

**UNIVERSITETET I OSLO**  
**Institutt for informatikk**

**Hvor ser brukeren? En  
analyse av eye tracking-  
opptak fra bruk av  
Skatteetatens nettsider**

**Masteroppgave**  
60 studiepoeng

Nora  
Raum

**4. mars 2013**





# **Hvor ser brukeren?**

**En analyse av eye tracking-opptak fra bruk  
av Skatteetatens nettsider**

Nora Raaum

Masteroppgave

Institutt for informatikk

UNIVERSITETET I OSLO

2013

© Nora Raaum

2013

Tittel

Nora Raaum

<http://www.duo.uio.no/>

Trykk: Reprosentralen, Universitetet i Oslo

# Abstrakt

Eye tracking kan oversettes til blikksporing på norsk, og er en teknologi som sporer menneskets øyebevegelser, og til enhver tid vet akkurat hvor mennesket som blir sporet ser.

Hensikten med denne oppgaven var å se på øyebevegelsene til brukere som utførte oppgaver på Skatteetatens nettsider. Basert på opptak av blikket ønsket man å finne ut om det var mulig å se forskjellige mønster hos personer med bakgrunn fra fagfeltet interaksjonsdesign og personer uten slik bakgrunn. Testdeltakerene bakgrunn ble bestemt etter egenopplyst erfaring fra interaksjonsdesign, webdesign, eller brukervennlighet på nett – i studier eller arbeid. Det var også ønskelig å reflektere rundt teknologien eye tracking.

For å kunne observere brukernes blikkmønster ble eye tracking-teknologi benyttet, en eye tracker kalt Tobii T60XL.

Det ble utført et eksperimentelt forsøk under kontrollerte forhold på en lab i Skatteetatens lokaler i Oslo. 13 testdeltakere utførte oppgaver på Skatteetatens nettsider mens eye tracking-utstyret gjorde opptak av deres øyebevegelser. Opptakene ble senere bearbeidet i programvaren Tobii Studio. Det ble ved hjelp av programvaren hentet ut visuell data og statistikk fra syv av oppgavene, samt at opptakene ble manuelt gjennomgått av testleder. Dette var grunnlaget for analysen. I analysen ble det sett på hvordan de forskjellige testdeltakerene løste oppgaven, hvilke områder av skjermen de fokuserte på og hvor lenge, samt at de to gruppene av brukere ble sammenliknet.

Resultatene fra denne studien kan gi ytterligere innsikt i forskjell mellom utvikler og vanlig databruker. Videre kan resultatene gi interessant innsikt i bruk av offentlige nettsteder som inneholder mye informasjon. Studien kan til slutt være relevant og interessant for alle som jobber med eye tracking, og kanskje spesielt for de som vurderer å bruke denne teknologien i forskningsprosjekter eller bruksundersøkelser.



# Forord

Denne masteroppgaven på 60 studiepoeng er skrevet av student Nora Raaum for forskergruppen Design av informasjonssystemer ved Institutt for Informatikk, Universitetet i Oslo. Jeg har arbeidet med oppgaven over en tidsperiode på et drøyt år, fra januar 2012 til mars 2013. Det praktiske forsøket beskrevet i oppgaven ble utført sommeren 2012. Det har vært spennende og lærerikt arbeid, og jeg håper lesere av oppgaven vil finne den interessant.

Jeg vil gjerne rette en stor takk til veilederene mine, Alma Leora Culén og Guri Birgitte Verne, for god hjelp i startfasen, verdifull tilbakemelding på oppgaven underveis i skriveprosessen, og tålmodighet med meg når skrivingen har gått sakte.

Jeg vil også takke Skatteetaten for lån av lab med eye tracking-utstyr. Spesielt vil jeg takke min kontaktperson hos Skatteetaten, Laura Arlov, for all praktisk hjelp med utstyr, opplæring og tilrettelegging så det fungerte for mitt forsøk, og hjelp med rekruttering av deltakere. Hun delte også av sin erfaring med eye tracking-opptak, noe som var til stor hjelp spesielt i oppgavens startfase.

Til slutt vil jeg takke familie og kjæreste for moralsk støtte, og spesielt søsteren min for hjelp med korrekturlesing.





# Innholdsfortegnelse

1	Introduksjon .....	1
1.1	Bakgrunn .....	1
1.2	Motivasjon .....	2
1.3	Oppgavens mål .....	4
1.3.1	Problemformulering .....	4
1.3.2	Tanker bak problemformulering .....	4
1.4	Oppgavens struktur .....	6
2	Bakgrunn .....	7
2.1	Eye tracking .....	7
2.1.1	Historie .....	7
2.1.2	Dagens teknologi .....	9
2.1.3	Blikk som inputmetode .....	10
2.1.4	Tobii .....	11
2.1.5	Visualisering av data .....	11
2.1.6	Kritikk .....	13
2.2	Skatteetaten på nett .....	15
2.2.1	Nettstedets oppbygging .....	16
3	Litteratur .....	21
3.1	HCI .....	21
3.1.1	Samtaleanalyse .....	21
3.1.2	WWW .....	24
3.2	Brukstesting .....	25
3.3	Annen relevant forskning .....	27
3.3.1	Ekspert versus novise .....	27
3.3.2	Erotiske bilder .....	28
3.4	Relevans .....	29
4	Metode .....	31
4.1	Tidslinje .....	31
4.2	Eksperimentell forskning .....	32
4.3	Brukstesting .....	35
4.4	Feilmarginer .....	36

4.5	Spørreundersøkelse.....	39
4.6	Analyse av data.....	41
4.6.1	Analyse av skjermvideo fra opptak .....	41
4.7	Etikk.....	43
4.7.1	Personlig informasjon .....	43
4.7.2	Samtykke .....	44
5	Gjennomføring av studien.....	45
5.1	Valg av deltakere til forsøket.....	45
5.2	Spørreundersøkelse.....	46
5.2.1	Resultater fra spørreundersøkelsen .....	46
5.2.2	Gruppering av testdeltakere basert på bakgrunn .....	48
5.3	Utforming av testoppgaver .....	50
5.4	Laboppsett .....	52
5.5	Pilottest .....	53
5.5.1	Resultat av pilottest .....	53
5.6	Gjennomføring av tester .....	54
6	Analyse.....	57
6.1	Innhenting og bearbeiding av data.....	57
6.1.1	Prestasjon .....	57
6.1.2	Tobii Studio.....	57
6.1.3	Utvalg av oppgaver for videre analyse.....	62
6.2	Resultater .....	62
6.2.1	Forklaring til presentasjon av resultater .....	62
6.2.2	Oppgave 4 .....	64
6.2.3	Oppgave 11 .....	67
6.2.4	Oppgave 14 .....	72
6.2.5	Oppgave 19 .....	76
6.2.6	Oppgave 21 .....	81
6.2.7	Oppgave 22 .....	86
6.2.8	Oppgave 26 .....	90
6.3	Diskusjon rundt resultatene .....	94
6.3.1	Generelle observasjoner .....	94
6.3.2	Gruppe 1 versus gruppe 3.....	96

6.3.3	Faktorer som kan påvirke resultatene.....	98
6.4	Refleksjoner om eye tracking .....	100
6.4.1	Erfaring.....	100
6.4.2	Styrker .....	101
6.4.3	Svakheter.....	102
7	Konklusjon .....	105
7.1	Sammendrag .....	105
7.2	Konklusjon.....	106
7.3	Videre arbeid .....	107
	Litteraturliste .....	109
	Vedlegg .....	113



# 1 Introduksjon

## 1.1 Bakgrunn

Eye tracking kan oversettes til blikksporing på norsk, og er en teknologi som sporer menneskets øyebevegelser, og til enhver tid vet akkurat hvor mennesket som blir sporet ser. Det være seg en person som leser på papir, ser seg rundt i en butikk, eller det som er vanligst å bruke teknologien til i dag: navigere seg gjennom menyer og informasjon på en dataskjerm. Teknologien er ikke ny, den har i skrivende stund fantes i over 100 år (Reed og Meyer, 2006), men den forandres og utvikles stadig og har spesielt de siste årene blitt mer tilgjengelig og avansert. Man trenger ikke lenger feste utstyr på brukerens øyne eller hode, som var nødvendig i begynnelsen, og teknologien er responsiv og rask slik at den kan brukes i sanntid. Slik er det blitt et populært verktøy både for kommersielle aktører og akademiske studier. Det kan brukes på mange forskjellige måter: for å observere og måle bruk av teknologi, eller som input-metode for navigering i programmer i stedet for mer konvensjonell teknologi som mus og tastatur. Det ble høsten 2012 skrevet i norske aviser om et dansk firma som i 2013 vil lansere eye tracking-teknologi for mobile enheter, slik at man kan spille på smarttelefoner og nettbrett kun ved bruk av øynene. De bruker store ord om den nye bruken av teknologien, som «å revolusjonere mobilbransjen» og «morgendagens superstjerner» (Dagbladet, 2012). Selv om teknologien ikke er ny er den altså i det siste utviklet slik at den oppleves som nyskapende og spennende – og kanskje svært nyttig?

Innenfor akademia er det flere forskingsfelt som mener at eye tracking er akkurat det, og som benytter seg av teknologien. Det har blitt en populær metode innenfor temaet menneske-maskin-interaksjon (Poole and Ball, 2005), forkortet HCI etter engelske human-computer-interaction. HCI er et forskningsfelt der man oftest studerer noe som kan være vanskelig å måle: mennesker, og deres interaksjon med teknologi. Dermed er det ikke rart at eye tracking i årevis har vært et viktig verktøy innenfor HCI-forskning. Du kan se hvor en bruker styrer musepekeren sin, du kan se hvor han klikker, du kan spørre ham hva han la merke til på skjermen, eller hvor han så før han valgte et menyvalg, men ingen av disse tingene vil gi deg et korrekt og helhetlig bilde av hva brukeren gjorde og hvor han så. Data fra eye tracking kan gi ytterligere, objektiv informasjon om akkurat dette, som i ettertid kan analyseres på lik linje med annen forskningsdata.

Hvorfor er det relevant å vite akkurat hvor brukere ser? Med utgangspunkt i øye-tankehypotesen kan vi anta at det brukeren ser på er det brukeren prosesserer (Just og Carpenter, 1980). Hypotesen sier at det er direkte sammenheng mellom tiden man ser på et ord og tiden det tar å prosessere det. Den sier også at det ikke er noen vesentlig forsinkelse fra man ser på ordet til det prosesseres i hjernen. Hypotesen er basert på lesing av tekst, men vil også være gjeldende når brukere løser konkrete oppgaver på skjerm og derfor konsentrerer seg.

Duchowski (2003) kaller dette for visuell oppmerksomhet. Denne sammenhengen gjør måling av øyebevegelser og blikk relevant for å kartlegge atferd fordi vi da kan vite hva brukeren fokuserer på, ikke bare med synet, men med tankene. En ting å merke seg i forhold til denne hypotesen er at selv om man vet hvor brukeren fokuserer vet man ikke nødvendigvis hvorfor. På en nettside kan det være at brukeren fokuserer på et område fordi det er interessant, men det kan også være fordi det er forvirrende eller vanskelig å forstå (Moore og Churchill, 2011). Kontekst er veldig relevant her for å kunne analysere betydningen av blikkbevegelsene.

Naturlig nok er ikke alle like entusiastiske i henhold til eye tracking, og jeg vil i oppgaven også presentere negative aspekter og kritikk mot eye tracking.

I området mellom akademisk forskning og spill på smarttelefoner finner vi en stor gruppe som bruker eye tracking som verktøy i sitt arbeid: de som jobber med brukskvalitet ved og utvikling av nettsteder og programvare. Det kan være alt fra å gjøre nettsidene mer kommersielt suksessfulle – ved for eksempel å finne ut hvor det er mest hensiktsmessig å plassere reklame for at flest mulig skal se den, til å gjøre navigering mer brukervennlig. Brukervennlig navigering og presentasjon av informasjon er spesielt nødvendig på nettsider med mye viktig innhold som mange er nødt til å bruke i hverdagen, slik som nettsidene til offentlige etater og tjenester. Det er dette siste man skal se på i dette forsøk med eye tracking.

## 1.2 Motivasjon

Min interesse for oppgavens tema kommer fra både studier, arbeid og egen bruk av teknologi. Jeg har tatt flere studiepoeng innenfor gruppen Design av Informasjonssystemer og spesielt fagene «INF4260 – Human-computer interaction» og «INF5072 – Design av interaktive nettsteder» var interessante og ga meg relevant bakgrunnskunnskap for å kunne gjennomføre oppgaven. I flere år har jeg arbeidet deltid med IT-support ved Universitetet, og har hjulpet mange brukere med veldig varierende bakgrunnskunnskaper. Det har gitt meg innblikk i hvor

viktig brukervennlighet er når man utvikler datasystemer og nettsider, og hvor viktig det er at det er en form for kommunikasjon, eller et bindeledd, mellom brukere og utviklere. Mange en gang når jeg har veiledet brukere over telefon har jeg nok også ønsket meg muligheten til å forstå hva brukeren faktisk gjør og hvor de ser.

Ideen til oppgaven kommer opprinnelig fra prosjektet «Autonomi og automatisering i et informasjonssamfunn for alle» som ligger under forskningsgruppen Design av Informasjonssystemer ved Institutt for Informatikk, Universitetet i Oslo. Prosjektet baserer seg på målet om at samfunnet i dag skal ha digitale løsninger og hjelpemidler som ikke er for kompliserte for menigmann å ta i bruk, og vil bidra til å lage enklere og bedre systemer hos de som tilbyr offentlige tjenester (Institutt for Informatikk, 2012). På nettsiden til prosjektet står det følgende om temaene autonomi og automatisering i prosjektet:

*«En stor utfordring i et inkluderende informasjonssamfunn er forståelsen mellom autonomi og automatisering: Skaper tekniske løsninger en avhengighet og tap av selvråderetten, eller styrker de menneskets uavhengighet gjennom muligheten til å gjennomføre flere oppgaver selv?»*

*Utgangspunktet for dette prosjektet er at det eksisterer en likhet og en balanse mellom automatikk og autonomi. Vi ønsker å ta for oss hvordan fokuset på selvesøket etter denne balansen kan skjule det egentlige samspillet mellom de to faktorene.»*

(Institutt for Informatikk, 2012)

En av flere tjenester som er i fokus i prosjektet er de digitale tjenestene til Skatteetaten, en etat som har ansvaret for folkeregistrering, skatter og avgifter i Norge. De har derfor et samarbeid med etaten som jeg fikk nytte av i denne studien da de tilbød å låne ut lab med eye tracking-utstyr til masterstudent ved IFI.

Jeg fant beskrivelse av oppgaveideen på IFIs nettsider og syntes det hørtes spennende ut, i tillegg til at det var relevant for mine interesser og faglig kompetanse. Det var helt nytt for meg å bruke eye tracking, så det var en del ny kunnskap å sette seg inn i fra starten av. Før jeg begynte på oppgaven var jeg og veiledere på besøk hos Skatteetaten der kontaktperson i avdeling for Brukskvalitet viste oss laben og utstyret, og ga en kort innføring i hvordan det kunne brukes.

Det viste seg da at Skatteetaten var åpne for å la meg låne utstyret deres men velge spesifikk problemstilling og fokus for oppgaven min temmelig fritt. Etter å ha satt meg nærmere inn hva eye tracking gikk ut på og lest tidligere forskning og litteratur rundt emnet, konfererte jeg

både med veiledere på IFI og med kontaktperson hos Skattetaten. Jeg bestemte meg så for et studie av Skatteetatens nettsider med et fokus på kvalitativ analyse av eye tracking dataene.

Jeg mener denne oppgaven kan være interessant for flere. Skatteetaten jobbet med utvikling av nye nettsider mens jeg gjennomførte testene på skatteetaten.no, men utviklerne kan kanskje likevel få konkret nytte av resultatene i forbindelse med brukervennlighet på nettsidene. De jobber stadig for å forbedre brukskvaliteten på nettsidene sine og er interessert i alle tilbakemeldinger som kan hjelpe dem. Resultater av studien kan også gi interessant innsikt i bruk av offentlige nettsted med mye informasjon, både for utviklere av liknende nettsteder og av personer som forsker eller jobber med hvordan mennesker bruker nettsider. Til slutt mener jeg den kan være relevant og interessant for alle som jobber med eye tracking, og kanskje spesielt for de som vurderer å bruke denne teknologien i forskningsprosjekter eller i bruksundersøkelser.

## **1.3 Oppgavens mål**

### **1.3.1 Problemformulering**

Fokuset i oppgaven er å analysere øyebevegelsene til brukere som har bakgrunn innenfor fagfeltet interaksjonsdesign og brukere som ikke har det, i forbindelse med at de løser oppgaver på Skatteetatens nettsider.

Det er også ønskelig å reflektere rundt teknologien eye tracking som metode i denne sammenhengen.

### **1.3.2 Tanker bak problemformulering**

Problemområdet er en versjon av den kjente metoden å sammenlikne ekspert med nybegynner, men det er ønskelig å se på kunnskap og bakgrunn innenfor mer spesifikke områder enn bare hvor flinke eller erfarne brukerne er med å bruke datamaskiner. Det ville helt sikkert også vært et interessant tema, men det ville vært vanskeligere å velge kriterier for å bestemme forskjellige kunnskapsnivåer og dermed gruppere brukere.

Hvorfor er det interessant å undersøke brukere med forskjellig bakgrunn og kunnskapsnivå fra interaksjonsdesign? Motivasjonen min var tanken på det kjente gapet mellom utvikler av



nettsteder versus brukere av nettsteder, og ønsket om å finne ut mer om realiteten av dette. Store forskjeller vil blant annet tilsi at brukstester med ”vanlige” brukere er svært nødvendig og god bruk av ressurser. Lite kunnskap om data generelt er én ting, men forskjellen på vanlige personer som bruker data mye og personer med bakgrunn i interaksjonsdesign er kanskje en mer subtil variabel jeg gjerne ville se nærmere på. Vil utdanning eller jobb med interaksjonsdesign føre til at man ser på en annen måte når man leter etter informasjon eller menyelementer på en nettside? Dette var ting jeg funderte over og brukte som basis for formulering av problemstilling.

Ønsket om refleksjon rundt teknologien eye tracking oppsto da jeg ved lesing av bakgrunnsinformasjon og tidligere forskning oppdaget at det ikke er utelukkende positive meninger om emnet. Med bakgrunn i teori vil jeg se på egne erfaringer fra studien for å diskutere fordeler og ulemper, og se på om resultatene kan gi konkret hjelp for å forbedre designet til et informasjonssystem. Det virket også interessant å gjøre meg opp noen tanker om det er et verktøy best egnet for brukertesting eller forskning.

For å finne svaret på problemstillingen har jeg utført en studie med Skatteetatens nettsider som stimuli. Jeg har brukt elementer fra flere forskningsmetoder, men har i hovedsak gjennomført et eksperimentelt forsøk. Gruppen med testpersoner ble en og en observert og målt av meg og av eye tracking-utstyr fra Tobii mens de utførte en rekke gitte oppgaver på skatteetaten.no. Jeg har sammenliknet deltakerenes øyebevegelser med hverandre og med data fra tidligere forskning. Basert på litteratur om emnet og egne erfaringer fra studien har jeg reflektert rundt teknologien eye tracking.

Spørsmål som ble vurdert rundt formulering av problemstilling og design av eksperimentet:

- Hvilke forskjeller kan man finne i øyebevegelserne til brukere med forskjellig kunnskapsnivå og bakgrunn?
- Hvordan kan øyebevegelserne si noe om hvorfor brukeren ikke var suksessfull, eller brukte lang tid i fullføring av oppgaven?
- Hvordan kan analyse av øyebevegelser forbedre design på områdene av siden der brukeren opplever problemer?
- Hvordan kan eye tracking-data brukes på en nyttig måte?

## 1.4 Oppgavens struktur

Kapittel 1 er en introduksjon til oppgaven. Det inneholder en innføring i hva eye tracking er, hva som var motivasjonen og målet for denne oppgaven, samt problemformulering.

Kapittel 2 gir bakgrunnsinformasjon om teknologien eye tracking, med historie og informasjon om hvordan det brukes i dag samt kritikk mot metoden. Kapitlet inkluderer også informasjon om Skatteetaten og deres nettsted skatteetaten.no.

Kapittel 3 gir innblikk i litteratur om eye tracking som metode, og om tidligere forskning der eye tracking er benyttet, både innen brukstesting og forskning på HCI og andre fagfelt.

Kapittel 4 omhandler forskningsmetoden brukt for design og gjennomføring av studien, med teori om eksperimentell design, spørreundersøkelser, og innhenting og analyse av eye tracking-data, samt hvordan metodene har blitt anvendt i denne studien.

Kapittel 5 beskriver hvordan de forskjellige delene av studien ble gjennomført, fra valg av testdeltakere til gjennomføring av pilottest og de praktiske testene for opptak. Det er også informasjon om laboppsettet, og om resultatene av den initiale spørreundersøkelsen.

Kapittel 6 presenterer forskjellige data fra opptak samt informasjon om hvordan den ble fremstilt. Det inneholder også diskusjon rundt resultatene og rundt eye tracking som metode.

Kapittel 7 oppsummerer oppgaven før det gis en konklusjon på oppgavens problemformulering, basert på resultater og diskusjon fra kapittel 6. Det avsluttes med et avsnitt om videre arbeid.

## 2 Bakgrunn

### 2.1 Eye tracking

Eye tracking er altså som navnet tilsier en type teknologi som måler øyebevegelser, enkelt forklart hvor en person ser. Selv om man på norsk kan oversette dette til «blikksporing» vil det videre i teksten brukes det engelske uttrykket eye tracking.

Teknologien utnytter seg av måten menneskeøyet fungerer på. Delen av øyet som gjør at vi ser skarpt kalles fovea (Gray, 1918). Fordi fovea er veldig liten ser vi bare skarpt i en del av synsfeltet vårt, der vi til en hver tid fokuserer blikket, alt utenfor kalles det perifere synsfeltet. På grunn av dette kan eye tracking måle punktet der man ser og vite at det er nokså nøyaktig det lille området som fokuseres på akkurat da. For å måle og lagre denne informasjonen ser man på fikseringer, og bevegelsene mellom. Fiksering er det norske ordet for *fixation*, som man i eye tracking kaller det øyeblikket øyet fokuserer. Den raske øyebevegelsen mellom to fikseringer kalles på engelsk *saccade*. Fordi vi må fokusere på et punkt for å kunne se vet vi at det øyet passerer iløpet av en bevegelse mellom to fikseringer ikke sees og oppfattes på samme måte. Fikseringer blir på den måten den info som er mest relevant og interessant for å analysere eye tracking-data (Nielsen og Pernice, 2010). Det personen som blir observert i en eye tracking-studie ser på – bildet, videoen, nettsiden, eller objektet – omtales gjerne som *stimulusen* eller flertallsformen *stimuli* i eye tracking-terminologi (Yusuf, Kagdi, og Maletic, 2007).

#### 2.1.1 Historie

Reed og Meyer (2006) forteller at studien av øyets bevegelser ved lesing begynte allerede på 1800-tallet, men da ved direkte observasjon av øyet. Først helt på slutten av århundret kom den første oppfinnelsen som skulle kunne måle øyebevegelsene - en slags kontaktlinse laget av Edmund Huey fra Pennsylvania, USA. Han brukte denne til å gjøre undersøkelser av personer som leste tekst på papir. Linsen ble festet direkte på øyet, og hadde et hull til å se gjennom. Til linsen var det festet en tynn metallstang. Denne stangen tegnet med sot på en papirrull mens forsøkspersonen leste tekst, og slik fikk han kartlagt øyebevegelsene. For at testpersonene skulle kunne bruke et slikt apparat fikk de bedøvelse og noen ganger dop.

Senere, i 1901 og 1905, kom det nye metoder som var mindre påtrengende. Dodge og Cline brukte refleksjon av lys i øyet for å måle horisontale øyebevegelser, som ble markert på en fallende fotografisk plate. Judd, McAllister og Steel brukte en teknikk der en liten bit hvitt materiale ble festet direkte på øyet og deretter ble dens bevegelse filmet. Videre utvikling av teknologien bygget på dette og kombinerte lysrefleksjon i øyet med video (Jacob og Karn, 2003).

Helt frem til 1948 var det en forutsetning at testpersonene holdt hodet helt stille, og det ble gjerne satt fast i en eller annen form for holder. Så oppfant Hartridge og Thompson det første eye tracking-apparatet som kunne festes på brukerens hode. Etter dette ble disse apparatene bedre og mindre (Jacob og Karn, 2003).

En viktig person innenfor forskning på øyebevegelser rundt midten av århundret blir fremhevet av Tatler et al. (2010). Det var den russiske vitenskapsmannen Alfred L. Yarbus, som skrev boken «Eye Movements and Visions» i 1965, oversatt til engelsk i 1967. Boken er fremdeles en av de mest siterte innenfor fagområdet. Han brukte en eye tracking-enhet han selv hadde oppfunnet, som ble festet direkte på øyet. Fordelen disse enhetene hadde var at de kunne gjøre stabile opptak av øyebevegelser over lengre tid, og de ga et stabilt netthinnebilde på grunn av en teknikk der bildet brukeren så på beveget seg med øynene. Yarbus gjorde mange observasjoner der testpersonene så på malerier og fotografier. Han diskuterte i detalj mønster i fikseringer og øyebevegelser, gjennom hele perioden brukeren så på bildet og spesielt i de første sekundene, og observerte for eksempel at øynene var det som ble først sett på i et fotografi av et ansikt. Mest kjent er han for observasjonen av at forskjellige oppgaver gitt til testpersonene hadde stor innvirkning på mønsteret av øyebevegelser han fikk.

Eye tracking hadde stor fremgang på 70-tallet. Selve teknologien ble videre utviklet og ble mye mer avansert enn før, og det var et populært hjelpemiddel for forskning, spesielt innenfor psykologien. Det ble særlig gjort undersøkelser av hvordan menneskeøyet fungerte og hvordan øyebevegelser kunne lenkes til kognitive prosesser og sanseinntrykk (Jacob og Karn, 2003). Først sent på 80-tallet og på 90-tallet ble det vanlig å bruke teknologien innenfor forskning på menneske-maskin-interaksjon og brukervennlighet i forhold til informasjon på skjerm.

Eye tracking har i nyere tid blitt brukt både til å kartlegge øyebevegelser på papir, på dataskjermer og på andre skjermer. I 1990 ble det i USA gjort en undersøkelse av personer som så på fotballkamp på TV for å se hvilke deler av kampen de typisk gikk glipp av (UXBooth.com). Det er også mulig å bruke eye-tracking uten skjerm, men for å se hvor en person ser mens en går rundt. For eksempel kan man kartlegge hvor en kunde ser når han handler i en butikk. Slik type eye-tracking krever at testpersonen har på seg spesielle briller eller annet hodemontert utstyr, og det krever nøye observasjon av miljøet og personen for å kunne koble data fra utstyret med akkurat hvor det ble sett, i for eksempel butikkhyllen (Nielsen og Pernice, 2010).

### **2.1.2 Dagens teknologi**

Moderne eye tracking-teknologi bruker som oftest noe som kalles «Pupil Center Corneal Reflection», som fungerer ved at et infrarødt lys rettes mot øyet. Lyset reflekteres i øyet og fanges opp igjen av eye tracking-utstyret slik at man kan se refleksjonen i pupillen, og det infrarøde lyset som reflekteres i et punkt i hornhinnen. Vektoren mellom disse to punktene brukes så til å regne ut øyeballens bevegelser. Etter en kalibreringsprosess der brukeren fokuserer på enkelte punkter av skjermen kan det videre regnes ut akkurat hvor på skjermen brukeren ser til en hver tid under opptaket (Poole and Ball, 2005). Det infrarøde kameraet kan være koblet til eller integrert i noe som ser ut som en normal dataskjerm, og man trenger derfor ikke lenger å feste noe på hodet eller øyet til brukeren (Moore og Churchill, 2011). Det er altså mulig å få en så normal bruksopplevelse som det er mulig å oppnå i en lab, fordi utstyret er så godt som usynlig for brukeren, samtidig som øyebevegelser måles med stor presisjon. Et av selskapene som selger slik teknologi idag er svenske Tobii (Tobii.com).

### **Kommersiell bruk**

Forskning og brukertesting innenfor IKT generelt og menneske-maskin-interaksjon spesielt, der eye tracking brukes som metode, er som regel konsentrert om bruk av programmer eller nettsider. Kanskje aller mest nettsider og nettapplikasjoner. Mens forskere gjerne er opptatt av å finne generelle bruksmønstre er utviklere mer opptatt av om deres produkt fungerer for brukere eller ikke.

Et litt annet område, som kanskje ikke er så interessant for forskere innen HCI, men desto mer interessant for det kommersielle markedet, er reklame. Det har blitt gjort mange undersøkelser

av reklame på nettsider ved hjelp av eye tracking. Nielsen og Pernice (2010) trekker frem noen resultater fra disse undersøkelsene. Det viser seg at en stor andel nettreklame blir fullstendig oversett av brukere, de blir rett og slett ikke sett på i det hele tatt. Reklame til høyre på skjermen blir oftere sett på, mens banner-reklame på toppen av siden oftest blir oversett. Man har også funnet at tekstreklame får flere blikk enn det grafisk eller animert reklame gjør. Unntaket er reklame som matcher det gjeldende designet på nettsiden godt. Data som dette ville vært vanskelig å få ved hjelp av andre undersøkelsesmetoder som intervjuer eller observasjon, fordi det er en type atferd brukere sjeldent er oppmerksomme på.

### **2.1.3 Blikk som inputmetode**

En siste, og ganske annerledes måte å bruke eye tracking-teknologi på er som input-metode. Med det menes at nettsiden eller programmet brukeren ser på reagerer på brukerens blikk. Dette er en ganske ny mulighet etter som teknologien blir bedre, og det har blitt forsket på innenfor menneske-maskin-interaksjon gjennom flere år, men det er ikke vanligvis implementert i praksis av flere grunner. En grunn kan være problemet med å forstå brukerens hensikt basert kun på hvor han ser. Si for eksempel at et menyvalg på skjermen reagerer når man ser på det ved å åpnes, det er ikke nødvendigvis det brukeren ville selv om det var der man så. I tillegg til det tekniske aspektet er det altså mange slike ting som må tas hensyn til om man skal bruke blikk som input. Et område der det har blitt satt i praksis å bruke blikket som input-metode er for funksjonshemmede. Selv om det for vanlige brukere nok vil være mer praktisk å bruke dagens input-teknologi (musepeker, touch osv.) er det lett å tenke seg at dette gir mange nye muligheter for noen som ikke kan bruke armer og fingre. Det kan også være et hjelpemiddel for personer som mangler språk eller tale. Det finnes flere slike hjelpemidler på markedet idag, de fleste fokuserer på å hjelpe brukerne å kommunisere med omverdenen. (Tobii.com - assistive-technology) Man kan anta at dette er et felt der vi vil se spennende utvikling ettersom teknologien blir ennå billigere, robust og mer tilgjengelig. Det er allerede indisier på at det vil skje: det danske selskapet «The Eye Tribe» har annonsert at de i 2013 vil lansere teknologi som gir mulighet for å spille på smarttelefoner og nettbrett kun ved bruk av øynene (Dagbladet, 2012).

## 2.1.4 Tobii

Tobii er et svenskt firma etablert i 2001 som i 2013 er markedsledende innen eye tracking, og teknologien deres blir brukt over hele verden av alt fra forskere til personer med nedsatt funksjonsevne (Tobii.com).

Eye tracking-utstyr og –programvare brukt for opptak i denne studien er fra selskapet Tobii. Modellen brukt heter Tobii T60XL og er integrert i en 24" skjerm. Denne modellen bruker en variant av metoden «Pupil Center Corneal Reflection» som er nevnt i kapittel 2.1.2., og er ikke-forstyrrende for brukeren i det henseende at de ikke trenger noe form for utstyr festet på seg (Tobii.com). Programvaren som ble brukt heter Tobii Studio, versjon 3.0. Alle testsesjoner ble gjennomført med en oppløsning på 1024x768 pixler og oppdateringsfrekvens på 60Hz. En fullstendig oversikt over tekniske spesifikasjoner for T60XL, samt bruksanvisning til programvaren, finnes som et nedlastbart pdf-dokument på Tobii sine nettsider.

## 2.1.5 Visualisering av data

Det er flere måter å visualisere eye tracking-data på. De to vanligste formene er *heat map* og *gaze plot*, som er bilder som henholdsvis viser data fra flere eller en bruker.

### Heat map

Heat map brukes som en måte å grafisk fremstille data fra flere brukere. Det er et skjermbilde som er markert med farger for å vise hvor flest brukere har sett. Forskjellige farger indikerer flere eller færre fikseringer i området, eller kortere eller lengre varighet på fikseringer i området. I programvaren jeg brukte, Tobii Studio, genererte jeg heat maps som viste relativ tidsbruk. Det vil si at det ble regnet ut og vist varigheten av fikseringer i de forskjellige områdene av skjermbildet relativt til tiden på opptaket, ved å dele fikseringsvarigheten på den totale tiden (Tobii Studio User Manual). Når det gjelder fargebruk er det vanlige er at man bruker sjatteringer av rød, gul og grønn der rødt er området flest brukere har fokusert på, deretter gult, deretter grønt. Se figur 2.1 for eksempel på et heat map. For at et heat map skal være nyttig i brukertest-sammenheng, mener Nielsen og Pernice (2010) at det bør være basert på minst 30 brukere. Dette er et tall de har kommet frem til fra sin egen erfaring med eye tracking og bruker-studier. De forklarer ikke mer utfyllende hvorfor de mener dette.

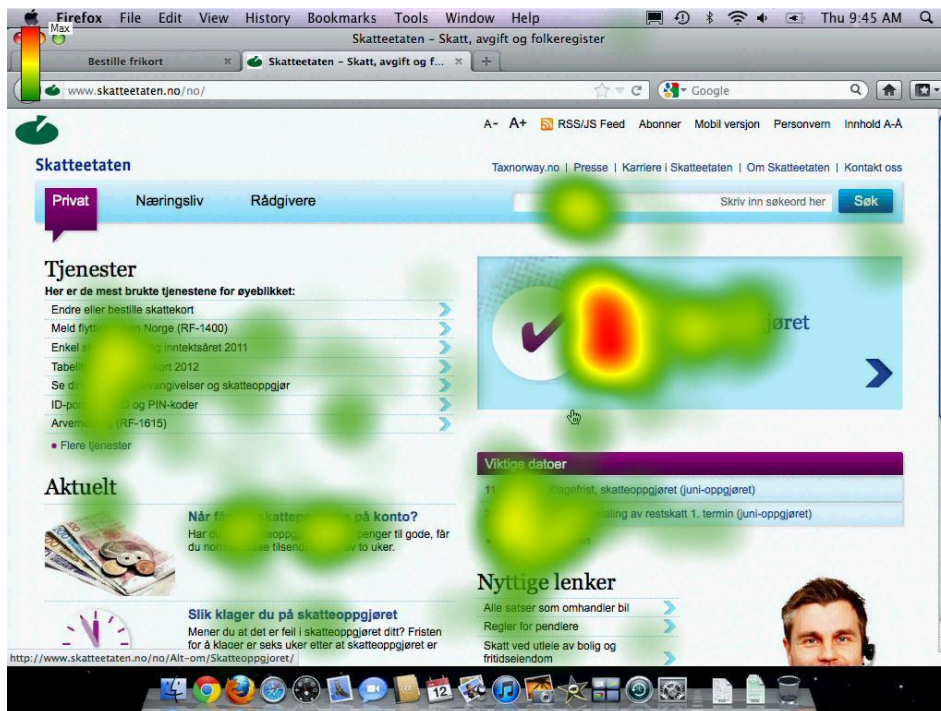


Fig. 2.1 – Eksempel på et heat map. Fremstilt i Tobii Studio, basert på data fra noen av forsøkets eye tracking-opptak.

## Gaze plot

Gaze plot brukes som en måte å vise data fra én bruker (Nielsen og Pernice, 2010). Det er et skjermbilde med grafisk fremstilling av en brukers øyebevegelser iløpet av en viss tidsperiode lagt over. Det kan være en kort periode eller hele perioden han brukte på det konkrete skjermbildet. Fikseringer vises som prikker eller sirkler. De kan vises med forskjellig størrelse avhengig av lengden på fikseringen, der større prikk tilsvare en lengre fiksering. Bevegelserne mellom fikseringene vises som rette, tynne streker. Fikseringene kan også nummereres. Figur 2.2 viser et eksempel på et gaze plot, fremstilt fra denne studiens data ved hjelp av programvaren Tobii Studio. Man kan vise gaze plot fra flere brukere i samme skjermbilde hvis ønskelig, men det ser fort rotete ut om man forsøker å vise mange brukere.



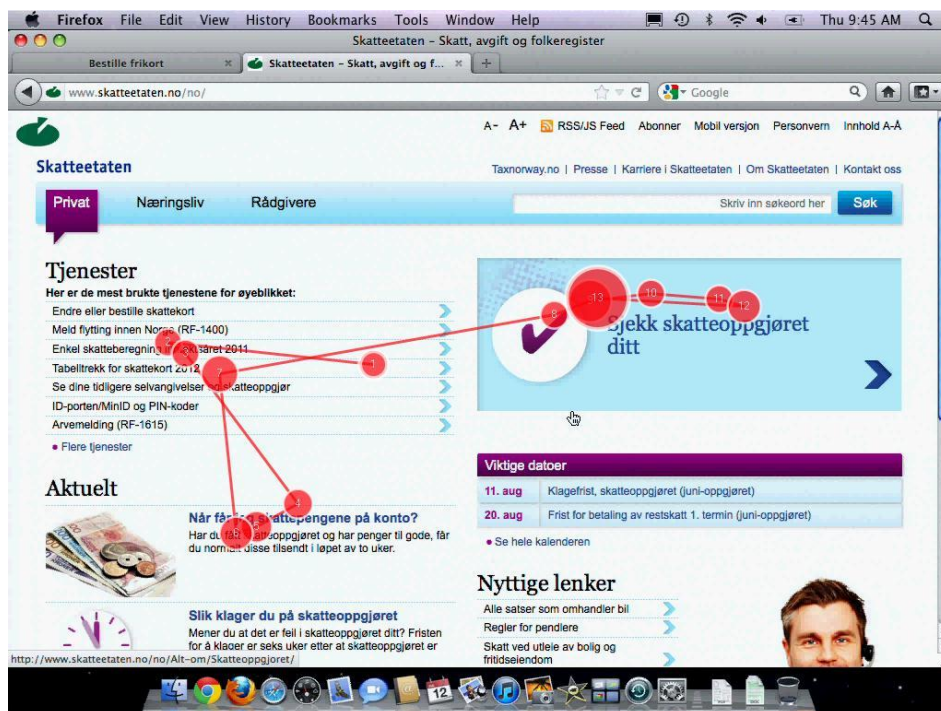


Fig. 2.2– Eksempel på et gaze plot. Fremstilt i Tobii Studio, basert på data fra et av forsøkets eye tracking-opptak.

## Statistikk

For å presentere kvantitative eye-tracking data kan man vise til statistikk. Dette kan man hente ut fra programvaren eller eventuelt måle og notere selv. Det kan være data som antall fikseringer, antall fikseringer eller tid på fikseringer innenfor et interessant område, tid brukt på å løse oppgaven og andre ting som kan fremstilles som tall og statistikk. Dette kan presenteres som all annen statistikk, i tabeller og forskjellige grafer.

### 2.1.6 Kritikk

#### Utstyr og utgifter

Eye tracking-forsøk krever en del for å kunne utføres, da man er avhengig av spesialutstyr og programvare for å gjøre opptakene og arbeide med dataene i ettertid. Det vil også være hensiktsmessig, kanskje til og med helt nødvendig, med en egen lab satt opp spesielt til utstyret, med forskjellige muligheter for lyd- og videoopptak og observasjon. Både å leie og kjøpe eye tracking-utstyr kan være svært kostbart (Toftøy-Andersen og Wold, 2011).

## Labmiljø

Eye tracking-opptak av personer som ser på en skjerm må utføres i en lab med nødvendig utstyr, fremfor i brukers naturlige miljø, med de problemer det måtte medføre. Det har blitt rapportert at brukere kan oppføre seg annerledes i en labsetting enn i en mer naturlig setting, gjerne på grunn av faktorer som labmiljøet eller det at de blir observert (Lazar et. al, 2010). Dette er en kjent faktor i eksperimentell forskning og vil også være gjeldene for eye tracking-forsøk fordi de gjøres i lab med mennesker som testsubjekter.

Utstyret i laben må videre kalibreres før opptak av hver enkelt bruker, noe som kan være tidkrevende eller enkelt, avhengig av typen eye tracking-utstyr man har tilgjengelig i laben. Selv om moderne eye tracking-utstyr ikke krever fysiske enheter festet på brukers hode vil det likevel være en bevegelsesbegrensning der brukere må sitte forholdsvis rolig og ikke kan flytte seg rundt som han ønsker (Jacob og Karn, 2003).

## Data

Eye tracking-data fra en bruker som ser på en nettside vil gi data om *hvor* brukeren har sett til et hvert tidspunkt, men for å kunne vite *hva* han så på må man synkronisere eye tracking-dataene med det skjermbildet brukeren så på. Man er altså avhengig av programvare som gjør dette dersom man skal se på opptakene i ettertid og kunne forstå dem (Lazar et. al., 2010).

Både Lazar et. al (2010) og Jacob og Karn (2003) peker på en utfordring når man bruker eye tracking, nemlig videre arbeid med data fra opptak. Det blir generert store mengder data, og det kan være data som kan være utfordrende å sortere, tolke og presentere. Toftøy-Andersen og Wold (2011) hevder at det for å tolke resultatene riktig kreves en del forkunnskaper. De sier også, basert på egen erfaring med brukstesting med Tobii eye tracking-utstyr, at analysen vil være tidkrevende dersom opptaket er av navigasjon og sideflyt på en nettside, eller av oppgaver som går over flere ulike skjermbilder.

I tillegg til utfordringene med å analysere eye tracking-dataene innenfor et enkelt eksperiment er mangelen på standardisering av terminologi og variabler med på å gjøre sammenlikning av forskjellige forskningsresultater vanskelig (Jacob og Karn, 2003).

## 2.2 Skatteetaten på nett

*«Underlagt Finansdepartementet er Skatteetaten hovedansvarlig for et oppdatert folkeregister, samt at skatter og avgifter blir fastsatt og innbetalt på riktig måte»*  
(skatteetaten.no, 2012)

Slik definerer Skatteetaten seg selv på sitt nettsted. De er altså en etat med et stort og omfattende ansvarsområde som angår så godt som alle som bor i Norge. Nettstedet deres forsøker å dekke alle disse temaene, både for privatpersoner og næringsliv som bruker tjenestene, samt personer som jobber med tjenestene – for eksempel som rådgivere rundt skattespørsmål. I tillegg til informasjon og veiledning finnes det både enkle og avanserte tjenester tilgjengelig fra nettsidene, som bestilling av skattekort, melding om flytting, levering av selvangivelse og mye annet. Noen av disse tjenestene ligger direkte på skatteetaten.no, mens andre ligger på den offentlige siden Altinn. I de tilfellene der et aktuelt skjema ligger på Altinn lenker skatteetatens nettsider videre dit, slik at brukeren kan logge på og komme til skjemaet.

For meg og mitt arbeid er Skatteetatens nettsted interessant av flere grunner. Først og fremst er det et nettsted som er aktuelt for alle. Det er også et omfattende nettsted som inneholder mye informasjon, og fra et utviklers eller designers synsvinkel er det interessant å se hvordan dette kan vises slik at brukeren får en best mulig opplevelse. I motsetning til mange andre nettsteder skal ikke skatteetaten.no underholde, selge noe eller nødvendigvis være spennende og fascinerende. Det som er viktig er at brukeren raskt finner frem til det han lurte på eller ønsker, og at det går smertefritt å gjennomføre ønskede tjenester.

Tjenestene som tilbys på skatteetaten.no er ikke nye tjenester, de har før vært tilgjengelige via skjemaer sendt i post eller mer vanlig ved direkte kontakt med Skatteetaten. Det er altså et godt eksempel på digitalisering og modernisering av personlige oppgaver og offentlig informasjon. Prosesser som tidligere var menneske-menneske-interaksjon er nå i større grad menneske-maskin-interaksjon. I tillegg har flere prosesser blitt automatisert og forenklet, og i noen tilfeller helt fjernet elementet der brukeren selv må gjennomføre en handling. For eksempel selvangivelsen, som nå kan sjekkes ferdig utfylt på nett, og ikke lenger må leveres dersom man ikke ønsker endringer. Dette kan få en til å lure på hva disse forandringene har å si for brukerne av tjenestene.

Skatteetatens brukerundersøkelse om lønnstakere og pensjonisters oppfatning av Skatteetaten for 2010 underbygger det at etaten har flere flittig brukte tjenester - 54% av de spurte svarer at de har vært i kontakt med Skatteetaten de siste 2 år. En tredjedel av disse kontaktet etaten via skatteetaten.no. Epost er ikke regnet med som kontakt via nettsidene, men kontakt via Altinn er telt med. Mer spesifikt hva «kontakt» innebærer i dette tilfellet er ikke oppgitt. Dette viser at nettsiden er mye brukt. Resultater fra undersøkelsen påviser dette ytterligere ved å vise at, av de som har tilgang til internett, har 8 av 10 besøkt skatteetaten.no en eller annen gang, og internett er den klart vanligste kanalen for å skaffe seg informasjon om skatte- og avgiftsspørsmål. Skatteetatens internettsider benyttes som hovedkilde til skatte- og avgiftsinformasjon av 29% i målgruppen, mens 19% benytter andre internettsider som hovedkilde.

Siden skatteetaten.no er et såpass stort nettsted med mange undersider og tjenester ble det valgt å fokusere på kun noen deler av siden, som ble antatt som relevant for gruppen av testpersoner. I kapittel 5 vil går man nærmere inn på design av testoppgavene. Det ble valgt å inkludere flere oppgaver rundt informasjonssidene om, og bestilling av, skattekort. Uformell tilbakemelding fra kontakt hos Skatteetaten tilsier at dette er en tjeneste de får mange «unødvendige» henvendelser om. Altså spørsmål og henvendelser der brukeren selv kunne funnet løsningen på skatteetaten.no, men har støtt på problemer eller ikke valgt å forsøke. Tjenesten er ikke begrenset til en snever aldersgruppe eller liknende, den er aktuell for alle med en form for inntekt, og den er mye brukt. Av de som har vært i kontakt med Skatteetaten de siste to årene – alle typer kontakt regnet med – oppgir 47 % at henvendelsen gjaldt skattekort, og over halvparten av de spurte foretrekker en nettløsning for bestilling av skattekort og skjemaer (Skatteetaten brukerundersøkelse, 2010).

### **2.2.1 Nettstedets oppbygging**

Skatteetatens nettsider har adresse <http://www.skatteetaten.no> og er tilgjengelig i dag. I tiden mellom gjennomføring av forsøket beskrevet i denne oppgaven, og ferdigstilling av teksten, har nettsidene blitt oppdatert til en ny og forandret versjon. Alle opptak ble utført på det som nå er en tidligere versjon av nettstedet, slik det så ut i perioden mars-august 2012. Denne oppgaven forholder seg dermed kun til den tidligere versjonen og bruker aldri nyere versjoner av nettstedet. Jeg har derfor inkludert denne forklaringen der jeg kort gjennomgår hvordan

sidene var lagt opp når testene ble gjennomført, inkludert noen skjermbilder fra siden. Skjermbildene er tatt i perioden mellom mars og august, 2012.

Forsiden kan oppfattes som en portal med både informasjon og lenker videre inn på nettstedet. Det er tre forskjellige portaler, med navn «Privat», «Næringsliv» og «Rådgivere». Man kan velge mellom disse i menyen på toppen av siden. En farget snakkeboble indikerer hvilken portal du ser på. Figur 2.3 viser forsiden tiltenkt privatpersoner, indikert ved den lilla snakkeboblen rundt menyvalget «Privat» i toppmenyen. Det er dette som er standard forsiden man kommer til når man taster inn Skatteetatens URL.

The screenshot shows the Skatteetaten website homepage. At the top, there is a navigation bar with the Skatteetaten logo on the left and links for RSS/JS Feed, Abonner, Mobil versjon, Personvern, and Innhold A-A on the right. Below the logo, the text 'Skatteetaten' is displayed, followed by 'Taxnorway.no | Presse | Karriere i Skatteetaten | Om Skatteetaten | Kontakt oss'. A search bar is located on the right side of the navigation bar with the text 'Skriv inn søkeord her' and a 'Søk' button. The main navigation menu is located below the search bar, with 'Privat' selected and highlighted in a purple box, and 'Næringsliv' and 'Rådgivere' also visible. The main content area is divided into several sections: 'Tjenester' (Services) with a list of services like 'Endre eller bestille skattekort' and 'Meld flytting innen Norge (RF-1400)'; 'Aktuelt' (Current) with two articles: 'Når må jeg betale hvis jeg har fått restskatt?' and 'Slik klager du på skatteoppgjøret'; 'Viktige datoer' (Important dates) with a table of dates: '24. sep' (Frist for betaling av restskatt 2. termin (juni-oppgjøret)), '17. okt' (Siste pulje av skatteoppgjøret er klar), and '19. okt' (Søk i skattelistene blir tilgjengelig); 'Nyttige lenker' (Useful links) with a list of links like 'Skjemaer', 'Lag KID-nummer', and 'Regler for pendlere'; and a 'Kontakt oss' button with a photo of a smiling man. At the bottom, there is a footer with the 'SVART ARBEID ER EN TV' logo, a list of services like 'Skattekort', 'Frikort', and 'Selvangivelsen', and a contact information section with the text 'Vi ønsker din hjelp for å gjøre skatteetaten.no til et bedre nettsted. Gi oss dine innspill her.' and social media links for Facebook and Twitter.

Fig. 2.3 – Forsiden. Nettsiden man kommer til når man taster URL <http://www.skatteetaten.no>. Som standardverdi vil man se forsiden for «Privat», men «Næringsliv» eller «Rådgivere» kan velges i meny øverst for å se forsidene for de respektive temaene.

Forsideportalen har noen deler som er statiske og ikke forandrer seg. Det gjelder menyen og søkefelt øverst på siden, menyen nederst på siden, elementet «Nyttige lenker» og bildet som

lenker til kontaktsiden. Andre deler av forsiden forandres jevnlig for å vise aktuell informasjon. Det gjelder elementene «Tjenester», «Aktuelt», «Viktige datoer» og grafiske elementer som lenker til tjenester som er ekstra aktuelle og som derfor fremmes på forsiden for å gjøre dem enkle å finne. I skjermbildet i figur 2.3 er dette to bannere som lenker til informasjon om henholdsvis skatteoppgjør og flyttemeldinger. Under «Tjenester» finner man det som er mest brukt for øyeblikket, mens «Aktuelt» og «Viktige datoer» endres basert på aktuelle temaer og frister i nær fremstående tid.

The screenshot shows the website for Skatteetaten (The Tax Authority) with the 'Skattekort' (Tax Card) page selected. The page layout includes a top navigation bar with links like 'A- A+', 'RSS/JS Feed', 'Abonner', 'Mobil versjon', 'Personvern', and 'Innhold A-A'. Below this is a secondary navigation bar with 'Privat', 'Næringsliv', and 'Rådgivere' tabs, and a search box. The main content area is titled 'Skattekort' and contains several sections: 'Du må selv kontrollere opplysningene i skattekortet og levere det til arbeidsgiveren din.', 'Skattekortet skal ikke leveres til NAV.', 'Skattekortet for 2012 ble sendt til din folkeregistrerte adresse mellom 15. og 31. desember 2011.', 'Endre eller bestille skattekort', 'Bestill kopi av skattekortet', 'Sommerjobb? Husk frikort', 'Spørsmål og svar', 'Brosjyrer og bøker', and 'Artikler'. The right sidebar features a 'Se også' section with links to 'Emner', 'Selvbetjening', 'Skjema', 'Tabeller og satser', and 'Grupper', followed by a 'Fant du det du lette etter?' section with a search form.

Fig. 2.4 – Dette er siden man ser om man velger menyvalget «Skattekort».

Ved å klikke på en av de mange lenkene på forsiden vil man komme videre inn i nettstedet. De fleste undersidene har en lik struktur. Toppfeltet med menyen der forblir likt. Hovedfeltet er delt opp i tre kolonner. Venstre kolonne er en statisk meny som er lik hele tiden, som kan brukes for å navigere til andre temaer. Midtkolonnen inneholder både informasjon og lenker.

Høyre kolonne har en meny som forandres basert på siden man er på. Her finnes lenker til relatert innhold, skjemaer, tabeller osv.

Flere steder finner man lenker til viktige tjenester og skjemaer, som for eksempel bestilling av frikort eller bestilling av nytt skattekort. I figur 2.5 ser man en slik side for å regne ut tabelltrekk. Disse lenkene er fremhevet ved å ha stor skrift og en lys grå boks rundt seg. Klikker man på en av disse lenkene får man enten opp et skjema for å utføre tjenesten, eller man blir sendt videre til Altinn. Figur 2.6 viser skjemaet man kom videre til om man trykket på lenken «Tabelltrekk for skattekort 2012» i figur 2.5. Figur 2.7 viser et annet av disse skjemaene, det for å bestille kopi av skattekort.

The screenshot shows the Skatteetaten website interface. At the top, there is a navigation bar with the Skatteetaten logo, a search bar, and links for RSS/JS Feed, Abonner, Mobil versjon, Personvern, and Innhold A-Å. Below this is a secondary navigation bar with 'Privat', 'Næringsliv', and 'Rådgivere' tabs, and a search bar with the text 'Skriv inn søkeord her' and a 'Søk' button. The main content area is titled 'Tabelltrekk for skattekort 2012' and includes a sub-header 'Her finner du en elektronisk versjon av trekktabellene. Trekktabellene viser hvor mye skatt arbeidsgiver skal trekke dersom du har et skattekort som viser en tabell.' Below this, a large grey box contains the link 'Tabelltrekk for skattekort 2012' in blue text. A list below the link shows '• Trekktabeller som pdf'. To the left of the main content is a vertical navigation menu with categories like 'Alt om', 'Folkeregistrering', and 'Bibliotek'. To the right is a sidebar with a 'Se også' section containing links to 'Emner', 'Tabeller og satser', and 'Grupper'. At the bottom of the sidebar is a feedback form titled 'Fant du det du lette etter?' with radio buttons for 'Ja' and 'Nei', a text input field for 'Hva ønsket du å finne?', and a 'Send' button.

Fig. 2.5 – Siden om tabelltrekk, med lenke til utregningsskjema og til en pdf av trekktabellen.



## Tabelltrekk for skattekort 2012

Oppgi tabellnummeret, inntektstypen og trekkperioden for å se hele trekktabellen

[Tabellnummer](#)    
[Inntektstype](#)  Lønn  Pensjon   
[Trekkperiode](#)

...og trekkgrunnlaget for å finne skattetrekket

[Trekkgrunnlag](#)

---

Har du spørsmål om skatt, ring Skatteopplysningen tlf 800 80 000.  
Du finner adressen til [skattekontoret her](#).

Fig. 2.6 – Skjema for å regne ut tabelltrekk.



## Bestill kopi av skattekort

Tast inn fødselsnummer (11 sifre)

---

Har du spørsmål om skatt, ring Skatteopplysningen tlf 800 80 000.

Fig. 2.7 – Skjema for å bestille kopi av skattekort.



# 3 Litteratur

## 3.1 HCI

*«Because eye movements provide a window onto so many aspects of cognition, there are also rich opportunities for the application of eye-movement analysis as a usability research tool in HCI and related disciplines such as human factors and cognitive ergonomics.»*

(Poole and Ball, 2005)

Poole and Ball (2005) diskuterer det som den gang var gjeldende status for eye tracking som metode i HCI-forskning. De fokuserer blant annet på betydningen av de forskjellige typer av blikkdata man sitter igjen med etter opptak med teknologien: fikseringer (eng. *fixations*), og raske bevegelser (eng. *saccades*), og basert på disse igjen: søkesti (eng. *scan path*) og blikk (eng. *gaze*). De nevner også pupilstørrelse og blunkefrekvens, men konkluderer med at dette er mindre brukte variabler grunnet ytre faktorer som vil påvirke dem. Fikseringer og raske bevegelser er nevnt i kapittel 2.1 av denne oppgaven, og er som Poole og Ball (2005) også poengterer, det man hovedsakelig måler i eye tracking-opptak. Søkesti definerer de som en sekvens satt sammen av fikseringer og raske bevegelser, mens blikk er det de kaller en samling av fikseringer innen et begrenset område. Videre ser de på forskjellige mønster i disse målingene, og hva de menes å bety basert på tidligere forskning innen HCI. Eksempler på slike mønster er repeterte fikseringer, antall fikseringer i et område, antall raske bevegelser, og lengde eller tidsbruk på søkesti. To sammenhenger som blir nevnt, blant flere, er at antall fikseringer i et område kan bety at området er mer merkbart eller viktig for en bruker, og at antall raske bevegelser kan indikere mer søking (Poole og Ball, 2005).

### 3.1.1 Samtaleanalyse

I tidlig bruk av eye tracking, som beskrevet i kapittel 2.1.1, var det fokus på observasjon av brukere som leste tekst på papir. Naturlig nok ettersom eye tracking-metoden er eldre enn den moderne datamaskin. Men interaksjon med moderne datamaskiner og nettsider baserer seg ikke kun på lesing av tekst på skjerm, men på interaksjon mellom bruker og maskin.

Maskinene er raske og man kan ha sanntids-interaksjon på et personlig nivå som på mange måter kan minne om sosial interaksjon mellom mennesker. Ved hjelp av tekstlig og muntlig

input samt musebevegelser og –klikk kommuniserer brukere med maskinen, og maskinen responderer med informasjon, spørsmål osv. Fordi maskiner er programmert vil de alltid svare på en logisk måte ut i fra et sett gitte verdier og muligheter, man vil ikke få det samme uforutsigbare aspektet man har i menneske-menneske-interaksjon. En av «svakhetene» hos maskinene er at de ikke kan lese situasjonen, kroppsspråk og underliggende meninger, de er avhengige av hva de får som input for å reagere. Videre vil en maskin ha problemer med å ordne opp i ting som misforståelser. På den positive siden vil en maskin aldri blande inn personlige følelser eller bli sliten (Moore og Churchill, 2011). Hvordan kan likheter og ulikheter mellom menneske-menneske- og menneske-maskin-interaksjon påvirke undersøkelser av det sistnevnte ved hjelp av eye tracking-teknologien? Kan man bruke metoder brukt innen analyse av sosiale regler og ritualer kombinert med samtaleanalyse til slik forskning?

Ideen om å bruke metoder for samtaleanalyse innen forskning på interaksjon mellom mennesker og datamaskiner er ikke ny, Suchman diskuterte det allerede i 1987 i sin bok «Plans and Situated Actions» (Suchman, 1987). Boken finnes i en nyere versjon fra 2007 (Suchman, 2007). Her brukte hun slike metoder for å analysere menneske-maskin-interaksjon, og hun sammenliknet også resultatene med de samme analysene av menneske-menneske-interaksjon. Hun ser på både mennesker og maskiner som deltagere i interaksjonen, «samtalen» mellom dem, men ikke som likeverdige deltagere, og hun fokuserer mye på kartlegging av hvor maskinene kommer til kort i interaksjonen.

### **«Computer interaction analysis»**

Moore og Churchill (2011) påpeker at det ikke finnes studier som kombinerer disse metodene, etnometodologi og samtaleanalyse, med en ekstra faktor: den nyeste eye tracking-teknologien. De påpeker også fokuset på kvantitativ behandling av eye tracking-data fremfor kvalitativ. På bakgrunn av dette har de gjennomført en undersøkelse som de introduserer slik:

*«We have identified two gaps in the HCI literature: qualitative, eye-tracking studies that focus on the temporal structure of individual users' actions and ethnomethodological and conversation analytic studies of single users and GUI-based computers. We attempt to fill these gaps with a novel approach, which we call «computer interaction analysis».»*

(Moore og Churchill, 2011. p. 504-505)

Metoden de videre beskriver innebærer å bruke eye tracking-utstyr for å observere en enkelt person som bruker en datamaskin, og deretter lage et transkript og en kvalitativ analyse av det brukeren gjorde. Analysen baseres på eye tracking-informasjonen samt video av hvor brukeren klikket, skrev osv. – såkalt «screen capture video». I tillegg kan man inkludere video av brukeren mens han utfører oppgavene, men Moore og Churchill mener dette er mindre interessant for analysen enn selve eye tracking-videoen. Metoden tar med seg kunnskap om vaner og systematisering i menneskelige relasjoner, den sekvensielle måten samtaler mellom mennesker kan transkriberes, samt sammenlikning mellom menneske-maskin-interaksjon og menneske-menneske-interaksjon. Formålet er å bruke en detaljert rapport over GUI-interaksjonen til å gjøre en kvalitativ analyse, for å så forstå bedre hvordan systemet brukes, eventuelle problemer mellom bruker og systemet, samt hvordan brukeren løser problemer i interaksjonen (Moore og Churchill, 2011).

Transkribering av GUI-interaksjon vil ha med både hva brukeren gjør, hva brukeren ser på, og informasjon om tid hver enkelt handling tar samt tidsperioder mellom handlingene. Dette er veldig likt som i samtaleanalyse. Slik får man en detaljert logg over deler av interaksjonen, som man så kan se nærmere på. Moore og Churchill (2011) skriver at det ikke meningen at dette skal brukes til videre analyse alene, det er meningen at det skal brukes som støtte, i tillegg til rådata (videostrøm, lydopptak osv.), og som et verktøy for å fremstille deler av rådata for andre.

Slik ser et eksempelutdrag fra et "Computer interaction analysis"-transkript ut:

```
(1) [Sj22_Web4]
001 <TYPES "activitied", 2.2>
002     (0.3) {{Glances at term "activitied"}}
003 <DELETES "activitied">
004     (0.2)
005 <TYPES "activities for kids ", 3.0>
006     (0.7) {{scans query}}
007     (1.1)
008 <TYPES "activities for kids in the bay area", 2.9>
009     (0.1)
010 <ENTER> (spinning icon & "Connecting..." appear in page tab)
011     (0.9)
012 //activities for kids in the bay area – Yahoo! Search Res//
013     (0.2) {{scans "also try: free activities for kids..."}}
```

(Moore og Churchill, 2011. p. 508)

Her er info om øyebevegelser inkludert i krøllparanteser. Det foreslås i artikkelen forskjellige måter å vise øyebevegelsene på, for eksempel ved å inkludere tekst fra skjermen i loggen med delene av teksten brukeren fokuserer på uthevet med understrek og/eller tykk skrift, gjerne for å skille mellom bevegelser og fikseringer. Slik detaljert info om øyebevegelser gjør loggen mye lenger og er ikke alltid nødvendig for analysens skyld, da man også alltid kan sjekke video for mer detaljer om øyebevegelsene. Logger som dette gir et bedre overblikk over tidsforhold mellom blick, musebevegelser, input, scrolling og annet som skjer på skjermen enn det grafiske fremstillinger gir. Grafisk fremstilling vil på den annen side gi bedre informasjon om avstandsforholdet mellom musepeker og blick (Moore og Churchill, 2011).

### **3.1.2 WWW**

Bruk av internett øker stadig, og brukere forventer i større grad en god brukeropplevelse når de besøker nettsider. Dermed er behovet for å teste brukskvaliteten på nettsider til stede, og eye tracking er en metode som blir mer vanlig (Ehmke og Wilson, 2007). På grunn av mangel på konkluderende studier om direkte sammenheng mellom forskjellige øyebevegelser og brukeropplevelsen blir forskere og brukstestere ofte nødt til å foreta subjektive vurderinger av eye tracking-dataene (Ehmke og Wilson, 2007). Dette ble adressert av Ehmke og Wilson (2007) som utførte en utforskende studie med den hensikt å konstruere et rammeverk for sammenheng mellom mønster i øyebevegelser og spesifikke problemer med brukervennlighet.

De presenterer to tabeller som resultat. Den ene er satt sammen på bakgrunn av forskjellig tidligere forskning, den andre er basert på deres egne forsøk. Forsøket de gjennomførte besto av 19 testpersoner som fikk to oppgaver på to forskjellige nettsider. Ved hjelp av eye tracking-utstyr fra Tobii ble øyebevegelsene deres tatt opp mens oppgavene ble utført. I tillegg ble det samlet inn tilbakemelding fra brukerne enten gjennom eller etter testen, som ble brukt sammen med observasjon for å identifisere problemer som oppsto. Til slutt ble disse dataene analysert i sammenheng. Figur 3.1 viser et utdrag fra tabellen de konstruerte basert på disse analysene.

Freq.	Usability problem generalisation	Usability problem	Eye-tracking pattern generalisation
12	Expected information missing (on page, area or subsection)	A user goes to a page on the site, expecting to find specific details which are not provided.	Many short fixations across page where information is expected.
5	Ineffective presentation	The search box draws the attention of visitors more than the actual link to the item that was looked for.	Very few fixations in general. No fixations on quickest link element. Some interface objects draw more attention.
5	Misleading element, unclear target of link	Users have problems distinguishing between interaction elements with slightly different behaviours (internal vs. external links).	High number of short fixations across page (scanning and reading), no long fixations.
5	Overloaded, ineffective presentation	Most participants find the page too crowded and not easy to read. They cannot easily grasp the concepts or how things are interlinked on the page or how they have to interact with it.	A lot of short fixations on single areas followed by longer saccades and regressions to elements.

Fig. 3.1 – Utdrag fra tabellen «Correlation of usability problems to eye-tracking patterns» (Ehmke og Wilson, 2007). Freq. står her for antall forekomster av det aktuelle problemet i Ehmke og Wilsons forsøk.

## Søkeresultater

Hvordan samhandler brukere med en liste resultater fra en søkemotor på nett? Dette forsøkte Granka, Joachims, og Gay (2004) å finne ut av ved hjelp av eye tracking-opptak. De så på opptak fra 26 studenter som ble gitt et sett oppgaver for å søke etter informasjon på søkemotoren Google. Observasjon av en bruker eller opptak av skjermen vil gi innsikt i hva brukeren søkte etter og hvilket resultat han klikket på, men i dette eksperimentet kunne de også se hva brukeren så på før han valgte et søkeresultat. De ønsket å se nærmere på oppførsel som hvorvidt brukeren scannet søkeresultatene fra topp til bunn, og hvor mange søkeresultater brukeren så på før han klikket på et. Forsøket fokuserte kun på perioden fra brukeren søkte etter svaret på en oppgave, til første resultat han klikket på. Data fra dette forsøket indikerte at brukere gjerne scanner søkeresultatene fra topp til bunn. De viste videre at mens det første søkeresultatet fikk klart flest klikk, så brukerne omtrent like lenge på de to første resultatene. De så også signifikant lenger på de to første resultatene enn resultater videre nedover på listen.

## 3.2 Brukstesting

En studie utført av Bojko i 2006 for å evaluere en ny nettside for ASCO (Bojko, 2006) er et typisk eksempel på en undersøkelse der målet er å finne ut om et nytt design for et spesifikt nettsted er bedre enn et tidligere, og hvorfor. De hadde en gruppe brukere som gjennomførte et antall oppgaver både på den originale ASCO-nettside og på det nye designforlaget. Det ble

målt suksessrate og tidsbruk på oppgavene i tillegg til at eye tracking målte hvor brukerne fokuserte på skjermen. Det ble blant annet sett på hvor lang tid brukeren brukte før han fokuserte på riktig menyelement, hvor mange ganger brukeren fikserte på riktig menyelement før han valgte det og hvor brukeren så i tilfeller der det ble valgt feil menyelement. De skilte mellom to forskjellige typer problemer: problemer med å finne frem til riktig sted å trykke for å løse oppgaven, og problemer med forståelsen – for eksempel navnet på et menyelement. I det siste tilfellet kunne de for eksempel se på en oppgave der mange valgte feil menyvalg og se om brukerne hadde fiksert på riktig menyvalg før de valgte feil. Dersom de hadde fokusert på riktig menyvalg en eller flere ganger før de valgte et annet kunne designerene konkludere med at det var forståelse som var problemet, ikke plassering på siden. Dette ville vært vanskeligere å dokumentere uten bruk av eye tracking og det blir påpekt i konklusjonen deres at det er i denne kvalitative analysen verdien av eye tracking ligger. I 17 av 25 oppgaver i studien noterte de at eye tracking ga fordelaktig ekstra informasjon for forståelse av problemer i designet.

Denne type studie er interessant å se på for å studere metodene brukt og hvordan eye tracking gir nyttige resultater, men selv om den gir resultater som er viktige for utviklerene og designerene av dette spesifikke nettstedet er det nok mindre interessant på generell basis. Det blir på mange måter en kontrollert og detaljert gjennomført brukstest. Dette er noe Nielsen og Pernice diskuterer (Nielsen og Pernice, 2010), og peker på noen områder der de mener eye tracking studier ikke gir korrekte eller generelt interessante resultater. For eksempel tester som kun bruker én gruppe brukere, som kanskje ikke er i målgruppen for nettsiden som testes. Da tenker de spesielt på resultater innen brukervennlighet på nett, «web usability», som er deres spesialitet.

*«Note the three important criteria:*

*Representative users*

*Realistic task performance*

*A wide variety of Web sites.*

*We believe all three are necessary to form conclusions that can be generalized and serve as useful design guidelines for other Web sites. »*

(Nielsen og Pernice, 2010, p. 35)

Spesielt punktet om realistiske oppgaver anser jeg som viktig også om man skal gjennomføre kontrollerte studier. Dersom du ber brukere se på en nettside og si om de synes den er

brukervennlig vil du nok kunne få nyttig tilbakemelding direkte fra brukerne, men eye tracking-dataene vil ikke gjenspeile normal bruk av siden fordi de færreste ser på en nettside med det formål å evaluere den.

## 3.3 Annen relevant forskning

### 3.3.1 Ekspert versus novise

Eye tracking brukes også innen mange akademiske fagfelt annet enn HCI, for eksempel innenfor psykologi. Et forskningsprosjekt utenfor fagfeltet IKT eller HCI som var med på å inspirere denne oppgaven er et som ble utført av Vogt på Psykologisk Institutt ved Universitetet i Oslo (Vogt og Magnussen, 2007). I dette forskningsprosjektet ble det sett på mønster for øyebevegelser ved betraktning av bilder og kunst, og hensikten var å sammenlikne kunstnere med lek menn (Vogt og Magnussen, 2007). Eksperimentet ble gjennomført med to brukergrupper: kunststudenter og psykologistudenter uten kunstbakgrunn. Begge gruppene så på en gitt rekke bilder med både naturlige og mer abstrakte motiver. De fikk se bildene to ganger, både uten noen annen oppgave enn å se og med hukommelsesoppgaver der de skulle huske bildene i ettertid. Forsøkspersonene fikk ikke vite at det var øyebevegelsene deres som ble observert, for å ikke påvirke hvordan de så på bildene. Under det praktiske eksperimentet ble det målt øyebevegelser, fikseringer og tid brukt på å se innenfor interesseområder av bildene definert på forhånd, det som på engelsk kalles «area of interest», forkortet AOI. Resultatene viste blant annet at de utrente hadde en høyere fikseringsfrekvens når de skulle se fritt på bildet, mens det var omvendt for kunststudentene, som hadde en høyere fikseringsfrekvens når de skulle huske bildet. Videre ble det konkludert at de utrente brukte mer tid enn kunststudentene på å se innenfor områder av bildet med gjenkjennbare elementer, som for eksempel ansikter. Etter å ha blitt bedt om å memorisere bildene brukte også kunstnerne mer tid på å se innenfor disse områdene.

Resultatet av den frie observasjonen her er interessant også i forhold til menneske-maskin-interaksjon: Kan dette overføres til dataskjermer? Basert på denne undersøkelsen er det naturlig å tro at om man kan bruke gjenkjennbare elementer som bilder av mennesker for å trekke oppmerksomheten til en ønsket del av for eksempel en nettside. Det kunne vært interessant å se på om man har de samme forskjellene med erfarne og uerfarne databrukere

som det her ble målt forskjell på kunstnere og lekmenn. Dette forsøk er en variant av dette, og som sagt definitivt inspirert av denne studien.

Å sammenlikne erfarne og uerfarne brukere innen et spesielt kunnskapsområde er ikke en uvanlig problemstilling i forskningsprosjekter. Et annet eksempel der eye tracking ble brukt i et eksperiment av denne typen ble utført av Law, Atkins, Kirkpatrick og Lomax (2004). De sammenliknet to grupper kirurger, en ekspertgruppe og en novisegruppe, som utførte en øvelse på en datamaskin-basert laparoskopisk kirurgisimulator. Øvelsen var en sikteøvelse som ble utført med en hånd, og eye tracking-utstyr målte hvor på skjermen kirurgene så mens de utførte oppgaven. Deretter ble eye tracking-dataene brukt for å få en bedre forståelse av hvorfor ekspertene var raskere og mer feilfrie enn novisene.

I 2012 ble det utført et liknende eksperiment i Finland, med nevrokirurger som så på bilder fra en operasjon (Eivazi et. al., 2012). Dette var en foreløpig studie der forskerene undersøkte eye tracking-data fra fire erfarne kirurger og fire uerfarne kirurger som fikk se på 4 nærbilder fra en operasjon for å fjerne en svulst.

### **3.3.2 Erotiske bilder**

Lykins, Meana og Kambe (2006) gjennomførte et forsøk der de undersøkte visuell oppmerksomhet i forbindelse med erotisk stimuli, ved hjelp av erotiske bilder og eye tracking-opptak. De presenterte to grupper, en gruppe med 10 menn og en med 10 kvinner, for 10 forskjellige bilder. Halvparten av disse var av erotisk karakter, halvparten var ikke. Mennene fikk se bilder av kvinner og kvinnene fikk se bilder av menn. De så annenhvert erotisk og ikke-erotisk bilde, og forsøket var balansert slik at halvparten så erotisk bilde som det første og halvparten ikke. De fikk se på hvert bilde i 15 sekunder, mens eye tracking-utstyr gjorde opptak av hvor de så.

Deretter kunne forskerene gjøre sammenlikninger av eye tracking-dataene når gruppene så henholdsvis på erotiske og ikke-erotiske bilder. De presenterte også observasjon av forskjellene mellom de to gruppene menn og kvinner, men ingen direkte sammenlikning da de ble presentert forskjellige bilder.

For å gjøre sammenlikningen delte de inn bildene i tre interesseområder: ansikt, kropp og kontekst (bakgrunn). De tre variablene som så ble sammenliknet var totalt antall fikseringer



innen et interesseområde, tiden på første fiksering i interesseområdet og total tid brukt i interesseområdet. De valgte disse variablene fordi total antall fikseringer i et område samt total tid brukt i et område begge antas å indikere generell interesse for området, eller at området tiltrekker seg oppmerksomhet. Tiden på første fiksering på et område er antatt å indikere hvor godt område fanger initiell oppmerksomhet. Dataene fra de to gruppene ble presentert separat.

Lykins et. al (2006) konkluderer med at opptakene deres viser forskjellige mønster for øyebevegelser når brukere ser på erotiske og ikke-erotiske bilder. Hovedforskjellen de så var preferansen for å fokusere på kroppen på bildet fremfor ansikt og kontekst. Fra et psykologisk perspektiv mener de dette bekrefter at visuell prosessering av erotisk stimuli er annerledes enn ikke-erotisk stimuli. Når det gjelder eye tracking som metode mener de at deres forsøk viser at teknologien har evnen til å fange denne forskjellen i visuell oppmerksomhet. Videre diskuterer de forskjellen på variablene som ble målt. De største forskjellene fant de i måling av de mer generelle variablene antall fikseringer og total tid, mens forskjellen på den mer subtile variabelen tid på første fiksering ikke var like fremtredende. Som mulig grunn viser de til Henderson, Weeks og Hollingworth (1999) som mener at denne variabelen påvirkes mer av visuelle faktorer som kontrast, lys og farge, mens antall fikseringer og totalt tid påvirkes mer av bildets innhold.

Når det gjelder forskjell mellom menn og kvinner i dette forsøket trekker ikke forskerene noen konklusjon, men de observerer at det – basert på deres data – kan se ut til at kvinner visuelle mønster endres mer mellom erotisk og ikke-erotisk stimuli enn menns. De noterer dette som et interessant tema for videre forskning. Et annet tema de anser som interessant å fortsette videre med er å se hvordan forskjellene de har påvist i øyebevegelsene henger sammen med brukerens tanker og vurderinger.

### **3.4 Relevans**

Metoden brukt i Lytkins et.al (2006) sin studie, og den praktiske gjennomføringen, likner mye på fremgangsmåten for dette prosjektet. En forskjell er at de i analysen har de hovedsaklig sammenliknet to forskjellige variabler innen gruppen, fremfor mellom to forskjellige grupper. En annen relevant forskjell er at det i denne studien, som i flere andre av studiene nevnt i dette kapitlet, er fokusert på analyse av brukere som sitter passive og ser på stillbilder på en

skjerm. Opptak av brukere på Skatteetaten derimot var av testpersoner som aktivt utførte oppgaver på skjermen, de navigerte og lette etter informasjon på en nettside, og hoppet gjerne mellom flere sider og scrollet opp og ned på skjermen. Metode og problemstilling kan sammenliknes med studiene til Lytkins et.al. (2006) og Vogt (2007), men tema og testoppgavens natur likner mer på studiene fra HCI, for eksempel Granka, Joachims, og Gay (2004) sin studie rundt søkerresultater og Ehmke og Wilson (2007) sine studier for å kartlegge øyebevegelser i forbindelse med bruksproblemer.

## 4 Metode

Når man skal gjennomføre et forskningsprosjekt er det mange forskjellige metoder for design av forsøk, og for innhenting av data. Innenfor fagfeltet HCI benyttes flere forskjellige forskningsmetoder, der tre som er mye brukt er eksperimentell forskning, case-studier og etnografi (Lazar et. al., 2010). Forskjellen på disse er blant annet antall testdeltakere. Der en case studie kan følge så lite som én enkelt bruker har man i eksperimentell forskning og etnografi mange brukere for å sikre mengder data til kvantitativ analyse. Miljøet for testing er et annet aspekt som er forskjellig, etnografi ser på testobjektene i deres naturlige miljø mens eksperimentell forskning utføres under kontrollerte forhold i en lab. I tillegg til forskningsmetodene nevnt her har man brukstesting, som også hører med innenfor HCI.

For innhenting av data har man mange alternativer, blant annet spørreundersøkelse, dagbøker, intervjuer, fokusgrupper, observasjonslogg, videoopptak og skjermopptak. Data kan tas opp og registreres både automatisk og manuelt. Dersom det brukes manuelle metoder kan det være både testbrukerene selv som registrerer, for eksempel ved bruk av dagbok, eller forskeren, for eksempel ved hjelp av observasjonslogg.

Denne studien er utført ved hjelp av et kontrollert eksperiment i lab, men med et lavt antall deltakere: 14 stykker. Hovedkilden for innhenting av data har vært eye tracking-utstyr som automatisk har gjort opptak av brukere, både lyd, video, skjermopptak og eye tracking-data. I tillegg har det blitt benyttet spørreundersøkelse på nett i forkant av labtesten, og fysisk skjema for tilbakemelding fra testbrukere underveis i labtesten. I dette kapittelet skal jeg gå gjennom hvilke metoder som er brukt for å planlegge denne studien, og for å hente inn og analysere data.

### 4.1 Tidslinje

Prosjektet besto av flere faser, som ble utført i omtrent denne rekkefølgen (med noe overlapping av enkelte punkter):

1. Innføring i bakgrunnsinformasjon om eye tracking og tidligere forskning.
2. Utforming av problemstilling.
3. Opplæring i utstyr for opptak og datahåndtering.
4. Design av eksperiment.

5. Utforming av spørreundersøkelse.
6. Utforming av testoppgaver.
7. Rekruttering av testdeltakere.
8. Gjennomgang av resultater fra spørreundersøkelse.
9. Gjennomføring av pilottest.
10. Gjennomføring av hovedforsøk.
11. Fremstilling og sortering av data og resultater.
12. Analyse av data og resultater.

### **Bakgrunnsinformasjon**

De to første punktene er dekket i oppgavens første kapitler. Jeg leste om teknologien eye tracking og om tidligere forskning, besøkte lab for demonstrasjon av utstyr og utformet problemområde. Den konkrete problemstillingen ble deretter formulert underveis i prosessen med å planlegge forsøket.

### **Opplæring og testing av utstyr**

Ved første besøk på Skatteetaten sammen med veiledere fikk vi en innføring i utstyret som var tilgjengelig på lab og litt om kontaktpersons erfaring med utstyret og eye tracking i brukstesting.

Senere hadde jeg to økter med kontaktperson hos Skatteetaten der jeg fikk opplæring i kalibrering, opptak og etterarbeid, både hvordan det fysiske oppsettet fungerte og programvaren.

I forkant av pilottest, og før ferdigstilling av testoppgaver, ble det også gjennomført en testsesjon med veiledere som testdeltakere der hensikten var å teste prosessen med kalibrering og opptak.

## **4.2 Eksperimentell forskning**

Man kan kort definere eksperimentell forskning som studier der man gjennomfører et empirisk eksperiment med gitte faktorer i et kontrollert miljø.

Hoveddelen av denne studien er et eksperiment som ble utført i en eye tracking-lab, og videre analyse er deretter basert på data som er produsert i eksperimentet ved hjelp av eye tracking-utstyret.

Som Lazar et.al. (2010) påpeker er det vanlig å starte en slik studie med en hypotese. De forklarer at en hypotese er et presist uttrykk for et problem, som kan testes gjennom en empirisk undersøkelse. Hypotesen formuleres gjerne på to måter i en studie: en nullhypotese og en alternativ hypotese som alltid er det motsatte av nullhypotesen. I dette eksperimentet kunne man formulert en nullhypotese slik: «Det er ingen forskjell i blikkmønster hos brukere med bakgrunn fra interaksjonsdesign og brukere uten denne bakgrunnen.» Da vil den alternative hypotesen bli formulert slik: «Det er en forskjell i blikkmønster hos brukere med bakgrunn fra interaksjonsdesign og brukere uten denne bakgrunnen». Eksperimentet videre går så ut på å motbevise nullhypotesen, fordi man da kan si at man har bevist den alternative hypotesen. Man kan gjerne undersøke flere hypoteser i ett eksperiment.

I et eksperiment har man flere variabler, og de deles inn i avhengige og uavhengige variabler. De uavhengige variablene er satt uavhengig av hva testbrukeren gjør, og det er ofte mulig for forskerene å endre på dem. De avhengige variablene på sin side, er man interessert i å se hva som skjer med, da de påvirkes og forandres av de uavhengige variablene. (Lazar et.al., 2010) Et eksempel på en nullhypotese kan være «Det er ingen forskjell på tiden det tar å skrive en gitt setning med QWERTY-tastatur enn med DVORAK-tastatur». Typen tastatur er da den uavhengige variabelen, mens tidsbruk er den avhengige variabelen. Det er mulig å ha flere uavhengige variabler. I eksemplet her kunne man lagt til alder på brukerne som en uavhengig variabel som også kan påvirke tiden, den avhengige variabelen. En uavhengig variabel kan ha to eller flere tilstander som må fastsettes når eksperimentet designes, for eksempel to typer tastatur eller fire typer nasjonalitet.

Uavhengig variabel i dette eye tracking-eksperimentet er testdeltakernes bakgrunn, og avhengig variabel er deres blikkmønster når de ser på nettsiden til Skatteetaten.

Lazar et.al. (2010) bruker antall uavhengige variable for å skille mellom forskjellige typer eksperimentdesign. Har man bare én uavhengig variable kaller de det enkel design, mens flere uavhengige variable blir kalt faktoriell design. Innenfor begge typene skilles det mellom det de kaller *within-group design* og *between-group design*, samt at man i faktoriell design har mulighet for *split-plot design*. Forskjellen her er hvor mange variabeltilstander hver deltaker

blir eksponert for. I *between-group design*, som er designet brukt i dette forsøket, deles deltakerene inn i grupper. Hver gruppe eksponeres for en gitt tilstand av en variabel, for eksempel en spesiell type meny, eller hører til en gitt tilstand av en variabel, for eksempel en aldersgruppe. Gruppene i denne studien er basert på variabelen erfaring. I *within-group design* er alle deltakerene i en gruppe, og i tilfeldig rekkefølge eksponeres de for alle tilstandene av en eller flere variabler. Dette vil naturlig nok ikke fungere når variabelen er erfaring, da en bruker ikke kan være både erfaren og uerfaren samtidig. I en situasjon der variabelen er en type meny, tastatur eller liknende kan det derimot være et fungerende design. *Split-plot design* er en kombinasjon av dette, som altså er mulig når man i faktoriell design har mer enn to uavhengige variable. Det går ut på at man undersøkes noen variable mellom grupper og noen variable innenfor grupper.

Når analyse og resultater av eksperimentell forskning presenteres er det vanlig å inkludere signifikanstester. Man vil alltid gjøre et eksperiment på et utvalg personer, som man gjerne ønsker å si at representerer en større gruppe av populasjonen. Signifikanstester sier noe om hvor sannsynlig det er at resultatene fra det begrensede utvalget kan generaliseres til hele befolkningen, eller befolkningsgruppen man har tatt utvalget fra. (Lazar et.al., 2010) Mange testdeltakere vil gi resultater som er mer signifikante. Denne studien brukte en begrenset mengde testdeltakere, og målet med dette forsøket var heller ikke å finne statistisk signifikant data for klart å bevise hypotesen om forskjellige blikkmønstre på grunn av erfaring. Målet var heller å se etter indikasjoner på hvorvidt dette stemte eller ikke, og utforske og diskutere problemformuleringen. Tidligere forsøk med eye tracking har gjerne også et begrenset antall deltakere, som man kan se på eksemplene i kapittel 3. Dette er sannsynligvis grunnet en kombinasjon av flere faktorer, som tidsbruk på tester og ikke minst den store datamengden som må lagres og etterarbeides.

Eksperimentell design har også noen svakheter. Lazar et.al. (2010) lister de største til å være vanskeligheter med å definere en hypotese som er mulig å teste, vanskeligheter med å kontrollere forskjellige faktorer, og forskjell i testbrukers oppførsel på grunn av at det kontrollerte miljøet er annerledes enn et naturlig miljø.

## 4.3 Brukstesting

Testing av brukervennlighet, eller brukstesting, betyr at man lar representative brukere teste et produkt for å finne ut hvor godt det fungerer. I HCI-sammenheng er dette gjerne en nettside eller en programvare, men det kan også være mobile enheter som telefoner som skal testes. Testene kan gjennomføres på flere stadier i utviklingen av et produkt, fra tidlige prototyper på papir til ferdige produkter. Lazar et.al. (2010) skriver at brukstesting på alle stadiene har et felles mål: å forbedre bruksopplevelsen ved å oppdage mangler og feil ved produktet. Hvor opplever brukeren problemer, og må det gjøres endringer i grensesnittet for å unngå problemene? Man vil også gjerne finne ut hva som fungerer slik at man er sikker på å beholde de gode funksjonene. Alt i alt ønsker man altså å teste produktet eller grensesnittet sitt slik at det blir bedre, og man oppnår en høy grad av brukervennlighet i sluttproduktet.

Brukstesting kan likne på tradisjonell forskning på flere måter, ettersom de gjerne bruker de samme metodene og de samme beregningene. Den største forskjellen er målet med testingen, som er å forbedre brukergrensesnittet (Lazar et.al., 2010). Denne forskjellen gjør at man kan gjennomføre brukstesting på en mindre kontrollert måte enn ved en akademisk undersøkelse. Dette fordi resultatene ikke skal publiseres men brukes til å forbedre eget produkt. Målet med dette forskningsprosjektet er heller ikke å finne feil og forbedre Skatteetatens nettsider, men planlegging og fremgangsmåten for det praktiske eksperimentet likner på måten man ville utført en liknende brukstest på nettsidene. Ved utforming av testoppgavene ble det til dels sett på aspekter ved nettsiden som man ville anta kunne skape problemer for brukeren. Testdeltakerenes oppfatning av vanskelighetsnivå på oppgaver og observasjon av problemer som oppsto, kanskje i forbindelse med dårlig grensesnitt, ble inkludert under vurdering av hvilke testoppgaver som skulle analyseres nærmere etter eksperimentet.

I tillegg til brukstesting med representative brukere har man testing med eksperter som bruker forskjellige gitte metoder for å finne feil i grensesnittet, eller programvare som gjør det samme etter et gitt sett instruksjoner. (Lazar et.al., 2010)

Brukstesting kan foregå hvor som helst: for eksempel i et laboratorium, på en arbeidsplass, hjemme hos brukeren, eller over nettet; det som er mest praktisk iforhold til testutstyr og brukere, samt best iforhold til hva man ønsker å undersøke. En tradisjonell labsetting er gjerne to rom der brukeren sitter i det ene rommet og en eller flere observatører sitter i det andre

rommet. Man har da enveis speil eller video som gjør at observatøren kan se hva brukeren gjør, i tillegg til opptaksutstyr som tar opp ønsket data fra brukerens sesjon. (Lazar et.al., 2010) Slik var også laben der dette forsøket ble utført – den er satt opp for brukstesting til vanlig. En mer utfyllende beskrivelse av laboppsettet, med bilde, finnes i kapittel 5.4.

I tillegg til testmiljø er det er mange ting som må planlegges og gjennomføres i forbindelse med en brukstest. Lazar et.al. (2010) lister opp flere viktige punktene som må planlegges og gjennomføres i tillegg til selve testutføringen: Hvor mange brukere man trenger, hvilke oppgaver brukeren skal gjøre og i hvilken rekkefølge, innhenting av samtykke og informasjon til brukere om testoppgaver, hva man skal måle, og hvordan man skal tolke dataene i ettertid. Disse tingene står det mer om i kapittel 5 i forbindelse med praktisk gjennomføring av eksperimentet, og i kapittel 4.7 om etikk.

## 4.4 Feilmarginer

Det som testes og samles data om i dette forsøket er mennesker og deres oppførsel, noe som er vanlig i studier om menneske-maskin-interaksjon. Som beskrevet i (Lazar et.al., 2010) gir det en del utfordringer som man ikke støter på innen andre fagfelt, for eksempel i kjemi og matematikk. Først og fremst gir mennesker som testobjekter problemet med variasjon i målingene, noe som fører til forskjellige typer feilmarginer. Vi kan dele disse inn i tilfeldige feil og systematiske feil.

Tilfeldige feil oppstår vilkårlig og betyr rett og slett at hvis man måler samme testperson i samme test flere ganger på rad, vil man få litt forskjellige resultater hver gang. Dette kan man ikke gjøre noe med. Den eneste måten å minimere tilfeldige feil er å ha en stor mengde testpersoner (Lazar et.al., 2010). Det er som nevnt ikke en stor mengde testpersoner i dette forsøket og man må derfor anta at tilfeldige feil vil være tilstede i større grad. Det er relevant å vite når man skal trekke konklusjoner fra resultatene, da man i den endelige analysen bør ta hensyn til at tilfeldige feil kan ha påvirket dataene uten at man har kontroll på det.

Systematiske feil er annerledes, og noe man må ta hensyn til og i så stor grad som mulig forsøke å unngå når man utformer og gjennomfører et eksperiment samt når man analyserer dataene. Dette er viktig for å få mest mulig pålitelige resultater.



*«Systematic errors, also called «Biases», are completely different in nature from random errors. While random errors cause variations in observed values in both directions around the actual value, systematic errors always push the observed values in the same direction. As a result, systematic errors never offset each other in the way that random errors do and they cause the observed mean to be either too high or too low.»*

(Lazar et.al., 2010. p. 57)

Siden systematiske feil som forklart her ikke utligner hverandre slik tilfeldige feil vil gjøre, men heller trekker alle feilmålinger i samme retning, kan de altså ha stor innvirkning på resultatene. Derfor blir det ekstra viktig å ta hensyn til og prøve å minimere det som gir systematiske feil. Lazar et.al. (2010) lister fem hovedgrunner til at systematisk feil oppstår. Her følger en forklaring om hvordan det ble forsøkt å ta hensyn til hver enkelt av disse grunnene i dette eksperimentet:

### **Feil på grunn av måleutstyr**

Mye av målingene i forsøket blir gjort med programvare, noe som utelukker menneskelige feil. Utstyret brukt til opptak var testet, konfigurert og vedlikeholdt av Skatteetaten, så man kunne stole på at de har sørget for at alt var presist og fungerte korrekt. Det ble også kjørt tester av utstyret før forsøket, slik forsikret man seg om at alt fungerte som ønsket.

### **Feil på grunn av prosedyrer og gjennomføring**

For å minimere feil i dette aspektet ble det brukt mye tid på formulering av testoppgaver og instruksjoner til deltakerene, slik at formulering ikke skulle påvirke hvordan de utførte oppgavene. Det ble satt opp en huskeliste for både gjennomføring av eksperimentet og for instruksjon til deltakeren, slik at det ble likt hver gang og slik at testleder husket alt og brukte riktige formuleringer.

Før hovedtesten ble det gjennomført en pilottest der dette var deler av fokuset, i tillegg til testing av utstyret. Mer om pilottesten kan leses i kapittel 5.5.

I noen tilfeller har ting som rekkefølgen på oppgavene i en test innvirkning på resultatene. Dette kan være fordi deltakerene lærer underveis og blir mer kjent med produktet eller testsituasjonen, eller blir slitne. For å unngå at en ting som rekkefølge blir en uavhengig variabel man ikke kan styre bruker man gjerne randomisering for å sørge for at rekkefølgen på deltakere, oppgaver, og tilordning av uavhengige variabler, blir tilfeldig (Tullis og Albert,

2008). Det ble i dette forsøket valgt å ikke bruke randomisering av testoppgavene, og de ble presentert i samme rekkefølge til alle deltakerene. Grunnen til at man valgte å gjøre det på denne måten var at man ønsket å kunne sammenlikne de forskjellige testdeltakerenes utføring av en gitt oppgave fremfor å sammenlikne data fra en oppgave med en annen. Derfor var det ønskelig at oppgaven skulle komme på samme stedet i testen for alle deltakerene, slik at faktorer som læring, hvor slitne de var osv. skulle være så likt som mulig på den aktuelle oppgaven, noe som ville føre til at man kunne se bort i fra disse faktorene i analysen.

### **Feil på grunn av testdeltakere**

Rekruttering av representative brukere for dette forsøket er det skrevet mer om i kapittel 5.

I testsituasjonen måtte testleder forsøke å sørge for at deltakerene var avslappet og ikke slitne eller nervøse. Det ble forsøkt å hjelpe testdeltakerene med dette ved å tilby dem kaffe/te/vann før testen begynte, og bruke god tid på introduksjon og forklaring, samt åpne for spørsmål. Deltakerene kunne også be om pauser underveis dersom de ønsket det. Alt for at de skulle føle seg komfortable og rolige.

### **Feil på grunn av testleders oppførsel**

Testleder hadde ikke tidligere erfaring med å gjennomføre tester i lab, men hadde erfaring som en av flere ledere for en liten fokusgruppe i forbindelse med et studieprosjekt. Det ble derfor gjennomført opplæring av utstyr for testleder i forkant av testene. Pilottest ble også brukt som øvelse for forsøkstestene. Testleder sørget for å møte opp og sette opp utstyret i god tid før testdeltakeren møtte opp. Huskelisten for gjennomføring ble memorert på forhånd i tillegg til at testleder støttet seg på den underveis i hver test.

Under opptaket fokuserte testleder på å være rolig og ikke forstyrre testdeltakeren, samt å ikke gi ekstra informasjon eller hjelp underveis så fremt det var helt nødvendig. Manuskriptet ble fulgt for å gi testoppgavene muntlig, slik at testleder ikke skulle være subjektiv eller ledende, eller gi oppgavene på forskjellig måte til de forskjellige deltakerene.

### **Feil på grunn av testmiljøet**

Denne testen ble gjennomført på en lab, og jeg hadde ikke innvirkning på hvordan laben var konstruert og satt opp. Testleder sørget for at det var ryddig og pent, og at det ikke var noen distraksjoner underveis i testen fra utenforstående. Rommet der testen foregikk hadde ikke

vindu ut mot gaten, så det var ikke problemer med sollys i skjerm eller bråk fra utsiden. Det var altså jevnt lys og rolige forhold. Fordi testene ble gjennomført om sommeren var det få mennesker på jobblassene rundt laben, noe som førte til ytterligere ro.

Det var litt varmt og noe dårlig luft i laben, men dette kunne man ikke gjøre noe med annet enn å tilby testdeltakerene vann og mulighet til pause.

Fordi det var valgt å gi oppgavene muntlig satt testleder i laben sammen med testdeltakerene, en faktor som kan påvirke deres oppførsel. Det var ingen andre personer som kunne se testen bli utført, det var kun en testdeltaker om gangen sammen med testleder.

## 4.5 Spørreundersøkelse

En spørreundersøkelse er et sett med definerte spørsmål som en bruker blir bedt om å svare på, gjerne for seg selv via en nettside eller på papir. Med en spørreundersøkelse har man muligheten til å rimelig raskt få et høyt antall svar fra brukere over større geografiske områder. Det er en av de vanligste metodene innenfor et bredt spekter av forskningsfelt, ikke bare innenfor HCI (Lazar et.al., 2010).

I forkant av den praktiske labtesten ble det gjennomført en spørreundersøkelse som alle testdeltakerene svarte på. Hensikten med denne undersøkelsen var å samle bakgrunnsinformasjon om gruppen med testdeltakere. Resultatene fra spørreundersøkelsen var altså ikke tenkt for direkte bruk i den endelige analysen eller til å svare på problemstillingen, men for å kartlegge bakgrunnen til testdeltakerene og gruppere dem før analysen av opptak.

Undersøkelsen ble utformet med tjenesten Nettskjema fra Universitetet i Oslo, som er en tjeneste for å lage nettbaserte undersøkelser. Det ble lagt ved en lenke til denne undersøkelsen i informasjonseposten deltakerene fikk etter at de ble rekruttert, og hver enkelt deltager ble bedt om å svare på den før oppmøte i testlokalet. Slik fikk deltagerene svare i fred og ro når de passet dem.

Spørreundersøkelsen inkluderte både praktiske spørsmål om demografi, samt spørsmål om bakgrunn, evner og interesse innenfor de aktuelle feltene man var interessert i. Den inkluderte også noen spørsmål om tidligere bruk av Skatteetatens nettsider.

SurveyMonkey (SurveyMonkey, 2010) forklarer i sin artikkel om smart undersøkelsesdesign hvor viktig det er å jobbe med både spørsmål og svaralternativer for å få en god spørreundersøkelse. De forklarer hvordan faktorer som rekkefølgen av spørsmål, typen spørsmål man tar med, lengden av undersøkelsen og formulering av spørsmål er viktig å fokusere på. Når man utformer spørsmål og svaralternativer må man først ha klart for seg hva man ønsker å finne ut og få svar på. Det hjelper også å kjenne brukergruppen som skal svare på undersøkelsen. Deretter må man passe på formuleringen, spørsmålene skal være objektive og tydelige, lette å forstå og aldri ledende. Slik øker man sjansene for å sitte igjen med relevante og korrekte resultater. Alle disse rådene ble brukt for å prøve å utforme en objektiv og direkte undersøkelse med god flyt.

SurveyMonkey (SurveyMonkey, 2010) forklarer også at det er flere typer spørsmål man kan bruke, med forskjellige måter å avgi svar på. Åpne spørsmål gir brukeren frihet til å skrive et svar, mens lukkede spørsmål har spesifikke svaralternativer der brukeren må krysse av for et eller flere. I tillegg har man forskjellige skalerte spørsmål der brukeren svarer på en skala for å uttrykke for eksempel hvor enige de er i en påstand, eller hvordan de vil rangere en rekke gitte alternativer. Skalaen brukeren skal svare på kan være gitt med tall, for eksempel 1 til 5, eller ord som «helt enig», «litt enig», «helt uenig» osv.

Denne undersøkelsen besto av to deler, og alle disse spørsmålstypene ble benyttet. Den første delen hadde 14 forskjellige spørsmål, der de fleste var lukkede spørsmål med ja/nei/vet ikke eller liknende varianter som alternativer. Et spørsmål var åpent, der man spurte om deres yrke eller studieretning. Den andre delen inneholdt 16 utsagn der svaret skulle gis på en Likert-skala (Tullis og Albert, 2008). En Likert-skala er en tallskala, gjerne fra 1 til 5 eller fra 1 til 7, der man angir hvor enige man er i en påstand. Man ba brukeren rangere hvor enig de var i de følgende utsagnene på en skala fra 1 til 5 der 1 betydde «Helt uenig» og 5 betydde «Helt enig». De kunne også velge «Vet ikke».

Selv om det kan virke enkelt å lage og gjennomføre en spørreundersøkelse anbefales det at man kjører en pilottest av undersøkelsen for å luke ut eventuelle problemer, finne ut reell tidsbruk for svareren og andre ting som kan være nyttig før den ferdige spørreundersøkelsen sendes ut til brukere (SurveyMonkey, 2010). Jeg fikk en venn til å gå gjennom undersøkelsen for å sjekke at den var rask å svare på og gi tilbakemelding på om spørsmålene var forståelige. Fordi denne undersøkelsen var såpass kort, og fordi den bare var et ledd i forsøket og ikke skulle gi data til den konkluderende analysen, følte jeg at dette var nok kvalitetssikring.

Tilbakemeldingen fra den ene testpersonen var at alle spørsmål var forståelig og at det kun tok 3-4 minutter å svare, så det ble valgt å ikke gjøre noen endringer før undersøkelsen ble sendt til alle testdeltakerene.

Resultatene fra spørreundersøkelsen ble som sagt brukt blant annet til å gruppere testdeltakerene basert på bakgrunn. Mer om resultatene av undersøkelsen og hvordan grupperingene ble gjort skrives det om i kapittel 5. En rapport med alle spørsmålene i undersøkelsen samt svar fra testdeltakerene er inkludert som vedlegg B.

## **4.6 Analyse av data**

Data kan samles inn både manuelt og automatisk og ved hjelp av en mengde metoder. Man kan velge å analysere kun én type data eller man kan ha data fra mange forskjellige kilder, noe som er aller vanligst i case studier (Baxter og Jack, 2008). Hovedkilden til data som skulle analyseres i dette eksperimentet kom fra eye tracking-utstyr som automatisk tok opp brukerens øyebevegelser sammen med skjermopptak, lyd, og video av testdeltakeren. Andre datakilder som spørreundersøkelse og tilbakemeldingsskjema fra deltaker ble brukt som støtte til analysen, for å gruppere deltakere, velge ut lette og vanskelige oppgaver og slike ting. Eye tracking-utstyret som ble brukt var eye trackeren TX60 XL og programvaren Tobii Studio, begge fra selskapet Tobii. Eye trackeren samlet inn rådata om fikseringer og bevegelsene mellom, og programvaren ble brukt for å lage visualiseringer og hente ut statistikk fra rådataene.

### **4.6.1 Analyse av skjermvideo fra opptak**

I HCI-forskning samler man gjerne både kvantitative data og kvalitative data, og etter innsamling må disse dataene analyseres. Kvantitative data som tidsbruk, feilrate osv. kan analyseres ved hjelp av statistikk og signifikanstester (Lazar et.al., 2010), men med kvalitative data må man bruke en annen tilnærming. Hva er kvalitative data? Eksempler på dette er svar fra undersøkelser med åpne svar, notater fra intervjuer eller observasjon, og kanskje spesielt video- og lydopptak fra intervjuer eller eksperimenter. Skjermopptak og eye tracking-opptak faller også under denne kategorien. Selv om mye av dataen fra opptak kan analyseres i programvare med tall og statistikk som resultater, bør selve live-visningen med skjermopptak og testbrukers øyebevegelser analyseres kvalitativt av mennesker. Nettopp dette er det som

virkelig skiller analyse av kvantitativ og kvalitativ data – det menneskelige aspektet. Statistikk og formler kan regnes ut av en datamaskin, men analyse av en intervjuvideo må gjøres av et menneske. Dette, sammen med mangel på numerisk og konkret data, gjør at kvalitativ analyse er mer utsatt for subjektiv eller partisk tolkning. (Lazar et.al., 2010)

Lazar et.al. (2010, kap. 11) forklarer grunnleggende aspekter ved kvalitativ analyse: de tre stadiene av analysen, og «grounded theory». Deretter diskuterer de en rekke kjente strategier for å analysere kvalitative data, både tekst og multimedia, samt andre ting man må tenke på, slik at resultatene man får skal bli gyldige og pålitelige. Disse strategiene inkluderer metoder for innholdsanalyse, å stille spørsmål om dataene, bruk av programvare, inndeling i kodekategorier, å se etter nøkkelinformasjon, og mye annet. Sammenlikning av data er en av metodene de diskuterer som er blitt brukt i analysen. Det går ut på at man sammenlikner dataene man har på flere forskjellige måter. Man kan sammenlikne forskjellige tilfeller innenfor datasettet, man kan sammenlikne en gruppe deltakere mot en annen gruppe deltakere, eller man kan sammenlikne egne data med data og resultater fra tidligere forskning og litteratur.

For å kunne si at resultatene er gyldige må man kunne fremme argumenter for sin tolkning av dataene, og kunne vise hvordan denne tolkningen er basert på dataene. For å kunne gjøre dette er det viktig med veldokumenterte data og prosedyrer. Et annet aspekt som vil gi resultatene bedre gyldighet er hvis flere individuelle kilder av data støtter den samme tolkningen. Stake (1995) kalte dette for «data source triangulation» (Lazar et.al., 2010). Kontroll av pålitelighet er noe som først og fremst er aktuelt når man har et stort datasett og derfor flere personer som analyserer forskjellige deler av datasettet. Man må da kontrollere at disse følger samme metode og har samme kvalitet på analysen slik at ikke resultatene blir forskjellig på grunn av analysen, istedenfor på grunn av dataene. I dette prosjektet var det kun én person som både var testleder og som analyserte dataene i ettertid.

De tre stadiene av analyse går ut på at man starter med et datasett med informasjon fra en undersøkelse. I denne oppgaven er det data fra eye tracking-opptakene. Man ønsker så å identifisere de viktigste komponentene av det datasettet viser informasjon om. Neste stadiet blir så å se nærmere på hvert komponent og analysere og forstå både innholdet i hvert komponent samt hvordan de relaterer til hverandre. Til slutt bruker man kunnskapen fra det andre stadiet til å si noe om det originale datasettet og gjøre slutninger rundt dette, noe som da utgjør konklusjonen.

«Grounded theory» er et uttrykk som forklarer en form for forskningsmetode og analyse som er forankret i dataene. Ved eksperimentell forskning starter man gjerne med en teori eller hypotese, deretter samler man data, som man så bruker for å bevise teorien sin. Man kan så gå tilbake til teorien og starte prosessen på nytt om nødvendig. I «grounded theory» starter man med et sett empirisk data, og målet er å komme frem til en teori ut i fra disse dataene. Man ser gjerne «grounded theory» innen forskningsmetodene etnografi, observasjon, case studier og intervjuer. Det kan også minne veldig om fremgangsmåten ved brukertesting. Denne studien begynner med en problemstilling man ønsker å finne svar på, men åpner også for andre observasjoner ved analyse av datasettet.

## **4.7 Etikk**

### **4.7.1 Personlig informasjon**

Det kan være aktuelt i et forsøk å ta hensyn til personvern, opptak av og oppbevaring av personlig og sensitiv informasjon.

Det ble ikke samlet inn noen form for sensitiv informasjon fra testdeltakerene i dette forsøket. I forbindelse med rekruttering og innhenting av bakgrunnsinformasjon ble for eksempel ikke spurt om personnummer. Kommunikasjon foregikk via epost.

Man må anta at en del av testpersonene ville følt at informasjon om sin egen lønn og skatt var personlig. De ble derfor aldri bedt om å taste inn dette underveis i opptaket. De ble heller ikke bedt om å taste inn personnummer noe sted, eller om å logge på sider som Altinn underveis i opptaket. Dette satte noen begrensinger for utforming av testoppgaver. En opprinnelig tanke var å inkludere oppgaver der testdeltakerene fylte inn skjema for bestiling av skattekort, men dette ble gått bort i fra grunnet problemstillingen med personlig informasjon og innlogging.

Alle testopptakene ble lagret i programvaren med et testnummer, ikke med navnet til testdeltakeren.

Rådata fra opptak ble oppbevart på en harddisk hos Skatteetaten, og backup på en ekstern harddisk hos testleder. All rådata ble slettet etter ferdigstilling av analyse og oppgavetekst.

## 4.7.2 Samtykke

Før rekruttering av testdeltakere ble det utformet et samtykkeskjema. Det var basert på et annet samtykkeskjema som kontaktperson hos Skatteetaten pleide å bruke for sine brukskvalitettester med eye tracking-utstyret. Skjemaet inneholdt punkter om hva som ble tatt opp i testen, hvem som hadde tilgang til å se opptakene og når de skulle slettes. Det var også et punkt om videre bruk av opptakene der det ble klarert at kortere klipp fra opptaket, samt stillbilder, kunne brukes i min trykkede masteroppgave, ved presentasjon av oppgaven, og eventuelt til andre artikler og konferanser i senere tid. De kunne her krysse av for hvorvidt de syntes det var ok eller ikke, og hvorvidt de ønsket at ansiktet deres kunne være synlig på disse bildene eller opptakene. Jeg formulerte punktet om bruk av gjenkjennelige bilder slik at de måtte krysse av for å gi eksplisitt samtykke, heller enn å be dem krysse av dersom det ikke var greit. Dette var bevisst fra min side for å unngå misforståelser, så jeg skulle være sikker på å ikke bruke gjenkjennelig bilde eller video av noen som ikke ønsket det.

Alle de frivillige testdeltakerene måtte underskrive på samtykkeskjemaet ved oppmøte på den praktiske labtesten. De fikk også tilsendt skjemaet på epost på forhånd, slik at de kunne lese det gjennom før de møtte opp. De som ønsket det fikk en kopi av skjemaet med seg hjem etter testen, med testleders underskrift og kontaktinformasjon.

Samtykkeskjemaet er lagt ved oppgaven, se vedlegg A.



# 5 Gjennomføring av studien

## 5.1 Valg av deltakere til forsøket

Ved en brukstest for et produkt eller en nettside brukes målgruppe som en viktig faktor ved utvalg av testdeltakere. På tross av at dette forsøket hadde et annet mål enn brukertest ble testdeltakere valgt ut i fra målgruppen til Skatteetatens nettsider. I Skatteetatens brukerundersøkelse (2010) kom det frem at de gruppene som er mest tilbøyelige til å benytte Skatteetatens internettsider er de som har lønnet arbeid på heltid, de i aldersgruppen 25-39 år, de som bor i Oslo og de som har høyere utdanning. Ergo var det naturlig å velge testbrukere som var en del av en eller flere av disse gruppene. Det ble altså fokusert på at alle skulle være voksne og i arbeid. En baktanke med dette er at de da vil ha omtrentlig de samme behovene ved bruk av skatteetaten.no, i motsetning til en 15-åring, en selvstendig næringsdrivende eller en pensjonist, som sannsynligvis har andre utfordringer i forhold til skattespørsmål. Dette var et poeng i forhold til å utelukke faktorer som ville gjøre sammenlikning av testresultatene vanskeligere. Når man skal sammenlikne to grupper basert på en uavhengig variabel, må man forsøke å sørge for at andre uavhengige variabler ikke er for forskjellige slik at de påvirker resultatet (Lazar et.al., 2010).

Testdeltakerene ble rekruttert på flere måter. Noen av dem jobbet på Skatteetaten, og ble rekruttert ved hjelp av kontaktpersonen der. Andre var fra USIT, UiOs sentrale IT-enhet. Til sist ble det brukt bekjente og ikke nære venner av meg som jeg tok kontakt med via sosiale medier. De stilte alle opp frivillig og gratis.

Det ble rekruttert 14 testdeltagere. Det var ønskelig med minst 10-12 deltagere, men man måtte også ta høyde for at noen av opptakene kunne bli dårlige, eller at det kunne skje noe så ikke alle var brukbare. Av de 14 opptakene oppsto det tekniske problemer med utstyret én gang, noe som førte til at mesteparten av det opptaket ikke ble lagret. Det var også én testdeltager som eye trackingen ikke klarte å ta opp like godt som de andre, muligens fordi personen brukte sterke briller. Opptaket ble likevel bra nok til å hente ut statistikk og visuell data, så det ble besluttet å inkludere det likevel. Etter forsøket satt man altså igjen med 13 opptak som kunne brukes til analyse. I videre omtale av testdeltakerene er de nummerert fra 1 til 13, og korresponderende opptak er nummerert likt.

Det var nødvendig å designe eksperimentet slik at det fungerte med færre deltagere. Det var praktiske grunner som var medfaktorer til at det ikke var mulig å gjennomføre et stort eksperiment med mange testdeltagere. Det tok tid å finne hver deltager, sette opp tid for testen og gjennomføre testen. Hvert opptak varte mellom 20 og 45 minutter, og besto av flere store filer, både lyd, video, skjermopptak og data om øyebevegelsene. Det tok altså opp mye lagringsplass på disken som var tilgjengelig. I ettertid av testene tok gjennomgang av hvert opptak ganske mye tid. Selv om programvaren som tilhørte Eye Tracking-utstyret hadde mulighet til å presentere dataene både visuelt og statistisk krevde det mye manuelt arbeid for å bearbeide rådataene, noe som forklares nærmere i kapittel 6.

## **5.2 Spørreundersøkelse**

Som beskrevet i kapittel 4 ble det utformet en kort spørreundersøkelse som alle testdeltakerene svarte på før de møtte opp til den praktiske testsesjonen. Undersøkelsens formål var å kartlegge testgruppen, deres bakgrunn og deres erfaring med databruk og Skatteetatens nettsider. Dataene fra denne undersøkelsen ble brukt til å gruppere deltakerene. En rapport med alle spørsmålene i undersøkelsen samt svar fra testdeltakerene er inkludert som vedlegg B.

### **5.2.1 Resultater fra spørreundersøkelsen**

#### **Demografi**

Alle testdeltagerene var arbeidstakere og jobbet enten fulltid eller deltid. I tillegg studerte to av testdeltakerene heltid ved siden av jobb.

De fordelte seg i alder slik: 5 stykker i aldersgruppen 25-39, 6 stykker i aldersgruppen 40-49, 2 stykker i aldersgruppen 50-59.

Det var 7 kvinner og 6 menn, en nokså jevn fordeling. Det ble ikke sett på kjønn som en påvirkende faktor i sammenlikning av data. Det ble gjort en antagelse om at kvinner og menn ikke bruker øynene forskjellig når de skal løse de samme oppgavene på en kjønnsnøytral nettside som skatteetaten.no er.

## **Bruk av datamaskin og internett**

Tre spørsmål i testen var ment for å kartlegge testdeltakerenes daglige bruk av data:

*Bruker du datamaskin når du jobber?(Til arbeidsrelaterte oppgaver.)*

*Bruker du datamaskin når du studerer? (Til studierelaterte oppgaver.)*

*Hvor mye surfer du på nett for informasjon eller underholdning? (Jeg tenker her på nettbruk utenom studie- og arbeidsrelaterte oppgaver.)*

Basert på det deltakerene svarte på disse spørsmålene kunne man konkludere at alle i testgruppe var normalt datakyndige og vant med å bruke datamaskin og Internett til både arbeidsoppgaver og personlige oppgaver.

## **Erfaring med Skatteetatens nettsider**

Deltakerene ble spurt om de hadde benyttet Skatteetatens nettsider før:

*Har du besøkt Skatteetatens nettsider før?*

*Har du bestilt skattekort via Skatteetatens nettsider før?*

*Har du levert selvangivelsen via Skatteetatens nettsider før?*

Av de 14 testdeltakerene som svarte på undersøkelsen svarte 12 at de hadde besøkt Skatteetatens nettsider «ofte» eller «noen ganger», og 1 svarte «vet ikke». 8 hadde bestilt skattekort via nettsidene og 12 hadde levert selvangivelsen via nettsidene.

De skalerte utsagnene til slutt innholdt også noe om Skatteetatens nettsider, for eksempel:

*Jeg føler at det er trygt å ordne med skattetjenester som bestilling av skattekort, selvangivelse osv. via nettet.*

Dette utsagnet svarte alle testdeltagerene 4 eller 5 på, på skalaen der 5 betydde svært enig. Fra dette kan man konkludere at de føler seg komfortable med bruk av nettsidene fordi de stoler på at det er trygt.

Jeg inkluderte også to spørsmål om kontakt, formulert på litt forskjellige måter for å se om de foretrakk telefon over andre metoder ved kontakt og ved behov for informasjon. Basert på

svarene på disse to spørsmålene kan jeg si at de fleste foretrekker kontakt med offentlige tjenester som Skatteetaten via epost heller enn telefon eller brev. Det var også veldig få som svarte at de foretrakk telefonkontakt når de lurte på noe. Dette bekrefter viktigheten av gode nettsider, da det er et naturlig sted å først søke informasjon.

I Skatteetatens undersøkelse var kontaktmetoden brukt blant dem som hadde kontaktet etaten iløpet av de siste to årene nokså jevnt fordelt mellom nett, telefon og personlig kontakt (Skatteetatens brukerundersøkelse, 2010). Sammenliknet med Skatteetatens egen undersøkelse antyder dette at denne gruppen er noe mer tilbøyelige til å bruke nettsiden og epost som kontakt enn den generelle bruker av Skatteetaten. Dette var ikke et overraskende resultat med tanke på at alle i testgruppen er vant til å bruke PC som arbeidsverktøy på daglig basis.

### **Bakgrunn fra økonomi og skatt**

Det var inkludert noen spørsmål om deltakerenes bakgrunn innen skatt og økonomi, om de hadde studert eller jobbet med disse fagfeltene. Grunnen til at deltakerene ble spurt om dette var på én side for å få mer informasjon om deres bakgrunn. På en annen side var det basert på et ønske om å kunne sammenlikne personer med god innsikt i den informasjonen som blir vist på nettsidene brukt i oppgavene, med personer uten spesiell kunnskap om dette. Det var en tanke før gjennomføring av eksperimentet at denne sammenlikningen kunne gjøres i analysen dersom det ble tid. Det ble i ettertid valgt å ikke anse denne informasjonen som en faktor i analysen av dataene. Grunnlaget for denne avgjørelsen var først og fremst tidsaspektet, men også utvalget. Kun 3 av 13 deltakerene hadde bakgrunn innen skatt og økonomi. Se forøvrig kapittel 7.3 om videre arbeid.

### **5.2.2 Gruppering av testdeltakere basert på bakgrunn**

Jeg stilte testdeltakerene to spørsmål om bakgrunn innenfor interaksjonsdesign, formulert slik:

*Studerer du på, eller har du tidligere gått studie, som inneholder emner om interaksjonsdesign, webdesign eller brukervennlighet på nett?*

Ja

Nei

*Har du en jobb, eller har du tidligere hatt en jobb, med arbeidsoppgaver inkludert interaksjonsdesign, webdesign eller brukervennlighet på nett?*

Ja, stor del av jobben

Ja, liten del av jobben

Nei

Basert på svarene her grupperte jeg testdeltakerene. For spørsmålet om studier ga jeg dem 1 poeng dersom de svarte ja og 0 poeng dersom de svarte nei. På spørsmålet om jobb ga jeg dem 1 poeng dersom de svarte «Ja, liten del av jobben», 2 poeng dersom de svarte «Ja, stor del av jobben», og 0 poeng dersom de svarte nei. Testdeltakeren der opptaket ble ufullstendig grunnet tekniske problemer ble fra nå av sett bort i fra og er ikke inkludert i tabellen.

Testdeltakerene kunne altså få til sammen 0, 1, 2 eller 3 poeng. Fordelingen av poeng ble da slik:

<b>Deltaker:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>
Jobb:	0	1	2	2	1	0	0	1	2	1	2	0	1
Studie:	0	0	0	0	0	0	0	0	1	1	1	0	0
Sum:	0	1	2	2	1	0	0	1	3	2	3	0	1

De ble deretter delt inn i fire grupper:

Gruppe 1 – deltakerene med 0 poeng, altså ingen erfaring. 4 stykker.

Gruppe 2 – deltakerene med 1 poeng, altså litt erfaring. 4 stykker.

Gruppe 3 – deltakerene med 2 eller 3 poeng, altså en del eller mye erfaring. 5 stykker.

Når resultatene mellom gruppene sammenliknes i analysen så man på gruppe 3 – som da ble ekspertgruppen – og sammenliknet dem mot gruppe 1 – som da ble novisegruppen.

## Interesse og egenoppfattet kunnskap

Blant påstandene de skulle rangere var det også inkludert noen om interesse og egenoppfattet kunnskap:

*Jeg interesserer meg for webdesign/interaksjonsdesign/brukervennlighet på nett.*

*Jeg kan mye om webdesign/interaksjonsdesign/brukervennlighet på nett.*

Det viste seg at interesse og egenoppfattet kunnskap korrelerte ganske godt med erfaring fra jobb og studier. Svarene på disse spørsmålene ble derfor ikke vurdert ved gruppering av testdeltakerene.

## 5.3 Utforming av testoppgaver

Før man utformer testoppgaver må man se på nettstedet man skal teste på og finne ut hva som er brukerens mål med besøke nettstedet og hvordan de vanligvis bruker nettstedet (Tullis og Albert, 2008). Hva er viktig for brukeren? Alle testoppgaver i dette forsøket blir gjennomført på Skatteetaten.no så jeg har sett nærmere på den siden og hvordan den brukes i kapittel 2.

Brukeropplevelsen kan deles inn i prestasjon og tilfredshet. Prestasjon går på hva brukeren gjør, hvor fort løser han oppgaven, hvor effektivt, hvor feilfritt. Tilfredshet er hvor fornøyd brukeren er med prosessen. Det er observert at disse to ikke nødvendigvis er forbundet med hverandre, brukeren kan rapportere at han er fornøyd selv om det var vanskelig å gjennomføre oppgaven, og omvendt kan de rapportere misfornøydhet selv om alt gikk bra. (Tullis og Albert, 2008) På grunn av dette gjorde jeg opptak som jeg skal analysere og gå gjennom etter testen, samt at jeg samlet noe tilbakemelding fra bruker underveis i testen. Det var altså viktig å huske på under analysen å se på begge aspektene dersom man ønsker å si noe om helhetlig måloppnåelse for brukeren på en enkelt oppgave. I en ren brukervennlighetsundersøkelse vil dette være svært viktig, da man sannsynligvis ønsker tilfredse brukere fremfor effektive brukere. I et eksperimentelt forsøk som her er det mindre viktig, men ble brukt som en faktor for å velge ut områder av opptak å se nærmere på i analysen.

Tullis og Albert (2008) forklarer hvordan man i en brukervennlighetsundersøkelse først bestemmer hvilke mål brukeren har ved bruk av produktet som skal testes, samt målet for undersøkelsen, for deretter å bestemme hvilke type beregninger og metoder man skal bruke,

og hva som skal måles. I dette studiet var det bestemt før eksperimentet ble utformet at man skulle bruke eye tracking-utstyr for å ta opp data om øyebevegelser. Basert på en de mest vanlige typene scenarioer for å teste brukervennlighet og hva som er aktuelt å måle for hvert enkelt (Tullis og Albert, 2008) la jeg til noen elementer til testen: En oppgave utenom skjermen for å kunne måle bevissthet/oppmerksomhet, og et skjema for selvrapporing fra brukerne. Det ble også bestemt å bruke tid som en måleenhet, men fordi det ble gjort opptak av hele testen kunne man beregne tidsbruk i ettertid. Det var altså ikke nødvendig å måle separat mens testen ble gjennomført. Andre konkrete ting det var ønskelig å måle men som man også kunne se på opptak i ettertid var oppgavesuksess og brukerfeil. Disse prestasjonsmålene gjorde at i utforming av testoppgavene måtte jeg tenke på å lage veldefinerte oppgaver med en konkret slutt eller et konkret mål.

Selv med oppgaver med konkrete mål var det ikke godt å vite hvor på nettstedet brukeren ville ende ved fullføring av en oppgave. Både på grunn av muligheten for at de gjorde feil eller ikke klarte oppgaven, og muligheten for at svaret kunne finnes mer enn ett sted på et såpass stort nettsted. Derfor ble det før flere av oppgavene opplyst at man skulle navigere tilbake til forsiden før oppgaven ble gitt. I listen over testoppgaver ble dette indikert så testleder ikke skulle glemme det.

Det man først og fremst ønsket å måle i denne testen var brukerenes øyebevegelser eller blikkmønster, og det måtte tenkes på ved utforming av testoppgaver. Artikler omtalt i bakgrunnskapitlene om tidligere forskning der eye tracking var brukt fokuserte i stor grad på resultater og noe mindre på detaljer rundt metode, og det var heller ikke brukt samme utstyr og programvare som i dette forsøket. Opplæring i utstyret av kontaktperson på Skatteetaten var derfor veldig nyttig, også med tanke på formulering av testoppgaver. Det samme var pilottest. I ettertid, ved gjennomgang av data fra opptak, kom det likevel frem noen aspekter som kunne blitt gjort bedre for å lette analysen. Det diskuteres noe rundt erfaring fra studien på dette punktet i kapittel 6.

Det fullstendige settet med testoppgaver besto av 27 oppgaver. Ikke alle oppgavene ble analysert i ettertid, mer om dette i kapittel 6.1. I kapittel 6 er også oppgavene der opptaket ble analysert nærmere gjengitt, sammen med en begrunnelse for inkludering av oppgaven og resultater fra den delen av opptaket. Hele settet med testoppgaver brukt i forsøket er inkludert i vedlegg C.

## 5.4 Laboppsett

Laben som ble brukt til testene befant seg i Skatteetatens lokaler på Tøyen, Oslo, i avdelingen for brukskvalitet. Den besto av to rom: et testrom og et observasjonsrom.

Observasjonsrommet inneholdt flere datamaskiner og skjermer, der man kunne se video fra webkamera, høre hva som foregikk i testrommet samt hva som foregikk på datamaskinen i testrommet og følge med på eye tracking-data underveis i testen. Dette rommet ble ikke brukt i denne studien. Det var kun en testleder, som satt i testrommet sammen med brukerne.

Testrommet hadde plass til to personer og inneholdt et skrivebord med en 24" skjerm som hadde integrert eye tracker (Tobii T60XL), og en datamaskin til en eventuell testleder.

Rommet hadde en dør som kunne lukkes, og et lite vindu ut til kontorlandskap. Vinduet var dekket med folie for å minske innsyn og utsyn. Begge plassene hadde mikrofoner.

Testdeltakeren satt foran eye tracking-skjermen, som også hadde et webkamera som filmet brukeren under opptak. Skjerm, tastatur og mus var koblet til en bærbar Mac som sto mellom testdeltaker og testleder. Dette måtte settes opp før hver testsesjon, da det kun var dette prosjektet som brukte oppsettet med den bærebare Macen – ansatte ved

Brukskvalitetsavdelingen brukte en tynnklient som også befant seg i testrommet. Testleder satt på en datamaskin ved siden av brukeren, og kunne følge eye tracking-dataene på sin skjerm underveis i testen. Fig. 5.1 viser testlaben med en testdeltaker. Testveileder satt på plassen til høyre for testbruker.

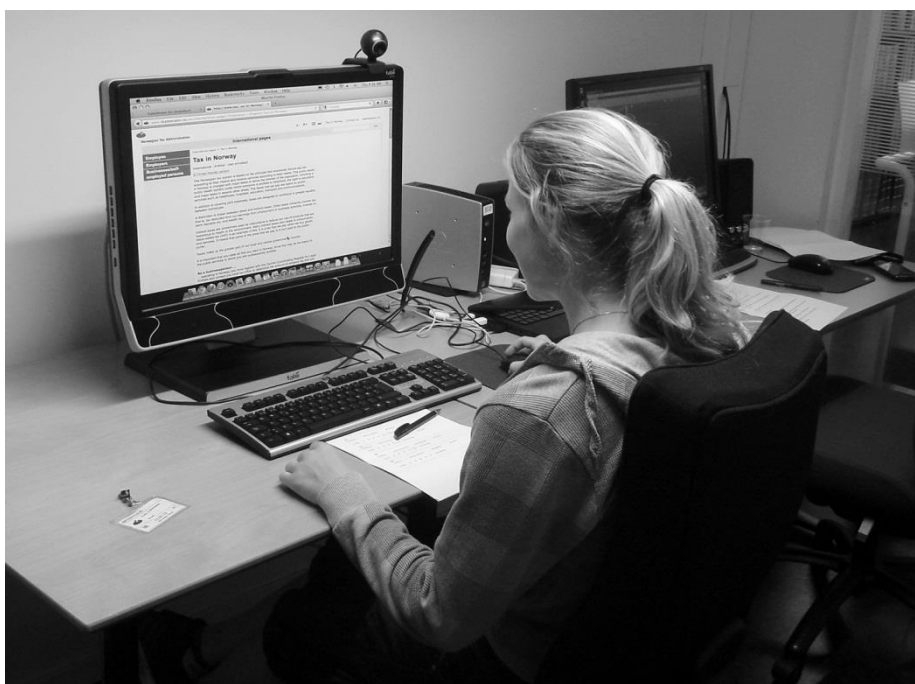


Fig. 5.1 – Bilde fra testlab, med en testdeltaker foran skjerm med integrert eye tracker, «Tobii T60XL».



## 5.5 Pilottest

Etter å ha designet studien, men før gjennomføring av de praktiske testene i laben, ble det satt opp en pilottest. Hensikten med pilottesten var som følger:

- Teste utstyret og se at alt teknisk fungerte som det skulle.
- Øvelse for testleder i å bruke utstyret og gjennomføre testen.
- Ta tiden på å klargjøre utstyret og på gjennomføring av testen.
- Tilbakemelding på formulering av testoppgavene, samtykkeskjema og informasjon.

Jeg fikk en venn til å stille opp som testdeltaker for pilottesten. Han var i samme aldersgruppe og med samme erfaring fra normal databruk som de andre testdeltakerene. Han hadde aldri vært testdeltaker i eye tracking-test før.

Pilottestdeltakeren ble bedt om å gi noe tilbakemelding underveis dersom han oppfattet oppgavene som uklare eller ikke forsto veiledningen han ble gitt. Etter gjennomføring av pilottesten kunne man da finpusse på formuleringen av testoppgavene slik at de ble bedre før man startet med ekte testdeltakere.

I tillegg ble det tatt tiden på alt som ble gjort under pilottesten. Klargjøring av laben og utstyret der, informasjonen til testbruker, og gjennomføring av opptaket.

Alt i alt var pilottesten svært nyttig. Det ble avdekket flere problemer som man da fikk mulighet til å rette opp i, testleder fikk øvelse i utstyret og testsituasjonen, og man fikk et tidsestimat på klargjøring og gjennomføring av test.

### 5.5.1 Resultat av pilottest

Først og fremst ble det oppdaget at det hadde blitt gjort oppdateringer på eye tracking-programvaren som skulle brukes sammen med eye trackeren, noe som gjorde at det oppsto litt tekniske problemer med oppsettet som måtte løses. Oppdateringen av programvare førte til at oppsett med egen Acer bærbare datamaskin ikke lenger fungerte optimalt, på grunn av skjermkort som ikke var kompatibelt med eye tracking-utstyret. Dette ble løst ved å få låne en bærbar Mac av kontaktperson hos Skatteetaten, som så ble brukt videre i alle testene.

Noen av oppgavene ble oppfattet som uklare av pilottesterene. Et par andre av oppgavene ble det observert at pilottesterene misforsto. Disse oppgavene ble så skrevet om i ettertid så de ble tydeligere formulert. Det ble også lagt til et par punkter i veiledningen som skulle gis til testbruker før opptaket, basert på observasjon fra pilottesten. Samtykkeskjema virket forståelig for pilottestbrukeren så dette ble beholdt slik det var.

Tidsestimat viste at klargjøring av lab, med oppsett og oppstart av alt utstyr, tok rundt 10 minutter. Derfor ble det alltid satt av minst et kvarter før første testdeltaker til å klargjøre laben når opptak skulle gjennomføres. Gjennomgang av testoppgavene tok omtrent like lang tid som forventet, bare noe raskere. Det ble derfor lagt et par korte oppgaver til i oppgavelisten i ettertid av pilottesten. Til aktuelle testdeltakere ble det anslått basert på pilottesten at testen ville ta 20 til 40 minutter, noe som stemte for de aller fleste.

## 5.6 Gjennomføring av tester

Praktisk sett besto hver test med testdeltaker av tre deler:

- Informasjon til testdeltaker og gjennomgang av samtykkeskjema.
- Kalibrering av utstyr.
- Gjennomføring av testoppgavene, med opptak.

### Informasjon og samtykke

Testen ble alltid startet med gjennomgang og underskrift på samtykkeskjema. Skjema hadde også blitt mailet til alle testdeltakere på forhånd, slik at de kunne lese gjennom det før de møtte opp. Samtykkeskjema inneholdt punkter om hva som ble tatt opp i testen, hvem som hadde tilgang til å se opptakene og når de skulle slettes.

Deltakerne fikk også en del informasjon før testen startet, om hva som skulle skje, hvordan utstyret fungerte, hvordan det var ønskelig at de oppførte seg under testen og hva de skulle gjøre i forskjellige situasjoner. All informasjonen ble gitt muntlig av testleder, basert på et manus slik at alle deltakerene skulle få samme informasjon. Kort oppsummert, gjengitt her i muntlig form men noe forkortet, ble de gitt disse retningslinjene:

«Hold kroppen noenlunde rolig, og ikke bytt sittestilling etter at kalibrering er gjennomført. Det er helt greit å bevege hodet.

Løs hver oppgave som om du skulle sittede alene, altså ikke tenk høyt, forklar hva du gjør eller spør spørsmål underveis, unntatt om oppgaven er uklar.

Etter hver oppgave, gi meg svaret muntlig eller si at du har fullført. Kryss så av på tilbakemeldingsskjema.

Dersom du vet svaret på en oppgave fra før, se om du likevel kan finne informasjonen på nettstedet.

Etter hver oppgave, stopp der du er på nettstedet og ikke gå tilbake med mindre jeg gir beskjed om det.»

Både i rekrutteringsprosessen og før hver test forklarte jeg at selv om vi brukte Skatteetatens utstyr og nettside drev jeg med et selvstendig studieprosjekt og jobbet ikke for Skatteetaten.

Det ble ikke gitt konkret informasjon om hva man ønsket å finne ut eller hva man skulle se på i analysen, utover det at øyebevegelsene deres ble registrert og tatt opp. Det ble presisert at de ikke ble vurdert i henhold til om de var «flinke» eller ikke, og bare skulle løse oppgavene i normalt tempo uten å føle press. Min oppfatning var at alle testdeltakerne forsto dette og følte seg komfortable tross labsituasjonen.

### **Kalibrering av utstyr**

Etter å ha gitt nødvendig informasjon måtte eye tracking-utstyret kalibreres for hver testperson. Ved å starte opptak fra Tobii Studio fikk bruker opp et vindu på sin skjerm, et sort felt med en hvit sirkel for hvert øye. Man måtte så finne en sittestilling slik at brukeren var passe langt unna skjermen og passe høyt/lavt i forhold til skjermen, noe man kunne se basert på de hvite sirkelene samt på en skala på hver side av feltet. Når dette var i orden startet kalibreringen, en rød prikk som brukeren skulle følge med blikket slik at utstyret fikk kalibrert hvor brukerens øyne var i forhold til kamera og i forhold til koordinatene til den røde prikken, for kalkulering av nøyaktige hvor på skjermen de så. Programmet ga så beskjed om kalibreringen var i orden eller ikke, og man kunne starte opptaket hvis alt var greit.

## **Gjennomføring av testoppgaver**

Etter kalibrering startet opptaket. Jeg sjekket så at alt fungerte som det skulle og at utstyret oppfattet øyebevegelsene til testdeltakeren før jeg ba deltakeren om å gå til Skatteetatens nettsider, og vi startet testen. Jeg som testleder satt i det samme rommet, ved siden av testdeltaker, men jeg satt ikke og så over skulderen deres. De ble så gitt en og en oppgave muntlig, og de sa ifra når de var ferdige. Etter hver oppgave krysset de av på tilbakemeldingsskjema før de ble gitt neste oppgave. Testdeltakerene fikk bruke så lang tid de ønsket på hver oppgave, og de fikk også velge selv dersom de ønsket å gi opp.

# 6 Analyse

## 6.1 Innhenting og bearbeiding av data

Tobii Studio gjorde opptak av testsesjonen som flere individuelle opptak: video av testbruker fra webkamera, skjermopptak fra testbrukers skjerm og rådata fra eye tracking-opptaket. Rådata besto av informasjon om deltakerens blikkmønster, lagret i programvarens filformat .tgd. Disse opptakene måtte aksesseres via programvaren Tobii Studio, som kombinerte dem og viste dem som et helhetlig opptak. Dette opptaket kunne så behandles videre i programvaren.

### 6.1.1 Prestasjon

Som nevnt tidligere kan man måle både brukerens prestasjon og tilfredshet i et forsøk med mennesker som testpersoner. Tilfredshet ble i dette forsøket til dels dekket ved et enkelt tilbakemeldingsskjema om oppfattet vanskelighetsgrad, som brukerne krysset av på underveis i testen. Videre analyse av hvorvidt testbrukerne var tilfredse med nettsidene var ikke relevant for målet med forsøket. Prestasjon ble målt i ettertid ved gjennomgang av opptak. Fem grunnleggende måter å måle prestasjon på er: oppgavesuksess, tidsbruk på oppgave, feil, effektivitet og læring (Tullis og Albert, 2008). I dette forsøket så man kun på de fire første av disse, og hovedsakelig på de tre første – altså hvorvidt testdeltakerene løste oppgaven de fikk korrekt og feilfritt, og hvor lang tid de brukte. Disse målingene ble så brukt for å støtte analysen av eye tracking-dataene. For eksempel kunne man se på tidsbruk på en spesiell oppgave og deretter sammenlikne øyebevegelsene til brukeren som brukte lengst og kortest tid på oppgaven. En annen mulighet var å sammenlikne deltakere som gjorde feil med de som løste oppgaven korrekt – kunne øyebevegelsene gi en indikasjon på hvorfor det skjedde? Direkte sammenlikning av de to gruppene er også mulig ved hjelp av disse målingene, sammenlikning av tidsbruk eller feilrate mellom ekspertgruppen og novisegruppen.

### 6.1.2 Tobii Studio

For å hente ut håndterbare data fra opptakenes rådata ble det brukt programvare fra produsenten av eye tracking-utstyret, Tobii. Programmet heter ”Tobii Studio” og var det

samme programmet som ble brukt for å gjøre selve opptaket. Her hadde man flere muligheter, de fire delene av programmet som ble brukt i analysen var ”Replay”, ”Visualisation”, ”Area of Interest” og ”Statistics”. Fig 6.1. viser et skjermbilde av programmet i delen ”Replay”. Her kunne man se gjennom hele opptaket med lyd, og med video fra webkamera som filmet brukeren under testen nede i hjørnet. Hovedbilde viser skjermopptak av brukerens skjerm underveis i opptaket, med øyebevegelsene som røde streker og sirkler. Sirklene viser fikseringer og har forskjellig størrelse basert på lengden på fikseringen, strekene viser øyebevegelsene mellom fikseringene. Det var her mulig å se på i raskere eller saktere tempo. Det var også her man måtte begynne for å kunne hente ut statistikk og for å lage visualiseringer av øyebevegelsene.

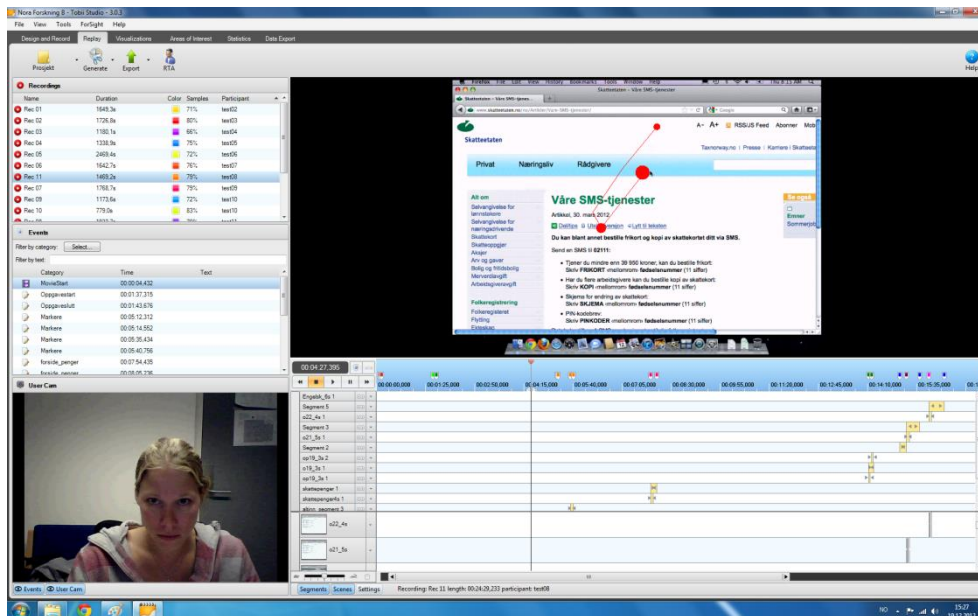


Fig. 6.1 – Tobii Studio, grensesnitt for å se gjennom opptak og generere segmenter og scener.

## Segmenter og scener

Under opptaket av brukeren er det en tidslinje, slik at man kan hoppe frem og tilbake i opptaket. På tidslinjen er alle såkalte hendelser markert. Disse hendelsene kunne legges inn underveis i opptaket eller, som det ble gjort i dette tilfellet, i ettertid. Hendelser kunne være ting som museklikk, start på en oppgave eller annet man måtte ønske å markere.

Under tidslinjen er det info om segmenter og scener. Segmenter er en kort eller lang del av opptaket. Man kan gi segmentet et eget navn og eksportere det som en avi-fil eller til en scene. Segmenter kan lages manuelt ved å markere med musen og dra segmentet kortere eller lengre, eller man kan be programvaren gjøre det ved hjelp av tidligere markerte hendelser. Jeg

la først inn hendelser, laget så segmenter basert på disse hendelsene, og la så segmenter til scener. Segmentene kunne man velge å lage automatisk for alle opptak med innlagte hendelser, men scener måtte manuelt opprettes i hvert enkelt opptak.

En scene er også en del av opptaket på samme måte som segment, men med ett konkret og statisk bakgrunnsbilde (Tobii Studio User Manual). I dette opptaket gjort på et nettsted vil det si deler av opptaket der brukerne så på en spesiell del av nettstedet uten å skifte side eller scrolle opp eller ned. Dersom man gjorde opptak med stillbilder ville ett stillbilde være én scene. Her møtte man på en utfordring i forhold til å hente ut talldata, da de fleste naturlige oppgaver på et stort nettsted innebærer at brukeren klikker seg gjennom flere undersider fra han får oppgaven til han finner svaret, samt scroller opp og ned på nettsiden. Dette ble løst på to måter: Noen av oppgavene var designet slik at brukeren burde kunne klare å løse dem uten å bevege på skjermbildet. Andre oppgaver ble det hentet ut talldata bare for deler av oppgaven, for eksempel de første 5 sekundene etter at oppgaven ble gitt, før brukeren begynte å scrolle nedover eller byttet til et annet skjermbilde.

Segmentene som skulle legges til scenene måtte være like lange for hver konkrete scene, for å kunne sammenlikne statistikk som antall fikseringer på den aktuelle oppgaven mellom brukere. På grunn av oppgaveløsningens dynamiske natur der brukerne kunne scrolle og skifte skjermbilde ble disse segmentene aldri særlig lange, de varierte fra 3 til 6 sekunder.

## **Visualisering og statistikk**

Når man hadde laget scener kunne disse brukes for å lage visualiseringer. Programmet genererte da gaze maps og heat maps. Når man først hadde scenene klare var dette en nokså enkel prosess. Man valgte scenen man ville visualisere, og fikk opp alle opptakene som hadde denne scenen i seg på venstre side av skjermen. Man haket så av de som skulle være med i visualiseringen, valgte type visualisering på toppen og endret eventuelt på innstillingene nederst. Når man var fornøyd var det bare å eksportere bildet. Figur 6.2 viser grensesnittet for å generere heat maps. For alle heat maps ble det valgt å vise relativ tidsbruk (Tobii Studio User Manual).

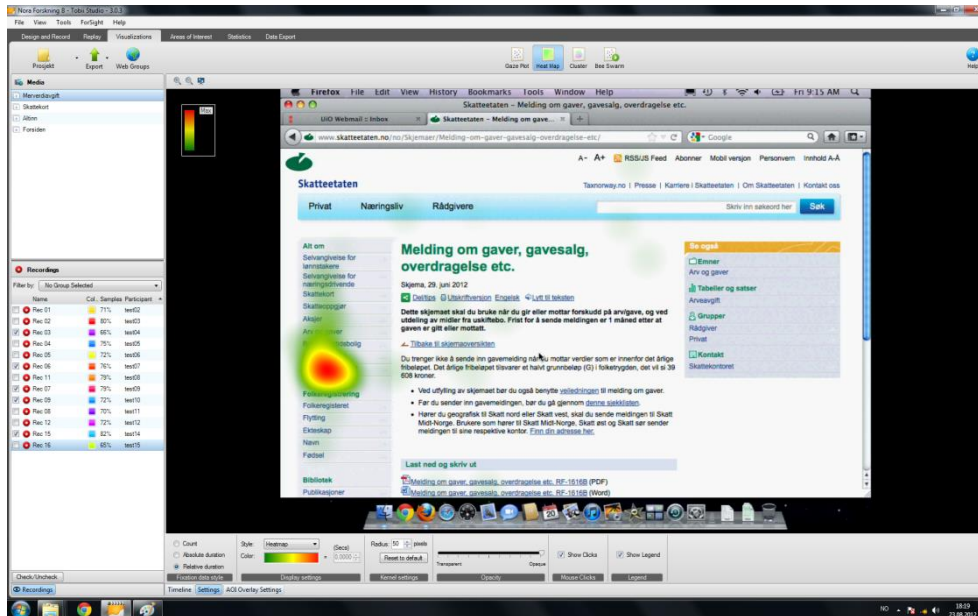


Fig. 6.2 – Tobii Studio, grensesnitt for å generere heat maps.

Area of Interest, AOI, var noe som måtte settes for å kunne få opp statistikk fra en scene. Man valgte da en scene og brukte verktøyene i programmet for å markere hvilke områder av skjermbildet som skulle være et AOI, altså et interessant område. En meny kunne for eksempel markeres som et interessant område. Fig 6.3 viser et grensesnittet for å angi AOI.

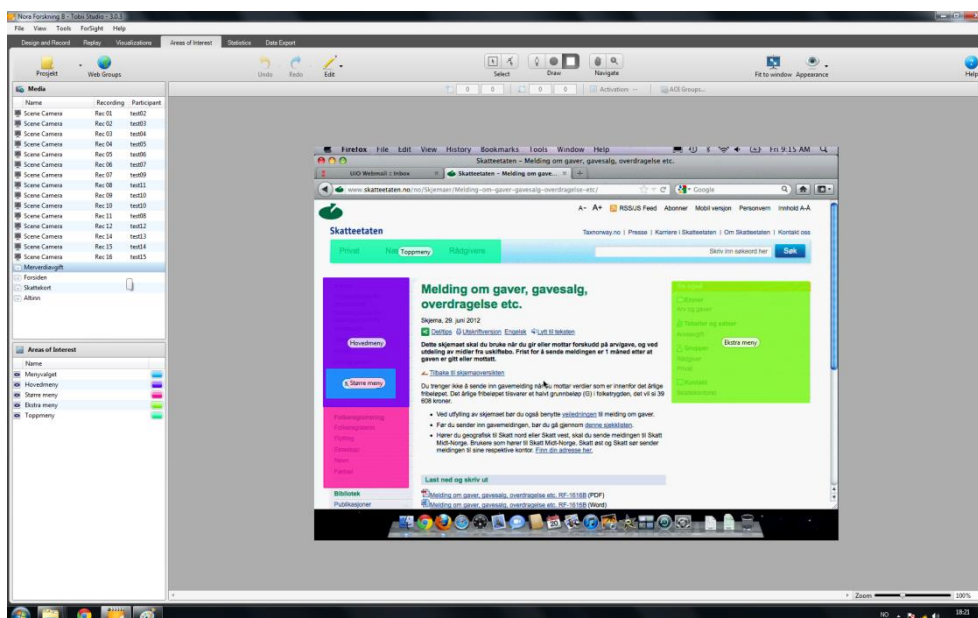


Fig. 6.3 – Tobii Studio, grensesnitt for å angi interesseområder (AOI).

Statistikkverktøyet ble brukt for å se på diverse data fra en spesiell scene med en eller flere definerte AOI. Først valgte man ut en eller flere scener. Deretter klikket man på de opptakene



man ønsket å ha med i statistikken. Til slutt valgte man hvilken statistikk som skulle vises, og eksporterte den som en tabseparert tekstfil. Denne tekstfilen kunne åpnes i Excel. Her ble det videre regnet ut sum og gjennomsnitt av hver måling, for hver gruppe brukere. Det ble også gjort en Student t-test i Excel med formelen TTEST(array1,array2,tails,type). «Tails» var satt til 2 for «tosidige fordeling» og «type» var satt til 2 for «to utvalg som antar lik varians» (Office, 2013).

Det ble valgt å generere statistikk om:

- *Time to First Fixation* – Tiden (i sekunder) fra starten av stimulu til deltakeren fikserer innenfor et gitt AOI (Tobii Studio User Manual).
- *Fixation Count* – Antall fikseringer deltakeren har innenfor et gitt AOI (Tobii Studio User Manual).
- *Total Visit Duration* – Samlet tid (i sekunder) på alle fikseringer til en deltaker innenfor et gitt AOI (Tobii Studio User Manual).
- *Percentage Fixated* – Prosentandel av deltakere som fikserte minst en gang innenfor et gitt AOI (Tobii Studio User Manual).

## Replay

I Replay kunne man se gjennom opptaket, eventuelt i saktere eller raskere tempo. I hovedvinduet i Tobii Studio kunne man da se øyebevegelsene som røde streker og sirkler over skjermvideo fra testsesjonen. Det var også mulig å eksportere deler av opptaket som en .avi-fil og se på i et annet program senere, med eller uten webkameraopptaket medfølgende.

Ved gjennomseeing av oppgavene konsentrerte testleder seg om en oppgave av gangen og så på alle brukerens løsning av oppgaven i både normal og saktere tempo. Det ble tatt notater om øyebevegelsene. Notatloggen liknet på en mer uformell versjon av Moore og Churchill (2011) sine transkript fra eye tracking-opptak. De var mindre formelle for eksempel i det henseende at de var uten konkrete tidspunkter for hver handling eller øyebevegelse. Det som ble observert og lagt merke til var fikseringer og sakkader, og spesielt søkestiene disse dannet.

Det ble ikke lagt vekt på videoopptaket av testdeltakeren som ble gjort med webkamera under opptak. Det ble ikke regnet som relevant eller interessant, en rask gjennomgang av opptakene

viser av testdeltakerene for det meste satt rolig og uten spesielle reaksjoner som kunne sees på videoen. Moore og Churchill (2011) sier også i forbindelse med sin studie at video av brukeren er mindre interessant for analysen.

### **6.1.3 Utvalg av oppgaver for videre analyse**

Hver testdeltaker fikk 27 oppgaver å løse. Resultatene fra syv av disse oppgavene ble gått gjennom og er presentert i neste delkapittel, 6.2. To ting ble lagt vekt på ved utvalg av hvilke oppgaver som skulle presenteres her:

- Mulighet for generering av visuelle data og statistikk i Tobii Studio.
- Oppgavens vanskelighetsgrad basert på tilbakemelding fra brukerne.

Som nevnt tidligere var det nødvendig for å kunne opprette scener, at alle testbrukerne oppholdt seg på det samme skjermbilde den samme perioden etter at oppgaven ble gitt. For å kunne gjøre visualiseringer og få frem statistikk måtte man ha en scene som grunnlag. På grunn av dette var det ikke alle oppgavene som egnet seg til å generere slike data fra i Tobii Studio. Det ble generert slik data fra oppgave 4, 11, 14, 19, 21, 22 og 26.

Etter hver oppgave brukerne løste krysset de av for hvor vanskelig de syntes oppgaven var, på en skala fra 1 til 5. Hele tabellen med testdeltakerenes tilbakemelding på vanskelighetsgrad finnes i vedlegg D. Ut i fra dette kunne man se at oppgave 2 og 19 ble ansett som enklest, mens oppgave 26 og 24 ble ansett som vanskeligst.

## **6.2 Resultater**

### **6.2.1 Forklaring til presentasjon av resultater**

I de følgende underkapittelene følger resultater fra de syv oppgavene, presentert en oppgave av gangen.

Oppgaveteksten testdeltakerene fikk kommer først i hvert avsnitt, i kursiv, etterfulgt av skjermbilde og forklaring rundt oppgaven. Deretter følger generelle observasjoner om oppgaven fra opptakene, visualiseringer, statistikk og en sammenlikning av gruppe 1 versus gruppe 3 på den aktuelle oppgaven. Gruppe 1 besto av deltakerene uten bakgrunn fra

interaksjonsdesign – novisegruppen. Gruppe 3 var deltakerene med denne bakgrunnen – ekspertgruppen. Gruppe 2 var deltakerene med kun noe bakgrunn. Deres opptak er også gjennomgått, men ikke tatt med i den direkte sammenlikningen av gruppe 1 og gruppe 3. Når det refereres til «alle testdeltakerene» menes kombinasjonen av gruppe 1, 2 og 3. Dersom det kun er snakk om data fra én av gruppene refereres det til den aktuelle gruppen ved tall. Mer om hvordan testdeltakerene ble delt opp i disse tre gruppene er forklart i kapittel 5.2.2.

Visualiseringene består av heat maps og gaze plots, som er beskrevet nærmere hva betyr og hvordan er fremstilt i henholdsvis kap. 2.1.5 og 6.1.2. Det er opplyst for hver oppgave hvor mange sekunder av opptaket som er brukt som grunnlag for visualiseringene. Det samme gjelder statistikken.

Statistikk er presentert som en slik tabell. Tallenes betydning er forklart nedenfor.

	Gruppe 1			Gruppe 3			T-test
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	p
<b>Fixation count</b>							
AOI	4	16	4,00	4	23	5,75	0,498

Rader: Hver rad i tabellen viser målinger fra et bestemt interesseområde, kalt AOI. Hva som måles står med fet tekst over AOI. For hver oppgave er det før tabellen inkludert en figur med skjermbilde der alle AOI er markert med farger. Alle AOI er ikke inkludert i tabellen, det ble først og fremst valgt ut de der  $n \geq 3$ .

Kolonner: Det er tre kolonner for hver gruppe, og en kolonne for t-test. De tre kolonnene i hver gruppe viser verdiene  $n$ ,  $sum$  og  $gj.snitt$ . Verdien  $n$  står for antall opptak som danner grunnlaget for sum og gjennomsnitt.  $Sum$  er summen av disse opptakene.  $Gj.snitt$  står for gjennomsnittet, altså  $sum/n$ . T-test indikerer hvor signifikant forskjellen på gruppene er, og der  $p \leq 0,05$  er det markert med fet skrift. Hvordan t-testen er regnet ut er beskrevet i kapittel 6.1.2.

Hensikten med denne statistikken er først og fremst sammenlinkning av gruppe 1 og 3, derfor er det ikke inkludert statistiske data for gruppe 2 i tabellen.

Sammen med visualisering av AOI for oppgaven er det inkludert verdier for *percentage fixated*. Dette tallet viser til prosentandel av  $n$  som fikserte minst en gang innenfor et gitt AOI

i løpet av tidsperioden. Verdien  $n$  viser også her til antall opptak som dannet grunnlaget for prosentverdien, som i utgangspunktet skal være alle testdeltakerene. Verdien for hvert AOI er uavhengig, da en testbruker kan ha fiksert på flere eller ingen AOI.

## 6.2.2 Oppgave 4

*På denne siden, finn menyvalget som tar deg til siden om moms og klikk på det.*

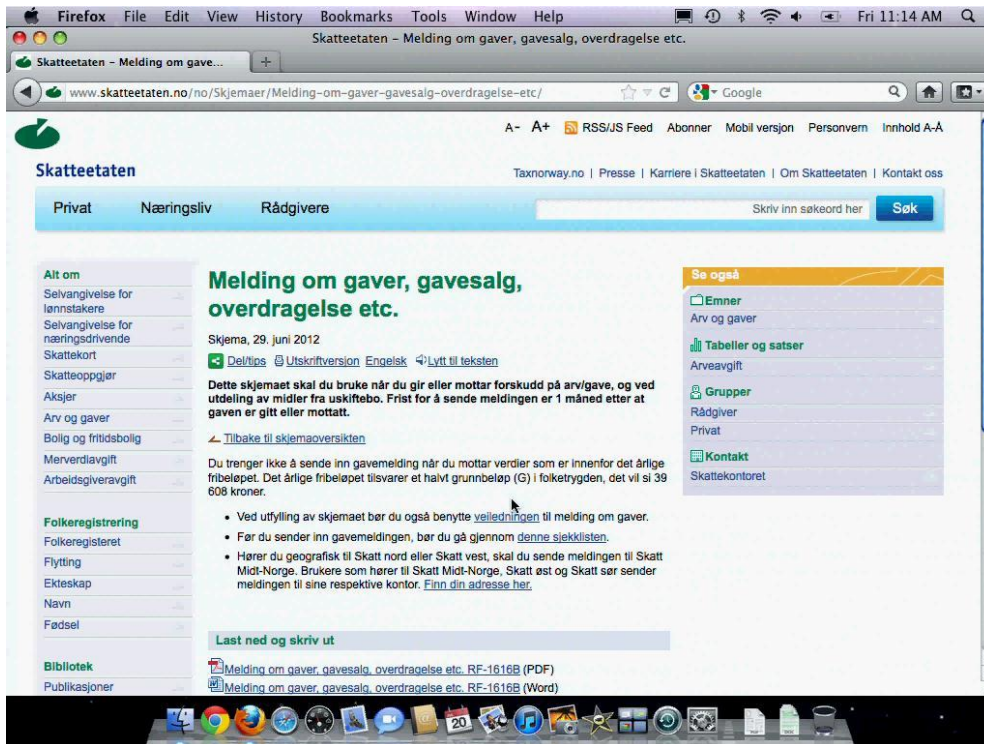


Fig. 6.4a – Skjermbildet brukeren sto på idet de fikk oppgave 4. Svaret på oppgaven kunne finnes i menyen til venstre.

Dette var en kort oppgave der brukeren ble bedt om å finne et menyvalg og klikke på det. Det var den andre oppgaven av denne typen i oppgavelisten, og begge oppgavene hadde svaret i hovedmenyen på venstre side. Forskjellen var at de i den første oppgaven ble brukt et ord fra menyvalget i oppgaven, «*På denne siden, finn menyvalget som tar deg til siden om arveavgift og klikk på det.*» (Menyvalget heter «Arv og gaver»), mens de her ble bedt om å finne valget for «moms» mens det i menyen står «Merverdiavgift».

Deltakerene sto på siden vist i figur 6.4a når oppgaven ble gitt.

## Generelle observasjoner

Testdeltakerene scorete oppgaven som ganske lett, 1,73 på en skala fra 1 til 5. Alle trykket på riktig lenke, uten å trykke på noe galt først.

9 av 13 personer brukte under 10 sekunder på å løse oppgaven. Av de resterende var det en som brukte betydelig lenger tid enn de andre med 31 sekunder. Raskeste svar var 3,6 sekunder.

Generelt fokuserte deltakerene raskt på venstre meny der svaret var, kalt hovedmenyen. De scannet deretter raskt opp og ned før de fant svaret og trykket på det. Man kunne kanskje forventet at de ville scanne sakte fra øverste menyvalg og nedover, en systematisk leting, men det var mer vanlig å scanne raskt opp og ned i menyen. Noen fokuserte noe lenger på svaret før de trykket enn andre, og noen fokuserte på det mer enn en gang, mens andre så ut til å bestemme seg ved første fokusering. Dette kan man anta kom av at noen var usikre på om Merverdiavgift var lenken som samsvarte med oppgavens spørsmål om å finne info om moms, grunnet forskjellig ordvalg. Testdeltakeren som brukte lengst tid fokuserte på riktig menyvalg flere ganger før han valgte, og oppga også muntlig at ordbruken forvirret ham.

## Visualiseringer

De følgende visualiseringene er fra de første 3 sekundene etter at oppgaven ble gitt.

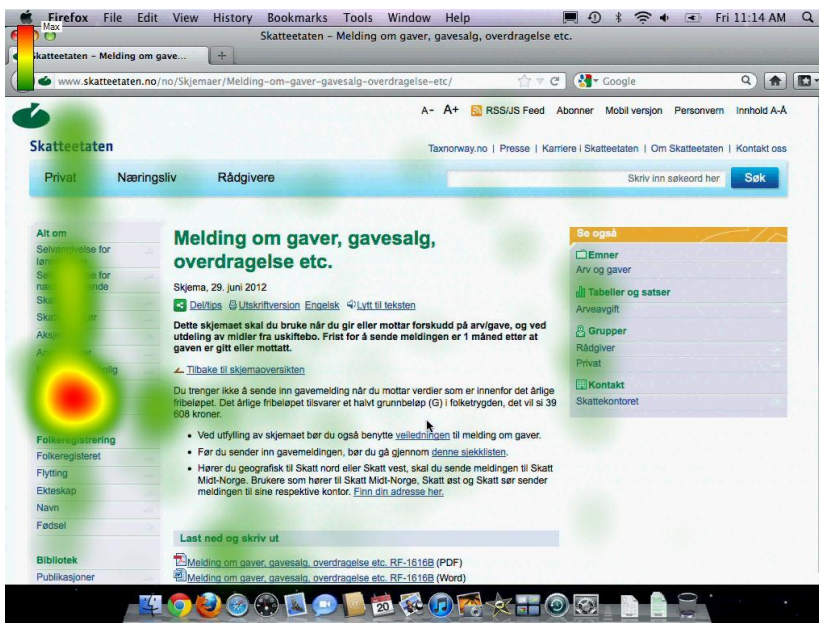


Fig. 6.4b – Heatmap. Alle. Fra deltakerene ble gitt oppgaven til de enten trykket riktig eller scrollet, altså ikke lik tid for alle. Heatmap viser klart at det ble fokusert hovedsaklig på meny og aller mest fokus der det riktige menyvalget «Merverdiavgift» befinner seg.

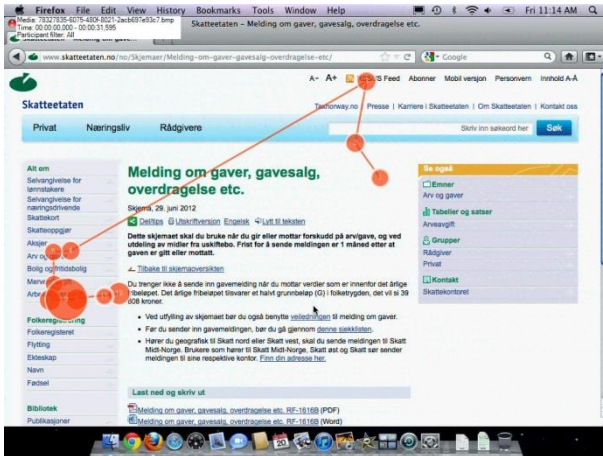


Fig. 6.4c – Gaze plot. Testdeltaker 6

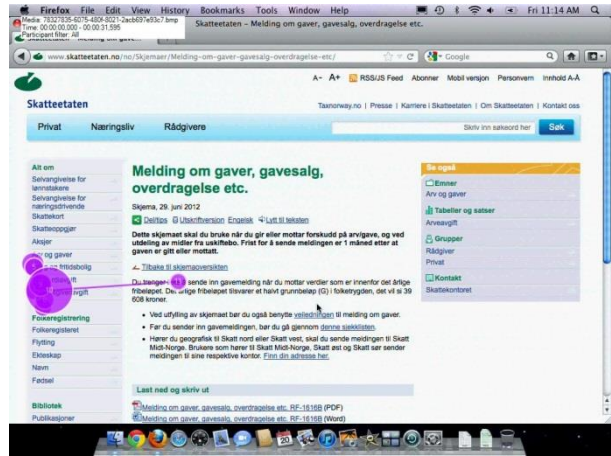


Fig. 6.4d – Gaze plot. Testdeltaker 3. Dette var testdeltakeren med raskest tid.

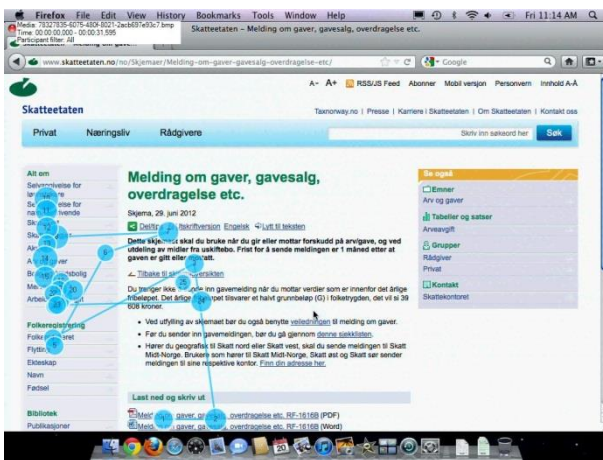


Fig. 6.4e – Gaze plot. Testdeltaker 12

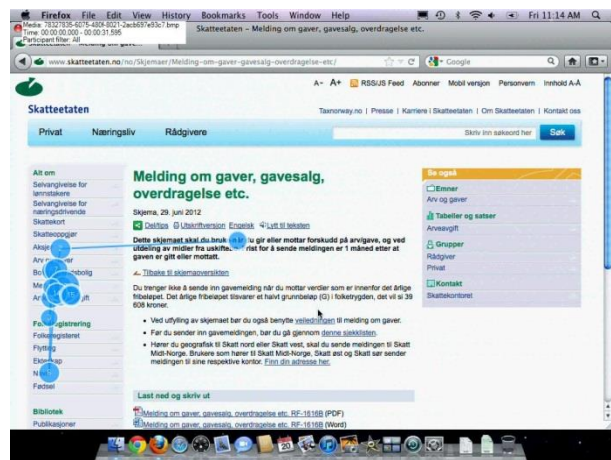


Fig. 6.4f – Gaze plot. Testdeltaker 9

Fordi mange av testdeltakerene scrollet noe ned raskt etter at oppgaven ble gitt var det ikke mulig å hente ut heatmaps eller statistikk for å sammenlikne gruppe 1 og gruppe 3.

### Gruppe 1 vs. gruppe 3

Gjennomgang av opptak ga ikke inntrykk av at gruppe 1 og 3 løste denne oppgaven forskjellig. Gruppe 1 løste den generelt noe raskere enn gruppe 3. Hvor lenge hver testdeltaker fokuserte på menyvalget før han klikket varierte noe mellom deltakerene, fra svært kort fiksering til medium lang fiksering, men det var variasjoner innenfor hver gruppe, ikke mellom gruppene. Ingen andre blikkmønster utpekte seg heller som forskjellig, med unntak av en detalj. Det var at ingen av de fire testdeltakerene fra gruppe 1 fokuserte på menyen «Se også» til venstre, mens tre av fem fra gruppe 3 gjorde dette. Hva kan dette bety? Det kan være tilfeldig, eller det kan være at gruppe 3 med sin erfaring i større grad er klar over at det kan være nyttig informasjon i en slik fremhevet ekstrameny.

## 6.2.3 Oppgave 11

*Du husker ikke passord for MinID. Finn telefonnummeret du kan ringe for å bestille eller få info om skattekort istedet.*

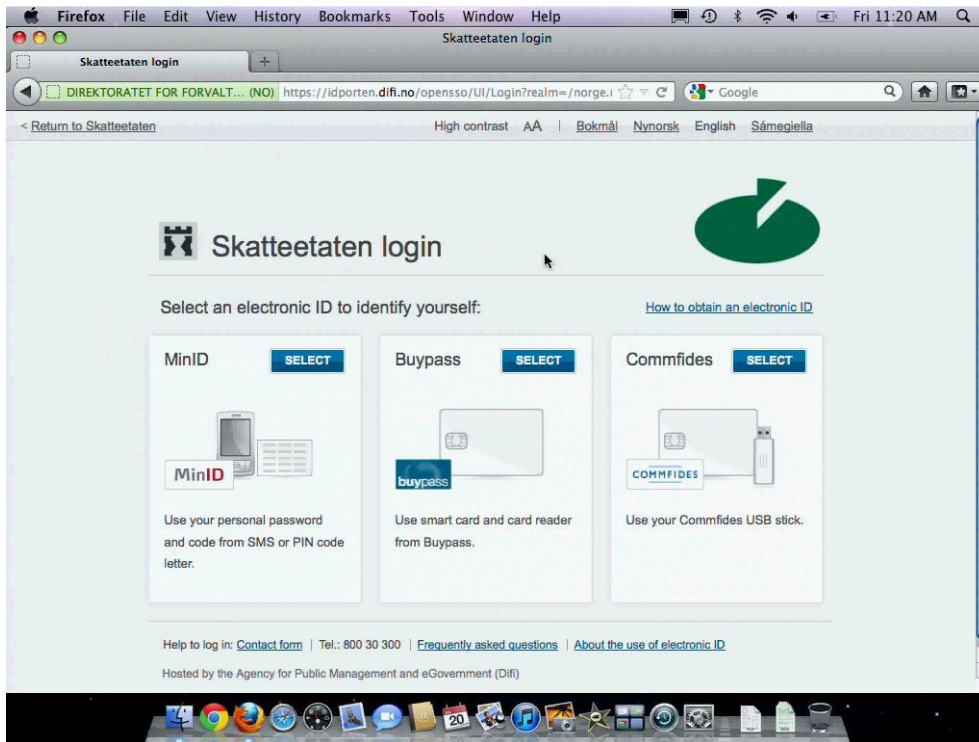


Fig. 6.5a. Skjermbildet for login på Altinn, der testdeltakerene befant seg når oppgaven ble gitt, fordi oppgaven før gikk ut på å finne skjema for å endre skattekort, som førte dem hit.

Denne oppgaven ble gitt etter at deltakerene hadde navigert til Altinn sin login-side, fordi de skulle finne skjema for å bestille skattekort. De sto altså på siden i figur 6. da oppgaven ble gitt.

Denne oppgaven ble inkludert fordi man antok at denne siden kan oppfattes som forvirrende. Brukere blir sendt til denne Altinn-siden når de trykker på lenken for å bestille skattekort på skatteetaten.no. Kontaktinformasjonen som står nederst på denne siden er altså til Altinn og gjelder problemer med login – ikke problemer med skattespørsmål/skattekort. Da må man navigere tilbake til skatteetaten.no for å finne riktig kontaktinformasjon. Lenken tilbake til Skatteetaten er med nokså liten skrift øverst i venstre hjørne, og det er ikke eksplisitt merket at man har gått fra Skatteetatens sider til Altinn. Dette er nok gjort hensiktsmessig for at brukeren skal oppfatte de to tjenestene som sømløse og integrerte, men kan kanskje også skape unødig forvirring.

Det er sannsynlig å anta at labsituasjonen påvirket resultater noe her, da testdeltakerene kanskje ønsket å finne svaret kjapt, og derfor ikke leste så nøye som de muligens ville gjort hvis de skulle finne ut av det samme hjemme. Man ville likevel gjerne sammenlikne øyebevegelsene til testdeltakerene som svarte nummeret nederst på skjermen og de som svarte riktig (altså returnerte til skatteetaten.no for å finne telefonnummeret der)

### **Generelle observasjoner**

Oppgavens vanskelighetsgrad ble rangert som 2,13 på en skala fra 1 til 5. Tidsbruk var jevnt fordelt mellom 4 og 20 sekunder.

Fem testdeltakere oppga nummeret nederst på siden som svar, altså galt. Syv av testdeltakerene forlot siden for å finne nummeret på riktig sted på Skatteetatens egne sider. En testdeltaker hadde ikke målbare resultater. Figur 6.5c og 6.5d viser heat map av de første 4 sekundene fra de som svarte galt og de som svarte riktig. Her ser man at de som svarte galt generelt hadde flere fokuseringer på midten av siden, og ingen øverst til venstre der man kunne returnere til Skatteetaten. Både de som svarte riktig og de som svarte feil fokuserte i stor grad på telefonnummeret nederst. Statistikken bekrefter dette, 75% av alle testdeltakerene fokuserte på AOI med telefonnummeret i løpet av de første 4 sekundene, men bare 25% fokuserte på AOI med tilbake-lenken. Dette kan tilsa at informasjonen tiltrakk alles oppmerksomhet, men at den ble prosessert forskjellig da noen oppfattet det som riktig svar og andre ikke.



## Visualiseringer

De følgende visualiseringene er fra de første 4 sekundene etter at oppgaven ble gitt.

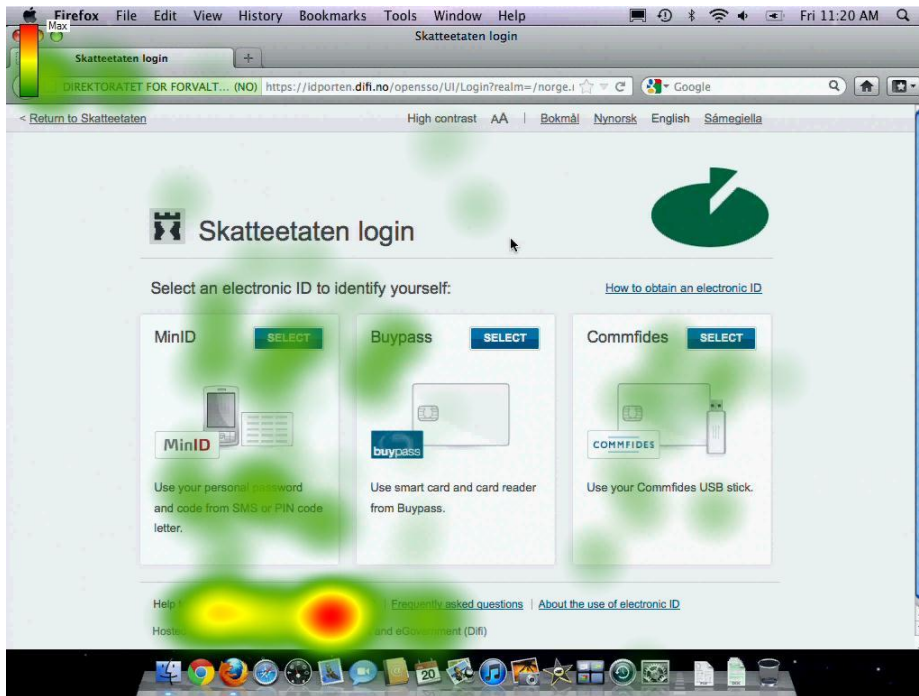


Fig. 6.5b – Heat map. Alle testdeltakerene.

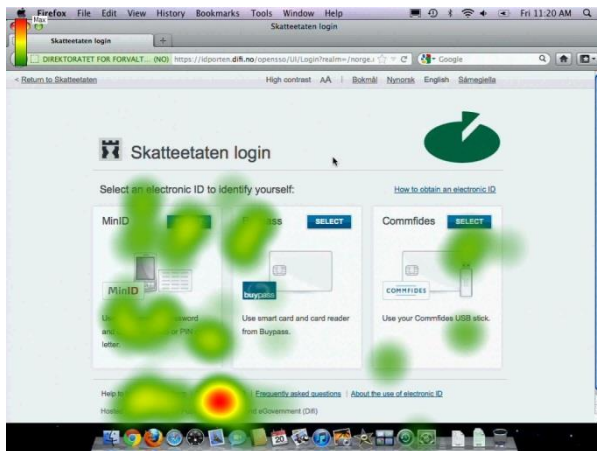


Fig. 6.5c – Heat map. Testdeltakerene som svarte feil.

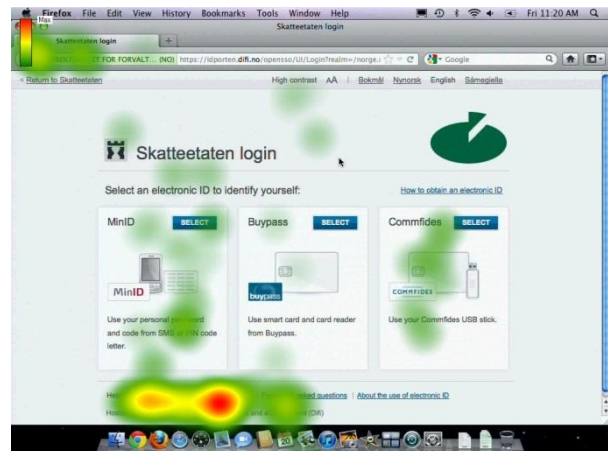


Fig. 6.5d – Heat map. Testdeltakerene som svarte riktig.

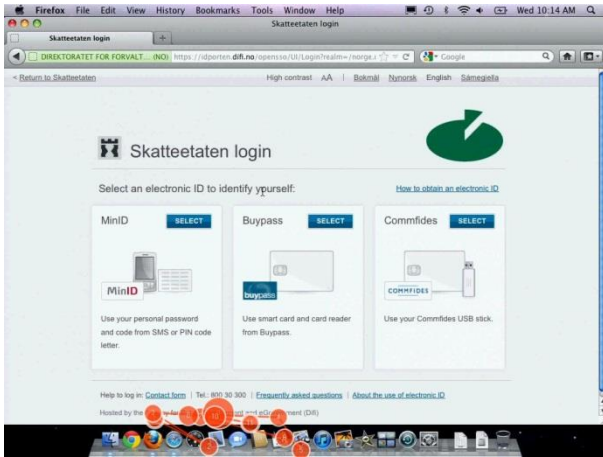


Fig. 6.5e – Gazeplot. Testdeltaker 6, som løste oppgaven korrekt.

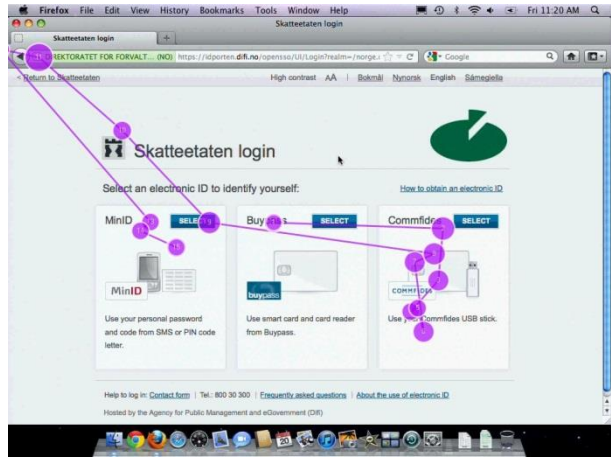


Fig. 6.5f – Gazeplot. Testdeltaker 4, som løste oppgaven korrekt.

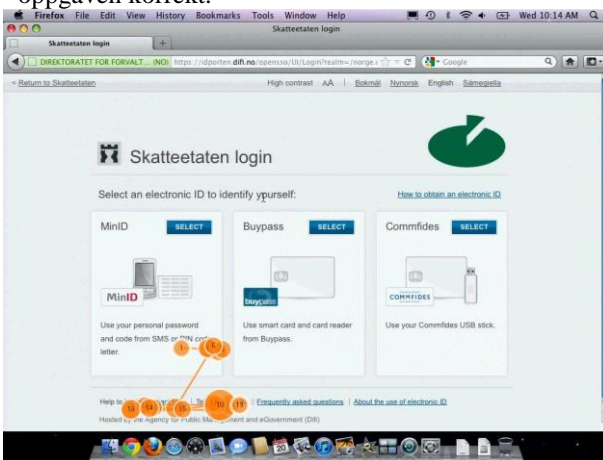


Fig. 6.5g – Gazeplot. Testdeltaker 7, som svarte feil.

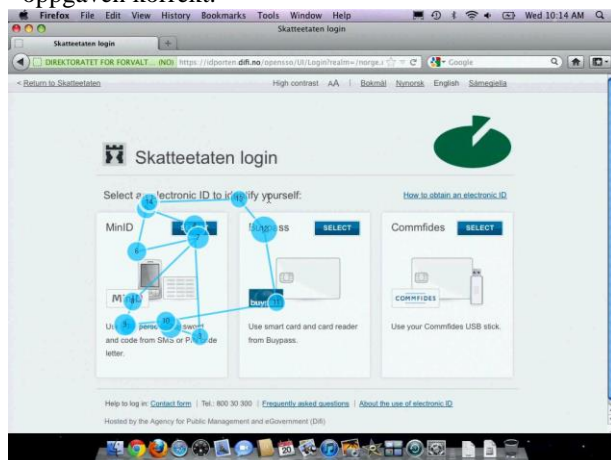


Fig. 6.5h – Gazeplot. Testdeltaker 12, som svarte feil.

## Statistikk

Følgende statistikk er fra de første 4 sekundene etter at oppgaven ble gitt.

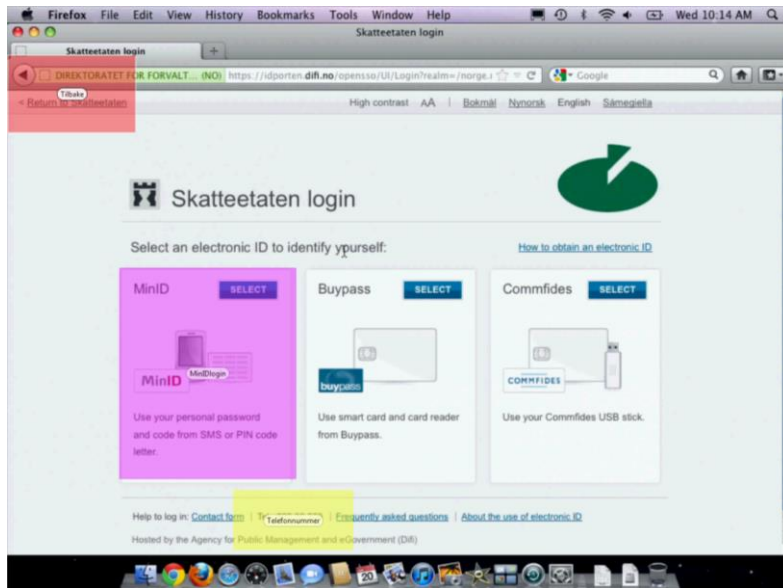


Fig. 6.5g – Oversikt over AOI.

### Percentage Fixated

$n=12$

AOI	Prosent
MinIDlogin	67 %
Telefonnummer	75 %
Tilbake	25 %

	Gruppe 1			Gruppe 3			T-test
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	p
<b>Fixation count</b>							
MinIDlogin	4	18	4,50	5	10	2,00	0,247
Telefonnummer	4	11	2,75	5	21	4,20	0,483
Hele	4	55	13,75	5	78	15,6	0,441
<b>Total visit duration (i sek)</b>							
MinIDlogin	4	3,79	0,95	5	2,05	0,41	0,298
Telefonnummer	4	3,27	0,82	5	4,94	0,99	0,746
<b>Time to first fixation (i sek)</b>							
MinIDlogin	3	1,85	0,62	3	5,31	1,77	0,341
Telefonnummer	2	3,26	1,63	4	1,64	0,41	<b>0,001</b>

### Gruppe 1 vs. gruppe 3

Når det gjaldt suksessrate var det en forskjell mellom gruppe 1 og gruppe 3. Alle fra gruppe 3 løste oppgaven på korrekt måte ved å forlate Altinn-siden og finne nummeret på Skatteetatens egne sider. Bare én fra gruppe 1 gjorde dette, de tre andre svarte feil.

Ser vi på statistikken er det et punkt der p tilsier en forskjell mellom gruppene med en verdi på 0,001. Det gjelder tid til første fiksering på AOI «Telefonnummer», der gruppe 3 i gjennomsnitt fokuserte på området tidligere enn gruppe 1. Det er derimot ingen signifikant forskjell mellom antall fikseringer og tid brukt innenfor AOI «Telefonnummer», som tilsier at begge gruppene fokuserte i like stor grad på disse områdene de første 4 sekundene.

Disse to observasjonene hentyder at gruppe 3 raskt oppfattet at det var informasjon nederst i skjermbildet, men klarte å prosessere denne informasjonen og trekke en korrekt konklusjon om at det ikke var den de var ute etter.

## 6.2.4 Oppgave 14

(Naviger til forsiden)

*Du har fått igjen på skatten og lurer på når du får pengene på konto. Finn informasjon om hvilken dato du får pengene dine.*

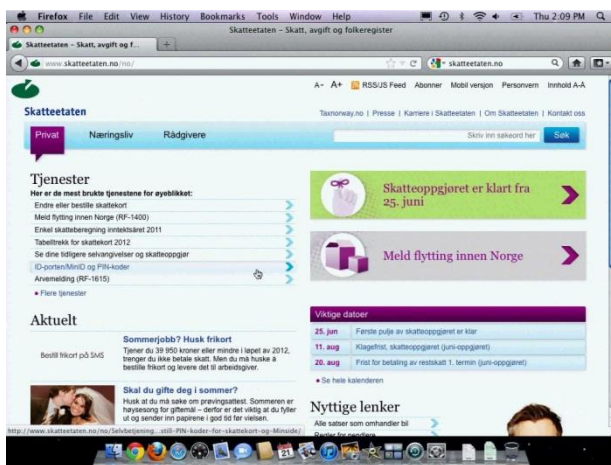


Fig. 6.6a – Versjon 1 av forsiden med info om skatteoppgjøret.

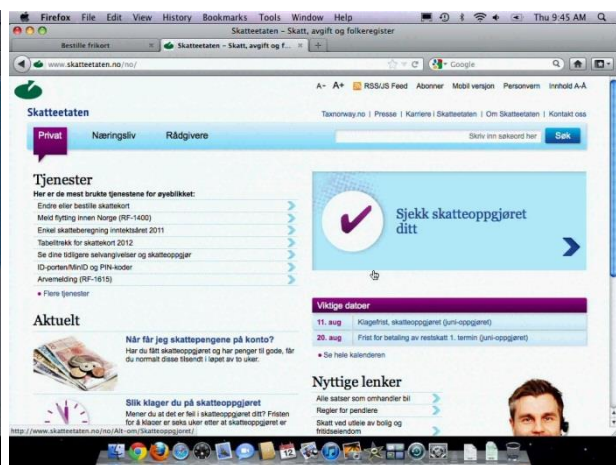


Fig. 6.6b – Versjon 2 av forsiden med info om skatteoppgjøret.

Her var det flere måter å finne svaret på. Man kunne finne det under Tjenester, under Viktige Datoer, under Aktuelt eller ved å klikke på den store boksen på forsiden. Skatteetaten endret forsiden sin noe midt i testen min så halvparten av testdeltakerene hadde den ene og halvparten den andre, de kan sees i henholdsvis figur 6.6a og 6.6b. Den store snarveien var plassert på samme sted, men i versjon 2 het lenken øverst i Aktuelt-feltet «Når får jeg skattepengene på konto?». Svaret kunne finnes på flere undersider, blant annet siden «Skatteoppgjør» og siden «Når får jeg pengene på konto?».

Et aspekt som var interessant her var å se hvorvidt den store fargede boksen trakk brukernes oppmerksomhet slik den nok er ment å gjøre.

## Generelle observasjoner

Oppgavens vanskelighetsgrad ble rangert som 1,67 på en skala fra 1 til 5. Tidsbruk var jevnt fordelt mellom 4,7 og 15 sekunder fra de fikk oppgaven til de valgte en lenke på forsiden som de trykket på. Deretter brukte alle litt tid på å lese teksten for å finne svaret.

Det var altså mulig å finne svaret på en av flere sider som var tilgjengelig via kun ett klikk fra forsiden. Alle unntatt en testdeltaker oppnådde dette.

Nesten alle testdeltakerene, 92 %, fokuserte på området med liste over tjenester iløpet av de første 4 sekundene. 62 % fokuserte på det store fargede feltet iløpet av samme tiden. Heat map fra alle deltakerene viser tre hovedpunkter for fokus, feltet med liste over tjenester, feltet med viktige datoer og den fargede snarveiboksen. Det er flest fikseringer på snarveien.

## Visualiseringer

De følgende visualiseringene er fra de første 4 sekundene etter at oppgaven ble gitt.

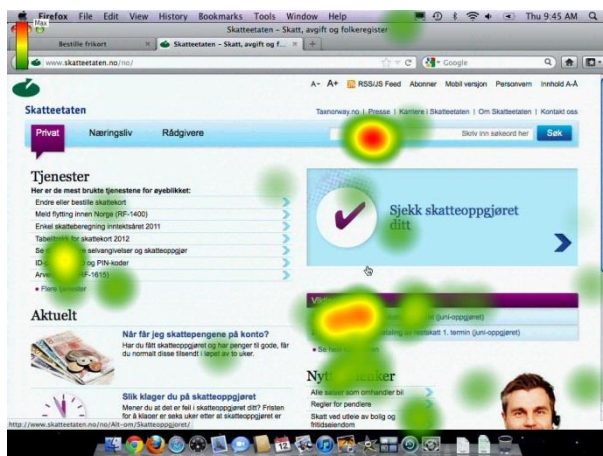


Fig. 6.6c – Heat map. Gruppe 1

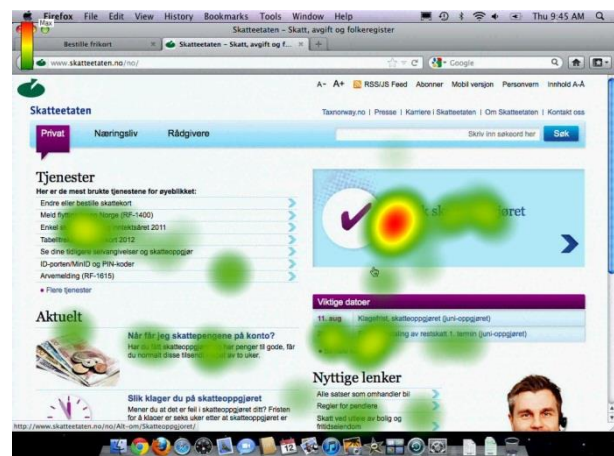


Fig. 6.6d – Heat map. Gruppe 3

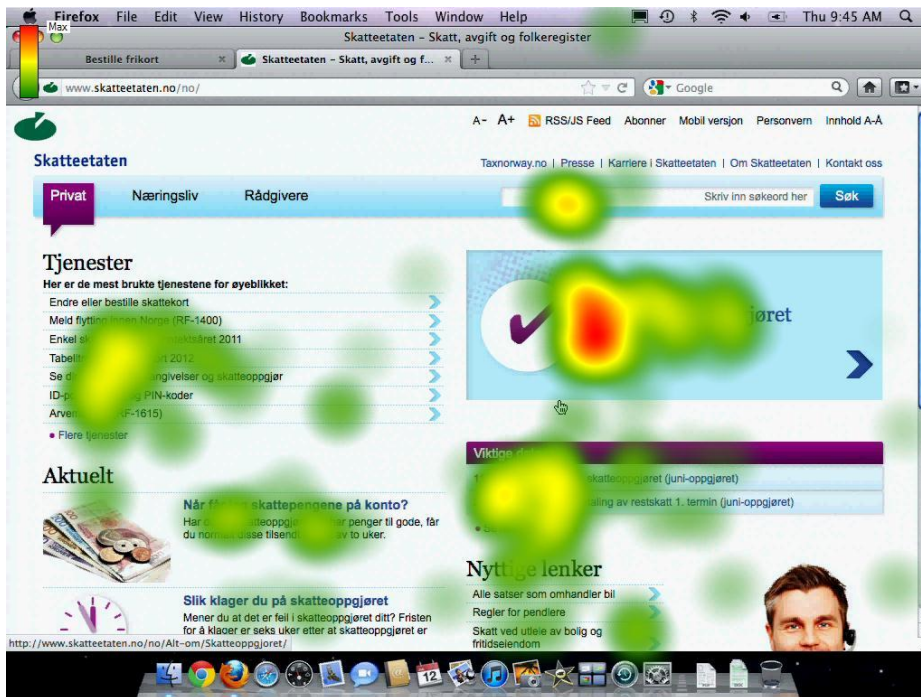


Fig. 6.6e – Heat map. Alle testdeltakerene.

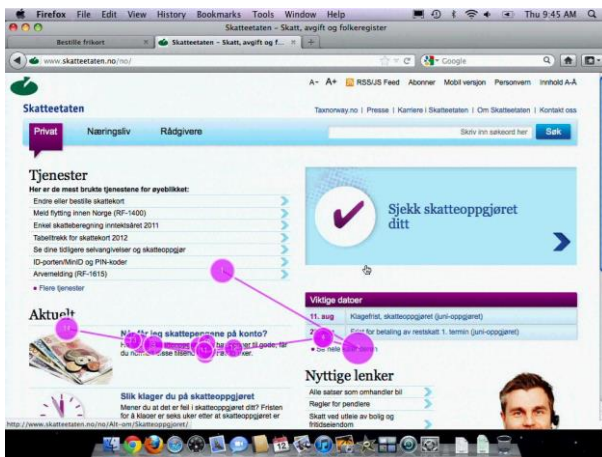


Fig. 6.6f – Gazeplot. Testdeltaker 11, fra gruppe 3. Testdeltakeren som brukte kortest tid fra forsiden til riktig underside.

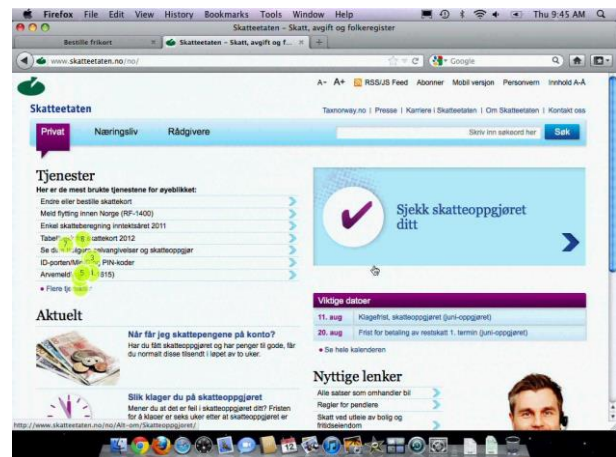


Fig. 6.6g – Gazeplot. Testdeltaker 13, fra gruppe 2. Testdeltakeren som brukte lengst tid fra forsiden til riktig underside.

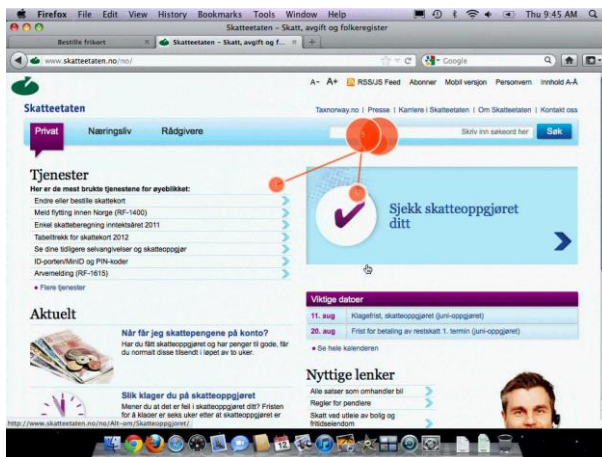


Fig. 6.6h – Gazeplot. Testdeltaker 6, fra gruppe 1. Den eneste som ikke klikket seg direkte inn på korrekt svar.

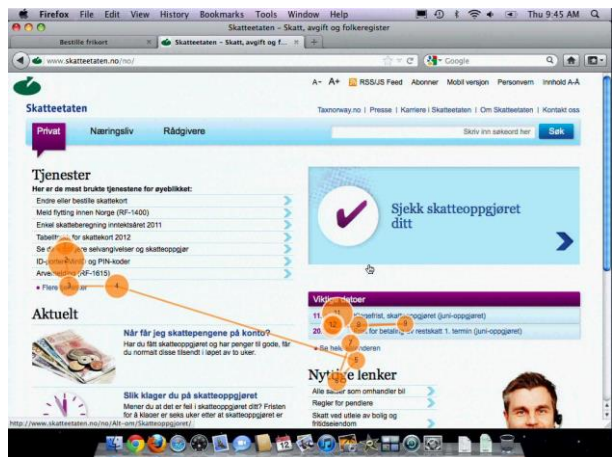


Fig. 6.6i – Gazeplot. Testdeltaker 7, fra gruppe 1.

## Statistikk

Følgende statistikk er fra de første 4 sekundene etter at oppgaven ble gitt.

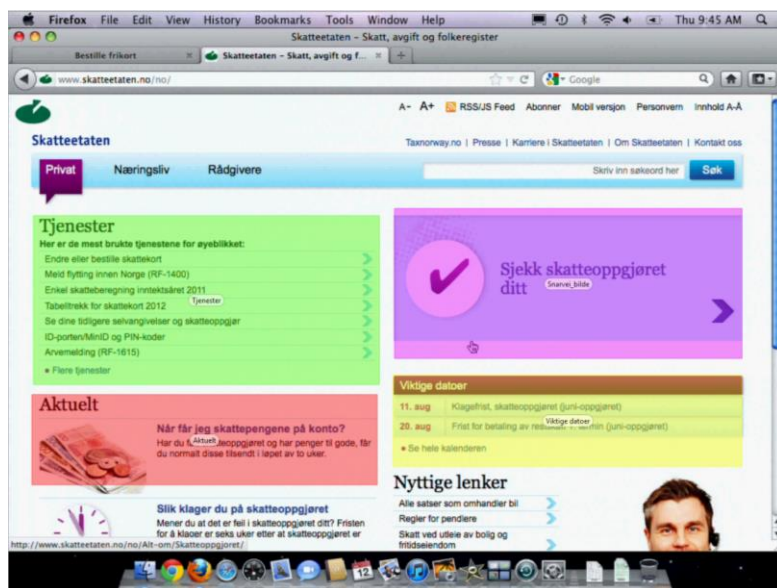


Fig. 6.6f– Oversikt over AOI.

## Percentage Fixated

$n=13$

AOI	Prosent
Aktuelt	31 %
Snarvei_bilde	62 %
Tjenester	92 %
Viktige datoer	38 %

	Gruppe 1			Gruppe 3			T-test
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	p
<b>Fixation count</b>							
Snarvei_bilde	4	2	0,50	5	31	6,20	0,117
Tjenester	4	6	1,50	5	15	3,00	0,550
Viktige datoer	4	19	4,75	5	13	2,60	0,437
Hele	4	45	11,25	5	75	15,00	0,109
<b>Total visit duration (i sek)</b>							
Snarvei_bilde	4	0,40	0,10	5	8,45	1,69	0,143
Tjenester	4	2,00	0,50	5	3,36	0,67	0,805
Viktige datoer	4	5,56	1,39	5	3,32	0,66	0,293
<b>Time to first fixation (i sek)</b>							
Snarvei_bilde	2	6,19	3,10	3	0,57	0,19	<b>0,012</b>
Tjenester	3	0,00	0,00	5	3,61	0,72	0,480
Viktige datoer	3	5,35	1,78	2	1,02	0,51	<b>0,004</b>

## Gruppe 1 vs. gruppe 3

Hensikten med et stort farget felt med snarvei til en aktuell tjeneste er som nevnt for å tiltrekke seg brukerens oppmerksomhet. Ser vi på tid til første fiksering i dette AOI, som er antatt som en verdi for hvor godt området trekker oppmerksomhet (Ehmke og Wilson, 2007),

ser vi at gruppe 3 har signifikant kortere tid til første fiksering enn gruppe 1,  $p = 0,012$  De har også noe høyere fikseringsantall i dette AOI og noe høyere total tid brukt innenfor AOIet iløpet av de første 4 sekundene. Heat map fra gruppene, figur 6.6c og 6.6d, støtter disse tallene. De viser videre at gruppe 1 heller fokuserte på menyen for viktige datoer, AOI «Viktige datoer».

En annen signifikant forskjell, der  $p = 0,004$ , er for tid til første fiksering på AOI «Viktige datoer». Selv om gruppe 1 har noen fler fikseringer i området fokuserte gruppe 3 på det raskere enn det gruppe 1 gjorde.

Det var ingen spesiell forskjell på hvor lang tid det tok for de to gruppene å klikke på en av lenkene på forsiden som førte dem til riktig svar.

## 6.2.5 Oppgave 19

(Naviger tilbake til siden med generell kontaktinformasjon til Skatteetaten)

Finn navnet på Skatteetatens Twitter-konto.

The screenshot shows the Skatteetaten website's contact page. At the top, there is a navigation bar with 'Privat', 'Næringsliv', and 'Rådgivere' tabs. A search bar is located to the right of these tabs. The main content area is titled 'Her finner du oss' and contains the following information:

- Telefon:** 800 80 000 (08:00 til 15:00). Det er gratis å ringe, men for mobiltelefon gjelder egne satser. Fra utlandet: +47 22 07 70 00.
- Jus dåntut veahki sáamegilli, dieđit telefuvdnávástideaddjái.**
- SMS:** Bestill frikort og kopi av skattekortet ditt på SMS.
- Sosiale medier:** Du finner "Skatten min" på Facebook og Twitter. Her svarer vi på generelle spørsmål og forteller om aktuelle skattesaker.
- Skatteoppkrever/kemner:** Har du spørsmål om blant annet utbetaling av skatt du har til gode, eller betaling av restskatt, må du ringe skatteoppkrever (kemner).
- Servicesenteret for utenlandske arbeidstakere (SUA):** Vi har kontorer i Stavanger, Oslo og Kirkenes.
- Skattekontoret:** 09:00 - 15:00. Enkelt skattekontor stenger dørene 14:30. Sjekk åpningstidene for ditt kontor. Det er ingen timebestilling hos skattekontorene. Du må møte opp og trekke kølapp. Finn post- og besøksadressen til ditt kontor i oversikten over kontorene nedenfor.
- Epost:** Send en generell henvendelse til Skatteetaten.

On the right side, there are two sections: 'Finn adressen direkte' with a form to enter a municipality or postal code, and 'Velg kontor' with a list of various tax offices and service centers.

Fig. 6.7a – Skjerm bilde av «Kontakt oss»-siden

Figur 1 viser kontaktsiden, og for å komme til Twittersiden kan man trykke på Twittersymbolet i teksten, i setningen om sosiale medier. Hvis man leser teksten ser man også at det står «Du finner «Skatten min» på [Facebook] og [Twitter].» Dette henviser til navnet



på kontoen, som er «Skatten min». Det er ingen tekstlenke til Twitter, kun et symbol, men det kom opp en liten infotekst i tooltip om man holdt musepekeren over symbolet. Setningen og mangelen på en informativlenke kan oppfattes som noe uklart, spesielt om man ikke er kjent med Twittersymbolet, og det var derfor ønskelig å se hva testdeltakerene gjorde når de ble bedt om å finne navnet på kontoen.

## Generelle observasjoner

Testdeltakerene rangerte denne oppgaven som lett, med en gjennomsnittscore på 1,27 var det den nest letteste oppgaven. Suksessraten var 100%, alle fant det korrekte navnet.

Hvis man ser på gazeplot fra testdeltaker 7 (Figur 6.7j) og testdeltaker 11 (Figur 6.7k) som begge svarte på oppgaven uten å trykke på Twitterikonet, ser vi at de hadde svært like øyebevegelser de første 3 sekundene av oppgaven.

## Visualiseringer

De følgende visualiseringene er fra de første 3 sekundene etter at oppgaven ble gitt.

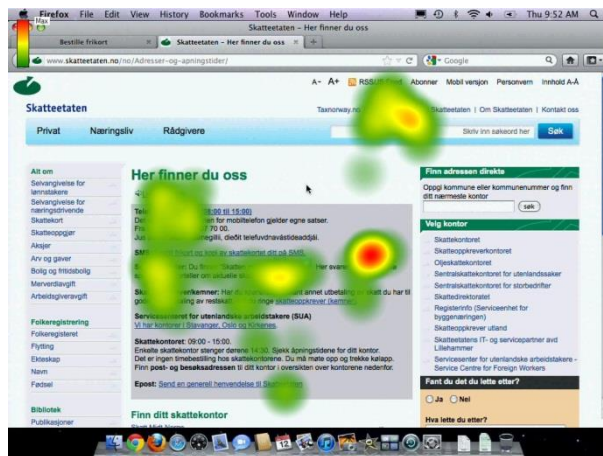


Fig. 6.7b – Heat map. Gruppe 1

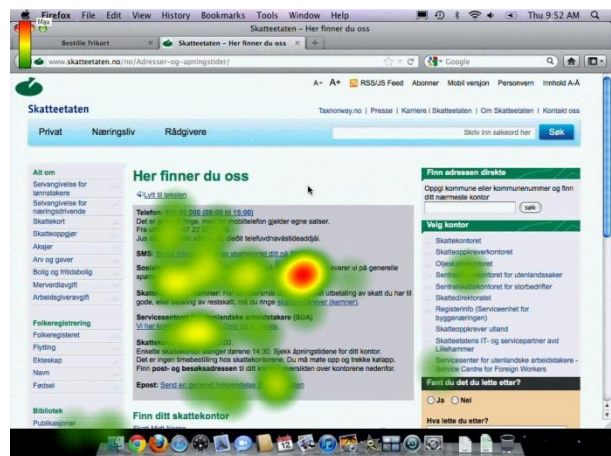


Fig. 6.7c – Heat map. Gruppe 3

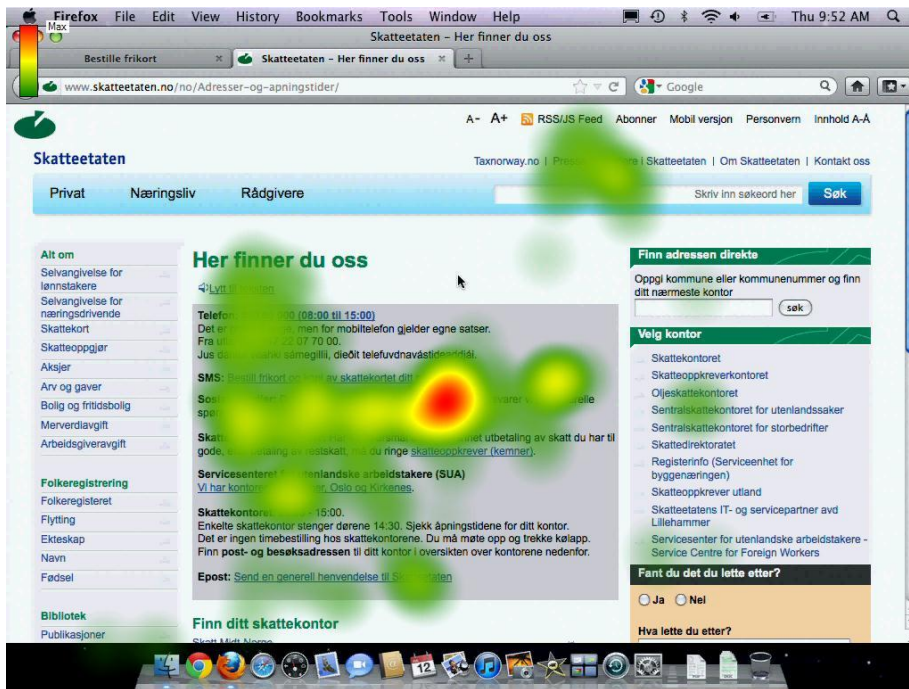


Fig. 6.7d – Heat map, alle testdeltakerene.

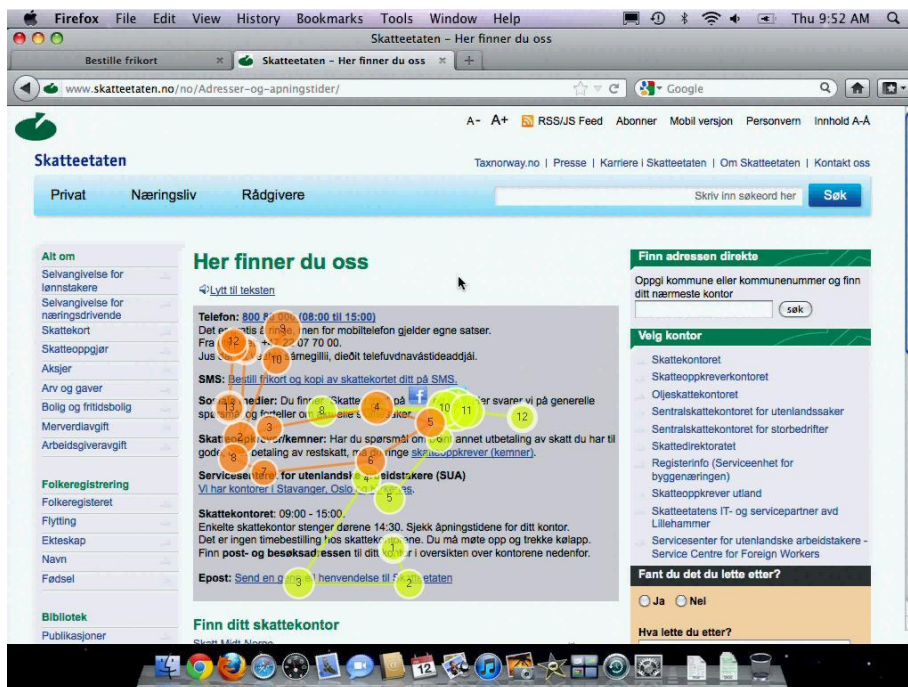


Fig. 6.7e – Gazeplot. Testdeltaker 7 (orange) fra gruppe 1, og testdeltaker 9 (gul) fra gruppe 3.

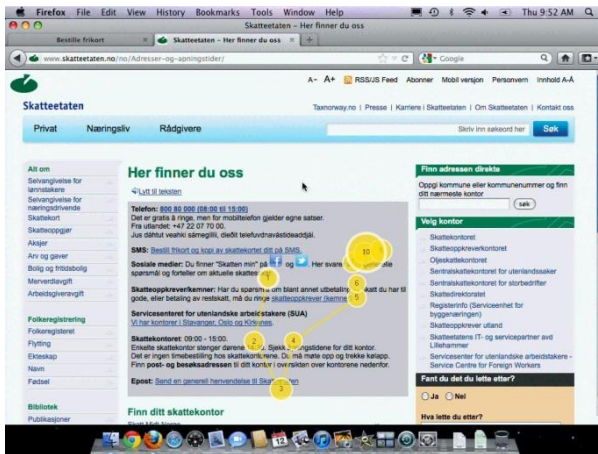


Fig. 6.7f – Gazeplot. Testdeltaker 1, fra gruppe 1.

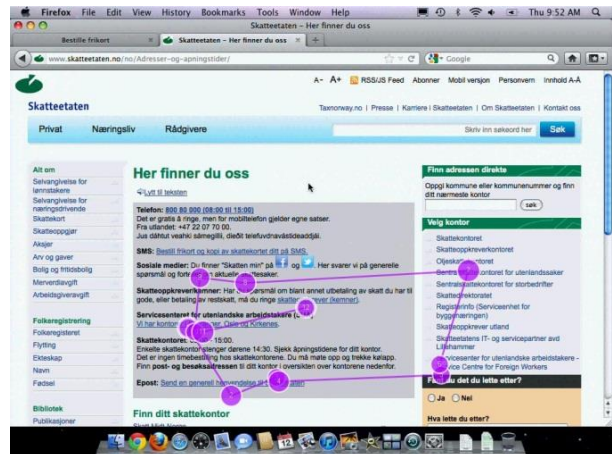


Fig. 6.7g – Gazeplot. Testdeltaker 3, fra gruppe 3.

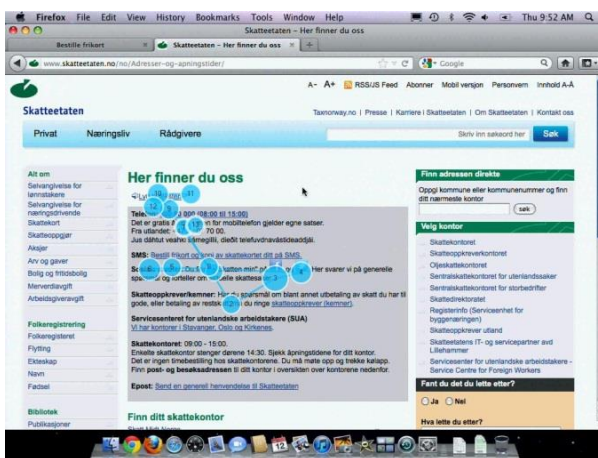


Fig. 6.7h – Gazeplot. Testdeltaker 12, fra gruppe 1.

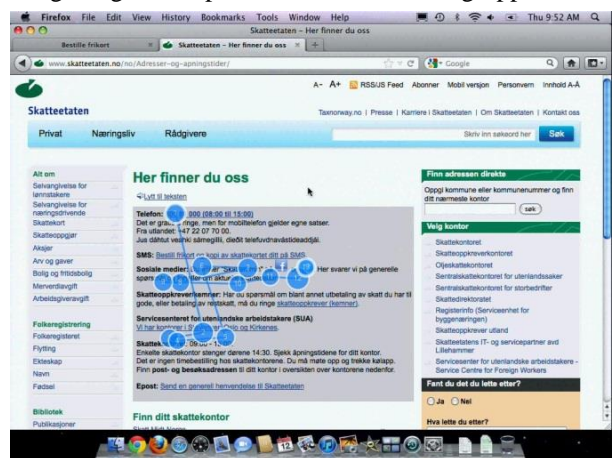


Fig. 6.7i – Gazeplot. Testdeltaker 4

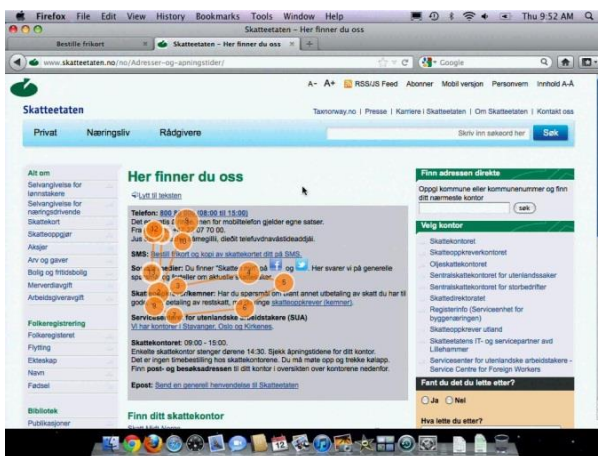


Fig. 6.7j – Gazeplot. Testdeltaker 7, fra gruppe 1.

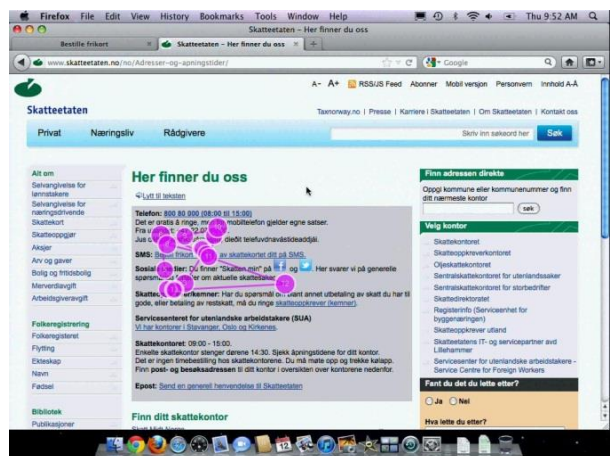


Fig. 6.7k – Gazeplot. Testdeltaker 11, fra gruppe 3.

## Statistikk

Følgende statistikk er fra de første 3 sekundene etter at oppgaven ble gitt.

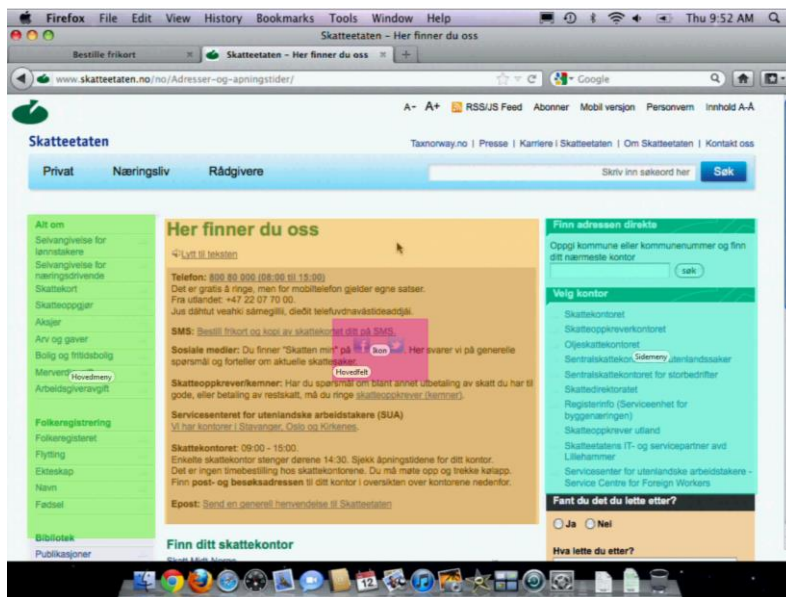


Fig. 6.71 – Oversikt over AOI.

## Percentage Fixated

$n=13$

AOI	Prosent
Hovedfelt	92 %
Hovedmeny	0 %
Ikon	77 %
Sidemeny	23 %

	Gruppe 1			Gruppe 3			T-test
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	p
<b>Fixation count</b>							
Hovedfelt	4	37	9,25	5	48	9,6	0,925
Ikon	4	4	1	5	8	1,6	0,461
Hele	4	46	11,50	5	51	10,2	0,648
<b>Total visit duration (i sek)</b>							
Hovedfelt	4	9,09	2,27	5	11,11	2,22	0,957
Ikon	4	1,13	0,28	5	2,01	0,40	0,682
<b>Time to first fixation (i sek)</b>							
Hovedfelt	4	2,91	0,73	5	2,31	0,46	0,751
Ikon	3	1,23	0,41	4	8,42	2,11	<b>0,009</b>

## Gruppe 1 vs. gruppe 3

Det ble sett nærmere på øyebevegelsene til gruppe 1, gruppe 2 og gruppe 3. En person fra hver gruppe løste oppgaven ut i fra teksten på siden, eller fra teksten i tooltip. Resten klikket på Twittersymbolet og fant svaret på twitter.com. To av de som trykket på Twitter-symbolet, en fra hver gruppe, fant svaret i URL istedet for på selve nettsiden. Det var altså ingen klar

forskjell på hvordan gruppene fant det riktige svaret. Det ble heller ikke observert noen utpreget forskjell i øyebevegelsene til de forskjellige gruppene.


Statistisk data viser en signifikant forskjell,  $p = 0,009$ , på tid til første fiksering på AOI «Ikon». Dette er området like rundt Twitterikonet. I motsetning til i oppgave 11 og 14, der gruppe 3 hadde signifikant kortere tid til første fiksering i relevante AOI, er det her gruppe 1 som hadde lavest gjennomsnittsverdi.

Heat map fra de 3 første sekundene viser at gruppe 3 sine øyebevegelser var noe mer fokuserte på midten av siden, rundt svaret.

## 6.2.6 Oppgave 21

*(Naviger til forsiden)*

*Anta at du har tabelltrekk med tabellnummer 7200 og månedslønnen din er 20000. Finn ut hvor mye du blir trukket hver måned.*



The screenshot shows the Skatteetaten website interface. At the top, there is a navigation bar with links for 'Privat', 'Næringsliv', and 'Rådgivere'. Below this is a search bar with the text 'Skriv inn søkeord her' and a 'Søk' button. The main content area is titled 'Tabelltrekk for skattekort 2012' and includes a sub-header 'Her finner du en elektronisk versjon av trekktabellene. Trekktabellene viser hvor mye skatt arbeidsgiver skal trekke dersom du har et skattekort som viser en tabell.' Below this is a link to 'Tabelltrekk for skattekort 2012' and a link to 'Trekktabeller som pdf'. On the left side, there is a sidebar with various menu items under 'Alt om', 'Folkeregistrering', and 'Bibliotek'. On the right side, there is a 'Se også' section with links to 'Emner', 'Tabeller og satser', and 'Grupper', and a 'Fant du det du lette etter?' section with radio buttons for 'Ja' and 'Nei' and a search input field.

Fig. 6.8a – Skjerm bilde av undersiden for tabelltrekk, med fremhevet lenke til beregningskjema for tabelltrekk for skattekort 2012 samt lenke til trekktabell som pdf.

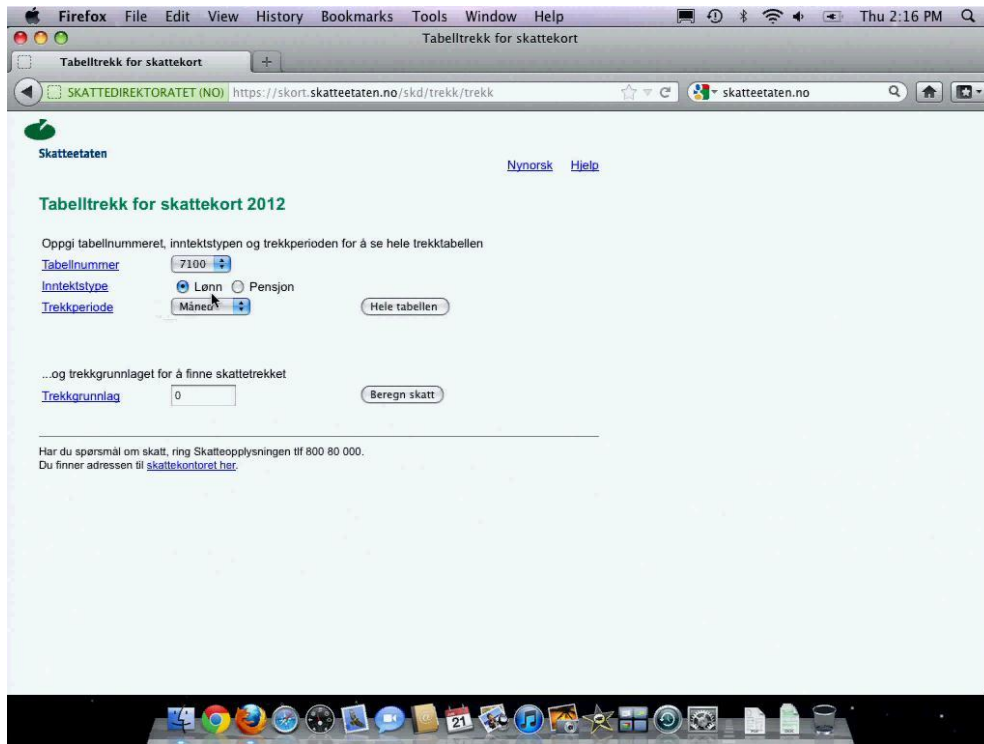


Fig. 6.8b – Skjerm bilde av siden med skjema for utregning av tabelltrekk.

Denne oppgaven startet på forsiden. Ønskelig var at testdeltakerene skulle finne frem til skjema for beregning av tabelltrekk, som vist i figur 6.8a, og deretter fylle ut skjemaet for å finne svaret på oppgaven. Raskeste måte å komme frem til skjemaet på fra forsiden var å trykke på Tabelltrekk-valget under Tjenester, som førte en til siden vist i figur 6.8a. Her trykket man på den fremhevede lenken i midten for å komme til skjemaet.

Det var ønskelig å se hvordan testdeltakerene forholdt seg til et slikt utfyllingsskjema, og om det var forskjell på hvordan de to brukergruppene så på skjemaet.

## Generelle observasjoner

Oppgavens vanskelighetsgrad ble rangert som 1,67 på en skala fra 1 til 5.

To av testdeltakerene kom aldri frem til skjemaet i figur 6.8b. Testdeltaker 5 scannet over feltet Tjenester, men selv med kort fiksering på riktig menyvalg valgte han å scrolle ned til meny nederst på siden og trykke seg inn på undersidene for Skattekort, der han videre navigerte rundt til hun fant trekktabellskjema som pdf. Han var den eneste testdeltakeren som ikke trykket på tabelltrekk-valget under «Tjenester». Testdeltaker 10 trykket på riktig lenke på forsiden, under «Tjenester», men istedet for å velge den store lenken «Tabelltrekk for

skattekort 2012» valgte han den lille lenken under, «Trekktabell som pdf». Han fikserte først på den store lenken, deretter den lille, deretter den store igjen og så den lille, så det var tydelig at han gjorde et bevisst valg, ikke at han overså den store lenken som ville ført ham til skjemaet. At en bruker fikserer gjentatte ganger på riktig lenke uten å velge den indikerer gjerne at de ikke forsto lenkens betydning (Bojko, 2006).

Av de 11 som kom frem til skjemaet var det kun en som brukte svært lang tid, 33 sekunder. De andre 10 varierte fra 7,2 til 15,6 sekunder fra de fikk oppgaven til de åpnet skjema for tabelltrekk. Etter at skjemaet var åpnet klarte 9 av 10 å fylle ut info riktig og trykke på «Beregn skatt» for å få opp riktig svar. De brukte fra 11 til 26 sekunder på dette. Den siste personen trykket på knappen «Hele tabellen» istedet for riktig knapp for å beregne svaret. Han fokuserte kun en gang på knappen for å beregne skatt, og det var den første fikseringen før utfylling av skjemaet. Man kan se gazeplot for de første sekundene i figur 6.8g. Han fylte heller ikke inn trekkgrunnlag, men scannet flere ganger over de andre valgene i skjemaet. Dette blikkmønsteret kan tyde på at brukeren syntes det var uklart hva han skulle gjøre (Ehmke og Wilson, 2007).

Det var kun små forskjeller i blikkmønsteret mellom forskjellige deltakere inne i skjemaet. Generelt scannet de skjemaet fra øverste til nederste valg, og lot musepekeren følge blikket fra punkt til punkt. Noen scannet innom alle elementene i skjemaet, valg og navn på feltene, mens andre scannet raskere og så ut til å hoppe over noe informasjon. Heat map viser fokus først og fremst på de øverste valgene, der nedtrekksmenyene er. Observasjon av opptak utover de 5 sekundene viste også at de fleste fokuserte raskt på begge knappene enten før de fylte ut trekkgrunnlag eller før de trykket «Beregn skatt».

Fikseringene ble mye lenger enn det som ellers ble sett idet deltakerene skulle velge riktig tall i dropdownmenyen. Dette kan komme av at de fokuserte på et punkt og scrollet menyen, fremfor å la blikket gli opp og ned valgene.

## **Visualiseringer**

De følgende visualiseringene er fra de første 5 sekunder etter at testdeltakerene klikket på lenken til skjemaet og det ble lastet på skjermen.

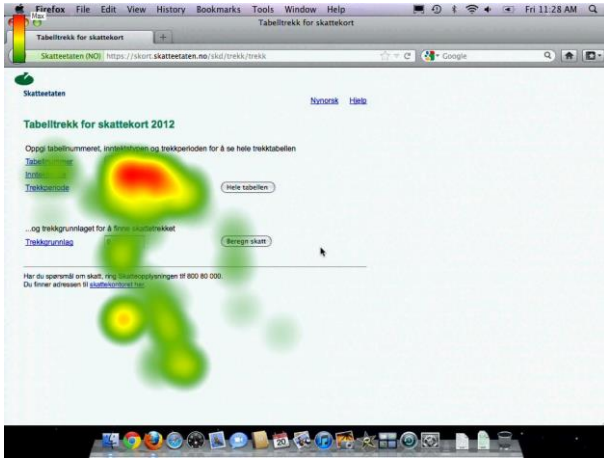


Fig. 6.8b – Heat map. Gruppe 1

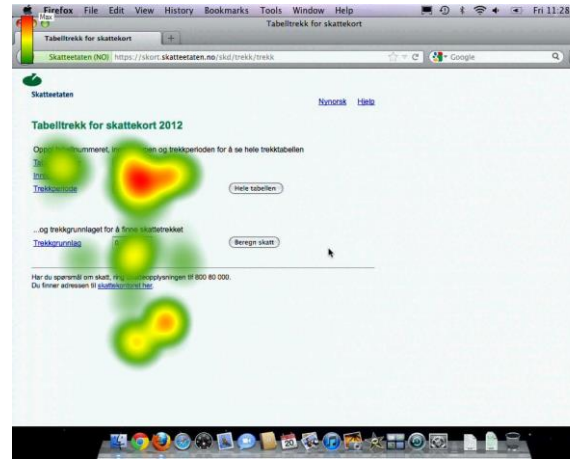


Fig. 6.8c – Heat map. Gruppe 3

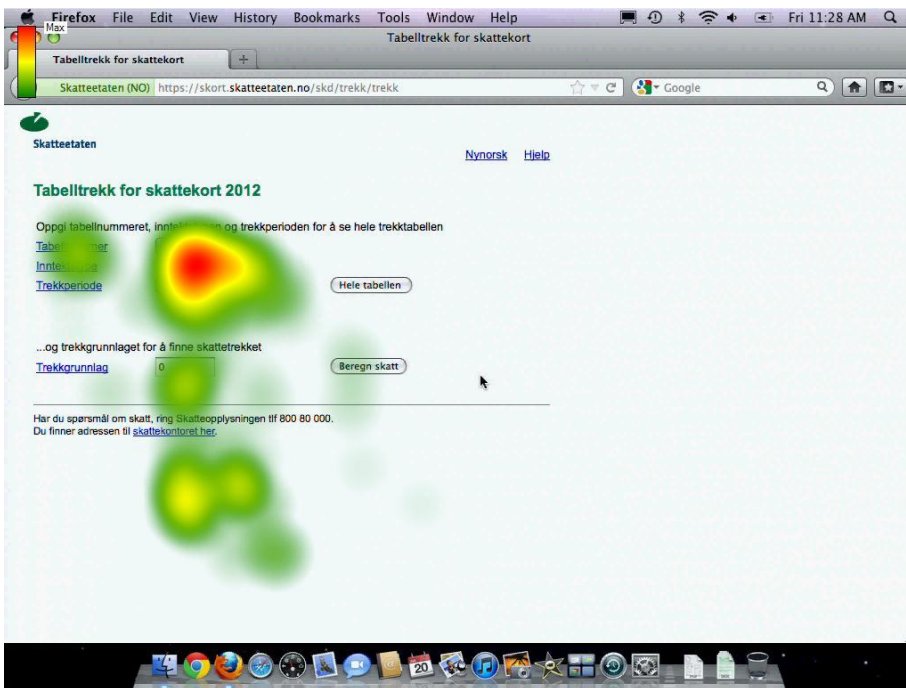


Fig. 6.8d – Heat map. Alle testdeltakerene.

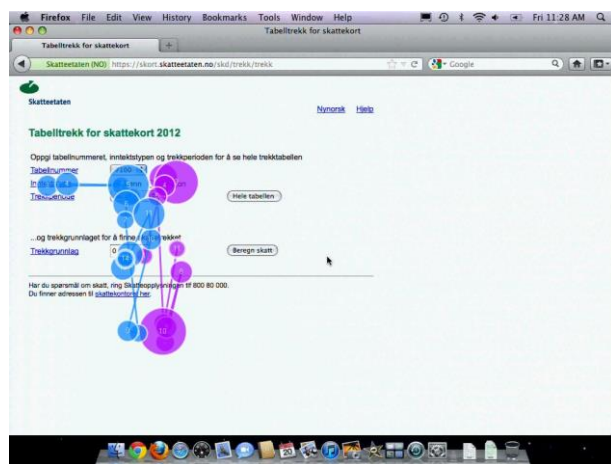


Fig. 6.8e – Gazeplot. Testdeltaker 12 (blå) fra gruppe 1, og testdeltaker 3 (lilla) fra gruppe 3.

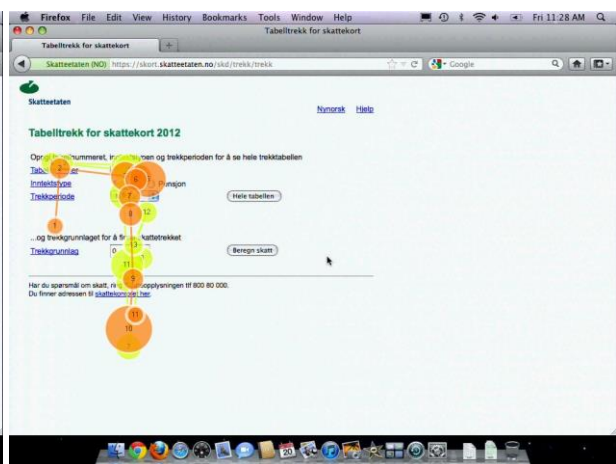


Fig. 6.8f – Gazeplot. Testdeltaker 7 (orange) fra gruppe 1 og testdeltaker 9 (gul) fra gruppe 2.



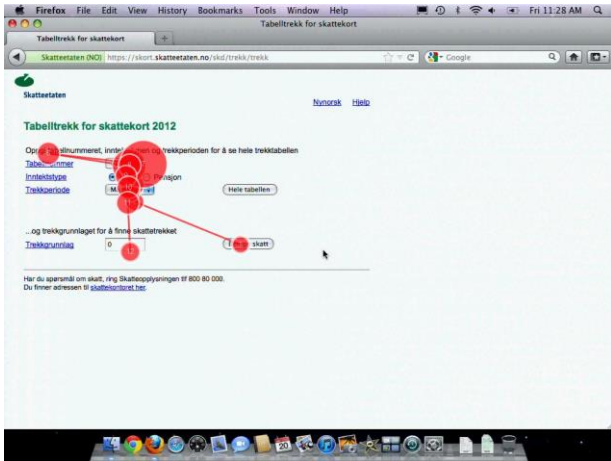


Fig. 6.8g – Gazeplot. Testdeltaker 2, som trykket på «Hele tabellen» istedet for korrekte «Beregn skatt».

## Statistikk

Følgende statistikk er fra de første 5 sekunder etter at testdeltakerene klikket på lenken til skjemaet og det ble lastet på skjermen.

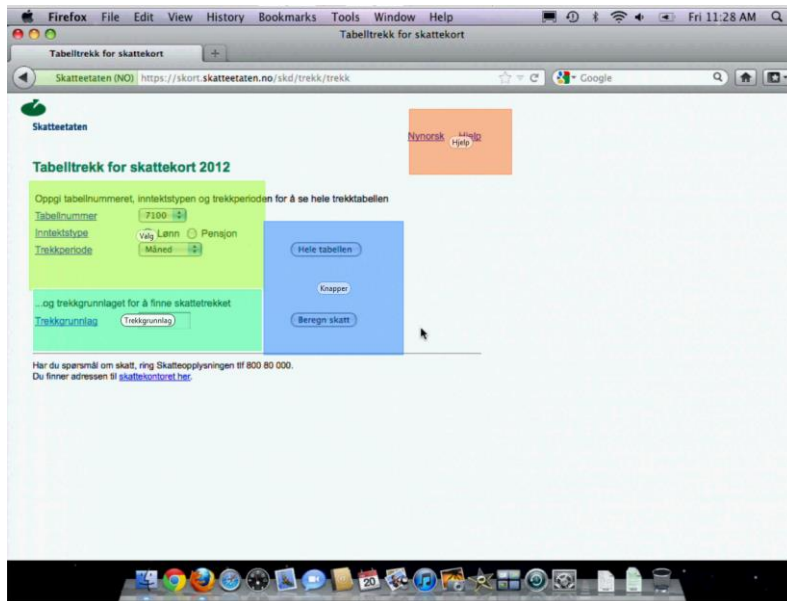


Fig. 6.8 – Oversikt over AOI.

## Percentage Fixated

$n=11$

AOI	Prosent
Hjelp	0 %
Knapper	18 %
Trekkgrunnlag	45 %
Valg	91 %

	Gruppe 1			Gruppe 3			T-test
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	p
<b>Fixation count</b>							
Trekkgrunnlag	4	4	1,00	4	10	2,50	0,300
Valg	4	30	7,50	4	29	7,25	0,908
Hele	4	54	13,50	4	49	12,25	0,368

<b>Total visit duration (i sek)</b>							
Trekkgrunnlag	4	1,24	0,31	4	2,63	0,66	0,477
Valg	4	11,66	2,92	4	11,80	2,95	0,953
<b>Time to first fixation (i sek)</b>							
Trekkgrunnlag	-	-	-	-	-	-	-
Valg	4	0,98	0,25		0,54	0,14	0,673

### Gruppe 1 vs. gruppe 3

Gruppe 3 brukte gjennomsnittlig noe kortere tid på utfylling av skjemaet enn det gruppe 1 gjorde.

Heat map for gruppe 1 og 3 ser svært like ut, som man kan se i fig. 6.8b og 6.8c. Gazeplot, som vist i figur 6.8e og 6.8f viser det samme, svært like øyebevegelser for testdeltakerene fra gruppe 1 og 3 iløpet av de første 5 sekundene i skjemaet. Tall fra statistikken viser heller ingen signifikant forskjell mellom de to gruppene, ingen av  $p$ -verdiene er  $\leq 0,05$ .

Som forklart var det små forskjeller på hvor nøye brukeren scannet skjemaet, og om de fokuserte på knappen for å vise hele tabell i tillegg til de viktige delene av skjemaet. Disse forskjellene var dog ikke mellom gruppe 1 og 3, det var ingen forskjeller mellom gruppene som utmerket seg når det gjaldt måten de scannet skjemaet på.

Av deltakerene som ikke kom frem til skjemaet i det hele tatt var han som valgte pdf-lenken fra gruppe 3, og han som ikke gikk via menyen «Tjenester» i det hele tatt fra gruppe 2. Den ene testdeltakeren som ikke valgte «Beregn skatt» var fra gruppe 2.

### 6.2.7 Oppgave 22

*Du er litt usikker på om du valgte rett alternativ på «Trekkperiode». Finn ut hva dette betyr uten å navigere ut av skjemaet for tabelltrekk-info.*

Denne oppgaven skulle løses i samme skjema som oppgave 21, som vist i figur 6. . Riktig svar her kunne man finne ved å trykke på ordet «Trekkperiode». Det var også mulig å gå via «Hjelp» øverst i høyre hjørne. Det var ønskelig å se om testdeltakerene forsto at navnet på feltene også var hjelpetekster man kunne trykke på for mer informasjon, og om dette var mer intuitivt for gruppe 3.

De to testdeltakerene som ikke åpnet riktig skjema i oppgave 21 fikk ikke denne oppgaven.

## Generelle observasjoner

Oppgavens vanskelighetsgrad ble rangert som 1,62 på en skala fra 1 til 5. Hvor lang tid testdeltakerene brukte på å finne svaret varierte en del, fra 4 til 26 sekunder.

To av testdeltakerene fikserte på menyvalget «Hjelp» øverst i venstre hjørne mens de søkte etter svaret. En av disse valgte å trykke på lenken og finne svaret der. En annen person svarte uten å åpne noen lenker, bare ved å se på valgene i nedtrekksmenyen og resonnere seg frem muntlig. Resten av testdeltakerene trykket på lenken «Trekperiode». Generelt fokuserte alle på dette området største del av tiden, og hadde gjentatte fikseringer på og raske sakkader mellom navneteksten, som også var lenken, og nedtrekksmenyen. Man ser disse fikseringene og sakkadene mellom de to punktene ganske godt illustrert i figur 6.9f og 6.9g.

## Visualiseringer

De følgende visualiseringene er fra de første 6 sekundene etter at oppgaven ble gitt.

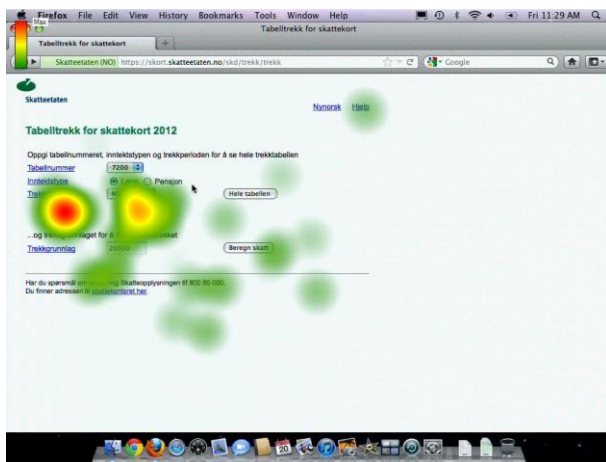


Fig. 6.9a – Heat map. Gruppe 1

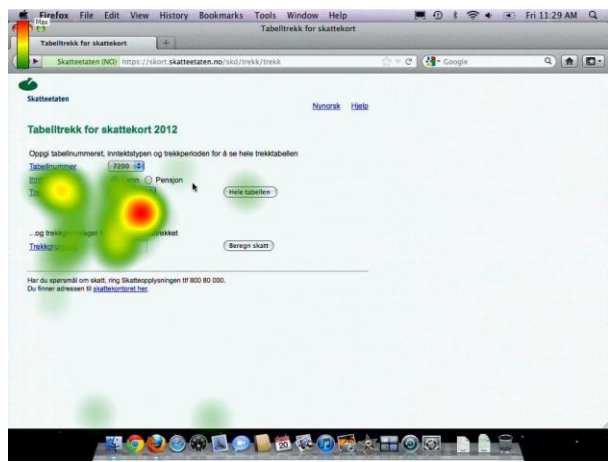


Fig. 6.9b – Heat map. Gruppe 3

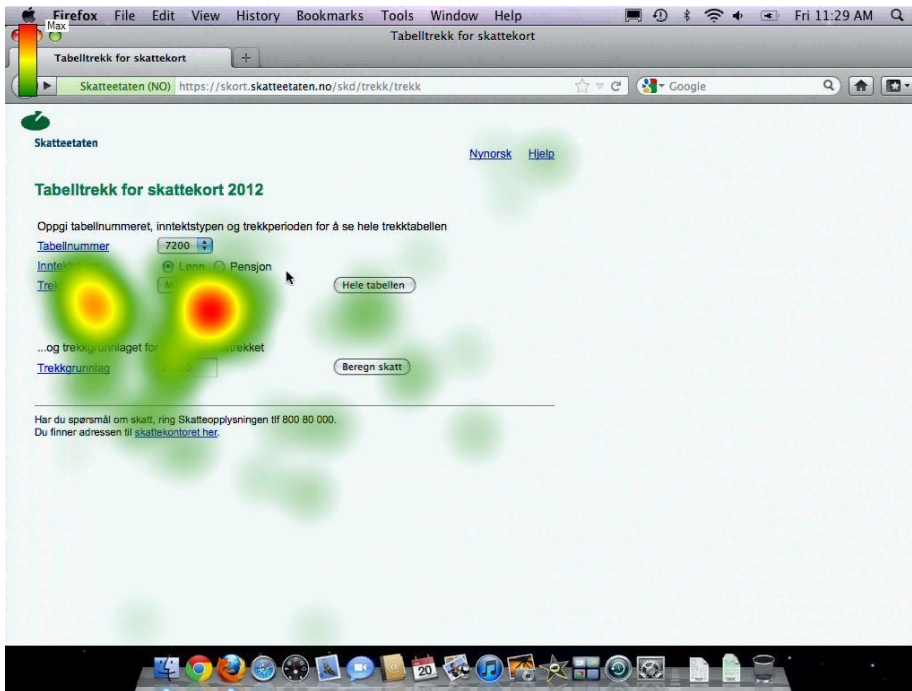


Fig. 6.9c – Heat map. Alle testdeltakerene.

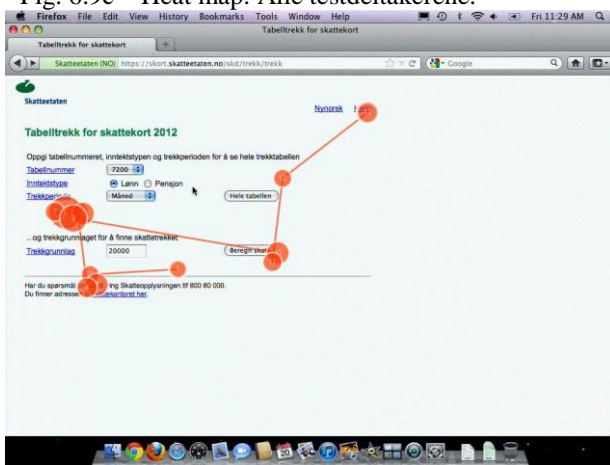


Fig. 6.9d – Gaze plot. Testdeltaker 6, fra gruppe 1.

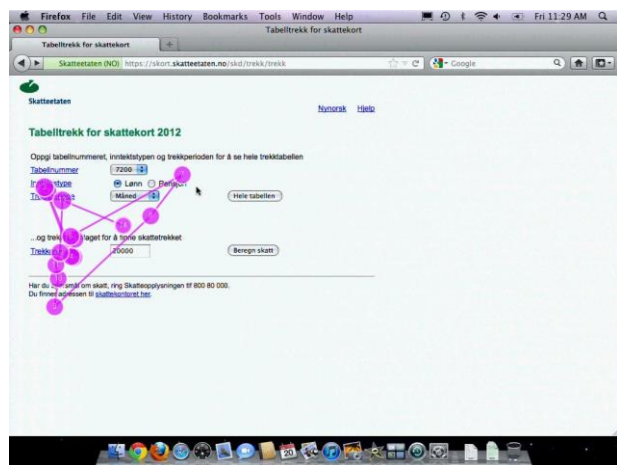


Fig. 6.9e – Gaze plot. Testdeltaker 11, fra gruppe 3.

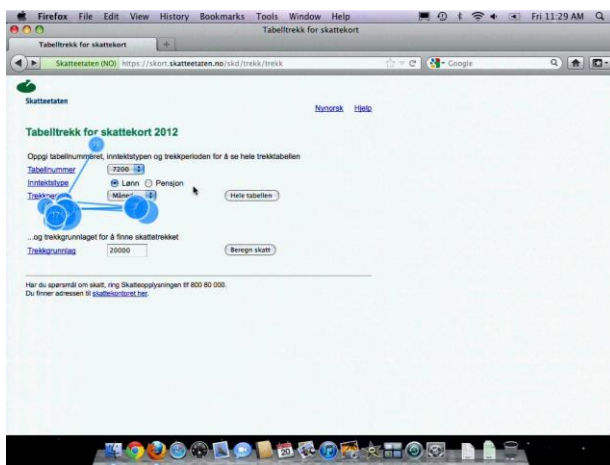


Fig. 6.9f – Gaze plot. Testdeltaker 12, fra gruppe 1.

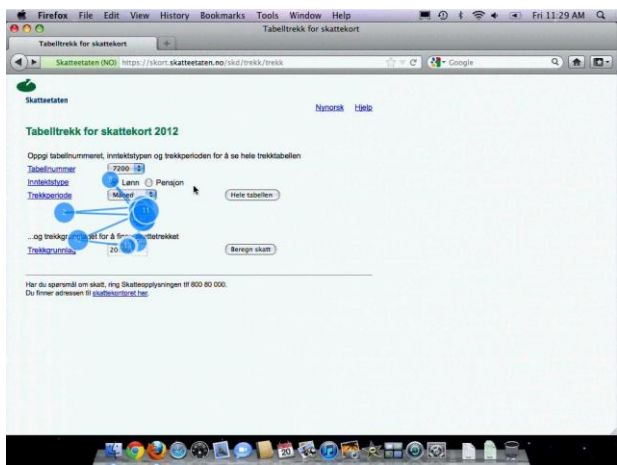


Fig. 6.9g – Gaze plot. Testdeltaker 4, fra gruppe 3.

## Statistikk

Følgende statistikk er fra de første 6 sekundene etter at oppgaven ble gitt.

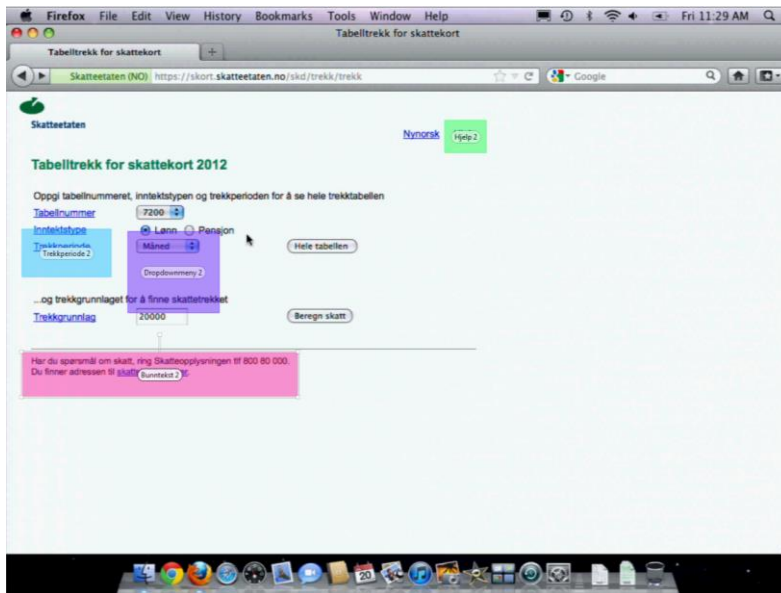


Fig. 6.9h – Oversikt over AOI.

### Percentage Fixated

$n=11$

AOI	Prosent
Bunntekst	36 %
Dropdownmeny	91 %
Hjelp	9 %
Trekkeperiode	82 %

	Gruppe 1			Gruppe 3			T-test
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	p
<b>Fixation count</b>							
Dropdownmeny	4	16	4,00	4	23	5,75	0,498
Trekkeperiode	4	17	4,25	4	11	2,75	0,628
Hele	4	55	13,75	4	50	12,5	0,562
<b>Total visit duration (i sek)</b>							
Dropdownmeny	4	5,70	1,43	4	6,00	1,50	0,935
Trekkeperiode	4	4,33	1,08	4	3,06	0,77	0,660
<b>Time to first fixation (i sek)</b>							
Dropdownmeny	3	2,80	0,93	4	4,72	1,18	0,829
Trekkeperiode	3	4,57	1,52	3	0,77	0,26	0,298

### Gruppe 1 vs. gruppe 3

Heat map fra gruppe 1 og gruppe 3 ser svært like ut, gruppe 3 sitt fokus er noe mer sentrert på midten, rundt lenken for trekkeperiode. Statistikken viser heller ingen tydelig forskjell mellom gruppene.

Det var ingen forskjell på hvor lang tid gruppene brukte på å løse oppgaven, halvparten fra gruppe 1 brukte lang tid og halvparten kort, og det samme så man i gruppe 3.

Heller ikke gjennomgang av opptak viste noen store forskjeller mellom gruppene. Den ene forskjellen som ble observert var at to personer fra gruppe 1 scannet innom og fikserte på menyvalget «Hjelp» øverst i hjørnet, det gjorde ingen fra gruppe 3. Ellers så man liknende blikkmønster i begge gruppene, der deltakeren fikserte på «Trekkgrunnlag» og tilhørende nedtrekksmeny vekselvis før det ble trykket på lenken. Blikkmønsterene her ga ikke inntrykk av at gruppe 3 forsto hensikten med at navnet på valget også var en hjelpelenke raskere enn det gruppe 1 gjorde.

## 6.2.8 Oppgave 26

*Finn ut om det er en engelsk versjon av sidene om utleie og salg av bolig.*



Fig. 6.10a – Forsiden til undersidene om utleie av bolig, med informasjon iforhold til skatt av utleieinntekter. Det var denne siden testdeltakerene sto på når de fikk oppgave 26.

Denne oppgaven kom etter to andre oppgaver om salg og utleie av bolig. Oppgave 24, «*Finn ut hvor lenge du må bo i egen bolig før du kan selge den uten å betale skatt.*», og oppgave 25, «*Finn ut hvor mye du kan tjene på utleie av hele boligen din uten å måtte skatte av det.*».

Det er vanlig at man finner et engelsk flagg eller ordet “English” et sted øverst i høyre hjørnet på sider som tilbyr oversettelse / engelsk versjon. Skatteetaten.no har ikke dette, men i den lille menyen øverst til høyre har man «Taxnorway.no». Dette er profilert som «Skatteetatens internasjonale sider» og inneholder en del informasjon om skattespørsmål på norsk, engelsk og polsk. Det finnes ingen engelsk versjon av sidene om utleie og salg av bolig. Vil de med bakgrunn fra interaksjonsdesign i gruppe 3 i større grad se til høyre hjørne først for å se etter engelsk? Vil testdeltakerene se lenken til Taxnorway.no og anse den som relevant eller korrekt svar?

### **Generelle observasjoner**

Dette var en oppgave testdeltakerene anså som vanskelig, noe som ikke er overraskende da den ikke hadde et klart riktig svar. Den fikk høyest score, 3,07 på en skala fra 1 til 5.

Ehmke og Wilson (2007) kom frem til i sin studie at når en bruker ser på en spesiell del av siden og forventer å finne en detalj der, som mangler, vil man se et blikkmønster som består av mange korte fikseringer i området der informasjonen er forventet. Dette mønsteret kunne man kjenne igjen hos alle testdeltakerene som fokuserte på øverste høyre hjørne av siden der de lette etter forventet ikon eller lenke med mange korte fikseringer.

Unntaket var at de fleste fokuserte noe lengere på lenken til Taxnorway.no. De fleste fokuserte på Taxnorway-lenken flere ganger før de trykket på den første gang, og flere lot musepekeren hvile på lenken mens de scannet rundt på siden, før de fokuserte tilbake på lenken og så trykket på den. Bare en av deltakerene fokuserte ikke på Taxnorway-lenken i det hele tatt.

Ehmke og Wilson (2007) fant videre at en overlesset og ineffektiv side ville føre til mange korte fikseringer på enkelte områder fulgt av lange sakkader og avsporinger til enkelte elementer. Dette var et blikkmønster man også kunne kjenne igjen hos de fleste testdeltakerene idet de scannet over siden, som inneholder mye tekst samt flere menyer og mange lenker, for å forsøke å finne ut hvordan de kunne endre språket. To eksempler på dette kan sees i gazeplot i figur 6.10e og 6.10f. Dette kan tyde på problemer med å forstå hvordan man skal interagere med siden for å løse oppgaven man har fått.

## Visualiseringer

De følgende visualiseringene er fra de første 6 sekundene etter at oppgaven ble gitt.

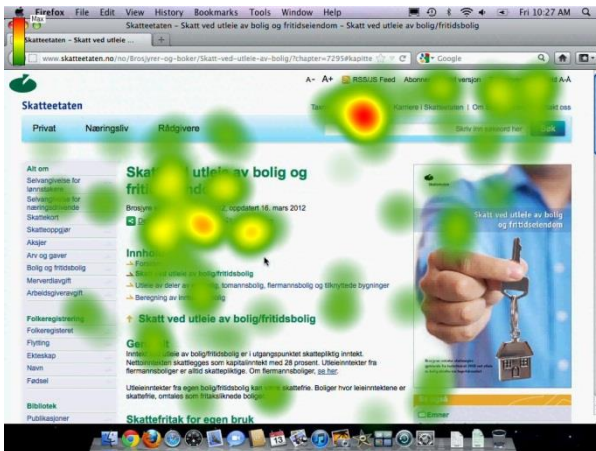


Fig. 6.10b – Heat map. Gruppe 1



Fig. 6.10c – Heat map. Gruppe 3

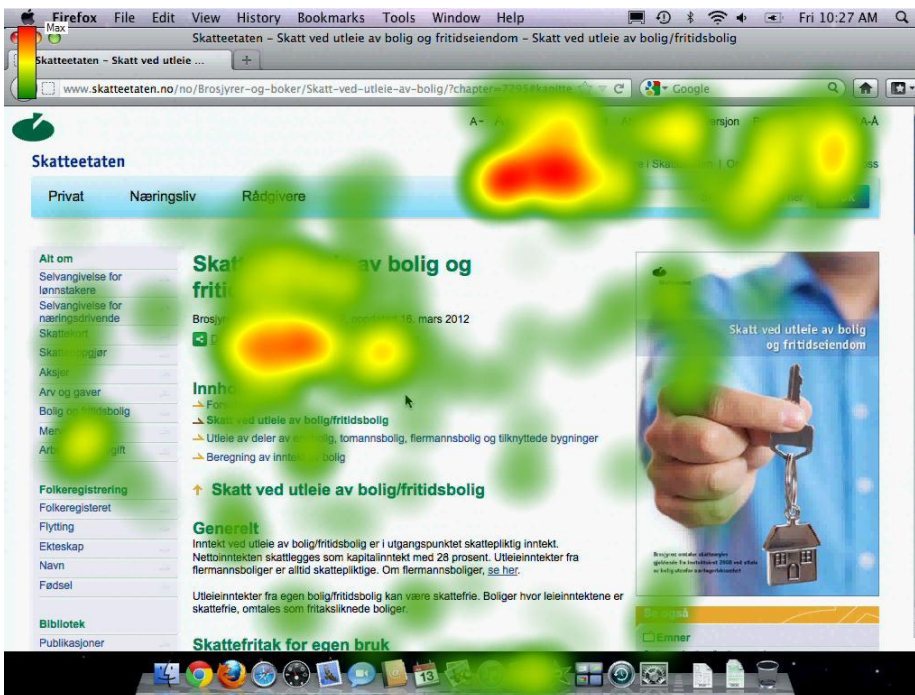


Fig. 6.10d – Heat map. Alle testdeltakerene. Her ser man at testdeltakerene fokuserte over store deler av siden, med noe ekstra vekt på øverste høyre hjørne.





Fig. 6.10e – Gazeplot. Testdeltaker 7, fra gruppe 1.

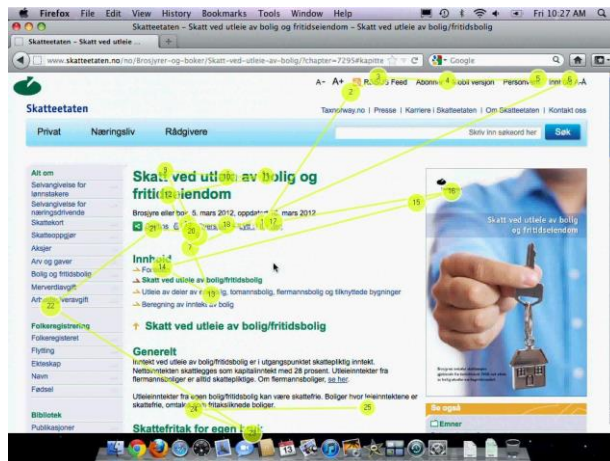


Fig. 6.10f – Gazeplot. Testdeltaker 9, fra gruppe 3.

## Statistikk

Følgende statistikk er fra de første 6 sekundene etter at oppgaven ble gitt.



Fig. – Oversikt over AOI. Hovedmeny er det grønne feltet til venstre. Høyre toppmeny er det gule feltet øverst til høyre, som er dekket av rosa og lilla.

## Percentage Fixated

$$n=11$$

AOI	Prosent
Hovedmeny	55 %
Høyre toppmeny	91 %
Høyre toppmeny1	64 %
Høyre toppmeny2	82 %
Taxnorway	55 %
Toppmeny	27 %

	Gruppe 1			Gruppe 3			T-test p
	n	Sum	Gj.snitt	n	Sum	Gj.snitt	
<b>Fixation count</b>							
Hovedmeny	4	3	0,75	4	7	1,75	0,518
Høyre toppmeny	4	27	6,75	4	31	7,75	0,769
Hele	4	94	23	4	72	18	0,090

<b>Total visit duration (i sek)</b>							
Hovedmeny	4	0,72	0,18	4	2,00	0,50	0,480
Høyre toppmeny	4	6,16	1,54	4	8,90	2,23	0,563
<b>Time to first fixation (i sek)</b>							
Hovedmeny	3	6,10	2,03	2	5,19	2,60	0,817
Høyre toppmeny	4	2,90	0,73	3	3,81	1,27	0,488

### Gruppe 1 vs. gruppe 3

Sammenlikning av heat maps gir ingen indikasjon på at det var forskjell på hvor de to gruppene fokuserte mest i løpet av de første 6 sekundene av oppgaven. Statistikken gir heller ingen klar indikasjon på forskjell. Gruppe 3 har noen flere fikseringer og noe mer tid brukt innenfor høyre toppmeny de første 6 sekundene av oppgaven, selv om de hadde færre fikseringer totalt enn det gruppe 1 hadde. Dette tilsier at de første 6 sekundene så gruppe 3 litt lenger på høyre toppmeny enn gruppe 1, men ikke signifikant lenger.

Gjennomgang av opptak viser heller ingen utpreget forskjell mellom gruppene, og støtter ikke teorien om at de erfarne i gruppe 3 ville fokusere på høyre toppmeny i større grad eller tidligere enn gruppe 1. Tvert imot var det flere personer fra gruppe 3 som scannet hovedfeltet noe først før de scannet høyre toppmeny, enn i gruppe 1, og den ene personen som aldri fokuserte på lenken til Taxnorway var fra gruppe 3. Testdeltakere fra begge gruppene søkte i større grad etter svaret øverst til høyre enn på resten av siden.

## 6.3 Diskusjon rundt resultatene

### 6.3.1 Generelle observasjoner

Det vil ikke være hensiktsmessig å gjøre en direkte sammenlikning av visualiseringer og statistisk data mellom oppgavene, av flere grunner. I sitt studie av visuell stimuli viste Lytkins et.al (2006) sine testdeltakere en rekke bilder der bildene var forskjellige men besto av de samme hovedelementene: bakgrunn og menneske. Alle bildene ble vist like lenge. Da kunne man sammenlikne antall fikseringer på AOI «kropp» mellom bilde 1 og bilde 2 om ønskelig. I denne studien ble oppgavene utført på forskjellige skjermbilder, og forskjellige AOI ble definert. Segmentene hadde også forskjellige lengde, fra 3 til 6 sekunder. Alle verdier kan

derfor ikke sammenliknes direkte fra oppgave til oppgave. Det er likevel mulig å se på store linjer mellom oppgavene, noe som gjentar seg eller er svært forskjellig. Dette blir basert på den manuelle gjennomgang av opptak, og i tillegg de generelle tendensene i de visuelle dataene. For sammenlikning av statistiske verdier kan man se på  $p$  i de forskjellige oppgavene.

Oppgavene hadde noe forskjellig form og mål, men som en generalisering kan man si at de alle handlet om å finne spesifikk informasjon på nettsidene. Oppgave 21 skilte seg noe ut da deltakerene skulle fylle ut et skjema på skjermen for å komme frem til svaret.

En ting som er verdt å nevne når det gjelder generelle tendenser er at man i flere av oppgavene ser at blikkmønsteret de første sekundene er veldig like fra deltaker til deltaker. Det er spesielt tydelige likheter i oppgave 19 (se figur 6.7e – 6.7k) og oppgave 21 (se figur 6.8e – 6.8f). Hva betyr det når så mange av testdeltakerene har såpass like blikkmønster? Poole and Ball (2005) referer til Goldberg & Wichansky (2003) som hevder det er store individuelle forskjeller mellom øyebevegelsene til deltakere på like oppgaver. Om dette er sant kan man argumentere videre at det har betydning når øyebevegelsene likevel viser seg å være ganske like på en oppgave. Det kan være fordi brukervennligheten er såpass dårlig at alle deltakerene opplever de samme problemene. Ehmke og Wilson (2007) viste i sin studie til spesifikke blikkmønstre som oppsto ved spesielle bruksproblemer. Det kan også være helt motsatt, designen er såpass god og forståelig at alle brukere ser der det er meningen at man skal se for å finne det man leter etter. I disse to oppgavene har testdeltakerene rapportert at de syntes oppgaven var lett, noe som fører til at det siste alternativet er mest sannsynlig. Øyebevegelsene likner hverandre fordi det er et forståelig design på stimulien de ser på. Videre undersøkelser viser at flere av oppgavene ble løst på tilsvarende måte hos flertallet av testdeltakerene – de navigerte seg til svaret på samme måte og via samme menyvalg og lenker. Dette igjen støtter det at navigasjonen er forståelig, men det kan også være et resultat av en testgruppe som er vant med daglig bruk av datamaskiner og det å finne informasjon på nettsider. I dette forsøket kan det også være en påvirkende faktor at nettsiden ikke er fullstendig ny og ukjent, men noe de fleste hadde besøkt og benyttet som privatpersoner tidligere.

En annen tendens som ble lagt merke til var hvordan testdeltakerene scannet menyer for å finne et oppgitt menyvalg, eller for å finne et som vil føre dem til ønsket informasjon de ble bedt om å finne. Man ser gaze plots for en slik type oppgave i figur 6.4c – 6.4f. Disse er

basert på opptak fra oppgave 4, men denne observasjonen ble også gjort på flere av de andre oppgavne ved gjennomgang av opptak. Tendensen var at testdeltakerene for det aller meste ikke scannet menyen fra topp til bunn, men startet på midten og scannet opp og ned ganske raskt. Noen ganger scannet de også nedenfra og opp, spesielt om de først hadde scrollet ned siden for å lete etter informasjon i hovedteksten, men ikke funnet det. Dette ble notert som interessant, og en oppførsel man gjerne kunne studert videre.

### 6.3.2 Gruppe 1 versus gruppe 3

#### Oppsummering av observerte forskjeller

Statistikk viste kun signifikante forskjeller i verdien *time to first fixation* – tid til første fiksering. Disse forskjellene ble kun sett på oppgave 11, 14 og 19. Med signifikante forskjeller menes her at  $p \leq 0,05$ .

Oppgave 11: Tid til første fiksering på AOI «Telefonnummer» –  $p = 0,001$ . Gruppe 3 hadde lavest gjennomsnittsverdi.

Oppgave 14: Tid til første fiksering på AOI «Snarvei» –  $p = 0,012$ . Gruppe 3 hadde lavest gjennomsnittsverdi. Tid til første fiksering på AOI «Viktige datoer» –  $p = 0,004$ . Gruppe 3 hadde lavest gjennomsnittsverdi.

Oppgave 19: Tid til første fiksering på AOI «Ikon» –  $p = 0,009$ . Gruppe 1 hadde lavest gjennomsnittsverdi.

Som poengtert før er det vanskelig å gjøre en direkte sammenlikning av forskjellige skjermbilder da de inneholder forskjellige AOI, men generelt kan man si at disse områdene med signifikante forskjeller var relevante for hver oppgave. I 3 av 4 tilfeller der det ble målt en forskjell mellom gruppene var det gruppe 3 som hadde lavest gjennomsnittsverdi. I det fjerde tilfellet var det gruppe 1 som hadde lavest gjennomsnittsverdi.

På oppgave 26 ble det observert at gruppe 1 fokuserte på det svært relevante AOI «Høyre toppmeny» noe raskere enn gruppe 3.

Tid til første fiksering ser altså ikke ut til å være konsekvent forskjellig.

Når det gjelder områder der kun den ene gruppen fikserte så man dette ganske tydelig i oppgave 4 og oppgave 22. I oppgave 4 fikserte tre av fem fra gruppe 3 på sidemenyen med

relevant tilleggsinformasjon (kalt «Se også»). I oppgave 22 derimot var det to av fire fra gruppe 1 som avvek fra de andre testdeltakerene, ved å fiksere på lenken «Hjelp» øverst i høyre hjørne.

Hva kan man forstå ut fra observasjonene om tid til første fiksering samt områder der kun en av gruppene fikserte? Fordi det ikke er konsekvent hvilken gruppe som hadde lavest verdi eller fikserte på et ekstra område, er det vanskelig å trekke konklusjoner. Man kan spekulere i at områdene der gruppe 3 fikserte raskere, eller der kun deltakere fra gruppe 3 fikserte, er områder som skiller seg ut og er ment å trekke oppmerksomhet. Dette er for eksempel menyen «Se også» eller grafisk snarvei til informasjon om skatteoppgjør. Kanskje fungerer denne type fremheving av elementer bedre for brukere som er klar over virkemiddelet?

På den ene oppgaven som inneholdt utfylling av skjema, oppgave 21, gjennomførte gruppe 3 utfyllingen gjennomsnittlig noe raskere enn gruppe 1. Ingen av de andre oppgavene viste betydelig forskjell i tidsbruk mellom gruppene, og det var heller ingen av oppgavene der gruppe 1 var klart raskere enn gruppe 3.

Når det gjaldt suksessrate var det kun oppgave 11 som skilte seg ut. Her løste alle testdeltakerene fra gruppe 3 oppgaven korrekt, mens kun en av fire fra gruppe 1 gjorde det samme. I de andre oppgavene som ble analysert var stort sett alle testdeltakerene suksessfulle.

Basert på disse to observasjonene om tidsbruk og suksessrate sammen med tidligere observasjoner av tid til første fiksering ser det ut til at gruppe 3 generelt var noe flinkere når det gjaldt å løse oppgavene, men det er kun en subtil forskjell. Man kan ikke trekke en definitiv konklusjon om at det var forskjell på gruppene basert på dette grunnlaget.

### **Generell sammenlikning**

En generell sammenlikning av gruppe 1 og 3 kan man gjøre ved å se på heat maps fra alle oppgavene og se om det er noen forskjeller som gjentar seg. I det store og hele er heat maps fra gruppe 1 og gruppe 3 nokså like, men man kan se en tendens til at gruppe 3 er noe mer fokusert. Dette ser man ved at gruppe 1 har noe større områder av grønn farge, som tilsvarer flere korte og spredte fikseringer. Dette kan indikere at gruppe 3 var noe mer tilbøyelig til å fokusere på korrekt område uten overflødige fikseringer i andre områder av skjermbildet. Det er ikke en såpass tydelig differanse mellom heat maps at man kan konkludere med at dette er en definitiv forskjell. Her er antall testdeltakere en faktor, da heat map for kun én gruppe er

basert på 4-5 brukere. Ifølge Nielsen og Pernice (2010) bør man ha et større datagrunnlag for heat maps.

I forkant av opptak hadde man en antagelse om at gruppen med bakgrunn fra interaksjonsdesign ville løse oppgavene noe kjappere og mer korrekt, men dette viste seg å ikke være tilfellet. Gruppe 3 brukte ikke konsekvent kortere tid enn gruppe 1 og hadde ikke en høyere suksessrate på oppgavene. Dette er basert på oppgavene som er gjennomgått i kapittel 6.2. Når det gjelder total tid på hele oppgavesettet er ikke disse tidene inkludert i resultatene, fordi det er en tid påvirket av mange faktorer, blant annet testdeltakers formidling av oppgaver, uventede pauser osv. Men generelt kan man si at det ikke var noen tydelig lavere gjennomsnittstid for gruppe 3 enn for gruppe 1.

Ved manuell gjennomgang av opptakene ble det sett på det generelle blikkmønsteret til brukeren på hver spesifikke oppgave. Det ble sett på hvor brukeren fikserte, om det var mange og lange fikseringer, om det var lange eller korte sakkader, om fikseringene var sentrert til ett området eller tilfeldig spredt over hele skjermbildet, samt rekkefølgen av elementer det ble fiksert på. Denne gjennomgangen førte til noen interessante generelle observasjoner, som beskrevet i forrige underkapittel. Ingen klare forskjeller i blikkmønsteret mellom de to gruppene ble observert. For eksempel kan man se på oppgave 4 og oppgave 26. Oppgave 4 ble ansett som ganske lett mens oppgave 26 ble ansett som vanskelig. Her så man forskjeller i blikkmønsteret mellom oppgavene. På oppgave 26 så man klart flere korte fikseringer over større deler av siden, samt lange sakkader og lange hopp fra ett element til et annet. Se for eksempel figur 6.4d kontra figur 6.10f. Disse to gazeplot er basert på forskjellige tidsperioder, 3 sekunder og 6 sekunder, så figur. 6.10f inneholder flere fikseringer. Likevel ser man tendensen til at fikseringene i figur 6.4d er mye mer konsentrert. Denne tendensen var lik for gruppe 1 og gruppe 3, noe som kan antyde at oppgavens natur var mer relevant for blikkmønster enn brukerens bakgrunn.

### **6.3.3 Faktorer som kan påvirke resultatene**

Det vil alltid være faktorer til stede som kan påvirke resultatene. I kapittel 4.4 er det beskrevet de tiltak som ble gjort for å begrense det som kalles Systematiske feil (Lazar et. al, 2010). Men andre ting kan også ha hatt innvirkning.

Den faktoren som nok har hatt størst innvirkning på resultatene i denne studien er det menneskelige aspektet, noe som alltid vil være en faktor når manuell observasjon er en del av datasettet som legges til grunn for en analyse. I forsøket her er også grunnlaget for visualiseringer, statistikk, og talldata som tidsbruk blitt gjort manuelt ved at testleder har laget segmenter og scener i programvaren Tobii Studio. Dette skaper et ekstra ledd i bearbeiding av data, som igjen øker risiko for feil. Dette kommer igjen av stimuliens natur i dette forsøket, da det er et sammenhengende opptak fra interaksjon med nettsiden. Forsøk med stillbilder som stimuli, for eksempel studien av kunst og bilder utført av Vogt og Magnussen (2007), har ikke dette problemet.

I ettertid ser man at det kunne vært hensiktsmessig å inkludere flere korte oppgaver i forsøket, der deltakerene ikke hadde måttet scrolle eller bytte skjermbilde, for å lette analyse og uthenting av data.

En annen faktor man bør se på som kan ha påvirket resultatene, er utvalget av testdeltakere. De ble gruppert etter problemformuleringens krav basert på egenopplyst bakgrunn fra studier eller jobb med interaksjonsdesign, webdesign, eller brukervennlighet på nett. Det ble ikke gjort videre undersøkelser av hva testdeltakerene la i svaret sitt her. For en *between-group* design med en mer tydelig distinkt uavhengig variabel kunne man lagt mer vekt på rekrutteringsprosessen og sørget for å rekruttere testdeltakere som kunne dokumentere lang bakgrunn fra fagfeltet. Det er en mulighet for at man da ville fått andre resultater og sett større forskjell mellom gruppene. Testene gjort på denne studiens deltakerutvalg viser dog ingen klar indikasjon på at dette ville være tilfellet.

Ved sammenlikning av gruppe 1 og gruppe 3 i denne studien er det blitt sett på data om antall fikseringer og hvorvidt spesifikke relevante AOI for oppgaven fattet testdeltakerenes oppmerksomhet og hvor mye tid de fokusert på disse AOI. Det ble videre sett på hvor raskt de løste oppgaven og hvordan de gikk frem for å finne svaret. Det ble også sammenliknet hvor på skjermen de så, hvor lenge de scannet etter informasjonen og liknende, og en generell oppfatning av blikkmønsteret deres. En person med god erfaring fra analyse av eye tracking-opptak kunne kanskje sett enda mer detaljerte mønster i blikk, søkesti, fikseringer og sakkader, som lengde og retning på sakkader, lengde på fikseringer, og hva sammensetningen av dem betød.

## 6.4 Refleksjoner om eye tracking

### 6.4.1 Erfaring

Man gjør seg en del erfaringer ved gjennomføring av et slikt forsøk, spesielt når man benytter seg av en teknologi man ikke var kjent med på forhånd. Det var spennende og lærerikt, og i ettertid kan man gjøre seg tanker om at noe burde vært gjort annerledes, basert på tilegnet kunnskap. Det var svært verdifullt å få opplæring i utstyr fra kontaktperson på Skatteetaten, som var godt kjent med det i forbindelse med sitt arbeid. Fordi eye tracking-studier er avhengig av spesielt utstyr og programvare kan det være en høyre terskel for å bruke det i forsøk fremfor mindre ressursavhengige metoder som spørreundersøkelse eller intervju. Det kan også være at uerfarne forskere vil gjøre flere feil i begynnelsen som kan påvirke resultatene, dersom man ikke har satt seg godt nok inn i utstyr og metode. Praktisk erfaring man kan ta med seg fra denne studien er blant annet typen oppgaver, slik det ble diskutert i 6.3.3, der oppgavene bør være enda mer konkrete og korte, og gjerne også færre oppgaver i testsesjonen. En annen ting er selve opptakene, ved sammenhengende opptak av en lang testsesjon ville det vært lurt å dele opp opptaket noe underveis. Dette ville gjort den praktiske delen av databehandlingen noe lettere, og også forhindre at et helt testopptak forsvinner dersom tekniske feil oppstår.

I kapittel 2.1.6 om kritikk mot eye tracking ble det blant annet trukket frem hvordan Toftøy-Andersen og Wold (2011) hevder at det kreves en del forkunnskaper for å tolke resultatene fra eye tracking-opptak. De tar ikke direkte feil, ved første øyekast kan dataene – spesielt videoopptakene der man ser øyebevegelsene over skjermvideo i vanlig fart – virke overveldende og vanskelig å forstå. Likevel, det finnes nå mer og mer litteratur om emnet så man har mulighet til å støtte seg på andres ekspertise. Og selv om en med mye erfaring nok kan se flere sammenhenger og trekke flere slutninger fra dataene, kan også en uerfaren forsker oppdage interessante ting. I denne studien valgte man også å se på opptakene i sammenheng med statistikken og visualiseringene (heat maps og gazeplots). En slik triangulering av data gjorde det lettere å se mønster og forskjeller, eller bekrefte tendenser man kunne se i opptak.

Toftøy-Andersen og Wold (2011) har erfaring med Tobii eye tracker i forbindelse med brukstesting, og har brukt denne til opptak av navigasjon og sideflyt på nettsider. De poengterer at analyse av slik data er tidkrevende, og mener det ikke er nok nytteverdi i det



iforhold til tidsbruk. Erfaring fra denne studien stemmer ikke helt overens med det de hevder her. Ja, analysen av slike opptak er tidkrevende. Selv med hensyn som konkrete og korte oppgaver vil analysen kreve tid. Det var likevel mange små ting man la merke til underveis i gjennomgang av opptak fra dette forsøket – like mønster i øyebevegelsene eller avvikende mønster for eksempel – som ga et inntrykk av at eye tracking har mange styrker som kan nyttegjøres i brukstesting.

### **6.4.2 Styrker**

En interessant observasjon fra opptak viste at det var variabelt hva testdeltakeren gjorde med musepekeren mens de scannet nettsiden. Noen lot musepekeren i perioder følge blikkets søkesti, men mange lot den også være i ro et sted på siden mens de scannet med blikket. Tidligere forskning bekrefter også at det ikke er gitt at musepekerens posisjon sier noe om hvor brukeren ser (Granka et. al, 2004). Ved noen tilfeller holdt brukeren musepekeren over en potensiell lenke til riktig svar mens de scannet videre på siden, andre ganger holdt de musepekeren over lenken mens de fokuserte på den. Uten eye tracking-dataene til å støtte skjermopptaket er det umulig å vite om brukeren holder musepekeren på lenken som et holdepunkt, eller om brukeren ser lenge på lenken for å vurdere riktighet. Disse tingene støtter nytteverdien av eye tracking-opptak for bedre innsikt i brukerens fokus, sammenliknet med kun skjermopptak.

Erfaring fra analyse av opptak i denne studien støtter også at det er et nyttig verktøy for å finne feil og problemområder på nettsider, og potensielle løsninger på dette. Et konkret eksempel fra denne studien er resultatene fra oppgave 11 der deltakerene ble bedt om å finne Skatteetatens telefonnummer når de sto på login-siden til Altinn. Når man trykker på en lenke på Skatteetaten, i dette tilfellet var det for å bestille nytt skattekort, blir man rutet videre til Altinn sin login-side, men det kommer ikke klart frem at man forlater Skatteetaten og havner på Altinn. Det er som nevnt tidligere en mulighet at det er bevisst fra utviklers side for å skape sømløshet og en helhetlig opplevelse for bruker, men eye tracking-opptakene fra denne oppgaven viser at det også kan skape problemer. Blant annet var det svært få som fikserte på lenken øverst der man kunne gå tilbake til Skatteetaten. Som utvikler kan man se på disse dataene og konkludere med at dette er et problem, og at en eventuell løsning kunne være at lenken bør gjøres mer synlig.

Eye tracking er ikke en helt ny teknologi, men utvikling av teknologien har gjort den mer tilgjengelig og aktuell som verktøy i forskning og brukstesting det siste tiåret. Altså vil det nok for mange være en ny metode. At noe er nytt kan såklart være en svakhet på grunn av terskelen for å bli en erfaren bruker, men det gjør også at det er spennende og interessant, og kan kanskje gjøre utviklere mer interessert i å se resultater fra brukstesting? Eye tracking gir en unik mulighet til å supplere brukstester av nettsider med visuelle data som viser veldig konkret hva brukeren har sett på.

### 6.4.3 Svakheter

Noe data fra eye tracking-opptak står dårlig alene. Et eksempel fra studien ser vi i oppgave 14. Der fokuserte gruppe 1 mer på området Aktuelle tjenester, mens gruppe 3 fokuserte mest på den grafiske snarveien. Man kan ved hjelp av eye tracking-opptakene observere og konstatere *at* dette skjer, men uten tilbakemelding fra brukerne er det ikke alltid man forstår *hvorfor*. Videre studier av eye tracking som metode, av typen Ehmke og Wilson (2007) utførte, vil medføre at man får mer kunnskap om dette slik at det blir lettere å trekke konklusjoner fra eye tracking alene.

En annen svakhet ved eye tracking-forsøk er den store variasjonen i utstyr og programvare. Dette er i hovedsak en svakhet i det man ønsker å sammenlikne data fra forskjellige studier. Presentasjon av dataene svært variabelt, noe man raskt ser ved å se på tidligere forskning der metoden er brukt. Noen forsøk presenterer resultatene som tabeller med statistikk over antall fikseringer og andre målte verdier (Lykins et. al (2006). Andre presenterer grafer av de samme verdiene (Vogt og Magnussen, 2007) (Granka et. al., (2004). Andre igjen støtter seg på grafiske presentasjoner som for eksempel gaze plots (Eivazi et. al., 2012).

Begrenset antall deltakere er ikke uvanlig i eye tracking-forsøk. Praktiske faktorer gjør det vanskeligere å gjennomføre forsøk med store grupper av testdeltakere. Dette kan anees som en svakhet dersom man ønsker statistisk signifikante data, og sikkerhet om at resultatene ens er gjeldende for den generelle befolkningen. Når det gjelder dette er det aktuelt å se på de forskjellige typer stimuli man kan bruke, og hvilke konsekvenser det har for praktisk gjennomføring av forsøk og analyse. Dersom stimuli er stillbilder og testbruker er passiv vil det være enklere å gjennomføre forsøk med større antall testdeltakere. I en studie av typen utført her, der stimuli er en nettside og brukeren er aktiv, vil det være mer utfordrende med et høyt antall.

En annen fremgangsmåte for å omgå denne utfordringen er å utføre en pre-studie. Basert på egen erfaring fra denne studien virker dette som en god løsning dersom man ønsker å utføre et eye tracking-forsøk. I en pre-studie kan man se på et begrenset antall deltakere. Man kan så trekke ut interessante observasjoner som man ønsker å studere videre i en større studie. Testoppgaver og utvalg av deltakere kan så tilspisses ytterligere basert på erfaringer fra pre-studien.



# 7 Konklusjon

## 7.1 Sammendrag

Hensikten med denne oppgaven var å se på øyebevegelsene til brukere som utførte oppgaver på Skatteetatens nettsider. Basert på opptak av blikket ønsket man å finne ut om det var mulig å se forskjellige mønster hos personer med bakgrunn fra fagfeltet interaksjonsdesign og personer uten slik bakgrunn. Testdeltakerene bakgrunn ble bestemt etter egenopplyst erfaring fra interaksjonsdesign, webdesign, eller brukervennlighet på nett – i studier eller arbeid. Det var også ønskelig å reflektere rundt teknologien eye tracking.

For å kunne observere brukerens blikkmønster ble eye tracking-teknologi benyttet. Eye tracking-teknologi kan gjøre opptak av akkurat hvor brukeren fokuserer mens de gjør oppgaver eller ser på bilder på en skjerm. Nyere teknologi kan gjøre dette uten at brukeren må ha forstyrrende elementer som hodeutstyr festet på seg. Det ble brukt slik ikke-forstyrrende teknologi i dette forsøket, en eye tracker kalt Tobii T60XL.

Eye tracking har blitt et populært verktøy innen HCI-forskning (Eivazi et. al., 2012). Det er også brukt innen andre fagfelt som for eksempel psykologi (Vogt og Magnussen, 2007). I tillegg til akademiske studier er det et verktøy som benyttes for å teste brukskvalitet, og til kommersielle formål som å kartlegge hensiktsmessig plassering av reklame (Nielsen og Pernice, 2010).

Det ble utført et eksperimentelt forsøk under kontrollerte forhold på en lab i Skatteetatens lokaler i Oslo. 14 testdeltakere ble rekruttert til testen. I forkant av testen svarte de på en spørreundersøkelse som ble brukt til å kartlegge deres bakgrunn og gruppere dem i henholdt til forsøkets problemstilling. Det ble deretter gjennomført opptak med en testdeltaker om gangen, der de utførte 27 oppgaver på Skatteetatens nettsider. Oppgavene ble gitt til dem muntlig av en testleder som oppholdt seg i samme rom som testdeltakerene. Opptakene ble senere bearbeidet i programvaren Tobii Studio. Det ble ved hjelp av programvaren hentet ut visuell data og statistikk fra syv av oppgavene, samt at opptakene ble manuelt gjennomgått av testleder. Dette var grunnlaget for analysen.

## 7.2 Konklusjon

Basert på data fra de syv oppgavene presentert og diskutert i kapittel 6.2 og 6.3 kan man ikke trekke en definitiv konklusjon om at gruppene har forskjellig blikkmønster. Tvert i mot viser flere heat maps og gaze plot fra de første sekunder av en oppgave svært like øyebevegelser. En nøye sammenlinkning av heat maps viste videre at gruppe 3 var noe mer fokuserte på relevante områder av skjermen, mens gruppe 1 hadde noen flere fikseringer i irrelevante områder av skjermbildet. Forskjellen var dog subtil. Statistikken viser ikke signifikante forskjeller på antall fikseringer eller tid brukt innen relevante AOI. De verdiene som var statistisk signifikante gjaldt tid til første fiksering i et relevant AOI. Her var det først og fremst gruppe 3 som hadde lave gjennomsnittsverdier, men det var ikke konsekvent gruppe 3 som hadde lavest. Også gruppe 1 hadde lavest verdi på to oppgaver. Det var i tillegg én oppgave gruppe 3 løste raskere, og én oppgave der gruppe 3 hadde en klar høyere suksessrate. Man kan altså spekulere i at gruppe 3 hadde en fordel på grunn av sin bakgrunn. Disse to observasjonene sier dog ikke noe om blikkmønster, kun noe om testdeltakerenes evne til å raskere løse oppgaven eller prosessere informasjonen de leste. Ved manuell observasjon av eye tracking-opptakene ble det ikke oppfattet klare forskjeller i blikkmønster. Det så derimot ut til at blikkmønsteret i større grad var relatert til type oppgave og eventuelle problemer med å finne løsninger.

Konklusjonen på problemformuleringen er at studien ikke påviste klare og konsekvente forskjeller i blikkmønsteret hos til brukere som har bakgrunn innenfor fagfeltet interaksjonsdesign, og brukere som ikke har det.

Når det gjelder eye tracking som metode er det definitivt en interessant teknologi. Erfaring fra gjennomføring av denne studien, og gjennomgang av opptak, støtter at dette er et nyttig verktøy for brukstesting av nettsider. Det er nyttig fordi det gir et godt innblikk i hva brukeren gjør, og gir inntrykk av å være et godt hjelpemiddel for å kunne forstå problemer med design eller informasjon på nettsiden, og hvordan problemene kan løses. I akademiske studier kan det også gi nyttige og interessante resultater, men det er viktig med godt forarbeid for å få data som er hensiktsmessige å jobbe videre med. Dette er en nødvendighet ved all type datainnsamling, men ved eye tracking møter man flere utfordringer: mange forskjellige metoder for opptak, forskjellig type utstyr, og varierte variabler for presentering av data.

Forskere som ikke har brukt teknologien før bør konferere med mer erfarne brukere, eller gjennomføre et pre-studie før det settes i gang med store forsøk.

## 7.3 Videre arbeid

Konklusjonen fra denne studien viser ingen betydelig forskjell mellom de to brukergruppene. Resultatene presentert viser noen forskjeller, men de er ikke konsistente. Med det menes at det ikke er observert noen forskjeller i blikkmønster som gjentar seg gjennom resultatene. Det er derfor ingen klare indikasjoner på at det kan være forskjeller som ville kommet tydeligere frem i et forsøk med et større antall testdeltakere. Dersom man ønsket å se videre på studiens problemstilling med nye forsøk kunne det definitivt vært aktuelt å bruke et bredere utvalg av nettstedet som grunnlag for opptak. En mer distinkt forskjell på novise- og ekspertgruppen ville også vært hensiktsmessig ved en eventuell videre undersøkelse. Man skal ikke se bort i fra at en ny studie med disse to endringene ville ført til et mer klart resultat.

Noen generelle tendenser og interessante observasjoner ble gjort, som nok kan skape nye problemstillinger for videre forsøk. Et eksempel som utmerket seg her var hvordan brukerne scannet menyer, slik det er diskutert i 6.3. En videre undersøkelse av dette vil være nyttig for utvikling av nettsteder. Dersom det viser seg at brukere ikke konsekvent scanner menyer fra topp til bunn, men har et annet mønster for å lete etter riktig menyvalg, kan det påvirke hvor man skal plassere de viktigste menyvalgene. En slik studie kan gjøres ganske enkel, eller den kan inkludere forskjellige menytyper og menyer med variabelt antall lenker.

Skattetaten er en side som inneholder mye informasjon om et spesielt emne, skatt og økonomi. I den initiale undersøkelsen ble testdeltakerene spurt om sin bakgrunn fra dette, om de hadde jobbet eller studert emnet. I dette utvalget testdeltakere var det kun en liten andel som hadde denne bakgrunnen, og det ble valgt å ikke legge vekt på det. Ved videre analyse av opptakene kunne det vært interessant å se på om dette var en relevant faktor. Ved indikasjoner på forskjeller kunne man vurdert en ny undersøkelse med et bedre utvalg for å undersøke dette.

Et scenario der man ønsker å sammenlikne eksperter og noviser åpner for mange mulige problemstillinger. I dette eksperimentet ble det sett på personer som hadde eller ikke hadde bakgrunn innenfor et spesifikt fagområde. Tidligere forskning henvist til i kapittel 3 der problemstillingen ekspert versus novise var i fokus så blant annet på noe liknende, da Vogt og

Magnussen (2007) sammenliknet kunststudenter – som da har studert fagfeltet kunst – med personer uten noen ekspertise på dette området. Det var også studier som så på forskjellige nivåer innenfor en allerede avgrenset gruppe, erfarne kirurger sammenliknet med nyutdannede kirurger (Eivazi et. al., 2012). Begge disse tilnærmingene kan overføres til andre fagfelt. Innenfor IKT er et åpenlyst eksempel erfarne og uerfarne databrukere.

Denne oppgaven var basert på bruk av offentlige tjenester, spesielt Skatteetatens nettsider. Offentlige nettsider skal brukes av en svært stor og variert gruppe mennesker, og det kan derfor tenkes å være nyttig om det blir gjort flere undersøkelser på personer med forskjellig kunnskapsnivå, for å bedre lage et godt tilbud for alle. Et eksempel er unge mennesker idag som er vokst opp med data, Internett og tilgang til store datamengder. De offentlige tjenestene på nett skal tilpasses til disse på lik linje med eldre personer som er nye brukere av slike verktøy. Eye tracking-forsøk kunne nok blitt brukt som et ledd i å finne ut av eventuelle forskjeller mellom slike grupper, og videre hvordan man kan bruke denne kunneskapen til å lage bedre tilpassede tjenester. Et annet eksempel som kanskje kunne vært interessant ville være å se på nye brukere av en tjeneste versus brukere som er godt kjent med tjenesten.

Den objektive tilbakemeldingen om hvor brukeren ser som man får fra eye tracking-data gjør det naturlig å tenke på mulighetene for å utføre forsøk på personer som ellers ikke kunne gitt nyttige eller korrekte tilbakemeldinger selv, for eksempel barn.

Poole og Ball (2005) konkluderte at den positive utviklingen av og den økte tilgjengeligheten på eye tracking-utstyr ville føre til videre bruk innen HCI-forskning. Denne utalelsen kan godt gjentas i dag. Generelt kan man anta at nyere eye tracking-utstyr, samt utvikling i kompetanse på både bruk av utstyr og analyse av eye tracking-data, vil gi større muligheter for å bruke det til et bredt spekter av brukstesting, kommersielle tjenester og akademiske forsøk i fremtiden.



# Litteraturliste

- Baxter, Pamela, and Susan Jack. "Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers." *The Qualitative Report*, volum 13, nr. 4 (2008): 544-559
- Bojko, Agnieszka. "Using Eye Tracking to Compare Web Page Designs: A Case Study." *EServer TC Library*, January 1, 2006.
- Dagbladet. "Lar deg styre mobilen med øyet". (Besøkt oktober 2012)  
<http://www.dbtv.no/?vid=1914806016001>
- Duchowski, Andrew T. *Eye Tracking Methodology: Theory and Practice*. New York: Springer, 2003.
- Ehmke, Claudia, Stephanie Wilson. "Identifying Web Usability Problems from Eye-tracking Data." 119-128. British Computer Society Swinton, UK. 2007
- Granka, Laura A., Thorsten Joachims, and Geri Gay. "Eye-tracking Analysis of User Behavior in WWW Search." 478. ACM Press, 2004.
- Gray, Henry. *Anatomy of the Human Body*. Philadelphia: Lea & Febiger, 1918; Bartleby.com, 2000. <http://www.bartleby.com/107/225.html> (Besøkt februar 2012)
- Institutt for Informatikk "Autonomi og automatisering i et informasjonssamfunn for alle". Besøkt oktober 2012.  
<http://www.mn.uio.no/ifi/forskning/prosjekter/autonomi-og-automatisering/index.html>
- Jacob, R.J.K., and K.S. Karn, "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary)," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, ed. by J. Hyona, R. Radach, and H. Deubel, pp. 573-605, Amsterdam, Elsevier Science (2003)
- Just, Marcel Adam, and Patricia A. Carpenter. "A theory of reading: From eye fixations to comprehension" *Psychological Review* 87 (1980): 329-354.
- Law, Benjamin, M. Stella Atkins, A. E. Kirkpatrick, Alan J. Lomax, and Christine L. Mackenzie. "Eye Gaze Patterns Differentiate Novice and Experts in a Virtual Laparoscopic Surgery Training Environment." 41-48. ACM Press, 2004.
- Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Chichester, West Sussex, U.K.: Wiley, 2010.

- Lykins, A.D., Meana M., and Kambe G., "Detection of differential viewing patterns to erotic and non-erotic stimuli using eye-tracking methodology." *Arch Sex Behav.* 2006;35(5): 569-75.
- Moore, Robert J., and Elizabeth F. Churchill. "Computer Interaction Analysis: Toward an Empirical Approach to Understanding User Practice and Eye Gaze in GUI-Based Interaction." *Computer Supported Cooperative Work (CSCW)* 20, no. 6 (September 21, 2011): 497–528.
- Nielsen, Jakob, and Kara Pernice. *Eyetracking Web Usability*. Berkeley, CA: New Riders, 2010.
- Office, "Funksjonen TTEST". (Besøkt februar 2013)  
<http://office.microsoft.com/nb-no/excel-help/funksjonen-ttest-HP010062550.aspx>
- Poole, Alex, and Linden J. Ball. "Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects." Chapter in C. Ghaoui (Ed.): *Encyclopedia of Human-Computer Interaction*. Pennsylvania: Idea Group, Inc. 2005.
- Reed, J.B., and Meyer, R.J. "Edmund Burke Huey (1870-1913): A Brief Life With an Enduring Legacy." In *S.E. Israel, & E. Monaghan (Eds.), Shaping the Reading Field* (pp. 159-175). Newark, DE: International Reading Association. 2006.
- Shahram Eivazi, Shahram, Roman Bednariky, Markku Tukiainen, Mikael von und zu Fraunbergx, Ville Leinonen, and Juha E Jääskeläinen. "Gaze Behaviour of Expert and Novice Microneurosurgeons Differs During Observations of Tumor Removal Recordings." 377-380. *ACM Press*, 2012.
- Skatteetaten. "Skatteetaten.no". (Besøkt januar-oktober 2012)  
<http://www.skatteetaten.no>
- Skatteetaten. "Brukerundersøkelse om lønnstakere og pensjonisters oppfatning av Skatteetaten – 2010", 2010.
- Suchman, Lucille Alice. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, 1987.
- Suchman, Lucille Alice. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, 2007.
- SurveyMonkey. "Smart Survey Design". 2010.
- Tatler, Benjamin W., Nicholas J. Wade, Hoi Kwan, John M. Findlay, and Boris M. Velichkovsky. "Yarbus, Eye Movements, and Vision." *i-Perception* 1, no. 1 (2010): 7–27.
- Tobii "Eye tracking by Tobii Technology." (Besøkt mars 2012)

<http://www.tobii.com>

Tobii "Assistive Technology, Norge, Produkter." (Besøkt mars 2012)

<http://www.tobii.com/no/assistive-technology/norway/produkter/hardware/>

Tobii Studio User Manual, Tobii. 2012.

Toftøy-Andersen, Eli, and Jon Gunnar Wold. *Praktisk brukertesting*. Oslo: Cappelen Damm akademisk, 2011.

Tullis, Tom, and Bill Albert. *Measuring the user experience collecting, analyzing, and presenting usability metrics*. Amsterdam; Boston: Elsevier/Morgan Kaufmann, 2008.

UX Booth, David Leggett. "A brief history of eye-tracking" Sist oppdatert 19. januar, 2010.

(Besøkt mars 2012)

<http://www.uxbooth.com/blog/a-brief-history-of-eye-tracking/>

Vogt, Stine, and Svein Magnussen. "Expertise in Pictorial Perception: Eye-movement Patterns and Visual Memory in Artists and Laymen." *Perception* 36, no. 1 (2007): 91 – 100

Yusuf, S., Kagdi, H., Maletic, J.I. , "Assessing the Comprehension of UML Class Diagrams via Eye Tracking". *Program Comprehension, 2007. ICPC '07. 15th IEEE International Conference on* , vol., no., pp.113-122, 26-29 June 2007



# Vedlegg

- A. Samtykkeskjema
- B. Spørreundersøkelse, spørsmål og resultater
- C. Testoppgaver
- D. Tabell, post task ratings.



## Vedlegg A – Samtykkeskjema

### Betingelser for eye tracking-test med opptak

1. **Du vet om opptak.** Du skal vite om opptaket og hva det inneholder. Du kan selv se på opptaket, og du kan be om at opptaket blir slettet før du går idag, dersom du ønsker det.
2. **Du vet om observasjon.** Intervjuet observeres kun av masterstudent Nora Raaum.
3. **Personvern.** Vi bruker kun fornavn under opptak. Du skal ikke oppgi sensitive personopplysninger, muntlig eller i det du taster inn på datasystemet. For videre analyse av opptaket vil det brukes et nummer istedet for navnet ditt. Ansiktet ditt og stemmen din kan gjenkjennes i opptaket hvis det er noen som kjenner deg som ser det.
4. **Hvem får se opptakene.** Opptak av ditt ansikt og stemme vil i utgangspunktet kun sees gjennom og analyseres av masterstudent Nora Raaum. I tillegg kan de sees av veiledere Alma Leora Culén og Guri B. Verne, samt kontakt hos skatteetaten, Laura Arlov. Skjerm bilde av eye tracking-opptak kan brukes i oppgaven. Skjermopptak og eye tracking-data kan brukes til presentasjon av oppgaven for sensor, samt på eventuelle konferanser eller andre akademiske arrangementer. Det vil isåfall ikke knyttes til ditt navn eller ansikt.

For tillatelse til å omgå dette punktet, vennligst kryss av på det som er greit for deg. Dette vil gjelde kortere klipp, ikke hele opptaket.

- a. Bilde av meg kan brukes i oppgaven og i presentasjon for sensor og veiledere.
  - b. Video og lydopptak av meg kan brukes i presentasjon av oppgaven for sensor og veiledere.
  - c. Video og lydopptak av meg kan brukes senere til presentasjon på konferanser eller andre akademiske arrangementer.
5. **Opptakene blir ikke spredd.** Opptakene eies av masterstudent Nora Raaum og oppbevares av Skatteetatens gruppe for Brukskvalitet. Kopi av kortere klipp vil oppbevares av meg. Jeg har ikke anledning til å gi kopi til andre eller legge ut kopier på internett eller Skatteetatens interne nett.
  6. **Opptakene slettes.** Opptakene fra dagens intervju blir slettet når jeg ikke lenger har behov for dem, altså etter sensur av masteroppgave. De vil senest slettes ved utgangen av 2012. Unntak er klipp fra opptak etter avtale under punkt 4.

**Jeg har lest betingelsene og gir mitt samtykke:**

Testnummer: \_\_\_\_\_

Dato: \_\_\_\_\_

Navn: \_\_\_\_\_

Signatur: \_\_\_\_\_

## Vedlegg B – Spørreundersøkelse, spørsmål og resultater

### Rapport fra «EyeTracking»

Innhentede svar pr. 18. februar 2013 14.45

Uten fritekstsvar Med fritekstsvar

#### Testnummeret ditt \*

Du finner dette i eposten jeg har sendt deg.

- 3
- 11
- 10
- 9
- 04
- 15
- 06
- 14
- 8
- 13
- 07
- 12
- 05
- 02

#### Hva er din alder? \*

Svar	Antall	Prosent
under 20	0	0.0 %
20-29	4	28.6 %
30-39	2	14.3 %
40-49	6	42.9 %
50-59	2	14.3 %
60 eller eldre	0	0.0 %

#### Hva gjør du til daglig? \*

Du kan her avgi flere svar.

Svar	Antall	Prosent
Jobber deltid	2	14.3 %
Jobber heltid	12	85.7 %
Studerer deltid	0	0.0 %
Studerer heltid	2	14.3 %
Jobber som selvstendig næringsdrivende	0	0.0 %
Permisjon/arbeidsledig/ufør/sykemeldt/pensjonist	0	0.0 %

#### Hva jobber du med/studerer du? \*

Kort forklart, for eksempel stillingstittel eller studieretning.

- mediekonsulent
- Saksbehandler pensjon
- IT-prosjekter
- Prosjektmedarbeider
- Interaksjonsdesigner
- Master i informatikk
- Tjenesteutvikling (Internettjenester)
- Senioringeniør, datasikkerhet
- HR-kordinator SITS. Bemanning, lønn, personal.
- Bibliotekar
- IT
- Funksjonell IT-arkitekt
- adjunkt, kontaktlærer i småskolen.
- Arbeid: Rådgiver forebyggende veiledning (utlånt til SITS som Altinn-tester) Studerer revisjonsfag

#### Studerer du på, eller har du tidligere gått studie, som inneholder emner om interaksjonsdesign, webdesign eller brukervennlighet på nett? \*

Svar	Antall	Prosent
Ja, ved universitet	3	21.4 %
Ja, ved høyskole	0	0.0 %
Ja, ved annet type studiested	0	0.0 %
Nei	11	78.6 %
Kan/vil ikke svare	0	0.0 %



**Studerer du på, eller har du tidligere gått studie, som inneholder emner om personlig økonomi eller skatteregler? \***

Svar	Antall	Prosent
Ja, ved universitet	2	14.3 %
Ja, ved høyskole	3	21.4 %
Ja, ved annet type studiested	0	0.0 %
Nei	9	64.3 %
Kan/vil ikke svare	0	0.0 %

**Har du en jobb, eller har du tidligere hatt en jobb, med arbeidsoppgaver inkludert interaksjonsdesign, webdesign eller brukervennlighet på nett? \***

Svar	Antall	Prosent
Ja, stor del av jobben	4	28.6 %
Ja, liten del av jobben	5	35.7 %
Nei	5	35.7 %
Kan ikke svare	0	0.0 %

**Har du en jobb, eller har du tidligere hatt en jobb, med arbeidsoppgaver om personlig økonomi eller skatteregler? \***

Svar	Antall	Prosent
Ja, stor del av jobben	3	21.4 %
Ja, liten del av jobben	1	7.1 %
Nei	10	71.4 %
Kan/vil ikke svare	0	0.0 %

**Har du besøkt Skatteetatens nettsider før? \***

Svar	Antall	Prosent
Ja, ofte	1	7.1 %
Ja, noen ganger	11	78.6 %
Ja, en gang	1	7.1 %
Nei	0	0.0 %
Vet ikke	1	7.1 %

**Har du bestilt skattekort via Skatteetatens nettsider før? \***

Svar	Antall	Prosent
Ja	8	57.1 %
Nei	6	42.9 %
Vet ikke	0	0.0 %

**Har du levert selvangivelsen via Skatteetatens nettsider før? \***

Svar	Antall	Prosent
Ja	12	85.7 %
Nei	2	14.3 %
Vet ikke	0	0.0 %

**Har du opplevd problemer med Skatteetatens nettsider tidligere? \***

Du kan her avgi flere svar.

Svar	Antall	Prosent
Har ikke brukt Skatteetatens nettsider	0	0.0 %
Nei	4	28.6 %
Ja, med å finne informasjon	7	50.0 %
Ja, med å finne skjema	3	21.4 %
Ja, med innlogging	5	35.7 %
Ja, med annet	0	0.0 %

### Bruker du datamaskin når du jobber? \*

Til arbeidsrelaterte oppgaver.

Svar	Antall	Prosent
Jobber ikke	0	0.0 %
Ja, litt	1	7.1 %
Ja, mye	13	92.9 %
Nei	0	0.0 %

### Bruker du datamaskin når du studerer? \*

Til studierelaterte oppgaver.

Svar	Antall	Prosent
Studerer ikke	9	64.3 %
Ja, litt	1	7.1 %
Ja, mye	4	28.6 %
Nei	0	0.0 %

### Hvor mye surfer du på nett for informasjon eller underholdning?

Jeg tenker her på nettbruk utenom studie- og arbeidsrelaterte oppgaver.

Svar	Antall	Prosent
Flere timer daglig	5	35.7 %
Litt hver dag	7	50.0 %
Flere timer ukentlig	0	0.0 %
Litt hver uke	2	14.3 %
Sjeldnere	0	0.0 %

### Enig/uenig

Hvor enig er du i de følgende utsagnene? Vennligst ranger på en skala fra 1 til 5 der 1 betyr "Helt uenig" og 5 betyr "Helt enig".

#### Svar fordelt på antall

	1	2	3	4	5	Vet ikke
Jeg interesserer meg for webdesign/interaksjonsdesign/brukervennlighet på nett. *	0	0	4	5	4	1
Jeg kan mye om webdesign/interaksjonsdesign/brukervennlighet på nett. *	2	5	6	0	1	0
Jeg interesserer meg for skatteregler i Norge. *	3	2	6	3	0	0
Jeg kan mye om skatteregler i Norge. *	3	5	3	3	0	0
Jeg foretrekker å kommunisere med offentlige tjenester via epost fremfor telefon eller brev. *	0	1	3	2	7	1

#### Svar fordelt på antall

	1	2	3	4	5	Vet ikke
Jeg finner ofte informasjon om økonomi og skatt via nettet. *	2	2	2	4	2	2
Jeg foretrekker å finne informasjon om skatteregler o.l. på Skatteetatens nettsider fremfor Googlesøk eller andre nettsider. *	2	1	3	1	5	2
Jeg foretrekker å kontakte Skatteetaten via telefon dersom jeg lurer på noe angående skatt. *	7	2	3	0	1	1
Jeg føler at det er trygt å ordne med skattetjenester som bestilling av skattekort, selvangivelse osv. via nettet. *	0	0	0	3	11	0
Jeg synes det er enkelt å ordne med skattetjenester som bestilling av skattekort, selvangivelse osv. via nettet. *	0	2	1	5	5	1
Jeg er flink til å navigere og finne informasjon jeg trenger på nettet. *	0	1	0	9	4	0
Jeg blir usikker når jeg må finne informasjon om skatt og økonomi på nettet. *	3	4	3	3	0	1

**Svar fordelt på antall**

	1	2	3	4	5	Vet ikke
Jeg har god oversikt over egen økonomi. *	0	1	1	6	6	0
Jeg blir sjeldent overrasket over beløpet på restskatt/skatt tilgode når selvangivelsen kommer. *	0	0	3	4	7	0
Det er ikke viktig for meg å ha kunnskap om skatteregler. *	5	3	3	2	1	0
Jeg stoler på at den ferdigutfylte selvangivelsen er korrekt. *	1	2	3	4	4	0

## Vedlegg C – Testoppgaver

Naviger til Skatteetatens hjemmeside.

1. Finn siden med generell informasjon om skattekort.
2. På denne siden, finn menyvalget som tar deg til siden om arveavgift og klikk på det.
3. På denne siden, finn menyvalg for å åpne skjema for registrering av gaver.
4. På denne siden, finn menyvalget som tar deg til siden om moms og klikk på det.

Naviger til forsiden

5. Finn ut hvor mye en ugift 20åring kan tjene i året før han eller hun må betale skatt.
6. H\*n forventer å tjene rundt 20 tusen. Åpne skjema for å bestille riktig type skattekort til denne personen.
7. Finn ut hva nummeret er for å bestille frikort via sms.
8. Gjør teksten på siden større og så tilbake til normalt igjen.

Naviger til forsiden

9. Du har byttet til en jobb der du får høyere lønn. Finn informasjon om hvordan du kan få tilsendt nytt skattekort i posten basert på ny lønn.
10. Finn skjema for å bestille det nye skattekortet du trenger via nettet.
11. Du husker ikke passord for MinID. Finn telefonnummeret du kan ringe for å bestille skattekort istedet.

Naviger til forsiden

12. Finn ut hvordan du kan bestille ny pinkode til MinID.

Naviger til forsiden

13. Finn skjema du kan printe ut og sende i posten, for å endre skattekortet ditt.

Naviger til forsiden

14. Du har fått igjen på skatten og lurer på når du får pengene på konto. Finn informasjon om hvilken dato du får pengene dine.

(14 1/2) Oppmerksomhetsoppgave, på papir

Informasjon til bruker: Forsøk å gjøre de neste oppgavene uten å bruke søkefunksjonen

Naviger tilbake til siden med generell info om skattekort. (Dersom de gjorde feil, be dem gå til siden via menyvalg nederst på skatteetaten.no)

15. Du har fått en ekstra jobb og har allerede gitt skattekortet du fikk i posten til hovedarbeidsgiver, og finner ikke den andre delen. Åpne riktig skjema for å få nytt skattekort tilsendt, uten å endre informasjonen på kortet. (via nett)
16. Du har ombestemt deg og vil ikke bestille kopi, gå tilbake til siden du var på.
17. Du har bestilt kopi av skattekort men fikk det ikke i posten. Se om du kan finne informasjon om hvorfor det kan være på sidene om skattekort.
18. Du vil sende epost til Skatteetaten for å spørre om ting du lurer på rundt bestilling av skattekort. Finn riktig epostadresse eller åpne kontaktskjema for å gjøre dette.

Naviger (tilbake) til siden med generell kontaktinformasjon til Skatteetaten

19. Finn navnet på Skatteetatens Twitter-konto.
20. Finn adressen til skattekontoret som er nærmest der du bor

Naviger til forsiden

21. Anta at du har tabelltrekk med tabellnummer 7200 og månedslønnen din er 20000. Finn ut hvor mye du blir trukket hver måned.
22. Du er litt usikker på om du valgte rett alternativ på "Trekperiode". Finn ut hva dette betyr uten å navigere ut av skjemaet for tabelltrekk-info.

Naviger til forsiden

23. Finn menyvalget for å vise "ofte stilte spørsmål" og klikk på det.

Naviger til forsiden

24. Finn ut hvor lenge du må bo i egen bolig før du kan selge den uten å betale skatt.
25. Finn ut hvor mye du kan tjene på utleie av hele boligen din uten å måtte skatte av det.
26. Finn ut om det er en engelsk versjon av sidene om utleie og salg av bolig.



## Vedlegg D – Tabell, post task ratings

Tabell med verdiene fra testdeltakerenes tilbakemeldingsskjema. Hver oppgave ble rangert etter opplevd vanskelighetsgrad.

Rader: Oppgavenummer.

Kolonner: Testdeltake

Skala: 1-5 der 1=Lett og 5=Vanskelig

	1	2	3	4	5	6	7	8	9	10	11	12	13	14*	Gj.snitt
1	2	2	1	1	1	3	1	2	2	3	2	1	1	3	1,67
2			1	1	1	4	1	1	1	1	1	1	1	2	1,23
3			1	1	2	4	1	2	2	2	3	1	1	2	1,69
4	1	3	1	1	1	4	4	1	1	1	3	1	1	3	1,73
5	4	3	1	2	3	4	1	3	1	4	2	2	4	4	2,53
6	1	1	2	3	5	4	2	1	1	3	1	1	1	2	1,87
7	1	1	1	2	1	3	1	1	1	1	1	1	1	4	1,33
8		1	1	1	1	4	1	3	1	1	1	1	1	3	1,43
9	1	3	1	1	2	4	1	1	1	1	3	1	1	1	1,47
10			1		2			1			2	1	1		1,33
11	2	3	1	1	1	4	1	1	3	3	5	1	1	5	2,13
12	1	3	1	1	1	4	1	2	1	1	2	1	1	1	1,4
13	2	2	1	1	4	3	3	1	2	1	3	1	1	2	1,8
14	1	3	1	2	2	4	2	1	1	1	2	2	1	2	1,67
15	1	2	1	1	3	4	1	1	1	2	2	1	1	3	1,6
16	1	2	1	5	4	5	5	3	3	1	1	1	1	5	2,53

<b>17</b>	5	2	1	2	1	2	5	5	3	2	4	3	1	4	2,67
<b>18</b>	1	2	1	1	3	3	3	3	2	3	4	4	2	5	2,47
<b>19</b>	1	2	1	1	1	4	1	1	1	1	2	1	1	1	1,27
<b>20</b>	1	4	1	1	5	4	1	2	2	4	4	1	1	5	2,4
<b>21</b>	1	4	1	1	2	4	1	1	1	3	2	1	1	2	1,67
<b>22</b>	1	5	1	1		4	1	1	2		1	1	1	2	1,62
<b>23</b>	1	4	1	1	4	4	1	1	1	2	1	2	1	3	1,8
<b>24</b>	2	3	1	2	5	4	5	4	4	4	5	1	2	3	3
<b>25</b>	2	3	1	1	3	4	4	2	2	1	1	2		5	2,21
<b>26</b>	1	5	1	2	5	4	5	3	4	2	3	5	2	4	3,07
<b>Gj.snitt</b>	1,55	2,74	1,04	1,48	2,52	3,8	2,12	1,85	1,76	2	2,35	1,5	1,24	3,04	

\*Deltakeren som ikke ble inkludert i analysen av eye tracking-data på grunn av tekniske problemer med opptak. Er inkludert i denne tabellen på grunn av at tilbakemeldingsskjemaene ble behandlet før eye tracking-dataene. Fordi hensikten med tabellen kun var å få en omtrentlig oversikt over opplevd vanskelighetsgrad ble det ansett som ok at det ikke ble gjort nye beregninger uten denne testdeltakeren. Hans tilbakemelding sier også noe om oppgavens vanskelighetsgrad selv om eye tracking-opptaket ikke ble med i analysen.





