, og er dreid i en retning av at vurdering «handler om å samle bevis for læring og å tolke og anvende denne informasjonen i beslutniger om undervisning og læring» (Brevik & Blikstad-Balas, 2014, s. 2). Slik ser vi at vurderingens formål har endret seg; den er ment å være styrende for *fremtidig* undervisning, ikke bare reflektere den *foregående* undervisningen. Et sentralt begrep i denne sammenhengen er *vurdering for læring* (VfL), som vitner om formålet med læringen (Brevik & Blikstad-Balas, 2014, s. 2). En av nøkkelfaktorene for slik vurdering er tidspunktet vurderingen kommer på – den må være en integrert del av den daglige undervisningen (Kunnskapsdepartementet, 2017). Informasjon som innhentes gjennom samtaler, demonstrasjoner og observasjoner danner et grunnlag for å både justere egen undervisning og samtidig gi elevene tilbakemeldinger underveis i læringsprosessen (Brevik & Blikstad-Balas, 2014, s. 3). *Formativ vurdering* er en type VfL, som sier noe om funksjonen vurderingen har; den innhentede informasjonen brukes for å tilpasse læringen til elevens behov (Brevik & Blikstad-Balas, 2014, s. 2). Dette prinsippet er også etablert i LK20 og er et viktig grep for å tilpasse opplæringen og øke kompetansen i fag (Kunnskapsdepartementet, 2017).

Diagnostiske oppgaver er en type formativ vurdering. Det primære formålet er å kartlegge hva elevene *ikke* kan, og gi en indikasjon på hva denne tilkortkommenheten skyldes – oppgavene skal *diagnostisere* (Brekke, 2002). Denne informasjon blir lærerens utgangspunkt når disse misoppfatningene skal overvinnes. Brekke (2002, s. 16) sier at denne typen vurderinger gjerne gis før elevene har hatt noe undervisning i det gjeldende temaet, og at de derfor ofte blir brukt til å:

1. identifisere og framheve misoppfatninger som elevene har utviklet, også uten at det trenger å ha vært noen formell undervisning i det en vil undersøke
2. gi læreren informasjon om løsningsstrategier elevene bruker for ulike typer av oppgave

Disse punktene kan gjenkjennes som å bygge på det som National Research Counsil (2001, s. 116) betegner som *konseptuell forståelse*, *prosedural flyt* og *strategisk kompetanse*. Disse ferdighetene er noe av det som kjennetegner matematisk dyktighet (National Research Counsil, 2001). Intensjonene med å identifisere dette er at denne informasjonen vil gjøre det lettere for læreren å tilrettelegge undervisningen, slik at misoppfatninger overvinnes, og løsningsstrategier utvikles (Brekke, 2002). På denne måten ser man sammenhengen mellom diagnostiske oppgaver og vurdering for læring. Etter at læreren har latt den innhentede informasjonen være med å forme undervisningen, vil resultatene kunne danne et sammenligningsgrunnlag om oppgavene gjøres på ny. Således er en annen fasett ved slike oppgaver at de kan gi innsikt i den faglige utviklingen om elevene vurderes både før og etter undervisningssekvensen (Brekke, 2002, s. 16). Som med de fleste andre vurderinger er oppgavenes validitet viktig å sikre om man vil ha diagnostisk informasjon. For diagnostiske oppgaver innebærer dette at resultatene tydelig skiller mellom misoppfatninger og tilfeldig feil som enhver elev kunne gjort (Brekke, 2002). Som nevnt overfor er misoppfatninger en konsekvens av feilaktig tenkning, og er derfor ganske konsekvente, i motsetning til tilfeldige feil (Brekke, 2002). Diagnostiske oppgaver må da klare å belyse disse tankesettene for å gi reliabel informasjon.

Utformingen av diagnostiske oppgaver kan også være opphav til mulige utfordringer. Smith (2009, s. 25) sier at motivasjon bygger på interesse, som igjen forutsetter en vilje til å lære. Denne læreviljen avhenger av man føler at man kan få noe positivt ut av å engasjere seg, uten at ens selvaksept står i fare (Smith, 2009, s. 25). Hver gang man står overfor en ny oppgave gjøres en slik vurdering basert på tidligere erfaringer med slike oppgaver. Elever utvikler ofte en forsvarsmekanisme hvor de ikke prøver sitt beste, for å bevare selvaksepten. Elever som har dårlige erfaringer med vurderingssituasjoner vil da kunne holde igjen, og ikke la svarene deres reflektere tankene, dersom oppgavene og oppsettet virker kjent – og avskrekkende. Om dette er tilfellet for en diagnostisk oppgave har det en negativ innvirkning på oppgavens indrevaliditet: oppgaven gir ikke informasjon om elevens misoppfatninger eller løsningsstrategier.

2.2.2 Bruk av diagnostiske oppgaver i sannsynlighet

Forskning viser at elever som har gjennomgått undervisning som bygger på misoppfatninger har prestert bedre på sannsynlighetstester enn dem som gjennomgår «vanlig» undervisning (Batanero & Sanchez, 2005, s. 262). Samtidig mener Hirsch og O'Donnell (2001) at lærere har for dårlige verktøy til å vurdere forståelse i sannsynlighet, og derfor overestimerer elevenes kompetanse. De sier videre at rene avkryssningsoppgaver blir for kvantitative til å være gode måleverktøy, og bør suppleres med mer kvalitative spørsmål som fanger opp *hvorfor* elevene svarer slik de gjør. Ved å tilføre en slik kvalitativ komponent ser man at mange av de elevene som krysser av for riktig alternativ, gjør dette på bakgrunn av misoppfatninger (Hirsch & O'Donnell, 2001). Hirsch og O'Donnell (2001) beregnet den gjennomsnittlige skåren uten oppklaringsspørsmålet til å bli 80 %, men med oppklaringen ville skåren ha sunket til 61 %. Dette viser at man kan ende opp med å svare riktig, uten å ha en rett oppfatning av et konsept. Dette peker også på behovet for å diagnostisere misoppfatninger. Dersom misoppfatninger utgjør en så stor andel av elevers svar kan det stilles spørsmålstegn ved selve hensikten med sannsynlighetsundervisningen. Det er tvilsomt om elever får et grunnlag for å ta gode valg og tenke kritisk dersom de ikke forstår konsept, men bare svarer riktig fordi de kjenner igjen utformingen.

Det har blitt forsket mye på elevers misoppfatninger innen sannsynlighet, og i det øyemed har det blitt utviklet spesifikke diagnostiske oppgaver for å kartlegge misoppfatninger (Batanero & Sanchez, 2005; Jones & Thornton, 2005). Piaget og Inhelder (Green, 1986) og Tversky og Kahneman (2013) var tidlig ute på dette feltet, og banet vei for etterfølgende forskning knyttet til misoppfatninger. I videre forskning har det primært blitt forsket på én-og-én misoppfatning, og det er hovedsakelig metastudier som sammenlignet flere misoppfatnigner (Green, 1986; Shaughnessy, 1977; Ang & Shahrill, 2014; Öçal, 2018; Batanero & Sanchez, 2005). Flere av disse diagnostiske oppgavene er utformet som skriftlige avkryssningsoppgaver, hvor hvert alternativ representerer en misoppfatning. I sammenheng med disse undersøkelsene trekkes det også frem en nødvendighet av å ha et kvalitativt element som supplerer undersøkelsen, enten ved intervju eller som skriflig svar. De undersøkelsene som ikke inkluderer slike element fremmer noen utfordringer ved dette. En utfordring er at det er vanskelig å konkludere med akkurat *hvilken* misoppfatning eleven har (Ang & Shahrill, 2014, s. 28). En annen utfordring er at det er vanskeligere å avgjøre om noen av dem som har krysset av et alternativ tilhørende en misoppfatning gjør dette fordi de har en misoppfatning, eller om det er en tilfeldig målefeil (Green, 1986, s. 291). I et klasserom er det også mulighet for at forskjellige elever sitter med forskjellige misoppfatninger. Behovet for å få en mer helhetlig oversikt over hvilke misoppfatninger som gjør seg gjeldende melder seg. Å bruke tester som utelukkende fanger opp én misoppfatning vil da kunne bli tidkrevende. Som Brekke (2002) poengterer har diagnostisike oppgaver sitt formål i å innhente informasjon om hvor eleven kommer til kort i sine matematiske ferdigheter. Dette peker på at dersom diagnostiske oppgaver skal tilgjengeliggjøres i klasserommet og dras nytte av i sannsynlighetsundervisning, er det behov for nytenkning rundt utformingen.

2.2.3 Seriøse spill og vurdering

Den didaktiske bruken av spill i klasserommet, og tilhørende forskning, har hatt en fremmarsj de siste tiårene (Farber, 2015, s. 22). Man ser at spill evner å motivere og stimulere ferdigheter hos elever som andre læringsaktiviteter ikke klarer i like stor grad (Apt, 1970, s. 13; McGonigal, 2010). Mye av den forskningen som blir gjort knytter seg til læring med *seriøse spill*, sett i en digital kontekst (Egenfeldt-Nielsen, Meyer & Sørensen, 2011; Farber, 2015). Denne formen for læring vil bli drøftet videre i avsnitt 3.3. Det som skal nevnes i denne seksjonen er at det meste av forskning har et fremadskuende perspektiv, hvor spillet skal fremme ny læring. Hvordan spill kan brukes på å kartlegge hva eleven allerede kan blir ikke drøftet i stor grad. I spill får ofte spilleren tilbakemeldinger på prestasjoner for å komme videre i spillet (Farber, 2015, s. 33). Dette kan forstås som en form for formativ tilnærming kan være spillets formativt vurderingsaspekt. Dette aspektet har blitt forsket på, men det blir mest knyttet opp mot digitale platformer, og settes sjelden sammen med analoge brettspill.

Forskning viser mot at det er mye jobb å utforme et spill med diagnostiske egenskaper. Noe av motivasjonen min i denne oppgaven var å undersøke om det ville være lettere å bruke et mer håndfast verktøy til dette, og at det på denne måten ville være enklere å ta i bruk i klasserommet.

3 Teori

Dette kapittelet vil videre aktualisere tidligere forskning og ta for seg og drøfte relevante teorier som danner bakteppe for metodens utforming og videre analyse av resultater. Første del av dette kapittelet vil knytte seg til å definere ulike misoppfatninger innenfor sannsynlighet og se på deres relevans i undervisningssammenheng. Deretter vil jeg gi en teoretisk ramme for å forstå spillets plass i klasserommet. Spilldesignet som ble brukt i undersøkelsen vil så bli presentert og tolket ut fra denne rammen.

3.1 Forventet at elevene kan

For å knytte misoppfatningene til skolen er det nyttig å se på de konkrete kompetansemålene for ungdomsskoletrinnet. I læreplanen som gjelder for årets 10. klassinger (Utdanningsdirektoratet, 2013, s. 9) sier kompetansemålene at elevene skal kunne:

- finne og diskutere sannsyn gjennom eksperimentering, simulering og berekning i daglegdagse samanhengar og spel
- beskrive utfallsrom og uttrykkje sannsyn som brøk, prosent og desimaltal
- drøfte og løyse enkle kombinatoriske problem

For å diskutere sannsynlighet i dagligdagse sammenhenger er det essensielt at man har en korrekt forståelse av hva sannsynlighet sier: hva betyr det at sannsynligheten er 75 %? En forståelse av dette må bygge på en at man forstår prinsippet bak *De store talls lov*, og at en tolkning av sannsynlighet må rotfestes i det store perspektivet, ikke øyeblikksbildet. Dette prinsippet møter derfor elever ofte når de skal lære sannsynlighet.

Å beskrive et utfallsrom krever ofte at man har kjennskaper til *kombinatoriske* teknikker som hjelper en å telle opp mulige og gunstige utfall, som er utgangspunktet for å kunne uttrykke noe som en brøk. I denne prosessen vil ulike oppgaveformuleringer legge føringer for hvilke utfall som er aktuelle å vurdere, man sier at utfallsrommet blir *betinget*. Elever må forstå hvilke implikasjoner dette har for kombinatorikken, og derfor forstå betydningen av begrepene *ordning* og *tilbakelegg* om de skal klare å løse kombinatoriske problem, beskrive utfallsrom, og fine sannsynligheter. Disse målene blir i stor grad videreført i komptansemålene for 9. klasse (Utdanningsdirektoratet, 2019a, s. 13). Man kan slik se at punktene spiller på hverandre, og eventuelle misoppfatninger kan bunne i flere av disse punktene. Dette vil sees mer på i den kommende delkapittelet.

3.2 Misoppfatninger

I kapittel 2, Tidligere forskning, kom det frem at det finnes mange ulike misoppfatninger innenfor sannsynlighet. Den ikke-åpenbare fremtoningen, sammen med lite kunnskap om misoppfatninger, gjør at lærere har vansker med å kartlegge dem, og følgelig undervise ut fra dem. Kompetanser innenfor sannsynlighet henger som sagt sammen, noe som gjør at misoppfatningene og deres fremtoning ofte er kompleks. Misoppfatningene i dette delkapittelet blir presentert fordi de har mye teoretisk forankring i tidligere forskning, og fordi de har en klar tilknytning til kompetansemålene i skolen. De tre største seksjonene tar for seg misoppfatninger knyttet til Kombinatorikk, Heuristikker og Betinget sannsynlighet.

3.2.1 Kombinatorikk

Kombinatorikken innehar en sentral rolle når det kommer til sannsynlighetsregning (Batanero & Sanchez, 2005, s. 242). Fra tidlig skolealder blir elever lært at man regner sannsynlighet ved å finne antall mulige og ønskede utfallene i en hendelse, og så se på hvor stor del av det mulige hendelsesrommet de ønskede opptar. For uniforme sannsynlighetsmodeller stemmer dette, og den påfølgende utfordringen er hvordan man skal klare å finne kardinaliteten til utfallsmengdene. Kombinatorikken er den delen av fagfeltet som er dedisert oppgaven med bestemme størrelsen på disse mengdene. Kombinatorikken gir en metode å finne antall mulige utfall, og antall måter et bestemt utfall kan opptre på (Wikipedia, 2019). Det er ikke mange konkrete misoppfatninger av kombinatorikk, men mange elever synes å ha vanskeligheter med metodikken, og gjør mange feil. Disse feilene påvirker videre hvordan elever forstår sannsynlighet (Shaughnessy, 1977, s. 295). Shaughnessy (1977, s. 295) sier at noen av misoppfatningene elevene har i sannsynlighet direkte skyldes at de undervurderer kombinatorisk vekst. Resten av denne seksjonen vil derfor ta for seg noen av disse feilene, slik at det vil være lettere å se å koble kombinatorikken til misoppfatninger i neste seksjon.

Før vi går videre må to begrep innføres; *kombinasjoner* og *permutasjoner*. Disse begrepene kan forstås på flere måter, noe som i seg selv kan være en kilde til misforståelse. En måte å forstå det på, som ligger nært elevenes kompetansemål avsnitt 3.1, er ved å relatere dem til ordning. Kombinasjoner er ekvivalenten til et uordnet utvalg, slik at utfallene {2,1} og {1,2} er bare én kombinasjon. Likeledes er permutasjoner ekvivalenten et ordnet utvalg, slik at utfallene {2,1} og {1,2} er to permutasjoner, av den samme kombinasjonen. Den andre faktoren som er med å definere utvalget er tilbakelegg: om elementene som velges blant kan trekkes på ny. Følgelig kan man ha fire typer utvalg; man kan ha kombinasjoner med og uten tilbakelegging, og permutasjoner med og uten tilbakelegging. Dersom utvalget består av n elementer og man skal trekke k ganger, vil de fire utvalgene kunne uttrykkes algebraisk som i tabell 1.

	Permutasjoner	Kombinasjoner
Med tilbakelegg	*Seleksjons-modellen* n^k	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$
Uten tilbakelegg	*Partisjons-modellen* ${}_nP_k = \frac{n!}{(n-k)!}$	*Distribusjons-modellen* ${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

Tabell 1: Kombinatoriske utvalg

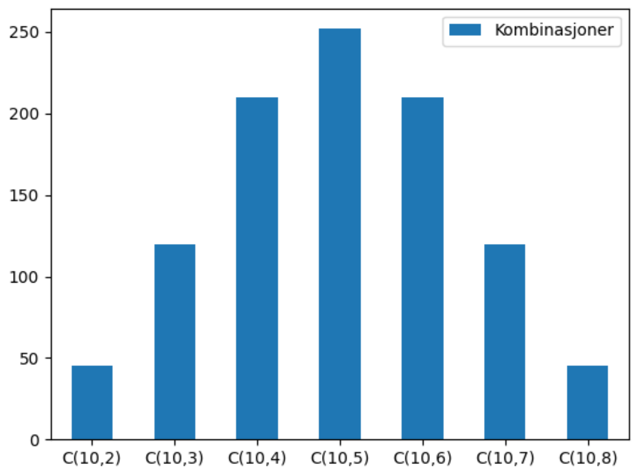
Dubois (Batanero & Sanchez, 2005, s. 242) navngav tre av utvalgene som vist i tabell 1, og betegner disse som enkle kombinatoriske konfigurasjoner. Det er disse tre som undervises på skolen. En studie gjennomført på 700 elever viste at elevene opplevde at oppgaver tilknyttet disse tre modellene var av ulik vanskelighetsgrad (Batanero & Sanchez, 2005, s. 244). Seleksjonsmodellen ble oppfattet som lettest av de tre, og elevene fant raskt en kombinatorisk formel for å svare på oppgavene. Oppgaver tilhørende de andre to modellene var mye vanskeligere å bearbeide, og i det hele tatt få gitt en matematisk representasjon. Partisjonsmodellen og distribusjonsmodellen henger sammen, ved at ${}_nP_k = {}_nC_k \cdot k!$, som man kan forstå som at når man har et utvalg uten tilbakelegging, vil hver kombinasjon gi $k!$ permutasjoner. Man kan forstå det som at det er en surjeksjon fra permutasjonene til kombinasjonene. Det virker som at denne sammenhengen henger tett sammen med de vanskene elevene har med kombinatorikken; de klarer ikke å skille disse to modellen. Restultatene viser noen av de vanligste feilene elevene gjør (Batanero & Sanchez, 2005, ss. 244-245):

- Ordningsfeil: glemmer hvorvidt ordningen har noe å si for spørsmålet, og ender opp med feil modell, eller en slags mellomting.
- Tilbakeleggsfeil: glemmer å ta hensyn til tilbakelegging, slik at de ikke ser på repetisjon som en mulighet, og da mister noen permutasjoner.
- Objektsfeil: misforstår om et element er unikt eller ei, slik at man ender opp med feil modell.
- Usystematisk telling: ramser opp usystematisk, og slik mister noen ordninger, eller teller noen flere ganger.
- Feilbruk av diagram

3.2.2 Heuristikker

Heuristikk forstås i denne sammenhengen som den underliggende metodikken personer bruker når de skal anslå hvor hyppig en hendelse inntreffer (Kahneman, 2013, s. 143).
En type heuristikk som gjør seg gjeldende innenfor sannsynlighet er *tilgjengelighetsheuristikk*. Kahneman (2013, s. 144) lagde begrepet for å forklare tendensen mennesker har til å ta ikke-statistiske beslutninger i situasjoner hvor man må fatte et valg. Det sentrale ordet her er *tilgjengelighet*, som omhandler hvor tilgjengelige egne eksempler er. Kahneman (2013, s. 143) definerer selv tilgjengelighetsheuristikken som «prosessen med å bedømme hyppighet ut fra hvor 'hvor lett man kommer på tilfeller'». Denne prosessen kan forklares ved at «hjernen» sliter med å gjøre statistiske beregninger for å besvare et spørsmål, og velger derfor å modifisere spørsmålet slik at det kan besvares utfra tilgjengelige erfaringer og hvor lette disse er å komme på. Kahneman (2013, s. 458) sier: «Tilgjengelighet er en fin ledetråd i vurdering av hyppighet og sannsynlighet, fordi tilfellet fra store kategorier som regel er lettere og raskere å komme på enn eksempler fra mindre hyppige kategorier». Når man sammenligner kardinaliteten til mengder vil dette kunne gi et effektivt estimat; blir man spurt om det er flere ord som begynner på q enn på s, vil nok de fleste svare nei, fordi det er lettere å komme på ord som begynner på s. I dette tilfellet vil tilgjengelighetsheuristikken bidra til et, trolig, hurtig og korrekt svar. Det som er problematisk er at *tilgjengelighetsskjevheter* gjør at man ikke alltid kommer på eksempler fordi de tilhører den største kategorier (Kahneman, 2013, s. 144). Personlige erfaringer kan være med å prege svaret; man vurderer risikoen for hjerteproblemer som stor dersom det finnes hjerteproblemer i nær omgangskrets – det er veldig tilgjengelig (Kahneman, 2013, s. 458).

At noe er *vanskeligere å se for seg* gjør at man minimerer denne kategorien (Kahneman, 2013, s. 459). Si at vi har 10 personer, og k av disse personene skal plukkes ut for å lage en komité, hvor $2 \leqslant k \leqslant 8$. Vi ønsker så å finne ut hvor mange personer komiteen skal bestå av, hva k må være, for at man skal ha flest muligheter til å lage en unik komité. I dette eksempelet er det tenkt at det ikke er noen dedikerte roller i komiteen. Da kan dette forstås som et uordnet utvalg uten tilbakelegging fordi i) personenes rekkefølge i komiteen er irrelevant, og ii) samme person



Figur 1: Kombinasjoner for ulik gruppestørrelse

ikke kan velges flere ganger til komiteen. Binominalkoeffisienten, $C(n,k) = \frac{n!}{k!(n-k)!}$, forteller oss akkurat dette; hvor mange kombinasjoner som kan lages når man plukker k fra n. Med $n = 10$, og k som definert tidligere blir fordelingen som i figur 1, som viser at man får flest forskjellige komiteer om man velger ut 5 personer, $k = 5$, som gir 252 mulige sammensetninger. Når intervjuobjekt har blitt bedt om å estimere hva k må være for å gi flest unike komiteer, viser resultatet at de fleste tror at k må være minst mulig (Kahneman, 2013, s. 460). Det er lettere å se for seg ulike komiteer med få personer enn med mange, fordi man da får noe overlapp (Shaughnessy, 1977, s. 297). Derfor tror man at den mengden må være større. I forsøket som ble gjort ble mediananslaget for komiteer med 2 personer mer enn tre ganger så høyt som for komiteer med 8 personer, hhv. 70 og 20. Dette er bemerkelsesverdig når $C(10,2) = C(10,8) = 45$. I seg selv kan man argumentere for at tilgjengelighetsheuristikk ikke er en misoppfatning, fordi den ikke er konsekvent feilaktig tanke, men snarere en kognitiv prosess. Grunnen til at den har blitt inkludert her likevel er at de ukorrekte tankene kan ha sitt opphav i en feilaktig tankeprosess, slik som tilgjengelighet (Kahneman, 2013, s. 150).

En annen type heuristikk som har blitt mye omtalt i litteraturen er *representativitet*, som er en misoppfatning som konsekvent har vist seg å være en hindring for elevers konseptuelle forståelse (Hirsch & O'Donnell, 2001). Hvis man har en populasjon B av mulige hendelser, og skal vurdere sannsynligheten for at et utvalg $A \in B$ skal forekomme, vurderes dette på bakgrunn av hvor representativ A er for B, dvs. hvor lik A er B (Kahneman, 2013, s. 450; Konold, 1991, s. 296). Denne likheten trenger ikke gjelde for hele mengdene, men noen aspekter ved dem (Batanero & Sanchez, 2005, s. 247). Kahneman (2013, s. 450) sier at denne tilnærmingen til sannsynlighetsestimering «fører til alvorlige feil, fordi likhet eller representativitet ikke påvirkes av andre faktorer som *burde* [min utheving] virke inn på sannsynlighetsbedømmingen». En faktor som i aller høyeste grad bør bli vektlagt er *utvalgsstørrelsen* (Kahneman, 2013, s. 452). Den sier hvor stor del av populasjonen som er med i utvalget. Man ser at elever ofte ofte neglesjerer denne faktoren (Batanero & Sanchez, 2005, s. 248). Bagatellisering av utvalgsstørrelsen følges ofte av misoppfatninger i sannsynlighet, fordi størrelsen på utfallsrommet samtidig bagatelliseres (Hirsch & O'Donnell, 2001). De neste avsnittene vil ta for seg noen av mest skoleaktuelle representativitets-misoppfatningene.

3.2.2.1 *De Små Talls lov*

Som nevnt i avsnitt 3.1 er det naturlig at elever lærer om Store Talls lov og sentralgrensesetningen, de skal ha en forståelse av hvordan et aritmetisk gjennomsnitt tilnærmes når utvalgsstørrelsen økes. Det kan fort overgeneraliseres til en misoppfatning: man tenker at gjennomsnittet skal være tilnærmet, uavhengig av utvalgsstørrelsen. Man kan si at utvalget skal være representativt for populasjonen, for å klargjøre hvorfor det er en del av representativitetsheuristikkene. Denne misoppfatningen blir beskrevet som at man er *ufølsom for utvalgsstørrelsen*, og det har blitt sagt at man da følger *De Små Talls lov*, som er en humoristisk vri på at det nettopp er de store talls lov som misoppfattes (Kahneman, 2013, s. 125). Personer som lider av denne misoppfatningen vil ha vansker med å forklare ekstreme tilfeller av statistiske avvik, og vil være tilbøyelig til å gi disse avvikene ikke-matematiske forklaringer. I små tettsteder vil man f.eks. finne ekstremitetene av både sykdomsrelaterte plager, og visse politiske oppslutninger. Om man tenker at det statistisk korrekte er at alle utvalg skal være representative for populasjonen, vil man kunne ende opp med å forklare slike avvik med en kausal fortolkning: folk som stemmer rødt har lettere for å få liktær! Dette kan man i tillegg gi en fiffig analyse av og si: «folk som stemmer rødt er mer miljøpositive, og går mer uten barføtt», som i beste fall er en forhastet slutning. Det som *er* tilfellet er at når man vurderer et lite utvalg vil det være mye mer sannsynlig å få et ekstremt utfall av enheter enn når man har et stort utvalg (Kahneman, 2013, s. 122). Dette er rent matematisk, og handler om hvordan permutasjoner ordner seg. Et eksempel på dette kan være hvorden fødsler fordeler seg:

> Et lite og et stort sykehus har henholdsvis 10 og 100 fødsler en dag. *Er det like sannsynlig at 80 % av de fødte er gutter på det lille som på det store?*
>
> For å svare på dette bruker vi formelen for binomisk fordeling for en stokastisk variabel $X =$ (gutter født), med n barn, k gutter, og med sannsynlighet $p = 0,5$ for gutt: $p_X(k) = \binom{n}{k} \cdot 0,5^k \cdot 0,5^{n-k}$.
>
> $$n = 10, k = 8 \Rightarrow p_X(8) = \binom{10}{8} \cdot 0,5^{10} \approx 0,044$$
> $$n = 100, k = 80 \Rightarrow p_X(80) = \binom{100}{80} \cdot 0,5^{100} \approx 4,2 \cdot 10^{-10}$$
>
> Her ser vi at det er ca. 100 millioner ganger mer sannsynlig å se et slikt utfall på det lille sykehuset enn det store. En representativ bedømming vil si at siden forholdstallet mellom andelen fødte og andelen gutter er lik på begge sykehusene – begge utfallene er like representative – vil sannsynligheten også være lik.

3.2.2.2 *Tilfeldighet*

Elevers evne til å skjelne mellom tilfeldige og deterministiske fenomen preger måten de forstår sannsynlighet på, spesielt når de skal vurdere sannsynligheten for forskjellige hendelser (Batanero & Sanchez, 2005, s. 245). Piaget og Inhelder var noen av de første som forsket på oppfattelsen av tilfeldighet, innenfor utviklingspsykologien, og *The Origin of the Idea of Chance in Children* har inspirert forskere i ettertid (Jones & Thornton, 2005, s. 65; Batanero & Sanchez, 2005, s. 245). Piaget og Inhelder (Green, 1986, s. 288) hevdet at ingenting er mer vanlig enn den tilfeldige fordelingen av regndråper som starter en regnskur, og brukte dette til å kartlegge elevers oppfattelse av tilfeldig distribusjon. Elevene skulle se på 16 taksteiner hvor 16 regndråper hadde fordelt seg, og skulle avgjøre hvilken fordeling som var mest sannsynlig, altså best representerte en tilfeldig fordeling. De kunne velge mellom tre ulike fordelinger: en helt jevn fordeling hvor hver takstein blir truffet på samme punkt; en semi-tilfeldig fordeling hvor hver takstein blir truffet, men i ulike punkter; og en helt tilfeldig fordeling hvor noen taksteiner blir truffet flere ganger med ulik dråpespredning, og noen ikke blir truffet i det hele tatt. Piaget og Inhelder (Jones & Thornton, 2005, s. 67) klassifiserte tre nivå for den forståelsen knyttet til tilfeldighet, hvor man først på det øverste kunne håndtere sannsynlighet. De hevdet også at mange av elevene var på dette nivået, og hadde en god nok forståelse av kombinatorisk argumentasjon til å kunne forstå sannsynlighet (Batanero & Sanchez, 2005, s. 246). På de lavere nivåene så man flere misoppfatninger hos elevene. Disse resultatetene har blitt bestridt i etterkant, og Green (1986, s. 291) fant at elevprestasjonene i stor grad avhang av oppgavenes utforming, ofte i en negativ forstand i forhold til Piaget og Inhelders funn. Selv om de fleste klarte å skjelne mellom den tilfeldige og den jevne fordelingen var forskjellen på den semi-tilfeldige og den tilfeldig spredning ikke like åpenbar.

Teorien om representativitet belyser hvordan personer har misoppfatninger knyttet til tilfeldige hendelsers fordeling. «Folk forventer at en serie av tilfeldige hendelser skal være representative for en tilfeldig prosess selv når sekvensen er kort» (Kahneman, 2013, s. 453). Hirsch og O'Donnell (2001, 5. avsnitt) eksemplifiserer dette: «De fleste mennesker tenker at når man kaster mynt og krone, vil en sekvens med bare mynt {M, M, M, M, M, M} være mindre sannsynlig enn en blandet sekvens som {K, M, M, K, M, K} fordi sekvensen med bare mynter ikke er representativ for fordelingen av hendelser. En tilsynelatende tilfeldig sekvens med mynt og krone, som den ovenstående, blir ofte ansett representativ fordi denne sekvensen ligner mer på den teoretiske fordeling med 50/50 mynt og krone.» [min oversettelse] Så sant man ser på ordenede utvalg vil disse to sekvensene være like sannsynlige; $1/2^6 \approx 1,6\ \%$. Om serien skal oppfattes som sannsynlig må den være representativ for selve prosessen (Batanero & Sanchez, 2005, ss. 246-247). Elever oppfatter en serie som representativ først når den hverken er for lik den teoretiske fordelingen eller for ulikt. Elever med representativitetsheuristikk vil heller ikke tenke at det er sannsynlig at ulike mønster, som f.eks. rekker av det samme utfallet eller et gjentagende mønster, fremkommer i tilfeldige fenomen. Men det er nettopp disse utfallene man kan vente seg fra en tilfeldig serie; »en serie som er representativ lokalt, avviker systematisk fra hva vi kan forvente av tilfeldigheter: Den inneholder for mange vekslinger og for få rekker med samme utfall.» (Kahneman, 2013, s. 454). Dette gir også en ny innfallsvinkel for å tolke elevenes forståelse av Piaget og Inhelders regndråper. Hverken den jevne eller den tilfeldige fordelingen er representativ. Den jevne er for ordnet til at den oppfattes som representativ for en naturlig fordeling. Den tilfeldige fordelingen etterlater seg tørre taksteiner, noe som heller ikke er representativt for erfaringene man har av at regn gjør alt vått.

At folk forstår tilfeldighet slik kan også sees sammen med tilgjegelighetsheuristikk og utvalgets ordning. I et ordnet utvalg vil utfallene {M, M, M, M, M, M} og {K, M, M, K, M, K} være to kombinasjoner av i alt 64 mulige kombinasjoner med en uniform sannsynlighetsfordeling. I et uordnet utvalg endres utfallsrommet, og utfallene {6M, 0K} og {3M, 3K} er to permutasjoner av i alt syv mulige utfall/kombinasjoner. Siden de syv kombinasjonene ikke har like mange permutasjoner har man nå en ikke-uniform sannsynlighetsfordeling, og kombinasjonene som har derfor forskjellige sannsynligheter, respektive $\approx 1,6$ % og $\approx 31,3$ %. En slik tankegang vil kunne være veldig tilgjengelig om man har spilt mye terningspill, slik at man bruker det som begrunnelse for å si at det blandede utfallet er mer sannsnynlig. Dette fremmer igjen viktigheten av å være, og gjøre elever, bevisst på hvilke matematiske konsekvenser språklig differering medfører. Abrahamson (2012) viser at en virkningsfull måte å adressere denne misoppfatningen er ved å la elevene selv oppdage denne distinksjonen gjennom utforskende undervisning.

Hasardspellerens feilslutning (Gambler's fallacy) er en side ved en representativ oppfatning av tilfeldighet hvor man tenker at en serie med like utfall vil ha en innvirkning på sannsynligheten til det påfølgende utfallet (Shaughnessy, 1977, s. 295; Kahneman, 2013, s. 454). Hvis man kastet mynt og har fått en lang rekke med krone tenker man: «siden det har kommet så mange kroner, er det mer sannsynlig at det kommer en mynt nå». Dette kan forstås som mer sannsynlig, fordi det gjør at sekvensen er mer representativ. Om man har erfaringer med å regne på dette vil man også kunne støtte denne intuisjonen med matematikk. Dersom man har kastet fem kron, og skal til å kaste den sjette, vil man kunne tenke at sannsynligheten for at den sjette også blir kron kan uttrykkes $P(6kron) = (\frac{1}{2})^6 = 1/64 \approx 0,012$. Slik kan det sees om lite sannsynlig å få den seks kron på rad. Det som her er feil er at når man allerede *vet* utfallet av de tidligere kastene er deres sannsynlighet lik 1, og man får da $P(6kron) = 1^5 \cdot \frac{1}{2} = \frac{1}{2}$: Sannsynligheten er 0,5, uanfektet av de forrige kastene. «Tilfeldighet betraktes ofte som en selvkorrigerende prosess der et avvik den ene veien utløser et avvik den andre veien for å gjenopprette likevekt. Men i virkeligheten blir ikke avvik 'korrigert' underveis i en tilfeldig prosess, de får bare mindre verdi» (Kahneman, 2013, s. 454). Derfor kan vi fremdeles stå fast ved at de tidligere hendelsene ikke påvirker det neste, og samtidig akseptere Store Talls lov.
Hasardspellerens feilslutning omfatter også tanken om at hendelsen nå er *mer* tilbøyelig for utfallet som har blitt gjentatt: det er *mer* sannsynlig at det utfallet inntreffer på ny. «På grunn

av at det har kommet så mange kroner allerede vil det nok komme en krone nå også.» Dette gir også tolkningsgrunnlag for å si at uttryk som «å være i sonen», «he's on a roll» og «never change a winning team», beror på statistiske misoppfatninger. Denne delen av misoppfatningen faller da utenom den representative heuristikken og hører nok mer til under tilgjengelighetsheuristikken, men er like fullt en misoppfatning.

3.2.2.3 Ekviprobabilitets bias

Ekviprobabilitets bias er en heuristisk misoppfatning hvor man har en tendens til å anta at enhver tilfeldig prosess må være uniform, og tenker følgelig at alle utfall har lik (ekvi) sannsynlighet (Gauvrit & Morsanyi, 2014). Dette kan settes i en kontekst av terningkast, når man skal vurdere de to hendelsene «én femmer og én sekser» og «to seksere». Den første hendelsen fremkommer av to forskjellige utfall, $\{(5, 6), (6, 5)\}$, og den andre hendelsen gis av ett unikt utfall, $(6, 6)$. Det betyr at den første hendelsen er dobbelt så sannsynlig som den andre. Likevel vil flertallet tolke dem som like sannsynlige (Batanero & Sanchez, 2005, s. 248). Det gjør seg spesielt gjeldende når man kan anta redskapene som hendelsene bygger på er uniforme, eksempelvis terninger, myntkast, barnefødsler, etc. (Gauvrit & Morsanyi, 2014). En annen tolkning ekviprobabilitets biasen er at man vurderer snittet av to utfall som like sannsynlig som de to utfallene for seg selv. Kahneman (2013, s. 177) gjennomførte et omfattende forsøk hvor de presenterte studenter for Linda, en fiktiv kvinne med et sett egenskaper. Studentene ble bedt om å vurdere hvorvidt det var mer sannsynlig at Linda var en bankkasserer eller en bankkasserer som også er aktiv i feministbevegelsen. En svimlende andel av de statistikktrente studentene, 85 %-89 %, svarte at det var mer sannsynlig at Linda var *både* kasserer *og* feministisk, noe som åpenbart ikke kan rettferdiggjøres utfra matematisk logikk.

3.2.3 Betinget sannsynlighet

«Sannsynlighet er et felt med en betydelig mengde paradokser, hvorav mange hviler på misforståelser om betinget sannsynlighet.» [min oversettelse] (Batanero & Sanchez, 2005, s. 254). Betinget sannsynlighet har en sentral rolle i prosessen med å samhandle med en usikker verden (Falk, 1986, s. 292). «Blant annet er utforsking av elevers misoppfatinger rundt betinget sannsynlighet viktig fordi det viser nødvendigheten av at løsninger må bygge på de forhåndsbestemte kriteriene. Derfor må elever tenke på betingelsene, før de kan gå videre med påfølgende ressonering og vurdering» [min oversettelse] (Fischbein & Schnarch, 1997, sitert i Öçal 2018, s. 88). Som sett i avsnitt 3.1 er på betinget sannsynlighet fastsatt i elevenes kompetansemål. Sannsynligheten for at en hendelse A inntreffer, gitt at hendelse B inntreffer er definert som $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$. Dette betegnes som at hendelse A er betinget av hendelse B. Denne definisjonen er i og for seg enkel, men det bli likevel funnet feilslutninger og misoppfatninger når man undersøker elevenes forståelse (Falk, 1986, s. 292). Disse vanskene kan i stor grad skyldes på en tungtvint terminologi (Batanero & Sanchez, 2005, s. 251). Noen av de mest omtalte misoppfatningene omtales i de følgende avsnittene.

3.2.3.1 *Misoppfatning av tidsaksen*

Denne misoppfatningen knytter seg til måten man resonnerer rundt hendelser knyttet til tidsaksen. Falk (1986, sitert i Batanero og Sanchez 2005, s. 251) fant at en vanlig, og feilaktig, oppfatning er at den betingende hendelsen B må opptre *før* den betingede hendelsen A. Oppgaven som illustrerte dette poenget gikk ut på at man har plassert fire kuler, to hvite og to sorte, i en urne. Det skal trekkes to kuler, uten tilbakelegging, hvor denotasjonen S_1 betyr at *første trekk er en sort kule*. Elevene ble spurt to spørsmål: a) hva er $P(S_2 \mid S_1)$, og b) hva er $P(S_1 \mid S_2)$. De fleste oppfatter at svaret i a) er $1/3$, fordi her opptrer trekkene i en kronologisk rekkefølge. I b) er det verre. Her mener mange at det ikke gir mening å betinge en hendelse med noe som inntreffer senere; det er ikke gyldig (Falk, 1986, s. 292). Mange svarer derfor at sannsynligheten må være $1/2$, siden på det tidspunktet S_1 ble trukket, var det to av hver kule i urnen. Falk (1986, s. 242) sier: «indeed the first ball doesn't care whether the second is white or black but – we do». Misoppfatningen bunner i at man tenker at det er en kausal sammenheng mellom de to hendelsene A og B; A blir forårsaket av B. Med en slik kausal resonnering gir det lite mening om A da kommer før B. Men dette er en feilaktig resonnering; det er ikke noen kausalitet. Selv om det ikke er noen kausal relasjon fra den andre hendelsen på den første, vil *informasjonen* om at 2. kule i utfallet er sort, redusere utfallsrommet for den første trekningen. Ved hjelp av et diagram, er det greit å se at informasjonen begrenser utfallsrommet til at det bare er seks utfall hvor den andre kule er sort, hvorav to av disse har en sort kule som første trekk; $P(S_1 \mid S_2) = \frac{1}{3} \Big(= P(S_2 \mid S_1) \Big)$. Et logisk lavterskel-argument mot denne kausale resonneringen kan dras fra hverdagslige, ikke-kronologiske, eksempler, som at nye arkeologiske funn gir informasjon om fortiden. Man kan også ta utgangspunkt i et sykdsomsforløp (Falk, 1986, s. 243): Sykdom forårsaker symptomer, sykdommen kommer før symptomene, men den diagnostiske fremgangsmåten estimerer sannsynligheten for sykdom, gitt symptomene. Ofte når en snakker om sannsynlighetens plass i medisinsk diagnostikk er det brukes Bayes sitt teorem for å gjøre utregninger og illustrere poeng som *falsk positiv*, og *type I feil*. Om elevene ikke forstår denne reversible komponenten av betinget sannsynlighet, vil det også være vanskelig å forstå denne fortolkningen av Bayes sitt teorem (Batanero & Sanchez, 2005, s. 253; Falk, 1986, s. 295). Falk (1986, s. 295) sier at dette medfører en mangelfull forståelse og en overdreven språklig tilnærming som fort kan bli roten til misoppfatninger.

En nærliggende tematikk til misoppfatning av tidsaksen er å skjelne mellom *synkrone* og *diakroniske* hendelser (Batanero & Sanchez, 2005, s. 253). En vanlig forståelse er at to synkrone hendelser skjer samtidig; man trekker en kule fra to forskjellige urner hvor det i den ene er én sort og én hvit kule, og det i den andre er to sorte og tre hvite kuler. To diakrone hendelser skjer etter hverandre; man har én urne med to hvite og én sort kule, og trekker først én kule, legger den tilbake, og trekker på ny. Det som er tilfellet er at disse to hendelsene er ekvivalente, siden det andre tilfellet kan forstås som at man har to like urner, 2H og 1S, og trekker samtidig (Batanero & Sanchez, 2005, s. 253). Hvis man ikke oppdager dette vil man kunne ende opp med å betinge en synkron hendelse som om den var diakronisk.
Man ser at *konteksten* har noe å si for hvordan elever oppfatter oppgavene. Tekstoppgaver som krever mer tolkning gjør det vanskeligere å oppfatte den matematiske strukturen i oppgavene (Batanero & Sanchez, 2005, s. 249). Også dette kan være en faktor som gjør det vanskelig å skjelne mellom diakroniske og synkrone hendelser. Man kan oppfatte det som mer sannsynlig å tre 5-ere på ett kast med tre terninger, enn å få tre 5-ere på rad med én terning, selv om dette, som i eksempelet over, er den samme hendelsen.

3.2.3.2 *Problem med å definere den betingende hendelsen*

For mange ligger vanskeligheten i å avgjøre *hvordan man skal betinge*; på hvilken måte skal utfallsrommet begrenses? (Batanero & Sanchez, 2005, s. 254) Dette illustreres kanskje best ved et eksempel.

> Man har tre identiske ekser, og i hver eske ligger to kuler. I én eske ligger det to blå kuler, i én ligger det to hvite kuler, og i den siste ligger det én blå og én hvit kule. Man åpner en av eskene og trekker en tilfeldig kule. Hvis man antar at kulen som trekkes er blå, hva er sannsynligheten for at den gjenværende kulen også er blå? $\Big($ Hva er $P(B_2 \mid B_1)$? $\Big)$

Her vil de fleste elever tenke at svaret er 1/2. Dette baseres på en antagelse om at esken med de to hvite kulene kan ignoreres når en blå er trukket. Da er det bare to esker igjen, hvorav én inneholdt to blå kuler; 50/50. Feilen her ligger i at man betinger på feil element; man betinger på eskene, når det er kulene som er den betingende faktoren (Falk, 1986, s. 294). Vi kan liste opp alle utfallene for å tydeliggjøre dette: Hakeparentes markerer hver pose, parentes tilsvarer utfall, indeksen teller utfall, den hevede skriften er en numerering av sjetongene, og det venstre elementet i hver parentes er den første sjetongen som blir trukket i hvert utfall. Utfallsmengden blir da: $\Big\{\Big[(B^1, B^2)_1, (B^2, B^1)_2\Big], \Big[(B^3, H^1)_3, (H^1, B^3)_4\Big], \Big[(H^2, H^3)_5, (H^3, H^2)_6\Big]\Big\}$. Her blir det tydelig at en betingning på pose kun vil fjerne utfall 5 og 6, mens å betinge på kuler gjør at man nøster vekk det éne av utfallene i blandingsposen også, utfall 4. Da står man igjen med at tre utfall, hvorav to gir oss to blå sjetonger; $P(B_2 \mid B_1) = 2/3$.

Dette kan også kobles til det kjente *Monty Hall problemet*. Nøkkelen til å forstå hvorfor man alltid bør bytte dør ligger i at man forstår at det ikke er dørene som betinger utfallene, men programlederen og hans dørvalg som betinger utfallene. Det er riktig at utfallene som involverer to hvite sjetoner faller bort, men det er også enda ett utfall som går bort. «What matters is not only what we know but also how do we know it» (Falk, 1986, s. 294). Lærdommen her er at sannsynligheten til en hendelse skal betinges av den umiddelbare hendelsen som er gitt som referanse i problemet, ikke på fortolket konsekvens av referansehendelsen.

3.2.4 Utfallstilnærming og ikke-rasjonell tro på ekstern variabel

Utfallstilnærming og *tro på ekstern variabel* er to misoppfatninger som bare så vidt skal nevnes her. Disse vil kunne opptre i sammenheng med de andre misoppfatningene drøftet i dette kapittelet.

Utfallstilnærming handler om hvordan man tolker en sannsynlighet. Man vil her ha et lokalt perspektiv hvor man vurderer sannsynligheten for at et utfall skal inntreffe én gang, og ikke som en global fordeling av utfall (Batanero & Sanchez, 2005, s. 249). Derfor tolker en kun sannsynligheter utfra hvor nære de er 0 %, 50 % og 100 % (Konold, 1991, s. 146). Er de nærmest 0 % vurderes hendelsen som *umulig*, er den nær 100 % er den *uungåelig*, og er den nær 50 % oppfattes de som helt tilfeldige. (Batanero & Sanchez, 2005, s. 249). Denne forståelsen kan knyttes opp til spørsmålets konteksten, nevnt i avsnitt 3.2.3.1. «Det har derfor blitt sett at spørsmålet 'Hva er den ordentlige tolkningen av sannsynlighet?' kan være villedende, og at det derfor kan være fordelaktig å endre det til 'Hva er de ordentlige tolkningene av sannsynlighetsutsagn i den-og-den konteksten?' »[min oversettelse] (Rowbottom, 2015, ss. 8-9). Spillet som undersøkelsen bygger på kan forstås å bygge på nettopp utfallstilnærming. I spillet blir elevene bedt om å estimere sannsynligheten for et utfall, og så vedde på utfallet. Om man da estimerer $p = 0,6$ eller $p = 0,8$, vil det ha samme implikasjon for hvordan man vedder, dersom man vedder stokastisk.

Troen på en ekstern variabel handler om at man tror at sannsynligheten påvirkes av ting som *flaks*, *gud*, eller *menneskelige attributter* (Ang & Shahrill, 2014, s. 27). Dette kan knyttes til heuristikker. Representativitet gjør at man ikke anser det som sannsynlig å få en serie med det samme utfallet, og tilgjengelighet kobler sammen den forrige erfarte serien, med at man f.eks. blåste på terningene eller ba til Gud.

3.3 Spill i skolen

Spill har en lang historie med å være informasjonsbærer og lære kommende generasjoner tro, kultur og teorier, som militære stratgier (Farber, 2015, s. 11). Denne rollene som informasjonsbærer har også aktualisert bruken spill i skolen (Farber, 2015, ss. 21-22). Til stadighet er det fremdeles er det bare et «eksotisk krydder», brukt av vågale lærere, og ikke en integrert del av den formelle opplæringen (Felicia & Egenfeld-Nielsen, 2011, s. 8). Clark Abt innførte begrepet *seriøse spill* i 1970 i boken sin med samme tittel (Farber, 2015, s. 17). Apt (1970, s. 13) sa at spill er «effektive redskap for læring elever i alle aldre og i mange situasjoner fordi de er høyst motiverende, og fordi de kommuniserer mange tematiske fakta og konsept effektivt.» [min oversettelse]. Apt (1970) definerte så seriøse spill som «spill som har et eksplisitt og nøye gjennomtenkt pedagogisk formål, og er ikke intendert å bli spilt primært for fornøyelsens skyld.» Det har blitt gjort mange forsøk på å utvikle spill med denne hensikten, og gjennom disse forsøkene har man sett at det er lettere sagt enn gjort (Farber, 2015, s. 10). Utfordringen ligger i at det bør oppleves autentisk og spennede, samtidig som det fremmer lærering. Skolen har en historie med kjedelige, ikke-autentiske spill, det Farber (2015, s. 16) kaller «sjokoladetrukket brokkoli», som hverken er motiverende eller kommuniserer konsepter. Konkret har slik brokkoli ofte en tendens til å være forsøk på å gjøre puggeoppgaver om til spill; drep en zombie med matematikk, eller pugg spørsmål og oppgi riktig svar for å flytte brikken videre. En læringssituasjon som vil oppleves mer autentisk vil være å ta utgangspunkt i *Angry Birds*, og for eksempel drøfte fuglenes parabelbanener som funksjoner av variabler som sprettert-spenn og tyngde. «When you can grab a student's interest at some level and they can build confidence, it's easier to extend and try something else.» (Farber, 2015, s. 22) For å få en bedre forståelse av spill i skolen vil jeg se på disse to delene av seriøse spill som Abt løfter frem – motivasjon og kommunikasjon – og knytte dem begge til de pedagogiske formålene.

3.3.1 Selvaksept og motivasjon

En av fordelene med læring knyttet til spill er at spillerne ofte har mer selvtillit og motivasjon i en spill-setting enn de har til vanlig i klasserommet (Felicia & Egenfeld-Nielsen, 2011, s. 22). McGonigal (2010) er en amerikansk spillutvikler som har forsket på positive effekter ved bruken av dataspill. I dataspill gis spillerne store oppdrag og delmål, men disse oppfattes som oppnåelige. Ofte gis spilleren en spesifikk rolle og ansvar i begynnelsen, tilpasset deres nivå og ferdigheter. Dette er med på å bygge elevers tro på egne evner; de kan få til det umulige (McGonigal, 2010). Tanken om at man kan klare å overkomme store hindringer er viktig i dagens samfunn. Det finnes store utfordringer, både lokalt og globalt, som elever skal bidra til å løse (Kunnskapsdepartementet, 2017, Tverrfaglige temaer). For å løse store problemer må

man ha en tro på at disse kan løses, og at man selv har en rolle i den løsningen. Det ble nevnt i avsnitt 2.2.1 at selvfølelsen spiller er en stor rolle for motivasjon og lærelyst (Smith, 2009, s. 25). Elevers erfaring med dataspill, og sine egne muligheter og roller i spill, vil kunne gjøre at de møter temaer med høyere selvfølelse og på derfor er mer motiverte for å lære. I spill kan man også prøve så mange ganger man vil; det gjør ingenting om man trenger åtte forsøk på å klare level 5 i Super Mario , det viktigste er at man prøver på ny, og tar *lærdom* fra de tidligere forsøkene (Thomas, 2018). Dette er også en faktor som er med å ufarliggjøre læringen med spill: det er rom for å mislykkes. Man ikke oppleve en like direkte trussel mot selvbildet, dersom selvbildet ikke avhenger av hvordan man presterer – man kan bare prøve på ny!
Når man spiller online samarbeider vanligvis flere spillere for å nå målene, og hver spillers rolle er viktig for å klare oppgaven (McGonigal, 2010). Slik ser man at spill bygger elevers samarbeidsevne; man får læringsfellesskap, fordi man sammen må lære for å komme videre (Sørensen, 2011). Det er rett og slett slik at folk liker å løse problemer sammen (Farber, 2015, s. 35). Om spill klarer å skape samarbeid vil det også være en motiverende faktor.

Siden undervisningsopplegget i denne oppgaven bygger på et analogt spill vil det være effekter som ikke like godt kommer frem. I det aktuelle spillet gis f.eks. ikke elevene noen spesifikk rolle tilpasset deres egenskaper. Det er likevel relevant å tenke at elevers erfaringer med spill vil kunne ha en innvirkning på deres lærevilje. Affeksjonsheuristikk er en heuristikk hvor man tolker hendelsen opp mot følelsene den vekker (Kahneman, 2013, s. 153). Da kan elevenes motivasjon bygge på spørsmålet «liker jeg spill?», fremfor «liker jeg matematikk». I tillegg er både samarbeid og muligheten til å lære av feil faktorer som gjør seg gjeldende i spilldesignet. Motivasjonen vil ikke være en faktor som drøftes noe mer heretter, fordi det ikke ble gjort noe forsøk på måle denne effekten av undervisningsopplegget. Det skal for øvrig nevnes at de deltagende elevene uttrykte stort engasjement i form av høylytte gledesutbrudd og ønsker om å være med å delta i forsøkene.

3.3.2 Læring

Emeritus ved Center for Undervisningsudvikling og Digitale Medier, Jessen (2011, s. 164), sier at man bør styre unna motivasjonsaspektet om man skal få nye perspektiv på læringen; det ikke er spillet i seg selv som er interessant, men læringspotensialet det har. En bør sette spørsmålstegn ved *hvordan* spillere lærer når de spiller. En måte å forstå dette på er ved å se på begrepene *formell* og *uformell* læring (Sørensen, 2011, s. 107). Formell læring er vanligvis knyttet til skolen hvor læring er målet, mens uformelt er ofte knyttet til fritid, hvor man lærer for å oppnå målet; man lærer seg en spesiell kode for å endre ens bloggs estetikk, eller lærer et engelsk ord for å forstå en sangtekst, et filmsitat, eller et oppdrag i *Battlefield*. Uformell læring finner man når pedagogikk og underholdning er velbalansert, hvor denne optimale balansen medfører at man nesten ikke er klar over at man lærer (Farber, 2015, s. 110). Dette gjør seg spesielt gjeldende når læringssituasjonen oppleves autentisk, fordi det presenteres læringssituasjoner hvor spilleren har et *ønske* om å lære for å komme seg videre i den fremmete situasjonen (Farber, 2015, s.110). Det er slike autentiske situasjoner spill kan tilføre undervisningen. Det er derimot ikke gitt at disse situasjonene fremmer læring av seg selv, og læreren er derfor en viktig brikke i spillbasert læring i klasserommet. Forskning viser at lærere fort blir passive når spill introduseres i klasserom (Sørensen, 2011, s.117; Jessen, 2011, s. 147). Lærerne later til å ha problemer med å finne sin rolle i spillklasserommet, og ender ofte opp med å klargjøre rammene, og deretter tre inn i en passiv observatør-rolle. Dette kan være fordi lærere har for stor tiltro til spills undervisningsevner. Farber (2015, s. 35) forklarer samspillet mellom læreren og spillet slik: «Games are great at creating teachable moment. Games are terrible at knowing they have successfully delivered the teacheable moment and then filling in the right information at the right time. But instructors are great at that!» Spill kan derfor ikke erstatte læreren, men kan være et godt verktøy for læreren, og bør brukes sammen med andre pedagogiske grep som et didaktisk verktøy (Felicia & Egenfeld-Nielsen, 2011, s. 30). Slik kan lærere adaptere prinsippene ved uformell læring til sin egne undervisning.

Didaktikken i et seriøst spill kommer frem gjennom spillets utforming og hvordan det anvendes (Jessen, 2011). Utformingen har vært, som tidligere nevnt, en hindring for læring, fordi balansen mellom pedagogiske perspektivene og underholdningaspektet ikke har vært gagnelig. Om man bommer på dette tyder det på at man ikke forstått hvorfor spill har andre læringspotensialer en andre verktøy (Farber, 2015, s. 31). På samme måte som at man kan gjøre en regneoppgave uten å lære av den, kan man også *gjøre* et spill uten å spille det, og dermed ikke lære av det som planlagt. Det har blitt sagt at en vanlig feil folk gjør er at de prøver å tilføye moro til en disiplin; man bør heller finne ut hva som er morsomt med en disiplin, og så finne måter å lære dette bort (Farber, 2015, s. 89). Det kreves en mer møysommelig utforming enn at man slår sammen et vilkårlig spill og en matematisk disiplin for å få en didaktisk tungvekter (Farber, 2015, s. 33). Hunicke, Leblanc og Zubek (2004) har laget MDE-rammeverket for utvikling av spill, og de presenterer *mekanikk*, *dynamikk* og *estetikk* som de tre hovedkomponentene når et spill skal utvikles. Spillets estetikk er de inntrykkene og den oppfatningen man ønsker at spilleren skal ha av spillet; mekanikken er reglene som sier hva man kan gjøre, og dynamikken er broen mellom mekanikk og estetikk, implementeringen av mekanikken slik at spillet oppfattes slik man vil (Hunicke et al., 2004; Farber, 2015, s. 33). Som lærer er kanskje estetikken den komponenten som er mest nærliggende læringsmålet, og mekanikken og dynamikken må fremme den. Om man vil at spille skal fremme samarbeid (estetikk), må det lønne seg å samarabeide fremfor å motarbeide hverandre (dynamikk), selv om begge deler er lov og fullt mulig å gjøre (mekanikk). Slike prinsipp kan brukes av læreren for å uttrykke det fengende ved en disiplin gjennom et spill. Når didaktikken først kommer til uttrykk i spillet, vil det også være lettere å anvende det i klasserommet. Klasseromsimplementereingen forutsetter læreren som den kapable partneren for å bygge på de didaktiske punktene som spillet fremmer, og slik støtte elevenes læring i den videre undervisningen.

3.4 Borel

Undervisningsopplegget i denne masteravhandlingen bygger på et spill kalt *Borel*, oppkalt etter den franske matematikeren og statistikeren Émile Borel. I spillet må spillerne forholde seg til ulike statistiske eksperiment med uniform eller binomisk fordeling. Spillet bruker terninger, mynter og urne-modeller i eksperimentene. Hvert eksperiment har et tilhørende kort, hvor eksperimentets utfall er formulert som et ja/nei-spørsmål, som: «Kommer du til å trille en sekser på neste terningkast?» Spillets mekanikk er ellers ganske ukomplisert:

1. Spillerne får en kapital på 4000 som man skal vedde for. Spilleren kan kun vedde 800, 300 eller 100 om gangen, og kan maks vedde halvparten av kapitalen.
2. Man får utdelt et «Ja»-kort og et «Nei»-kort.
3. En av spillets påstander blir lest opp.
4. Spillerene skal vedde på hvilket utfall eksperimentet får, utfra spørsmålsformuleringen på kortet.
   Spillerne vedder ved å velge et Ja- eller Nei-kort, som legges med forsiden ned, og så velge seg et av de tre beløpene.
   Spillerne avslører samtidig hva de veddet.
5. Eksperimentet gjennomføres.
6. Spillerne får gevinst som legges til kapitalen ettersom de veddet; oddsen som gis er 1:1.
7. Spilleren med størst kapital etter siste forsøk vinner.

Grunnen til at nettopp Borel ble valgt var fordi det ble oppfattet som å ha et stort læringspotensiale. Siden dette er et autentisks spill, som enhver mann kan kjøpe uten å ha formål om didaktsik bruk, trenger det nødvendigvis ikke å regnes som et seriøst spill, etter Abts definisjon. Definisjonen peker på spillets pedagogiske formål, og jeg mener derfor at når spillet brukes i skolen som et læringsverktøy, vil det kunne forstås som et seriøst spill. Masterveilderens empiri tilsier at spillet er en god ressurs for uformell læring, noe som også peker på egnetheten for bruk i klasserommet. Mye av denne læringen har skjedd ved at engasjerte og faglige diskusjoner blir fasiliteret gjennom spillet. Dette har vært en indikator på at spillets estetikk er *følelsesmessig engasjerende* og er faglig *utfordrende*, siden det ikke er innlysende hvordan man bør vedde. Disse to estetikkene er blant flere beskrevet av Hunicke et al. (2004). Tidspress er en dynamisk faktor som gjør spillet mer utfordrende (Hunicke et al., 2004). Når man skal plassere veddemålet har man en tidsbegrensning, og det vil derfor være vanskelig å regne nøyaktig på alle sannsynlighetene. Veddemålene vil derfor bygge på både matematiske ferdigheter og sannsynlighets-intuisjon. Spillerne kan oppleve dette som en utjevnende faktor: man trenger ikke være flink i sannsynlighet for å vinne, man trenger bare en god intuisjon og litt flaks! Spillets dynamikk balanserer ferdighet og flaks på denne måten. Et rent ferdighetsspill, som sjakk, krever at spillerne er på likt nivå om det skal bli spennende; et spill som utelukkende bygger på flaks, som tvungen Yatzy, neglesjerer spillernes ferdigheter helt. I Borel vil de matematisk dyktige spillerne ha en fordel med å bedømme utfalls sannsynlighet, men en slik beregning er ingen garanti for at eksperimentet faktisk får det utfallet. Man kan spørre seg om det vil kunne virke mot sin hensikt at man ikke får full uttelling for egne ferdigheter, men i en undervisningssituasjon tror jeg at dette elementet bidrar til mer spenning, og i seg selv illustrer sannsynlighetens natur.

Spillets kan også justeres med flere dynamikker, og slik få andre estetikker. Ved at en samarbeider i par eller grupper vil spillet oppleves å *bygge fellesskap* (Hunicke et al., 2004). At samarbeidsevnen bygges er som sagt en av de positive attributtene som spill støtter. I Borell kommer samarbeidet frem når man *sammen* skal bestemme seg for et veddemål. I gruppene vil man drøfte oppfatninger av eksperimentene og de mulige utfallene før man plasserer veddemålet. Selv om en lener seg på intuisjon, vil man bli mer bevisst på egne prediksjoner, og veddemålene vil i større grad gjengi en bevisst tanke, ikke bare et instinkt. Mekanikken er derimot ikke helt tilpasset å fremme samarbeid. Farber (2015, s. 41) sier at spill hvor det kåres en vinner egner seg dårligere som et samarbeids-spill. Elevene vil kunne oppleve det som viktigere å vinne enn å lære. Helt konkret kan det bety at elevene kun bryr seg om utfallet, uten å ønske å reflektere rundt hvorfor man fikk det utfallet, og om det var sannsynlig. Dette peker nok en gang på lærerens funksjon i spill-klasserommet, om spillet faktisk skal fremme læring.

Spillet kan også gis andre estetikker. Det vil være nærliggende å gi spillet et *narrativ* ved å for eksempel sammenligne veddemålene med å gjøre aksjeinvesteringer. Dette vil kunne gjøre at spillet oppfattes som mer relevant og autentisk. Elevene vil da kunne engasjere seg på en annen måte, og oppleve det som viktigere å forvalte kapitalen ansvarlig, og slik bli mer bevisste på å ta *fornuftige* valg.

Man hadde allerede sett at spillet var en god ressurs for læring, og ideen for masteravhandlingen var at man kunne videreutvikle bruken av spillet for å kartlegge misoppfatninger. Antagelsen var at man kunne bruke innsatsen i veddemålene som et mål på elevers oppfatninger av sannsynlighet. Dersom man kunne se avjgøre hva som var en fornuftig innsats, vil man også kunne vurdere hva som motiverer den ufornuftig innsatsen. Dermed kan man knytte ulike eksperiment til ulike misoppfatninger, basert på hvordan elevene vedder. Neste seksjon vil derfor drøfte hva som gjør en innsats fornuftig.

3.4.1 Hvordan bør man vedde?

Kelly-kriteriet er et kritierie som kan brukes til å forstå hvordan man objektivt bør plassere et veddemål. Kelly-kriteriet uttrykkes ved en formel, hvor innsatsen avhenger av spillets odds, utfallets sannsynlighet og egen kapital (Wikipedia, 2021). For et veddemål med odds 1:1, kapital W, med sannsynlighet p, sier Kelly-kriteriet at den optimale innsatsen er gitt:

$$I_k(p, W) = (2 \cdot p - 1) \cdot W \tag{1}$$

I Borel er alltid oddsen 1:1, og man kan derfor bruke likning (1) slik den står for å beregne hva som er fornuftig å vedde. For det første eksperimentet er $W = 4000$, og for et eksperiment med $p = 5/6$ får man: $I_k(5/6, 4000) = (2 \cdot 5/6 - 1) \cdot 4000 \approx 2667$. Man bør altså gå inn med 2667, men siden det er en øvre grense i Borel på 800 må man nøye seg med det. Det er også verdt å merke seg at likning (1) sier at dersom $p = 0,5$ bør man vedde 0. Algebraisk kan man se at det også er lett å utlede andre formler med praktiske tolkninger og bruksområder. Siden det i Borel er diskrete innsatser kan det eksempelvis være relevant, utfra egen kapital, å se på hvor stor sannsynligheten må være før det er fornuftig å gå inn med ulike innsatser: $p(I_k, W) = \frac{1}{2}(I_k/W + 1)$. Kelly-kriteriet er kjent som den *vitenskapelige vedde-metoden* og blir også brukt på aksjemarkedet dersom visse kriterier er oppfylt. Dette kan derfor være en fin inngang til et aksje-narrativ, som overnevnt. Kelly-kriteriet utledes ved at man maksimerer forventningsverdien av logaritmen av kapitalen, og vil derfor, i det lange løp, gi størst avkastning (Wikipedia, 2021). Dersom man alltid vedder på det utfallet med størst sannsynlighet, sier forventningsverdien at man vil vinne på sikt. Det kreves derimot at man har nok kapital til å kunne holde ut i lengden og ikke går konkurs tidlig i løpet; man må ta høyde for at den stokastiske variansen kan påføre noen tap. Ideen var at dette skulle bli målestokken for å dømme innsatser; elevenes innsats skulle tilsi hvor sannsynlige utfallet var, og sett opp mot Kelly-kriteriet ville det vurderes som fornuftig eller ikke.

En utfordring var at elevenes vedding kunne også bygge på mindre matematiske betraktninger, slik som *spillteori* og *risikoadferd*, og være situasjonsbetinget og subjektivt fornuftige. Spillteori handler om hvilke utfall man får når to eller flere «rasjonelle» personer spiller sammen, ut fra de matematiske og sosiale overveiningene disse personene gjør seg i spillet (Farber, 2015, s. 40). I Borel skal man kåre en vinner, og da vil spillteori kunne ha en effekt på veddingen, slik at den ikke reflekterer sannsynlighetsoppfatningen. En rasjonell person som ligger på andre plass vil kanskje vedde på det motsatte av det han tror at førsteplassen vedder, i håp om å ta førsteplassen. Førsteplassen kan også vedde mindre enn det hun tenker er det mest logiske, fordi hun da ikke står i fare for å tape fullt så mye, og på den måten sikrer førsteplassen. Å vedde slik vil gå imot Kelly-kriteriet, men vil i den situasjonen kunne være mer fornuftig for å sikre seieren.

Noen av de første til å beskrive egenskaper ved en rasjonell persons beslutninger var von Neumann–Morgenstern, men man har sett at disse normative beskrivelsene sjelden gjør seg gjeldende i reelle situasjoner (Kahneman, 2013, s. 475). Personers adferd overfor risiko har i liten grad matematisk forankring, og bygger i stor grad på subjektive oppfatninger, slik som hvilke konnotasjoner situasjonens formuleringen har, og generelt hvorvidt det oppleves som et mulig tap eller gevinst (Kahneman, 2013, ss. 471-492). Dette er kanskje forventet utfra hva man har sett påvirke elevers sannsynlighetsoppfatninger. En elev vil derfor ha kunnet ressonert seg frem til en korrekt sannsynlighet og regnet ut forventet gevinst, men ha holdt tilbake på innsatsen fordi det mulige tapet oppleves som større en den mulige gevinsten.

Begge disse to faktorene kan ha spilt inn på de innsamlede dataene. Det ble derimot antatt at de fleste elevene ville bedømme det som mer gunstig å satse på det de tenkte var *mest sannsynlig* jevnt over, og at effekten av spillteoretiske motivasjoner ikke ville ha stor påvirkning på helheten. Siden det ikke var snakk om ekte penger, og spillet ikke hadde noen videre konsekvens, ble det også antatt at det ikke var en stor overvekt av taps- eller gevinstfølelse, og at ulike elevers risikoadferd ville veie opp for hverandre. I de tilfellene hvor risikosøking og risikoaversjon er gjeldende ble det også antatt at elevenes innsats vil reflektere *hva* elevene tror er mest sannsynlig, selv om det ikke fanger opp *hvor sannsynlig* de tror det er.

4 Metode

Denne studien intervjuet 107 elever gjennom en deduktiv spørreundersøkelse for å undersøke om måleverktøyet kunne fange opp misoppfatnigner. Dette kapittelet vil først redegjøre for undersøkelsens forskningsdesign ved å presentere utvalget, spørsmål og gjøre rede for analyseverktøy. Deretter vil undersøkelsen sees i lys av rådende forskningsnormer.

4.1 Forskningsdesign

Forskningensdesignet har som mål å besvare forskningsspørsmålet, introdusert i avsnitt 1.3, på en effektiv måte (Furuseth & Everett, 2012, s. 128, Silverman, 2011, s. 42). Forskningsdesignet kunne blitt utformet på mange forskjellige måter. Man kan ikke av prinsipp si at én metode er bedre enn en annen, det avhenger både av forskningsspørsmålet som skal besvares, og pragmatikk rundt forskerens ressurser (Furuseth & Everett, 2012, ss. 128-129). For denne oppgaven måtte derfor metoden undersøke hvordan Borel fungerer som ressurs for å gi informasjon om elevers misoppfatninger innenfor sannsynlighet. Som redegjort for tidligere er dette et felt hvor det er begrenset med forskning fra før. Masteravhandlingen er derfor et forsøk på utfylle tidligere forskning ved å slå sammen bruken av seriøse spill i klasserommet, og diagnostikk av misoppfatninger i sannsynlighet (Furuseth & Everett, 2012, s. 130). Et annet aspekt ved forskningsspørsmålet er at Borel ikke bare skal undersøkes som et måleverktøy, men det skal vurderes som en ressurs for kartlegging *i klasserommet*. Om dette skal være tilfellet bør den samme metoden som kartlegger Borels diagnostiske egnethet også være en adapterbar metode som kan implementeres i et annet klasserom for diagnostisering. Denne faktoren var også med å forme forskningsmetoden.

4.1.1 Utvalg

Undersøkelsens utvalg bestod av elever på 10.-trinn, primært aldersgruppen 15 til 16 år, og alle elevene gikk på samme ungdomsskole. Det var i alt 107 respondenter, noe som tilsvarer i overkant av 90 % av hele årskullet. Skolen lå sentrumsnært i en storby, i et område som ansees å ha høy sosioøkonomisk standard. Alle elevene hadde arbeidet med sannsynlighet og temaene fremmet i avsnitt 3.1 i løpet av de siste månedene, og hadde derfor et likt utgangspunkt for å kunne løse oppgavene. Det ble vurdert som mest hensiktsmessig at elever utgjorde enhetene i utvalget. Ved å bruke elevene som respondenter kunne jeg forsøke å måle *deres* oppfattelse av ulike stokastiske forsøk. En annen mulighet kunne vært å bruke lærere som respondenter, og for eksempel intervjue dem om hvordan de opplever at spillet kartlegger elevenes misoppfatninger. Ved å bruke lærere som utvalgsenheter ville de blitt gitt en funksjon som fortolkende sekundærkilder av elevenes misoppfatninger (Furuseth & Everett, 2012, s. 133). En åpenbar svakhet ved et slikt utvalg ville vært at mange lærere, som nevnt i avsnitt 2.1.1, har dårlig oversikt over elevenes misoppfatninger, og deres feilaktige fortolkninger ville kunne påvirket oppgavens indrevaliditet, og oppgaven samplet derfor elever.

Piloteringen ble gjennomført ved en videregående skole i samme geografiske område som den faktiske datainnsamlingen. Koronarestriksjonene begrenset derimot mulighetene og tilgangen på en videre datainnsamling ved den skolen, og for så vidt andre videregående skoler i distriktet. Innhentingen ble derfor gjort på en annet trinn enn først intendert. Siden funn viser at forekomst av sannsynlighetsmisoppfatninger er ganske lik på de ulike årstrinnene ble de konkludert med at undersøkelsen kunne gjennomføres på en annet årstrinn i stedet. Det var ikke forventet at dette skulle ha en stor effekt på utvalget. De eneste endringene som ble antatt å måtte gjøres knyttet seg til oppgavesettet, og hvilke kunnskaper elevene var forventet å ha i forkant.

4.1.2 Forskningsmetode

For å svare på forskningsspørsmålet valgte jeg å bruke en kvantitativ metodikk. Kvantitativ metodikk kan beskrives som å «(...) forklare et fenomen ved å samle numeriske data som blir analysert ved matematisk baserte metoder»[min oversettelse] (Aliaga & Gunderson i Jopling, 2019, s. 56). Det var flere motiv for å velge en slik metodikk. For det første tok forskningen min utgangspunkt i alleredeeksisterende begrep og teorier, noe som ga en mer deduktiv tilnærming til undersøkelsen (Larsen, 2017, s. 24). Ved å bruke spillet kunne elevenes oppfatninger forsøkes og tallfestes, og så måles utfra den eksisterende teorien. Forskningsspørsmålet fremmet også det som kan kalles en *deskriptiv undersøkelse*, fordi den søkte å *beskrive* ulike styrker og svakheter ved spillet som måleverktøy, ikke finne noen årsaksforklaringer (Larsen, 2017, s. 22). Larsen (2017, s. 26) sier at ved deskriptive undersøkelser er det naturlig å se til kvantitative metoder, fordi man da måler effekter. Konkret for min undersøkelse betød det også at flere elever nås med undervisningsopplegget, og utvalget kunne være både større og bredere (Larsen, 2017, s. 28). Store utvalg er med på å sikre at undersøkelsen ikke lider av tilfeldige målefeil, som har en innvirkning på forskningens indre validitet (Solbakken, 2019, s. 35). Dette vil drøftes mer i de kommende seksjonene.

Innenfor kvantitativ metodikk finnes det flere ulike metoder man kan velge for å innhente informasjon. En mye brukt metode er strukturerte spørreundersøkelser, som ofte kalles surveyundersøkelser (Grønmo, 2004, s. 191; Silverman, 2011, s. 43). Jopling (2019, s. 59) sier at surveyer er gode verktøy for å fange opp karakteristikker ved en populasjon, for eksempel oppfatninger og adferder, fordi de søker å vise korrelasjon og ikke kausalitet. Slik kunne denne metoden være et godt verktøy for å fange opp misoppfatninger, uten å forsøke å begrunne hvorfor respondentene har slike. Det er mange måter å utforme surveyen på, og det kreves nøye bearbeiding av spørsmålene for å være sikker på at de måler det de er tiltenkt å måle (Jopling, 2019, s. 59; Grønmo, 2004, s. 193). Å klare å utforme surveyen slik at samspillet mellom den og undervisningsopplegget ble funksjonelt var utfordrende, fordi jeg ønsket at surveyen skulle forstås som en naturlig del av undervisningsopplegget. Dersom undervisningsopplegget skulle ha en integrerbar diagnostisk effekt i et klasserom burde vurderingsverktøyet, surveyen, fremstå som en del i oppleggets helhet. Surveyen fikk derfor det som kan kalles et noe ukonvensjonell utforming.

Oppgave 1: Petter har 3 kuler i en boks. 2 av dem er røde, og 1 er grønn. Uten å se trekker Petter opp en av kulene fra boksen. Hva er sannsynligheten for at han trekker den grønne?

Svar:_________________

Oppgave 2:

a) En PIN-kode på en mobil består av 4 tall som er fra 0-9. Hvor mange ulike kombinasjoner av PIN-koder finnes det?

   Svar:_______________

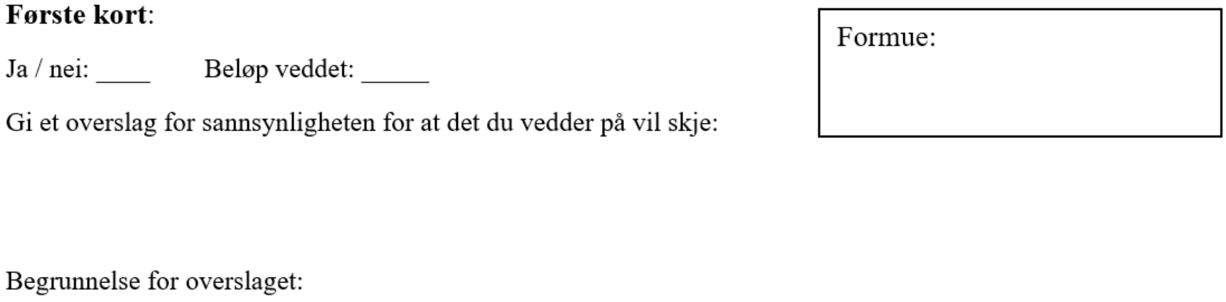
b) Tora har glemt PIN-koden til mobilen sin. Hun har 3 forsøk på å gjette koden. Hva er sannsynligheten for at hun klarer å gjette riktig kode?

   Svar:_______________

Figur 2: Del 1

Surveyen bestod av to deler som elevene fikk utdelt fysisk på papir. Del 1, vist i figur 2, inneholdt tre matematiske kartleggingsspørsmål. Disse spørsmålene skulle brukes for å kunne vurdere om elevens feilsvar var knyttet til manglende matematiske ferdigheter og ikke misoppfatninger. Oppgave 1) og 2b) knyttet seg til sannsynlighet. Oppgave 2a) var et kombinatorisk spørsmål knyttet til seleksjonsmodellen, som skulle brukes for å beregne utfallsrommet i flere av de kombinatoriske eksperimentene. Alle disse oppgavene knyttes tydelig til kompetansemålene, nevnt i avsnitt 3.1, og oppgaveneformuleringene ble vurdert som gjenkjennelige for elevene. Elever som ikke klarte å svare på disse spørsmålene ble i analysen tolket som å lene seg på intuisjonen når de plasserte innsatsene.

Del 2 bestod av ti sammensatte spørsmål eller batteri med spørsmål (Grønmo, 2004, s. 195). Hvert batteri korresponderte til et av eksperimentene som skulle gjennomføres. Et eksempel på et slikt batteri kan sees i figur 3.



Første kort:

Ja / nei: ____ Beløp veddet: _____

Gi et overslag for sannsynligheten for at det du vedder på vil skje:

Formue:

Begrunnelse for overslaget:

Figur 3: Del 2, med spørsmålsgruppe 1

Disse batteriene var den delen av surveyen som brukte spillet til å generere data om elevenes misoppfatninger. Hvert batteri bestod av fire spørsmål som elevene skulle svare på; «Formue»-rubrikken var bare et hjelpemiddel for å støtte elevenes kontroll på egen kapitalen, slik at de kunne vedde i henhold til spillets regler, og hadde ingen videre effekt på dataene eller analysen. For at det skal være et poeng å bruke surveyer som kvantitative verktøy forutsettes det også at dataene som genereres kan tallfestes (Solbakken, 2019, s. 14). Det ble derfor nødvendig å spørsmålene verdier viss *målenivå* kunne rangeres (Kleven, 2014, s. 31). Elevenes innsats var en kombianasjon av *Ja/Nei*- feltet og *Beløp veddet*-feltet. Ja/nei er en dikotom variabel med kun to mulige verdier; i undersøkelsen ble Ja gitt verdien «1» og Nei verdien «-1» (Solbakken, 2019, ss. 20, 51). Beløpet er en variabel med tre forskjellige verdier: 800, 300 og 100. Kombinasjonen av disse to variablene produserte selve *innsatsvariabelen*, som ble brukt som et uttrykk for elevers sannsynlighetsoppfatning, og ble følgelige gitt seks mulige verdier: -800, -300, -100, 100, 300, 800. Sett opp mot Kelly-kriteriet vil det være mulig å rangere disse verdiene som at noe var mer fornuftig enn noe annet. Som nevnt tidligere bygde denne initielle ideen om bruk på at elevenes innsats representerte hvor sannsynlig de oppfattet noe å være. I lys av piloteringen ble det klart at dette var en dårlig antagelse. Innenfor spillets rammer ble Kelly-kriteriet et for lite nyansert verktøy; generelt for en sannsynlighet p slik at $|p - 0,5| > 0,1$, kreves det veldig lite kapital før det er fornuftig å vedde spillets maksbeløp, 800. Dette førte til en ny tolkning av beløpsvariabelen: verdiene var uttrykk for en subjektiv sikkerhet, ikke en tilnærming av en objektiv fornuft! Å være denne erfaringen rikere var nyttig, og pekte på et behovet for en kvalitativ komponent også i denne undersøkelsen, slik drøftet i avsnitt 2.2.2, for å kunne vurdere om elevene veddet fornuftig.

Del 1 og de to resterende feltene ble lagt til i etterkant av piloteringen. I piloteringen gjennomgikk 63 elever et likt undervisningsopplegget og svarte på en noe mindre survey.

Surveyen bestod av syv batteri, og bestod som sagt bare av innsatsvariabelen. Det ble laget tre oppgavesett med kombinasjoner av i alt 12 eksperiment. Den overnevnte innsatsvariabelen målte det som ble tolket som misoppfatninger elevene hadde. Det var derimot vanskelig å være sikker på at feilaktige innsatser skyldtes misoppfatninger, og at de ikke bare var forårsaket av eksempel utregningsfeil, misforståelse av spørsmål, eller risikoadferd. Kelly-kriteriet fremstod som for lite nyansert innenfor spillrammene til at innsatsvariabelen alene skulle kunne avdekke misoppfatninger. Slik avtegnet det seg også her et behov for et mer nyanserende aspekt ved undersøkelsen.

En måte å veie opp for en slik svakhet er ved å supplere med en annen metode, såkalt *metodetriangulering* (Larsen, 2017, s. 30). Det er mange måter å gjøre dette på, men det vanlige er å bruke en annen metodikk (Creamer, 2016). Styrken ligger i at de forskjellige metodikkene kan ordnes slik at de utfyller hverandre og kan slik gi større forståelse av det man undersøker. Siden elevene allerede arbeidet med en survey ble det naturlig å legge til mer åpne og ustrukturerte spørsmål med mer kvalitative egenskaper i skjemaet som allerede var i bruk. De to spørsmålene som omhandler sannsynlighetsoverslag ble derfor lagt til surveyen. Disse spørsmålene er mer åpne enn innsatsvariablene, ved at de ikke har oppgitt svaralternativ (Grønmo, 2004, s. 193). Elevenes overslag er begrenset til å ligge mellom 0 og 1, men mellom der fins det uendelig mange alternativ, og elevene står fritt til å begrunne overslaget som de vil. Selv om hvert spørsmål har ett rett svar, er det elevenes oppfattning av korrekt svar som spørsmålene ønsker å belyse, og de forstås derfor som åpne. Informasjonen som gis her er med på å skille om elevene gjør kombinatoriske feil, eller om de baserer innsatsene sine på misoppfatninger om hva som er fornuftig. Grønmo (2004, ss. 194-195) sier at fordelen med lukkede spørsmål er at de er enklere å håndtere for respondentene og intervjuer, mens åpne spørsmål tar mer vare på de ulike nyansene i svarene fra ulike respondenter. Slik kan hver av disse to typene spørsmål belyse to forskjellige spørreord; de lukkede spørsmålene adresserer på *hva* elevene tror, mens de åpne spørsmålene prøver å forstå *hvorfor* elevene tror det (Creamer, 2016, s. 6). En viktig presisering er at det er ønskelig å få svar på hvorfor elevene svarer slik de gjør, slik at man kan stadfeste at en misoppfatning også opptrer, men det vil være vanskelig å kunne konkludere ut fra dette en kausal effekt. For slike sammensatte spørsmål sier Grønmo (2004, s. 196) at hvert av svarene har liten verdi i seg selv, og blir først meningsfylte når de sees i sammenheng med svarene på de andre spørsmålene. Det hadde vært fullt mulig å triangulere ved bruk av andre metoder, som intervju eller observasjon, noe som kunne fanget opp andre aspekter ved elevenes oppfatninger, og kanskje til og med kunne gjort dette bedre. Silverman (2011, s. 45) sier at for enkelhetsskyld bør man som novise fokusere på én metode, og for å kunne fokusere på analysen og ikke ende opp med for mye data, og dette var også medvirkende til at jeg bare bygget videre på spørreundersøkelsen. I neste seksjon vil selve oppgavesettet som surveyen bestod av presenteres.

4.1.3 Eksperimentene og oppgavesett

Oppgavesettet jeg brukte bestod av totalt 17 eksperiment, som kan sees i tabell 2.

	Eksperiment
1	Trill to 6-sidede terninger til du enten får: a) minst en 2-er og ingen 5-er, eller b) en 2-er og en 5-er. Kommer du til å få alternativet b) med 2 og 5 først?
2	Kast en mynt 9 ganger. Bestem om det er mynt eller kron som har kommet flest ganger. Ved neste kast, kommer du til å få det som kom færrest ganger i de 9 første kastene?
3	Kast en mynt 10 ganger. Kommer du til å få akkurat 5 kron og 5 mynt?
4	Trill en 6-sidet terning 5 ganger. Kommer det samme resultatet til å komme minst 2 ganger på rad?
5	Trill en 6-sidet terning 8 ganger. Kommer det samme resultatet til å komme minst 2 ganger på rad?
6	Trill to 6-sidede terninger. Kommer summen av terningene til å bli 7?
7	Trill tre 6-sidede terninger. Kommer summen av to eller flere av terningene til å bli 7?
8	Trill fire 6-sidede terninger. Kommer summen av to eller flere av terningene til å bli 7?
9	Du har tre poser og det er to sjetonger i hver pose. I en pose er begge blå, i en er begge hvite, og i den siste er det en sjetong av hver farge. Trekk opp én sjetong fra av en av posene og se på den. Kommer den gjenværende sjetongen i posen til å ha samme farge som den du trakk opp?
10	Du har en pose med fire sjetonger; 2 hvite og 2 blå. Trekk en sjetong og se på den. Når den neste trekkes, kommer denne til å ha samme farge som den første?

11	Du har en pose med fire sjetonger; 2 hvite og 2 blå. Trekk en sjetong og hold den skjult. Når den neste trekkes, kommer denne til å være hvit?
12	Du har en pose med fire sjetonger, 2 hvite og 2 blå. Trekk én sjetong og hold den skjult. Trekk en ny og se på den. Var den første hvit?
13	Du har en pose med tre sjetonger, 2 hvite og 1 blå. Du skal trekke to sjetonger samtidig. Kommer du til å trekke én hvit og én blå?
14	Du har en pose med tre sjetonger; 2 hvite og 1 blå. Du skal trekkee én sjetong, legge den tilbake og trekke på ny. Kommer du til å trekke to sjetonger med forskjellig farge?
15	Du skal trille en 6-sidet terning til du får en 1-er, kommer du til å ha fått en 3-er eller 4-er før du stopper?
16	Du skal trille en 6-sidet terning til du enten får en 1-er eller en 2-er, kommer du til å ha fått en 3-er eller 4-er før du stopper?
17	Du skal kaste et kronestykke og skriver ned resultatene helt til du får en av de to sekvensene [1. Kron, 2. Kron] eller [1. Mynt, 2. Kron]. Kommer sekvensen [1. Kron, 2.Kron] til å komme først?

Tabell 2: Oversikt over eksperimentene

Grønmo (2004, s. 193) sier at «å formulere spørsmål innebærer både å velge ut hvilke spørsmål som skal stilles, og å gi hvert enkelt spørsmål en språklig utforming». Den første delen av å formulere spørsmålene bygget altså på å velge ut riktig eksperiment. Eksperimentene i tabell 2 ble valgt ut på bakgrunn av flere ting. Noen av eksperimentene var allerede i Borel-kartoteket, som Eksperiment 3, og det fremstod som ganske klart at feilsvar på dette eksperimentet kunne ha en sammenheng med misoppfatninger knyttet til representativitets-heuristikk:

> Spørsmålet fordrer en forståelse av stokastikk, og hva en kan forvente fordelingen av et lite utvalg. Noen som har en representativitetsheuristikk vil her lene seg på at sannsynligheten for hvert av kastene er $0,5$, det mest sannsynlige utfallet skal være representativt for dette; det mest fornuftige er å vedde på «Ja». Som diskutert tidligere er en slik oppfatning feilaktig, og man bør vedde på «Nei», som faktisk har

en sannsynligheten på ≈ 75 %.

Dette samme gjaldt også Eksperiment 9, som er ekvivalent med eksempelet drøftet i avsnitt 3.2.3.2, og handler om misoppfatninger rundt betingning. Andre eksperiment, slik som Eksperiment 1, ble *utviklet* for å kunne fange opp ekviprobabilitetsbias, avsnitt 3.2.2.3. I dette eksperimentet vil personer som har en slik bias tenke at det vil være like sannsynlig å få en toer og en femmer som å få en toer og noe annet, uten å følge denne tankerekka noe videre. Prosessen med å velge ut eksperimentene kan derfor beskrives som en deduktiv prøv-og-feil-prosess, hvor teori og tidligere eksperiment lå i bunnen; så måtte det testes hvilke teorier som kunne integreres i Borel, og hvilke Borel-oppgaver som man ved teori kunne forstå som passende for forskningen. I tabell 3 gis det en oversikt over hvilke eksperiment som ble tenkt å gi informasjon om de ulike misoppfatningene. Hvordan alle eksperimentene henger sammen med misoppfatningene kan forstås ut fra den presenterte teorien, og virkegraden vil drøftes videre avsnitt 5.2.

Misoppfatning tilknyttet	**Eksperiment**
Representativitets-heuristikk	2, 3, 4, 5
Tilgjengelighets-heuristikk	3, 5, 7, 8, 14
Sensitivitet for utvalgsstørrelsen	5, 7, 8
Betinget sannsynlighet	9, 10, 11, 12
Ekviprobabilitet	1, 15, 17
Ikke-rasjonell overbevisning	16
Kombinatorikk	13, 14

Tabell 3: Oppgaver og misoppfatninger

Det var i alt fem klasser som skulle delta i datainnsamlingen. De forskjellige eksperimentene ble fordelt i tre oppgavesett slik at hvert oppgavesett bestod av ti eksperiment. Både oppgavesett 1 og 2 ble gitt til to klasser hver, og oppgavesett 3 ble gitt til én klasse. Dette medførte at noen av eksperimentene ble gjennomført i mange klasser, mens noen bare ble gjennomført i én klasse. Denne fordelingen bygget på funn i piloteringen. Piloteringen pekte på noen eksperiment og kombinasjoner av eksperiment som vakte oppsikt, ved at vi fikk unaturlige fordelinger av «Ja» og «Nei», og at elevenes innsatser ikke var som forventet. Eksempelvis var Eksperiment 3 et

slikt eksperiment, og ble inkludert i alle oppgavesettene.
Tanker om interne forhold påvirket også formuleringen av oppgavesettene. For noen av misoppfatningene var det avgjørende at noen bestemte eksperimentene hørte til i samme oppgavesett. For eksempel: i forsøk på å undersøke om respondentene var følsomme for utvalgsstørrelsen ble både Eksperiment 4 og 5, hvor bare antallet kast er forskjellig, gitt til gruppe 3. Av samme grunn ble gruppe 1 og 2 gitt både Eksperiment 6 og 7, og gruppe 3 ble gitt Eksperiment 6 og 8, hvor 8 var tenkt å være en enda mer ekstrem variant av 7. Ved å ha disse Eksperimentene etter hverandre var tanken at man kunne se om Eksperimentenes distinksjon medførte endrede vedde-mønster hos elevene.
Det var også en intensjon å organisere oppgavesettene slik at noen av gruppene kunne funksjonere som kontrollgrupper. Siden gruppe 2 hadde Eksperiment 4 og gruppe 1 hadde eksperiment 5, kunne man se om svarene fra gruppe 3 på de to oppgavene bar lik karakteristikk, eller om det kunne se ut som om de hadde blitt påvirket av det tidligere eksperimentet. Styrker og svakheter ved hvilke eksperiment som ble brukt, og hvordan de ble organisert drøftes videre i avsnitt 4.2.

Piloteringen hadde også påvirkning på spørsmålenes og eksperimentenes formulering. Et ord som lot til å forvirre og underbygge tanken om at tilfeldighet er en selvkorrigerende prosess var ordet «vil», til dømes: «vil summen av to terninger bli 7?». Dette kom frem gjennom den pågående diskursen i klasserommet under piloteringen, og motiverte den nåværende formuleringen, med bruken av «kommer». Gjennom piloteringen erfarte jeg noen slike formuleringer var med å forvirre respondentene, og fikk endret dem til datainnsamlingen. For noen eksperiment ble ikke endringen tilstrekkelig. For eksperiment 17 var det mange som ikke oppfattet kronologiens viktighet, noe som skapte forvirring. (Grønmo, 2004, s. 192) sier at en av fordelene med personlige intervju er at man kan oppklare spørsmål når respondentene har vanskeligheter med å forstå spørsmålet, noe som er spesielt viktig når respondenten selv fyller ut skjemaet. Dette kunne jeg dra fordel av når jeg var i klasserommet, slik at de ikke-systematiske misforståelsene tilknyttet formuleringen kunne kjapt oppklares. Ut fra egen erfaring vil jeg påstå at det tilhører sjeldenhetene at *alle* elevene får med seg *alle* beskjedene, og det å nøste opp slike forvirringer er noe lærere er kjent med.

4.1.4 Datainnsamling

Selve datainnsamlingen ble gjort gjennom at jeg kom og gjennomførte et undervisningsopplegg sentrert rundt spillet i de klassene. Dette ble gjort i det tidsrommet hvor elevene skulle hatt vanlig matematikkundervisning. Alle elevene som skulle vært i den aktuelle timen deltok i undersøkelsen. Elevene ble informert om min rolle som masterstudent og forsker i denne settingen. Alle elevene fikk utdelt spørreskjemaet, og det ble satt av 10 minutt til at elevene skulle fylle ut del 1. Det ble så gitt en innføring i spillet og dets regler. For hvert eksperiment leste jeg opp eksperiment-teksten, og elevene kunne selv lese teksten på tavla på en PowerPoint-slide. Etter at jeg hadde presentert eksperimentet fikk elevene litt tid, $<$ 20 sekund, til å stille oppklaringsspørsmål i plenum. Deretter fikk elevene 1,5 minutt til å tenke og fylle ut skjemaet. Når tiden var gått ble eksperimentet gjennomført, og for det meste ble gjennomføringen gjort eller assistert av frivillige elever. Elevene ble under hvert spørsmål bedt om å svare på hele spørsmålsbatteriet. Underveis ble konkurranse-aspektet bygget under ved å kartlegge hvem som hadde størst kapital. Etter at alle eksperimentene var gjennomført ble skjemaene samlet inn. Den resterende tiden av timen ble brukt til å forklare motivasjonen bak et korrekt veddemål, og til å fremme viktigheten av en stokastisk forståelse.

4.1.5 Analyse av data

Dataene i undersøkelsen ble genereret av spørsmålsbatteriene. Hvert spørsmålsbatteri kan forstås som en *multivariat variabel*, men for det meste av analysen var det mer naturlig å dekomponere batteriet til enkeltspørsmålene, og så se på hvert spørsmål som en *univariat variabel* (Solbakken, 2019, s. 21) For å analysere slike data foretas det en univariat analyse, hvor det mest nærliggende er å telle opp hvor mange enheter som har markert hver av verdiene (Christoffersen & Johannesen, 2010, s. 141). Variablene som var tallfestede og tellbare var hvor mange som svare Ja og Nei; hvor mange som veddet maksimalt av hva de kunne; og hvor mange som estimerte $p < 0,5$, $p = 0,5$ eller $p = 0,5$. De to første variablene ble ordnet som *dikotome variabler*, som betyr at de bare hadde to verdier; Ja eller Nei; maks eller ikke maks (Solbakken, 2019, s. 51). Estimat-variabelen var er en *kontinuerlig variabel*, siden det finnes uendelig mange mulige estimat å oppgi, men den ble kategorisert i de tre ovennevnte kategoriene. Forekomsten av disse har blitt fremstilt i frekvenstabeller på side 50 og 53. Verdiene skulle si noe om hva respondentene anså som mest sannsynlig, og hvor sikre de var på dette. Frekvensen av de forskjellige verdiene ville derfor være informativt for å se stor andel av respondentene som svarer rett og galt (og da kunne vurderes for en misoppfatning), og hvordan selvsikkerheten variert for de ulike eksperimentene.

For de eksperimentene hvor det var relevant å vurdere $p = 0,5$, enten fordi dette var riktig sannsynlighet eller fordi det var indikasjonen på en misoppfatning, vil frekvensen av Ja og Nei ikke være like åpenbart informativ. For å tolke denne frekvensen ble fordelingens skjevhet vurdert ved hypotesetesting (Solbakken, 2019, s. 154). Nullhypotesen i slike tilfeller var at fordelingen ville være jevn: H_0: Fordelingen av Ja og Nei er sannsynlig gitt at respondentene vet at $p = 0,5$; H_{Alt}: Fordelingen Ja og Nei er ikke sannsynlig gitt at respondentene vet at

$p = 0,5$. Hypotesen ble gitt et signifikansnivå på 10 %. Dette signifikansnivået ble valgt i samråd med masterveileder, skulle balansere presise resultater, og samtidig være ganske sikker på at det er riktig å forkaste nullhypotesen. Når nullhypotesen blir forkastet vil tolkningen være at det er sannsynlig at respondentene ikke vet at $p = 0,5$.
Selv om hovedvekten vil ligge på den univariate analysen, vil det være et innslag av *bivariat analyse* mellom noen av variablene (Solbakken, 2019, s. 85). I avsnitt 5.2.1 vil det sees på sammenhengen mellom eksperimentenes sannsynlighet og hvor sikre respondentene later til å være, og i avsnitt 5.2.8 vil det drøftes hvordan størrelsen på sannsynligheten korrelerer med hvordan respondentene vedder.

4.1.6 Etiske aspekt

Etter samråd med *Norsk senter for forskningsdata* (NSD) ble det klart at undersøkelsen aldri fordret behandling av personopplysninger, og derfor kunne ansees som en anonym behandling (senter for forskningsdata, u.d.-a). Den eneste personopplysningen jeg kunne ende opp med å måtte forholde meg til var respondentenes navn. For å unngå å havne i en situasjon hvor jeg måtte håndtere navn ble det gitt eksplisitt beskjed om å la være å skrive navn på surveyene. Dette stod i tillegg med uthevet skrift fremst på skjemaet. Dette ble gjentatt på slutten av undersøkelsen, slik at respondenter som hadde skrevet navn kunne viske det ut, rive det av, eller streke over det til det ugjenkjennelige. I etterkant av innsamlingen ble alle skjemaene sjekket, og der noen hadde skrevet navn ble skjemaet makulert. Dette var tilfellet for to respondenter. Siden det ikke ble behandlet personopplysninger var det mindre krav til informering. Det ble likevel vurdert som hensiktsmessig for undersøkelsen å gi informasjon om hva deltagelse i forskningen ville innebære, utfra det NSD sier informasjonen skal omfatte (senter for forskningsdata, u.d.-b). I hver gruppe ble derfor undersøkelsens formål presentert, sammen med informasjon om spillet som metode, og hvilke opplysninger som ville kartlegges, og hvordan disse ville bli behandlet. Det ble understreket at alle dataene som jeg kom til å behandle kom til å være helt anonyme. Min rolle som uavhengig forsker i klasserommet ble også tydelig avklart, og for å etablere tillit til forskningen ble det også poengtert at jeg kom som student under Universitetet som institusjon. Selv om tematikken i forskningen ble poengtert, valgte jeg å ikke informere om forskningsspørsmålet. Dette var fordi jeg ikke ville at respondentene skulle ha noen forutinntattheter som kunne prege metaforståelsen av egen sannsynlighetsoppfatning, og således veddingen deres. Kanskje det viktigste aspektet her var å avklare at deltakelse i undersøkelsen var helt frivillig. Siden det ble viet tid til undersøkelsen i undervisningstiden måtte alle være til stede under innsamlingen, men utfyllingen og innleveringen av skjemaet var helt frivillig; å levere inn skjemaet ble vurdert som et samtykke til å være deltakende i undersøkelsen. I denne sammenhengen var det viktig å få frem at undersøkelsen var uavhengig av faglærer, og at opplysninger som ble samlet hverken ville eller kunne få konsekvenser for undervisningen eller resultatene i faget. Noen av elevene hadde kjennskap til meg fra praksisperioden, og i tillegg satt faglærer også i klasserommet under gjennomføringen. Dette kan ha påvirket deres oppfatning av frivillighetsgraden, og de kan ha følt et press for at de måtte delta.

Av etiske hensyn må det sies at jeg ikke er sponset av Borel, og har ikke noen tilknyttning til spillmakerne. Jeg har ingenting å tjene på å bygge undersøkelsen min på dette spillet, og spillets sentrale plass er bare fordi jeg ble inspirert av det og så det som et potensielt godt diagnostisk verktøy.

4.2 Troverdighet

Formålet med metoden er at den skal generere data som gjennom analysen skal besvare forskningsspørsmålet. For at dette skal være tilfellet er det viktig at dataene er troverdige. For å sikre dette må man kontrollere at metoden faktisk måler det den forsøker å måle, og at den måler dette konsekvent (Johnson, 2003). Dette blir også viktig for selve analysen av dataene; selve veien fra data til funn må også være troverdig. Disse aspektene vil drøftes videre under.

4.2.1 Validitet

En undersøkelses *validitet* handler om undersøkelses gyldighet (Solbakken, 2019, s. 39). Det skilles ofte mellom indre og ytre validitet. Den indre validiteten beskriver i hvilken grad resultatene fra undersøkelsen er gyldige for utvalget (Solbakken, 2019, s. 41). Den ytre validiteten handler om hvorvidt resultatene fra utvalget er gyldige, og kan generaliseres, for selve populasjonen (Solbakken, 2019, s. 40).

4.2.1.1 *Ytre validitet*

Formålet med oppgaven er å si om bruken av Borel kan bidra til å få informasjon om elevers misoppfatninger. Dette peker på at den ønskede effekten av Borel skal gjelde for hele elevpopulasjonen, men gitt masteroppgavens rammebetingelser ble det utfordrende å sikre at utvalget skulle være generaliserbart for hele denn populasjonen. Generaliserbarheten svekkes blant annet av at utvalget var et *bekvemmelighetsutvalg*, hvor utvelgelse av enheter bygget på bekvemmelige og praktiske årsaker, fremfor en helt tilfeldig utvelgelse. Dette er en form for *ikke-sannsynlighetsutvelging*, fordi sannsynligheten for at akkurat disse enhetene ble valgt ut fra populasjonen ikke er kjent (Solbakken, 2019, s. 36). Årsaken til at nettopp dette utvalget

var mer bekvemmelig er at jeg har gjennomført en praksisperiode ved skolen hvor undersøkelsen ble gjennomført. Praksisveilederen min fikk da rollen som *portvakt*, og gjorde det enklere å få innpass i klasserommene for å samle data (Fangen, 2011, s. 53). Slik bekvemmelighetsutvelging går på bekostning av utvalgets representativitet av populasjonen (Firebaugh, 2008, s. 21); det vil være vanskelig å kunne konkludere om spillets virkning vil være gjeldende overalt. Likevel er det positivt for representativiteten at utvalget favner en betydelig del av et helt årskull (Solbakken, 2019, s. 151). Gjennom portvakten ble jeg gitt tilgang til alle klassene på trinnet, og nesten alle elevene i hver klasse deltok i undersøkelsen. Siden dette var på 10. trinn avhenger ikke respondentenes skoleplass av karaktersnitt. Det vil derfor være en spredning av faglig sterke og svake elever i utvalget. Det fremstod også som en jevn fordeling av kjønn, noe som logisk bør stemme for en så stor andel av et helt trinn.
At utvalget var fra én skole betød at andre eksterne variabler kunne spille inn og være en svekke den ytre validiteten. Skolens geografiske plasseringen vil ha en innvirkning på hvordan skolen styres, og sammen med den tilhørende sosioøkonomiske standarden vil dette kunne være ledende for hvilke lærere som jobber på skolen. Disse faktorene vil påvirke undervisningen av sannsynlighet og misoppfatninger, vil ha noe å si for hvilke konseptuelle verktøy elevene hadde i møte med undervisningen. De sosioøkonomiske forholdene vil også kunne ha en innvirkning på hva elevene har med hjemmefra, generelt knyttet til verdier, ferdigheter og læring, og spesielt rundt spillerfaringer og økonomisk forvaltningstankegang. Dette vil påvirke heuristikkene matematiske avgjørelsene som ligger bak veddingen.

Et av de viktigste prinsippene når det kommer til innsamling av data er at det ikke har så mye å si hvor stort utvalget er, men hvordan det ble valgt (Firebaugh, 2008, s. 21). Det er ikke åpenbart hvordan studiens utvalg forholder seg til dette prinsippet. Bekvemmelighetsutvelgingen gjør det vanskelig å rettferdiggjøre en generalisering av funn for hele elevpopulasjonen i Norge. At utvalget er en så stor del av populasjonen ved skolen gjør at man kan anta at det i det minste er representativt for skolen. I tillegg kan det også tenkes at funn vil kunne generaliseres til skoler med lignende demografi. Dette i seg selv var ikke hovdegrunnen for å la utvalget bli så stort som det ble, relativt til populasjonen: grunnen var å sikre å få nok data til å få rik analyse. En av de største begrensningene med et lite utvalg er den manglende analytiske slagkraften (Firebaugh, 2008, s. 23). For små utvalg vil man ha vanskeligere for å kontrollere for effekter og korrelasjon enhetene i mellom. Så selv om utvalgets representativitet hadde sine begrensninger for den ytre validiteten, var det likevel et ønske å ha høy indrevaliditet, og fremstod derfor som lettere med et stort utvalg.

4.2.1.2 Indre validitet

Det er flere faktorer som er med på å bestemme om man kan stole på resultatene fra undersøkelsen (Johnson, 2003, s. 280). For undersøkelsen min ble det tatt grep for å styrke den indre validiteten, og i etterkant fremkom det likevel noen svakheter ved undersøkelsen. For å styrke målingen ble det opplegget gjennomført i flere klasserom. Dette var ikke bare for å samle nok data, men også for å begrense virkningen bakenforliggende variabler har på dataene (Johnson, 2003, ss. 284-290). Variabler som dagsform, faglærer eller statistisk avvik fra forventingen vil kunne påvirke dataene, og ved flere målinger vil slike påvirkninger ha mindre effekt. En sikring for at måleinstrumentet, spillet og oppgavesettet, faktisk målte det som var tiltenkt lå i den induktive metoden, hvor eksperimentene hele tiden ble sett opp mot den eksisterende teori. I tillegg ble opplegget utviklet i et nært samarbeid med veileder, som er mer erfaren hva gjelder forskning. Slik ekstern validering skulle være med å sikre at egne forutintattheter og inkompetanse ikke påvirket undersøkelsen negativt (Dewilde, 2020, slide 23).

Det som står frem som mest svekkende for undersøkelsens resultat er nok knyttet til måleapparatet og samspillet med de operasjonelle definisjonene. Når en operasjonaliserer et begrep bestemmer en «hvilke indikatorer som skal tas som tegn på begrepet» (Kleven, 2014, s. 28). I denne undersøkelsen ble misoppfatningene operasjonalisert som ulike operasjoner i spillet (Johnson, 2003, s. 296); hvordan respondentene veddet, og hvordan de beskrev motivasjonen for disse innsatsene. Det finnes mange måter å operasjonalisere et begrep på, og måten det blir målt vil alltid medføre noen feil (Johnson, 2003, s. 297). Ved å bruke survey kunne misoppfatningene operasjonaliseres i forkant av innsamlingen, og begrense informasjonsmengden til akkurat det som var ønskelig, som var en styrke (Larsen, 2017, s. 28). For at denne fordelen skal gjøre seg gjeldende må den *definisjonsmessige validiteten* være sterk: den operasjonelle definisjonen må dekke det teoretisk definerte begrepet (Solbakken, 2019, s. 43). I undersøkelsen ble det aldri sikret at måten misoppfatningene ble operasjonalisert på faktisk var god; dette var bare en forventing tuftet på teorien. Selve spørsmålsbatteriet skulle fange opp og klare å skille mellom de ulike misoppfatningene, men i praksis ble dette vanskelig. I flere tilfeller vil det være aktuelt å vurdere om flere misoppfatninger gjør seg gjeldende samtidig, og måten de opptrer på vil ikke nødvendigvis være mulig å skille med dette redskapet; at noen vedder på Ja kan både skyldes at de har estimert $p = 2/3$ og $p = 0,5$, og dersom de har estimert $p = 0,5$ kan det i noen tilfeller både skyldes representativitet, tilgjengelighet, ikke-stokastisk tro, eller generelt regnefeil. Dette går imot en av de viktigste reglene for surveyer: man skal kun spørre om én ting om gangen (Kleven, 2014, s. 36). For å veie opp for dette skulle flere eksperiment belyse de samme misoppfatningene (Grønmo, 2004, s. 198). Det var derimot ikke en sikring mot at de triangulerende eksperimentene ikke led av den samme svakheten. Dette kunne kanskje vært unngått om forskningen hadde bygget på et rammeverk, men da dette feltet ikke var veldig utforsket før, var det vanskelig å finne noe slik. Slik fremstår operasjonaliseringen som for upresis og dette svekket den definisjonsmessige validiteten.

Selve impelemteringen av måleapparatet i undersøkelsen var også med å svekke resultatenes gyldighet. Siden det ikke ble gjennomført noen pretester som kartla misoppfatningenes forekomst mer nøyaktig, fantes det ikke noen «fasit» å se undersøkelsens data opp mot. Slik sett ble måleapperatet selv brukt til å vurdere måleapparatets evne til å avdekke misoppfatninger. Dette gjør at man ikke kan vite om resultatene er representative for misoppfatningenes faktiske forekomst. Det kan være at spillet ikke fanger opp misoppfatninger som absolutt er tilstedet, og det kan være at noe tolkes som en misoppfatning, som faktisk ikke er det. En konsekvensiell mellomting er at apparatet kartlegger de tenkte misoppfatningene uten at selve gir et avbildning av den faktiske tilstanden.

En annen svakehet ved spillets mekanikk er knyttet til tidsbegrensningen. Selv om dette er en dynamikk som er med å skape spenning, kan det påvirke svarene. Dette kan bidra til at spørreskjemaet ikke kartlegger konsekvente misoppfatninger, men heller øyeblikksbilder av sannsynlighetsintuisjonen respondentene hadde i møte med eksperimentene. O

4.2.2 Reliabilitet

Reliabilitet kan forstås som målesikkerhet og handler om hvor konsekvente resultatene er: vil man få de samme resultatene om man gjør undersøkelsen på ny? (Johnson, 2003, s. 279) Det er flere kilder til reliabilitetsproblemer, og for denne undersøkelsen var noen av disse kildene meg selv, respondentene og selve surveyen (Solbakken, 2019, s. 45).

For det første kan min egen tilstedeværelse under datainnsamlingen ha påvirket respondentenes svar, noe som kalles *intervjueffekt* (Solbakken, 2019, s. 46). En av fordelene med personlige intervjue er at man kan hjelpe respondentene dersom noen begrep er uklare, og kan motivere dem underveis. For å motivere deltakelse kan jeg ha utvist et fremtredende engasjement, og dette kan ha resultert i mer vågale innsatser enn det respondentene ville hatt uten min påvirkning. Jeg kan også ha sagt ting for å motivere som kan å påvirket veddingen. Til og med måten spørsmålene ble presentert vil ha kunnet gi repsondentene inntrykket av at det ene eller det andre utfallet var mer fornuftig å vedde på. For å styrke reliabiliteten ble det laget noen faste rammer for å prøve å begrense min påvirkning. Det ble satt en lik tidramme for alle eksperimentene, og eksperimentteksten hadde en fast formulering som respondentene selv kunne lese. Dette gjorde at respondenter i forskjellige grupper skulle få like oppfatninger av eksperimentene.

En ting som senket undersøkelsens reliabilitet var knyttet til frafall av respondenter. Dette er et vanlig problem i undersøkelser hvor respondentene ikke vil eller klarer å svare (Grønmo, 2004, s. 209). I undersøkelsen var dette tydelig for de spørsmålene i batteriet som krevdes mest av respondentene. For spørsmål 3, hvor respondentene skulle gi et sannsynlighetsestimat, var frafallsprosenten 64 % relativt til hvor mange som bidro på selve veddingen i eksperimentene. For spørsmål 4, hvor de skulle begrunne estimatene sine, var frafallet enda større: i gjennomsnitt 91 %. Dette har stor påvirkning for resultatenes reliabilitet. Om flere hadde svart ville dataene blitt vektlagt på en annen måte, fordi de hadde utgjort en mindre andel av svarmengden. Om frafallet hadde vært likt i et annet klasserom kunne man likevel fått et helt annerledes utvalg når bare 9 % i snitt svarer, og helt andre aspekt hadde da blitt belyst av de aktuelle respondentene.
Hvordan oppgavesettene ble utformet kan også ha påvirket undersøkelsens reliabilitet. Rekkefølgen av spørsmålene vil kunne forme måten respondentene svarer (Grønmo, 2004, s. 206). Om svarene på et bestemt spørsmål påvirket svarene på et annet spørsmål, såkalte *konteksteffekter*, var noe som var forventet og nødvendig for å vurdere tilgjengelighetsheuristikker og sensitivitet for utvalgsstørrelse. Men for noen av eksperimentene kan en slik effekt oppstått uintensjonelt, og da ha vært en bakenforliggende variabel som ikke var redegjort for. Alle respondentene gikk gjennom Eksperiment 6, og hadde enten Eksperiment 7 eller 8 i etterkant, men ville man sett de samme mønstrene om man hadde byttet rekkefølgen på disse? Utformingen tok heller ikke høyde for at rekkefølgen eksperimentene kom i kunne påvirke hvordan respondentene veddet. Respondentenes risikoadferd kunne opptre ulikt på forskjellige tidspunkt i spillet; man har kanskje mer tapsaversjon i begynnelsen? Hadde det da vært mer rullering av eksperimentenes kronologi ville man kanskje fått andre målinger.

5 Funn

Dette kapittelet har som mål å forklare hvilke resultat spørreskjemaet genererte, og hvordan disse bidrar med svar til forskningsspørsmålet. Det første delkapittelet, avsnitt 5.1, vil fremvise og redegjøre for de innsamlede dataene. I det påfølgende delkapittelet, avsnitt 5.2 vil disse dataene bli analysert etter analysemetoden presentert i forrige kapittel for å lete etter resultat som kan forstås som relevante funn for oppgaven.

5.1 Data

Dataene fra undersøkelsen har i all hovedsak blitt delt inn i to grupper, hvor hver gruppe har blitt dedisert en egen tabell, og det vil bli referert til disse to tabellene videre i analysen. Tabell 4 er en organisering av responsene på de to første spørsmålene i spørsmålsbatteriet som utgjorde innsatsvariabelen, og hver rad representerer det tilhørende eksperimentet fra tabell 2. For å ha et tolkningsgrunnlag viser den andre og tredje kolonnen henholdsvis sannsynligheten for at eksperimentets utfall skal tilsvare «Ja», og hvor mange respondenter det var i hvert eksperiment. Som man kan se er antallet respondenter forskjellig for de ulike eksperimentene, som avhenger av de praktiske vurderingene fremmet i avsnitt 4.1.3. Tabellen gir et overblikk over respondentenes tendenser når de vedder; om det er et eksperiment hvor mange vedder *fornuftig*, eller et der få gjør det. Dette vurderes ut fra de siste tre kolonnene; hvordan fordelingen av ja-og-nei fortoner seg, hvor stor innsats respondentene går inn med, og hvor mange som, ut fra Kelly-kriteriet, vedder riktig.

Eksp.	**P(Ja)**	**Respondenter**	**Fordeling Ja:Nei**	**Størst mulig innsats**	**Riktige innsatser**
1	2/11	40	13:27 (0,33)	17 (0,43)	0,18 (-800)
2	1/2	67	38:29 (0,57)	25 (0,37)	0,22 (±100)
3	0,246	107	8:101 (0,07)	82 (0,77)	0,73 (-800)
4	0,52	67	29:39 (0,43)	33 (0,49)	0,24 (100)

5	0,72	63	45:19 (0,71)	33 (0,52)	0,32 (800)
6	1/6	107	25:82 (0,23)	67 (0,63)	0,56 (-800)
7	0,51	84	50:34 (0,60)	49 (0,58)	0,05 (100)
8	0,83	23	18:5 (0,78)	14 (0,61)	0,48 (800)
9	2/3	44	32:13 (0,73)	30 (0,68)	0,50 (800)
10	1/3	63	10:54 (0,16)	35 (0,56)	0,44 (-800)
11	1/2	44	30:14 (0,68)	30 (0,68)	0,02 (±100)
12	1/3 eller 2/3	63	45:18 (0,71) (Riktig: Feil)*	17 (0,68)	0,33 (-800)
13	2/3	63	48:15 (0,76)	43 (0,68)	0,46 (800)
14	4/9	67	23:44 (0,34)	54 (0,81)	0,34 (kapital-avhengig)
15	2/3	40	30:10 (0,75)	31 (0,78)	0,58 (800)
16	1/2	44	34:10 (0,77)	35 (0,80)	0,02 (±100)
17	1/4	84	38:46 (0,45)	73 (0,87)	0,44 (-800)

Tabell 4: Data og innsatsvariabel. Tallene i parentes viser hvor stor andel av respondentene som hhv. svarte «Ja», gikk inn med alt, og hva som var en riktig innsats. *I dette eksperimentet beskrives svarene som riktig:feil, i stedet for Ja:Nei.

Tabell 5 er en organisering av hvordan elevene estimerer sannsynligheten, som svarer til det tredje spørsmålet i batteriet. Denne tabellen brukes for å forstå hvorfor respondentene vedder slik de gjør. Som forklart i avsnitt 4.1.5 var det gunstig å kategorisere estimatene relativt til

$p = 0,5$. Elevene ble i spørsmål 4 bedt om å gi en forklaring på hvor de estimerte slik de gjorde. Svarene på dette spørsmålet har ikke blitt kategorisert. Noe av grunnen til at det ikke fremstod som hensiktsmessig å kategorisere var at de ikk eha

Disse har blitt kategorisert etter hvilke misoppfatninger de later til å lene seg på.

Det ble ikke vurdert som nyttig å prøve å fremstille disse på noe tallmessig vis; noen av dem vil i stedet bli brukt for å supplere dataene ved å illustrere hvordan noen elever kan argumentere for sitt syn.

Frafall av respondenter er som sagt et kjent problem når man bruker spørreundersøkelser, og spesielt når respondentene fyller ut skjemaet selv. For denne undersøkelsen ble dette tydeligst når det kom til disse to siste spørsmålene i hvert batteri: mange lot være å svare på disse spørsmålene. Dette vises i den tredje kolonnen i, som viser hvor mange som svarte på spørsmål 3, og hvor stor andel det utgjorde av alle respondentene for det eksperimentet. Dette varierte stort fra eksperiment til eksperiment. I de tilfellene hvor det var færre enn 20 respondenter på disse spørsmålene blir ikke anslagsfordelingen angitt med desimaler, for at det skal være tydeliggjøre når en skjevfordeling er mer uforventet. Siden det endte med at ikke alle respondentene svarte, men bare de som selv ville, vil dette være et tilfelle hvor fordelingen ikke nødvendigvis er ensbetydende med en lik fordeling av estimat for alle respondentene. En slik antagelse vil ikke være reliabel. Resulatetene vil likevel bli brukt for å kunne gi en pekepinn for hvorfor fordelingen av veddemål ble som den ble.
Det må også påpekes at oppslutningen om spørsmål 4, hvor respondentene skulle begrunne anslaget sitt, hadde en veldig liten oppslutning. Eksperimentet hvor flest begrunnet anslaget sitt var Eksperiment 3 hvor 13 personer, omtrent 18 %, ga en begrunnelse. Ellers var gjennomsnittlig andel av besvarelser på spørsmål 4 bare 8,6 %.

Eksp.	**P(Ja)**	**Svar**	**Estimert sannsynlighet**		
			P(Ja)<0,5	P(Ja)=0,5	P(Ja)> 0,5
1	2/11	8 (0,20)	7	1	–
		3:5	2:5	1:0	–
2	1/2	57 (0,85)	2 (0,04)	55 (0,96)	–

		31:26	1:1	30:25	–
3	0,246	48 (0,44)	20 (0,42)	28 (0,58)	–
		5:43	1:19	4:24	–
4	0,52	29 (0,41)	19 (0,66)	4 (0,14)	6 (0,20)
		14:15	7:12	2:2	5:1
5	0,72	9 (0,14)	6	1	2
		5:4	2:4	1:0	2:0
6	1/6	38 (0,35)	30 (0,79)	7 (0,18)	1 (0,03)
		14:24	9:21	5:2	0:1
7	0,51	20(0,23)	12 (0,60)	3 (0,15)	5 (0,25)
		9:11	6:6	1:2	2:3
8	0,83	6 (0,26)	3	1	2
		3:3	0:3	1:0	2:0
9	2/3	28 (0,62)	5 (0,18)	11 (0,39)	12 (0,43)
		21:7	4:1	8:3	9:3
10	1/3	30 (0,47)	26 (0,87)	1 (0,03)	3 (0,1)
		3:27	1:25	0:1	2:1
11	1/2	23 (0,51)	5 (0,22)	18 (0,78)	–
		13:10	3:2	10:8	–

12	1/3	9	8 (2:6)	–	1 (0:1)
	2/3	14	3 (2:1)	6 (5:1)	5 (5:0)
	Totalt	23 (0,37)	Feil_{1^*}: 4(0,17)	$\text{Feil}_{2^{**}}$: 6 (0,26)	Riktig: 13(0,57)
13	2/3	11 (0,17)	4	5	2
		10:1	4:0	4:1	2:0
14	4/9	24 (0,35)	14 (0,58)	2 (0,08)	8 (0,33)
		8:16	4:10	2:0	2:6
15	2/3	3 (0,07)	2	–	1
		3:0	2:0	–	1:0
16	1/2	18 (0,41)	2	11	5
		16:2	1:1	10:1	5:0
17	1/4	22 (0,26)	8 (0,36)	14 (0,64)	–
		13:9	5:3	8:6	–

Tabell 5: Fordeling av respondentenes sannsynlighetsestimat. *Feil_1 er et estimatavvik fra $p > 1/6$. **Feil_2 er et estimatavvik fra $p \leqslant 1/6$.

For at tabellen skulle prøve å holdes enkel ble det valgt å ikke inkludere alle estimatene respondentene bidro med. Tabellen gir en pekepinn på hvor mange som hadde et estimat som ga en utfallstilnærming til de respektive sidene av Ja og Nei. Der det blir aktuelt for analysen å spesifiserer estimatene ytterligere vil disse dataene bli konkretisert.

5.2 Analyse

Dette delkapittelet vil presentere hvordan dataene vitner om hvordan spillet belyste ulike misoppfatninger. Misoppfatningene er gruppert som i tabell 3 side 40, og ellers vil analysen ta utgangspunkt i tabell 4 og tabell 5.

5.2.1 Forståelse av veddet beløpet

Formålet med beløpet var å få et inntrykk av respondentenes selvsikkerhet på at det de hadde estimert var korrekt. Vi begynner med å se bort ifra de seks eksperimentene hvor det, i følge Kelly-kriteriet, er mest fornuftige å ikke gå inn med størst mulig innsats (dvs. de eksperimentene hvor $p = 0,5 \pm 0,1$). I de resterene eksperimentene veddet i gjennomsnitt 65 % av respondentene fullt beløp, og i gjennomsnitt veddet 31 % av disse på feil utfall. Denne prosentandelen varierte fra eksperiment til eksperiment, og i det eksperimentet hvor flest bommet, Eksperiment 1, veddet bare 40 % av de som veddet maksimalt på riktig utfall. Dette var i tillegg det eksperimentet hvor færrest veddet maks, og totalt var det bare 18 % av respondentene som veddet på både riktig utfall og riktig beløp. Dette kan derfor forstås som et eksperiment hvor respondentene var usikre, og mange gjorde feil. Eksperiment 3 var det eksperimentet hvor flest veddet riktig, og her veddet 95 % av dem som veddet maks på riktig utfall. Generelt er det en sterk korrelasjon mellom variablene *riktig innsats* og *maks innsats*, $r = 0,65$, som betyr at ofte når respondentene går inn med maksimal innsats, så satses det på det mest sannsynlige utfallet (Solbakken, 2019, s. 113). Akkurat innsatsstørrelsen kan tenkes å avhenge av respondentens spillferdighet og risikoadferd, og ikke nødvendigvis deres forståelse av sannsynlighet. En respondent som har tapsaversjon vil kunne forstå at sannsynligheten for et utfall er $2/3$, og likevel ikke vurderer denne sannsynligheten som *god nok* til å satse det størst mulige beløpet.

Videre kan det sees at korrelasjonen mellom hvor «stor» sannsynligheten er, størrelsen $|p - 0,5|$, og andelen respondenter som veddet maks er veldig svak: $0,04$. Det kan også sees at korrelasjonen mellom å vedde maksimalt og å vedde på det mest sannsynlige utfallet også er veldig svak: $0,16$. Dette er et overraskende funn, da det var forventet at de maksimale innsatsene skulle opptre sammen med høye sannsynligheter og ikke med de lave. Dette kan peke på at spillmekanikken begrenset forholdet til innsatsene: det var for lett å vedde maks. Dette resultatet støtter også ideen om at innsatsene har lite å gjøre med den objektive sannsynligheten, og bør tolkes som et mer subjektivt mål. Dette kan sees i de tre eksperimentene, Eksperiment 2, 11 og 16, hvor $p = 0,5$. Her veddet henholdsvis 37 %, 68 % og 80 % maksimalt beløp, og i anslagene estimerte henholdsvis 96 %, 78 % og 61 % helt riktig i disse eksperimentene. Det virker som om jo flere som oppdaget at sannsynligheten var 0,50, jo mer konservative innsatser ble gjort. Den lave korrelasjonen mellom maksimal innsatsstørrelsen og å vedde på riktig utfall gir rom for enda en tolkning; nemlig at respondentene var sikre på sannsynlighetsestimatene sine, men at denne antatte sannsynligheten var et dårlig estimat i forhold til den reelle sannsynligheten.

At det veddede beløpet tilsynelatende har så liten sammenheng med sannsynlighetene er interessant, men gjør at det ikke er så mye fortolkningsverdi i denne variabelen.

5.2.2 Representativitetsheuristikk – 2, 3, 4 og 5

Det var i alt fire eksperiment som ble brukt for å se om repsondentene hadde basert innsatsene sine på representativitetsheuristikk, Eksperiment 2, 3, 4 og 5.

De to førstnevnte eksperimentene så på fordelingen av myntkast, og om respondentene tenker at det mest sannsynlige er at det blir fem av hvert utfall. I Eksperiment 2 skulle respondentene avgjøre om et tiende kast ville være påvirket av de ni foregående, og gjenkjennes som Hasardspellerens feilslutning. Feilslutningen kan gi en tilbøyelighet til både å svare Ja og Nei. Det måtte derfor vurderes om denne fordelingen var usannsynlig skjev, da det var en indikator på at representativitetsheuristikk hadde vært styrende for veddingen. Ved å ta utgangspunkt i nullhypotsen for eksperiment med $p = 0,5$ ble det vurdert om det var presedens for å si at hypotesen kunne forkastes på grunn av skjevfordeling av Ja- og Nei-verdiene. Fordelingen var innenfor konfidensnivået, og nullhypotesen kunne derfor ikke forkastes. De andre spørsmålene må derfor brukes som indikatorer på misoppfatninger. Tabell 5 viser at de fleste, 96 %, av dem som estimerte sannsynligheten anslo at $p = 0,5$. Ut fra dette later det til at representativitet ikke preger måten respondentene vedder, men når man ser på begrunnelsesresponsene kan man se tendenser til feilslutninger likevel. Her er to responser som illustrer dette:

> «Det er 50/50, men siden det er flest krone er det litt større sjanse at det blir mynt.»
> «Sannsynligheten er 40/60. Det er litt mindre sjanse for mynt.»

I alt fem av de åtte som ga begrunnelser begrunnet det på lignende vis, og alle utenom én anslo $p = 0,5$. Utfra dette kan det tenkes at flere av dem som anslo $p = 0,5$ også begrunner det slik.

I Eksperiment 3 spørres det om det er mest sannsynlig at man får akkurat fem krone og fem mynt. En representativitetsheuristikk vil her komme til uttrykk ved at man setter innsatsen på Ja. Det er det veldig få som gjør her, bare 7 %, og man kan ikke si at innsatsene tyder på representativitet. Det som er verdt å merke seg er at 58 % av de estimerende respondentene estimerer $p = 0,5$, som åpenbart ikke er riktig, og vil kunne være tegn på representativitet. Blant disse var det en usannsynlig høy andel som valgte Nei – 24/28. Sannsynligheten for en slik fordeling, dersom alle tror at $p = 0,5$, er bare $0,02$ %, som er signifikant. Dette gir grunn til

å forkaste nullhypotesen, og si at det er sannsynlig at ikke alle som oppgir $p = 0,5$ faktisk tror det, men sier det av en annen grunn. Denne grunnen kan for eksempel være at respondentene har en følelse av at en slik fordeling er usannsynlig, men ikke vet hvordan dette skal uttrykkes matematisk. Da vil det være lett å trekke linjene til sannsynlighetsfordelingen for ett enkelt myntkast, og bare si at $p = 0,5$, selv om man ikke tror det. Hvis dette er tilfellet er det færre enn 58 % som oppgir p=0,5 på grunn av at de har en representativ oppfatning.
Eksperiment 4 og 5 vuderer om elevene forventer rekker av samme utfallet med en terning. Spesielt i Eksperiment 4, hvor man kaster terningen fem ganger – færre ganger enn antallet terningsider – virker det ikke representativt at man får ett tall flere ganger på rad, samtidig som minst to tall uteblir helt. For begge eksperimentene er det mest sannsynlig å få en slik rekke, og man bør vedde på Ja. Dersom elevene hadde regnet på det, ville de sett at for Eksperiment 4 var sannsynligheten nær 0,5, noe som kunne rettferdiggjort å vedde på Nei, og i Eksperiment 5 kunne man brukt en lignende fremgangsmåte og sett at det er stor sannsynlighet for at en slik rekke oppstår. Ingen av estimatene eller begrunnelsene tydet på at elevene gjorde en slik utregningen, og det ble ikke oppgitt noen matematisk forankrede begrunnelser for noen av eksperimentene, og svarene kan derfor antas å hvile fullt på intuisjon. I Eksperiment 4 svarer 57 % Nei, og i Eksperiment 5 svarer 28 % det samme. I begge eksperimentene anslår omtrent 2/3 av respondentene $p < 0,5$. Dette antyder at respondentene ikke ser på rekker som sannsynlige, noe som kan indikere representativitetsheuristikk.

5.2.3 Tilgjengelighetsheuristikk og sensitivitet for utvalgsstørrelse – 3, 5, 7, 8, 14

Noen av eksperimentene var nesten indentiske, med bare noen små endringer i variabelverdiene. For disse eksperimentene er det aktuelt å se etter mønster som indikerer tilgjengelighetsheuristikk eller liten sensitivitet for utvalgsstørrelse. Begge disse misoppfatningene sees i sammenheng med tidligere erfaringer, hvor man forsøker å vurdere om respondentenes innsatser er preget av disse erfaringene. Sensitivitet for utvalgsstørrelse avhenger av at denne variabelen endres fra det ene eksperimentet til det neste, mens tilgjengelighetsheuristikk bare bygger på tidligere erfaringer. For totalt fem eksperiment er det vurdert at dette kan være innvirkende på veddingen: Eksperiment 3, 5, 7, 8 og 14.
Tre klasser hadde både Eksperiment 2 og 3. Her er eksperimentet nesten helt likt, men spørsmålet blit bare stilt på forskjellige tidspunkt i gjennomføringen av eksperimentet. Én klasse hadde både Eksperiment 13 og 14, hvor estetikken er ganske lik: man har tre sjetonger og to skal trekkes. På grunn av dette tolkes det som at tilgjengelighetsheuristikk kan ha spilt inn på svarene i Eksperiment 3 og 14.
Én klasse hadde Eksperiment 4 og 5, hvor det er ulikt antall terningkast, og alle gruppene hadde

Eksperiment 6 og enten Eksperiment 7 eller 8, hvor det er terningantallet som varierer. Her kan det forstås som at både tilgjengelighetsheuristikk og utvalgsstørrelsen kan ha innvirkning på innsatsene. I tabell 6 vises en oversikt over hvor stor andel av respondentene som svarte Ja på de ulike eksperimentene.

Eksperiment	**Respondentgruppe**		
	1, 40 respondenter	2, 44 respondenter	3, 23 respondenter
Eksperiment 2	–	x	x
Eksperiment 3	0,13	0,05	0,04
Eksperiment 4	–	0,43	0,45
Eksperiment 5	0,75	–	0,65
Eksperiment 6	0,15	0,32	0,22
Eksperiment 7	0,68	0,52	–
Eksperiment 8	–	–	0,78
Eksperiment 13	0,83	–	0,65
Eksperiment 14	–	0,41	0,21

Tabell 6: Andel av positive innsatser på gruppenivå for spørsmål adresserende tilgjengelighetsheuristikk og sensitivitet for utvalgsstørrelse

Når utvalget øker blir det i alle disse tilfellene en større andel gunstige utfall, og det ser ut som at respondentene oppfatter denne virkningen. I samtlige eksperiment kan det sees at flere vedder på Ja når utvalget øker. Totalt går det fra at 23 % vedder på Ja i Eksperiment 6, til at 60 % og 78 % vedder Ja i hhv. Eksperiment 7 og 8. Dette er en rent statistisk indikator på at respondentene er bevisste på utvalgsstørrelsen. Dette tolkningen støttes av hvordan noen av respondentene begrunner innsatsene sine:

> «Jo flere ganger, jo større sannsynlighet»
> (stemte Ja) «Fordi det er større sannsynlighet enn i forrige oppgave»
> «Større sjanse jo mer terninger»
> «Magefølelsen sier ja, fordi mer terninger»

Her kan man se at elevene legger stor vekt på endringen fra forrige eksperiment. Måten de vurderer endringen fremstår derimot som noe ukritisk; de vedder på Ja fordi sannsynligheten øker, men de utviser liten betenkning rundt hvorvidt sannsynligheten faktisk er *størst* for Ja, selv om har økt. For Eksperiment 7 er det faktisk bare *litt* mer sannsynlig at utfallet blir tilsvarende Ja – 51 %.

Tilgjengelighetsheuristikker bygger på tidligere erfaringer. De utenom-klasseromlige erfaringene er umulige å ta høyde for, og derfor ble misoppfatningen kun vurdert ut fra hvilke eksperiment respondentene hadde hatt tidligere. Det var det få klare indikatorer på at elevens innsatser bygde på tilgjengelighet fra de foregående eksperimentene, men noe data tilsa at en slik ressonering kan ha tatt sted. I Eksperiment 3 kan man se at de to gruppene som hadde erfart Eksperiment 2 svarte mer Nei enn den siste gruppen. I gjennomsnitt veddet i underkant av 4,5 % av de med erfaringer på Ja, mens omlag 13 % av de uten erfaringer veddet på Ja. At nesten tre ganger så mange vedder på Ja når de ikke har erfaring kan være en effekt av tilgjengelighetsheuristikk, men kan også skyldes andre ikke-kartlagte effekter, eller ren tilfeldighet. En annen ting som kan legges merke til er at respondentgruppe 3 svarte ganske likt gruppe 2 i Eksperiment 4, mens for Eksperiment 5 var det 10 prosentpoeng færre som svarte Ja i gruppe 3. Når gruppe 3 gjennomførte Eksperiment 4 resulterte utfallet i Nei, og dette kan ha preget innsatsene i det neste eksperimentet. Et lignende funn kan sees i Eksperiment 14, hvor gruppe 3, som også gjorde Eksperiment 13, svarte mindre Ja enn gruppe 2. Dette kan være fordi denne gruppen fikk erfarninger som gjorde det lettere å differensiere mellom de ulike kombinatoriske modellene, og dermed svare mer rett. For de gjeldende eksperimentene var det ingen begrunnelser som viste mot tilgjengelighetsheuristikker. Dette trenger ikke bety at slik heuristikk ikke preget veddingen, da det kan være underbevisst. Det gjør det likevel vanskelig å konkludere med at disse eksperimentene kartla tilgjengelighetsheuristikk.

5.2.4 Betingede hendelser – 9, 10, 11 og 12

Eksperimentene her tar for seg misoppfatninger som henger sammen med betingede hendelser. Mesteparten av misoppfatningene kommer til uttrykk ved en feilaktig kombinatorisk modell, men stammer fra misoppfatninger rundt hva som er den betingende faktoren.
Eksperiment 10 ble utformet slik at det skulle være ganske enkelt å se hva som var den betingende faktoren, og lignet i så måte på spørsmål 1 i Del 1 av surveyen. Av de gruppene som gjennomførte Eksperiment 10 svarte 87 % riktig på kartleggingsspørsmålet. Det ser ut som at de fleste klarte å anvende den samme matematiske ferdigheten også i spillet, og 86 % veddet på Nei, som var det mest sannsynlige utfallet (med $p = 1/3$ for Ja). 47 % av respondentene estimerte sannsynligheten, og av disse estimerte 87 % $p < 0,5$, som var riktig. Totalt estimerte 63 % av de estimerende helt riktig, $p = 1/3$. Disse resultatene skiller seg ut fra de resterende

resultatene, og det kan virke som om dette eksperimentet var enklere å avkode hva gjelder betingende faktor.

Eksperiment 11 er ganske likt Eksperiment 10, kun forskjellig i at man ikke får vite hva som blir trukket først. Eksperiment 12 er en tredje variant hvor man får vite hva som blir trukket sist, og skal så bestemme hva som sannsynligvis ble trukket først. Det som er den avgjørende faktoren i alle disse tre eksperimentene er at man må avgjøre om man vet noe nyttig; gjør man det vil man kunne betinge utfallsrommet. I Eksperiment 11 vet man ingenting, og derfor er $p = 0,5$ selv etter trekket. Den tilhørende fordeling av Ja og Nei var 30:14, og en slik fordeling er veldig usannsynlig. Dersom alle disse visste at $p = 0,5$ var sjansen for en slik fordeling $2,3$ %, som er en signifikant lav verdi for vår nullhypotese. Nok en gang er dette en statistisk indikasjon på at respondentene ikke oppfatter at $p = 0,5$. Dette støttes av sannsynlighetsanslagene, hvor 22 % estimerer $p < 0,5$; av disse tror 75 % at $p = 1/3$. Dette kan forstås som at respondentene har betinget på trekket, og baserer seg på at det bare er tre sjetonger igjen. De resterende 78 % estimerer at $p = 0,5$, som er riktig, og fordelingen av Ja og Nei blant disse er innenfor toleransenivået for hypotesen. Om den estimerende andelen av respondentene er representativ for resten, kan det tyde på at mange faktisk forstår at $p = 0,5$.

Den tredje varianten, Eksperiment 12, utfordrer respondentenes forståelse av kronologi. Som forklart tidligere vil en misoppfatning her bygge på at det ukjente trekket ble gjort først, og derfor er upåvirket av det trekket som man vet hva er; sannsynligheten er derfor 50/50. Men siden man vet noe mer kan utfallsrommet likevel begrenses ved betingning, og man kan se at sannsynligheten må være 1/3 eller 2/3, alt ettersom hvilken farge som ble trukket sist. Et ikke-rigorøst logisk argument for tilfellet hvor sannsynligheten var 2/3 ble gitt av en respondent: «Siden den andre var blå vil jeg tro at det mest sannsynlig var flest blå igjen i posen som betyr at han trakk hvit først». Motivasjonen for dette resonnementet er ekvivalent med det i Eksperiment 10, men det kan late til at færre ser denne sammenhengen: færre vedder riktig. Ca. 71 % vedder riktig, som er en reduksjon på 15 prosentpoeng i forhold til Eksperiment 10. 57 % har rett når de anslår hva som er mest sannsynlig utfall, men bare 26 % estimerer riktig sannsynlighet; som er en nedgang på 37 prosentpoeng fra Eksperiment 10. I tillegg kan det sees at like mange, 26 %, estimerer at $p = 0,5$, som kan vitne om en misoppfatning. Siden misoppfatningen går ut på at man tror $p = 0,5$ vil også noen av de med misoppfatning ende opp med å vedde på riktig utfall, og dermed bidra til en mer positiv statistikk, til tross for misoppfatningen. Her er det ingen tolkning å hente for nullhypotesen.

Det siste eksperimentet som dediseres til betingnings-misoppfatnigner er Eksperiment 9. 73 % vedder på riktig utfall, men også her gjør misoppfatningen at man betinger slik at man tror $p = 0,5$, og dette kan gjøre at man vedder riktig, uten å gjøre det av rett grunn. Respondentenes

estimat støtter mistanken om at dette er tilfellet. Dette er et av de eneste eksperimentene hvor det er flere feilaktige utfallstilnærminger enn riktige: 39 % estimerer $p = 0,5$ og 22 % tilnærmer $p < 0,5$. Av de som tilnærmer $p > 0,5$ estimerer 75 % riktig sannsynlighet. Dette tilsvarer 32 % av alle som oppgir estimat. Mindre enn 1 av 3 vedder altså riktig på dette eksperimentet av rett grunn, dersom de 62 prosentene som anslo sannsynligheten er representative. Av de 25 prosentene som tilnærmet riktig, men estimerte feil, svarte alle at $p = 3/4$, som kan forstås som at de har misforstått om det var med tilbakelegging eller ikke.
Av de som svarte $p = 0,5$ gis det noen begrunnelser som er med på å stadfeste at feilsvaret faktisk skyldes en misoppfatning:

> «Når man trekker blå er HH ute»
> «Like stor sannsynlighet for å få hvit og blå»

Begge disse sitatene viser at respondentene her har betinget på poser, og bare går ut fra at det vil være like stor sjans for at den gjenværende sjetongen er hvit som blå. Det fremstår derfor som tydelig at elevenes misoppfatninger blir avslørt.

Man kan også se at vansker med å betinge har gjort at respondenter, som ikke har misoppfatninger, likevel estimerer feil sannsynlighet, fordi de ikke definerer utfallsrommet og mulighetene riktig. I Eksperiment 1 er det flere som oppgir at $p = 1/36$, som kan stamme fra at de kjenner igjen at med to terninger finnes det 36 muligheter. Dette eksperimentet krever derimot at minst én 5-er dukker opp, og da er ikke lenger den ønskede sannsynligheten $\mathrm{P}(2 \cup 5) = 2/36$, men $\mathrm{P}(2|5) = 2/11$. Lignende tilfeller finnes i Eksperiment 15, hvor mange oppgir $p = 1/3$. Dette som kan stamme fra at man forstår spørsmålet som $\mathrm{P}(3 \cap 4) = 2/6$, men spørsmålets ordlyd betinger utfallene til å kun inkludere 1, 3 og 4 $\Rightarrow \mathrm{P}(3 \cap 4|1 \cap 3 \cap 4) = 2/3$.

5.2.5 Ekviprobabilitet – 1, 15 og 17

Eksperimentene i denne seksjonen var utformet slik at det ikke skulle kreve noen utregninger for å vurdere hvilke utfall som var mest sannsynlig. Feilsvar ble da tiltenkt å skyldes at man ikke oppfattet nyansene og anså alle utfall som like sannsynlige. For Eksperiment 1, 15 og 17 var sannsynligheten for et utfall tilsvarende ja henholdsvis 2/11, 2/3 og 1/4, og den forventede indikatoren på en misoppfatning i alle disse tilfelle var da at man enten veddet på det motsatte eller estimerte $p = 0,5$. Fordelingen av innsatser viste at disse nyansene ikke var åpenbare for alle, og henholdsvis 33 %, 25 % og 45 % veddet feil på disse oppgavene. For Eksperiment 1 og 15 gis det lite oppklaring i dette fra estimat og begrunnelser. For Eksperiment 1 tilnærmet 7 av 8 $p < 0,5$. Riktig nok klarte ingen å estimere eksakt, men uansett var ikke dette resultatet representativt for de 33 prosentene som veddet på Ja. For Eksperiment 15 var det for få respondenter til å gjøre noen statistiske tolkninger. Svarene i Eksperiment 17 belyste derimot litt mer av respondentenes vedding. 22 personer ga estimat og 8 av disse, 36 %, tilnærmet $p < 0,5$, som var riktig. 7 av disse 8 estimerte helt riktig sannsynlighet. De resterende 14 respondentene estimerte at sannsynligheten var 0,5. Dette viser at de fleste estimerende respondentene tolket to ulike sannsynligheter som like store. Om dette er representativt for alles vedding, vil det si at selv om «bare» 45 % vedder feil, er det mørketall som skjules av misoppfatningens fordeling.

5.2.6 Ikke-rasjonell overbevisning – 16

Det ingen bestemte misoppfatninger som knyttes til Eksperiment 16, men formålet var å se hvorvidt respondentene klarte å dedusere at sannsynligheten var 0,5. Det krevdes ingen utregninger, og svarene var heller ikke antatt å bygge på representativitetsheuristikker. Som med de andre eksperimentene med $p = 0,5$ må dette sees ut fra fordelingen av Ja og Nei. Det man kan se er at denne fordelingen er en usannsynlig skjev, 34 Ja og 10 Nei. Sannsynligheten for en slik fordeling, dersom alle visste at $p = 0,5$, er $0,04$ %, en signifikant lav verdi etter det bestemte signifikansnivået. Fra estimatene kan det sees at 11 av 18 respondenter gjenkjenner sannsynligheten som 0,5, mens 7 av 18 bommer. Skjevfordelingen er lik i begge klassene i den aktuelle gruppa, som peker på at dette ikke var et ekstremt tilfelle i én av klassene. En respondent viser en ikke-rasjonell tro når vedkommende estimerte $p = 0,7$ og begrunnet: «Tror du må prøve flere ganger for å få 1 eller 2.»

5.2.7 Kombinatorikk – 13 og 14, m.m.

Det var spesielt to eksperiment, 13 og 14, som var myntet på å vurdere respondentenes kombinatoriske ferdigheter og misoppfatninger. Begge disse eksperimentene kan forstås som permutasjonsmodeller, hvor Eksperiment 13 tilsvarer uten tilbakelegging og Eksperiment 14 tilsvarer med tilbakelegging. For begge eksperimentene faller kanskje naturlig å tolke dem som kombinasjoner, men siden to av sjetongene var identiske ville man måtte se det som en ikke-uniform sannsynlighetsfordeling. Dette er utenfor hva de fleste elevene har erfaring med, og det ble derfor lettere å forstå det som permutasjoner hvor man skjelner mellom de to identiske sjetongene når man lister opp mulighetene. Misoppfatningen i denne oppgaven springer ut fra nettopp denne problematikken, hvor respondentene estimerer $p = 0,5$ fordi de ser det som to kombinasjoner (B, H) og (H, H), uten å se at den første vil opptre dobbelt så hyppig som den andre. Jeg har ikke regnet dette som ekviprobabilitetsbias siden det kan ha opphav i utregningene, ikke den stokastiske forståelsen. En slik feilaktig kombinatorikk forstås som ordningsfeil, forårsaket av objektsfeil, jamføre avsnitt 3.2.1. 76 % veddet på riktig utfall, men ut fra estimatene virker dette heller tilfeldig. Kun én av de 11 estimerende klarte å regne ut sannsynligheten, og bare én annen tilnærmet $p > 0,5$ som var riktig. Av de resterende estimerte 5 personer $p = 0,5$, indikerende for misoppfatningen.

Eksperiment 14 var ganske likt Eksperiment 13, men var til forskjell med tilbakelegging. Sannsynligheten for Ja var $4/9$, og man burde derfor vedde på Nei. Her ville en misoppfatning tilsi at respondenten svarte Ja, og begrunnet det med et estimat av $p = 2/3$. Å estimere 2/3 utfra manglende tilbakelegging vil kunne se ut som at $\mathrm{P}(1.H, 2.B|1.H) + \mathrm{P}(1.B, 2.H|1.B) = 2/3 \cdot 1/2 + 1/3 \cdot 1 = 2/3$, som er det riktige resultatet for Eksperiment 13. 34 % av respondentene veddet på Ja, noe som stemmer overraskende godt med at 33 % av de estimerende tilnærmet $p > 0,5$. 7 av disse 8 % anslo $p = 2/3$, og en av respondentene begrunnet dette anslaget med: «Siden det er flere like enn ulike tror jeg det er større sannsynlighet for at han trekker to like.» Hele 58 % tilnærmet $p < 0,5$, men bare 3 av 14 klarte å regne ut sannsynligheten eksakt. Av de som ikke klarte å regne eksakt estimerte tre av dem $p = 2/9$, som kan bety at respondentene har oppfattet ordningen og tilbakelegget, men har gjort dette usystematisk, og dermed glemt to permutasjoner. Også her svarte to stk. at $p = 0,5$, noe som tyder på samme resonnement som i Eksperiment 13, og kan peke mot at eleven har blandet både tilbakelegging og ordning.

I flere av de eksperimentene som har blitt gått gjennom tidligere kan det se ut som at det ikke bare var respondentenes misoppfatninger som stod i veien, men også deres kombinatoriske ferdigheter. I Eksperiment 17 for eksempel vil de kombinatoriske ferdighetene ha kunnet stå i veien for veddingen. Her var det to som estimerte $p = 1/3$, og dette kan tolkes som å bygge på at man så på ordninger av de oppgitte permutasjonene, og da så tre mulige ordninger – $(K, K), (M, K)$ og (K, M).

5.2.8 Respondentenes forhold til anslag

Fra tabell 4 fremkommer det at 71 % av respondentene i snitt vedder på riktig utfall i hvert eksperiment. Det er også en moderat korrelasjon, 0,46, mellom hvor stor sannsynligheten er og hvor mange som vedder riktig. Alt dette peker på at elevene har en oppfatning av hva som er mest sannsynlig, og hva dette medfører for gjettingen. Likevel fremstår det som om elevenes forståelse av egne estimat er dårlig; i mange av eksperimentene kan det sees at respondenter vedder mot egne estimat! Dette kan eksempelvis sees i Eksperiment 7, hvor halvparten av dem som estimerer $p < 0,5$ vedder på Ja. Likeens ser en at 6 av 8 vedder på Nei i Eksperiment 14, selv om de estimerte $p > 0,5$. Som sett tidligere opptrår det usannsynlige fordelinger hos dem som estimerer $p = 0,5$ også. I Eksperiment 16 var det en skjev fordeling blant alle som veddet, men det så man også hos dem som faktisk gjennomskuet at sannsynligheten var 0,5. Av 11 som oppga dette, veddet 10 av dem på Ja. Sannsynligheten for en slik fordeling forekommer naturlig er på bare 2 %, som er innenfor signifikansnivået.

En tolkning av dette fenomenet er at respondentene ikke tror på egne estimat, og bare oppgir et estimat som *oppleves* å inneha riktige element utfra konteksten, uten at de tenker at de er korrekte. Dette vil i så fall kunne henge sammen med beløpet som veddes og kommer til uttrykk i den tredjedelen som i snitt ikke vedder maksbeløpet. En annen tolkning er at man her kan se mønster av spillteoretiske handlinger; man har vurdert sannsynligheten for et utfall korrekt, men er villig til å ta sjansen på det minst sannsynlige, fordi man vurderer det som mer gunstig for egen posisjon i spillet.

6 Konklusjon

Undersøkelsens funn løfter frem flere interessante aspekter ved undervisningsopplegget. Et av de positive poengene er at noen av eksperimentene gir veldig konkrete indikatorer på misoppfatninger. I noen av eksperimentene, spesielt de tilknyttet betingning, indikterte Ja-/ Nei-variabelen alene at elevene ikke oppfattet sannsynligheten, og dette ble bekreftet av de oppgitte estimiatene. Generelt oppfattes Ja-/ Nei-variabelen som å være en god indikator i de tilfellene hvor misoppfatningen tilsvarer et estimat $p \neq 0,5$. I tilfellene hvor man kunne forvente at elever med misoppfatninger anslo at $p = 0,5$ var man avhengig av estimatene for å kunne gjøre en tolkning av elevenes forståelse. Selv om estimatene med data som gjorde det lettere å se mønster i svarene, kunne også disse oppfattes som mangelfulle. I noen tilfeller så man en usannsynlig skjev fordeling av Ja-og Nei-svar blant dem som estimerte sannsynlighetene til å være $0,5$, uten at det gikk an å konkludere med hva som ledet til denne fordelingen. I andre tilfeller ble det gitt sannsynlighetsestimat som hverken tilsvarte utfallets sannynlighet eller estimatet som indikerte den forventede misoppfatningen. Estimatene var belysende for innsatsene, men var for tvetydig til å gi konklusjoner om forekomster av misoppfatninger. Slik ble det tydelig at de tallfestede dataene i seg selv ikke ga grunnlag for troverdige tolkninger av misoppfatninger i klasserommet. I eksperiment hvor større andeler av elevene ga begrunnelser for estimatene sine var dette til hjelp for tolkningene. Slike begrunnelser var med å belyse hendelser som analysen ikke hadde fanget opp uten. Hverken innsatsene eller estimatene hadde fanget opp representativitetsheuristikken, men ved hjelp av begrunnelsene kunne det konkluderes med at noen elever hadde misoppfatninger som kunne tilegnes representativitet.
At dataene generelt åpner opp for en del tolkning uten å være konkluderende gjør at spillet ikke oppfattes som et godt diagnostisk verktøy. Noe av formålet med dette verktøyet var at det skulle være en effektiv måte å kartlegge flere misoppfatninger på. Denne effektiviteten vurderes som å ha gått på bekostning av forskningsdesignets kvalitet. Dette oppfattes som problematisk, fordi man ikke vet hvilke del av forskningsdesignet som skal tilegnes de ulike funnene. Skyldes de uforventede fordelingene av innsatser misoppfatninger, eller forvirring

rundt en uklar eksperimentformulering? Gir elevene bestemte estimat fordi det er det de tror, eller fordi det var det tidsbegrensningen tillot dem å beregne? Veddet elevene store beløp fordi utfallet opplevdes veldig sannsynlig, fordi deres spillintuisjon tilsa at sannsynligheten ikke trengte å være stor før det var fornuftig å vedde maks, eller var beløpet helt vilkårlig? Om en slik undersøkelse skal gjøres på ny vil det derfor være en fordel å sikre en jevnt større svarprosent hvor elevene må redegjøre for valgene sine.

Dersom undervisningsopplegget skal legge stor vekt på estimat og begrunnelser vil det kunne fremstå som en hvilken som helst annen vurdering, og da er spørsmålet om det er verdt å innføre. Selv om spillet står frem som et tvetydig og unøyaktig diagnostisk verktøy, betyr det ikke at det er uten verdi for undervisningen. Slik de enkeltstående resultatene er nå fremstår de med svak troverdighet om de skal brukes for å vurdere kompetanser. Sett at resultatene settes i en annen kontekst vil det likevel kunne være en ressurs. Om man gjennomfører undervisningsopplegget i mindre bolker, gjerne seksjonert etter misoppfatninger, og stopper opp og snakker om eksperimentene forløpende vil dette være en frisk inngang til en faglig samtale. Samtalen kan starte så enkelt som «hvorfor veddet du slik?» Når noen da vurderer eksperimentet til å betinge på den ene faktoren, og andre vurderer eksperimentet til å betinge på den andre faktoren vil dette kunne ende opp i en eksplorerende samtale for å finne ut hva som er riktig tanke. En slik samtale vil kunne ende opp med å adresse misoppfatninger direkte, og være en ressurs for å overkomme dem. Som for andre seriøse spill settes læreren her i en sentral rolle for å utnytte de lærende øyeblikkene som spillet fasiliterer. Spillets estetikk kan også virke fascinerende og avvæpnende for elever med lav faglig selvtillit, og åpne opp for innspill fra elever som vanligvis ikke ville bidratt til diskursen. Om dette ble gjort i begynnelsen av en periode vil man kunne refere til spillet og bruke det som referansepunkt for den formaliserte teorien: «slik kunne sett at din innsats var den mest fornuftige!»
Ved en slik bruk gis det også et utgangspunkt for å samtale om kritistk tenkning og å fatte beslutinger. For noen vil det kunne være nyttig å se hendelser ikke trenger å være så komplekse før det er lett å ta dårlige valg. Engasjementet som ble utvist av respondentene vitner om at spillet vil kunne fungere godt som et engasjerende tilskudd til undervisningen. Anvendelsen av et slik verktøy vil også kunne øke lærerens bevissthet på behove for mer målrettet undervisning for å håndtere misoppfatninger.

At man var avhengig av estimat og begrunnelser for å tolke svarene peker på at spillet trenger disse modifiseringene for å gi læreren informasjon om de rådende oppfatningene i sannsynlighet. For å gjøre en videre vudering av spillet som ressurs for diagnostikk vurderer jeg det som fornuftig at spillet trianguleres med et annet verktøy for å få referansepunkt å se funnene opp mot. På en slik måte vil man kunne vurdere funnenes troverdighet, og kunne avgjøre om dataene spillet genererer kan brukes for å kartlegge misoppfatninger i sannsynlighet.

7 Referanser

Abrahamson, D. (2012, 01). Seeing chance: perceptual reasoning as an epistemic resource for grounding compound event spaces. in r. biehler and d. pratt (eds.), probability in reasoning about data and risk [special issue]. *ZDM: The international Journal on Mathematics Education*, *44*, 869–881.

Ang, L.H. & Shahrill, M. (2014). ”identifying students’ specific misconceptions in learning probability”. *”International Journal of Probability and Statistics”*, *3*(2), 23-29.

Angell, C., Bungum, B., Henriksen, E.K., Kolstø, S.D., Persson, J. & Renstrøm, R. (2018). *Fysikkdidaktikk*. Cappelen Damm Akademisk.

Apt, C.C. (1970). *serious games”*. University press of America.

Batanero, C. & Sanchez, E. (2005). What is the nature of high school students’ conseptions and misconceptions about probability? I G.A. Jones (red.), *Exploring probability in school: Challenges for teaching and learning* (s. (241-266)). New York: Springer.

Biesta, G. (2009, 12). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*(21), 33-46.

Brekke, G. (2002). *Introduksjon til diagnostisk undervisning i matematikk: Kartlegging av matematikkforståelse*. Utdanningsdirektoratet.

Brevik, L.M. & Blikstad-Balas, M. (2014). Blir dette vurdert, lærer?”. Om vurdering for læring i klasserommet. I E. Elstad & K. Helstad (red.), *profesjonsutvikling i skolen”* (s. 1-13). Universitetsforlaget.

Christoffersen, L. & Johannesen, A. (2010). Kap. 15: Analyse av kvantitative data. fordeling av én egenskap - univariat analyse. I *Forskningsmetode for lærerutdanningene* (s. 141-149). Abstrakt forlag.

Creamer, E.G. (2016). A primer about Mixed Methods Research in an Educational Context. *International Journal of Learning, Teaching, and Educational Research*, 1-13.

Dewilde, J. (2020). *Normer for kvalitet i forskning.* (PowerPoint-lysbilder)

Egenfeldt-Nielsen, S., Meyer, B. & Sørensen, B.H. (red.). (2011). *Serious games in education.* Aarhus University Press.

Engh, R. (2011). *Vurdering for læring i skolen – på vei mot en bærekraftig vurderingskultur.* Høyskoleforlaget.

Falk, R. (1986).

I R. Davidson & J. Swift (red.), (s. 292-297). The Second International Committee on Teaching Statistics. Hentet fra https://iase-web.org/documents/papers/icots2/Falk.pdf

Fangen, K. (2011). Deltagende observasjon. I K. Fangen & A.-M. Sellerberg (red.), *Mange ulike metoder* (s. 37-56). Gyldendal Akademisk.

Farber, M. (2015). *Gamify your classroom. A field guide to game-based learning.* Peter Lang Publishing, Inc.

Felicia, P. & Egenfeld-Nielsen, S. (2011). State of the art. I Egenfeldt-Nielsen, Simon and Meyer, Bente and Sørensen, Birgitte Holm (red.), *Serious games in education* (s. 21-46). Aarhus University Press.

Firebaugh, G. (2008). Ch. 1: The First Rule. There Should Be the Possibility of Surprise in Social Research. I *Seven Rules for Social Research* (s. 1-30). 2008 Princeton University Press.

Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *28*(1), 96–105. Hentet fra http://www.jstor.org/stable/749665

Furuseth, I. & Everett, E. (2012). Kap. 8: Lettere sagt enn gjort – å utforme et metodisk opplegg for oppgaven. I *Masteroppgaven. hvordan begynne og fullføre* (s. 127-144). Universitetsforlaget.

Gauvrit, N. & Morsanyi, K. (2014). The equiprobability bias from a mathematical and psychological perspective. *Advances in cognitive psychology*, *10*(4). doi: \url{https://doi.org/10.5709/acp-0163-9}

Green, D.R. (1986). *Childrens's understanding of randomness: report of a survey.* https://www.stat.auckland.ac.nz/~iase/publications/icots2/Green-2.pdf/.

Grønmo, S. (2004). Kap. 12: Strukturert utspørring. I *Samfunnsvitenskapelige metoder* (s. 191-211). Fagbokforlaget.

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81-112.

Hirsch, L.S. & O'Donnell, A.M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education*, *9*(2), 1-22.

Hopfenbeck, T.N. (2016). Kap. 5: Vurderingskompetanse & kap. 11: Bruk og misbruk av vurderingsdata. I *"Å lykkes med elevvurdering"* (s. 75-84, 140-154). Fagbokforlaget.

Hunicke, R., Leblanc, M. & Zubek, R. (2004, 01). Mda: A formal approach to game design and game research. *AAAI Workshop - Technical Report*, *1*.

ILS. (2020, 22. 12). *Om TIMSS.* https://www.uv.uio.no/ils/forskning/prosjekter/timss/om-timss/index.html/. (Institutt for læring og skoleforskning v/ UiO)

Jessen, C. (2011). Learning games and the disruptive effects of play. I Egenfeldt-Nielsen, Simon and Meyer, Bente and Sørensen, Birgitte Holm (red.), *Serious games in education* (s. 153-169). Aarhus University Press.

Johnson, B.R. (2003). ”validity of research results in quantitative, qualitative and mixed research”. I B.R. Johnson & L. Christensen (red.), *educational research: Quantitative, qualitative, and mixed approaches”* (s. 277-316). Sage.

Jones, G.A. & Thornton, C.A. (2005). An Overview of Research into the Teaching and Learning of Probability. I G.A. Jones (red.), *Exploring Probability in School* (s. 65-92). Springer.

Jopling, M. (2019). Chp. 6: Using quantitative data. I M. Lambert (red.), *Practical research methods in education. an early researcher's critical guide* (s. 55-66). Routledge.

Kaarstein, H., Radišić, J., Lehre, A., Nilsen, T. & Bergem, O. (2019). *Timss 2019. kortrapport* (Teknisk rapport). Oslo: Institutt for lærerutdanning og skoleforskning, Universitetet i Oslo.

Kahneman, D. (2013). *Tenke, fort og langsomt.* Oslo: Pax Forlag A/S. ((Opprinnelig utgitt i 2011))

Klette, K. (2013). Hva vet vi om god undervisning? Rapport fra klasseromsforskningen. I R.J. Krumsvik & R. Säljö (red.), *Praktisk-pedagogisk utdanning – en antologi* (s. 173-193). Fagbokforlaget.

Kleven, T.A. (2014). data og datainnsamlingsmetoder”. I T.A. Kleven, F. Hjardemaal & K. Tveit (red.), *”innføring i pedagogisk forskningsmetode (2. utg.)”* (s. 27-47). Fagbokforlaget.

Konold, C. (1991). Understanding students' beliefs about probability. I E. von Glasersfeld (red.), *Radical constructivism in mathematics education* (vol. 7, s. (139-156)). Dordrecht: Springer.

Kunnskapsdepartementet. (2017). *Overordnet del – verdier og prinsipper for grunnopplæringen.* https://www.udir.no/lk20/overordnet-del/.

Larsen, A.K. (2017). Om samfunnsvitenskapelig metode. I A.K. Larsen (red.), *En enklere metode. veiledning i samfunnsvitenskapelig metode* (s. 17-31). Fagbokforlaget.

Lesh, R.A. & Zawojewski, J. (2007). problem solving and modeling”. I *Second Handbook of Research on Mathematics and Teaching and Learning* (s. 763-804). Information Age Publishing.

Manger, T. (2013). Motivasjon for skulearbeid. I R.J. Krumsvik & R. Säljö (red.), *Praktisk-pedagogisk utdanning – en antologi* (s. 145-165). Fagbokforlaget.

McGonigal, J. (2010). *Gaming can make a better world.* https://www.ted.com/talks/jane_mcgonigal_gaming_can_make_a_better_world#t-167303/.

Mellin-Olsen, S. (1981, 7). Instrumentalism as an educational concept. *Educational Studies in Mathematics*, *12*(3), 351-367.

National Research Counsil. (2001). THE STRANDS OF MATHEMATICAL PROFICIENCY. I J. Kilpatrick, J. Swafford & B. Findel (red.), *Adding it up. Helping children learn* *mathematics.*

Niss, M. (1996). Goals of mathematics teaching. I A.J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (red.), *International handbook of mathematics education* (vol. 4, s. 11-47). Dordrecht: Springer.

Norsk Lektorlag. (2018, 29. 12). *Nye læreplaner – tilbakemelding fra lektorene.* https://www.norsklektorlag.no/nyheter/nye-laereplaner-tilbakemelding-fra-lektorene/.

Nygaard, O. & Zernichow, A.G. (2006). Den blokkerende misoppfatning. *Spesialpedagogikk*(4), 34-38. https://home.uia.no/olavn/blokkerende.pdf/.

Regjeringen. (2019, 11). *Nye læreplaner skal gi elevene tid til mer fordypning.* https://www.regjeringen.no/no/aktuelt/nye-lareplaner-skal-gi-elevene-tid-til-mer-fordypning/id2678138/. ([Pressemelding])

Regjeringen. (2021). *Lov om grunnskolen og den vidaregåande opplæringa.* https://lovdata.no/dokument/NL/lov/1998-07-17-61.

Rowbottom, D.P. (2015). *Probability.* Cambridge: Polity.

senter for forskningsdata, N. (u.d.-a). *Hvordan gjennomføre et prosjekt uten å behandle personopplysninger?* https://www.nsd.no/personverntjenester/oppslagsverk-for-personvern-i-forskning/hvordan-gjennomfore-et-prosjekt-uten-a-behandle-personopplysninger. (Besøkt 28.05.21)

senter for forskningsdata, N. (u.d.-b). *Informasjon til deltakerne.* https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/sjekkliste-for-informasjon-til-deltakerne. (Besøkt 28.05.21)

Shaughnessy, J.M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at collage level. *Educational Studies in Mathematics*, 295-316.

Shoenfeld, A.H. (1987). What's all the fuss about metacognition. I *"cognitive science and mathematics education"* (kap. 8). LEA.

Silverman, D. (2011). Chp. 2: Designing a research project. I *Interpreting qualitative data* (4. utg., s. 27-56).

Sjøberg, S. (2014). *Naturfag som allmenndannelse - en kritisk fagdidaktikk.* Oslo: Gyldendal.

Smith, K. (2009). Samspillet mellom vurdering og motivasjon. I Dobson, S. and Eggen, A. B. and Smith, K. (red.), (s. 23-39). Gyldendal Akademisk.

Solbakken, S.S. (2019). *Statistikk for nybegynnere.* Fagbokforlaget.

Säljö, R. (2013). Støtte til læring – tradisjoner og perspektiver. I R.J. Krumsvik & R. Säljö (red.), *Praktisk-pedagogisk Utdanning – En Antologi* (s. 53-78). Fagbokforlaget.

Sørensen, B.H. (2011). Educational design for serious games. I Egenfeldt-Nielsen, Simon and Meyer, Bente and Sørensen, Birgitte Holm (red.), *Serious games in education* (s. 101-124).

Aarhus University Press.

Thomas, A. (2018). *the effective use of game-based learning in education".* https://www.youtube.com/watch?v=-X1m7tf9cRQ&ab_channel=TEDxTalks/.

Ulvik, M. & Sæverot, H. (2017). Pedagogisk danning. I R. Krumsvik R. J.; Säljö (red.), *Praktisk-pedagogisk utdanning – en antologi* (s. 31-48). Bergen: Fagbokforlaget.

Utdanningsdirektoratet. (2013). *Læreplan i matematikk fellesfag (mat1-04).* http://data.udir.no/kl06/MAT1-04.pdf.

Utdanningsdirektoratet. (2019a, 15. 11). *Læreplan i matematikk 1–10 (mat01-05).* https://data.udir.no/kl06/v201906/laereplaner-lk20/MAT01-05.pdf?lang=nob&fbclid=IwAR1RFcDvmgJ62I07-Pz58LPN6RZBrOQGxa0N4bTa5AvRtQqVzFlya9dmJnI.

Utdanningsdirektoratet. (2019b, september). *Matematikk fellesfag 1-10 – oppsummering av høringen.* https://www.udir.no/laring-og-trivsel/lareplanverket/fagfornyelsen/oppsummeringer-av-7000-horingssvar-om-nye-lareplaner/matematikk-grunnskole--oppsummering-av-horingen/. (Besøkt 15.03.21)

Utdanningsdirektoratet. (2020, 26. 5). *Læreplan i matematikk for samfunnsfag.* https://data.udir.no/kl06/v201906/laereplaner-lk20/MAT04-02.pdf?lang=nob.

Wikipedia. (2019). https://no.wikipedia.org/wiki/Kombinatorikk.

Wikipedia. (2021, 04). *Kelly criterion.* https://en.wikipedia.org/wiki/Kelly_criterion.

Öçal, M.F. (2018). The case of time axis fallacy: 11th grade students' intuitively-based misconception in probability and teachers' corresponding practices. *Journal of Qualitative Research in Education*, *6*(3), 86-105.