STATSVETENSKAPLIGA INSTITUTIONEN

Artificiell rättvisa?

En experimentell studie på effekten av automatiserat beslutsfattande på proceduriell rättvisa

Elsa Höök

Examensarbete:	15 hp
Program:	Statsvetarprogrammet
Nivå:	Grundnivå
Termin:	Höstterminen 2021
Handledare:	Jacob Sohlberg
Antal ord:	11 960

Abstract

We are currently facing a dramatic change in the public sector. Technological advances in recent years have made it possible for allowing algorithms and artificial intelligence to take over what used to be human tasks. For example an algorithm could soon be the one deciding whether or not someone is qualified for social assistance or financial aid. Needless to say this change from human to automated decision making is a change in procedures, and there is hence a compelling need to investigate how this in turn might affect trust and satisfaction among citizens. This study uses an experimental design with 187 respondents to examine the effect of automated decision making on procedural justice. The study was performed as a survey experiment and data was collected in Gothenburg, Sweden. Results of the study show a significant negative effect of automated decision making on the perceived ethicality and reliability, but a positive effect on the perceived efficiency and impartiality. Furthermore the study concluded that a decision made by an algorithm was perceived by the respondents as significantly less fair than a decision made by a social worker.

Nyckelord: automatiserat beslutsfattande, proceduriell rättvisa, digitalisering

Innehållsförteckning

1. Inledning 2

1.1 Forskningsproblemet 2

1.2 Disposition 3

2. Tidigare forskning och teori 3

2.1 Konkurrerande definitioner av Quality of Government 3

2.2 Förtroende och betydelsen av proceduriell rättvisa 5

2.3 Vad utgör proceduriell rättvisa? 6

2.4 Algoritmer och artificiell intelligens inom offentlig förvaltning 8

2.5 Attityder till automatiserat beslutsfattande 10

2.6 Syfte och frågeställning 11

3. Hypoteser 13

4. Metod 15

4.1 Val av metod 15

4.2 Design 16

4.3 Urval och distribution av enkät 17

4.4 Beroende variabel och mekanismer 18

4.5 Behandling 19

4.6 Randomisering 20

5. Etiska överväganden 22

6. Kontroll och behandling av data 22

6.1 Randomiseringskontroll 23

6.2 Urvalets representativitet 24

6.3 Index för mekanismer 24

7. Resultat 25

7.1 Test av hypoteser och samband 25

7.2 Test av sambandets interaktion med kön och ålder 31

8. Slutsats och diskussion 32

9. Referenslista 35

10. Appendix 40

1. Inledning

1.1 Forskningsproblemet

Vi står just nu inför en dramatisk förändring av den offentliga sektorn. Enorma teknologiska genomslag de senaste decennierna har öppnat upp för möjligheten att låta algoritmer och artificiell intelligens ta över mänskliga arbetsuppgifter och det handlar inte längre enbart om maskiner som beslutsstöd utan även om maskiner som faktiska beslutsfattare. En utredning som gjorts på uppdrag av Sveriges regering rekommenderar att det nu ska göras lagligt att automatisera kommunala beslut (Finansdepartementet 2021). Detta innebär att en algoritm mycket snart kan vara det verktyg som hanterar din eller någon annans ansökan om till exempel försörjningsstöd. Samtidigt framgår inte i någon större utsträckning av utredningen hur en sådan laglighet skulle se ut eller vilka krav som behöver ställas på algoritmernas utformning. Nu finns därför ett stort behov av att vetenskapligt ta sig an frågan om algoritmbaserat beslutsfattande och vidare att anpassa vår teoretiska och empiriska förståelse av vad som utgör en god förvaltning till en mer komplex verklighet.

> Technology-driven disruption is taking place at a pace and scale not witnessed before in history. […] Many of the existing government structures and processes that have evolved over the last few centuries will likely become irrelevant in the near future. There is a compelling need to lay the groundwork for governments to rethink how they will be able to best serve their constituents. (Agarwal 2018, 917)

Två återkommande argument för den pågående teknologiseringen och automatiseringen är dels att utvecklingen anses ge förutsättningar för en mer effektiv offentlig förvaltning, dels att en algoritm tros kunna leverera mer tillförlitliga och opartiska beslut än en människa (Cordella & Tempini 2015). Samtidigt, på andra sidan om debatten, finner man dem som menar att detta riskerar ske på bekostnad av förvaltningens legitimitet och förtroende (Wirtz et al. 2020). Många av de offentliga beslut som kan komma att tas över av datorprogrammerade algoritmer bygger på individuella avvägningar och berör dessutom människor i högst utsatta situationer. Flera forskare frågar sig därför nu om en algoritmerna kan hantera sådan komplexitet, och vidare om det över huvud taget är etiskt att låta den försöka (Svensson 2020).

Huruvida det är rimligt att tro att teknologiseringen kan fortgå utan att hota förtroendet för våra myndigheter beror i någon mening på vilka kvaliteter man väljer att tillskriva konceptet god förvaltning. Förtroende mellan medborgarna och samhällets institutioner har många gånger om kunnat konstateras vara av yttersta vikt för ett välfungerande samhälle, såväl ekonomiskt som socialt (OECD 2017), men medan somliga menar att det viktigaste för att uppnå detta förtroende är att myndigheterna är effektiva och levererar resultat menar andra att det viktigaste är att myndighetsbeslut och beslutsprocedurer kan betraktas som rättvisa. Vidare förknippas rättvisa procedurer ibland med opartiskhet och saklighet, och andra gånger med transparent beslutsfattande väglett av moraliska principer (Rothstein 2018; Agnafors 2013). Oavsett vad man tror står emellertid två saker klara. Automatiseringen är en utveckling som pågår i detta nu och den kommer att innebära proceduriella förändringar för offentligt beslutsfattande. Mot bakgrund av detta ämnar följande studie att med ett experiment ta oss ett steg närmare dels en förståelse av huruvida automatiserat beslutsfattande faktiskt påverkar hur människor uppfattar ett myndighetsbeslut, dels hur människor värderar de egenskaper som förknippas med en algoritm som beslutsfattare.

1.2 Disposition

Direkt nedan följer en redogörelse av tidigare forskning och teori. Detta avsnitt behandlar för forskningsproblemet relevanta teorier och presenterar därefter det nuvarande forskningsläget om automatiserat beslutsfattande inom offentlig förvaltning med huvudsakligt fokus på en svensk kontext. Mot bakgrund av teori och tidigare forskning introduceras sedan studiens syfte, frågeställning och i nästföljande avsnitt dess hypoteser. Därefter följer ett avsnitt om metod, ett avsnitt om etiska överväganden samt ett avsnitt om kontroll och behandling av data. I uppsatsens resultatdel redovisas sedan den statistiska analysen och uppsatsen avlutas med en slutsats och diskussion där även framtida forskningsmöjligheter tas upp.

2. Tidigare forskning och teori

2.1 Konkurrerande definitioner av Quality of Government

Diskussionen om betydelsen av en god förvaltning, som kännetecknas av det nu väletablerade begreppet *Quality of Government*, tog ordentlig fart i samband med att flera oberoende studier uppmärksammade att välfungerande samhällsinstitutioner tycks spela en avgörande roll för ekonomisk tillväxt och social välfärd (Knack & Keefer 1995; Mauro 1995). Det är därför inte

förvånande att många tidiga definitioner av Quality of Government har utgått från effektivitet som grundpremiss. En effektiv förvaltning anses helt enkelt vara en god förvaltning. I en definition av Quality of Government presenterad av La Porta et al. (1999, 222) inkluderas exempelvis effektivt tillhandahållande av kollektiva nyttigheter och effektiv förvaltning av resurser som indikatorer. Kritiker menar dock att effektivitet som definition av Quality of Government saknar en normativ grund och dessutom inte tillräckligt väl fångar upp vad en god förvaltning faktiskt är.

Några som har uttryckt just detta är Bo Rothstein och Jan Teorell (2008). De förespråkar istället en definition av Quality of Government som opartiskt myndighetsutövande och menar att effektivitet snarare kan förstås som ett resultat av detta. Exempelvis har meritokratisk rekrytering visat sig ha en stor positiv inverkan på staters byråkratiska effektivitet (Rothstein & Teorell 2008, 183). Vidare motiverar Rothstein och Teorell normativt sin definition av Quality of Government som opartiskhet genom att placera fenomenet i en större kontext av politisk legitimitet. Enligt Rothstein och Teorell (2008, 170) avgörs en stats legitimitet av två huvudfaktorer, dels tillgång till politisk makt och dels hur den politiska makten utövas. En vedertagen förståelse av legitimitet vad gäller tillgång till politisk makt är demokratiskt folkstyre. Demokratier kan organiseras på en rad olika sätt, med parlamentarism eller presidentsystem och med olika typer av valsystem, men anses ändå legitima så länge de vägleds av politisk jämlikhet. På samma sätt, menar Rothstein och Teorell, kan en förvaltning organiseras på en rad olika sätt men bör för att anses politiskt legitim vägledas av opartiskhet. Med andra ord vägledas av universalism, saklighet och pålitlighet.

Idealet om administrativ opartiskhet går att spåra tillbaka till den kanske mest centrala gestalten inom byråkratisk teori, Max Weber. Webers (1922) byråkratiska ideal bygger på specialisering, meritokratisk rekrytering, tydliga hierarkier och opartiskt myndighetsutövande väglett av universella regler. Han drar helt enkelt en skarp skiljelinje mellan den politiska nivån och den administrativa nivån. Förvaltningen ska tjäna som ett verktyg för att förverkliga de politiska målen men ingen faktisk politik ska bedrivas utanför den politiska sfären. Detta ideal har dock kommit att ifrågasättas av dem som menar att en opartisk byråkrati i realiteten är omöjlig att uppnå eftersom en människa varken kan eller bör förväntas vara fullständigt neutral (Mendus 2002; Young 1993). ”No one can adopt a point of view that is completely impersonal and dispassionate, completely separated from any particular context and commitments” (Young 1993, 127).

Också Rothstein och Teorell medger att det är svårt att kräva fullständig opartiskhet av en människa men understryker samtidigt att deras krav på opartiskhet enbart gäller i tjänstepersonens yrkesutövande och vidhåller att detta är avgörande för en god förvaltning (2008, 174). En ytterligare fråga är dock om opartiskhet som vägledande princip ensamt kan förväntas föra med sig de många positiva samhällskonsekvenser som vi förknippar med Quality of Government. Marcus Agnafors (2013, 436-441) menar att så inte är fallet, konceptet kräver en mer komplex definition. För att skapa en god relation mellan medborgare och samhällets institutioner krävs utöver opartiskt myndighetsutövande också rättssäkerhet och ett regelrätt, transparent beslutsfattande väglett av moraliska principer. Agnafors lägger dessutom vikt vid det faktiska mötet mellan medborgare och förvaltning och understryker att myndighetsutövande inte ska bygga på envägskommunikation.

> This requirement also emphasizes the interaction between citizens and public agents – that governance is not merely a unilateral exercise of public authority. Public agents must perform their obligations for the right reasons and they must allow us to see what they do. (Agnafors 2013, 439).

2.2 Förtroende och betydelsen av proceduriell rättvisa

Någonting som Agnafors (2013) och Rothstein och Teorell (2008) däremot tycks vara fullständigt eniga om är att en offentlig förvaltning inte kan anses fungera väl utan att inneha medborgarnas förtroende. Rothstein och Teorell inleder sin text med en berättelse ur verkligheten. Alldeles intill flygplatsen på den tropiska ön Saint Lucia ligger två nedgångna baracker där mat och dryck serveras. Trots deras fantastiska läge alldeles vid stranden och en mängd potentiella kunder inom räckhåll på flygplatsen ekar stolarna tomma. Kvinnorna som arbetar i barackerna har varken vågat investera i ordentliga möbler eller i att köpa loss marken barackerna står på. De räknar med att mutorna de då hade behövt betala för att klara hälsoinspektionen hade varit långt högre än vad de har råd med. Rothstein och Teorell (2008, 183) använder berättelsen om kvinnorna i Saint Lucia för att illustrera betydelsen av förtroende för förvaltningen och dess anställda. Hade kvinnorna kunnat lita på opartiskhet hos de institutioner som utövar politiskt delegerad makt hade de kanske vågat satsa på sina verksamheter, vilket i förlängningen med största sannolikhet hade lett till en förbättrad ekonomisk situation såväl för dem som för landet i stort. I samhällen som domineras av korruption och partiska samhällsstrukturer hämmas istället den ekonomiska och sociala utvecklingen (Rothstein & Teorell 2008, 166). Rothstein utvecklar dessa idéer ytterligare i sitt kapitel i boken *Politik som organisation* (2018), där han beskriver hur människor som interagerar med myndigheter

inte enbart bryr sig om slutresultatet. Det är minst lika viktigt för det institutionella förtroendet att den process som leder fram till det offentliga beslutet uppfattas som rättvis (Rothstein 2018, 49).

Dessa argument bygger på socialpsykologiska tankegångar om så kallad proceduriell rättvisa. Proceduriell rättvisa handlar i breda drag om att förstå varför människor följer lagar som tillämpas av olika myndigheter, också när de går emot deras intressen (Lind & Tyler 1988). Förespråkare av betydelsen av proceduriell rättvisa menar att när myndighetsutövningen uppfattas som rättvis upplever människor inte enbart ett ökat förtroende för myndigheter, utan tenderar dessutom i större utsträckning att betrakta myndighetsbeslut som rättvisa också när de går emot deras intressen. Bland de första att driva denna tes var John Thibaut och Laurens Walker (1975). Deras idéer uppfattades då av många som kontraintuitiva men efterföljande forskning har, i en rad kontexter, kommit att ge starkt och utbrett stöd för betydelsen av proceduriell rättvisa (Folger & Cropanzano 1998; Tyler & Blader 2000). Tesen brukar kallas ”the procedural justice effect” vilket alltså syftar till den effekt beslutsprocedurer har på människors nöjdhet och uppfattning om myndighetsbeslutet som rättvist. Somliga menar till och med att procedurer är av större betydelse än det faktiska utfallet för hur människor bedömer ett myndighetsbeslut.

> My studies of people's evaluations of legal and political authorities indicate that both performance evaluations and attitudes about the legitimacy of legal authorities are linked to evaluations of the fairness of the procedures used by authorities to make decisions and determine policies. This procedural justice effect suggests that the roots of public support for law and government lie in citizen assessments of the manner in which government authorities exercise their authority. (Tyler 2001, 233)

Att människor upplever förvaltningsbeslut som rättvisa ökar i förlängningen deras institutionella förtroende vilket har visat sig ha starkt positiva konsekvenser för samhället i stort, såväl ekonomiskt som socialt (OECD 2017; Rothstein & Eek 2009), och de procedurer som leder fram till besluten tycks alltså spela en avgörande roll. Men vad menar människor egentligen när de säger att en procedur är rättvis eller orättvis? Vilka funktioner skapar detta viktiga förtroende?

2.3 Vad utgör proceduriell rättvisa?

En av de första att rikta forskningsvärldens uppmärksamhet mot dessa frågor var Gerard Leventhal (1980). Enligt Leventhal är människor långt ifrån enkelspåriga i sina bedömningar av beslutsprocedurer. En rättvis beslutsfattare ska vara lika delar hänsynsfull och moralisk som

opartisk och konsekvent. Vidare vill man se att varje individ blir lyssnad på, samtidigt som misstag och felbedömningar minimeras. Med utgångspunkt i Leventhals teorier genomförde Tom R. Tyler (1988) en intervjustudie med 652 personer som nyligen varit i kontakt med polismyndigheter eller rättsväsendet. Han kom dels, i enlighet med tidigare studier, fram till att proceduriell rättvisa spelar roll för utvärdering och tillfredsställelse, men nöjde sig inte med att enbart fastställa effekten av proceduriell rättvisa utan undersökte vidare hur medborgare definierar rättvisa procedurer. Resultaten var komplexa och visade att faktorer som påverkar rättviseuppfattningen sträcker sig från beslutskvalitet och etikalitet till bedömning av beslutsfattarens ärlighet och förmåga att undertrycka partiskhet (1988, 128-132). Proceduriell rättvisa tycks alltså vara långt ifrån okomplicerat att precisera, men fyra komponenter som återkommer hos såväl Leventhal (1980) och Tyler (1988) som i senare bidrag till forskningsområdet (Blader & Tyler 2003; Rothstein & Eek 2009; Schafer 2013) är opartiskhet tillförlitlighet, etikalitet och transparens.

Opartiskhet syftar till beslutsfattarens förmåga att agera objektivt och sakligt. Bo Rothstein och Daniel Eek (2009) lät, i en experimentstudie på mellanmänskligt förtroende och proceduriell rättvisa, behandlingsgruppen läsa om en situation där beslutsfattaren tagit emot en muta. De kunde konstatera att förekomsten av korrupta handlingar påverkade såväl respondenternas förtroende för myndigheten som deras uppfattningar om myndighetsbeslutet negativt. Viktigt att notera är dock att opartiskhet omfattar mer än enbart förekomsten av mutor och andra korrupta beteenden. Till opartiskt myndighetsutövande räknas också bland annat fördomsfrihet och frånvaro av diskriminering. Exempelvis operationaliserar Tyler (1988, 112) opartiskhet genom att fråga sina respondenter huruvida någon form av diskriminering hade förekommit i deras möte med polisen eller rättsväsendet. *Tillförlitlighet* i sin tur syftar helt enkelt till myndighetens förmåga att leverera tillförlitliga beslut. Det vill säga undvika fel och misstag. Enligt Leventhals (1980, 27-28) definition av tillförlitlighet eller ”accuracy” ökar den proceduriella rättvisan om beslutsfattaren innehar expertkunskap och myndigheten kan intyga att besluten bygger på korrekt information.

Etikalitet är ett mindre konkret koncept men innebär i breda drag en uppfattning om myndighetsutövningen som etiskt och moraliskt försvarbar. Tyler (1988, 113) operationaliserar i sin intervjustudie etikalitet med hjälp av två frågor. Dels huruvida myndigheterna har varit artiga och hänsynsfulla i sitt bemötande, dels huruvida myndigheterna har visat respekt för individens medborgerliga rättigheter. Leventhal (1980, 33) å andra sidan beskriver etikalitet i mer abstrakta termer, som någonting subjektivt format av individuella uppfattningar. Detta sätt att tänka går att

likna vid socialpsykologen Jonathan Haidts idéer om etik och moral. Enligt Haidt (2012) kommer uppfattningar om vad som är moraliskt rätt eller fel ur en magkänsla snarare än ett rationellt tänkande. Man bör därför inte söka universella regler för och definitioner av etik och moral. ”Understanding the simple fact that morality differs around the world, and even within societies, is the first step toward understanding your righteous mind” (Haidt 2012, 4). Slutligen, *transparens*, som är ett relativt omfattande koncept. Transparent myndighetsutövande kräver dels att individen har insyn i beslutsprocessen och de regler som ligger till grund för beslutet, dels att en förklaring till beslutet erhålles (Schafer 2013, 131). Blader och Tyler (2003, 757) operationaliserar bland annat goda procedurer med påståendet ”The rules require that I get an honest explanation for how decisions are made”.

2.4 Algoritmer och artificiell intelligens inom offentlig förvaltning

Algoritmer har funnits länge men i och med den teknologiska utvecklingen har dess betydelse och närvaro i samhället ökat exponentiellt. I denna studie syftar algoritmer till en programvara eller kodade regler som ersätter eller hjälper mänskliga arbetsuppgifter genom att samla in data och utifrån den generera ny data, antingen som beslutsstöd eller som ett faktiskt beslut. Steget från algoritm till artificiell intelligens (AI) görs när programvaran själv, med hjälp av stora mängder information, lär sig nya saker och vidare skapar egna beslutsregler (Agarwal 2018; Akademikerförbundet SSR 2021). Som läget ser ut i Sverige idag är AI enligt denna definition inte en aktuell del av den offentliga förvaltningens arbete i någon större utsträckning, även om den teknologiska utvecklingen ser ut att kunna vara påväg i en sådan riktning (Regeringskansliet 2018). Algoritmer som beslutsstöd och beslutsfattare däremot är alltså ett växande fenomen, det vill säga att datorer istället för handläggare fattar beslut i ärenden som berör medborgarna. Detta fenomen brukar kallas för *automatiserat beslutsfattande* och har förekommit hos statliga myndigheter sedan flera år tillbaka. Redan 2006 automatiserades exempelvis handläggningen av tillfällig föräldrapenning hos Försäkringskassan (Finansdepartementet 2021, 83).

Inom kommunal förvaltning har utvecklingen gått något långsammare (Svensson 2020). I en rapport från 2019 framkommer att 16 av Sveriges 290 kommuner har infört digital automatiserad handläggning av försörjningsstöd (Svensson 2020, 343). Försörjningsstöd, tidigare kallat socialbidrag, är en selektiv välfärdsrättighet tänkt att fungera som ett yttersta skyddsnät i det universella välfärdssystemet och kommunerna är skyldiga att individuellt pröva rätten till bistånd

för samtliga som söker. Under 1990-talet ökade den administrativa och ekonomiska belastningen på försörjningsstödet, då antalet bidragstagare blev fler och dessutom sökte stöd för längre tidsperioder. Handläggningsprocessen för försörjningsstöd har sedan dess blivit mer formaliserad och processen har i viss mån visat sig passa för automatiserat beslutsfattande och beslutsstöd (Svensson 2020, 342). Samtidigt framhåller Socialstyrelsen, i en uppföljning av den nationella digitala strategin, att det ännu inte finns lagstöd för ett helt automatiserat beslutsfattande inom socialtjänstens område (Socialstyrelsen 2019). I och med Finansdepartementets utredning från i fjol, som gjordes på uppdrag av Sveriges regering och som rekommenderar att det nu ska göras lagligt att automatisera kommunala beslut, togs dock ett betydande steg mot en legalisering (2021). Dessutom har ett par kommuner valt att gå händelserna i förväg. Ett exempel är det uppmärksammade fallet Trelleborg, där man alltså valt att låta en datorprogrammerad algoritm ta över handläggningen av ärenden om försörjningsstöd i kommunen (Trelleborgs Allehanda 2017).

Algoritmer och AI är inte ett nytt fenomen inom den samhällsvetenskapliga forskningen (Matthias 2004; Lin, Bekey & Abney 2008), men de frågor som rör hur den nya teknologin förhåller sig till staten och mer specifikt hur den kan komma att användas inom förvaltningen är ännu relativt outforskade (Wirtz et al. 2020, 818). En förklaring till bristen på tillfredsställande litteratur, utöver det faktum att algoritmer och AI inom förvaltningen är en relativt ny fråga, är fenomenets ytterst komplexa natur. Det ställer helt nya krav på vår förståelse av såväl etik och rättvisa som förvaltningsarbetets själva utformning. Den litteratur som i dagsläget finns på området drar därför i flera olika riktningar, såväl till fördel för teknologiseringen som till dess nackdel. Ett flertal studier stödjer exempelvis resultat som visar att automatiserat beslutsfattande i offentlig förvaltning ger ett mer neutralt beslutsfattande, minimerar risken för korruption och ökar tillförlitligheten (Ranerup & Henriksen 2019; Cordella & Tempini 2015) Vidare är ett återkommande argument för en automatisering att effektiviteten ökar. I en intervjustudie med handläggare inom socialtjänsten konstaterar Lupita Svensson att:

> Automatisering är en effektivitetshöjande insats, i den bemärkelsen att mindre resurser används för att utföra mera arbete. Utifrån formell rättssäkerhet och de regler som i dag styr handläggningen av försörjningsstöd är det högst sannolikt att en robot också gör ett mer rättssäkert arbete, det vill säga att reglerna tillämpas lika. (Svensson 2020, 358).

Samtidigt är resultaten från Svenssons studie inte entydigt positiva. Hon beskriver nämligen också hur effektiviseringen kan riskera ske på bekostnad av försörjningsstödets funktion som selektiv välfärdsrättighet. Handläggning av försörjningsstöd bygger på individuella avvägningar och berör dessutom människor i utsatta situationer. Handläggare som Svensson talat med känner oro inför huruvida en datorprogrammerad algoritm ska kunna hantera sådan komplexitet (2020, 359). Just detta är en vanligt återkommande invändning mot automatiserat beslutsfattande. Berndt Wirtz, Jan Weyerer och Benjamin Sturm (2020 821-822) frågar sig hur vi kan få algoritmer och AI att ta hänsyn till etiska normer och konstaterar att det är mycket är svårt att förutse alla de etiska dilemman som kan uppstå i en beslutsprocess för att sedan programmera en algoritm därefter. Vidare ligger angränsande utmaningar enligt Wirtz et al. (2020) dels i risken för att en algoritm innehåller inbyggda bias, det vill säga att algoritmen speglar fördomar eller partiskhet hos den som ligger bakom programmeringen, dels att det är svårt att få god insyn i automatiserat beslutsfattande.

2.5 Attityder till automatiserat beslutsfattande

Även om det råder delade meningar kring exakt hur, kan vi alltså konstatera att den pågående teknologiseringen och automatiseringen av förvaltningen kommer att förändra våra offentliga beslutsprocesser på mer än ett sätt. Vi vet också att de processer som leder fram till myndighetsbeslut påverkar huruvida besluten uppfattas som rättvisa och i förlängningen det institutionella förtroendet hos medborgarna (Tyler 1988; Tyler 2001). Trots detta har anmärkningsvärt få studier tagit sig an att studera hur människor faktiskt ställer sig till den pågående förändringen. Sedan 2019 har vi dock en något bättre bild av de svenska medborgarnas attityder till automatiserat beslutsfattande. I SOM-undersökningen 2019 ställdes nämligen frågor till respondenterna om deras inställning till fenomenet och i ett kapitel i 2019 års SOM-bok presenteras och analyseras resultaten av Thomas Denk, Karin Hedström och Fredrik Karlsson.

De konstaterar att endast en minoritet av medborgarna, 20 procent för att vara precis, har kännedom sedan tidigare om att kommuner och myndigheter i allt större utsträckning låter beslut som rör medborgarna fattas av datorer istället för tjänstepersoner. Vidare visar studien att en majoritet tror att automatiseringen gör beslutsfattandet mer opartiskt, men inte att besluten blir mer tillförlitliga. En majoritet tror också att en dator som beslutsfattare tar mindre hänsyn till individens situation och dessutom att den ger sämre insyn i beslutsfattandet (Denk, Hedström & Karlsson 2019, 183). Vidare visar studien att kännedom om automatiserat beslutsfattande sedan tidigare ser ut att variera inom befolkningen. Exempelvis har äldre mindre kännedom, liksom personer som bor i mindre tätort och

på landsbygd. Detta är intressant eftersom de personer som uppger sig ha tidigare kännedom om automatiserat beslutsfattande tycks vara mer positivt inställda till fenomenet i vissa avseenden. Bland dem anser nämligen en majoritet att automatiserat beslutsfattande ger mer tillförlitliga beslut och dessutom anser en större majoritet än bland dem som saknar tidigare kännedom att besluten blir mer opartiska. Vad gäller hänsyn till individens behov och insyn i beslutsfattandet finns däremot ingen signifikant skillnad mellan de två grupperna (Denk, Hedström & Karlsson 2019, 187-191).

Bland den svenska befolkningen tycks alltså inställningen till automatiserat beslutsfattande varken vara entydigt positiv eller negativ. Denk, Hedström och Karlsson (2019) frågar sig därför avslutningsvis hur medborgarna kan tänkas värdera de egenskaper som de uppfattar hos en dator som beslutsfattande jämfört med en tjänsteperson och konstaterar att en framtida forskningsuppgift nu handlar om att undersöka relationen mellan attityder till automatiserat beslutsfattande och förtroendet för offentlig förvaltning.

> Medborgarna kan även uppfatta att automatiserat beslutsfattande ger såväl fördelar som nackdelar. Vad som då kommer att påverka förtroendet för offentlig förvaltning är hur medborgarna väger fördelarna med automatiserat beslutsfattande i förhållande till dess nackdelar. Det finns därmed anledning att följa utvecklingen och återkomma till frågor om hur medborgarna uppfattar och värderar automatiserat beslutsfattande. (Denk, Hedström & Karlsson 2019, 194)

2.6 Syfte och frågeställning

Den pågående automatiseringen av offentligt beslutsfattande sätter nytt ljus på frågor om god förvaltning, förtroende och rättvisa men också på de teorier som under lång tid utarbetats för att besvara dessa frågor. Enligt Rothstein och Teorell (2008) bör en god förvaltning vägledas av opartiskhet och den ideala tjänstepersonen vara fullständigt opåverkad av fördomar och yttre omständigheter. Någonting som de samtidigt medger är mycket svårt att kräva av en människa. Har vad som tidigare ansetts vara omöjligt nu blivit möjligt? En fullständigt opartisk tjänsteperson.

Om man lyckas programmera en algoritm som tar hänsyn till alla avgörande aspekter av ett myndighetsbeslut och samtidigt undviker inbyggda bias har man kommit mycket nära. Detta är inte att säga att Rothstein och Teorell skulle föredra en förvaltning där tjänstepersonerna är utbytta mot algoritmer, men det går inte att förneka att man med en dator som beslutsfattare undviker mänsklig godtycklighet och fördomsfullhet. Vilket leder oss tillbaka till Agnafors (2013) fråga. Är opartiskhet

nog? Enligt teorierna om proceduriell rättvisa tar människor hänsyn till såväl opartiskhet och tillförlitlighet som etikalitet och transparens när de bedömer ett myndighetsbeslut (Levanthal 1980; Tyler 1988) och resultaten från SOM-undersökningen 2019 antyder att människor förväntar sig att algoritmbaserade beslut blir mer opartiska men också att besluten tar mindre hänsyn till individens situation och skapar mindre insyn i beslutsfattandet. Så vad innebär då detta? Kommer det institutionella förtroendet att kunna upprätthållas allteftersom teknologiseringen fortgår? Kommer det rent av att stärkas? För att närma sig ett svar på dessa frågor behövs först undersökas huruvida automatiserat beslutsfattande faktiskt har en kausal effekt på de proceduriella egenskaper som är av betydelse för rättviseuppfattningen och vidare hur människor värderar dessa egenskaper. Denna studie syftar, mot bakgrund av detta, till att i en svensk samtida kontext av automatiserat förvaltningsarbete bidra till förståelsen av proceduriell rättvisa genom att med ett experiment söka svar på frågeställningen *Påverkas proceduriell rättvisa av automatiserat beslutsfattande?*

Mer konkret avser jag undersöka huruvida automatiserat beslutsfattande om försörjningsstöd påverkar uppfattningen om opartiskheten, tillförlitligheten, etikaliteten, transparensen och effektiviteten i en beslutsprocess. Det vill säga studera fenomenets effekt dels på de komponenter som av Leventhal (1980) och Tyler (1988) konstaterats påverka rättviseuppfattningen, dels på effektiviteten som har varit det kanske mest framträdande argumentet för en automatisering. Detta är särskilt intressant att göra eftersom tidigare studier antyder att automatiserat beslutsfattande påverkar vissa proceduriella egenskaper negativt och andra positivt. I vissa avseenden tros automatiseringen medföra bättre procedurer, i andra inte. Vidare avser jag undersöka huruvida ett beslut fattat av en datorprogrammerad algoritm uppfattas som mer eller mindre rättvist än ett beslut fattat av en människa. Detta kan nämligen ge en indikation på hur hur medborgarna värderar de egenskaper som de uppfattar hos en dator som beslutsfattare. Studien kommer emellertid inte att med ett experiment göra anspråk på att belägga hela orsakskedjan från beslut till mekanismer till rättviseuppfattning. Givet forskningsläget behöver vi i ett första steg se huruvida automatiserat beslutsfattande över huvud taget har en kausal effekt på den proceduriella rättvisan och på de egenskaper som av tidigare forskning konstaterats utgöra viktiga komponenter till konceptet.

Valet att studera fenomenet i en svensk kontext bygger på frågans aktualitet i landet. Att i Sverige mötas av en dator som beslutsfattare när man exempelvis vänder sig till kommunen för ekonomiskt bistånd är inte en teoretisk situation utan en reell möjlighet. Dessutom arbetar Sverige mot målet att ”vara bäst i världen på att använda digitaliseringens möjligheter” (Regeringskansliet 2018).

3. Hypoteser

Tidigare forskning har kunnat konstatera att automatiserat beslutsfattande fungerar mer neutralt och opartiskt (Ranerup & Henriksen 2019; Cordella & Tempini 2015). Enligt Svensson (2020) gör en robot högst sannolikt ett mer rättssäkert arbete, det vill säga tillämpar reglerna mer lika. Dessa slutsatser tycks dessutom, enligt Denk, Hedström och Karlsson (2019), överensstämma med svenskarnas attityder till frågan. Mot bakgrund av detta lyder studiens första hypotes:

Hypotes (1): *Personer som läser om ett myndighetsbeslut fattat av en algoritm kommer att uppfatta beslutet som* ***mer opartiskt*** *än personer som läser om ett beslut fattat av en socialsekreterare.*

Enligt Denk, Hedström och Karlsson (2019) tror inte en majoritet av den svenska befolkningen att automatiserat beslutsfattande ger mer tillförlitliga beslut, däremot tror en majoritet av de personer som hade kännedom om fenomenet sedan tidigare detta. I och med att användandet av datorprogrammerade algoritmer som beslutsfattare blivit vanligare (Finansdepartementet 2021) de senaste åren och att fenomenet tagit större plats i samhällsdebatten (Karlberg 2021; Erkers & Vinge 2020) finns det därför skäl att tro att fler skulle betrakta en dator som beslutsfattare som mer tillförlitlig idag än 2019. Dessutom argumenterar bland annat Antonio Cordella och Niccolò Tempini (2015) för att automatiserat beslutsfattande genererar mer tillförlitliga beslut. Mot bakgrund av detta lyder studiens andra hypotes:

Hypotes (2): *Personer som läser om ett myndighetsbeslut fattat av en algoritm kommer att uppfatta beslutet som* ***mer tillförlitligt*** *än personer som läser om ett beslut fattat av en socialsekreterare.*

Flera tidigare studier har lyft de etiska utmaningar som kommer med automatiserat beslutsfattande (Svensson 2020; Wirtz et al. 2020), bland annat hur det är mycket svårt att förutse alla de etiska dilemman som kan uppstå i en beslutsprocess för att sedan programmera en algoritm därefter. Vidare ingår i Tylers (1988) operationalisering av etikalitet bland annat ett hänsynsfullt bemötande, någonting som svenskar enligt SOM-undersökningen 2019 tror att en dator gör sämre än en mänsklig handläggare. Mot bakgrund av detta lyder studiens tredje hypotes:

Hypotes (3): *Personer som läser om ett myndighetsbeslut fattat av en algoritm kommer att uppfatta beslutet som* ***mindre etiskt*** *än personer som läser om ett beslut fattat av en socialsekreterare.*

Bland annat Wirtz et al. (2020) lyfter brist på transparens som ett av de stora problemen med automatiserat beslutsfattande. Det är enligt dem svårare att med en algoritm veta exakt vilka avvägningar som gjorts i en beslutsprocess. Dessa slutsatser tycks dessutom, enligt Denk, Hedström och Karlsson (2019), överensstämma med svenskarnas attityder till frågan. Mot bakgrund av detta lyder studiens fjärde hypotes:

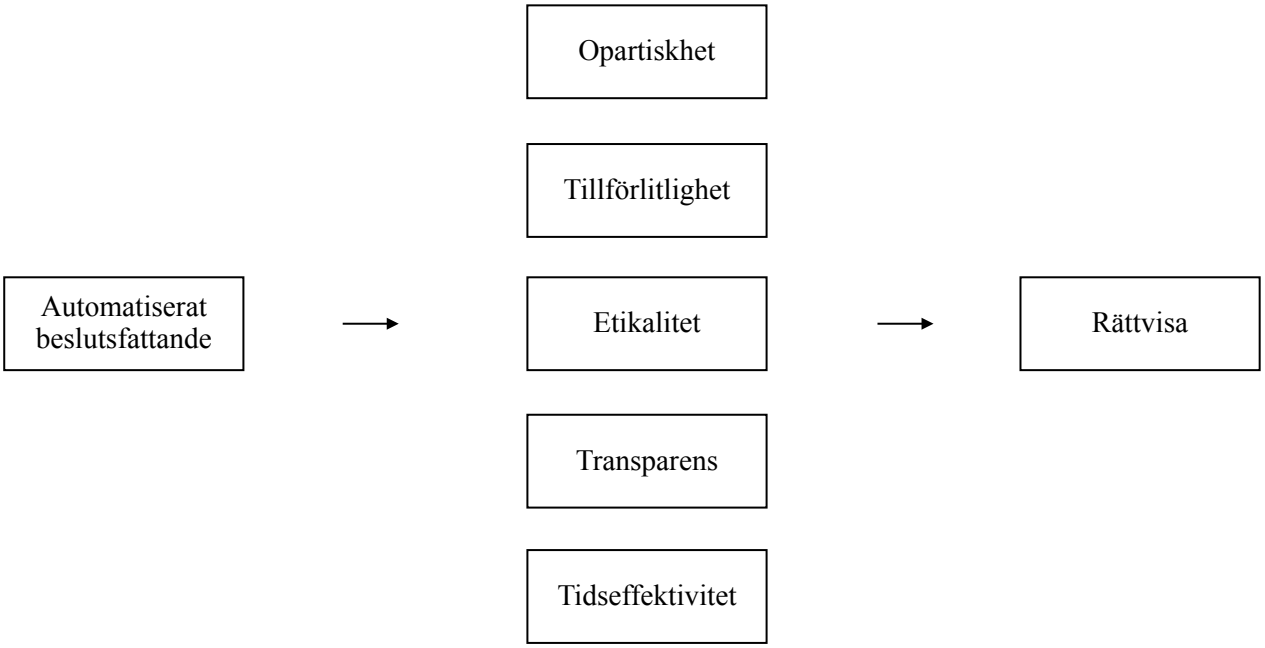
Hypotes (4): *Personer som läser om ett myndighetsbeslut fattat av en algoritm kommer att uppfatta beslutsprocessen som* ***mindre transparent*** *än personer som läser om ett beslut fattat av en socialsekreterare.*

Svensson (2020) konstaterar att automatisering i mångt och mycket är en effektivitetshöjande insats. Mindre resurser används för att utföra mera arbete och flera av de handläggare som intervjuas i hennes studie beskriver hur mer tid har frigjorts. Detta är dessutom ett av de mest framträdande argumenten för den pågående automatiseringen av förvaltningens arbete (Finansdepartementet 2021). Mot bakgrund av detta lyder studiens femte hypotes:

Hypotes (5): *Personer som läser om ett myndighetsbeslut fattat av en algoritm kommer att uppfatta beslutsprocessen som* ***mer tidseffektiv*** *än personer som läser om ett beslut fattat av en socialsekreterare.*

Tidigare forskning har kunnat konstatera att proceduriell rättvisa fungerar som ett samlande koncept för goda procedurer som skapar ökad nöjdhet (Tyler 2001). Resultatet av en studie som undersöker automatiserat beslutsfattandes effekt på rättviseuppfattningen skulle därför kunna tolkas som en indikation på hur medborgarna värderar de egenskaper som de uppfattar hos en dator som beslutsfattare. I och med att tidigare forskning, som beskrivet ovan, antyder att automatiserat beslutsfattande påverkar vissa proceduriella egenskaper negativt och andra positivt kommer jag emellertid inte att hypotisera riktningen på ett eventuellt kausalt samband mellan automatiserat beslutsfattande och proceduriell rättvisa utan istället behandla detta som en öppen fråga.

Figur 1: Teoretisk modell



4. Metod

4.1 Val av metod

I likhet med flera tidigare studier på proceduriell rättvisa (van den Bos et al. 1997; van den Bos et al. 2014) används en experimentell design för att besvara frågeställningen. Fördelen med ett experiment är att det erbjuder goda möjligheter att dra slutsatser om kausalitet, det vill säga orsak och verkan, då den oberoende variabeln tillåts isoleras och manipuleras (Esaiasson et al. 2017, 93). I detta fall indelas respondenterna slumpmässigt i en kontrollgrupp respektive behandlingsgrupp, där den ena får läsa om en situation där ett myndighetsbeslut har fattats av en socialsekreterare och den andra om en situation där ett myndighetsbeslut har fattats av en datorprogrammerad algoritm. I övrigt är den text som läses av behandlingsgrupp och kontrollgrupp i alla avseenden lika. På så vis kan eventuella skillnader i utfallsvariablerna hänvisas till den enda systematiska skillnaden grupperna emellan, behandlingen. I detta fall beslutsfattaren.

Ett ytterligare skäl att välja en experimentell design är att studien behandlar ett relativt nytt fenomen. Lind och Tyler (1988) konstaterar i en diskussion om lämpliga metoder för att studera proceduriell rättvisa att det bästa sättet att dra meningsfulla slutsatser om svårobserverade fenomen, såsom attityder och åsikter, är att testa sina hypoteser med flera olika metoder.

> The diversity of methods allows research to compensate for the weakness inherent in any one method with the strengths of other methods. Where the findings of studies using different methods converge we can be certain that the phenomena discovered are not artifacts of any particular research method. (Lind & Tyler 1988, 41)

Med Denk, Hedström och Karlssons (2019) kvantitativa analys av SOM-undersökningen 2019 togs ett första steg mot att förstå hur attityder till den offentliga förvaltningens arbete och beslutsfattande kan komma att förändras i takt med att automatiserat beslutsfattande får en allt större roll. Till min kännedom har inga efterföljande studier av experimentell karaktär gjorts för att pröva deras resultat och undersöka huruvida automatiserat beslutsfattande faktiskt har en kausal effekt på huruvida ett myndighetsbeslut betraktas som opartiskt, tillförlitligt, transparent och etiskt. Vidare kan ett experiment ta sig an frågan om hur medborgarna värderar en dator som beslutsfattare jämfört med en tjänsteperson. I och med att tidigare forskning har kunnat konstatera att proceduriell rättvisa fungerar som ett samlande koncept för goda procedurer som skapar ökad nöjdhet (Tyler 2001) kan resultatet av en studie som undersöker automatiserat beslutsfattandes effekt på rättviseuppfattningen tolkas som en första indikation på hur hur medborgarna värderar de egenskaper som de uppfattar hos en dator som beslutsfattare. Med det sagt kan jag inte med denna design empiriskt belägga att orsakskedjan ser ut som teorin säger, det vill säga att automatiserat beslutsfattande för med sig en rad proceduriella egenskaper som i sin tur påverkar den proceduriella rättvisan, och inte heller hur olika nivåer av dessa proceduriella egenskaper skulle kunna påverka hur medborgarna värderar beslutet. Exempelvis skulle man kunna tänka sig att ett automatiserat myndighetsbeslut med en kort förklaring uppfattas som mer rättvis än ett konventionellt myndighetsbeslut med en mer omfattande förklaring men med två veckor längre handläggningstid. Inom ramen för denna kandidatuppsats, och utifrån det nuvarande kunskapsläget, anser jag det emellertid lämpligt att börja i denna ände.

4.2 Design

Många studier av proceduriell rättvisa bygger på experiment där människor ombeds förställa sig att de befinner sig i en situation och vidare hur de skulle reagera i en sådan. Lind och Tyler (1988, 47) konstaterar dock att denna typ av experiment kan skapa validitetsproblem. Det är nämligen svårt att veta huruvida respondenternas föreställning faktiskt överensstämmer med hur de hade agerat eller reagerat i verkligheten. Jag försöker undvika denna fallgrop genom att istället låta den text som utgör behandling och kontroll vara utformad likt en nyhetsartikel. Respondenterna får alltså ta

ställning till hur de upplever ett myndighetsbeslut som drabbat någon annan. Denna typ av design är vanlig bland annat i psykologisk forskning om så kallad priming (Valentino 1999). Där har den använts för att visa hur man med tillspetsad information eller värdeord kan styra människors tankar i en viss riktning, men i grunden syftar den till att kunna isolera och manipulera den oberoende variabeln vilket är just vad jag önskar göra.

Studien genomförs som ett surveyexperiment, det vill säga med en enkät som besvaras av respondenterna (Esaiasson et al 2017, 240). Enkäten är utformad så att respondenten först i en kort text får information om studien och försäkras anonymitet, därefter får den läsa en kort nyhetsartikel som för behandlingsgruppen innehåller behandlingen. Efter att ha läst nyhetsartikeln får respondenten besvara ett antal kryssfrågor. Dessa frågor utgör mätindikatorer genom vilka studiens beroende variabler operationaliseras. Därefter ställs tre kontrollfrågor och slutligen, efter att enkäten återlämnats, får respondenten ta del av mer ingående information om studiens syfte. Experimentet har alltså en endast-efterdesign vilket innebär att mätningen av de beroende variablerna görs endast efter behandlingen. En nackdel med endast-efterdesign är att man inte med fullständig säkerhet kan veta att en skillnad mellan grupperna har orsakats av behandlingen, risken för att skillnaden är ett resultat av slumpen kvarstår. Med en före-efterdesign eliminerar man detta problem och kan med än större säkerhet uttala sig om kausaliteten i sambandet, men risken med en sådan design är å andra sidan att den första mätningen påverkar respondenternas svar i någon riktning inför den andra mätningen (Teorell & Svensson 2007, 75).

4.3 Urval och distribution av enkät

Målpopulationen för denna studie är Sveriges befolkning. Målet är alltså att utifrån studiens resultat i så stor utsträckning som möjligt kunna uttala sig om den svenska befolkningen, vilket ställer krav på att urvalet ska vara så representativt som möjligt (Teorell & Svensson 2007, 85). För att uppnå detta distribueras enkäten i fält i Göteborg. Mer specifikt på bibliotek, caféer, bussterminaler samt på pendeltåg mellan Göteborg och Alingsås. Förhoppningen är att detta ska ge ett mer representativt urval än vad en digital enkät som sprids med hjälp av självselektion på sociala medier har potential att göra. Risken med det andra alternativet är nämligen, i denna studies fall, en överrepresentation av framförallt unga människor i och med att urvalet då i stor utsträckning baseras på den egna bekantskapskretsen. En brist som kvarstår vad gäller representativitet i urvalet är emellertid att majoriteten av de som svarar lär vara bosatta i stad eller storstad.

4.4 Beroende variabel och mekanismer

Likt flertalet tidigare studier på proceduriell rättvisa ombeds respondenterna ange hur rättvist de upplever ett beslut (Tyler 1988; van den Bos et al. 2014). Detta med frågan ”Hur rättvist upplever du beslutet?” där 4 = ”Mycket rättvist” och 1 = ”Inte alls rättvist”. I och med att en dator som beslutsfattare innebär en förändring i beslutsproceduren kan en skillnad i kontrollgrupp respektive behandlingsgrupp vad gäller rättviseuppfattning tolkas som en effekt av automatiserat beslutsfattande på proceduriell rättvisa.

De variabler som teoretiskt betraktas som mellanliggande variabler eller mekanismer i sambandet mellan automatiserat beslutsfattande och proceduriell rättvisa utgörs av opartiskhet, tillförlitlighet, etikalitet, transparens och tidseffektivitet. Rent empiriskt utgör de dock, i denna studie, experimentets utfallsvariabler och resultatet kommer inte att bekräfta den kausala kedjan. Vardera mekanism operationaliseras genom frågor i enkäten som samlas till ett litet index på två till tre frågor vardera. Index hjälper studiens reliabilitet, det vill säga förmåga att undvika osystematiska fel, i och med att risken att fel sak mäts på grund av att en fråga missförstås av respondenten minskar när flera frågor mäter samma fenomen (Esaiasson et al. 2017, 397) Detta kan vara särskilt viktigt när svårobserverade fenomen så som åsikter studeras. För att öka studiens validitet bygger frågorna på beprövade operationaliseringar från tidigare forskning. Om än omskrivna för att bättre passa studiens kontext och syfte.

Opartiskhet operationaliseras med hjälp av frågan ”Beslutet kan betraktas som opartiskt” (Denk, Hedström & Karlsson 2019) där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls” samt frågan ”Hur sannolikt tror du att det är att paret utsattes för någon form av diskriminering?” (Tyler 1988) där 4 = ”Mycket sannolikt” och 1 = ”Inte alls sannolikt”.

Tillförlitlighet operationaliseras med hjälp av frågan ”Beslutet kan betraktas som tillförlitligt” (Denk, Hedström & Karlsson 2019) där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls” samt frågan ”Hur sannolikt tror du att det är att någonting har blivit fel eller ett misstag har begåtts hos kommunen?” (Tyler 1988) där 4 = ”Mycket sannolikt” och 1 = ”Inte alls sannolikt”.

Etikalitet operationaliseras med hjälp av frågan ”Paret behandlas med respekt” (Blader & Tyler 2003) där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls” samt frågan ”Beslutsfattaren tar hänsyn till parets situation” (Tyler 1988) där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls”.

Enligt Haidts (2012) psykologiska teorier kan moraliska uppfattningar härledas till en magkänsla som är svår att operationalisera med andra frågor, därför ställs också helt enkelt frågan ”Beslutet fattas på ett etiskt och moraliskt sätt ” där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls”.

Transparens operationaliseras genom frågan ”Förklaringen som ges till beslutet är bristfällig” (Blader & Tyler 2003) där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls” samt frågan ”Beslutsprocessen är transparent” (Schafer 2013) där 4 = ”Instämmer helt” och 1 = ”Instämmer inte alls”.

Effektivitet operationaliseras med frågan ”Hur sannolikt tror du att det är att paret fick besked inom två veckor efter ansökan?” där 4 = ”Mycket sannolikt” och 1 = ”Inte alls sannolikt”.

Validitet avser hur väl man lyckas översätta begrepp från den teoretiska nivån till någonting som går att mäta. Med god begreppsvaliditet undviker man alltså systematiska fel i sina mätningar (Teorell & Svensson 2007, 56). Det faktum att algoritmer inom förvaltningen är en relativt ny fråga för många riskerar att innebära problem med just begreppsvaliditet för den beroende variabeln. Risken finns att en observerad negativ effekt av algoritmbaserat beslutsfattande på utfallsvariablerna i viss mån också speglar en mer generell skepsis eller oro för ny teknik. Detta behöver emellertid inte göra resultatet mindre intressant. Också det faktum att människor upplever en otrygghet på grund av okänd teknik kan i förlängningen påverka det institutionella förtroendet vilket många gånger om har konstaterats vara av största vikt för socialt och ekonomiskt välstånd.

4.5 Behandling

Som beskrivet ovan får både kontrollgrupp och behandlingsgrupp innan de besvarar enkäten läsa en tidningsartikel om ett myndighetsbeslut, närmare bestämt ett beslut om försörjningsstöd. Behandlingen utgör operationaliseringen av den oberoende variabeln, automatiserat beslutsfattande. I texten innebär detta att det på de ställen där beslutsfattaren omnämns står ”algoritm” eller ”datorprogrammerad algoritm” för behandlingsgruppen och ”socialsekreterare” för kontrollgruppen. I övrigt är texterna identiska för att på så vis kunna hänvisa all variation i de beroende variablerna till det enda som systematiskt skiljer de två grupperna åt, behandlingen. Det är därmed givetvis viktigt för studien att respondenterna faktiskt uppfattar att beslutet är fattat av en socialsekreterare respektive algoritm. Man vill helt enkelt i så stor utsträckning som möjligt undvika att vissa

respondenter läser nyhetstexten utan att lägga märke till vem som har fattat beslutet. Beslutsfattaren nämns därför i inledningen samt två gånger i texten.

Gruppen som får läsa om ett automatiserat beslut benämns alltså som behandlingsgrupp, men i och med att såväl behandlingsgrupp som kontrollgrupp läser en text innan de besvarar enkätfrågorna utsätts på sätt och vis båda två grupper för en behandling. En utmaning ligger därför i textens utformning. För att fånga upp en variation i mina utfallsvariabler är det viktigt att texten i övrigt är relativt neutral, eller snarare att den inte konstanthåller de fenomen jag är intresserad av att mäta. Skulle nyhetsartikeln exempelvis beskriva en situation där paret som söker försörjningsstöd helt uppenbart har blivit illa behandlade, diskriminerade mot eller där ett misstag har begåtts kan man inte förvänta sig någon särskilt stor variation i dessa utfallsvariabler och därmed inte heller någon signifikant skillnad grupperna emellan. Såväl behandlingsgrupp som kontrollgrupp skulle oberoende av beslutsfattare ange att etikaliteten, opartiskheten och tillförlitligheten var låg.

Nyhetsartikeln utgår från en text i Tranås Tidning (Molin 2021). För att uppfattas som trovärdig är artikelns utformning och typsnitt inspirerat av Dagens Nyheters digitala texter (se Appendix). Nyhetstexten beskriver ett par som sökt försörjningsstöd och fått avslag med hänvisning till att en av makarna inte sökt CSN för hela den period som hennes studier omfattade vilket gjorde att familjen gick miste om inkomst. Texten beskriver ett avslag eftersom detta tros ge tydligare resultat. Effekten av procedurer på bedömningar av beslut har nämligen visat sig vara större vid negativa utfall (van den Bos, van der Velden & Lind 2014, 12). I och med att respondenterna får veta att en förklaring har getts till beslutet konstanthålles delvis variabeln transparens. Detta är en avvägning som behövt göras för att ge respondenterna så goda förutsättningar som möjligt att besvara de efterföljande frågorna. Om texten innehåller så gott som ingen information om beslutet är det svårt att förvänta sig av respondenterna att ta ställning och risken är att det får konsekvenser dels på respondenternas villighet att delta (Esaiasson 2017, 248), dels på variationen i de övriga utfallsvariablerna. Dessutom är det intressant att undersöka möjligheten att transparensen upplevs olika grupperna emellan trots att de läser samma förklaring till beslutet.

4.6 Randomisering

Enkäterna delas ut i en randomiserad ordning som slumpas fram med hjälp av Stata. Avgörande för att en experimentstudie ska fylla sin funktion, att dra slutsatser om orsak och verkan, är nämligen att respondenterna indelas fullständigt slumpmässigt i behandlingsgrupp och kontrollgrupp. Skapar

man tillräckligt stora av slumpen sammansatta grupper kan man så gott som räkna med att dessa är identiska på aggregerad nivå, det vill säga som helhet lika trots att de består av olika individer, och då blir det enda som kan förklara en skillnad i den beroende variabeln mellan de två grupperna behandlingen (Teorell & Svensson 2007, 75).

För att säkerställa att randomiseringen har lyckats och utesluta andra förklaringar till skillnader i den beroende variabeln ställs därför tre kontrollfrågor till respondenterna i slutet av enkäten. Valet av vilka frågor som ska ingå är förankrat i litteraturen om faktorer som påverkar attityder till automatiserat beslutsfattande och teknisk utveckling mer generellt. För att få god validitet är samtliga frågor hämtade från SOM-undersökningen 2019. Den första frågan gäller kön och respondenterna får ange ”kvinna”, ”man” eller ”annat”. Denk, Hedström och Karlsson (2019, 187) har visserligen inte specifikt undersökt sambandet mellan kön och inställning till automatiserat beslutsfattande, men tidigare studier har kunnat visa på en större skepsis hos kvinnor gentemot ny teknik (Sáinz & López-Sáez 2010; Sundström & McCright 2016). Detta antas ha att göra med de roller som kvinnor och män socialiseras in i, bland annat hur kvinnor socialiseras in i mer altruistiska roller och tenderar vara mindre riskbenägna. Den andra frågan anger födelseår. Denk, Hedström och Karlsson (2019, 186) konstaterar nämligen en signifikant mindre kännedom om automatiserat beslutsfattande hos äldre. Ett ytterligare skäl att ta hänsyn till ålder är att sannolikheten att man har kommit i kontakt med ett selektivt välfärdssystem som försörjningsstöd lär öka med åldern, vilket man kan tänka sig påverkar hur respondenterna uppfattar behandlingen och besvarar de efterföljande frågorna.

Avslutningsvis får respondenterna svara på en fråga för att kontrollera huruvida de hade kännedom om automatiserat beslutsfattande inom förvaltningen sedan tidigare. Detta eftersom personer med kännedom om automatiserat beslutsfattande har visat sig ha en mer positiv inställning till fenomenet (Denk, Hedström & Karlsson 2019, 186). Skulle randomiseringen misslyckas och det finnas en betydande skillnad i kännedom mellan behandlingsgrupp och kontrollgrupp kan detta alltså skapa skillnader i den beroende variabeln som inte är resultat av behandlingen. Eftersom frågan ställs i slutet av enkäten, efter att respondenterna har tagit del av tidningsartikeln, finns dock en risk att viss skevhet uppstår grupperna emellan. De som precis har läst om en algoritm som fattat ett myndighetsbeslut skulle kunna tänkas vara mer benägna att ange att detta är ett fenomen de kände till sedan tidigare. Alternativet hade varit att ställa frågorna för randomiseringskontroll i början av enkäten men då är risken istället att dessa frågor påverkar hur respondenten därefter tar till sig

tidningsartikeln och besvarar frågorna. Tidigare studier, bland annat experimentstudier på så kallade ”racial cues”, har nämligen visat att en inledande fråga om respondentens fördomar neutraliserade effekten av den efterföljande behandlingen (Mendelberg 2009, 109). Mot bakgrund av detta väljer jag därför att ställa kontrollfrågorna sist, men det bör alltså noteras att en eventuell skillnad i angiven kunskap om automatiserat beslutsfattande mellan kontrollgrupp och behandlingsgrupp kan vara en effekt av behandlingen snarare än av en misslyckad randomisering.

5. Etiska överväganden

Syftet med en experimentell design är att manipulera den oberoende variabeln för att undersöka om det påverkar respondenterna i någon riktning. Detta är inte försvarbart i alla sammanhang, därför behöver vissa etiska överväganden göras (Teorell & Svensson 2007, 19-20). En viktig etisk riktlinje är samtyckeskravet. Respondenterna behöver ge sitt medgivande till att delta i studien och dessutom ha möjlighet att avbryta enkätundersökningen. En andra viktig riktlinje är informationskravet, respondenterna ska göras uppmärksamma på studiens syfte. För att inte påverka respondenternas svar kan det dock finnas en idé med att först efteråt informera dem om att de deltagit i ett experiment. Inledningsvis presenteras därför studien som ”en enkätundersökning som handlar om hur personer reagerar på olika nyhetsartiklar”. Efter att enkäten besvarats delges deltagarna mer utförlig information, ”Denna studie syftar mer specifikt till att jämföra hur folk reagerar på att en algoritm fattar ett myndighetsbeslut jämfört med att en människa fattar ett myndighetsbeslut. Du hamnade slumpmässigt i en av grupperna.” Det görs också tydligt för respondenterna att deras svar är anonyma och endast kommer att presenteras som del av statistiska resultat i en kandidatuppsats. Avslutningsvis är det viktigt, av etiska skäl, att informera respondenterna om att den nyhetsartikel de ska läsa är en fingerad text, det vill säga något omskriven.

6. Kontroll och behandling av data

Allt som allt tillfrågades 273 personer att delta i studien. 70 personer avstod från att delta. Detta resulterade i 203 besvarade enkäter, 103 i kontrollgruppen och 100 i behandlingsgruppen. Det totala antalet användbara observationer var dock 187, detta efter att 16 ofärdiga svar tagits bort, 11 från kontrollgruppen och 5 från behandlingsgruppen. Datan insamlades mellan datumen 2021-12-01 och 2021-12-22.

6.1 Randomiseringskontroll

I tabell 1 presenteras medelvärdena av randomiseringskontrollerna i kontrollgrupp respektive experimentgrupp tillsammans med p-värdena för vardera jämförelse. P-värdena för jämförelsen mellan kön och för jämförelsen mellan tidigare kännedom är hämtade från en logistisk regression eftersom dessa är dikotoma utfallsvariabler och dess fördelning därmed inte är normal.

Tabell 1: Medelvärdet av randomiseringskontrollerna i vardera grupp

	Kvinna	Födelseår	Tidigare kännedom
Kontrollgrupp	.522	1981.902	.402
Behandlingsgrupp	.537	1980.916	.526
p-värde	0.836	0.666	0.090
N	187	187	187

Kommentar tabell 1: Kvinna är baserat på en dummyvariabel där kvinna = 1 och man = 0. Tidigare kännedom är en dummyvariabel för tidigare kännedom om automatiserat beslutsfattande 1 = ja och 0 = nej.

För samtliga randomiseringskontroller är medelvärdena relativt lika mellan grupperna. Andelen kvinnor i kontrollgruppen är 52,2% och i behandlingsgruppen 53,7%. Den genomsnittliga födelseåret i kontrollgruppen är 1982 och i behandlingsgruppen 1981 Störst skillnad mellan kontrollgrupp och behandlingsgrupp syns i frågan om tidigare kännedom om automatiserat beslutsfattande. 40,2% i kontrollgruppen anger att de har tidigare kännedom om fenomenet och motsvarande andel i behandlingsgruppen är 53,6%. I samtliga tre jämförelser är $p > 0,05$, det vill säga ingen av skillnaderna mellan grupperna är statistiskt signifikant. Detta innebär att grupperna kan anses vara homogena och att randomiseringen lyckats. Det bör dock noteras att p-värdet för jämförelsen mellan tidigare kännedom i kontrollgrupp och behandlingsgrupp inte är större än 0,1. Sannolikheten att skillnaden är ett resultat av slumpen är därmed relativt låg om än inte tillräckligt för statistisk signifikans. Som konstateras ovan behöver detta dock inte vara tecken på en misslyckad randomisering utan kan snarare antas bero på att behandlingen haft viss effekt på hur kontrollfrågan besvarats.

6.2 Urvalets representativitet

För att se urvalets representativitet presenteras i tabell 2 datan från tabell 1 i en jämförelse med motsvarande siffror för hela den svenska populationen. Detta ger en bild av i vilken utsträckning det kan vara möjligt att dra mer generella slutsatser utifrån studiens resultat.

Tabell 2: Skillnad i medelvärden mellan urvalet och populationen nationellt

	Andelen kvinnor (%)	Ålder (födelseår)	Tidigare kännedom (%)
Urval	52,9	1981	46,5
Nationellt	49,7	1978	20
Differens	3,2	3	26,5

Enligt SCB (2021a) bestod Sveriges befolkning i december 2020 av 49,7% kvinnor och 50,3% män. Medelåldern för Sveriges befolkning i december 2020 var 41,4 år, vilket innebär att det genomsnittliga födelseåret var 1978 (SCB 2021b). För nationella siffror på frågan om tidigare kännedom om automatiserat beslutsfattande använder jag SOM-institutets data som presenterat av Denk, Hedström och Karlsson (2019, 187). Det ska dock sägas att en del har hänt i samhället sedan denna data samlades in som kan förklara den skillnad vi ser i kännedomen om automatiserat beslutsfattande i urvalet jämfört med populationen. Dels har användandet av datorprogrammerade algoritmer som beslutsfattare blivit vanligare (Finansdepartementet 2021), dels har fenomenet tagit större plats i samhällsdebatten (Karlberg 2021; Erkers & Vinge 2020). Anledningen till att mer uppdaterad data inte används är att detta inte finns att tillgå i dagsläget. En ytterligare anledning till skillnaden kan emellertid vara att studien är genomförd i Göteborg samt på pendeltåg mellan Göteborg och Alingsås. Enligt Denk, Hedström och Karlsson (2019, 187) var kännedomen om automatiserat beslutsfattande signifikant större bland personer som bor i storstad och stad jämfört med personer som bor i mindre tätort och på landsbygd. Detta är någonting som bör tas hänsyn till när mer generella slutsatser dras utifrån studiens resultat, men det går likväl sammanfattningsvis att konstatera att urvalets representativitet är relativt god. Både vad gäller kön och ålder ligger urvalet nära de nationella medelvärdena.

6.3 Index för mekanismer

Index skapas för de fyra mekanismerna opartiskhet, tillförlitlighet, etikalitet och transparens. Varje Index aggregerar svaren på de två eller tre frågor som används för att operationalisera respektive

mekanism. För att skapa dessa index har frågan om huruvida förklaringen är otillräcklig, frågan om sannolikheten för diskriminering samt frågan om sannolikheten för misstag kodats om för att riktningen på svaren för samtliga frågor ska vara lika.

Tabell 3: Index för respektive mekanism med Cronbachs Alpha

Mekanism	Namn på index	Cronbachs Alpha
Opartiskhet	opart_index	0.59
Tillförlitlighet	tillförlit_index	0.66
Etikalitet	etik_index	0.82
Transparens	transp_index	0.47

Med hjälp av Cronbachs Alpha kan man testa lämpligheten för variabler att ingå i ett index. Cronbachs Alpha går från lägst lämplighet 0 till höst lämplighet 1 och baseras på hur väl variablerna korrelerar med varandra. Den konventionella gränsen för lämplighet ligger på 0,7 (Field 2017, 823). Utifrån tabell 3 kan man alltså konstatera att endast index för etikalitet uppfyller lämplighetskravet. Resterande frågor för mekanismerna korrelerar alltså inte i den utsträckning som teoretiskt förväntades. I analysen används trots detta samtliga index som planerat men där presenteras även en regressionsanalys av skillnaden mellan grupperna för varje enskild fråga.

7. Resultat

7.1 Test av hypoteser och samband

Respondenterna har läst en i alla avseenden identisk nyhetsartikel med undantag för behandlingen, att beslutet har fattats av en datorprogrammerad algoritm eller i kontrollgruppens fall en socialsekreterare. Nedan undersöks huruvida detta skapar en signifikant skillnad mellan grupperna i hur de uppfattar de individuella proceduriella egenskaperna opartiskhet, tillförlitlighet, etikalitet, transparens och tidseffektivitet. Detta blir alltså test av studiens fem hypoteser. En första regressionsanalys görs för att testa Hypotes (1), sambandet mellan automatiserat beslutsfattande (behandling) och opartiskhet (opart_index). Jag använder Ordinary Least Squared regression med robusta standardfel. Robusta standardfel hanterar risken för att heteroskedastiska residualer i datan skulle skapa något för låga p-värden och något för smala konfidensintervall (Ruist 2021, 540).

Tabell 4: Regressionsanalys av skillnad mellan grupperna i opartiskhet

	(1) opart_index
behandling	0.184
	(1.84)
_cons	3.005***
	(43.90)
N	187
R^2	0.018

Kommentar tabell 4: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Regressionsanalysen i tabell 4 visar på ett positivt (**β** = 0,184) samband mellan automatiserat beslutsfattande och opartiskhet. Den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm, betraktade i genomsnitt beslutet som mer opartiskt (index 3,189 av 4) än den grupp som läste om ett beslut fattat av en socialsekreterare (index 3,005 av 4). Sambandet är däremot, med ett p-värde på 0,067, inte statistiskt signifikant. Konventionella regler för att betrakta ett samband som statistiskt signifikant kräver ett p-värde på maximalt 0,05 vilket innebär att sannolikheten är maximalt 5% att den observerade skillnaden mellan behandlingsgrupp och kontrollgrupp skulle vara ett resultat av slumpen (Teorell & Svensson 2007, 212). Ett p-värde på 0,067 ligger nära 0,05 och skulle kunna betraktas som marginellt signifikant men bör tolkas med försiktighet. Hypotes (1) får således visst stöd av resultatet.

En andra regressionsanalys görs för att testa Hypotes (2), sambandet mellan automatiserat beslutsfattande (behandling) och tillförlitlighet (tillförlit_index). Jag använder Ordinary Least Squared regression med robusta standardfel.

Tabell 5: Regressionsanalys av skillnad mellan grupperna i tillförlitlighet

	(1) tillförlit_index
behandling	-0.281**
	(-2.64)
_cons	2.891***
	(39.70)
N	187
R^2	0.036

Kommentar tabell 5: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Regressionsanalysen i tabell 5 visar på ett statistiskt signifikant ($p < 0,01$) negativt ($\beta = -0,281$) samband mellan automatiserat beslutsfattande och tillförlitlighet. Den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm, betraktade i genomsnitt beslutet som mindre tillförlitligt (index 2,61 av 4) än den grupp som läste om ett beslut fattat av en socialsekreterare (index 2,891 av 4). Hypotes (2) får således inte stöd av resultatet, istället tycks det omvända gälla mot hypotesens kausala antagande.

En tredje regressionsanalys görs för att testa Hypotes (3), sambandet mellan automatiserat beslutsfattande (behandling) och etikalitet (etik_index). Jag använder Ordinary Least Squared regression med robusta standardfel.

Tabell 6: Regressionsanalys av skillnad mellan grupperna i etikalitet

	(1) etik_index
behandling	-0.664***
	(-5.98)
_cons	2.801***
	(35.02)
N	187
R^2	0.162

Kommentar tabell 6: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Regressionsanalysen i tabell 6 visar på ett statistiskt signifikant ($p < 0{,}001$) negativt ($\beta = -0{,}664$) samband mellan automatiserat beslutsfattande och etikalitet. Den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm, betraktade i genomsnitt beslutet som mindre etiskt (index 2,137 av 4) än den grupp som läste om ett beslut fattat av en socialsekreterare (index 2,801 av 4). Hypotes (3) får således stöd av resultatet.

En fjärde regressionsanalys görs för att testa Hypotes (4), sambandet mellan automatiserat beslutsfattande (behandling) och transparens (transp_index). Jag använder Ordinary Least Squared regression med robusta standardfel.

Tabell 7: Regressionsanalys av skillnaden mellan grupperna i transparens

	(1) transp_index
behandling	-0.173
	(-1.55)
_cons	2.641***
	(32.25)
N	187
R^2	0.013

Kommentar tabell 7: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Regressionsanalysen i tabell 7 visar på ett negativt ($\beta = -0{,}173$) samband mellan automatiserat beslutsfattande och transparens. Sambandet är emellertid inte statistiskt signifikant, p-värdet är 0,122. Det går alltså inte att se en statistiskt signifikant skillnad mellan behandlingsgrupp och kontrollgrupp i hur transparent de upplever beslutet. Hypotes (4) får således inte stöd av resultatet.

En femte regressionsanalys görs för att testa Hypotes (5), sambandet mellan automatiserat beslutsfattande (behandling) och uppskattad tidseffektivitet (svarstid). Jag använder Ordinary Least Squared regression med robusta standardfel.

Tabell 8: Regressionsanalys av skillnaden mellan grupperna i tidseffektivitet

	(1) svarstid
behandling	0.475***
	(3.74)
_cons	2.304***
	(26.89)
N	187
R^2	0.070

Kommentar tabell 8: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Regressionsanalysen i tabell 8 visar på ett statistiskt signifikant (p < 0,001) positivt (**β** = 0,475) samband mellan automatiserat beslutsfattande och uppskattad tidseffektivitet. Den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm, såg det i genomsnitt som mer sannolikt att paret fick besked inom två veckor efter ansökan än den grupp som läste om ett beslut fattat av en socialsekreterare. 2,779 respektive 2,304 där 1 är ”inte alls sannolikt och 4 är ”mycket sannolikt”. Hypotes (5) får således stöd av resultatet.

Tabell 9: Regressionsanalys av skillnaden i varje enskild fråga för opartiskhet och tillförlitlighet

	(1) opartiskt	(2) diskriminering	(3) tillförlitligt	(4) misstag
behandling	0.0801	0.288**	-0.411**	-0.151
	(0.60)	(2.87)	(-3.15)	(-1.30)
_cons	3.130***	2.880***	3.011***	2.772***
	(34.63)	(41.84)	(34.18)	(34.45)
N	187	187	187	187

Kommentar tabell 9: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Tabell 10: Regressionsanalys av skillnaden i varje enskild fråga för etikalitet och transparens

	(1) respekt	(2) etikmoral	(3) hänsyn	(4) förklaring	(5) transparens
behandling	-0.571***	-0.764***	-0.657***	-0.215	-0.131
	(-4.48)	(-5.44)	(-5.12)	(-1.61)	(-0.92)
_cons	3.087***	2.848***	2.467***	2.467***	2.815***
	(36.68)	(29.92)	(25.10)	(26.09)	(29.58)
N	187	187	187	187	187

Kommentar tabell 10: Behandling kodas 1 (behandling) 0 (kontroll). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Tabell 9 och 10 visar resultatet av regressionsanalyser på sambandet mellan opartiskt myndighetsutövande och varje enskild fråga. Frågan om förklaringen var otillräcklig, samt frågorna om sannolikheten för misstag och diskriminering är likt i indexet omkodade i omvänd riktning för att stämma med riktningen på de andra frågorna. Utifrån resultatet i tabell 9 och 10 kan konstateras att den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm i genomsnitt såg det som mindre sannolikt att paret utsattes för någon form av diskriminering. Denna skillnad är statistiskt signifikant ($p < 0,01$). Däremot syns ingen signifikant skillnad mellan grupperna i frågan om huruvida beslutet kan betraktas som opartisk, även om behandlingsgruppen i genomsnitt höll med i något större utsträckning i påståendet än kontrollgruppen. Detta signifikanta samband med en enskilda fråga, som inte gav signifikanta resultat med sitt motsvarande index, bör dock tolkas med viss försiktighet i och med att reliabiliteten som beskrivet ovan ökar med index. Vidare går också att se att den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm i genomsnitt betraktade beslutet som signifikant ($p < 0,01$) mindre tillförlitligt. Däremot syns ingen signifikant skillnad mellan grupperna i frågan om hur sannolikt respondenten tror att det är att ett misstag har begåtts, även om behandlingsgruppen betraktade ett misstag som något mer sannolikt.

Avslutningsvis, för att testa huruvida behandlingen har skapat en signifikant skillnad i hur rättvist respondenterna uppfattar myndighetsbeslutet görs en regressionsanalys på sambandet mellan den oberoende variabeln automatiserat beslutsfattande (behandling) och den beroende variabeln rättviseuppfattning (rättvisa). Jag använder Ordinary Least Squared regression med robusta standardfel.

Tabell 11: Regressionsanalys av skillnad mellan grupperna i rättviseuppfattning

	(1) rättvisa
behandling	-0.370**
	(-3.11)
_cons	2.728***
	(30.23)
N	187
R^2	0.050

Kommentar tabell 11: Behandling kodas 1 (behandling) 0 (kontroll). Rättvisa kodas 1 (inte alls rättvist) till 4 (mycket rättvist). T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Regressionsanalysen i tabell 11 visar på ett statistiskt signifikant ($p < 0,01$) negativt ($\beta = -0,370$) samband mellan automatiserat beslutsfattande och rättviseuppfattning. Den grupp som läste om ett beslut fattat av en datorprogrammerad algoritm, svarade i genomsnitt 2,358 av 4 på frågan hur rättvist de upplever beslutet, där 4 var ”mycket rättvist” och 1 ”inte alls rättvist”. Den grupp som läste om ett beslut fattat av en socialsekreterare svarade i genomsnitt 2,728 av 4 på frågan.

7.2 Test av sambandets interaktion med kön och ålder

Tidigare forskning antyder att attityder till automatiserat beslutsfattande och ny teknologi mer generellt varierar med avseende på kön och ålder. Behandlingens effekt skulle därför kunna tänkas vara olika för olika värden på dessa variabler. För att testa detta görs två regressionsanalyser på sambandet mellan automatiserat beslutsfattande (behandling) och rättviseuppfattning (rättvisa) med kön respektive ålder som interaktionsvariabler. I mitt urval har jag emellertid en god spridning av åldrar med den äldsta respondenten född 1941 och den yngsta respondenten född 2004 samt en relativt jämn fördelning av kvinnor och män, så interaktionsanalysens resultat ska inte påverka studiens generaliserbarhet. Hade jag däremot exempelvis haft en stor majoritet äldre i studien och effekten av automatiserat beslutsfattande på rättviseuppfattning visat sig vara signifikant större hos denna grupp hade det minskat resultatets generaliserbarhet.

Tabell 12: Regressionsanalys med interaktion mellan kön och behandling

	(1) rättvisa
1.behandling	-0.273 (-1.57)
1.kön	0.220 (1.30)
1.behandling# 1.kön	-0.188 (-0.79)
_cons	2.614*** (21.33)
N	187
R^2	0.059

Kommentar till tabell 12: T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.

Tabell 13: Regressionsanalys med interaktion mellan ålder och behandling

	(1) rättvisa
1.behandling	-12.08 (-0.80)
födelseår	0.000805 (0.14)
1.behandling# c.födelseår	0.00591 (0.78)
_cons	1.133 (0.10)
N	187
R^2	0.058

*Kommentar till tabell 13: T-värde inom parantes, * p <0.05, ** p <0.01, *** p <0.001.*

Utifrån tabell 12 och13 kan jag konstatera att interaktionseffekterna inte är signifikanta, det vill säga effekten av behandlingen på rättviseuppfattningen är inte signifikant olika för kvinnor och män eller för olika åldrar. Detta tyder på att den negativa effekten av automatiserat beslutsfattande på rättviseuppfattning som påvisats är relativt allmängiltig.

8. Slutsats och diskussion

Den här studien har sin utgångspunkt i den förändring som offentlig sektor står inför i och med att datorprogrammerade algoritmer tillåts ta över vad som tidigare varit mänskliga arbetsuppgifter, en förändring som enligt vissa kommer att effektivisera förvaltningens arbete avsevärt och dessutom tillhandahålla mer tillförlitliga och opartiska beslut (Cordella & Tempini 2015), men som samtidigt enligt andra riskerar göra detta på bekostnad av medborgarnas förtroende och förvaltningens legitimitet (Wirtz et al. 2020; Svensson 2020). Utifrån studiens resultat kan man konstatera att svenskarnas inställning till automatiserat beslutsfattande varken är entydigt negativ eller entydigt positiv men också att Wirtz et al. (2020) och Svensson (2020) i viss mån har fog för sin oro.

Studien har med ett experiment studerat effekten av automatiserat beslutsfattande på proceduriell rättvisa. 187 respondenter indelades i en kontrollgrupp respektive behandlingsgrupp där den ena fick läsa om hur ett myndighetsbeslut fattats av en socialsekreterare och den andra om hur ett myndighetsbeslut fattats av en datorprogrammerad algoritm. Därefter fick respondenterna besvara frågor om deras uppfattning om beslutet och beslutsprocessen. Svaren på frågorna utgör studiens

utfallsvariabler. Den nyhetsartikel som lästes av behandlingsgrupp och kontrollgrupp var i alla avseenden lika, med undantag för behandlingen, och respondenterna indelades slumpmässigt i grupperna. Därmed kan alla eventuella skillnader i utfallsvariablerna hänvisas till den enda systematiska skillnaden grupperna emellan, behandlingen, det vill säga beslutsfattaren.

Den empiriska analysen ger sammantaget signifikant stöd för två av studiens fem hypoteser, och marginellt signifikant stöd för en tredje. Till att börja med uppfattade de personer som läste om ett beslut fattat av en algoritm i genomsnitt beslutet som mindre etiskt än de personer som läste om ett beslut fattat av en socialsekreterare. Där fanns en statistsik signifikant skillnad mellan grupperna för samtliga tre frågor som ställdes för att operationalisera etikalitet. Detta innebär stöd för Hypotes (3) och kan tolkas som att automatiserat beslutsfattande har en negativ effekt på de etiska aspekterna av en beslutsprocess. Vidare betraktade de personer som läste om ett myndighetsbeslut fattat av en algoritm i genomsnitt det som mer sannolikt att paret fick besked inom två veckor. Också denna skillnad mellan grupperna var statistiskt signifikant och ger stöd för Hypotes (5), personer som läser om ett myndighetsbeslut fattat av en algoritm kommer att uppfatta beslutsprocessen som mer tidseffektiv. Slutligen visar resultatet ett marginellt signifikant positivt samband mellan automatiserat beslutsfattande och opartiskhet. Gruppen som läste om ett beslut fattat av en algoritm såg det i genomsnitt som mindre sannolikt att paret utsattes för någon form av diskriminering. Däremot kunde ingen signifikant skillnad konstateras mellan grupperna i frågan om huruvida beslutet kan betraktas som opartiskt, även om behandlingsgruppen i genomsnitt höll med i något större utsträckning i påståendet än kontrollgruppen. Hypotes (1) fick således delvis stöd av resultatet.

Vad gäller effekten av automatiserat beslutsfattande på tillförlitligheten går sambandet i motsatt riktning mot vad som hypotiserades. Automatiserat beslutsfattande har alltså en signifikant negativ effekt på tillförlitligheten. Hypotesen byggde dels på tidigare forskning (Cordella & Tempini 2015) men också på tesen att en större kännedom om fenomenet idag än 2019 skulle medföra en mer positiv inställning, i och med att en majoritet av personer med kännedom om automatiserat beslutsfattande i SOM-undersökningen 2019 angav att de trodde att algoritmer genererar mer tillförlitliga beslut. En möjlig förklaring till att sambandet inte ser ut som förväntat, trots att kännedomen verkar ha ökat, är att de personer som hade kännedom om fenomenet redan för tre år sedan inte är representativa. Hos dem skulle exempelvis ett stort intresse för teknik kunna tänkas förklara både kännedomen om automatiserat beslutsfattande och en positiv inställning till

fenomenet. Avslutningsvis, för effekten av automatiserat beslutsfattande på transparensen fanns däremot ingen signifikans i någon riktning. Detta var ett delvis förväntat resultat i och med att det i nyhetstexten gavs en förklaring till beslutet, men det går alltså att konstatera att ingen signifikant effekt av automatiserat beslutsfattande på transparens kunde påvisas.

Så, *påverkas proceduriell rättvisa av automatiserat beslutsfattande?* Svaret på studies frågeställning är ja, på mer än ett sätt. De utfallsvariabler som har studerats utgår från komponenter som av tidigare forskning konstaterats vara viktiga för konceptet proceduriell rättvisa (Tyler 1988; Levanthal 1980) och sammanfattningsvis kan konstateras att automatiserat beslutsfattande har en positiv effekt på uppfattningen om opartiskhet och tidseffektivitet men en negativ effekt på uppfattningen om etikalitet och tillförlitlighet. Dessutom visar resultatet av studien att automatiserat beslutsfattande har en negativ effekt på hur rättvist beslutet uppfattas. För att återkomma till debatten mellan Rothstein och Teorell (2008) och Agnafors (2013) om vad som skapar en god förvaltning skulle detta resultat kunna tolkas som ett tecken på att människor faktiskt kräver mer än opartiskhet och saklighet för att känna förtroende för förvaltningen och dess beslutsfattande, i linje med Agnafors tes. I och med att tidigare forskning har kunnat konstatera att proceduriell rättvisa fungerar som ett samlande koncept för goda procedurer som skapar ökad nöjdhet och i förlängningen ökat förtroende (Tyler 2001) skulle nämligen den negativa effekten av automatiserat beslutsfattande på rättviseuppfattningen kunna tolkas som en första indikation på att människor värderar de proceduriella egenskaper som förknippas med mänskligt beslutsfattande högre än de egenskaper som förknippas med automatiserat beslutsfattande.

Samtidigt ger denna studie inga empiriska belägg för att de teoretiska mekanismerna faktiskt har kausal effekt på rättviseuppfattningen. Det skulle alltså kunna vara någonting annat som orsakar människor att uppleva automatiserat beslutsfattande som mindre rättvist. Här finns därför stort utrymme för framtida studier att ta vid. Med en faktoriell design i ett framtida experiment skulle man exempelvis ha möjlighet att säga någonting om vilka mekanismer som påverkar rättviseuppfattningen. En ytterligare fråga för framtida forskning är hur automatiserat beslutsfattande kan anpassas och utformas på ett sätt som möter medborgarnas krav och förväntningar på rättvist beslutsfattande. Min förhoppning är att resultaten av denna studie ska kunna inspirera ny forskning och tjäna som ett bidrag till det framtida arbetet för en ansvarsfull digitalisering.

9. Referenslista

Agarwal, P.K., 2018. Public Administration Challenges in the World of AI and Bots. *Public administration review*, 78(6), pp.917–921.

Agnafors, M., 2013. Quality of Government: Toward a More Complex Definition. *The American political science review*, 107(3), pp.433–445.

Akademikerförbundet SSR. 2021. *All makt åt algoritmen, vår befriare?* Stockholm: Akademikerförbundet SSR.

Blader, S.L. & Tyler, T.R., 2003. A Four-Component Model of Procedural Justice: Defining the Meaning of a "Fair" Process. *Personality & social psychology bulletin*, 29(6), pp.747–758.

Cordella, A. & Tempini, N., 2015. E-government and organizational change: Reappraising the role of ICT and bureaucracy in public service delivery. *Government information quarterly*, 32(3), pp.279–286.

Denk, T., Hedström, K & Karlsson, F. 2019. Medborgarna och automatiserat beslutsfattande. I Andersson, U., Rönnerstrand, B., Öhberg, P. & Bergström, A. (red) *Storm och stiltje*. Göteborgs Universitet: SOM-institutet.

Erkers, Heike & Vinge, Simon. 2020. "Obehörig algoritm tar beslut i socialtjänsten". *Svenska Dagbladet*. 18 januari. https://www.svd.se/obehorig-algoritm-tar-beslut-i-socialtjansten (Hämtad 2021-12-26).

Esaiasson, P. et al., 2017. Metodpraktikan : konsten att studera samhälle, individ och marknad Femte upplagan.

Field, A.P., 2018. *Discovering statistics using IBM SPSS statistics* 5. ed. London: Sage Publications.

Finansdepartementet. 2021. *En väl fungerande ordning för val och beslutsfattande i kommuner och regioner* (SOU 2021:16). Stockholm: Finansdepartementet.

Folger, R., & Cropanzano, R. 1998. Organizational justice and human resource management. Thousand Oaks, CA: Sage.

Karlberg, Erik. 2021. Svårt få insyn i algoritmerna som styr robotarna i jobbet. *Dagens Nyheter*. 31 januari. https://www.dn.se/ekonomi/svart-fa-insyn-i-algoritmerna-som-styr-robotarna-i-jobbet/ (Hämtad 2021-12-26).

Knack, S. & Keefer, P., 1995. Institutions and economic performance: Cross-country tests using alternative institutional measures. *Economics and politics*, 7(3), pp.207–227.

La Porta, R. et al., 1999. The quality of government. *Journal of law, economics, & organization*, 15(1), pp.222–279.

Leventhal, Gerald S. "What should be done with equity theory?." *Social exchange*. Springer, Boston, MA, 1980. 27-55.

Lin, P., Bekey G., & Abney K., 2008. Autonomous military robotics: Risk, ethics, and design. California Polytechnic State Univ San Luis Obispo.

Lind, E.A. & Tyler, T.R., 1988. The social psychology of procedural justice, New York ; London: Plenum.

Matthias, A., 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and information technology*, 6(3), pp.175–183.

Mauro, Paolo. 1995. "Corruption and Growth," 110 Quarterly Journal of Economic.

Mendus, Susan. 2002. Impartiality in Moral and Political Philosophy. Oxford: Oxford University Press.

OECD. 2017. *OECD Guidelines on Measuring Trust.* Paris: OECD Publishing.

Ranerup, A. & Henriksen, H.Z., 2019. Value positions viewed through the lens of automated decision-making: The case of social services. *Government information quarterly*, 36(4), p.101377.

Regeringskansliet. 2018. *Nationell inriktning för artificiell intelligens*. https://www.regeringen.se/informationsmaterial/2018/05/nationell-inriktning-for-artificiell-intelligens/ (Hämtad 2022-01-02).

Rothstein, B. & Eek, D., 2009. Political Corruption and Social Trust. *Rationality and society*, 21(1), pp.81–112.

Rothstein, B. & Teorell, J., 2008. What is Quality of Government: A Theory of Impartial Political Institutions. Governance. *International Journal Of Policy And Administration, 2008, Vol. 21, Iss. 2, pp. 165-.190*, 21(2), pp.165–190.

Rothstein, B. i Dahlström, C., 2018. Politik som organisation : förvaltningspolitikens grundproblem Sjätte upplagan. s. 35-56.

Sáinz, M. & López-Sáez, M., 2010. Gender differences in computer attitudes and the choice of technology-related occupations in a sample of secondary students in Spain. *Computers and education*, 54(2), pp.578–587.

SCB. 2021a. *Sveriges befolkning*. https://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/sveriges-befolkning/ (Hämtad 2021-12-27).

SCB. 2021b. *Befolkningens medelålder efter region och kön. År 1998-2020*. https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101B/BefolkningMedelAlder/ (Hämtad 2021-12-27).

Schafer, J.A., 2013. The role of trust and transparency in the pursuit of procedural and organisational justice. *Journal of policing, intelligence and counter terrorism*, 8(2), pp.131–143.

Socialstyrelsen. 2019. *E-hälsa och välfärdsteknik i kommunerna 2019*. Stockholm: Socialstyrelsen.

Sundström, A. & McCright, A., 2016. Women and Nuclear Energy: Examining the Gender Divide in Opposition to Nuclear Power among Swedish Citizens and Politicians. *Energy Research & Social Science, 2016, Vol. 11*, 11, pp.Energy Research & Social Science, 2016, Vol. 11.

Svensson, L., 2020. Automatisering – till nytta eller fördärv? *Socialvetenskaplig tidskrift*, 26(3-4), p.341.

Teorell, J. & Svensson, T., 2007. Att fråga och att svara : samhällsvetenskaplig metod 1. uppl.

Thibaut, J., & Walker, L. (1975). *Procedural justice*. Hillsdale, NJ: Lawrence Erlbaum.

Trelleborgs Allehanda. 2017. Robot fattar beslut om stöd. 7 april. https://www.trelleborgsallehanda.se/trelleborg/robot-fattar-beslut-om-stod/ (Hämtad 2021-11-25).

Tyler, Tom R. 2001. "A Psychological Perspective on the Legitimacy of Institutions and Authorities." i *The Psychology of Legitimacy*, redigerad av J. T. Jost och B. Major: Cambridge University Press.

Tyler, T. R., & Blader, S. L. 2000. Cooperation in groups: Procedural justice, social identity and behavioral engagement. Philadelphia: Psychology Press.

Valentino, N.A., 1999. Crime News and the Priming of Racial Attitudes during Evaluations of the President. *Public opinion quarterly*, 63(3), pp.293–320.

van den Bos, K., van der Velden, L. & Lind, E.A., 2014. On the role of perceived procedural justice in citizens' reactions to government decisions and the handling of conflicts. *Utrecht law review*, 10(4), pp.1–26.

van den Bos, K. et al., 1997. How Do I Judge My Outcome When I Do Not Know the Outcome of Others? The Psychology of the Fair Process Effect. *Journal of personality and social psychology*, 72(5), pp.1034–1046.

Weber, Max. 1922. ”Bureaucracy” I Economy and Society. Weber, Max (red) 956- 1005. Hämtad från https://archive.org/details/VOL12_201904/page/n1067/mode/2up

Wirtz, B.W., Weyerer, J.C. & Sturm, B.J., 2020. The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration. *International journal of public administration*, 43(9), pp.818–829.

Young, Iris M. 1990. Justice and the Politics of Difference. Princeton, NJ: Princeton University Press.

10. Appendix

10.1 Enkät

Enkätundersökning för kandidatarbete
Statsvetarprogrammet vid Göteborgs Universitet

Detta är en enkätundersökning som handlar om hur personer reagerar på olika nyhetsartiklar. Undersökningen tar inte mer än några minuter att genomföra och svaren är fullständigt anonyma, de kan inte spåras tillbaka till dig. Resultatet kommer endast presenteras som siffror i en tabell som del av en kandidatuppsats i statsvetenskap på Göteborgs Universitet.

Undersökningen består av tre sidor. På den första sidan kommer du att få läsa en något omskriven och nedkortad nyhetsartikel. På den andra sidan ombeds du därefter svara på ett antal korta frågor om den text du precis har läst och på den tredje sidan får du svara på tre korta avslutande frågor.

I vilken utsträckning instämmer du i följande påståenden?

	Instämmer helt	Instämmer delvis	Instämmer knappast	Instämmer inte alls
A. Paret behandlas med respekt	☐	☐	☐	☐
B. Beslutet kan betraktas som opartiskt	☐	☐	☐	☐
C. Beslutet kan betraktas som tillförlitligt	☐	☐	☐	☐
D. Beslutet fattas på ett etiskt och moraliskt sätt	☐	☐	☐	☐
E. Förklaringen som ges till beslutet är bristfällig	☐	☐	☐	☐
F. Beslutsfattaren tar hänsyn till parets situation	☐	☐	☐	☐
G. Beslutsprocessen är transparent	☐	☐	☐	☐

Hur rättvist upplever du beslutet?

Mycket rättvist	Ganska rättvist	Inte särskilt rättvist	Inte alls rättvist
☐	☐	☐	☐

Hur sannolikt tror du att det är att…

	Mycket sannolikt	Ganska sannolikt	Inte särskilt sannolikt	Inte alls sannolikt
A. …paret utsattes för någon form av diskriminering?	☐	☐	☐	☐
B. …någonting har blivit fel eller ett misstag har begåtts hos kommunen?	☐	☐	☐	☐
C. …paret fick besked inom två veckor efter ansökan?	☐	☐	☐	☐

Tre avslutande frågor:

Kön:

☐ Man

☐ Kvinna

☐ Annat

Vilket år är du född? _____

Kommuner och myndigheter låter allt fler beslut som rör medborgarna fattas av datorer istället för av tjänstemän. Känner du till sedan tidigare att en del beslut inom offentlig sektor fattas av datorer?

☐ Ja

☐ Nej

Stort tack för din medverkan!

Denna studie syftar mer specifikt till att jämföra hur folk reagerar på att en algoritm fattar ett myndighetsbeslut jämfört med att en människa fattar ett myndighetsbeslut. Du hamnade slumpmässigt i en av grupperna.

Att kommuner och myndigheter låter allt fler beslut som rör medborgarna fattas av datorer istället för av tjänstemän är alltså helt sant. Den nyhetsartikel du har läst är något omskriven och nedkortad men baserad på en text ur Tranås Tidning 2021.

Dina svar är fullständigt anonyma och resultatet kommer endast presenteras som siffror i en tabell som del av en kandidatuppsats i statsvetenskap på Göteborgs Universitet.

Om du har frågor får du gärna mejla: **guselsho@student.gu.se**

10.2 Behandling

SVERIGE

Par ansökte om försörjningsstöd från kommunen – algoritmen gav avslag

9 augusti 2021 17:03

Paret ansökte om ekonomiskt bidrag till hyra och uppehälle för maj månad 2021, samt för hemförsäkring och läkarvård. Ansökan behandlades av en datorprogrammerad algoritm som används för att hantera socialbidragsärenden på kommunens socialkontor. Efter en tid kom besked. Kommunen avslog ansökan.

Beslutet motiverades av algoritmen med att en av makarna inte hade ansökt om CSN-lån för hela den period som hennes studier omfattade vilket gjorde att familjen gick miste om inkomst. Makarna är inte nöjda och vill att beslutet ska omprövas med en fiktiv inkomst motsvarande den som vederbörande skulle ha fått om hon sökt studielån för hela perioden.

SVERIGE

Par ansökte om försörjningsstöd från kommunen – socialsekreteraren gav avslag

9 augusti 2021 17:03

Paret ansökte om ekonomiskt bidrag till hyra och uppehälle för maj månad 2021, samt för hemförsäkring och läkarvård. Ansökan behandlades av en socialsekreterare anställd på kommunens socialkontor. Efter en tid kom besked. Kommunen avslog ansökan.

Beslutet motiverades av socialsekreteraren med att en av makarna inte hade ansökt om CSN-lån för hela den period som hennes studier omfattade vilket gjorde att familjen gick miste om inkomst. Makarna är inte nöjda och vill att beslutet ska omprövas med en fiktiv inkomst motsvarande den som vederbörande skulle ha fått om hon sökt studielån för hela perioden.