



GÖTEBORGS
UNIVERSITET

SEGERSTEDTINSTITUTET

RAPPORT

RAPPORT 15

ONLINERADIKALISERING

En kartläggning av forskningsläget
och digitala dynamiker

Anton Törnberg

Göteborgs universitet
Segerstedtinstitutet

November 2025

Innehåll

Exekutiv sammanfattning.....	1
1. Introduktion.....	4
1.1 Syfte.....	6
1.2 Centrala utmaningar inom forskning om radikaliserings.....	7
1.3 Att definiera radikaliserings.....	9
2. Teoretiska perspektiv på radikaliserings.....	14
2.1 Strukturella faktorer.....	14
2.2 Individuella faktorer.....	15
2.3 Relationella och interaktionistiska perspektiv.....	17
2.4 Den nya generationen radikaliseringsforskning.....	19
3. Vilken roll har internet och sociala medier?.....	26
3.1 Individens utveckling mot ökad radikaliserings.....	27
3.1.1 Statistiska och experimentella studier av medieexponering.....	27
3.1.2 Fallstudier av radikaliserade individer.....	28
3.1.3 Uppluckring online/offline?.....	30
3.2 Digitala plattformar och mediemiljöer.....	31
3.2.1 Ekokammare och filterbubblor.....	32
3.2.2 Ekokammare eller skyttegravskrig?.....	33
3.2.3 Plattformsaffordanser.....	34
3.2.4 Engagemangsdrivna algoritmer och algoritmisk förstärkning.....	35
3.2.5 Ilskans ekosystem.....	36
3.2.6 Skillnader mellan plattformar.....	37
3.2.7 Gamification.....	38
3.3 Sammanfattning.....	39
4. Preventiva strategier.....	41
4.1 Innehållsreglering.....	41
4.2 Faktagranskning och motberättelser.....	42
4.3 Plattformsåtgärder: algoritmisk design och teknisk arkitektur.....	43
4.4 Digital resiliens och medieläskunnighet.....	44
5. Från datan till gatan: en syntesmodell.....	46
Mekanism 1: diskursiv normalisering och legitimering (makronivå).....	46
Mekanism 2: affektiv interaktion (mesonivå).....	47
Mekanism 3: instrumentell koordinering (mikro-/mesonivå).....	48

Exekutiv sammanfattning

Begreppet onlineradikalisering används för att beskriva en bred grupp av aktiviteter på nätet – från att ta del av extremistiskt material, att interagera med det och med andra som gör det, till att planera våldsdåd i slutna forum. Forskningen visar ett tydligt samband mellan digitala miljöer och radikalisering, särskilt när individer aktivt söker upp och interagerar med extremistiskt innehåll. Samtidigt är det viktigt att understryka att kontakt med sådant material i sig inte gör någon radikal. Internet fungerar snarare som en katalysator – en miljö som kan förstärka, fördjupa och påskynda redan pågående processer hos vissa individer. Nätet förändrar hur människor exponeras för och interagerar med våldsbejakande budskap, gemenskaper och berättelser. Denna förändring i våra sätt att kommunicera har sänkt tröskeln för interaktion och organisering – det som tidigare krävde planerade fysiska möten och hierarkiska organisationer kan nu ske med några knapptryck.

Radikalisering beskrivs ofta som en gradvis process där en person utvecklar radikala eller extremistiska åsikter (åsiktsradikalisering) och/eller börjar stödja eller använda våld (beteenderadikalisering). Forskningen om onlineradikalisering fokuserar främst på den första delen – hur attityder, idéer och övertygelser förändras. Det betyder att man intresserar sig för hur människor tolkar sina erfarenheter och känslomässigt förankrar dem i relation till sin sociala och politiska omgivning. Detta perspektiv kallas ofta den nya generationens radikaliseringsforskning och betonar att radikalisering inte bara handlar om ideologier eller individuella sårbarheter – riskfaktorer, utan också om människors behov av mening, tillhörighet och erkännande i ett visst samhälleligt sammanhang.

Det betyder dock inte att individuella faktorer saknar betydelse. Tvärtom visar forskningen att vissa riskfaktorer – som exponering för våldsbejakande budskap eller kopplingar till likasinnade – kan förstärka upplevda orättvisor eller känslan av att stå utanför. Det viktiga är att förstå att människor inte ses som passiva mottagare av propaganda, utan som aktiva meningsskapande individer som tolkar sina upplevelser i ljuset av berättelser som ger svar på frågan ”varför händer detta mig, och vem bär skulden?”. Sådana berättelser – exempelvis i form av konspirationsteorier, hypermaskulina ideal eller våldsbejakande ideologier – erbjuder enkla förklaringar och tydliga fiendebilder. Dessa kan i sin tur förstärka känslan av att det är rätt, eller till och med nödvändigt, att agera mot den eller det som uppfattas som hotet. Sett ur detta perspektiv utgör de digitala plattformarna, genom sin infrastruktur, en särskilt gynnsam miljö för radikalisering – en plats där berättelser om orättvisa tillstånd cirkulerar och får fäste, där känslor kopplade till dessa tillstånd ges extrema förklaringar och uttryck, och där förutsättningar finns för att ta steg mot att omsätta dessa känslor i handling.

En viktig del av denna infrastruktur handlar om de stora plattformarnas engagemangsdrevna algoritmer – system som lyfter fram känsloladdat, polariserande och konfliktinriktat innehåll.

Dessa algoritmer förändrar vilka budskap som får spridning och hur människor exponeras för dem. Genom att rekommendera grupper, konton eller videor driver algoritmerna användarens uppmärksamhet mot innehåll som väcker starka känslor, ofta mer extremt än det individen först sökte efter. I exempelvis TikToks *For You*-flöde anpassas innehållet snabbt efter vad användaren tittar eller klickar på. Om en användare visar intresse för konspirationsteorier, hypermaskulina ideal eller våldsbejakande budskap, kan flödet efter bara några timmar domineras av sådant innehåll.

Hur kan vi förstå vilka mekanismer som påverkar människor som radikaliseras online?

För personer som redan känner sig marginaliserade, orättvist behandlade eller hotade – som individer eller som del av en grupp – kan denna typ av algoritmiskt driven exponering bidra till att forma ett flöde med innehåll där extremism eller våld framstår som både logiska och legitima svar på upplevelsen av orättvisa – på frågor som ”varför händer detta mig, och vem bär skulden?”. Rekommendationssystemet leder användaren bort från etablerade berättelser till subkulturella och potentiellt extrema idévärldar, exempelvis konspirationsteorier. Denna förflyttning sker ofta gradvis och omärkligt för användaren, vilket resulterar i en normaliseringsspiral där idéer och budskap som tidigare uppfattades som extrema så småningom framstår som begripliga och rationella. Det här kan förstås som en **accelerationmekanism** för onlineradikalisering.

Internet och sociala medier är inte bara kanaler för kommunikation utan en miljö som människor lever i och där identitet, tillhörighet och känslor formas och omformas. Här möter deltagarna inte bara subkulturella och ibland extrema idévärldar – de bidrar också till att skapa gemenskaper med egna symboler, språk och normer. Genom dessa tolkar man sin omvärld och de berättelser som förklarar varför orättvisa tillstånd uppstår och vilka hot som riktas mot den egna gruppen. När människor i dessa gemenskaper samlas kring ett gemensamt fokus – som invandringskritik, kvinnohat eller olika konspirationsteorier – och delar känslor av ilska, frustration eller hat, uppstår en känslomässig samstämmighet som stärker både den gemensamma identiteten och övertygelsen om ett närvarande och påtagligt hot. När starka känslor delas och förstärks i en sådan gemenskap, kan en känslomässig spiral skapas där ilska och hat växer. Den algoritmiska logiken innebär att de mest extrema uttrycken får mest uppmärksamhet, vilket ytterligare eldar på denna dynamik. Resultatet blir att känslor, identitet och idéer binds samman på ett sätt som successivt gör det lättare att rättfärdiga våld. Om personen dessutom känner stöd från gruppen, kan våld börja framstå som både rimligt och nödvändigt. Detta kan förstås som en **mobiliseringsmekanism** för onlineradikalisering.

Internets olika miljöer – sociala medier och forum – möjliggör inte bara en accelererande spridning av idéer och mobilisering av känslor. Även möjligheten att agera på dessa idéer och känslor kan förstärkas genom plattformarnas struktur. Anonyma och oreglerade plattformar är i princip ”gränslösa arkiv” för kunskap om hur ett attentat kan planeras och genomföras. Dessutom kan anonyma och oreglerade plattformar fungera som platser för samordning av allt från nätmobbning till hot och rena sabotage – de kan till och med användas för att koordinera aktiviteter i realtid. Denna möjlighet fyller som regel en funktion i ett senare skede av onlineradikaliseringsprocessen, när mening, motivation och legitimitet för en attack redan

etablerats. Det här fungerar som en instrumentell **koordineringsmekanism** i online-radikalisering.

Som vi kan se av de tre mekanismerna karaktäriseras radikalisering online av andra dynamiker än de som präglar det fysiska livet. Trots det fyller dessa digitala dynamiker i betydande utsträckning samma mening och funktion – de erbjuder gemenskap, mening och motivation. Därför bör onlineradikalisering inte förstås som ett isolerat livsområde, utan som en del av människors vardagliga livsvärld, där det digitala och det fysiska livet samspelar och ömsesidigt formar varandra.

Avslutningsvis kan rapportens resultat ge tre viktiga medskick till praktiker:

1. Var uppmärksam på gradvisa förskjutningar i språk, världsbild och normer mot konspirationsteoretiska förklaringar, tydliga fiendebilder (t.ex. feminister, judar eller muslimer) och fascination för våld.
2. Bedöm exponering för extremistiskt material i sin kontext – vilken mening och funktion fyller det för individen?
3. Var medveten om att vissa anonyma och oreglerade plattformar i princip är gränslösa arkiv för kunskap som kan fungera som praktiskt stöd vid planering och genomförande av våld.

1. Introduktion

Under de senaste två decennierna har internet och sociala medier fått en alltmer central roll i radikaliseringsprocesser. Digitala miljöer används idag av ett brett spektrum av extremistiska grupper – från våldsbejakande jihadister till högerradikala aktivister, misogyna subkulturer och klimatförnekande konspirationsteoretiker. Gemensamt för dessa miljöer är att nätet fungerar som en arena för spridning, mobilisering och identitetsskapande, där ideologiska budskap kan förstärkas, normaliseras och i vissa fall bana väg för politiskt våld.

I västvärlden har denna utveckling varit särskilt tydlig inom den högerextrema miljön, som också utgör denna rapports huvudsakliga empiriska fokus. Under det senaste dryga decenniet har vi kunnat bevittna en tydlig våg av högerradikal mobilisering i stora delar av Europa och Nordamerika – dels genom parlamentariska framgångar för radikalnationalistiska partier i länder som Italien, Frankrike, Polen och USA, dels genom omfattande gatumobiliseringar och rasistiskt motiverade våldsdåd. Parallellt med detta har den högerradikala närvaron på nätet vuxit kraftigt genom ett omfattande digitalt ekosystem där högerextrema grupper och individer samordnar sig, rekryterar nya sympatisörer och sprider sina budskap. Nätverk av detta slag växer fram på etablerade sociala medier som YouTube, Reddit, X, TikTok och Facebook samt på spelplattformar som Discord – men också på mer nischade och extrema forum som Voat, Gab, 8chan och Stormfront. Samspelet mellan dessa plattformar är både dynamiskt och taktiskt. Användare rör sig ofta mellan olika delar av nätverken i takt med att reglerna förändras eller att innehåll censureras.

Efter Donald Trumps valseger 2016 införde flera stora plattformar hårdare modereringsregler, inklusive faktagranskning och censur, för att begränsa näthat och våldsbejakande extremism. Detta ledde till att många extremistiska grupper och användare migrerade till mer privata och mindre reglerade plattformar som Gab och Telegram, där de kunde verka mer ostört. I vissa fall ledde detta till än mer radikaliserade sätt att uttrycka sina åsikter än tidigare. Modereringsreglerna medförde också vad som kan beskrivas som en taktisk anpassning inom rörelsen, där användare i allt större utsträckning använder memes, kodord och subtil ironi som strategier för att kunna uttrycka sina budskap utan att fastna i plattformarnas filter.

Den senaste tidens avregleringar, såsom avvecklingen av faktagranskning på flera stora plattformar efter Trumps senaste valseger 2024, kan innebära att extremistiska rörelser återigen är på väg in i mer öppna, mainstream-forum.

Men det som sker i dessa digitala miljöer stannar inte där. Tvärtom tenderar digital extremism att "läcka" – att röra sig från nätet ut i samhället, där budskapen ibland får högst påtagliga konsekvenser. Här kan två särskilt tydliga uttryck för onlineradikalisering urskiljas: ensamagerande våldsdåd och kollektiv mobilisering.

Ensamagerande gärningspersoner

Under de senaste åren har vi sett en markant ökning av ensamagerande gärningspersoner, inom både jihadistiska och högerradikala miljöer. Enligt europeisk statistik utgörs numera en majoritet av dödliga högerextrema attacker i Europa av individer utan formell koppling till organisationer (Ravndal, 2016). I dessa fall sker radikaliseringsen ofta online. Enligt en kommande FBI-rapport har även gärningspersonernas genomsnittsålder sjunkit dramatiskt, från 38 till 28 år på bara några få år (McCaulley, 2023).

Ett av de mest uppmärksammade fallen på ensamagerande gärningsmän är Anders Behring Breivik, som före dåden i Oslo och på Utøya var aktiv på det högerextrema forumet Stormfront. Under användarnamnet *year2183* uttryckte han hat mot feminister, judar, korrupta medier och de ”folkförrädare” som enligt honom möjliggjorde muslimsk massinvandring. Strax före dådet mejlade han sitt manifest till andra medlemmar på forumet, och mejlet spreds sedan över internet. Breiviks dåd har sedan dess fått en närmast mytologisk status i vissa nätmiljöer, och det har inspirerat efterföljare. Ett av de tydligare exemplen är Brenton Tarrant, som 2019 dödade 51 personer i Christchurch, Nya Zeeland. I sitt manifest hyllade han Breivik och refererar till honom som en ideologisk föregångare. Tarrant hade också på sina vapen skrivit namn och citat från tidigare högerextrema attentatsmän, inklusive Breivik.

Genom manifest, estetik och strategi – livestreaming, memes, estetiserat våld – etablerades en formel för attacker som är både individuellt genomförda och kollektivt kodade. Det så kallade ensamagerandet är därmed sällan ensamt – det sker inom en kulturell och digital gemenskap där våldshandlingar beundras, imiteras och cirkuleras (Lindekilde et al., 2019; Malthaner et al., 2024).

Digitala mobiliseringar och kollektiva våldsdåd

Parallellt med dessa individuella gärningsdåd fungerar internet också som motor för kollektiva våldsytringar. Efter ett terrordåd mot en dansstudio i England sommaren 2024, där tre barn dödades, började ett rykte cirkulera på X. En högerextrem profil hävdade felaktigt att gärningsmannen var en muslimsk invandrare. Inlägget fick stor spridning och förstärktes av influerare som Andrew Tate och Elon Musk. Under de följande dagarna och veckorna utbröt kravaller, där byggnader sattes i brand och ett stort antal personer skadades.

Flera omfattande studier visar på direkta samband mellan digital aktivitet och politiskt våld. Exempelvis fann Müller & Schwarz (2020a) att omkring 10 procent av alla hatbrott i vissa tyska regioner kunde kopplas till aktivitet på Alternativ för Tysklands (AfD) Facebooksidor. I en annan studie (2020b) visade samma forskare att Trumps antimuslimska inlägg under valkampanjen 2016 sammanföll med en ökning av antimuslimska hatbrott. Forskning från Storbritannien har på liknande sätt visat tydliga geografiska och tidsmässiga samband mellan hatretorik online och fysiskt våld, särskilt i samband med uppmärksammade händelser som fungerar som utlösande faktorer (Arcila Calderón et al., 2024; Williams et al., 2020). I Sverige noterades ett liknande mönster under flyktningmottagarkrisen 2015–2016, då en våg av hatretorik i sociala medier följdes av en ökning av brandattacker mot asylboenden (Lundstedt et al., 2025; Wahlström & Törnberg, 2019).

Dessa exempel visar hur våldshandlingar kan uppstå i nätverksbaserade mediemiljöer, utan att det nödvändigtvis förekommer direkta uppmaningar eller någon centraliserad organisering. Detta fenomen benämns ibland som *stokastisk terrorism*, ett begrepp som fångar hur spridningen av ideologiskt motiverat hat i digitala miljöer – ofta från inflytelserika personer som politiker, medieprofiler eller influerare – kan skapa en atmosfär där vissa individer känner sig manade att agera. Det tycks inte vara enskilda budskap i sig som är avgörande, utan den sammantagna effekten av retorik, normförskjutningar och emotionell mobilisering. Det handlar om en slags statistiskt förutsägbar – men konkret oförutsägbar – våldsspiral: man vet inte vem, var eller exakt när något kommer att ske, men man vet med hög sannolikhet att någon kommer att agera. Denna process kan liknas vid en tryckkokarlogik, där retorikens affektiva och ideologiska laddning skapar en överhettad atmosfär som kan få våldsamma konsekvenser. På detta vis fungerar det digitala ekosystemet både som spridningsmekanism och som drivkraft för våld – även när ingen aktör uttryckligen har planerat eller beordrat ett dåd.

Sammantaget visar forskningen att digitala miljöer spelar en aktiv roll i att forma både individuella och kollektiva våldshandlingar. Trots att de flesta studier fokuserar på enskilda plattformar eller grupper, pekar resultaten mot tydliga samband av närmast kausal karaktär – det vill säga att det inte bara finns ett statistiskt samband, utan att sociala medier i flera fall verkar ha drivit på eller förstärkt specifika våldsamma handlingar. Det finns alltså fog för att anta att den sammanvägda effekten av det digitala ekosystemet är ännu starkare än vad enskilda studier visar.

Samtidigt är relationen mellan internet och politiskt våld inte linjär. Om internet i sig vore en drivande kraft bakom våld, borde våldet ha ökat i takt med digitaliseringen. Men så ser inte verkligheten ut. Terrorismen i västvärlden har snarare gått i vågor. Under 1970-talet ökade antalet attacker markant, nådde sin kulmen under 1980-talet, och minskade därefter under 1990-talet. Efter 11 september-attackerna 2001 ökade antalet igen med en ny topp runt 2010, innan antalet återigen sjönk efter 2014 (START, 2025). Detta antyder att digitaliseringen inte orsakar våld i sig, men att den omformar villkoren för hur radikalisering sker genom att möjliggöra nya vägar, tempon och kulturella uttryck för extremism.

Mot denna bakgrund uppstår behovet av att bättre förstå de mekanismer genom vilka digitala miljöer påverkar radikaliseringsprocesser – både på individnivå och i mer kollektiv skala. Även om forskningen kring onlineradikalisering har vuxit snabbt under senare år, återstår många obesvarade frågor: Hur samverkar tekniska plattformslogiker med psykologiska och sociala drivkrafter? Var går gränsen mellan ideologisk radikalisering och faktisk beredskap till våld? Och vilka motstrategier har visat sig verksamma? Det är dessa frågor som står i centrum för denna rapport.

1.1 Syfte

Syftet med denna rapport är att ge en översikt av den existerande forskningen om hur internet och sociala medier påverkar radikaliseringsprocesser och förekomsten av politiskt motiverat våld. Fokus i rapporten ligger huvudsakligen på den högerradikala miljön, eftersom det är här

onlineradikaliseringens uttryck och konsekvenser blivit mest synliga i västvärlden under det senaste decenniet. Samtidigt inkluderas även exempel och forskningsstudier som rör andra former av extremism, i syfte att identifiera gemensamma mönster och kontrasterande mekanismer. Ambitionen är således att bidra till en bredare förståelse av onlineradikalisering, med högerextremismen som det huvudsakliga empiriska fönstret.

Genom klargörandet av centrala begrepp och forskningsperspektiv samt genom att placera onlineradikaliseringen i en bredare teoretisk kontext, eftersträvas en nyanserad förståelse av hur digitala och sociala krafter samverkar i radikaliseringens dynamik. Avsikten är att rapporten därigenom kan bidra till ett kunskapsunderlag som kan nyttjas för fortsatt forskning och samhällsdebatt, men också som ett underlag för myndigheters strategiarbete kopplat till extremism.

Disposition

Rapporten är strukturerad så att den inledande delen behandlar centrala utmaningar i forskningen om radikalisering, särskilt när det gäller onlineradikalisering. Därefter följer en begreppsgenomgång och en distinktion mellan olika typer av radikalisering. Kapitel 3 ger en översikt över teoretiska perspektiv inom radikaliseringsforskningen, med syftet att kontextualisera betydelsen av internet och sociala medier i en bredare teoretisk förståelseram.

Huvuddelen av rapporten behandlar internets och sociala mediers påverkan på radikaliseringsprocesser. Den är indelad i tre tematiska områden: individbaserad forskning, studier av digitala plattformar och mediemiljöer, samt forskning om preventiva strategier och interventioner. Avslutningsvis presenteras en syntetiserande modell som för samman olika teoretiska perspektiv för att förklara internets roll och betydelse i radikalisering.

1.2 Centrala utmaningar inom forskning om radikalisering

Radikalisering är ett komplext och svårfångat fenomen, vilket gör det särskilt utmanande att studera – inte minst i dess digitala uttrycksformer. Nedan presenteras några av de mest framträdande metodologiska och begreppsliga hinder som forskningsfältet har att brottas med.

Ett fragmenterat forskningsfält

Radikalisering har historiskt studerats inom flera olika discipliner – terrorismforskning, social rörelseforskning, kriminologi, antropologi – liksom inom det framväxande, tvärvetenskapliga fält som fokuserar på att förebygga extremism och terrorism, ofta benämnt PVE (*prevention of violent extremism*) eller CVE (*countering violent extremism*). Mycket av radikaliseringsforskningen bedrivs dessutom utanför, eller i gränslandet till, den akademiska världen: bland myndigheter, tankesmedjor och fristående forskningsinstitut. Resultatet är ett forskningslandskap präglad av skilda teoretiska traditioner, varierande begreppsanvändning och brist på enhetliga definitioner.

Denna disciplinära fragmentering blir särskilt påtaglig inom forskningen om online-radikalisering. Här tillkommer ytterligare fält som medie- och kommunikationsvetenskap, med en uppsjö av närliggande termer som *polarisering*, *desinformation*, *gamification*, *hatretorik*, *toxic language*, *filterbubblor* och *ekokammare*. Dessa termer används ofta parallellt men sätts sällan i relation till varandra, trots att de i många fall beskriver likartade dynamiker. Denna begreppsmångfald försvårar översiktsarbeten och metaanalyser samt bidrar till en begreppslig oklarhet. Detta gäller inte minst själva kärnbegreppet *radikalisering*, som ofta fylls med olika betydelser inom olika discipliner, något som diskuteras mer utförligt i nästa avsnitt (1.3).

Komplexa och flernivåbaserade processer

Nutida forskning betraktar i regel radikalisering som en mångfasetterad process där individuella, sociala och strukturella faktorer samverkar över tid. Detta innebär att man sällan kan identifiera enskilda radikaliseringsfaktorer, utan snarare behöver förstå samspelet mellan omständigheter. En och samma strukturella faktor – exempelvis arbetslöshet – kan få olika effekter beroende på individens situation, sociala nätverk och ideologiska referensramar. På motsvarande sätt fungerar digitala miljöer inte som isolerade orsaksfaktorer, utan förstärker och modifierar redan befintliga processer i samverkan med andra variabler. Denna komplexitet utgör en tydlig metodologisk utmaning, eftersom den kräver teorier och metoder som kan hantera icke-linjära och kontextberoende samband.

Sammanflätning online och offline

En vanlig slutsats inom samtida forskning är att digitala och analoga radikaliseringsmiljöer inte går att separera empiriskt. Individer som sprider extremistiskt innehåll på nätet är ofta samtidigt verksamma i fysiska miljöer där de deltar i möten, demonstrationer eller lockar anhängare genom sina sociala nätverk. Radikalisering bör därför förstås som en process som utspelar sig i en hybrid miljö, där online- och offlinevärldar samverkar. Detta gör det metodologiskt problematiskt att analysera internet som en självständig orsak till radikalisering. I stället måste nätet förstås som en delkomponent i ett större, socialt och kulturellt ekosystem. Det innebär också att digitala uttryck måste analyseras i ljuset av den etablerade forskningen om radikalisering i bred bemärkelse.

Begränsad empirisk tillgång

En av de mest påtagliga utmaningarna i fältet är bristen på tillförlitliga och heltäckande data. Att kartlägga hela radikaliseringsresor – från första kontakt online till eventuell våldsutövning – är svårt av både etiska och praktiska skäl. Forskning om våldsbejakande miljöer innebär ofta säkerhetsrisker och kräver omfattande metodologiska och etiska överväganden. Resultatet är att studier ofta faller inom två kategorier: kvalitativa fallstudier som erbjuder djup och detaljerad förståelse av enskilda fall eller miljöer, men med begränsade möjligheter för generalisering i statistisk mening, och storskaliga kvantitativa analyser som ger bredd på bekostnad av individnära insikter. Experimentella studier förekommer, till exempel kring effekten av olika former av medieexponering, men de mäter vanligtvis attityder snarare än faktiska handlingar. Sammantaget finns en risk att forskningen överskattar vissa variabler och

underskattar samspelet mellan dem, vilket gör det svårt att skapa en helhetsbild av radikaliseringsprocessen.

Dessa sammantagna utmaningar visar tydligt varför det är svårt att genomföra en systematisk litteraturöversikt över forskningen om onlineradikalisering. Fältet är tvärvetenskapligt och begreppsligt fragmenterat, där studier ofta utgår från olika begrepp, definitioner och teoretiska ramverk. Denna begreppsliga spretighet medför betydande tolkningsproblem och gör det svårt att fastställa jämförbara urvalskriterier – vad som ska inkluderas eller exkluderas – i en systematisk översikt. Mot denna bakgrund har jag valt att inte göra en systematisk litteraturöversikt, trots att sådana ofta ses som standardmetod i forskningssammanställningar. I stället har jag valt att genomföra en kartläggande översikt – en så kallad *scoping review* – som är särskilt lämplig när ett forskningsfält är mångfasetterat och fragmenterat. Syftet är inte att besvara en avgränsad forskningsfråga, utan att identifiera övergripande teman, forskningsluckor och teoretiska inriktningar inom det aktuella kunskapsfältet.

Rapporten är strukturerad med detta syfte i åtanke. Den klargör centrala teoretiska begrepp och perspektiv, placerar onlineradikalisering i en bredare teoretisk kontext och sammanfattar vad vi idag vet om hur digitala miljöer påverkar radikaliseringsprocesser. Genom att integrera insikter från flera olika discipliner eftersträvar rapporten att bidra till en nyanserad förståelse av hur digitala och sociala krafter samspelar i radikaliserings dynamik.

1.3 Att definiera radikaliserings

Forskning om radikaliserings inleds ofta med konstaterandet att begreppet saknar en entydig definition. Vad som avses med radikaliserings varierar inte bara mellan akademiska discipliner – såsom terrorismstudier, kriminologi och forskning om sociala rörelser – utan även beroende på syfte och sammanhang. Definitionerna skiljer sig särskilt tydligt mellan akademiska och policyorienterade fält, där det senare – exempelvis inom PVE (*Preventing Violent Extremism*) – ofta bygger på pragmatiska behov snarare än analytisk stringens. Denna mångfald i användning är i sig inte ett problem, men när olika perspektiv blandas utan tydlig distinktion kan det skapa förvirring snarare än fördjupning.

I denna rapport utgår jag från en begreppsanvändning där två former av radikaliserings särskiljs: *kognitiv* respektive *beteendemässig* radikaliserings. Denna distinktion – mellan att å ena sidan tillägna sig extrema idéer som kraftigt avviker från rådande normer i samhället, och att å andra sidan agera ut dessa idéer i form av handlingar – är etablerad i litteraturen (McCauley & Moskalenko, 2008; Wolfowicz et al., 2021) och fungerar som ett analytiskt ramverk för rapportens resonemang.

Kognitiv radikaliserings

Kognitiv radikaliserings, även kallad ideologisk eller attitydbaserad radikaliserings, avser den process genom vilken individer internaliserar idéer som uppfattas som extrema eller radikala. Det är dock viktigt att understryka att detta är ett normativt och kontextberoende begrepp: vad som betraktas som ”radikalt” eller ”extremt” i ett auktoritärt samhälle kan vara fullt legitimt i

ett demokratiskt, och vice versa. I praktiken har begreppet använts för att beskriva allt från intresse för konspirationsteorier – som Flat Earth eller QAnon – till engagemang i demokrati-rörelser i repressiva regimer.

Flera forskare har därför föreslagit breda definitioner av kognitiv radikaliserings. Anja Dalgaard-Nielsen (2010), tidigare chef för den danska säkerhetspolisen, definierar det exempelvis som ”en växande beredskap att stödja långtgående samhällsförändringar som utmanar eller hotar den rådande ordningen” (s. 797). På liknande sätt beskriver Horgan & Braddock (2010) radikaliserings som ”en social och psykologisk process där individer gradvis engagerar sig i en extremistisk politisk eller religiös ideologi” (s. 279). Gemensamt för dessa definitioner är att de betonar kognitiv radikaliserings som en successiv förskjutning bort från idéer och världs-uppfattningar som uppfattas som mainstream i det omgivande samhället.

Andra forskare vill dock avgränsa begreppet till att endast omfatta idéer som uttryckligen legitimerar eller förespråkar våld. McCauley & Moskalenko (2008) definierar exempelvis radikaliserings som ”en förändring i tro, känslor och beteenden som rättfärdigar våld mellan grupper och kräver uppoffringar i försvaret av den egna gruppen” (s. 416). Här avgränsas således kognitiv radikaliserings till att handla om idéer som i sig inte nödvändigtvis leder till våld, men som kan utgöra en grund för att normalisera eller rättfärdiga våldsanvändning.

Beteendemässig radikaliserings

Beteendemässig radikaliserings, även kallad handlingsradikaliserings, syftar däremot på övergången från idéer till handling. Det innefattar dels våldsamma uttryck som attentat, misshandel eller planering av våldsdåd, dels andra typer av handlingar som på olika sätt syftar till att främja eller realisera extremistiska mål. Det kan röra sig om hot, trakasserier, spridning av propaganda, eller deltagande i strukturer eller organisationer som underminerar demokratiska institutioner. I denna rapport ligger dock fokus huvudsakligen på våld, eftersom det är kring detta som forskningsfältet har samlat mest empiri och teoriutveckling.

Detta handlingsfokuserade perspektiv är vanligt inom forskning som inspireras av sociala rörelser, där radikaliserings förstås som en kollektiv eskalering från icke våldsamma protestformer till alltmer våldsamma repertoarer. Donatella Della Porta (2018) beskriver exempelvis denna process som en ”eskalering från ickevåld till alltmer våldsamma handlingsrepertoarer, en utveckling som sker genom komplexa och ömsesidiga interaktioner över tid” (s. 462). Hos Della Porta och andra rörelseforskare är radikaliserings dock i första hand ett *kollektivt* begrepp, med tyngdpunkten snarare på grupp-dynamik och interaktion än på individuella övertygelser.

Skilda men delvis överlappande processer

Att olika forskningsfält använder olika definitioner av radikaliserings är inte i sig ett problem – de speglar olika analytiska syften. Problemen uppstår först när begreppen används växelvis utan att deras skillnader tydliggörs. Det finns därför goda skäl att initialt hålla begreppen kognitiv och beteendemässig radikaliserings analytiskt åtskilda, för att på så vis kunna undersöka hur de samspelar.

Forskningen är i stort sett enig om att radikala idéer i sig inte utgör en tillförlitlig indikator på våldspotential (se även Bjørge & Horgan, 2009). Wolfowicz et al. (2021) visar i sin omfattande litteraturöversikt över riskfaktorer för radikalisering att radikala åsikter visserligen är en viktig komponent i förståelsen av vägen till våld, men som enskild faktor kan åsikter inte förklara våldsanvändning i särskilt stor utsträckning. Enligt en ofta återgiven uppskattning är det endast 1 procent av dem som hyser radikala åsikter som faktiskt begår våldshandlingar (se även Schuurman & Taylor, 2018). En nypublicerad studie av Schuurman & Carthy (2024) visar dessutom att deltagande i extremistiska miljöer i vissa fall kan fungera som en skyddsfaktor, under förutsättning att miljön inte aktivt förespråkar våld.

Samtidigt pekar ett stort antal studier på att kognitiva och beteendemässiga processer ofta samverkar, om än på indirekta och kontextberoende sätt. Som diskuteras senare i rapporten är det vanligt att ideologiska övertygelser – även om de inte i sig uppmanar till våld – kan bidra till att normalisera hatretorik och öka acceptansen för våld i vissa grupper. Sådana processer är ofta kollektiva snarare än individuella, och medierade av sociala, tekniska och kulturella faktorer. I dessa fall blir det alltså inte fråga om en rak väg från idé till våld, utan om ett komplex samspel där idéer, känslor och sociala relationer samverkar i en dynamisk process.

Det ovan nämnda begreppet *stokastisk terrorism* är användbart för att förstå detta gränsland mellan idé och handling. Som tidigare nämnts syftar det på situationer där offentlig våldsbejakande retorik – ofta uttryckt av inflytelserika opinionsbildare, politiker eller medieaktörer – skapar en miljö där vissa individer uppfattar budskapen som en uppmaning till våldshandlingar. Det stokastiska består i att det inte går att förutsäga exakt vem som kommer att agera – eller när – men sannolikheten att någon gör det ökar i takt med att det ideologiska klimatet förändras. På så sätt utgör stokastisk terrorism en analytisk bro mellan kognitiv och beteendemässig radikalisering, eftersom den belyser hur idéer kan påverka handling genom ett komplext socialt och medialt ekosystem – utan direkt uppvigling eller planering.

Ytterligare ett skäl till att göra en analytisk åtskillnad mellan begreppen är att forskning visar att kognitiv och beteendemässig radikalisering tenderar att drivas av delvis olika orsaksfaktorer. Beteendemässig radikalisering är ofta kopplad till grupptryck, socialisering, statusjakt eller känslomässiga drivkrafter som hat, hämnd eller frustration – det vill säga interaktionistiska och situerade processer. Kognitiv radikalisering kan förstås också involvera interaktionistiska faktorer, men fokuserar ofta mer på frågor om individens identitet, världsbild och ideologiska övertygelser.

Dessa skillnader har både metodologiska och praktiska implikationer. Som Robin Andersson Malmros (2025) påpekar: att förstå radikalisering som en åsiktsförskjutning förutsätter studier av hur och varför individer utvecklar ett radikalt språk i relation till andra. Att förstå radikalisering som vägen till handling innebär i stället att studera vad individer faktiskt gjort – när, hur och varför – i relation till våldshandlingar eller andra former av extremistiskt beteende.

Sammantaget har man således mycket att vinna på att hålla begreppen kognitiv och beteendemässig radikalisering analytiskt åtskilda. En sådan distinktion är viktig inte bara för att förstå hur de samspelar, utan också för att kunna utvärdera deras konsekvenser. Ideologisk radikalisering kan nämligen vara samhällsskadlig även i frånvaro av fysiskt våld. När extrema

idéer – till exempel högerextrema, misogyna eller antivetenskapliga – får bredare fäste i samhället kan det bidra till att urholka demokratiska värden, minska tilliten till institutioner och öka diskriminering och samhällelig polarisering. Klimatförnekelse eller antivetenskapliga rörelser kan få långtgående konsekvenser för folkhälsa och klimatpolitik, på samma sätt som normaliseringen av främlingsfientliga och auktoritära idéer kan bana väg för inskränkningar i yttrandefrihet, minoritetsrättigheter och rättssäkerhet – även utan att något våld ägt rum. Att analytiskt skilja mellan olika typer av radikalisering är därför avgörande för att kunna förebygga både politiskt våld och demokratisk erosion.

Kollektiv och individuell radikalisering

En annan viktig distinktion – som dock fått mindre uppmärksamhet i forskningslitteraturen – är den mellan individuell och kollektiv radikalisering. Denna skillnad handlar inte om huruvida radikalisering leder till våld, utan om var själva processen är förankrad: i individen eller i ett socialt sammanhang.

Individuell radikalisering avser processer där enskilda individer utvecklar radikala attityder eller beteenden, ibland i avsaknad av fysiska gruppmiljöer. Detta fenomen har särskilt uppmärksammats i forskning om ensamagerande gärningsmän, där radikaliseringen i många fall sker genom digital påverkan snarare än genom deltagande i organiserade miljöer. Processen tenderar att vara idiosynkratisk, där individens psykologi, personliga erfarenheter och upplevda orättvisor samverkar med ideologisk exponering – exempelvis via sociala medier, manifest eller digitala gemenskaper. Som vi kommer att se i senare kapitel (avsnitt 3.1) är detta perspektiv centralt inom såväl terrorismforskning som litteratur om onlineradikalisering.

Kollektiv radikalisering, å andra sidan, handlar om gruppbaseade processer där interpersonella relationer, gruppdynamik och gemensam identitet spelar avgörande roller för att förstärka och normalisera radikala uppfattningar eller handlingar. Här är radikaliseringen inte primärt ett inre, psykologiskt förlopp, utan snarare ett resultat av deltagande i en social miljö – exempelvis organiserade rörelser, subkulturella gemenskaper eller löst sammansatta digitala nätverk. Inom social rörelseteori har denna kollektivt situerade förståelse betonats särskilt. Här ses radikalisering som en dynamisk process, formad av social interaktion och kontext, snarare än som uttryck för en individuell predisposition – även när det gäller ensamagerande gärningspersoner. Betydelsen av institutionellt bemötande lyfts ofta fram som en utlösande eller förstärkande faktor – exempelvis polisrepression, bristande politiskt gehör eller upplevd marginalisering. Kollektiv radikalisering kan därmed utvecklas genom eskalerande cykler av konfrontation mellan aktivister, motdemonstranter och myndigheter, där våld successivt legitimeras och betraktas som en rimlig eller nödvändig strategi.

Radikal och extremist

I denna rapport används begreppen radikal och extremist huvudsakligen synonymt, utan att en strikt begreppslig åtskillnad görs. Visserligen finns det teoretiska försök att särskilja dem, exempelvis genom att definiera radikalism som en strävan efter djupgående samhällsförändring inom ramen för demokratiska principer, medan extremism som ett motstånd mot dessa principer (Mudde, 2007). I praktiken är dock gränsdragningen ofta flytande, särskilt i empiriska studier

där fokus ligger på aktörers faktiska handlingar snarare än deras explicita ideologiska deklamationer. Att använda begreppen som utbytbara är därför metodologiskt motiverat i en kontext där intresset ligger i hur idéer och handlingar tar sig uttryck, snarare än i idealtypiska definitioner.

Därtill bör man beakta den historiska och politiska glidningen i begreppsanvändningen, särskilt inom policyfältet och den offentliga debatten, där *radikal* och *extremist* ofta används omväxlande och utan tydliga definitioner. I linje med detta intar rapporten en pragmatisk hållning: båda begreppen används för att beskriva idéer, uttryck och handlingar som står i konflikt med grundläggande demokratiska värden, och där våld eller hot om våld används – eller förespråkas – som medel för att uppnå politiska mål.

Jag har i möjligaste mån strävat efter att tydliggöra och särskilja mellan centrala begrepp och analytiska distinktioner i denna rapport. Samtidigt måste man vara medveten om att många empiriska studier inom fältet inte tillämpar enhetliga eller väldefinierade begrepp. Det innebär att vissa termer måste tolkas kontextuellt. Distinktionerna som den mellan individuell och kollektiv radikaliserings, mellan kognitiv och beteendemässig radikaliserings, samt mellan radikala och extrema aktörer, är analytiskt viktiga – men inte alltid skarpt urskiljbara i praktiken. Rapportens ambition är att bidra till en fördjupad förståelse av dessa gränsdragningar och deras betydelse för hur vi tolkar radikaliserings uttryck, orsaker och konsekvenser.

2. Teoretiska perspektiv på radikaliserings

Detta kapitel ger en överblick över några av de centrala teoretiska perspektiv som präglar forskningen om radikaliserings. Syftet är inte att erbjuda en uttömmande genomgång av fältets hela bredd – en sådan genomlysning skulle kräva en betydligt mer omfattande studie – utan snarare att skapa en teoretisk ram för att förstå onlineradikaliserings. Detta är viktigt eftersom forskning om digital radikaliserings ofta saknar en tydlig förankring i etablerad forskning om radikaliseringsprocesser i bredare bemärkelse.

Trots att radikaliserings som forskningsområde har teoretiska rötter i flera discipliner, kan man urskilja två dominerande traditioner: terrorismstudier och forskning om sociala rörelser. Terrorismstudier har främst fokuserat på individers väg till våldsbejakande extremism och på betydelsen av både strukturella och individuella faktorer. Forskningen om sociala rörelser har å andra sidan betonat kollektiv mobilisering, betydelsen av ideologi samt bredare politiska möjlighetsstrukturer som skapar förutsättningar för radikaliserings. Under lång tid utvecklades dessa forskningsfält parallellt, ofta med begränsad kontakt. På senare år har dock en allt större sammansmältning skett, där forskare försöker integrera och syntetisera perspektiv från båda fälten.

2.1 Strukturella faktorer

En betydande del av den tidiga radikaliseringsforskningen – särskilt under 1990-talet och det tidiga 2000-talet – var starkt präglad av terrorismstudiernas fokus på strukturella orsaksfaktorer. Denna forskning belyste hur bredare samhällsförhållanden, som strukturell ekonomisk ojämlikhet, snabba moderniseringsprocesser eller auktoritärt styre i ett land, kunde bidra till att skapa förutsättningar för politiskt våld.

Ett av de mest inflytelserika teoretiska ramverken i denna tradition är *strain-teorin*, med rötter i kriminologin (Agnew, 1992; Merton, 1938). Enligt denna teori genererar systemisk social ojämlikhet – som brist på inkomst, utbildning eller andra resurser – frustration och känslor av alienation. För vissa individer kan våld framstå som ett utlopp för denna frustration, ett sätt att återta kontroll eller försöka framtvinga samhällsförändringar.

Trots sitt inflytande är dock strain-teorin, liksom närliggande strukturella förklaringsmodeller, inte utan svagheter. En av de mest grundläggande invändningarna är att dessa teoretiska perspektiv har svårt att förklara varför endast ett fåtal individer väljer våld, trots att många upplever liknande strukturella villkor. Majoriteten av människor som befinner sig i ekonomisk

eller social utsatthet blir inte våldsamma, utan söker i stället icke våldsamma vägar till förändring.

En annan kritik är att strukturella modeller tenderar att fokusera på individens omedelbara livsvillkor, och därmed förbiser att radikalisering också kan motiveras av *upplevd* strain. Denna strain kan handla om förväntad framtida exkludering, eller om solidaritet med andra grupper som utsätts för förtryck och orättvisa. Individen behöver alltså inte själv vara direkt drabbad – identifikation med andra eller oro för en framtid av marginalisering kan vara tillräckligt mobiliserande.

Vidare har dessa modeller fått kritik för att implicera en förenklad och reaktiv förklaringslogik, där radikalisering betraktas som en direkt följd av svåra förhållanden. En sådan kausalitet, menar dessa kritiker, lämnar litet utrymme för att förstå radikalisering som en långsiktig och ofta icke-linjär process, där idéer, tolkningar och sociala interaktioner utvecklas över tid. Den linjära modellen – från strukturell ojämlikhet till våld – bygger enligt kritikerna på en föreställning om automatisk respons, som idag i allt högre grad ifrågasätts till förmån för mer processuella förståelser.

Empiriskt är stödet för en direkt koppling mellan strukturell ojämlikhet och radikalisering dessutom begränsat. Exempelvis bestod flera av 1970-talets vänsterterroristiska grupper i hög grad av personer från medelklassen, vilket talar emot att fattigdom i sig skulle vara en avgörande eller nödvändig förklaring (McCauley & Moskalenko, 2017). Likaså har demografiska faktorer som utbildningsnivå visat sig ha svag förklaringskraft för politiskt våld (Kis-Katos et al., 2011; Krueger & Malečková, 2002).

Vissa studier har dock pekat på att traditionella rasistiska organisationer ofta rekryterar individer från arbetarklassen, särskilt bland dem som upplever ekonomisk marginalisering (Dobratz & Shanks-Meile, 2000; Ezekiel, 2002; Mudde, 2019). Samtidigt visar nyare forskning att den samtida extremhögern har en bredare social räckvidd och även attraherar individer från medelklassen. Ett tydligt exempel är stormningen av den amerikanska kongressen den 6 januari 2021, där en majoritet av deltagarna var medelålders, väletablerade vita män – omkring 40 procent hade medelklassyrken och många var egenföretagare (Kyler, 2020).

Sammantaget kan man konstatera att strukturella modeller har ett fortsatt värde för att identifiera kontextuella riskfaktorer och kartlägga miljöer där radikalisering är mer sannolik. Men de räcker inte för att förklara varför vissa individer radikaliserar medan andra i samma eller liknande kontext inte gör det. För att förstå dessa skillnader krävs mer dynamiska och processinriktade modeller, som fångar hur upplevelser, idéer, identiteter och sociala relationer samverkar i en ofta långsamt eskalerande radikaliseringsprocess.

2.2 Individuella faktorer

Efter terroristattackerna den 11 september 2001 skedde ett tydligt skifte inom terrorismstudier. I den policy- och säkerhetsinriktade diskurs som följde trängdes de tidigare strukturella och sociopolitiska förklaringsmodellerna i viss mån undan till förmån för mer individcentrerade

perspektiv på radikaliserings. Detta innebar ett ökat intresse för psykologiska, socialpsykologiska och mikrosociologiska teorier. I denna kontext kom forskningen att i högre grad uppmärksamma hur individer radikaliserar i specifika miljöer – ofta i interaktion med ideologi, nätverk och sociala gemenskaper. Fokus låg inte längre främst på psykopatologi, vilket delvis hade präglat fältet under 70-, 80- och 90-talen med den tidens fokus på individuella faktorer, utan också på hur ideologisk dragningskraft, känslomässiga tillstånd och social identifikation samverkade i individers väg mot politiskt våld.

En tidig men inflytelserik riktning inom denna forskning handlade om att identifiera en särskild "terroristpersonlighet" – en uppsättning stabila psykologiska egenskaper som antogs kunna förutsäga benägenheten att begå extremistiska våldsdåd. Denna inriktning har dock kritiserats för att individualisera och reducera ett i grunden politiskt och socialt fenomen till en fråga om individuell avvikelse, samt för att sakna robust empiriskt stöd.

Psykologen Randy Borum har sammanfattat över 40 års forskning i slutsatsen att det inte finns någon enhetlig eller meningsfull terroristprofil. Tvärtom visar forskningen att personer som begår våldsdåd i politiska syften uppvisar stor variation i bakgrund och personlighet och ofta framstår som "helt normala" (Borum, 2004). På liknande sätt menar Ruby (2002) att terrorister sällan är psykiskt sjuka, utan snarare rationella aktörer som använder våld som ett strategiskt verktyg i avsaknad av andra medel för påverkan (se även Crenshaw, 2007).

Trots att idén om en terroristprofil idag i stort sett har övergetts, fortsätter forskningen att undersöka individuella egenskaper som potentiella riskfaktorer. Resultaten är dock ofta motstridiga. Vissa studier finner exempelvis ett svagt samband mellan introverta personlighetsdrag och benägenheten att omfamna extrem ideologi (Feddes et al., 2023), medan andra inte lyckats replikera dessa resultat. Det finns även viss evidens för att personer med hög *social dominance orientation* (SDO) – alltså en tendens att förespråka hierarkier och ojämlikhet – i högre grad accepterar eller förespråkar våld för att försvara den egna gruppen (Pratto et al., 1994; Sidanius & Pratto, 2001).

Sambandet mellan religiositet och radikaliserings har varit omdiskuterat. Vissa studier pekar på samband mellan fundamentalistiska trosföreställningar och acceptans för våld, medan andra inte finner några tydliga samband, särskilt när man skiljer mellan yttre religiös praktik och inre övertygelser. Vad gäller psykisk ohälsa finns inget generellt stöd för att exempelvis depression i sig skulle öka risken för radikaliserings. Däremot har vissa studier identifierat ett samband mellan posttraumatisk stress och antisocialt beteende i specifika grupper (Dixon et al., 2005; Greenwald, 2014).

Ett viktigt undantag från den generella bilden är forskningen om ensamagerande gärningspersoner. Till skillnad från kollektivt organiserade aktörer har ensamagerande ofta mer särpräglade drivkrafter och svagare ideologisk förankring. Här finns ett tydligare empiriskt stöd för att individuella sårbarheter – som psykisk ohälsa, social isolering, personlig kris eller känslor av misslyckande – spelar en mer framträdande roll i radikaliseringsprocessen (Corner & Gill, 2015; Gill et al., 2014; Schuurman & Carthy, 2024). Enligt Corner & Gill (2015) är det nästan 14 gånger vanligare att ensamagerande terrorister har en psykiatrisk diagnos jämfört med gruppbaseerade terrorister. Deras studie visar att 32 procent av de ensamagerande

individerna hade dokumenterad psykisk ohälsa, där diagnoser som schizofreni var överrepresenterade. Det är dock viktigt att tolka dessa resultat med viss försiktighet, då urvalet i studien var begränsat (119 individer i respektive grupp).

En ofta förbisedd aspekt i diskussionen om individuella förklaringar är de kulturella tolkningsramar som individer använder för att legitimera våld. Efter 9/11 växte en litteratur fram som betonade kulturella faktorer såsom heder, martyrskap, kollektiv skam eller föreställningar om civilisatoriska konflikter som mobiliserande krafter. Dessa faktorer bör dock inte förstås som statiska kulturella egenskaper, utan som symboliska resurser som individer aktiverar för att ge sina handlingar mening och betydelse. Även om dessa perspektiv har påverkat både forskning och policyutformning, finns också kritik mot att de riskerar att ”kulturförklara” våldet. Charles Tilly (2004) har varnat för just detta och argumenterat för att terrorism bör förstås som en *taktik* – inte som ett uttryck för essentialistiska kulturella eller psykologiska predispositioner. I sin text *Terror, Terrorism, Terrorists* argumenterar han för att forskningen bör fokusera på de organisatoriska och politiska sammanhang där våld uppstår. Samtidigt bör det påpekas att vad Tilly kallar ”specialister på våld” också kan verka inom statliga organisationer, såsom militär och polis, där mobiliseringsprocesser delvis liknar dem som identifierats i militanta politiska grupper (jfr Moskalenko, 2010). Av pragmatiska skäl avgränsas dock denna studie till icke-statliga aktörer.

Sammantaget är sambandet mellan individuella faktorer och radikaliseringsprocesser generellt svagt – vad gäller både ideologisk och beteendemässig radikaliseringsprocess. Föreställningen om en generell personlighetsprofil har till stor del förkastats inom forskningen, även om den i hög grad lever kvar i den offentliga debatten. Detta innebär dock inte att individuella faktorer är irrelevanta. Tvärtom kan de fungera som medierande faktorer som bidrar till att förklara varför vissa personer tycks vara särskilt mottagliga för radikaliserande budskap i vissa situationer och kontexter. Men de är sällan tillräckliga för att ensamma förklara – eller än mindre förutsäga – radikaliseringsprocesser.

2.3 Relationella och interaktionistiska perspektiv

Som tidigare nämnts har forskningen om sociala rörelser utvecklats parallellt med terrorismforskningen, men med ett delvis annorlunda fokus. Medan terrorismstudier traditionellt har betonat individers drivkrafter eller strukturella bakgrundsfaktorer, har rörelseforskningen främst bidragit med insikter om radikaliseringsprocesser som en relationell och processuell dynamik. Inom detta perspektiv ses radikaliseringsprocesser inte som en linjär förändring hos enskilda, isolerade individer utan som ett resultat av komplexa växelspel mellan individer, grupper och institutionella aktörer.

Till skillnad från mer statiska förklaringsmodeller betonar rörelseperspektivet att valet mellan våld och ickevåld uppstår i ett sammanhang där flera faktorer samverkar: tillgång till resurser, rörelsens relation till omgivande samhällsaktörer, institutionellt bemötande (exempelvis i form av polisrepression eller politiskt gehör för rörelsens krav) samt kollektiva tolkningar av händelser och situationer. Här spelar teorin om politiska möjlighetsstrukturer (*political*

opportunity structures) en central roll. Den betonar att förändringar i det politiska landskapet kan skapa nya möjligheter eller hot, vilket i sin tur påverkar om – och hur – radikaliserade praktiker eskalerar eller avtar (Alimi et al., 2015).

Ett viktigt bidrag från denna tradition är synen på radikalisering som ett *kollektivt fenomen*, även när den yttrar sig i så kallat ensamagerande våld. Även ensamagerande aktörer formas i sociala miljöer, influeras av gruppnormer och exponeras för ideologiskt material som cirkulerar i olika sociala gemenskaper – inte minst i digitala sammanhang. Denna processinriktade ansats som dominerar inom rörelseforskningen har därmed varit avgörande för att utveckla mer dynamiska modeller, där individuella och kollektiva faktorer integreras i analysen av radikaliseringsförlopp.

Framing och upplevda hot

En bärande tanke inom rörelseforskningen är att externa händelser eller hot i sig inte automatiskt leder till radikalisering. Det är i stället hur dessa händelser tolkas – det vill säga hur de ramas in (framing) – som avgör hur människor reagerar. Särskilt avgörande är hur grupper uppfattar hot mot sin kollektiva identitet.

Aktörer inom sociala rörelser producerar berättelser, symboliska tolkningar och narrativ som konstruerar upplevelsen av orättvisa, pekar ut ansvariga aktörer samt föreslår lösningar och uppmanar till handling för att förändra detta. Dessa så kallade diagnostiska, prognostiska och mobiliserande ramar är centrala för att förstå varför vissa grupper radikaliserar i vissa kontexter.

Robert Pape (2021) har exempelvis visat att rädslan för att icke-vita grupper ska få ökad makt och status är en av de starkaste förklaringsfaktorena bakom deltagande i rebellgrupper och upprorsrörelser i USA. Denna rädsla artikuleras ofta genom konspirationsteorier, såsom idén om ett ”folkutbyte”, där vita påstås bli ersatta av invandrade grupper (se även Waller et al., 2022). Polletta & Callahan (2017) har visat hur högerextrema miljöer mobiliserar så kallade *deep stories* – berättelser präglade av nostalgi, förlust och förräderi – som legitimerar våld som en form av självförsvar. Dessa berättelser förstärks i digitala miljöer, där de bidrar till att skapa en känsla av akut hot, normalisera extrema ståndpunkter och sänka tröskeln för våldsanvändning.

Relationer och kollektiv identitet

Som tidigare framgått sker radikalisering sällan isolerat. En stor mängd studier visar att etablerade sociala nätverk, vänskapsband och känslan av tillhörighet till en kollektiv identitet ofta är avgörande faktorer.

Marc Sageman (2004) betonar i sin klassiska teori om ”a bunch of guys” att radikalisering inte i första hand sker genom ideologisk indoktrinering uppifrån, utan genom informella gruppdyamiker mellan vänner och bekanta. På liknande sätt visar Doug McAdam (1986) i sin forskning om högriskaktivism och Jocelyn Viterna (2006) i sina studier om kvinnors deltagande i El Salvadors gerillarörelse att sociala relationer är avgörande för att skapa tillit, normförstärkning och handlingsbenägenhet. Genom gemensamma språkbruk, symboler och ritualer utvecklas en intern moral som ofta bygger på en stark polarisering: ”vi mot dem”. Denna form

av relationell polarisering är således inte bara ett resultat av radikaliseringsprocessen – den är också en aktiv drivkraft i den.

Trots sina styrkor har även det rörelseorienterade perspektivet vissa begränsningar. Det har utvecklats främst för att förstå kollektiv mobilisering inom avgränsade konfliktepoker – så kallade *episodes of contention* – vilket gör det mindre lämpat för att förstå långsamma, individuella radikaliseringsförlopp som sker utanför organiserade rörelser. Det saknas även ofta analytiska verktyg för att analysera hur individers psykologiska sårbarheter, digital miljöpåverkan och kulturella tolkningsramar samverkar över tid.

Dessa begränsningar har varit en viktig drivkraft bakom nyare försök att kombinera socialpsykologiska, ideologiska och relationella perspektiv i mer syntetiserande modeller. Dessa sammansatta angreppssätt, som vi nu ska titta närmare på, syftar till att överbrygga klyftan mellan individuella och kollektiva förklaringar, samt att bättre fånga radikaliseringens komplexitet i ett samtida, digitalt präglad sammanhang.

2.4 Den nya generationen radikaliseringsforskning

Nutida forskning kring radikalisering kännetecknas av en tydlig förskjutning från mer eller mindre statiska orsak-verkan-modeller till mer dynamiska, processinriktade och relationella ramverk. I stället för att betrakta radikalisering som ett resultat av beständiga faktorer – som fattigdom, psykisk ohälsa eller särskilda personlighetsdrag – ligger fokus numera på hur sådana faktorer får betydelse inom specifika tidsmässiga och sociala kontexter. Det handlar alltså inte om att förkasta tidigare insikter, utan snarare om att placera dem i tiden. Forskningen intresserar sig för hur strukturella, individuella och sociala villkor samverkar över tid, och hur de påverkar – och påverkas av – individens handlingar, val och omgivning.

Radikalisering ses inte längre som en linjär eller förutsägbar väg från punkt A till B, utan som en öppen och ibland motsägelsefull process som kantas av kritiska händelser, sociala interaktioner och meningsskapande. I stället för att förstå våld som en mekanisk reaktion på objektiva omständigheter, betonas hur individer successivt knuffas i en viss riktning, där olika faktorer – som upplevelsen av orättvisa, grupptryck eller identitetssökande – kan förstärka eller bromsa utvecklingen. En central insikt i detta perspektiv är att samma strukturella förhållanden, exempelvis socioekonomisk marginalisering, kan leda till olika utfall beroende på hur de tolkas och internaliseras av individen, samt vilka sociala relationer och möjligheter till kollektiv identifikation som finns tillgängliga i situationen.

Den alltmer komplexa syn på radikalisering som vuxit fram under senare år tydliggör två saker. För det första visar den att radikalisering är en (ofta) lång och oförutsägbar process, där olika förklaringsfaktorer gör sig gällande vid olika tidpunkter. För det andra visar den att olika förklaringsfaktorer kan ha olika betydelse för olika individer. För vissa individer är ideologin det primära; för andra är det social tillhörighet, sökande efter mening, status eller hantering av personliga trauman som är i förgrunden. Dessutom kan liknande mekanismer i olika kontexter leda till olika utfall – vilket betonar vikten av att förstå radikalisering som situerad snarare än

universell. En viktig förskjutning i samtida forskning är därmed erkännandet av individens aktiva roll. Individer ses inte längre som passiva mottagare av propaganda, utan som meningsskapande aktörer som tolkar sina erfarenheter i ljuset av sociala och politiska narrativ. Radikala ideologier erbjuder ofta både förklaringsmodeller och existentiell mening – de ger svar på hur orättvisor uppkommer, definierar fiender och ger individen en plats i ett större kollektivt ”vi mot dem”-projekt.

Därmed förstås radikaliserings inte bara som en kognitiv process, utan involverar även emotionella och symboliska processer. Affektiva dimensioner – som ilska, sorg, lojalitet och hopp – spelar en avgörande roll. Bilder, symboler och språkliga uttryck cirkulerar i sociala nätverk och formar subkulturer där specifika idéer och åsikter förstärks genom kollektiv bearbetning. I denna kontext handlar rekrytering till extremistiska grupper ofta inte om manipulation, utan snarare om processer där individens livsberättelse synkroniseras med gruppens berättelse. Hur stor betydelse denna typ av narrativpsykologiska perspektiv ges varierar dock i olika teoretiska perspektiv.

Pathway-modeller

Denna teoretiska förskjutning inom forskningen återspeglas i olika teoretiska perspektiv. Detta gäller både i pathway-modeller som Moghaddams (2005) *Staircase to Terrorism* och McCauleys & Moskalenkos (2017) *Pyramid Model*, men även i Hafez & Mullins (2015) *Radicalization Puzzle*, som betraktar radikaliseringsprocessen som en gradvis och kontextuell process (se tabell 1). Dessa modeller är olika försök att pedagogiskt illustrera radikaliseringsprocessen genom metaforer och scheman som visar hur individer gradvis rör sig från ett utgångsläge av missnöje till ett eventuellt stöd för, eller deltagande i, våld.

Moghaddams (2005) *staircase to terrorism*-modell är en av de tidigaste och mest spridda pathway-modellerna. Den beskriver radikaliseringsprocessen som en metaforisk trappa i sex steg. Trappan är linjär och bygger på antagandet att varje steg innebär en ökad risk för att individen rör sig mot terrorism. Grundidén är att människor (1) befinner sig i bottenplanet, med missnöje och upplevd orättvisa, men att bara några få klättrar uppåt, där man (2) kopplar detta missnöje till samhället eller det politiska systemet, (3) börjar se andra sociala grupper som ansvariga för sin prekära situation, (4) anammar en extremistisk ideologi som rättfärdigar våld, (5) integreras i en våldsbejakande grupp, och slutligen (6) blir en fullfjädrad terrorist.

McCauleys & Moskalenkos (2017) pyramidmodell bygger vidare på Moghaddams tankar men betonar snarare fördelningen av individer i olika stadier än en linjär rörelse mellan dem. Pyramiden illustrerar att de allra flesta som hyser missnöje befinner sig i basen och inte går vidare mot våld. Bara en liten minoritet befinner sig i toppen, där våldsutövning sker. Viktigt är att individerna inte nödvändigtvis rör sig uppåt i pyramidens struktur, utan att modellen snarare illustrerar en hierarki av engagemang och handling. McCauley & Moskalenko lägger större vikt vid gruppdynamik och social påverkan, i kontrast mot Moghaddams mer individcentrerade fokus. En annan skillnad är att pyramidmodellen implicit öppnar för att individer kan röra sig både uppåt och nedåt i pyramiden, beroende på omständigheter, vilket gör den mindre deterministisk. Det är också värt att notera att Moskalenko (2010) i en annan studie visar hur

liknande mobiliseringsmekanismer kan identifieras i statliga kontexter, exempelvis inom militär utbildning, vilket understryker att processerna inte är unika för militanta politiska grupper.

Hafez & Mullins (2015) *The Radicalization Puzzle* är i strikt mening inte en pathway-modell, eftersom temporala aspekter inte betonas. Till skillnad från trappan och pyramiden är detta inte en sekventiell pathway-modell utan en mer nätverkslik syntes. Radikalisering beskrivs här som ett pussel, där olika bitar måste samverka för att radikalisering ska ske. Dessa bitar inkluderar *grievances* (upplevd orättvisa eller frustration), *sociala nätverk* (som förstärker och normaliserar idéer), *ideologi* (som ger kognitiva och moraliska ramar för handling), samt *kontext* (till exempel konflikter, politisk repression eller kultur). Här ligger fokus på ömsesidigt förstärkande faktorer snarare än en linjär progression. Modellen är därför mer dynamisk och flexibel, och den fångar bättre komplexiteten i att vissa individer trots liknande missnöje inte radikaliseras, eftersom alla pusselbitar inte faller på plats.

Dessa pathway-baserade modeller används ofta som konceptuella ramverk snarare än som empiriskt bevisade modeller. Man kan beskriva dem som en brygga mellan äldre statistiska modeller och nyare icke-linjära ramverk. Hafez & Mullins modell är mer icke-linjär och individanpassad, där olika faktorer kan förstärka varandra. Detta öppnar upp för mer komplexa samband, då fokus ligger på interagerande faktorer, och modellen representerar ett försök att sammanföra strukturella, sociala och individuella mekanismer utan att låsa dem i en tidsordning. Snarare ligger fokus på hur strukturella frustrationer tolkas genom individens egna berättelser och narrativ.

Ett något annorlunda ramverk från dessa inträdesfokuserade modeller är Bjørgo & Horgan (2009) perspektiv, som i stället riktar uppmärksamheten mot processer för utträde och avradikalisering. Denna processmodell kan sägas komplettera inträdesmodellerna och förstärker förståelsen av radikalisering som icke-linjär. De visar att vägen ut ur våldsbejakande extremism sällan är en spegelbild av vägen in, och betonar särskilt betydelsen av praktiska omständigheter, sociala relationer och kognitiv distansering från våld snarare än ideologisk förändring. Här skiljer författarna mellan två separata men relaterade processer: å ena sidan *disengagement*, som innebär att individen upphör med våldsamt eller aktivt deltagande oavsett om de ideologiska övertygelserna kvarstår, och å andra sidan *deradicalization*, som är en mer djupgående process där individen också omvärderar och överger sin radikala ideologi. Ideologisk förändring är sällan den primära drivkraften till avradikalisering; i stället ligger fokus på sociala relationer (förändrade nätverk, brist på gruppsammanhållning, alternativa tillhörigheter), livsomständigheter och praktiska faktorer (som utmattning, familjeskäl, brottsmisstankar, brist på framgång).

Risk- och skyddsfaktorer

Parallellt med de teoretiska pathway-modellerna har en mer policyinriktad forskning vuxit fram med fokus på att kartlägga risk- och skyddsfaktorer som ökar respektive minskar sannolikheten för radikalisering (Borum, 2011a, 2011b; Da Silva et al., 2024; Feddes et al., 2015; Lösel & Bliesener, 2021; Schuurman & Carthy, 2024; Van den Bos, 2018; Wolfowicz et al., 2021). Dessa

ansatser påminner till viss del om Bjørgos & Horgans perspektiv, i det att de betonar multifaktoriella samband: det är sällan en enskild variabel som förklarar radikaliserings, utan snarare en kombination av olika faktorer som tillsammans ökar risken.

Risikfaktorer kan återfinnas på flera nivåer. På individnivå rör det sig ofta om psykisk ohälsa, trauma, identitetskriser eller låg självkänsla. På den sociala nivån kan faktorer som ensamhet, brist på tillhörighet eller negativa kamratgrupper spela roll, medan strukturella faktorer kan handla om diskriminering, fattigdom, marginalisering eller politisk repression. Skyddsfaktorer kan i sin tur bestå av starka befintliga sociala nätverk, familjeförankring, tillgång till utbildning och arbetsmarknad, samt kognitiva resurser som kritiskt tänkande och källkritisk förmåga (Bjørge & Gjelsvik, 2015; Bjørge & Horgan, 2009).

Denna typ av modeller har i regel inte primärt förklarande ambitioner, utan bör snarare förstås som strukturerande ramverk för empiriska studier och förebyggande arbete. De är ofta mer empiriskt grundade än pathways-modellerna, då de bygger på forskningsöversikter och metaanalyser. Samtidigt är de huvudsakligen deskriptiva och korrelationsbaserade, vilket innebär att de sällan fastställer kausala samband. Det gör att de är användbara för policy och praktik, men mindre lämpade som teoretiska förklaringsmodeller. En återkommande kritik är att de tenderar att överbetona individnivån, delvis eftersom det är den nivå som är mest mätbar i kvantitativa studier, medan kontextuella och diskursiva faktorer riskerar att hamna i bakgrunden.

Push-and-pull-modeller

Mellan pathway-modeller och risk-/skyddsfaktoransatser kan man placera Vergani et al. (2020) och deras *push-and-pull*-modell. Den kan ses som en vidareutveckling av riskfaktorperspektivet, men med ambitionen att syntetisera sociala, individuella och strukturella faktorer i ett ramverk som är lätt att tillämpa praktiskt. Push-and-pull-modellen bygger på en distinktion mellan två typer av drivkrafter:

- **Push-faktorer** syftar på de mekanismer som stöter bort individen från det etablerade samhället, exempelvis social exkludering, upplevd orättvisa, diskriminering, politiskt förtryck, fattigdom eller brist på sociala band.
- **Pull-faktorer** avser däremot det som drar individer in i radikala miljöer: känslan av mening och syfte genom ideologi, gemenskap och lojalitet, karismatiska ledare, status eller löften om äventyr och handlingskraft.

Modellen integrerar dessutom personliga sårbarhetsfaktorer, såsom livskriser, marginalisering eller psykologiska behov, som påverkar mottagligheten för push-and-pull-mekanismerna.

Push-and-pull-logiken har en längre historia inom migrationsforskning och social rörelseteori, och Vergani et al. anpassar den till radikaliseringsfältet för att ge en mer dynamisk bild än vad en ren riskfaktorkatalog kan erbjuda. En styrka är att modellen uppmärksammar både avvisande krafter (det som alierar individer) och attraktiva krafter (det som gör radikala miljöer tilltalande). Samtidigt har även denna ansats begränsningar: den riskerar att förenkla komplexa

processer till binära drivkrafter och fångar inte alltid de interaktiva eller sekventiella processerna som pathway-modellerna lyfter fram.

Modell	Perspektiv	Typ av modell	Användning
Moghaddam (2005): Staircase to Terrorism	Individcentrerat och psykologiskt, med fokus på perceptuella och kognitiva mekanismer som driver stegvis engagemang i våld	Heuristisk processmodell (stegvis progression)	Illustration av psykologisk förskjutning mot terrorism genom successiva stadier; används i forskning, undervisning och policy som pedagogiskt ramverk
McCauley & Moskaleiko (2008): Pyramid-modellen	Fokus på attityder vs. beteenden och sociala mekanismer som grupptryck, personlig moral och identitet; populationsperspektiv	Processmodell med nivåer (populationsfördelning)	Analys av olika grader av radikaliserings i en befolkning; används för att skilja mellan stöd för, legitimering av och faktiskt deltagande i våld
Hafez & Mullins (2015): Radicalization Puzzle	Multifaktoriell och integrativ, betonar hur strukturella, sociala och individuella faktorer samverkar och förstärker varandra	Syntesramverk (pusselmetafor)	Förståelse av radikaliserings komplexitet utan linjär tidsordning; används för att analysera konfigurationer av ömsesidigt beroende faktorer
Vergani et al. (2020): Push/pull-modell	Teoretisk syntes med fokus på samspelet mellan push-faktorer (avståndstagande från samhället) och pull-faktorer (attraktion till radikala miljöer), inklusive personliga sårbarheter	Hybridmodell av drivkrafter och mottaglighetsfaktorer	Policy- och praktisk nära analys av varför individer dras till, eller pressas in i, radikala miljöer; används i förebyggande arbete

Modell	Perspektiv	Typ av modell	Användning
Risk- och skyddsfaktorer	Identifiering av faktorer som ökar (risk) och minskar (skydd) sannolikheten för radikalisering; bygger på empirisk forskning och meta-analyser	Deskriptiv faktormodell (empiriskt kategoriseringsramverk)	Riskbedömning och prevention; används för att strukturera empiriska fynd och identifiera målområden för intervention
Bjørge & Horgan (2009): Väger ut ur våldsbejakande extremism	Fokus på utträde (disengagement) och avradikalisering, med betoning på praktiska, sociala och existentiella faktorer snarare än enbart ideologisk förändring	Processmodell för utträde och avradikalisering	Förståelse av varför och hur individer lämnar våldsbejakande extremism; vanligt verktyg i policy och exitprogram

Tabell 1. Översikt av centrala teoretiska perspektiv inom nutida radikaliseringsforskning.

Sammanfattningsvis betonar den nya generationens radikaliseringsforskning i allt högre grad sociala processer, individers aktiva meningsskapande samt berättelser och emotioner som centrala mekanismer. Detta innebär en förskjutning från tidigare mer individpsykologiska förklaringar, där radikaliseringsprocessen ofta framställdes som en linjär process eller en följd av statiska riskfaktorer, till ett fokus på hur människor själva tolkar, berättar om och känslomässigt förankrar sina erfarenheter i relation till en social och politisk omgivning.

Detta perspektiv är – som vi kommer att se i nästa kapitel – centralt för att förstå sociala mediers och internets betydelse för radikaliseringsprocessen. I denna förståelseram blir sociala medier och internet inte enbart neutrala kommunikationskanaler, utan sociotekniska arenor där identitet, tillhörighet och affekt ständigt förhandlas, förstärks och omformas. Individer möter här inte bara information, utan deltar själva aktivt i produktionen och reproduktionen av sociala gemenskaper, symboler och normer. Plattformarna fungerar därför både som informationsmiljöer och som emotionella och relationella rum.

Som vi kommer att se i nästa kapitel kan dock de digitala plattformarnas underliggande logik – genom exempelvis algoritmer som systematiskt belönar känsloladdat, polariserande och konfliktorienterat innehåll – i sig bidra till att forma radikaliseringsprocessen. Detta innebär inte att internet i sig radikaliserar, men att det fungerar som en katalysator: en miljö som kan förstärka, accelerera och fördjupa redan pågående processer genom att förändra villkoren för synlighet, interaktion och känslomässig mobilisering.

Mot denna bakgrund vänder vi oss nu till den existerande forskningen om vilken roll digitala plattformar spelar i radikaliseringsprocessen.

3. Vilken roll har internet och sociala medier?

Frågan om hur internet och sociala medier påverkar radikaliseringsprocessen har blivit alltmer central, både inom forskning och i den offentliga debatten. Trots detta råder fortfarande oenighet kring internets roll. Om radikalisering i sig är ett svåravgränsat begrepp, så är online-radikalisering än mer fragmenterat och svårfångat. Det studeras inom flera olika vetenskapsfält, ofta med delvis överlappande men osynkroniserade begrepp som *affektiv polarisering*, *online disinhibition effect*, *grooming*, *memetisk propaganda*, *gamification*, *desinformation/-misinformation*, *hatretorik* och *toxisk kommunikation*. Även om dessa begrepp fångar viktiga aspekter av digitala radikaliseringsdynamiker, behandlas de ofta inom separata forskningsfält och kopplas sällan samman eller relateras till etablerad forskning inom psykologi, sociologi eller kriminologi. Därtill görs sällan en tydlig åtskillnad mellan kognitiv och beteendemässig radikalisering, vilket försvårar jämförelser och samlade slutsatser.

I praktiken fungerar därmed onlineradikalisering som ett paraplybegrepp som rymmer en stor variation av nätbeteenden med radikaliseringspotential – från konsumtion av extremistisk propaganda till planering av våldsdåd i slutna forum.

En återkommande kritik inom fältet gäller bristen på robust empirisk forskning. I en genomgång av 200 vetenskapliga artiklar om onlineradikalisering, visade det sig att endast 6,5 procent byggde på empiriska data – och enbart 2 procent använde primärdata (Gill et al., 2017). Även om de senaste åren har sett en tydlig ökning av empiriska studier, domineras fältet fortfarande av huvudsakligen teoretisk forskning (Cottee, 2024). Majoriteten av studierna är antingen småskaliga och kvalitativa – ofta i form av fallstudier – eller bygger på anekdotisk evidens, vilket begränsar möjligheten att generalisera resultaten. Samtidigt tenderar storskaliga studier, exempelvis enkäter eller experiment, att förenkla den komplexitet som kännetecknar individers radikaliseringsprocesser. Som en följd av detta blir det svårt att dra meningsfulla slutsatser om mekanismer och utfall.

Fältet präglas även av en tydlig obalans i forskningsfokus. Merparten av studierna undersöker *tillgångssidan* – det vill säga innehållet i extremistiskt material och dess spridning – medan betydligt färre undersöker *mottagarsidan*, det vill säga hur individer interagerar med detta innehåll och vilka effekter det har. Denna skevhet förklaras delvis av metodologiska och etiska utmaningar; det är svårt att samla in tillförlitliga data om hur människor påverkas av digitala miljöer över tid, särskilt i relation till känsliga frågor som radikalisering och politiskt våld.

Sammantaget präglas forskningen om onlineradikalisering av brokiga och ibland motstridiga resultat, vilket speglar fältets metodologiska och teoretiska heterogenitet. Det saknas därför en

samlad och empiriskt förankrad bild av *hur* och *varför* internet kan bidra till radikaliseringsprocesser, samt vilka mekanismer som driver dessa processer.

I det följande avsnittet sammanfattas forskningsläget utifrån tre övergripande perspektiv: (1) internets betydelse för individers utveckling mot ökad radikaliseringsprocess, (2) digitala plattformars och mediemiljöers roll i att forma radikaliseringsdynamiker, och (3) en kort reflektion kring det aktuella forskningsläget.

3.1 Individers utveckling mot ökad radikaliseringsprocess

Forskning på individnivå fokuserar på internets roll i enskilda individers radikaliseringsprocess. Det handlar om psykologiska och sociala faktorer som kan göra individer mottagliga för extremistiska idéer, samt om hur digitala miljöer fungerar som förmedlare och förstärkare i dessa processer. Metodologiskt används både kvantitativa och experimentella studier av effekterna av att konsumera radikalt innehåll, men även fallstudier och intervjuer med gärningspersoner.

3.1.1 Statistiska och experimentella studier av medieexponering

Kvantitativ forskning om onlineradikaliseringsprocess har i hög grad förlitat sig på enkäter, registerdata och experimentella studier, ofta med syftet att undersöka hur exponering för radikalt innehåll påverkar individers attityder och benägenhet att stödja eller delta i politiskt motiverade våldsdåd. Resultaten är dock delvis motstridiga, vilket kan förklaras av variationer i definitioner, metodologiska skillnader och olikheter i datainsamling.

En central distinktion i forskningen rör skillnaden mellan *passiv* och *aktiv* mediekonsumtion. Passiv konsumtion avser situationer där individer exponeras för extremistiskt innehåll utan att själva aktivt ha sökt upp det, exempelvis via algoritmiska rekommendationer i sociala medier. Aktiv konsumtion innebär däremot ett medvetet engagemang: att individen söker upp, interagerar med eller sprider radikalt material. Denna distinktion är viktig eftersom den fångar skillnaden mellan oavsiktlig och avsiktlig exponering, vilket kan ha olika psykologiska och sociala konsekvenser.

Forskningen visar att båda formerna kan bidra till radikaliseringsprocesser, men genom olika mekanismer. Passiv exponering kan främja en gradvis normalisering av extrema idéer, särskilt vid upprepad eller subtil påverkan (Hassan et al., 2018), medan aktiv exponering oftare förstärker befintliga ideologiska övertygelser och bekräftar individens världsbild genom deltagande i ideologiskt homogena nätverk.

I en större litteraturöversikt av 53 studier, varav fyra experimentella, undersökte Wolfowicz et al. (2022) hur olika former av mediekonsumtion hänger samman med radikaliseringsprocesser. Resultaten visade att traditionella medier som tv och radio hade liten påverkan, medan digitalt innehåll – särskilt sådant som aktivt söks upp – tydligare korrelerade med radikala attityder.

Effekten var starkare när det gällde beteendemässig radikaliserings än när det gällde attitydnivåer.

Liknande resultat återfinns i Hassan et al. (2018), som visade att konsumtion av extremistiskt och våldsamt material på nätet ökade stödet för radikala ståndpunkter och politiskt motiverat våld, framför allt i högerextrema och islamistiska miljöer.

I en nyare enkätstudie från 2024 (Schumann et al., 2024), undersöktes hur olika användarmönster hänger samman med attityder till våldsbejakande extremism. Den grupp som både aktivt sökte upp och passivt exponerades för radikalt innehåll uppvisade de mest extrema attityderna och högst benägenhet att använda våld. De som inte exponerades alls hade de lägsta nivåerna av stöd för extremism. Även om det inte är förvånande – då aktivt sökande kan spegla ett redan existerande engagemang – är det viktigt att detta mönster bekräftas i systematisk forskning.

Trots metodologiska och etiska begränsningar – forskare kan inte gärna medvetet radikalisera deltagare – finns även experimentella studier inom fältet. I Wolfowicz et al. (2022) metaanalys ingår flera experiment där deltagare exponerades för radikalt material i kontrollerade former. Resultaten visade i genomsnitt en liten men märkbar ökning i stöd för radikala attityder efter exponering, särskilt hos individer med högre grad av aggressivitet eller frustration. Detta tyder på att vissa individer är mer mottagliga än andra. Samtidigt visar studierna också att de flesta som exponeras för radikalt innehåll inte radikaliseras. Effekterna finns – men de är måttliga och starkt beroende av sammanhanget.

En annan aspekt som fått ökad uppmärksamhet är den roll som sociala nätverksstrukturer och algoritmer spelar. I en studie av Wolfowicz et al. (2023) undersöktes hur informationsflöden – styrda av algoritmer – påverkar tolerans för våld. Studien visade att personer som placerades i nätverk med hög åsiktsdiversitet, där man exponerades för olika åsikter och nya kontakter, visade mindre förståelse för våldshandlingar som självmordsbombningar, jämfört med dem i mer ensidiga och algoritmstyrda nätverk. Detta antyder att så kallade ekokammare, särskilt när de förstärks av plattformars algoritmer, kan bidra till att sänka trösklar för våldsanvändning.

Sammantaget visar forskningen att det finns ett tydligt – men inte deterministiskt – samband mellan digitala mediemiljöer och radikaliserings, särskilt när individer aktivt söker upp och interagerar med extremistiskt innehåll i slutna nätverk. Samtidigt påverkas processen av individuella förutsättningar, sociala relationer och samhällsliga omständigheter. För bättre förståelse av dessa komplexa samband krävs fler longitudinella studier med fokus på orsakskedjor, interaktion och mottagarperspektiv.

3.1.2 Fallstudier av radikaliserade individer

Detaljerade fallstudier av radikaliserade individer – exempelvis ensamagerande gärningspersoner – ger en djupare förståelse av hur onlineaktiviteter samspelar med personliga och sociala faktorer i radikaliseringsprocessen. Denna typ av forskning bygger ofta på databaser med dömda eller kända terrorister, eller på intervjuer med personer som varit involverade i

extremistiska miljöer. Syftet är att kartlägga radikaliseringsvägar (pathways) och att analysera vilken roll digitala miljöer spelar i dessa processer.

I en omfattande studie av Kenyon & Baker-Beall (2022) analyserades 490 extremistiska brottslingar i England och Wales mellan 2007 och 2021. Forskarna identifierade tre huvudsakliga radikaliseringsvägar: (1) individer som huvudsakligen radikaliserats online, (2) individer som främst radikaliserats offline, samt (3) individer med en blandad väg, där online- och offlinefaktorer samspelar.

Resultaten visar att internets betydelse i radikaliseringsprocesser ökat markant över tid: andelen individer som radikaliserades via internet steg från 43 procent 2010–2012 till 92 procent 2019–2021. Studien pekar också på ett skifte i plattformsanvändning – från dedikerade extremist-webbplatser till mer svårövervakade arenor som sociala medier, krypterade appar (till exempel Telegram, WhatsApp) och spelplattformar som Discord.

Skillnader mellan grupperna framträder tydligt. De som radikaliserats online tenderar att begå fler brott i digitala miljöer och i lägre grad använda fysiskt våld. Denna grupp uppvisar också högre förekomst av psykisk ohälsa. De som radikaliserats offline är ofta äldre, har tidigare kriminell bakgrund och tenderar att begå våldsbrott i mindre grupper snarare än som ensamagerande.

En liknande analys genomfördes av Whittaker (2021) med en databas innehållande 231 IS-relaterade förövare i USA mellan 2012 och 2020. Studien visar att 92 procent av dessa använde internet, främst via mainstreamplattformar som Facebook, X och YouTube, och att 20 procent även nyttjade krypterade tjänster som Telegram och WhatsApp. Internet användes i flera syften: planering av aktiviteter (84 %), konsumtion av ideologiskt material (70 %), praktiska förberedelser inför aktioner (60 %), kommunikation med likasinnade (80 %), diskussioner som förstärkte befintliga övertygelser (55 %), samt rekrytering av nya anhängare genom att sprida extremistiskt material (50 %).

Studien visar även ideologiskt betingade skillnader: högerextrema gärningspersoner var mer än tre gånger mer benägna än al-Qaida-inspirerade att inhämta extremistiska idéer online och mer än två gånger mer benägna att kommunicera med likasinnade via digitala plattformar. Detta understryker vikten av att särskilja data efter ideologi, attackmetodik (till exempel ensamagerande kontra gruppaktioner) och roll (till exempel bombtillverkare kontra propagandister).

Enskilda fallstudier ger ytterligare nyanser. Mølmen & Ravndal (2023) analyserade tre högerextrema ensamagerande gärningsmän: Peter Mangs, Anders Behring Breivik och Dylann Roof. Studien visar att samtliga använde internet intensivt under radikaliseringsprocessen. Genom att spåra och jämföra deras utvecklingsförlopp identifierade forskarna hur digitala miljöer underlättade radikaliseringsprocessen genom att tillhandahålla propaganda och extremistisk information, förstärka grupptänkande och legitimera såväl ideologi som våld. Samtidigt framhölls offlinefaktorer – som personliga kriser och social frustration – som ökade mottagligheten för onlinepåverkan. Radikalisering framstår därmed som ett resultat av samspelet mellan digital exponering och individuella sociala sårbarheter.

Intervjubaserade studier bidrar med deltagarperspektiv. I en studie av Gaudette et al. (2022) intervjuades avhoppare från våldsbejakande högerextrema miljöer. Resultaten bekräftar bilden av en mixad radikaliseringsväg: ungefär en tredjedel av deltagarna kom först i kontakt med extremism via internet, medan majoriteten först exponerades i fysiska sammanhang – genom familj, vänner eller lokala grupper. Oavsett ingång blev dock internet en avgörande faktor för fördjupningen i extremism. Det erbjöd kontinuerlig tillgång till radikalt innehåll och möjligheten att etablera relationer och kontakter med likasinnade, vilket samverkade för att påskynda radikaliseringsförloppet. Dessa vittnesmål stöder bilden av internet som möjliggörare och accelerator i radikaliseringsprocessen.

Intressant nog visar flera studier att omfattande internetanvändning i vissa fall kan *förhindra* eller *försvåra* genomförandet av terrorattacker. Studier av Benson (2014); Jensen et al. (2020); Whittaker (2021) visar att individer med hög digital aktivitet tenderar att genomföra färre attacker, och när de försöker misslyckas de oftare. En möjlig förklaring är att digital kommunikation och innehållsspridning lämnar spår, vilket underlättar för säkerhetstjänster att övervaka och identifiera misstänkta.

Samtidigt väcker detta frågan om konsekvenserna av att extremistiskt innehåll censureras på öppna plattformar. När etablerade aktörer som YouTube, Facebook eller X tar bort material eller stänger konton finns risk att radikala användare migrerar till mer slutna krypterade miljöer – såsom och dark web-forum – som ofta är mer radikaliserade, mindre reglerade och betydligt svårare att övervaka. Detta betyder dock inte att censur saknar fördelar: att begränsa spridningen av extremistiskt innehåll på breda plattformar kan minska rekryteringsmöjligheterna och stödet från sympatisörer.

Det handlar därför om en balansgång: att begränsa tillgången till radikalt innehåll utan att samtidigt driva extremistmiljöer djupare in i slutna forum där både ideologisk radikalisering och våldspotentialen kan förstärkas. Denna avvägning diskuteras mer ingående i rapportens avslutande del om preventiva strategier.

3.1.3 Uppluckring online/offline?

Sammantaget ger forskningen om individers radikaliseringsprocesser begränsat stöd för uppfattningen att digitala interaktioner ersätter fysiska – en idé som ibland framhålls i samhällsdebatten. Tvärtom visar studier att radikaliserade individer oftast är verksamma i både digitala och analoga sammanhang. Online- och offlinemiljöer tenderar att förstärka varandra snarare än att utgöra separata sfärer. Individer som planerar extremistiska handlingar online är ofta också aktiva i fysiska miljöer – och vice versa.

Rekryteringsprocesser följer ofta denna växelverkan: en person kan först exponeras för propaganda på nätet, därefter diskutera innehållet med vänner offline och sedan återvända till digitala källor för att fördjupa och förkovra sig ytterligare. I vissa fall sker detta även kollektivt – exempelvis när grupper samlas för att tillsammans titta på propagandavideor (Baugut & Neumann, 2020).

Sådana mönster har fått forskare att ifrågasätta om det över huvud taget är meningsfullt att upprätthålla en strikt åtskillnad mellan online och offline i analysen av radikaliseringsprocesser. I praktiken följer dessa processer en *hybridlogik*, där digitala och fysiska dimensioner samspelar på olika sätt för olika individer. Det är därför mer fruktbart att analysera hur digital interaktion flätas in i människors vardagliga, fysiska livsvärldar (Herath & Whittaker, 2023), än att se cyberrymden som en isolerad domän. Digitala plattformar existerar inte fristående från sociala, kulturella och politiska kontexter – de formas av dessa sammanhang och påverkar dem i sin tur.

Samtidigt bör kritiken mot online/offline-distinktionen inte gå till överdrift. Gränsen är visserligen flytande, men det innebär inte att skillnaden saknar analytiskt värde. Som diskuteras i avsnitt 2.3 har digitala plattformar unika egenskaper – till exempel algoritmer som styr exponeringen av information – som på specifika sätt påverkar hur radikaliseringsprocesser utvecklas. För att förstå dessa mekanismer krävs därför att vi temporärt särskiljer digital och fysisk påverkan, inte som separata världar, utan som samverkande analysnivåer.

3.2 Digitala plattformar och mediemiljöer

Den digitala arenan – omfattande sociala medier, diskussionsforum, bloggar, videoplattformar och nyhetssajter – utgör idag en central miljö för radikalisering. Inom medie- och kommunikationsvetenskapen har forskningen i ökande grad fokuserat på hur plattformarnas tekniska arkitektur och användarlogik påverkar, förstärker och ibland omformar radikaliseringsprocesser.

Om tidigare sektioner i första hand behandlat *mottagarsidan* – alltså hur individer påverkas av extremistiskt innehåll – ligger fokus här på *tillgångssidan*: hur radikala idéer sprids, förstärks och får fäste genom specifika plattformsdynamiker. I praktiken är dock dessa två sidor förstådda inte helt åtskilda, eftersom personer som är involverade i exempelvis onlinediskussioner är både tillgång och mottagare för varandra.

Forskningen på tillgångssidan är ofta mer teoretiskt orienterad och syftar till att förstå hur digitaliseringen förändrar själva logiken bakom radikalisering. Medan mycket tidigare forskning, inte minst om individers radikaliseringsvägar, beskrivit internet som ett verktyg för spridning av propaganda, rekrytering och organisering, framhåller denna forskning att digitala plattformar utgör miljöer som möjliggör nya former av radikaliseringsdynamik. Det handlar alltså om mer än tekniska verktyg – det är fråga om ett potentiellt paradigmskifte där teknologisk infrastruktur, användarbeteenden och plattformslogik samverkar för att forma nya vägar till radikalisering.

En viktig förändring är språkets ökade betydelse. Tidigare manifesterades politisk identitet ofta genom fysiska markörer – exempelvis klädstil, tatueringar, eller deltagande i demonstrationer och vitmaktkonserter. I den digitala kulturen har detta delvis ersatts av *språkliga* och *diskursiva* uttrycksformer. På nätforum och sociala medier formas identitet och samhörighet främst genom konversationer: ordval, memes, intern jargong och ironiska uttryck blir centrala markörer för gruppstillhörighet.

I många digitala miljöer får språket närmast en rituellt funktion. Vissa uttryck fungerar som kodord som signalerar tillhörighet, medan andra fungerar gränsdragande gentemot ”de andra”. Extremistiskt innehåll kan dessutom paketeras i humoristisk eller ironisk form, vilket sänker tröskeln för engagemang, särskilt hos yngre användare. Det diskursiva rummet blir därmed en plats för socialisering, gränsdragning och ideologisk förankring, ofta utan att innehålla några explicita uppmaningar till våld.

En annan förändring gäller radikaliseringsorganisatoriska form. Digitaliseringen har kraftigt minskat behovet av formella organisationer. Funktioner som tidigare förutsatte en organiserad rörelse – såsom socialisering, ideologisk skolning eller taktisk planering – kan idag uppstå i löst sammansatta och ofta tillfälliga onlinegemenskaper.

Detta har långtgående konsekvenser. Tidigare kunde formella organisationer kanalisera missnöje in i kollektiva och delvis reglerade uttrycksformer – såsom demonstrationer, politiska kampanjer eller opinionsbildning – vilket samtidigt fungerade som ett slags filter. I deras frånvaro har vi sett en ökning av mer flytande, svårförutsägbara och ofta mer individualiserade former av radikalisering och politiskt motiverade handlingar.

Detta skifte har möjliggjort framväxten av fenomen som *stokastisk terrorism* och *hive terrorism*, där politiskt våld eller annan extremistisk handling inte planeras i organiserad form, utan uppstår ur ett digitalt informationsflöde präglad av hatretorik, affektivt laddad symbolik och normaliserad våldsjargong. Här fungerar språket inte som ett direkt kommando, utan som en ideologisk och emotionell infrastruktur där vissa individer – ofta utan formell anknytning till grupper – uppfattar det som moraliskt rätt eller nödvändigt att agera.

Våldet – eller andra politiskt motiverade handlingar – blir därmed svårare att förutse: statistiskt sannolikt, men med oklar tidpunkt, metod och mål. På detta sätt bidrar sociala medier inte bara till att sprida radikalt innehåll – de förändrar även *våldets karaktär*. Plattformarnas design, interaktionsmönster och algoritmiska logik skapar en resonansstruktur där vissa idéer inte bara får spridning, utan också emotionellt gensvar och upprepade förstärkning. Det innebär att radikaliseringsprocesser inte bara går snabbare – de kan också bli mer improviserade, affektiva och svårare att motverka med traditionella strategier.

3.2.1 Ekokammare och filterbubblor

Tidiga studier inom medieforskning pekade på att internet och sociala medier kan driva på polarisering och radikalisering genom att skapa slutna informationsbubblor – så kallade *ekokammare* – där individer omger sig med likasinnade och nästan uteslutande exponeras för åsikter som bekräftar deras egna (Sunstein, 2002, 2007, 2008). När individer isoleras från motstridiga idéer och alternativa perspektiv minskar möjligheten till kognitiv friktion, vilket enligt denna teori ökar risken för självradikalisering. Åsikter och världsbilder cementeras genom ständig upprepning, medan bristen på ifrågasättande förstärker känslan av att ”alla tycker så här”. Resultatet kan bli en successiv förskjutning av individens attityder, där radikala idéer successivt internaliseras och upplevs som legitima (McPherson et al., 2001; Sunstein, 2007).

Ett närliggande begrepp är *filterbubbla*, populariserat av Eli Pariser (2011). Det syftar på hur personaliseringsteknik, exempelvis rekommendationsalgoritmer på YouTube eller Facebook, tenderar att visa användare innehåll som överensstämmer med deras tidigare beteenden. Om en person börjar klicka på invandringsfientliga videor kan plattformen gradvis presentera mer liknande innehåll, vilket kan förstärka och radikalisera redan existerande attityder. Filosofen Mark Alfano benämner denna process ”technological seduction” – där tekniska system subtilt knuffar användaren i en viss riktning genom sin design (Alfano et al., 2018). Effekten är särskilt märkbar i ett tidigt skede av radikaliseringsprocessen, då individer är mottagliga för nya idéer men saknar en utvecklad kritisk distans.

3.2.2 Ekokammare eller skyttegravskrig?

Flera studier har försökt kartlägga algoritmers roll i radikaliseringsprocesser, särskilt på YouTube, som ofta pekats ut som en radikaliseringsmaskin. Denna forskning är dock metodologiskt utmanande. Algoritmer förändras ständigt och är ofta affärshemligheter, vilket gör deras inre logik svår att studera direkt. Forskarna tvingas därför att analysera användarbeteenden och innehållsflöden för att indirekt dra slutsatser om algoritmers effekter.

En systematisk översikt av Yesilada & Lewandowsky (2022), som analyserade 23 studier om YouTubes rekommendationssystem, visade att 14 av dem indikerade att algoritmen underlättar vägen till mer extremistiskt innehåll. Sju studier fann blandade resultat, medan två inte fann något samband alls. Sammantaget antyder översikten att algoritmiska rekommendationer *kan* bidra till radikalisering, men att evidensen är varierande.

Trots att begreppen ekokammare och filterbubblor blivit centrala för att förstå onlineradikalisering, har de också kommit att ifrågasättas i takt med att nyare empiriska studier nyanserat bilden. Teorin bygger på två antaganden: (1) att användare främst exponeras för likasinnade perspektiv, och (2) att denna ensidighet i sig leder till radikalisering. Båda har visat sig vara förenklade.

För det första visar ny forskning att digitala radikala miljöer sällan är helt isolerade. Politiskt innehåll cirkulerar ofta mellan olika grupper och plattformar, vilket innebär att argument och idéer ofta sprids utanför de egna kretsarna (Bail, 2022; Bail et al., 2018; Dubois & Blank, 2018; Jungherr et al., 2020). Facebookexperiment har visat att algoritmer kan öka både in- och utgruppsinnehåll i användarnas flöden, om än med en viss slagsida mot det egna lägret (Nyhan et al., 2023). Internet kan därmed även bidra till *ökad* mångfald av politisk exponering. Samtidigt kan algoritmernas rankning ändå prioritera polariserande och känsloladdat material inom respektive läger. Här spelar kvalitativa skillnader i innehållet roll – det är skillnad på att se ett sakligt nyhetsinslag om en motståndarideologi och att se ett hånfullt utspel om densamma.

För det andra är det oklart huruvida ensidig informationskonsumtion automatiskt leder till radikalisering – och om kontakt med motstridiga perspektiv nödvändigtvis har en dämpande effekt. Tvärtom har flera studier visat att strategier för att öka exponering för oppositionella åsikter ibland får motsatt effekt (Bélanger et al., 2020; Lewandowsky et al., 2012). Chris Bail och hans forskargrupp (Bail et al., 2018) visade i ett experiment med 1 200 X-användare att

republikanska deltagare som exponerades för demokratiska inlägg på plattformen blev mer konservativa än kontrollgruppen. Ju mer uppmärksamhet de ägnade åt innehållet, desto starkare blev effekten. Många deltagare upplevde exponeringen som ett hot mot sin identitet.

Liknande resultat har noterats inom forskning om avradikaliseringsstrategier, såsom *counter messaging* och *debunking*. Studier har visat att försök att rätta felaktig information, exempelvis myten att vaccin orsakar autism, kan förstärka oron hos individer som redan tror på ryktet eller konspirationen och därmed minska deras benägenhet att vaccinera sina barn (Berinsky, 2017; Nyhan et al., 2014). Detta tyder på att exponering för alternativa perspektiv ibland förstärker, snarare än minskar, radikaliserade övertygelser – särskilt när dessa upplevs som identitetshotande. Samtidigt bör det noteras att många av dessa studier mäter kortsiktiga effekter. Det är därför möjligt att debunkingstrategier fungerar bättre på längre sikt, och att polarisering i grunden drivs av långvarig ensidig exponering snarare än av att konfrontation i sig misslyckas.

I ljuset av detta har vissa forskare föreslagit att vi snarare bör tala om en form av skyttegravsdynamik (*trench warfare*) än om ekokammare (Karlsen et al., 2017; Törnberg & Törnberg, 2024; Törnberg & Wahlström, 2018). Där ekokammare bygger på isolering, bygger skyttegravsdynamik på negativ och konfliktpräglad interaktion mellan ideologiska läger. Deltagarna söker inte konsensus, utan bekräftelse genom konfrontation. De ekon som uppstår handlar mindre om att förstärka den egna gruppens världsbild och mer om att marginalisera motståndarna. Som stöd för denna hypotes visar flera studier att högerextrema grupper ofta interagerar med innehåll från politiska motståndare – inte för att ta till sig deras idéer, utan för att kritisera och håna dem. Dessa aktörer länkar flitigt till traditionella nyhetsmedier, sociala medier och offentliga uttalanden för att peka på vad de uppfattar som hyckleri eller inkonsekvens hos meningsmotståndaren (Haller & Holt, 2019; Heft et al., 2021; Törnberg & Nissen, 2022; Törnberg & Vallström, 2025).

Sammanfattningsvis har teorier om ekokammare och filterbubblor spelat en viktig roll för att belysa hur internet kan förstärka extremism. Men nyare forskning pekar på att dessa modeller är alltför förenklade. Radikalisering online är snarare en komplex process där teknologiska, sociala och psykologiska faktorer samverkar på dynamiska sätt – inte alltid genom isolering, utan ofta genom konflikt.

3.2.3 Plattformsaffordanser

Även om teorin om internet som en sluten ekokammare har begränsat empiriskt stöd, visar forskningen tydligt att digitala plattformar påverkar radikaliseringsprocesser på flera sätt. Det handlar inte bara om vilket innehåll användare ser, utan också om hur plattformarnas design och underliggande arkitektur strukturerar interaktionen och möjliggör spridning av innehåll.

Begreppet *affordans* (eng. *affordance*) syftar på de handlingsmöjligheter som en viss teknologi eller plattform erbjuder sina användare. I digitala miljöer omfattar detta bland annat möjligheterna att vara anonym, nå ut till breda publikgrupper och skapa gemenskaper. Sociala

medier har ett antal designegenskaper – exempelvis anonymitet, synlighet, nätverkskopplingar och spridningsmekanismer – som kan utnyttjas för att förstärka och sprida radikala budskap.

I en översikt av digitala miljöer som Telegram, WhatsApp och X identifierade Schulze et al. (2024) sjutton olika affordanser kopplade till radikaliseringsprocesser online. Tre av dessa framstod som särskilt centrala: anonymitet, synlighet och kollektivitet. Anonymitet minskar risken för sociala sanktioner och gör det möjligt att uttrycka extrema åsikter utan omedelbara konsekvenser. Synlighet – genom hashtags, delningsfunktioner och öppna grupper – underlättar spridning av extremistiskt innehåll till breda publikgrupper. Kollektivitetsskapande funktioner, såsom diskussionsgrupper, chattgrupper och kommentarstrådar, skapar sociala miljöer där likasinnade kan hitta varandra och förstärka varandras övertygelser.

Plattformarnas tekniska affordanser kan alltså skapa miljöer som gynnar radikalisering – inte genom aktiv styrning, utan genom att underlätta socialisering, normalisering och ideologisk förstärkning. Effekterna uppstår ofta gradvis, men kan på sikt ha stor betydelse för individers attitydförändringar och grupptillhörighet.

3.2.4 Engagemangsdrivna algoritmer och algoritmisk förstärkning

En annan central aspekt av plattformars inverkan på radikalisering rör de algoritmer som avgör vilket innehåll användare exponeras för. De flesta stora medieplattformarna använder engagemangsdrivna algoritmer som är optimerade för att maximera användaraktivitet. Detta inkluderar interaktioner som gillamarkeringar, delningar, kommentarer och visningstid.

Denna optimering har dock en baksida: det som genererar engagemang är inte nödvändigtvis det som främjar saklighet och måttfullhet. Tvärtom premierar algoritmer ofta innehåll som är sensationellt, kontroversiellt eller emotionellt laddat. I en uppmärksam essä observerade tekniksociologen Zeynep Tufekci att YouTube:s rekommendationssystem tenderar att föreslå allt mer extrema videoklipp efter att en användare börjat konsumera politiskt innehåll (Tufekci, 2018). Hon liknade algoritmen vid en restaurang som serverar allt sötare och fetare rätter – när användaren vant sig vid en viss nivå av stimuli, ökar algoritmen intensiteten för att behålla uppmärksamheten.

Förklaringen är enkel: innehåll som provocerar, upprör eller chockerar genererar fler klick, längre visningstid och fler interaktioner – vilket i sin tur gynnar plattformens affärsmodell. Särskilt innehåll som väddar till moraliskt laddade känslor, såsom ilska, rädsla eller indignation, tenderar att spridas snabbt.

Fenomenet är inte unikt för YouTube. I en experimentell studie av X:s algoritmiska flöde visade Milli et al. (2023) att det algoritmstyrda innehållet i högre grad än ett kronologiskt flöde lyfte fram politiska tweets med uttryck för ilska och fientlighet mot utgrupper. Deltagare som exponerades för detta flöde rapporterade efteråt högre nivåer av negativa känslor, inklusive irritation och ilska mot politiska motståndare. I en annan omfattande studie av miljontals inlägg på X visade Brady et al. (2017) att varje ytterligare ord som signalerade moraliskt laddad vrede ökade sannolikheten för att ett politiskt inlägg skulle retweetas med 17–20 procent.

Även interna dokument från plattformsföretagen visar att dessa mönster är kända. En läckt Facebook-rapport från 2016 visade att 64 procent av användarnas anslutningar till extremistiska grupper skedde via plattformens egna rekommendationssystem (Lauer, 2021). Rapporten konstaterade att ”rekommendationssystemet förvärrar problemet” genom att leda användare mot mer extrema nätverk. Detta fenomen, ofta benämnt *algoritmisk amplifikation*, innebär att plattformarnas design oavsiktligt – men systematiskt – förstärker spridningen av politiserande och radikaliserande innehåll.

När användare engagerar sig i en viss typ av material, föreslår algoritmen mer av samma slag. Detta skapar en självförstärkande återkoppling där extremistiska narrativ får oproportionerligt stor räckvidd. Det är inte en medveten propagandaapparat – men effekten kan i vissa fall likna just detta.

3.2.5 Ilskans ekosystem

En konsekvens av de mekanismer som tidigare beskrivits är att sociala medier kan skapa ett ”ilskans ekosystem”, där algoritmisk förstärkning successivt drar in användare i alltmer polariserade och affektivt laddade informationsmiljöer. Ribeiro et al. (2020) dokumenterade denna typ av progression på YouTube, där användare som konsumerade högerpopulistiskt eller anti-etablissemangsinnehåll ofta – via algoritmiska rekommendationer – gled över i mer högerextrema kanaler. Plattformens sekventiella rekommendationslogik innebar att användarna gradvis exponerades för alltmer radikaliserat innehåll, utan att själva aktivt ha sökt efter det.

Radikalhögern har visat sig särskilt skicklig på att utnyttja denna virala logik. Budskapen kännetecknas ofta av starkt emotionellt laddad retorik, där negativa känslor riktas mot utpekade fiender eller utgrupper – såsom invandrare, politiska motståndare eller religiösa minoriteter. Denna affektiva kommunikationsstil är särskilt väl anpassad till algoritmernas logik, som belönar provocerande innehåll. Wahlström et al. (2021) visar exempelvis att våld retorik särskilt ofta uppstår i samband med narrativ om kriminalitet där (föreställda) invandrare pekats ut som förövare och där offren framställs som särskilt sårbara, som oskyldiga barn, djur och kvinnor. Sådana berättelser kopplas ofta till en problembeskrivning av staten som korrupt och oförmögen att hantera situationen, samt till krav på deportationer. Den avhumaniserande retoriken fungerar både legitimerande för politiskt våld och känslomässigt mobiliserande, genom att väcka negativa affekter som i sin tur kan öka motivationen till våldshandlingar.

Studier av affektiv polarisering visar att när politisk kommunikation domineras av avsky, vrede och misstro gentemot motståndarsidan, fördjupas den känslomässiga klyftan mellan grupper. Diskussionerna rör sig då från sakfrågor till att handla om att hata ”de andra”. Som tidigare visats (till exempel Milli et al., 2025) tenderar engagemangsdrivna algoritmer att förstärka just sådana affektiva uttryck.

Samtidigt pågår en debatt i forskarvärlden om huruvida algoritmisk radikalisering är en oundviklig konsekvens av plattformarnas design, eller om det främst är användarnas egna val och benägenheter som driver utvecklingen. Flera studier pekar på att individer som aktivt söker upp extremt innehåll ofta redan har en förförståelse eller motivation som gör dem mottagliga

(Chen et al., 2023; Haroon et al., 2023; Ribeiro et al., 2020). Detta tyder på att *självselektion* ofta är en viktigare drivkraft än algoritmisk manipulation.

En mer nyanserad slutsats är därför att plattformarna sällan radikaliserar från ett nolläge, men att de skapar en miljö där mottagliga individer lättare leds vidare in i extremistiska sammanhang.

3.2.6 Skillnader mellan plattformar

Digitala plattformar är inte homogena utan skiljer sig markant i sina tekniska affordanser och algoritmiska strukturer. Dessa skillnader påverkar både hur radikalt innehåll sprids och vilken typ av genomslag det får.

Sociala medier som Facebook och X bygger på personliga nätverk och delningsfunktioner. Nyhetsflödesalgoritmer och grupprekommendationer spelar en central roll för hur användare exponeras för innehåll. På X kan trendalgoritmer och följ-förslag lyfta fram kontroversiella profiler eller polariserade hashtags, vilket gör att extremistiskt material kan spridas både viralt och genom befintliga nätverk.

Videoplattformar som YouTube och TikTok använder rekommendationssystem som kontinuerligt formar flödet av innehåll baserat på användarinteraktion. YouTube's *Up Next*-funktion och autouppspelning skapar en sekventiell logik där varje klick påverkar nästa rekommendation. Detta kan leda till en kaninhålleffekt, där användare gradvis leds bort från mainstreaminnehåll. TikToks hyperpersonliga *For You*-flöde går ännu längre: med hjälp av AI anpassas innehållet snabbt efter användarens reaktioner, vilket möjliggör en accelererad förskjutning mot extremism om användaren visar intresse för radikalt material (Shaw, 2023). Flera analyser har visat att redan efter några timmars exponering kan en användares TikTok-flöde domineras av innehåll med konspirationsteorier, hypermaskulinitet och våldsbejakande budskap. Efter kritik har vissa plattformar infört vissa policyändringar. YouTube har exempelvis justerat sina rekommendationssystem för att minska spridningen av radikaliserande innehåll.

Diskussionsforum och alternativa plattformar som Reddit, 4chan/8chan, Gab och Parler fungerar ofta utan avancerade rekommendationsalgoritmer. Här navigerar användare manuellt genom trådar och kategorier, och affordanser som anonymitet och brist på moderering är särskilt avgörande. 4chan och 8kun har blivit viktiga arenor för högerextrema och konspirationsteoretiska subkulturer (bland annat QAnon), just på grund av total anonymitet och tolerans för extremistiskt innehåll. Reddit har ett mer komplext system, där användarbetyg (*upvotes/downvotes*) fungerar som en form av kollektiv eller *crowdsourcing*-styrd algoritm. Vissa *subreddits* – som r/The_Donald – radikaliserades kraftigt innan de stängdes, genom att tusentals användare kollektivt röstade upp extremistiskt innehåll så att det dominerade gruppens flöde, i praktiken en användardriven amplifikation.

Meddelande- och chatappar som WhatsApp, Telegram och Discord saknar öppna algoritmiska flöden, men erbjuder andra affordanser som är relevanta för radikalisering. Telegram kombinerar anonymitet med möjlighet att bygga stora publika kanaler där meddelanden kan nå tiotusentals följare. Plattformens end-to-end-kryptering och brist på moderering gör den

attraktiv för extremistiska grupper som vill undgå övervakning. Liknande gäller slutna Facebook-grupper och Discord-servrar: även om de inte är öppna, möjliggör de spridning och förstärkning av extremt innehåll i avskärade miljöer, där ideologisk radikaliseringsprocesser kan ske ostört.

Sammanfattningsvis fyller olika plattformar olika funktioner i det digitala extremism-ekosystemet. Öppna, algoritmstyrda plattformar fungerar ofta som spridningsmotorer där nya sympatisörer nås via engagemangslogik och möjligheter till viral spridning. Mer slutna miljöer fungerar däremot som fördjupningskammare, där redan övertygade individer kan radikaliserats ytterligare i avsaknad av motstånd eller insyn. Dessa miljöer samverkar i praktiken och skapar en dynamisk infrastruktur för digital radikaliseringsprocesser.

3.2.7 Gamification

En aspekt som fått ökad uppmärksamhet under senare år är *gamification* – eller spelifiering – det vill säga användningen av spelelement såsom poäng, nivåer, belöningar och tävlingsmoment i icke-spelrelaterade sammanhang för att öka användarengagemang och motivation. Digitala plattformar har i allt högre grad integrerat sådana aspekter, och forskare har börjat undersöka hur detta påverkar politisk mobilisering, extremistiska rörelser och radikaliseringsprocesser online.

Forskningsfältet är dock fortfarande ungt och präglas av betydande metodologiska utmaningar. Dels är det svårt att empiriskt fastställa kopplingar mellan spelifierade mekanismer och faktisk radikaliseringsprocesser, dels är fenomenet ofta dolt i kodade uttryck, intern jargong eller slutna onlinemiljöer. Sammantaget innebär detta att området är underutforskat – även om indikationer på betydande effekter finns.

På många sociala plattformar används spelifiering som en integrerad del av användarupplevelsen. Gilla-markeringar, delningar och följarsiffror fungerar som belöningssystem som förstärker användarbeteenden, medan *badges*, troféer eller poängsystem – som på Reddit, Stack Overflow eller Discord – skapar sociala hierarkier, lojalitet och tillhörighet. När dessa mekanismer samspelar med extremistiska kulturer kan de bidra till att normalisera radikala idéer, belöna hatisk retorik och uppmuntra till konkurrens i spridandet av propaganda.

Vissa extremistiska rörelser har medvetet utnyttjat dessa mekanismer. Inom högerextrema och incel-relaterade miljöer har forskare dokumenterat hur användare tävlar i att skapa virala memes, ironiska referenser till våld eller extremistiska utmaningar – vilket skapar en miljö där våldsbejakande idéer gradvis trivialiseras och förstärks (Kallio & Törnberg, 2025). Denna typ av spelifiering kan även sänka tröskeln till fysiskt våld genom att rama in våldsdåd som en sorts strävan efter *high score* – en digital merit snarare än ett brott.

Sådana tendenser har bland annat observerats inom den så kallade No Lives Matter-rörelsen och i den digitala våldssekten 764, där mycket unga användare hetsar varandra till att begå till synes slumpmässiga, brutala våldsdåd. I flera fall har gärningsmän spelat in sina attacker i ett förstapersonsperspektiv som påminner om *Let's Play*-videor från datorspel. Dessa filmer sprids

därefter i onlinegemenskaper där de diskuteras och används som inspiration för framtida attacker (Lakhani & Wiedlitzka, 2023; Schlegel & Kowert, 2024).

Vissa grupper har dessutom gått längre genom att producera eller modifiera datorspel i ideologiskt syfte. Exempelvis utvecklade Al-Qaida en version av Battlefield 2 där spelaren intar rollen som jihadistisk krigare. IS har använt sekvenser från Grand Theft Auto och Call of Duty i propagandafilmer riktade till unga rekryter, och lanserade även ett eget mobilspel där användaren kunde träna i en virtuell jihadistvärld. Högerextrema grupper har gjort liknande försök: den amerikanska National Alliance tog fram ett *first-person shooter*-spel där spelaren dödar personer från minoritetsgrupper, och den accelerationistiska gruppen Atomwaffen Division har spridit spel som simulerar rasistiska attacker och ”rasrevolutioner”.

Trots dessa exempel tyder befintlig forskning på att syftet med spelifiering sällan är direkt rekrytering, utan snarare normalisering, socialisering och ideologisk förstärkning inom redan existerande miljöer. Genom meningsfulla belöningssystem och gruppgemenskap förstärks lojalitet och inbördes validering inom rörelsen (Davey, 2021; Vaux et al., 2021).

Sammantaget är spelifieringens roll i radikaliseringsprocessen ett angeläget men fortfarande underbelyst forskningsområde. Här finns ett behov av nya metodologiska verktyg för att förstå hur spelelement används i komplexa digitala kulturer – och hur de kan påverka individers benägenhet att först internalisera och senare agera på extremistiska idéer.

3.3 Sammanfattning

Forskningen visar med allt större tydlighet att digitala plattformars design – både deras tekniska affordanser och algoritmiska logik – spelar en avgörande roll i spridningen av radikalt och extremistiskt innehåll online. Funktioner som anonymitet, global sammankoppling och låg tröskel för delning har skapat nya möjligheter för extremistiska aktörer att nå ut, bilda gemenskaper och förstärka sina idéer. Samtidigt har de stora plattformarnas engagemangsdrivna algoritmer, designade för att maximera användaraktivitet och plattformstid, visat sig systematiskt prioritera innehåll som är polariserande, emotionellt laddat och ofta negativt – det vill säga samma typ av innehåll som återfinns i extremistisk retorik och propaganda.

Denna dynamik skapar självförstärkande återkopplingar där kontroversiellt innehåll får oproportionerligt stor synlighet, och där användare som visar visst intresse för sådant innehåll snabbt kan dras in i alltmer extrema idévärldar. Studier av bland annat X, Facebook och YouTube pekar tydligt på att algoritmerna tenderar att förstärka affekter som ilska och hat i flödena, samt att de aktivt rekommenderar grupper, konton eller videor som är mer extrema än vad användaren själv initialt sökt efter.

Det är dock viktigt att understryka att merparten av denna forskning har fokuserat på kognitiv radikaliseringsprocessen – det vill säga förändringar i attityder, åsikter och ideologiska övertygelser. Betydligt färre studier har lyckats belägga en direkt koppling mellan kognitiv radikaliseringsprocessen online och faktisk beteenderadikaliseringsprocessen, såsom deltagande i våldsdåd. Idén att digital konsumtion av extremt innehåll *automatiskt* leder till fysisk radikalism saknar i dagsläget robust

empiriskt stöd. Detta visar på ett metodologiskt glapp i forskningen: att spåra och verifiera kausala samband mellan digitala influenser och konkreta våldshandlingar är både etiskt och tekniskt svårt, vilket lämnar viktiga frågor öppna.

Radikalisering bör därför fortsatt förstås som en mångfasetterad och situerad process, där digitala teknologier fungerar som möjliggörande och accelererande faktorer, snarare än som autonoma orsaker. Individens mottaglighet, sociala omständigheter, erfarenheter offline och närvaron av ideologiska entreprenörer spelar alla centrala roller. Internet och algoritmer hjärntvättar inte passiva användare, men de skapar förutsättningar för en gradvis kognitiv förskjutning – genom ständig exponering, normalisering och social bekräftelse – som i kombination med andra faktorer kan bidra till radikaliseringsprocesser.

Det finns också ett växande intresse för att förstå hur teknik och psykologi samverkar: plattformarnas algoritmer belönar engagemang, och engagemang främjas av affektstarkt och ofta extremistiskt innehåll. Denna dynamik skapar ett slags förstärkning av förstärkningen, som både radikaliserar diskussioner och breddar extremistiska aktörers räckvidd i samhällsdebatten.

Samtidigt har medvetenheten om dessa problem vuxit, inom både forskarsamhället och tekniksektorn. Initiativ för att motverka skadliga effekter – såsom förändringar i algoritmiska mål, förbättrad innehållsmoderering, transparens kring prioriteringsmekanismer och design för så kallad meningsfull interaktion – är idag föremål för intensiv debatt. Än så länge är dock insatserna ojämnt implementerade, och forskningen om deras långsiktiga effekter är fortfarande i sin linda.

Sammanfattningsvis ger den samlade litteraturen en tydlig signal: digitala plattformars tekniska och affektiva arkitektur gynnar systematiskt det mest känslöväckande, polariserande och ofta extremistiska innehållet. Denna insikt utgör en central utgångspunkt för vidare forskning, policyutveckling och samhällsreflektion kring digitaliseringens roll i samtida radikalisering.

4. Preventiva strategier

Medvetenheten om att digitala miljöer kan fungera som drivkrafter i radikaliseringsprocesser har under det senaste decenniet lett till ett ökat fokus på preventiva strategier. Dessa åtgärder syftar i huvudsak till att minska tillgången till extremistiskt material, samt att på olika sätt påverka användares attityder och tolkningar. Strategierna rör sig därmed både på strukturell nivå – exempelvis via innehållsmoderering och algoritmisk justering – och på diskursiv nivå, genom faktagranskning, kontra-narrativ och strategisk kommunikation.

Forskningen på området är fortfarande relativt ung och präglas av betydande metodologiska utmaningar. Effekterna av specifika ingrepp är svåra att isolera, data är ofta svåråtkomliga och insatserna utspelar sig i snabbt föränderliga miljöer. I det här kapitlet presenteras några av de mest studerade preventiva spåren: innehållsreglering, faktagranskning och kontra-narrativ, samt algoritmiska åtgärder.

4.1 Innehållsreglering

En av de mest etablerade metoderna för att motverka spridningen av våldsbejakande extremism på nätet är att identifiera, moderera och i vissa fall ta bort innehåll eller användare. Under 2010-talet har stora teknikplattformar som Facebook, YouTube och X successivt ökat sina insatser, ofta efter påtryckningar från stater och mellanstatliga aktörer. Ett centralt initiativ är EU:s Internet Forum (2016), där teknikföretag, myndigheter och forskare samarbetat för att effektivisera så kallade *takedowns* av extremistiskt material.

Syftet med denna strategi är att begränsa extremistiska aktörers möjligheter att sprida propaganda, rekrytera nya anhängare och organisera sig digitalt. Plattformarna använder idag avancerade tekniska lösningar som maskininlärning och automatisk bild- och videogranskning, i kombination med mänskliga moderatorer, för att upptäcka och stoppa innehåll som bryter mot reglerna.

Trots dessa insatser kvarstår flera grundläggande utmaningar. Det råder ofta oklarhet kring definitionerna av vad som utgör extremistiskt innehåll, vilket försvårar tillämpningen. De automatiserade systemen är även oprecisa: de riskerar att både missa relevant material och felaktigt censurera legitima uttalanden. Dessutom uppstår lätt en *whack-a-mole*-effekt: när ett konto stängs eller innehåll tas bort, dyker det ofta upp igen på andra plattformar eller i modifierad form. Ett uppmärksammat exempel är Christchurch-attentatet 2019, där gärningsmannen filmade och lade upp massmordet i realtid. Innehållet spreds blixtnabbt vidare trots plattformarnas försök att begränsa distributionen. En relaterad risk är att användarna börjar

använda mer subtila budskap, visuella signaler och kodord (så kallade *dog whistles*) för att kringgå modereringssystemen (Åkerlund, 2022).

En alternativ strategi är att stänga ned hela användarkonton som är kopplade till extremism. Effekterna är dock blandade. Flera studier visar att detta försvårar rekrytering och spridning via mainstreamplattformar, men också att det tenderar att driva de mest engagerade användarna till alternativa, mindre reglerade plattformar som Telegram, Gab och Parler. Dessa miljöer präglas ofta av högre grad av ideologisk renhet, lägre transparens och sämre moderering – faktorer som kan öka graden av radikaliserings och risken för våld.

Exempelvis visade Berger & Perez (2016) att antalet IS-relaterade konton på X minskade dramatiskt efter insatser mot terrorismpropaganda – från mellan 46 000 och 70 000 år 2014, till endast 3 000 konton ett år senare. Samtidigt noterades en ökad aktivitet bland jihadistiska konton inom slutna grupper på Telegram (Clifford & Powell, 2019). Detta försvårar både för forskare som studerar dessa miljöer och för säkerhetstjänster som försöker kartlägga extremistiska nätverk.

En ytterligare risk är att repressiva åtgärder kan förstärka känslan av förföljelse eller martyrskap inom radikaliserade grupper. Narrativet om att ”systemet tystar oss” är vanligt i såväl jihadistisk som högerextrem propaganda, och ingripanden riskerar därmed paradoxalt nog att bekräfta och förstärka de föreställningar som driver radikalisering.

Det är dock inte entydigt negativt. Viss forskning visar att innehållsreglering också kan bidra till att förändra normer på plattformarna. En studie av Chandrasekharan et al. (2017) visade att när Reddit stängde ned särskilt hatiska underforum, minskade både förekomsten av hatretorik på plattformen i stort och den individuella benägenheten hos användare att uttrycka sig hatiskt. Det tyder på att reglering inte bara påverkar spridning utan även användarnas beteende och diskursklimatet som helhet.

4.2 Faktagranskning och motberättelser

En annan vanlig strategi i det förebyggande arbetet mot radikalisering online är att bemöta extremistiska budskap med fakta, alternativa perspektiv eller så kallade motberättelser (*counter-narratives*). Dessa insatser kan syfta till att demaskera falska påståenden, ifrågasätta centrala idéer inom en viss ideologisk diskurs eller erbjuda positiva exempel som undergräver stereotypa fiendebilder och konspiratoriska föreställningar. Strategin vilar ofta på antagandet att radikalisering drivs av ensidig information i slutna miljöer – och att det därför är möjligt att påverka individers tänkande genom att tillföra nyanser, ifrågasättanden och emotionellt engagerande alternativ.

Trots populariteten hos denna typ av strategier – särskilt inom perspektiv som P/CVE (*preventing/countering violent extremism*) – är forskningsläget osäkert. Flera översikter visar att det är svårt att utvärdera effekten av motberättelser på djupare kognitiva och affektiva förändringar hos mottagarna (Alava et al., 2017; Bélanger et al., 2020; Brouillette-Alarie et al., 2022; Rosand & Winterbotham, 2019). Utvärderingar baseras ofta på ytliga mått som visningar,

klick, eller interaktioner – indikationer som inte nödvändigtvis säger något om attitydförändring, än mindre om beteendepåverkan.

En metaanalys av 19 studier (Carthy et al., 2020) visar att vissa typer av motberättelser kan ha effekt, särskilt när de förmedlas av en trovärdig avsändare eller bygger på så kallade *counter-stereotypical exemplars* – individer eller berättelser som utmanar befintliga fiendebilder. Effekten är störst hos mottagare i ett tidigt stadium av radikaliserings. Hos mer ideologiska övertygade individer är effekten däremot ofta liten eller obefintlig.

I vissa fall kan sådana insatser till och med vara kontraproduktiva. Om budskapet uppfattas komma från en fientlig eller maktbärande källa – exempelvis staten, forskare eller etablerade medier – riskerar det att förstärka känslan av konspiration och utanförskap (Bélanger et al., 2020; Braddock & Horgan, 2016). Detta fenomen, ofta kallat *backfire-effekten*, har dokumenterats i flera sammanhang, bland annat i studier av vaccinationsmotståndare (Nyhan & Reifler, 2010; Nyhan et al., 2014). Exponering för vetenskapligt korrekta fakta om vaccins säkerhet kunde i vissa fall göra personer ännu mer skeptiska. Liknande mönster har noterats i extrema politiska grupper.

Dock har senare forskning nyanserat bilden av *backfire-effekten* och pekat på att dess förekomst är mindre vanlig än tidigare antagits (Swire-Thompson et al., 2020).

Ett närliggande angreppssätt är så kallad *redirection*, där användare som söker efter extremistiskt innehåll omdirigeras till motnarrativ eller avradikaliserande material. Ett exempel är Googles *redirect method*, där användare som söker efter jihadistrelaterade termer på YouTube i stället visas videor med exempelvis avhoppare som kritiserar extremism. En utvärdering (Helmus & Klein, 2018) visar att metoden nådde stor spridning, men att endast 2–3 procent av användarna klickade på innehållet – vilket väcker frågan om strategin verkligen når sin tänkta målgrupp. Det är sannolikt att just de mest radikaliserade individerna är de minst mottagliga för denna typ av motkommunikation.

4.3 Plattformsåtgärder: algoritmisk design och teknisk arkitektur

En annan väg att begränsa radikaliserings genomslag online är att modifiera den tekniska och algoritmiska infrastruktur som formar användares informationsflöden. Särskilt har forskare och teknikexperter diskuterat möjligheten att omformulera algoritmernas logik – från att enbart maximera engagemang till att även prioritera innehållets kvalitet och sociala konsekvenser. Bakgrunden är välkänd: starka känslor såsom ilska och hat tenderar att generera högre engagemang, vilket i sin tur belönas av algoritmernas uppmärksamhetslogik.

En rad förslag har lagts fram för att mildra detta. Detta inkluderar så kallade friktionsmekanismer, exempelvis varningar innan kontroversiellt innehåll delas, eller fördröjd delning där användaren måste bekräfta att innehållets har lästs innan det kan spridas vidare. Andra förslag handlar om avpersonifierade flöden eller algoritmisk diversitet, där innehåll inte

anpassas utifrån tidigare beteenden, utan snarare slumpmässigt blandar åsikter och perspektiv från olika håll.

Vissa plattformar, som YouTube och Facebook, har experimenterat med sådana lösningar, men i regel i begränsad skala. Som har diskuterats ovan pekar flera studier på att plattformarnas grundläggande affärsidé bygger på att maximera tid, engagemang och visningar, vilket skapar ett strukturellt incitament att lyfta fram känslomässigt laddat innehåll (Ribeiro et al., 2020; Tufekci, 2018).

En annan diskussion rör anonymitetens roll. En del forskare och beslutsfattare har föreslagit att minska möjligheten till anonyma konton genom identitetsverifiering eller multifaktorautentisering. Resonemanget bygger på teorin om *online disinhibition effect* (Demetriou & Silke, 2003): avsaknaden av ansvar och synlig identitet sänker tröskeln för aggressivt beteende.

Förslaget är dock omstritt. Kritiker menar att ökad identifiering kan hota yttrandefriheten, särskilt i repressiva stater där anonymitet är avgörande för opposition, aktivism och visselblåsning. Plattformar som Telegram har i flera fall spelat en central roll för demokratiörelser, exempelvis under den arabiska våren eller i samband med avslöjanden som Panamadokumenten. Ett generellt förbud mot anonymitet riskerar därför att underminera legitima former av protest och demokratisk organisering.

4.4 Digital resiliens och medieläskunnighet

I kontrast till tekniska och repressiva insatser har det under senare år vuxit fram en mer kapacitetsorienterad strategi som fokuserar på att stärka individers och samhällens motståndskraft mot radikalisering. Denna resiliensstrategi utgår från idén att medborgare inte bör ses som passiva mottagare av extremistiskt budskap, utan som aktiva informationskonsumenterna som – med rätt verktyg – kan lära sig att identifiera, ifrågasätta och stå emot manipulation.

Centralt i denna ansats är utbildning i källkritik, mediekunskap och digital kompetens. Det handlar inte enbart om att avslöja falska nyheter eller extremistisk propaganda, utan även om att förstå hur algoritmer, plattformsdesign och digitala affärsmodeller påverkar vilka budskap vi exponeras för, och varför vissa typer av innehåll premieras framför andra. Studier visar att sådana förmågor – särskilt hos unga – kan fungera som skyddsfaktorer mot att internalisera extremistiska idéer (Feddes & Gallucci, 2015; Feddes et al., 2015). Men även här är evidensläget osäkert: sambandet mellan digital kompetens och faktisk attityd- eller beteendeförändring är svårsmitt och ofta långsiktigt. De flesta interventionsstudier lider av små urval, korta tidshorisonter eller bristande kontrollgrupper.

Trots detta betraktas mediekunskap allt oftare som en grundläggande och långsiktig strategi i det förebyggande arbetet. I stället för att försöka kontrollera informationsflöden syftar strategin till att stärka individens förmåga att navigera i ett digitalt informationslandskap präglad av komplexitet, känslomässig intensitet och ideologisk polarisering.

En närliggande inriktning fokuserar på att stärka informella motkrafter och autentiska motröster online. Det kan handla om att stötta lärare, fritidsledare, influerare, gräsrotsorganisationer, före detta extremister eller unga användare att själva producera alternativa berättelser som motverkar radikaliserande idéer. Dessa budskap kan vara avväpnande i sin ton – exempelvis genom humor, personliga berättelser eller skildringar av våldets konsekvenser – och ses ofta som mer trovärdiga än statliga kampanjer (Braddock & Horgan, 2016; Freear & Glazzard, 2020). Plattformar kan här spela en proaktiv roll genom att synliggöra och lyfta fram sådana röster. Ett exempel är Peer to Peer: Facebook Global Digital Challenge, där unga i olika länder producerade egna kampanjer mot hatbudskap och extremism, ofta med betydande räckvidd och engagemang.

Även om dessa initiativ är lovande, bör de inte romantiseras. Effekten är svår att mäta, räckvidden begränsad och det krävs ofta långvarigt engagemang för att bygga trovärdighet i målgrupper med låg tillit till etablerade institutioner. Dessutom finns risken att även motberättelser tolkas som del av en statlig eller liberal agenda – särskilt om de distribueras genom plattformar som uppfattas som censurbenägna.

Sammantaget illustrerar denna strategi vikten av att kombinera teknologiska reformer med sociala och pedagogiska åtgärder. Digital resiliens bygger inte på att ta bort problematiskt innehåll, utan på att stärka individens förmåga att känna igen det, stå emot det – och i vissa fall aktivt bemöta det. Det handlar om att göra användare till kritiska aktörer snarare än passiva mål.

5. Från datan till gatan: en syntesmodell

Onlineradikalisering är, som visats i denna rapport, ett komplext och mångfasetterat fenomen som kräver ett tvärvetenskapligt angreppssätt. Den samlade forskningen visar tydligt att ingen enskild disciplin eller teoretisk modell ensam förmår förklara hur och varför vissa individer dras mot extremism – och i vissa fall till politiskt våld – via digitala miljöer. Psykologiska och sociologiska teorier om behovet av mening, identitet och tillhörighet måste därför kombineras med sociotekniska perspektiv på algoritmer, plattformsdesign och gruppdynamik.

Mot denna bakgrund föreslås här en syntetiserande modell bestående av tre huvudsakliga kausala mekanismer, lokaliserade på olika förklaringsnivåer: makro (struktur och kontext), meso (interaktion och nätverk), samt mikro/meso (koordinering och informationsspridning). Dessa mekanismer är analytiskt åtskilda men i praktiken ofta tätt sammanflätade. Vilken mekanism som dominerar beror på kontext, individ och miljö, och de är tillämpbara på såväl individuella som kollektiva uttryck för extremism och våld. Modellen är baserad på en artikel publicerad i *Terrorism and Political Violence* (Wahlström & Törnberg, 2019).

Mekanism 1: diskursiv normalisering och legitimering (makronivå)

En grundläggande förutsättning för politisk radikaliserings är upplevelsen av orättvisa eller hot, ofta riktat mot vad som upplevs som den egna gruppen. Det kan handla om reella erfarenheter av strukturella orättvisor såsom marginalisering, ekonomisk utsatthet och social exkludering, men också om föreställningar om framtida, potentiella hot, exempelvis som kretsar kring identitet, nation eller kultur. Dessa upplevelser är dock inte i sig mobiliserande. Det är först när de tolkas politiskt som uttryck för systematisk orättvisa, som de får mobiliserande kraft.

För att sådana tolkningar ska få genomslag och uppfattas som legitima krävs det som Koopmans och Olzak (2004) kallar en diskursiv möjlighetsstruktur (*discursive opportunity structure*, DOS): ett offentligt och medieburet diskursivt landskap där vissa problemformuleringar, perspektiv och reaktioner upplevs som rimliga och trovärdiga. Tre dimensioner är centrala här: synlighet (hur ofta och tydligt ett budskap eller en aktör förekommer i offentligheten), resonans (i vilken grad budskapet väcker genklang hos publiken och knyter an till etablerade värderingar eller känslor) samt legitimitet (i vilken utsträckning det uppfattas som rimligt eller värt att lyssna på).

Ursprungligen var denna diskursiva struktur starkt kopplad till traditionella massmedier, som genom redaktionella urval kunde påverka just vilka diskurser som blev synliga, väckte resonans och uppfattades som legitima hos allmänheten. Idag har dock sociala medier i stor utsträckning tagit över denna funktion, vilket förändrar villkoren för vilka tolkningar som får fäste i det offentliga samtalet.

Digitala plattformar påverkar inte bara vad som sprids, utan också hur idéer får fäste. Algoritmer premierar innehåll som väcker starka känslor – särskilt ilska, rädsla eller indignation – vilket innebär att just de budskap som är mest mobiliserande ofta också är de som mest effektivt sprids på plattformarna. Detta bidrar till att ytterligare förstärka sambandet mellan synlighet och resonans. Samtidigt förstärks även upplevelsen av legitimitet, då användare i specifika slutna eller halvoffentliga forum och grupper kan uppleva lokal legitimitet, genom att användarna bekräftar varandras tolkningar och beskrivningar, även om de inte har bred samhällelig legitimitet för sina idéer utanför dessa slutna rum.

En ytterligare eskalerande mekanism är algoritmernas tendens att leda användare mot alltmer specialiserat och radikalt innehåll. Rekommendationssystem skapar därmed en gradvis förskjutning från etablerade berättelser till subkulturella och potentiellt extrema idévärldar. Denna rörelse sker ofta gradvis och omärkligt för användaren, vilket resulterar i en normaliseringsspiral, där budskap som tidigare uppfattades som extrema successivt blir begripliga, rationella – och i vissa fall moraliskt nödvändiga.

Dessa kausala mekanismer på makronivå kan beskrivas som *en diskursiv accelerationsprocess*, där digitala miljöer genom känslolofokuserad förstärkning och algoritmisk selektion bidrar till att forma en kontext där radikalism och ibland våld ter sig som logiska och legitima svar på ett hot eller en upplevd orättvisa.

Mekanism 2: affektiv interaktion (mesonivå)

Utöver att fungera som diskursiv arena för ideologiska budskap fungerar sociala medier även som interaktiva miljöer där användare är aktiva medproducenter av mening. Här framträder den andra mekanismen: affektiva grupprocesser som förstärker radikaliserande dynamiker genom emotionell intensitet, kollektiv identifikation och moralisk mobilisering.

Dessa affektiva processer kan förstås som digitala motsvarigheter till vad Randall Collins kallar interaktionsritualer. Enligt Collins (2004a, 2004b) uppstår social sammanhållning och emotionell energi genom upprepade interaktioner där deltagare delar fokus och känsloläge. Den emotionella energin som alstras i dessa interaktioner blir drivkraften att återvända till liknande situationer och stärker individens band till gruppen.

I digitala miljöer, särskilt på slutna eller halvoffentliga forum, kan sådana ritualer uppstå även i asynkrona och textbaserade former. Intern jargong, memes, symboler och gemensamma fiendebilder förstärker sammanhållningen. När deltagare interagerar kring ett gemensamt fokus – som invandringskritik, misogyni eller konspirationsteorier – och uttrycker gemensamma känslor som ilska eller hat, uppstår en form av affektiv samstämmighet som förstärker både

identitet och ideologisk övertygelse. De som deltar i dessa ritualer får emotionell energi genom att uttrycka sin åsikt, få respons från andra användare (likes, svar, memes), och upplever därmed en samhörighet. Ibland sker detta i motstånd mot en upplevd mainstream eller politisk korrekthet, vilket förstärker känslan av att vara del av något autentiskt.

Collins teori om våld bygger på att våld oftast är svårt att utföra. Det kräver att man övervinner en naturlig tröskel av rädsla, osäkerhet och tvekan. Denna tröskel kan i vissa fall överskridas om individen befinner sig i en upptrissad interaktionskedja som ger tillräcklig emotionell energi, särskilt om man samtidigt känner stöd eller legitimitet från sin grupp. I en onlinekontext kan detta ta formen av eskalerande diskussioner som innefattar våldsretorik, martyrskap och glorifiering av tidigare gärningsmän. Symboler, politiska manifest och memes blir då inte bara abstrakta idébärare, utan även emotionella verktyg som för samman gruppen och sänker tröskeln till handling.

I många onlineforum uppstår affektiva spiraler, där deltagare övertrumfar varandra i känslouttryck, radikalitet och aggressivitet. De mest extrema rösterna får ofta mest respons i form av likes och delningar, vilket förstärker deras status och drivkrafter. Detta kan förklara varför vissa individer går från engagemang online till fysiskt våld: de har genomgått en lång rituell process där emotionell energi, identitet och ideologi gradvis vävts samman, vilket sänkt deras våldströskel och gett dem upplevelsen av moraliskt stöd från en imaginär publik.

Collins teori synliggör därmed hur radikaliserings handlar om såväl idéer som deltagande i emotionellt laddade interaktioner. Det är *i själva interaktionen*, i ritualens dynamik, som lojalitet, moralisk motivation och handlingsberedskap formas. Sociala medier förstärker dessa interaktionsdynamiker genom *affektiv acceleration* – återkopplingsslingor där starka känslor reproduceras, intensifieras och får allt större mobiliserande kraft.

Mekanism 3: instrumentell koordinering (mikro-/mesonivå)

Den tredje mekanismen rör den praktiska dimensionen av digital radikaliserings – hur nätmiljöer inte bara möjliggör idéernas spridning (mekanism 1) eller känslomässig mobilisering (mekanism 2), utan också tillhandahåller de verktyg och instruktioner som krävs för att omvandla intention till handling. Till skillnad från de tidigare mekanismerna är denna mindre ideologisk eller affektiv till sin karaktär och mer funktionell och instrumentell: den handlar om hur konkreta åtgärder förbereds, organiseras och genomförs.

Digitala plattformar – särskilt de mer anonyma, oreglerade och slutna – fungerar idag som gränslösa arkiv för operativ kunskap. Här delas detaljerade instruktioner om hur man bygger vapen, tillverkar sprängämnen, saboterar infrastruktur, anlägger bränder eller identifierar specifika mål. Det kan röra sig om kartor över flyktingboenden, adresser till politiska motståndare, inköpsguider till vapen, eller tips om hur man undviker upptäckt vid aktioner. Detta innehåll sprids via slutna Telegram-kanaler, krypterade chattplattformar eller på särskilt

avskärmda delar av forum som Reddit, Flashback eller 4chan, där moderering i vissa fall saknas och anonymitet är norm.

Men denna mekanism handlar inte enbart om individuella våldsrepertoarer. Den rymmer även kollektiva samordningsprocesser – från organiserad nätmobbning och trakasserier till samordnade sabotageaktioner eller koordinerade hot. I många fall fungerar digitala miljöer som operativa centraler där aktiviteter planeras, kommuniceras och synkroniseras i realtid. Det som tidigare krävde fysiska möten och hierarkiska strukturer kan nu genomföras med några knapptryck. Även löst sammanfogade nätverk utan formell organisation kan därmed uppnå en hög grad av taktisk koordination.

Det är viktigt att förstå att denna instrumentella funktion uppstår sällan i ett vakuum. Den aktiveras vanligtvis i ett senare skede av radikaliseringsprocessen, när individen eller gruppen redan har internaliserat en specifik världsbild (mekanism 1) och genomgått affektiva socialiseringsprocesser (mekanism 2). När mening, motivation och moralisk legitimitet väl har etablerats, blir tillgången till praktisk kunskap, teknisk infrastruktur och social uppmuntran avgörande för att handlingen ska realiseras – oavsett om det handlar om våld, hot eller sabotage, eller andra typer av våldshandlingar.

Denna tredje mekanism opererar alltså på mikronivå och besvarar konkreta frågor om hur, var och när en viss politiskt motiverad handling eller våldsaktion kan genomföras. Den är särskilt viktig för att förstå övergången från idé till handling.

Tillsammans utgör dessa tre mekanismer – (1) diskursiv normalisering, som förser individer med tolkningsramar och problemformuleringar, (2) affektiv interaktion, som laddar dessa idéer med emotionell energi och kollektiv förankring, och (3) instrumentell koordinering, som omvandlar intention till konkret handling – ett dynamiskt digitalt ekosystem för radikalisering. Dessa mekanismer verkar visserligen på olika nivåer men är ofta sammanflätade och ömsesidigt förstärkande. Radikalisering är därmed sällan en linjär process där idéer leder till känslor, som i sin tur leder till handling. I praktiken sker ofta parallella processer, återkopplingar och situationsvariationer beroende på individ, sammanhang och syfte.

Det viktiga är att förstå internet inte bara som en kanal för idéer, utan som en hel infrastruktur för radikalisering: en miljö där diskursiv, emotionell och operativ energi samverkar och där extrema uttryck, vare sig de är våldsamma eller ej, kan konkretiseras och få verkan i den sociala världen.

Litteraturförteckning

- Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, 30(1), 47-88.
- Alava, S., Frau-Meigs, D., & Hassan, G. (2017). *Youth and violent extremism on social media: mapping the research*. UNESCO publishing.
- Alfano, M., Carter, J. A., & Cheong, M. (2018). Technological seduction and self-radicalization. *Journal of the American Philosophical Association*, 4(3), 298-322.
- Alimi, E. Y., Bosi, L., & Demetriou, C. (2015). *The dynamics of radicalization: A relational and comparative perspective*. Oxford University Press.
- Andersson Malmros, R. (2025). Radikalisering – från myt till sanning? In C. Mattsson, Andersson Malmros, R. & Johansson, T. (Ed.), *Den levda demokratins utmaningar: Extrema rörelser, antisemitism och prevention* (pp. 15-35). Bokförlaget Korpen.
- Arcila Calderón, C., Sánchez Holgado, P., Gómez, J., Barbosa, M., Qi, H., Matilla, A., Amado, P., Guzmán, A., López-Matías, D., & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities. *Humanities and Social Sciences Communications*, 11(1), 1-14.
- Bail, C. (2022). *Breaking the social media prism: How to make our platforms less polarizing*. Princeton University Press.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216-9221.
- Baugut, P., & Neumann, K. (2020). Online news media and propaganda influence on radicalized individuals: Findings from interviews with Islamist prisoners and former Islamists. *New Media & Society*, 22(8), 1437-1461.
- Bélanger, J. J., Nisa, C. F., Schumpe, B. M., Gurmu, T., Williams, M. J., & Putra, I. E. (2020). Do counter-narratives reduce support for ISIS? Yes, but not for their target audience. *Frontiers in psychology*, 11, 1059.
- Benson, D. C. (2014). Why the internet is not increasing terrorism. *Security Studies*, 23(2), 293-328.
- Berger, J., & Perez, H. (2016). Occasional Paper The Islamic State's Diminishing Returns on Twitter: How suspensions are limiting the social networks of English-speaking ISIS supporters. *GW Program Extremism El Akkad, Omar (2012) 'Why Twitter's Censorship Plan Better you Think'*. *The Globe and Mail*. Berger, Morgan (2015) 'Defining Describing Population of ISIS Supporters Twitter' *Brookings Inst*.
- Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British journal of political science*, 47(2), 241-262.

- Bjørger, T., & Gjelsvik, I. M. (2015). Forskning på forebygging av radikaliserings og voldelig ekstremisme: En kunnskapsstatus. (Norwegian research on the prevention of radicalisation and violent extremism: A status of knowledge).
- Bjørger, T., & Horgan, J. (2009). Leaving terrorism behind. *Individual and collective*.
- Borum, R. (2004). *Psychology of terrorism*.
- Borum, R. (2011a). Radicalization into violent extremism I: A review of social science theories. *Journal of strategic security*, 4(4), 7-36.
- Borum, R. (2011b). Rethinking radicalization. *Journal of strategic security*, 4(4), 1-6.
- Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in conflict & terrorism*, 39(5), 381-404.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313-7318.
- Brouillette-Alarie, S., Hassan, G., Varela, W., Ousman, S., Kilinc, D., Savard, É. L., Madriaza, P., Harris-Hogan, S., McCoy, J., & Rousseau, C. (2022). Systematic review on the outcomes of primary and secondary prevention programs in the field of violent radicalization. *Journal for Deradicalization*(30), 117-168.
- Carthy, S. L., Doody, C. B., Cox, K., O'Hora, D., & Sarma, K. M. (2020). Counter-narratives for the prevention of violent radicalisation: A systematic review of targeted interventions. *Campbell Systematic Reviews*, 16(3), e1106.
- Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on human-computer interaction*, 1(CSCW), 1-22.
- Chen, A. Y., Nyhan, B., Reifler, J., Robertson, R. E., & Wilson, C. (2023). Subscriptions and external links help drive resentful users to alternative and extremist YouTube channels. *Science advances*, 9(35), eadd8080.
- Clifford, B., & Powell, H. (2019). Encrypted extremism: Inside the English-speaking Islamic state ecosystem on telegram. *The George Washington University Program on Extremism*, 27-53.
- Collins, R. (2004a). *Interaction ritual chains*. Princeton University Press.
- Collins, R. (2004b). *Violence: A micro-sociological theory*. Princeton University Press.
- Corner, E., & Gill, P. (2015). A false dichotomy? Mental illness and lone-actor terrorism. *Law and human behavior*, 39(1), 23.
- Cottee, S. (2024). Radicalization discourse: Consensus points, evidence base and blind spots. *Studies in conflict & terrorism*, 1-25.
- Crenshaw, M. (2007). The logic of terrorism. *Terrorism in perspective*, 24, 24-33.

- Da Silva, C., Amadio, N., Sarg, R., Domingo, B., & Benbouriche, M. (2024). Risk and Protective Factors Associated with Violent Extremism. *Perspectives on Terrorism*, 18(3), 69-89.
- Dalgaard-Nielsen, A. (2010). Violent radicalization in Europe: What we know and what we do not know. *Studies in conflict & terrorism*, 33(9), 797-814.
- Davey, J. (2021). Gamers who hate: An introduction to ISD's gaming and extremism series. *Institute for Strategic Dialogue*, 12.
- Della Porta, D. (2018). Radicalization: A relational perspective. *Annual Review of Political Science*, 21, 461-474.
- Demetriou, C., & Silke, A. (2003). A criminological internet 'sting'. Experimental evidence of illegal and deviant visits to a website trap. *British Journal of Criminology*, 43(1), 213-222.
- Dixon, A., Howie, P., & Starling, J. (2005). Trauma exposure, posttraumatic stress, and psychiatric comorbidity in female juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(8), 798-806.
- Dobratz, B. A., & Shanks-Meile, S. L. (2000). *The white separatist movement in the United States: "White power, white pride!"*. JHU Press.
- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729-745.
- Ezekiel, R. S. (2002). An ethnographer looks at neo-Nazi and Klan groups: The racist mind revisited. *American Behavioral Scientist*, 46(1), 51-71.
- Feddes, A. R., & Gallucci, M. (2015). A literature review on methodology used in evaluating effects of preventive and de-radicalisation interventions. *Journal for Deradicalization*(5), 1-27.
- Feddes, A. R., Mann, L., & Doosje, B. (2015). Increasing self-esteem and empathy to prevent violent radicalization: a longitudinal quantitative evaluation of a resilience training focused on adolescents with a dual identity. *Journal of Applied Social Psychology*, 45(7), 400-411.
- Feddes, A. R., Nickolson, L., van Bergen, N. R., Mann, L., & Doosje, B. (2023). Extremist thinking and doing: A systematic literature study of empirical findings on factors associated with (de) radicalisation processes. *International journal of developmental science*, 17(1-3), 7-18.
- Freear, M., & Glazzard, A. (2020). Preventive communication: Emerging lessons from participative approaches to countering violent extremism in Kenya. *The RUSI Journal*, 165(1), 90-106.
- Gaudette, T., Scrivens, R., & Venkatesh, V. (2022). The role of the internet in facilitating violent extremism: insights from former right-wing extremists. *Terrorism and Political Violence*, 34(7), 1339-1356.

- Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M., & Horgan, J. (2017). Terrorist use of the Internet by the numbers: Quantifying behaviors, patterns, and processes. *Criminology & Public Policy*, 16(1), 99-117.
- Gill, P., Horgan, J., & Deckert, P. (2014). Bombing alone: Tracing the motivations and antecedent behaviors of lone-actor terrorists. *Journal of forensic sciences*, 59(2), 425-435.
- Greenwald, R. (2014). *Trauma and juvenile delinquency: Theory, research, and interventions*. Routledge.
- Hafez, M., & Mullins, C. (2015). The radicalization puzzle: A theoretical synthesis of empirical approaches to homegrown extremism. *Studies in conflict & terrorism*, 38(11), 958-975.
- Haller, A., & Holt, K. (2019). Paradoxical populism: how PEGIDA relates to helmmainstream and alternative media. *Information, Communication and Society* 22(12), 1665-1680.
- Haroon, M., Wojcieszak, M., Chhabra, A., Liu, X., Mohapatra, P., & Shafiq, Z. (2023). Auditing YouTube's recommendation system for ideologically congenial, extreme, and problematic recommendations. *Proceedings of the National Academy of Sciences*, 120(50), e2213020120.
- Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., Varela, W., Borokhovski, E., Venkatesh, V., & Rousseau, C. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International journal of developmental science*, 12(1-2), 71-88.
- Heft, A., Knüpfer, C., Reinhardt, S., & Mayerhöffer, E. (2021). Toward a Transnational Information Ecology on the Right? Hyperlink Networking among Right-Wing Digital News Sites in Europe and the United States. *The International Journal of Press/Politics*, 26(2), 484-504.
- Helmus, T. C., & Klein, K. (2018). Assessing Outcomes of Online Campaigns Countering Violent Extremism. *Product Page*.
- Herath, C., & Whittaker, J. (2023). Online radicalisation: Moving beyond a simple dichotomy. *Terrorism and Political Violence*, 35(5), 1027-1048.
- Horgan, J., & Braddock, K. (2010). Rehabilitating the terrorists?: Challenges in assessing the effectiveness of de-radicalization programs. *Terrorism and Political Violence*, 22(2), 267-291.
- Jensen, M. A., Atwell Seate, A., & James, P. A. (2020). Radicalization to violence: A pathway approach to studying extremism. *Terrorism and Political Violence*, 32(5), 1067-1090.
- Jungherr, A., Rivero, G., & Gayo-Avello, D. (2020). *Retooling politics: How digital media are shaping democracy*. Cambridge University Press.
- Kallio, P., & Törnberg, A. (2025). Rituals of resentment: How dark emotional energy fuels incel identity and solidarity. *Acta Sociologica*, 00016993251317963.

- Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. *European journal of communication*, 0267323117695734.
- Kenyon, J., & Baker-Beall, C. (2022). The Internet and Radicalization Pathways: Technological Advances, Relevance of Mental Health and Role of Attackers.
- Kis-Katos, K., Liebert, H., & Schulze, G. G. (2011). On the origin of domestic and international terrorism. *European Journal of Political Economy*, 27, S17-S36.
- Koopmans, R., & Olzak, S. (2004). Discursive Opportunities and the Evolution of Right-Wing Violence in Germany. *American Journal of Sociology*, 110(1), 198-230.
- Krueger, A. B., & Malečková, J. (2002). Education, poverty and terrorism: Is there a causal connection? *Journal of Economic perspectives*, 17(4), 119-144.
- Kyler, A. M. (2020). *Alt-Right as a Social Movement: Ethnic Competition Theory and Internet Searches for Online Content*.
- Lakhani, S., & Wiedlitzka, S. (2023). “Press F to pay respects”: an empirical exploration of the mechanics of gamification in relation to the Christchurch attack. *Terrorism and Political Violence*, 35(7), 1586-1603.
- Lauer, D. (2021). Facebook’s ethical failures are not accidental; they are part of the business model. *AI and Ethics*, 1(4), 395-403.
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological science in the public interest*, 13(3), 106-131.
- Lindekilde, L., Malthaner, S., & O’Connor, F. (2019). Peripheral and embedded: Relational patterns of lone-actor terrorist radicalization. *Dynamics of Asymmetric Conflict*, 12(1), 20-41.
- Lundstedt, M., Törnberg, A., Vallström, V., & Wahlström, M. (2025). *Främlingsfientligt våld i Sverige 2009–2022* (9187876795). <https://hdl.handle.net/2077/86880>
- Lösel, F., & Bliesener, T. (2021). Extremism, radicalization and terrorism. *Monatsschrift für Kriminologie und Strafrechtsreform*, 104(3), 179-183.
- Malthaner, S., O’Connor, F., & Lindekilde, L. (2024). Scattered attacks: The collective dynamics of lone-actor terrorism. *Perspectives on Politics*, 22(2), 463-480.
- McAdam, D. (1986). Recruitment to high-risk activism: The case of freedom summer. *American Journal of Sociology*, 92(1), 64-90.
- McCauley, C., & Moskaleiko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence*, 20(3), 415-433.
- McCauley, C., & Moskaleiko, S. (2017). Understanding political radicalization: The two-pyramids model. *American psychologist*, 72(3), 205.
- McCauley, K. (2023). Preventing Attacks from Lone-Actors. *Violent Extremism and Terrorism – Challenges and Responses*, Stockholm.

- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444.
- Merton, R. K. (1938). Social Structure and Anomie. *American Sociological Review*, 3(5).
- Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2023). Engagement, user satisfaction, and the amplification of divisive content on social media. *arXiv preprint arXiv:2305.16941*.
- Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS nexus*, 4(3), pgaf062.
- Moghaddam, F. M. (2005). The staircase to terrorism: A psychological exploration. *American psychologist*, 60(2), 161.
- Moskalenko, S. (2010). Civilians into warriors: Mechanisms of mobilization in US Army recruitment and training. *Dynamics of Asymmetric Conflict*, 3(3), 248-268.
- Mudde, C. (2007). *Populist radical right parties in Europe* (Vol. 22). Cambridge university press Cambridge.
- Mudde, C. (2019). *The far right today*. Polity.
- Müller, K., & Schwarz, C. (2020a). Fanning the Flames of Hate: Social Media and Hate Crime. Available at <http://dx.doi.org/10.2139/ssrn.3082972>.
- Müller, K., & Schwarz, C. (2020b). From hashtag to hate crime: Twitter and anti-minority sentiment. Available at SSRN 3149103.
- Mølmen, G. N., & Ravndal, J. A. (2023). Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists. *Behavioral Sciences of Terrorism and Political Aggression*, 15(4), 463-487.
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303-330.
- Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: a randomized trial. *Pediatrics*, 133(4), e835-e842.
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., & Dimmery, D. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137-144.
- Pape, R. (2021). Understanding American domestic terrorism: Mobilization potential and risk factors of a new threat trajectory. *UChicago CPOST*.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.
- Polletta, F., & Callahan, J. (2017). Deep stories, nostalgia narratives, and fake news: Storytelling in the Trump era. *American Journal of Cultural Sociology*.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of personality and social psychology*, 67(4), 741.

- Ravndal, J. A. (2016). Right-wing terrorism and violence in Western Europe: Introducing the RTV dataset. *Perspectives on Terrorism*, 10(3), 2-15.
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020). Auditing radicalization pathways on YouTube. Proceedings of the 2020 conference on fairness, accountability, and transparency,
- Rosand, E., & Winterbotham, E. (2019). Do counter-narratives actually reduce violent extremism? *Brookings*. <https://www.brookings.edu/articles/do-counter-narratives-actually-reduce-violent-extremism/>
- Ruby, C. L. (2002). Are terrorists mentally deranged? *Analyses of Social Issues and Public Policy*, 2(1), 15-26.
- Sageman, M. (2004). *Understanding terror networks*. University of Pennsylvania press.
- Schlegel, L., & Kowert, R. (2024). *Gaming and extremism: the radicalization of digital playgrounds*. Taylor & Francis.
- Schulze, H., Greipl, S., Hohner, J., & Rieger, D. (2024). Social media and radicalization: An affordance approach for cross-platform comparison. *M&K Medien & Kommunikationswissenschaft*, 72(2), 187-212.
- Schumann, S., Clemmow, C., Rottweiler, B., & Gill, P. (2024). Distinct patterns of incidental exposure to and active selection of radicalizing information indicate varying levels of support for violent extremism. *PloS one*, 19(2), e0293810.
- Schuurman, B., & Carthy, S. L. (2024). Understanding (non) involvement in terrorist violence: What sets extremists who use terrorist violence apart from those who do not? *Criminology & Public Policy*, 23(1), 119-152.
- Schuurman, B., & Taylor, M. (2018). Reconsidering radicalization: Fanaticism and the link between ideas and violence. *Perspectives on Terrorism*, 12(1), 3-22.
- Shaw, A. (2023). Social media, extremism, and radicalization. *Science advances*, 9(35), eadk2031.
- Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.
- START. (2025). *Global Terrorism Database* (
- Sunstein, C. (2002). *Republic. com*. Princeton University Press.
- Sunstein, C. (2007). *Republic.com 2.0po*. Princeton University Press.
- Sunstein, C. (2008). Neither hayek nor habermas. *Public Choice*, 134(1-2), 87-95.
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of applied research in memory and cognition*, 9(3), 286-299.
- Tilly, C. (2004). Terror, terrorism, terrorists. *Sociological theory*, 22(1), 5-13.
- Tufekci, Z. (2018). YouTube, the great radicalizer. *The New York Times*, 10(3), 2018.

- Törnberg, A., & Nissen, A. (2022). Mobilizing against Islam on social media: hyperlink networking among European far-right extra-parliamentary Facebook groups. *Information, Communication & Society*, 1-19.
- Törnberg, A., & Törnberg, P. (2024). From echo chambers to digital campfires: The making of an online community of hate in Stormfront. In J. B. Walther, & Rice, R. E (Ed.), *Social processes of online hate* (pp. 93-119). Taylor & Francis.
- Törnberg, A., & Vallström, V. (2025). Asymmetric Alliances in Climate Misinformation: A Network Analysis of the Swedish Climate Change Countermovement. *SocArXiv*. https://doi.org/https://valldoi.org/10.31235/osf.io/wvgqe_v1
- Törnberg, A., & Wahlström, M. (2018). Unveiling the radical right online: Exploring framing and identity in an online anti-immigrant discussion group. *Sociologisk forskning*, 55(2-3), 267-292.
- Van den Bos, K. (2018). *Why people radicalize: How unfairness judgments are used to fuel radical beliefs, extremist behaviors, and terrorism*. Oxford University Press.
- Vaux, P., Gallagher, A., & Davey, J. (2021). *The extreme right on Steam* (Gaming and Extremism Series, Issue. <https://www.isdglobal.org/wp-content/uploads/2021/08/02-revised-gaming-report-steam.pdf>)
- Vergani, M., Iqbal, M., Ilbahar, E., & Barton, G. (2020). The three Ps of radicalization: Push, pull and personal. A systematic scoping review of the scientific evidence about radicalization into violent extremism. *Studies in conflict & terrorism*, 43(10), 854-854.
- Viterna, J. S. (2006). Pulled, pushed, and persuaded: Explaining women's mobilization into the Salvadoran guerrilla army. *American Journal of Sociology*, 112(1), 1-45.
- Wahlström, M., & Törnberg, A. (2019). Social media mechanisms for right-wing political violence in the 21st century: Discursive opportunities, group dynamics, and co-ordination. *Terrorism and Political Violence*, 1-22.
- Wahlström, M., Törnberg, A., & Ekbrand, H. (2021). Dynamics of violent and dehumanizing rhetoric in far-right social media. *New Media & Society*, 23(11), 3290-3311.
- Waller, B. Y., Maleku, A., Quinn, C. R., Barman-Adhikari, A., Sprague Martinez, L. S., Traube, D., & Bellamy, J. L. (2022). Just research: Advancing antiracist and antioppressive social work research. *Journal of the Society for Social Work and Research*, 13(4), 637-644.
- Whittaker, J. (2021). The online behaviors of Islamic state terrorists in the United States. *Criminology & Public Policy*, 20(1), 177-203.
- Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-Black and Anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, 60(1), 93-117.
- Wolfowicz, M., Hasisi, B., & Weisburd, D. (2022). What are the effects of different elements of media on radicalization outcomes? A systematic review. *Campbell Systematic Reviews*, 18(2), e1244.

- Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2021). Cognitive and behavioral radicalization: A systematic review of the putative risk and protective factors. *Campbell Systematic Reviews*, 17(3), e1174.
- Wolfowicz, M., Weisburd, D., & Hasisi, B. (2023). Examining the interactive effects of the filter bubble and the echo chamber on radicalization. *Journal of Experimental Criminology*, 19(1), 119-141.
- Yesilada, M., & Lewandowsky, S. (2022). Systematic review: YouTube recommendations and problematic content. *Internet policy review*, 11(1), 1652.
- Åkerlund, M. (2022). Dog whistling far-right code words: the case of 'culture enricher' on the Swedish web. *Information, Communication & Society*, 25(12), 1808-1825.