

# Normer og samarbeid

Jon Reiersen

Tønsberg: Høgskolen i Vestfold, 2004

Notat 5/2004

Notat 5/2004 Høgskolen i Vestfold  
Copyright: Høgskolen i Vestfold/Jon Reiersen  
ISSN: 0808-131x

# Normer og samarbeid

JON REIERSEN

## Sammendrag

Mange situasjoner krever samarbeid mellom to eller flere parter for å realisere et godt resultat. Selv om partene i slike tilfeller er innstilt på å samarbeide, så følger det ikke at samarbeid nødvendigvis vokser fram. En viktig årsak til dette er at det i mange samarbeidssituasjoner er en konflikt mellom handlinger som fremmer den enkeltes interesser og handlinger som fremmer partenes felles interesser. Artikkelen studerer nærmere hvordan en norm om samarbeid kan løse en slik konflikt, og hvilke forhold som kan forklare at normer om samarbeid “overlever”, og eventuelt sprer seg over tid, i et samfunn.

**JEL klassifisering:** 017, Q12, Q15

**Emneord:** Sosiale dilemma, normer, samarbeid

*“It is not possible to understand the peculiarities and the forces behind human cooperation unless we understand social norms. (...) Despite some recent progress in the analysis of social norms, a long journey still remains until science will be able to provide a satisfactory understanding.”* (Fehr og Fischbacher, 2004 s. 189)

## 1 Innledning

Det er vanskelig å tenke seg at et samfunn kan fungere uten normative standarder for akseptabel adferd. Sagt på en annen måte, ingen samfunn kan fungere uten *normer*. Det er derfor ikke overraskende at; (...) *social scientists, at least those outside of economics, invoke no other concept more frequently than that of ‘norms’.* (Fehr og Fischbacher, 2004 s. 64). Noen vil til og med gå så langt som til å si at evnen til å etablere og håndheve normer er et av de viktigste kjennetegn på mennesket som art (Gintis, 2003).

Som sitatet innledningsvis indikerer, så er normer spesielt viktig for å forklare samarbeid. Mange samarbeidsrelasjoner bærer nemlig i seg en viktig kime til konflikt. Egoistisk sett framstår det ofte som best å motta en tjeneste uten å måtte yte noe i gjengjeld. Det minst gunstige er å gi men ikke få noe tilbake. Samtidig kommer alle dårligere ut dersom ingen bidrar enn om alle yter en innsats. Det som fremstår som best for den enkelte leder altså til et resultat som for partene samlet sett er dårlig. Dette til tross, både hverdagserfaringer og eksperimenter gjennomført av psykologer, økonomer og andre samfunnvitere peker i retning av at folk flest er innstilt på å samarbeide i situasjoner der egeninteresse trekker i retning av å velge ‘ikke samarbeid’.<sup>1</sup> Sosiologer har i lang tid pekt på *normer* for å forklare dette. Normer om samarbeid holder våre opportunistiske tendenser i sjakk, og leder oss til å handle i tråd med fellesinteresser framfor egeninteressen. Økonomer har imidlertid vært lite begeistret for ideen om å peke på normer for å forklare økonomisk eller sosial adferd. En vanlig innvending er at enhver handling i prinsippet kan forklares ved å peke på en eller annen norm. Det finnes også sosiologer som påpeker problemet ved å anta normer som gitt. James Coleman er en av disse. Han tar til orde for at eksistensen av normer bør kunne forklares på en eller annen måte, og skriver: *“Much sociological theory takes norms as given and proceeds to examine individual behavior or the behavior of social systems when norms exist. Yet to do this*

---

<sup>1</sup>Se Fehr m.fl. (2002), Gintis m.fl. (2003) og Richerson m.fl. (2003) for en oversikt over denne litteraturen.

*without raising at some point the question of why and how norms come into existence is to forsake the more important sociological problem in order to address the less important.*” (Coleman, 1990 s. 244).<sup>2</sup>

Målsettingen med denne artikkelen er å forsøke å forklare framveksten av en norm om samarbeid som et resultat av en pågående sosial interaksjon. Dette gjøres ved å anta et samfunn bestående av to typer aktører. Den ene typen, som jeg kaller *samarbeidere*, er i utgangspunktet motivert av en norm om samarbeid. Den andre typen, som jeg kaller *opportunist*, er ikke opptatt av å handle i tråd med normer om samarbeid. De velger ‘ikke samarbeid’ når dette fremmer deres egoistiske interesser. Disse to typene aktører utgjør samfunnet som vi skal betrakte. Aktørene møtes innenfor dette samfunnet for å spille ulike varianter av et fangenes dilemma spill, der partene velger handling i henhold til sine subjektive preferanser. Disse preferansene endres imidlertid over tid via en sosial evolusjonær prosess. Det antas at hver enkelt forsøker å tilpasse sine handlinger over tid til den handlingen som gir størst utbytte i den sosiale interaksjonen som studeres. Implikasjonene av dette er at “overlevelsesevnen” til en norm om samarbeid er bestemt av utbyttet som følger av den handlingen som normen tilskriver. Dersom handlinger som bryter med normen om samarbeid generer et høyere utbytte, så vil det være en tendens til at normen brytes også av de som i utgangspunktet var “bærere” av den. Det motsatte vil skje dersom handlinger som er i tråd med normen om samarbeid genererer et høyere utbytte enn handlinger som bryter med normen. Normer om samarbeid vil da spre seg over tid blant aktørene i samfunnet.<sup>3</sup>

I analysen nedenfor betrakter jeg en norm om samarbeid som en handlingsregel. Når det er en konflikt mellom interessene til den enkelte og interessene til en gruppe eller et samfunn, så er det tvilsomt om en regel har noen innvirkning på hvordan den enkelte handler dersom avvik fra regelen ikke føre til noen som helst form for “straff”. En norm vil derfor bare ha

---

<sup>2</sup>En måte å forklare eksistensen av normer er å gå veien om såkalte funksjonelle forklaringer, der en norm (eller andre fenomener) blir forklart ved hjelp av den gunstige effekten normen (eller fenomenet) har for en gruppe eller for et samfunn. Problemet med en slik forklaring er at et fenomen forklares ved hjelp av et annet fenomen som inntreffer senere i tid. Fra et metodisk perspektiv bør en kunne peke på en forklaring som ligger forut i tid for fenomenet som søkes forklart. Det skal altså ikke være nødvendig å avvente virkningene av et fenomen for å forklare det. Det faktum at en norm om samarbeid er nyttig fordi den leder folk til å samarbeide er interessant, men altså ikke tilstrekkelig som en forklaring på at normen eksisterer. Se Elster (1983) for en mer utfyllende kritikk av funksjonelle forklaringer.

<sup>3</sup>Se også Basu (1995), Sethi (1996) og Bendor og Swistak (2001) som benytter en lignende tilnærming for å forklare eksistensen av normer.

en effekt dersom det finnes en form for håndhevelsesmekanisme som bidrar til at en den også følges opp i form av handling. Det skilles gjerne mellom to slike håndhevelsesmekanismer; intern og ekstern sanksjonering (Elster, 1989; Horne, 2001).

Dersom en norm er internalisert så vil de som er “bærere” av normen handle i tråd med den selv om et eventuelt avvik fra normen ikke blir oppdaget, og dermed heller ikke sanksjonert. Motivet for å handle i tråd med normen er å unngå en ubehagelig følelse av skam, skyld, tap av selvspekt o.l., som kommer innenfra en selv. Det er dette jeg vil referere til som intern sanksjonering. Med ekstern sanksjonering menes at andre vil straffe avvik fra en norm. Dette kan ta form av sosial utstøting, verbal sjikane, sinte blikk o.l. Poenget med å handle i tråd med en norm er i dette tilfellet å unngå en slik uformell straff fra andre.

Nedenfor skal jeg betrakte både situasjonen med intern og ekstern sanksjonering av normer av normer. Spørsmålet jeg skal forsøke å besvare under begge situasjonene er: Under hvilke betingelser vil adferd styrt av en norm om samarbeid overleve i situasjoner hvor egeninteressen best fremmes ved å være gratispassasjer på andres samarbeid?

Artikkelen er disponert på følgende måte. I avsnitt 2 og 3 presiserer jeg nærmere den sosiale interaksjonen som danner utgangspunktet for analysen. I avsnitt 3 redegjør jeg for hvordan preferanser endres over tid, mens i avsnitt 4 studerer jeg overlevelsessevnen til en internalisert norm om samarbeid. I avsnitt 5 er jeg opptatt av å forklare hvordan ekstern sanksjonering kan stabilisere normer om samarbeid. Noen avsluttende kommentarer følger i avsnitt 6.

## 2 Den sosiale interaksjonen

Anta et samfunn bestående av et stort antall aktører som møtes parvis. Det er tilfeldig hvem en møter og hvert møte er kun en engangsforeteelse. Spillerne velger derfor handling uavhengig av forventninger om framtidig interaksjon. Mulige handlinger for den enkelte aktør (som heretter kalles spiller) og utbyttet som følger av de ulike handlingene er angitt i spillmatrisen nedenfor.

	Samarbeid	Ikke samarbeid
Samarbeid	$b, b$	$d, a$
Ikke samarbeid	$a, d$	$c, c$

Anta at  $a > b > c > d$ , det vil si at spillerne er involvert i et fangenes dilemma spill. Spillet kan for eksempel tolkes som en markedstransaksjon der den ene spilleren er selger og den andre spilleren en kjøper. Dersom begge velger ‘samarbeid’ så betyr det at begge parter oppfyller sine forpliktelser. Selgeren leverer varen han har forpliktet seg til å levere, mens kjøperen betaler den avtalte prisen. Begge mottar et utbytte på  $b$ . Dersom begge velger ‘ikke samarbeid’ så betyr det at selgeren ikke leverer varen og kjøperen heller ikke betaler den avtalte prisen. Det vil si at et ikke finner sted noen transaksjon, og begge får  $c$  i utbytte. Dersom selgeren ikke leverer varen (velger ‘ikke samarbeid’), mens kjøperen betaler den avtalte prisen (velger ‘samarbeid’), så får selgeren utbyttet  $a$ , mens kjøperen får det lavest mulige utbyttet  $d$ . Dette fordi selgeren sparer kostnaden ved å framstille varen, samtidig som han får betalt for den. Kjøperen kommer da dårlig ut siden han betaler for en vare han ikke mottar. Dersom det motsatte tilfellet inntreffer, altså at selgeren leverer varen men ikke får betaling fra kjøperen, så er det selgeren som får  $d$  mens kjøperen får det høyeste utbyttet  $a$ .

Dette gir oss det velkjente resultatet fra fangenes dilemma spill. Siden ‘ikke samarbeid’ gir det høyeste utbyttet for hver av spillerne uansett hvilken handling den andre spilleren velger, blir utfallet at ingen av partene samarbeider.<sup>4</sup> Begge får dermed  $c$  i utbytte som er lavere enn utbyttet de hadde fått dersom begge hadde samarbeidet (som ville gitt  $b$ ). Individuell rasjonalitet leder altså til kollektiv urasjonalitet.

For å forenkle notasjonen og regningen så skal vi heretter anta at  $c = 0$  og  $d = -a$  i spillmatrisen ovenfor. Vi får da følgende spillmatrise

	Samarbeid	Ikke samarbeid
Samarbeid	$b, b$	$-a, a$
Ikke samarbeid	$a, -a$	$0, 0$

Merk imidlertid at disse endringene i notasjon ikke endrer noe på spillet. Utfallet vil fortsatt være at begge spillerne velger ikke samarbeid. Innenfor rammen av fangenes dilemma spillet ovenfor skal vi nå undersøke nærmere muligheten for at samarbeid kan vokse fram over tid. For å kunne gjøre dette er det imidlertid nødvendig å si noe mer om hva som kan motivere den enkelte til å velge samarbeid i en situasjon som den beskrevet ovenfor, og dernest hvordan denne motivasjonen kan tenkes å bli endret over tid. Jeg starter med det siste først.

---

<sup>4</sup>Husk at en underliggende antakelse her er at partene ikke har anledning til å skrive bindende avtaler som straffer den som avviker samarbeid.

### 3 Kulturell evolusjon

Inspirert av evolusjonær spillteori skal jeg anta at handlingsmønstre overføres fra et individ til et annet gjennom imitasjon eller en annen form for sosial læring. Spesielt vil jeg anta at handlingsmønstre som oppfattes som suksessfulle av medlemmene i samfunnet vil spre seg over tid.<sup>5</sup> Med suksessfull vil jeg videre mene at vedkommende handling gir et høyt utbytte i den sosiale interaksjonen (spillet) som medlemmene av samfunnet er involvert i.

En slik evolusjonær prosess kan presiseres mer formelt på følgende måte. Anta et samfunn bestående av et stort antall individer som kan deles opp i to typer, *samarbeidere* ( $S$ -typer) og *opportuniste* ( $O$ -typer). Hvilken type en er bestemmer hvilken handling en velger i spillet en er involvert i. La  $p \in (0, 1)$  være andelen samarbeidere i samfunnet der  $\dot{p}$  angir veksten i andelen samarbeidere over tid.<sup>6</sup> Dersom spillerne møtes parvis, og det er tilfeldig hvem en møter, så kan forventet utbytte til de to typene skrives

$$U(S) = pm(S, S) + (1 - p)m(S, O) \quad (1)$$

$$U(O) = pm(O, S) + (1 - p)m(O, O) \quad (2)$$

der  $m(i, j)$  er utbyttet til en  $i$ -type når han spiller mot en  $j$ -type. Relasjon (1) leses på følgende måte: Med sannsynlighet  $p$  møter en  $S$ -type en annen  $S$ -type noe som gir utbyttet  $m(S, S)$ , mens med sannsynlighet  $(1 - p)$  møter den samme  $S$ -typen en  $O$ -type noe som gir utbyttet  $m(S, O)$ . Relasjon (2) har en tilsvarende tolkning. Med sannsynlighet  $p$  møter en  $O$ -type en  $S$ -type noe som gir utbyttet  $m(O, S)$ , mens med sannsynlighet  $(1 - p)$  møter den en  $O$ -type en annen  $O$ -type noe som gir utbyttet  $m(O, O)$ .

På ethvert tidspunkt vil veksten i andelen samarbeidere i samfunnet være positiv eller negativ avhengig av om forventet utbytte til samarbeiderne er større eller mindre enn forventet utbytte til opportunistene, det vil si<sup>7</sup>

$$\dot{p} \begin{matrix} \geq \\ \leq \end{matrix} 0 \text{ alt etter som } U(S) \begin{matrix} \geq \\ \leq \end{matrix} U(O) \quad (3)$$

Relasjon (3) sier altså at andelen samarbeidere vil vokse eller avta over tid avhengig av om de har et høyere eller lavere forventet utbytte enn oppor-

---

<sup>5</sup>Se Boyd og Richerson (1985), og Richerson, Boyd og Henrich (2003) og Bowles (2003, kap. 2) for en mer utfyllende diskusjon og analyse av en slik "kulturell" evolusjonær prosess.

<sup>6</sup>Andelen opportuniste blir da  $(1 - p)$ .

<sup>7</sup>Weibull (1995) kaller dette antakelsen om *payoff monotonicity*. For å studere payoff monoton dynamikk trenger vi med andre ord kun å sammenligne forventet utbytte til de to typene.



tunistene.

Vi kan tolke dynamikken skissert ovenfor på følgende måte: Samfunnet starter ut med en gitt fordeling av samarbeidere og opportunister som møtes tilfeldig. Hver spiller velger handling i henhold til hvilken type de er noe som igjen tilskriver hver av spillerne et bestemt utbytte. Etter at den sosiale interaksjonen har funnet sted vurderer en andel av spillerne om de skal endre sitt valg av handling. Gitt at en spiller vurderer å endre handling, så er denne beslutningen basert på en sammenligning av eget utbytte med utbyttet som den alternative handlingen genererer. Hvis den alternative handlingen genererer et høyere utbytte enn det spilleren får selv, så vil spilleren bytte til den alternative handlingen. Hvis den alternative handlingen generer et lavere utbytte så vil spilleren holde på den opprinnelige måten å handle på.

Likevekt vil i denne sammenhengen da bety at vi ikke får noen endring i andelen samarbeidere eller opportunister over tid ( $\dot{p} = 0$ ). I henhold til relasjon (3) ser vi at likevekt oppnås når  $U(S) = U(O)$ , det vil si at forventet utbytte til de to typene er den samme. En slik likevekt ( $p^*$ ) kan være en av følgende tre typer. Dersom  $p^* = 0$  har vi et samfunn kun bestående av opportunister, dersom  $p^* = 1$  har vi et samfunn kun bestående av samarbeidere og dersom  $p^* \in (0, 1)$  har vi et blandet samfunn bestående av både samarbeidere og opportunister. En likevekt kan videre være stabil eller ustabil. Dersom  $p^*$  tilfredsstillers  $U(S) = U(O)$ , så er dette en stabil likevekt dersom en liten bevegelse bort fra  $p^*$  leder til en endring i andelen samarbeidere tilbake til  $p^*$ . Stabilitet krever altså at dersom andelen samarbeidere øker, så gir en økning i forventet utbytte for opportunistene relativt til samarbeiderne, noe som igjen leder til en økning i andelen opportunister (en dermed en reduksjon i  $p$ ). Kravet til stabilitet er altså at

$$\frac{d(U(S) - U(O))}{dp} < 0 \quad (4)$$

## 4 Internaliserte normer om samarbeid

Hva kan forklare at noen vil velge å samarbeide innenfor et fangenes dilemma spill? Dilemmaet i dette spillet er jo nettopp at for den enkelte vil det være best å velge 'ikke samarbeid' uansett hva den andre parten gjør. Som nevnt har sosiologer lenge vist til at normer kan begrense folks tilbøyelighet til

å handle egennyttig og opportunistisk.<sup>8</sup> La oss derfor i spillet ovenfor introdusere ideen om at noen av aktørene er motivert av en norm om samarbeid som sier at “*det er galt å ikke samarbeide*”. La oss kalle de som er “bærere” av denne normen for *samarbeidere* (*S*-typer). La oss videre anta normen om samarbeid er fullstendig internalisert hos de som bærer den, det vil si at de vil ta hensyn til den selv om handlinger som representerer et brudd med normen ikke blir oppdaget og sanksjonert av andre. Motivet for å handle i tråd med normen om samarbeid er ønsket om å unngå følelsen av skyld, skam og tap av selvrespekt.<sup>9</sup> La  $\tau > 0$  angi den subjektive sosiopsykologiske kostnaden ved å bryte normen om samarbeid. Når to samarbeidere møtes vil dermed fangenes dilemma spillet endres til spillet angitt i spillmatrisen nedenfor

	Samarbeid	Ikke samarbeid
Samarbeid	$b, b$	$-a, a - \tau$
Ikke samarbeid	$a - \tau, -a$	$0 - \tau, 0 - \tau$

Spillmatrisen viser at hvis  $\tau > a$  så blir ‘samarbeid’ en dominerende strategi for spillerne. Når to samarbeidere møtes vil altså begge velge ‘samarbeid’ og få  $b$  i utbytte. De som ikke har internalisert normen om samarbeid har  $\tau = 0$ , og vil dermed ha ‘ikke samarbeid’ som dominerende strategi. Jeg kaller disse for *opportunist* (*O*-typer). Når to opportunist møtes vil derfor begge velge ‘ikke samarbeid’ og få 0 i utbytte.

Hvordan vil andelen samarbeidere og opportunist utvikle seg over tid når vi antar at spillernes handlinger utvikler seg i tråd med den evolusjonære prosessen beskrevet i avsnitt 3, og spillerne møtes tilfeldig.<sup>10</sup> Som nevnt ovenfor vil samarbeiderne alltid velge ‘samarbeid’, mens opportunistene alltid velger ‘ikke samarbeid’. En samarbeider vil videre møte en annen samarbeider med sannsynlighet  $p$  (som er andelen samarbeidere i samfunnet). Disse vil da samarbeide og få utbyttet  $b$ . En samarbeider vil møte en

---

<sup>8</sup>Se Horne (2001) for en gjennomgang av sentrale bidrag til den sosiologiske litteraturen som behandler begrepet ‘normer’. Borgstede (2002) analyserer normer i sosiale dilemma situasjoner mer spesielt, mens Elster (1989) og Torsvik (2003) gir en mer generell diskusjon av sammenhengen mellom normer, rasjonalitet og egennyttig handling.

<sup>9</sup>Bowles og Gintis (2001) gir en interessant diskusjon av forhold som kan forklare hvorfor vi er utstyrt med evnen til å internalisere normer, og hvilken rolle følelser som skyld og skam spiller for økonomisk handling.

<sup>10</sup>Husk at den evolusjonære prosessen beskrevet ovenfor innebærer at selv om samarbeiderne bryr seg om normen om samarbeid, så er “overlevelsesevnen” til normen bestemt av utbyttet som følger av den handlingen normen tilskriver. Det betyr altså at dersom handlinger som bryter med normen om samarbeid generer et høyere utbytte, så er det en tendens til at normen brytes også av de som i utgangspunktet var “bærere” av den.

oppportunist med sannsynlighet  $1-p$  (som er andelen oppportunister i samfunnet). Samarbeideren vil da få  $-a$  mens oppportunisten får  $a$ . En oppportunist vil møte en annen oppportunist med sannsynlighet  $1-p$ , som gir begge 0 i utbytte. Forventet utbytte til henholdsvis samarbeiderne og oppportunistene blir da

$$U(S) = pb + (1-p)(-a) = pb - (1-p)a \quad (5)$$

$$U(O) = pa - (1-p)0 = pa \quad (6)$$

Siden  $a > b$  ser vi med en gang at forventet utbytte til for samarbeiderne er mindre enn forventet utbytte for oppportunistene ( $U(S) < U(O)$ ) for alle  $p \in (0, 1)$ . I henhold til (3) vet vi da at andelen samarbeidere vil avta over tid inntil  $p = 0$ . Situasjonen er illustrert i figur 1.

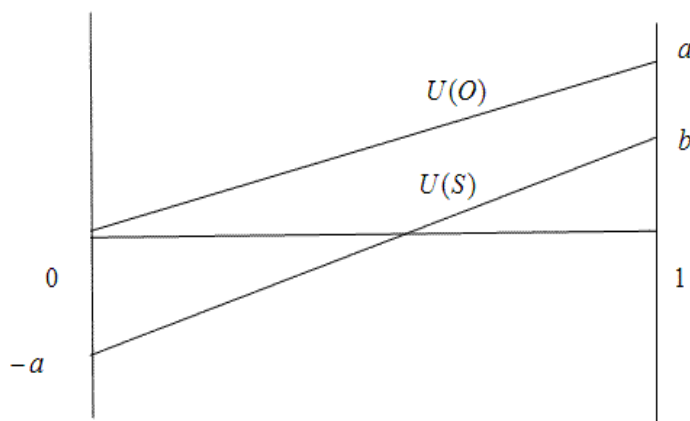


Figure 1:

Den eneste stabile likevekten er altså en situasjon der ingen i samfunnet til slutt samarbeider. Dette resultatet kan forklares med at når samarbeiderne er drevet av en betingelsesløs norm om samarbeid, så vil de samarbeide uansett hvem de møter. Dette høster oppportunistene en fordel av siden de da får det høyeste utbyttet  $a$  hver gang de møter en samarbeider. Oppportunistene utnytter samarbeidernes udiskriminerende samarbeid, og får som et resultat av det et høyere forventet utbytte. Når samarbeiderne observerer dette vil de etterhvert slutte å samarbeide og heller velge 'ikke samarbeid' (samarbeiderne blir oppportunister). Denne prosessen vil pågå helt til det ikke er noen igjen som samarbeider ( $p = 0$ ).

Konklusjonen er altså at en ubetinget norm om samarbeid ikke er levedyktig i situasjonen studert ovenfor. En betinget norm som sier at “*det er galt å ikke samarbeide med andre som også samarbeider*” er imidlertid mindre utsatt for å bli utnyttet av opportunistene.<sup>11</sup> Vi kan framstille en slik betinget norm om samarbeid i fangenes dilemma spillet ovenfor på følgende måte. La som før  $\tau > 0$  angi den subjektive sosiopsykologiske kostnaden ved å bryte normen om samarbeid, der det kun er samarbeiderne som er bærere av en slik norm. For opportunistene er  $\tau = 0$ . Når to inspektører med en betinget norm om samarbeid møtes vil fangenes dilemma spillet endres til

	Samarbeid	Ikke samarbeid
Samarbeid	$b, b$	$-a, a - \tau$
Ikke samarbeid	$a - \tau, -a$	$0, 0$

Vi ser at dersom  $\tau > a - b$  så vil begge spillerne ønske å samarbeide bare dersom den andre samarbeider. Det opprinnelige spillet endres fra å være et fangenes dilemma spill til å bli et koordineringsspill. Dette spillet har to Nash-likevekter. Enten velger begge spillerne ‘samarbeid’ og får  $(\beta, \beta)$ , eller begge velger ‘ikke samarbeid’ og får  $(0, 0)$ .<sup>12</sup> I det følgende skal jeg anta at når to samarbeidere møtes så vil de samarbeide.<sup>13</sup> Siden opportunistene, som før, har  $\tau = 0$  vil de spille fangenes dilemma spillet når de møtes. Begge velger ‘ikke samarbeid’ og får 0 i utbytte.

Med en betinget norm om samarbeid så vil samarbeiderne samarbeide med andre samarbeidere og velge ‘ikke samarbeid’ når de møter en opportunist. De får dermed et forventet utbytte på  $U(S) = pb - (1 - p)0 = pb$ . Opportunistene blir møtt med ‘ikke samarbeid’ uansett om de møter en samarbeider eller en opportunist, og får dermed 0 i utbytte. Utbyttet til samarbeiderne er dermed større enn utbyttet til  $(U(S) < U(O))$  for alle  $p \in (0, 1)$ . I henhold til (3) har vi da at andelen samarbeidere vil avta

<sup>11</sup>Se Opp (2001) for en diskusjon av forskjellen mellom en *betinget* og en *ubetinget* norm.

<sup>12</sup>Det eksisterer også en likevekt i blandede strategier der begge spillerne velger å samarbeide med sannsynlighet  $\alpha / (\beta + \tau)$ .

<sup>13</sup>Denne antakelsen kan begrunnes mer formelt ved å introdusere begrepet *risiko dominant* likevekt (se for eksempel Young (1998)). Anta at begge spillerne er usikre på hvilken strategi den andre vil velge. Dersom begge tror at den andre velger samarbeid med sannsynlighet 0.5, så blir forventet utbetaling ved å samarbeide  $(\beta - \alpha) / 2$ , mens forventet utbetaling ved ikke å samarbeide blir  $(\alpha - \tau - 0) / 2$ . Dersom spillerne er opptatt av å maksimere forventet nytte så vil de velge ‘samarbeid’ dersom  $(\beta - \alpha) / 2 > (\alpha - \tau - 0) / 2$ , eller dersom  $\tau > 2\alpha - \beta$ . Dette er det samme som å si at samarbeid er risikodominant. Dersom  $\tau > 2\alpha - \beta$  er oppfylt så vil altså to samarbeidere velge ‘samarbeid’ når de møtes.

over tid inntil  $p = 1$  (alle samarbeider). En betinget norm om samarbeid er altså levedyktig, og vil over tid spres seg til alle medlemmer av samfunnet. Dette resultatet hviler imidlertid på antakelsen om at samarbeiderne har informasjon om hvilken type de møter. Dette kan være en urealistisk antakelse i mange sammenhenger. La oss derfor anta at samarbeiderne i utgangspunktet ikke vet hvilken type de møter, men at de kan skaffe seg denne informasjonen gjennom undersøkelser (inspeksjon).

#### 4.1 Inspeksjon

Anta nå som ovenfor at samarbeiderne er innstilt på å samarbeide, men at de først vil undersøke eller inspisere nærmere hvem de møter før de faktisk velger å samarbeide. Anta videre at en slik inspeksjon med sikkerhet avslører hvem en står overfor. Det vil altså si at ved å undersøke vil en samarbeider vite om han spiller mot en annen samarbeider eller en opportunist. Men siden inspeksjon som regel tar tid og krefter så skal vi anta at en slik handling påfører samarbeiderne en kostnad  $K$ .<sup>14</sup>

Som ovenfor skal vi la  $p$  være andelen samarbeidere i samfunnet, mens  $1-p$  er andelen opportunister. Dynamikken i spillet er som følger. Samfunnet starter ut med en gitt fordeling av opportunister og samarbeidere. Disse møtes parvis, der det er tilfeldig hvem en møter. Samarbeiderne undersøker nærmere hvem de møter (noe som koster  $K$ ) og velger handling betinget av hvilken type de møter. Etter at den sosiale interaksjonen har funnet sted endrer sammensetningen av samfunnet seg i henhold til det relative utbyttet til de to typene, som beskrevet i avsnitt 3.

Forventet utbytte for samarbeiderne og opportunistene blir

$$U(S) = pb + (1-p)0 - K = pb - K \quad (7)$$

$$U(O) = 0 \quad (8)$$

Likevekt oppnås når  $U(S) = U(O)$ , som gir

$$p^* = K/b \quad (9)$$

der  $p^* \in (0,1)$  hvis  $K < b$ . Vi får altså en indre likevekt hvor både samarbeidere og opportunister eksisterer dersom kostnaden ved å undersøke hvem en samhandler med en mindre enn  $b$ .<sup>15</sup> Denne indre likevekten er imidlertid

<sup>14</sup>Denne ideen er tatt fra Güth og Kliemt (1998).

<sup>15</sup>Dersom  $K > b$  vil eneste stabile likevekt være et samfunn bestående av bare opportunister.

ustabil siden

$$\frac{d(U(S) - U(O))}{dp} = b > 0 \quad (10)$$

En liten økning i  $p$  bort fra  $p^*$  vil altså øke det forventede utbyttet til samarbeiderne relativt til opportunistene, noe som igjen leder til en ytterligere økning i andelen inspektører. Situasjonen er illustrert i figur 2.

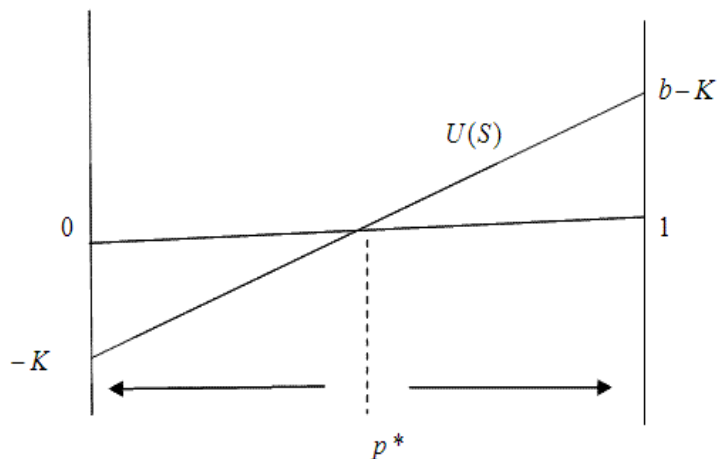


Figure 2:

Vi får dermed tre likevekter, nemlig 0,  $p^*$  og 1, hvorav den første og siste er stabile. Dette kan forklares på følgende måte. Anta et samfunn der ingen i utgangspunktet samarbeider ( $p = 0$ ) og at det av en eller annen grunn dukker opp en lite gruppe samarbeidere. Disse vil da gjøre det dårligere enn opportunistene siden det er så få andre å samarbeide med ( $U(S) < U(O)$ ). Den lille gruppen av samarbeidere vil da gradvis endre strategi til ‘ikke samarbeid’ inntil det ikke er noen igjen som samarbeider. La oss så betrakte et samfunn der alle i utgangspunktet samarbeider ( $p = 1$ ) og at det av en eller annen grunn dukker opp en liten gruppe opportuniste som ikke samarbeider. Disse opportunistene vil da bli møtt med ‘ikke samarbeid’ av samarbeiderne, mens samarbeiderne samarbeider med hverandre. Samarbeiderne vil da gjøre det bedre enn opportunistene ( $U(S) > U(O)$ ) noe som vil føre til at opportunistene gradvis endrer strategi til ‘samarbeid’. Denne prosessen vil pågå helt til alle i samfunnet samarbeider. Den indre likevekten  $p^*$  definerer derfor et kritisk “vippepunkt” i den forstand at straks andelen

samarbeidere er større enn  $p^*$  så vil andelen samarbeidere vokse inntil  $p = 1$  (alle samarbeider).

Fra (9) får vi videre at

$$\frac{dp^*}{dK} = \frac{1}{b} > 0 \quad (11)$$

$$\frac{dp^*}{db} = -\frac{K}{b^2} < 0 \quad (12)$$

En økning i kostnaden ved å undersøke flytter  $p^*$  mot høyre. Det kreves dermed en større andel samarbeidere i utgangspunktet for å starte en utvikling i retning av den stabile likevekten  $p = 1$ . En økning i utbetalingen ved gjensidig samarbeid flytter  $p^*$  mot venstre, noe som krever en mindre andel samarbeidere i utgangspunktet for å starte en utvikling i retning av den stabile likevekten  $p = 1$ .<sup>16</sup>

Modellen illustrerer dermed at den samme underliggende sosiale interaksjonen kan lede til to svært forskjellige stabile situasjoner. Hvor samfunnet ender opp avhenger av hvor det starter. Dersom den initiale andelen samarbeidere er liten (mindre enn  $p^*$ ) så vil samfunnet bevege seg i en retning der ingen samarbeider. Dersom den initiale andelen samarbeidere er stor (større enn  $p^*$ ) så vil samfunnet bevege seg i en retning der alle samarbeider. Begge situasjonene er mulig, det vil si at den evolusjonære prosessen er *historie-betinget*.<sup>17</sup>

## 4.2 Gjentatt interaksjon<sup>18</sup>

Ovenfor er det antatt at partene kun møtes en gang, eventuelt at den enkelte ikke husker hvem de har møtt tidligere. La oss nå endre på denne antakelsene

<sup>16</sup>Et talleksempel kan illustrere dette nærmere. Anta at  $a = 4$ ,  $b = 2$  og  $K = 1$ . Fra (4) får vi da at  $p^* = 0.5$ . Dersom  $K$  øker til 1.5 (og  $b$  holdes konstant) så øker  $p^*$  til 0.75. Dersom  $b$  øker til 3 (og  $K$  holdes konstant) så synker  $p^*$  til 0.3.

<sup>17</sup>En slik historiebetiget evolusjonær prosess står sentralt i Marx sin historieoppfatning, noe som blant annet følgende kommentar i *The Eighteenth Brumaire of Louis Bonaparte* (1852) uttrykker: “*Men make history, but they do not make it just as they please; they do not make it under circumstances chosen by themselves but under circumstances (...) given and transmitted from the past. The tradition of all the dead generations weights like a nightmare on the brain of the living.*” (Marx, 1852(1963) s. 15). Modellen ovenfor, og variantene av den samme modellen som analyseres nedenfor, kan betraktes som et eksempel på hvordan en slik prosess kan modelleres mer formelt. Rollen et samfunns historie spiller for utviklingen av normer og institusjoner, og for hvordan et samfunnet ser ut “i dag”, er også studert nærmere av North (1990), Putnam (1993), Platteau (2000), Rothstein (2000) samt Hechter og Opp (2001). Se også Bowles 2004 (kap. 1 og 12).

<sup>18</sup>Avsnitt 4.2 og 4.3 bygger på Bowles og Gintis (1998).

ovenfor ved å anta at partene møtes gjentatte ganger og at de er i stand til å huske (uten å undersøke) hvilken handling motparten valgte i forrige runde av spillet. En slik situasjon kan framstilles formelt ved å anta at etter hver runde spillet er spilt så fortsetter spillet med sannsynlighet  $\beta$ . Forventet antall runder spillet spilles blir da  $1 + \beta + \beta^2 + \beta^3 + \beta^4 + \dots + \beta^n + \dots$ . Dersom vi bruker formelen for en uendelig geometrisk rekke så vil dette konvergere mot  $1/(1 - \beta)$ . Anta videre at de sosiale typene følger det som gjerne blir kalt en *tit-for-tat* strategi:<sup>19</sup> Start ut i første periode med å samarbeide. I etterfølgende perioder, velg den strategien som partneren valgte i den forrige periode. Som før velger opportunistene alltid 'ikke samarbeid'. Dette betyr at dersom to sosiale typer møtes, noe som skjer med sannsynlighet  $p$ , så vil de starte ut med å samarbeide og fortsette å samarbeide i alle etterfølgende perioder. Dette gir dem et forventet utbytte på  $b/(1 - \beta)$ . Dersom en sosial type møter en opportunist, noe som skjer med sannsynlighet  $(1 - p)$ , så vil den sosiale typen samarbeide mens opportunisten velger 'ikke samarbeid'. Samarbeiderne får  $-a$ , mens opportunistene får  $a$ . Dette vil føre til at den sosiale typen bytter til 'ikke samarbeid' i neste periode og at begge parter holder seg til ikke samarbeid i alle etterfølgende perioder, noe som gir begge et utbytte på 0. Dersom to opportunister møtes vil begge velge 'ikke samarbeid' i første periode og i alle etterfølgende perioder, noe som gir et utbytte på 0. Forventet utbytte til de to typene kan da skrives

$$U(S) = p \left( \frac{b}{1 - \beta} \right) + (1 - p)(-a + 0) = \frac{pb}{1 - \beta} - (1 - p)a \quad (13)$$

$$U(O) = p(a + 0) + (1 - p)0 = pa \quad (14)$$

Likevekt har vi når  $U(S) = U(O)$  som gir

$$p^* = \frac{a(1 - \beta)}{b} \quad (15)$$

der  $p^* \in (0, 1)$  dersom  $b > a(1 - \beta)$  eller

$$\beta > \frac{a - b}{a} \equiv \beta^* \quad (16)$$

(16) forteller oss at vi får en indre likevekt hvor både samarbeidere og opportunister eksisterer dersom sannsynligheten for at spillet fortsetter etter

---

<sup>19</sup>Denne strategien er blitt kjent spesielt via arbeidene til Axelrod (1984).



hver periode spillet er spilt er større enn  $\beta^*$ .<sup>20</sup> Men som i avsnitt 4.1 så vil denne likevekten være ustabil siden

$$\frac{d(U(S) - U(O))}{dp} = \frac{b}{1 - \beta} > 0 \quad (17)$$

En liten økning i  $p$  bort fra  $p^*$  øker det forventede utbyttet til samarbeiderne relativt til opportunistene, noe som igjen leder til en ytterligere økning i andelen inspektører. Situasjonen er illustrert i figur 3.

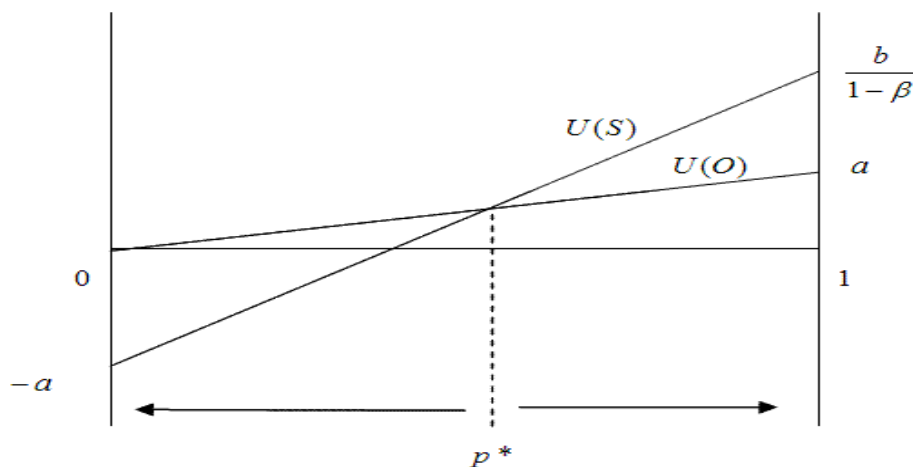


Figure 3:

Som i avsnitt 4.1 får vi tre likevekter, nemlig 0,  $p^*$  og 1, hvorav den første og siste er stabile. Den indre likevekten  $p^*$  angir igjen hvor stor andelen samarbeidere må være for å sette i gang en prosess der andelen samarbeidere vokser inntil  $p = 1$  (alle i samfunnet samarbeider).

Fra (15) får vi videre at

$$\frac{dp^*}{d\beta} = -\frac{a}{b} < 0 \quad (18)$$

$$\frac{dp^*}{db} = -\frac{a(1 - \beta)}{b^2} < 0 \quad (19)$$

$$\frac{dp^*}{da} = \frac{1 - \beta}{b} > 0 \quad (20)$$

<sup>20</sup>Dersom  $\beta < \beta^*$  vil eneste stabile likevekt være et samfunn bestående av bare opportunistene.

(18) og (19) sier altså at en økning i sannsynligheten for at spillet fortsetter, eller en økning i gevinsten ved at begge samarbeider, flytter  $p^*$  mot venstre. Det kreves dermed en mindre andel samarbeidere i utgangspunktet for å starte en utvikling i retning av den stabile likevekten der alle samarbeider ( $p = 1$ ). (20) sier at en økning i gevinsten ved å avvike samarbeid dersom motparten samarbeider flytter  $p^*$  mot høyre, noe som vanskeliggjør samarbeid.<sup>21</sup>

### 4.3 Segregering

Til nå har vi antatt at de to typene spillere møtes tilfeldig for å spille spillet definert ovenfor (det vil si at sannsynligheten for at en samarbeider møter en annen samarbeider avhenger av hvor mange samarbeidere som finnes i samfunnet). Dette er en urealistisk antakelse i mange sammenhenger. Som nevnt innledningsvis er det grunn til å tro at de som er innstilt på å samarbeide utvikler strategier for å unngå de som handler opportunistisk. De som ønsker å samarbeide til alles beste, og som ikke har som mål å sikre seg selv på bekostning av andre, kan for eksempel oppsøke markeder/butikker/organisasjoner/steder der de av erfaring vet at det er stor sannsynlighet for å møte personer av samme type som en selv. Dette kan vi ta hensyn til ved å anta at sannsynligheten for at en samarbeider møter en annen samarbeider ikke lenger er  $p$ , men  $r + (1 - r)p$ , hvor  $r \in (0, 1)$  er graden av segregering i samfunnet. Dersom  $r = 1$ , har vi fullstendig segregering i den forstand at samarbeiderne aldri møter en opportunist. Dersom  $r = 0$  har vi ingen segregering, det vil si at vi er tilbake til situasjonen hvor det er tilfeldig hvem en møter. Sannsynligheten for at en opportunist møter en annen opportunist er  $r + (1 - r)(1 - p)$ . Forventet utbytte til den samarbeiderne og opportunistene blir da

$$U(S) = [r + (1 - r)p]b + [(1 - r)(1 - p)](-a) \quad (21)$$

$$U(O) = [(1 - r)p]a + [r + (1 - r)(1 - p)]0 \quad (22)$$

---

<sup>21</sup>Et talleksempel kan igjen illustrere dette nærmere. Anta at  $a = 4$  og  $b = 2$ . Av (16) får vi da at  $\beta > 0.5$ . Dersom vi antar at  $\beta = 0.7$  får vi at  $p^* = 0.6$ , som betyr at andelen samarbeidere må være større enn 0.6 for å starte en utvikling der alle i samfunnet til slutt samarbeider. La så sannsynligheten for at spillet forsetter øke til 0.9. Av (15) får vi da at  $p^*$  reduseres til 0.4. Likeledes finner vi at  $p^*$  reduseres dersom  $b$  øker, mens  $p^*$  øker dersom  $a$  øker.

Likevekt oppnås når  $U(S) = U(O)$  som gir

$$p^* = \frac{(1-r)a - rb}{(1-r)b} \quad (23)$$

Nevneren i (23) er alltid positiv. For at  $p^*$  skal være positiv må da telleren i (23) også være positiv, det vil si at  $r > (a-b)/a$ . Videre må vi ha at  $(1-r)b > (1-r)a - rb$  for at  $p^* \in (0,1)$ . Den siste betingelsen gir at  $r < a/(a+b)$ . Til sammen betyr dette at  $p^* \in (0,1)$ , noe som gir en indre likevekt, dersom

$$\frac{a-b}{a} < r < \frac{a}{a+b} \quad (24)$$

Vi kan da etablere følgende resultat: Dersom graden av segregering er liten, det vil si dersom  $r < (a-b)/a$ , så vil forventet utbytte til samarbeiderne være lavere enn forventet utbytte til opportunistene ( $U(S) < U(O)$ ) for alle  $p^* \in (0,1)$ , noe som gjør  $p = 0$  (ingen samarbeider) til eneste stabile likevekt. Dersom graden av segregering er stor, det vil si dersom  $r > a/(a+b)$ , så vil forventet utbytte til samarbeiderne være høyere enn forventet utbytte til opportunistene ( $U(S) > U(O)$ ) for alle  $p^* \in (0,1)$ , noe som gjør  $p = 1$  (alle samarbeider) til eneste stabile likevekt. Dersom  $r$  verken er liten eller stor, det vil si dersom (24) er oppfylt, så vil vi ha en indre likevekt. Men siden

$$\frac{d(U(S) - U(O))}{dp} = (1-r)b > 0 \quad (25)$$

så er den indre likevekten ustabil. En liten økning i  $p$  bort fra  $p^*$  øker det forventede utbyttet til samarbeiderne relativt til opportunistene, noe som igjen leder til en ytterligere økning i andelen inspektører. Situasjonen er illustrert i figur 4.

Som i avsnitt 4.1 og 4.2 får vi igjen tre likevekter, nemlig 0,  $p^*$  og 1, hvorav den første og siste er stabile. Den indre likevekten  $p^*$  angir hvor stor andelen samarbeidere må være for å sette i gang en prosess der andelen samarbeidere vokser inntil  $p = 1$  (alle i samfunnet samarbeider).

Fra (23) får vi videre at

$$\frac{dp^*}{dr} = -\frac{1}{b(1-r)^2} < 0 \quad (26)$$

som forteller oss at en økning i graden av segregering flytter  $p^*$  mot venstre. Det kreves dermed en mindre andel samarbeidere i utgangspunktet for å starte en utvikling i retning av den stabile likevekten der alle til slutt

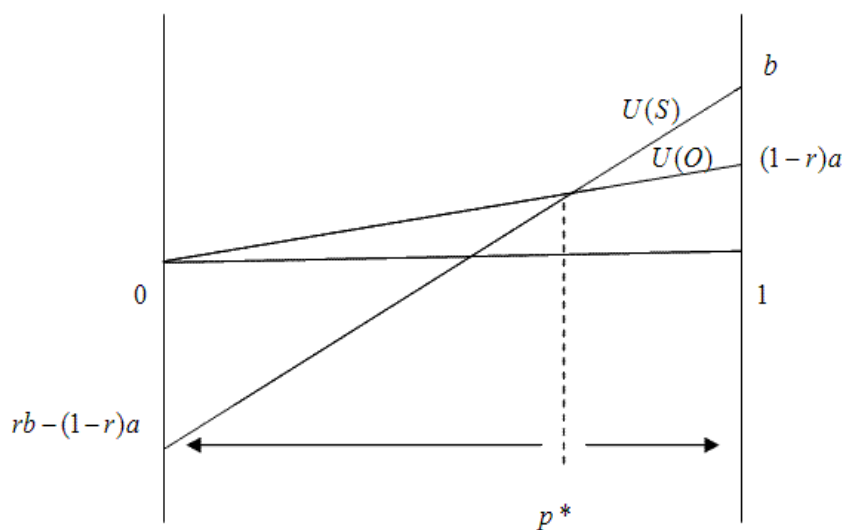


Figure 4:

samarbeider ( $p = 1$ ).

## 5 Ekstern sanksjonering av normer

Det som karakteriserte en samarbeider i analysen ovenfor var at vedkommende hadde internalisert en norm som sier at "det er galt å ikke samarbeide". Når den sosiopsykologiske kostnaden ved å bryte denne normen var stor nok, så motiverte det den enkelte til å velge samarbeid innenfor en sosial dilemmasituasjon. Nedenfor skal vi droppe antakelsen om at normen om samarbeid er internalisert. I stedet skal vi betrakte et samfunn hvor det eksisterer en gruppe individer som ønsker å samarbeide (samarbeidere), og at disse lever ved siden av en annen gruppe som ikke ønsker å samarbeide (opportunistene). Grunnen til at samarbeiderne ønsker å samarbeide skal vi anta ligger i at de ønsker å unngå og bli straffet av andre medlemmer av samfunnet. De som velger 'ikke samarbeid' når de møter en som samarbeider vil bli straffet med størrelsen  $\gamma$ . Vi skal imidlertid anta at dette kun inntreffer med en viss sannsynlighet  $\sigma$ .<sup>22</sup> Sannsynligheten  $\sigma$  er ment å måle samfunnets tilbøyelighet til å straffe opportunistisk adferd. En stor verdi på

<sup>22</sup>Denne ideen er tatt fra Sacco (1997).

$\sigma$  (dersom  $\sigma \rightarrow 1$ ) kan tolkes som at normen om samarbeid er dypt rotfestet i samfunnet, og at medlemmene er svært opptatt av å straffe brudd på normen. En liten verdi på  $\sigma$  (dersom  $\sigma \rightarrow 0$ ) kan tolkes som at medlemmene av samfunnet bare er moderat opptatt av å straffe brudd på normen om samarbeid. Til slutt skal vi som en forenkling anta at det å straffe noen for brudd på normen om samarbeid ikke koster noe for de som straffer.<sup>23</sup>

Oppportunistene bryr seg i utgangspunktet ikke om å bli straffet for brudd på normen om samarbeid. Men, som vi skal se, så kan tapet ved å bli straffet bli så stort for oppportunistene at de likevel vil velge å bytte strategi til 'samarbeid'. Målet med analysen nedenfor er å vise mer presist under hvilke betingelser dette vil gjelde.

Som ovenfor skal vi anta at medlemmene av samfunnet er involvert i fangenes dilemma spillet beskrevet i avsnitt 2. Spillerne møtes parvis, og det er tilfeldig hvem en møter. Andelen samarbeidere i samfunnet er  $p$ , mens andelen oppportunister er  $(1 - p)$ . Det vi er opptatt av er å studere hvordan andelen oppportunister og samarbeidere vil utvikle seg over tid når spillernes handlinger utvikler seg i tråd med den evolusjonære prosessen beskrevet i avsnitt 3. Med antakelsene ovenfor blir forventet utbytte til samarbeiderne og oppportunistene

$$U(S) = pb + (1 - p)(-a) \quad (27)$$

$$U(O) = pa + (1 - p)0 - \sigma\gamma \quad (28)$$

Likevekt oppstår når  $U(S) = U(O)$  som gir

$$p^* = \frac{a - \sigma\gamma}{b} \quad (29)$$

For at  $p^*$  skal være positiv må  $a - \sigma\gamma > 0$ , det vil si at  $\sigma < a/\gamma$ . Videre må vi ha at  $b > a - \sigma\gamma$  for at  $p^* \in (0, 1)$ . Den siste betingelsen gir at  $\sigma > (a - b)/\gamma$ . Til sammen betyr dette at  $p^* \in (0, 1)$  dersom

$$\frac{a - b}{\gamma} < \sigma < \frac{a}{\gamma} \quad (30)$$

---

<sup>23</sup>Hvorfor vil medlemmene av samfunnet straffe andre som bryter en norm om samarbeid? Jeg skal la dette spørsmålet ligge åpent, og kun nøye meg med å vise til en omfattende litteratur som viser at folk gjennomgående straffer antisosial adferd og brudd på normer, selv i tilfeller der slik straff innebærer betydelige kostnader for de som straffer, og hvor straffing av brudd på normer heller ikke tjener de langsiktige interessene til den som straffer (Fehr og Gächter, 2002; Fehr m.fl., 2002; Gintis m.fl., 2003; Fehr og Fischbacher, 2004).

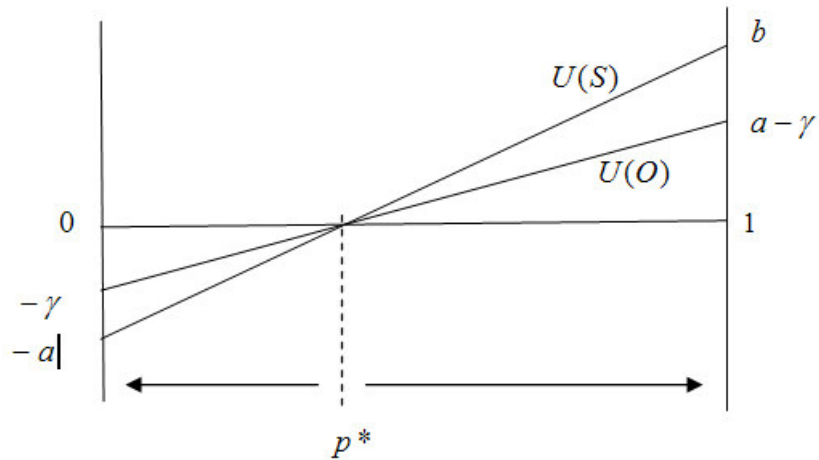


Figure 5:

For studere den evolusjonære prosessen nærmere er det hensiktsmessig først å betrakte situasjonen hvor  $\sigma = 1$ , det vil si at de som bryter normen om samarbeid alltid blir straffet med størrelsen  $\sigma$ .

### 5.1 Normbrytere blir alltid straffet ( $\sigma = 1$ )

Anta at de som bryter normen om samarbeid alltid blir straffet med størrelsen  $\gamma$ . Fra (29) har vi da at  $p^* \in (0, 1)$  dersom  $a - b < \gamma < a$ . Denne situasjonen er illustrert i figur 5.

Når  $\sigma = 1$  blir forventet utbytte til samarbeiderne  $U(S) = pb - (1 - p)a$ , mens opportunistene får  $U(O) = pa - \gamma$ . Dersom  $p > p^*$  er  $U(S) > U(O)$ , mens  $U(S) < U(O)$  for  $p < p^*$ . Den indre likevekten  $p^*$  er altså en ustabil likevekt. En liten økning i  $p$  bort fra  $p^*$  vil øke det forventede utbyttet til samarbeiderne relativt til opportunistene, noe som igjen leder til en ytterligere økning i andelen samarbeidere via den evolusjonære prosessen beskrevet i avsnitt 3. Av (29) har vi videre at en økning i straffen ved ikke å samarbeide (opp til  $a$ ) flytter  $p^*$  mot venstre. Det skal dermed en mindre andel samarbeidere til for å starte en utvikling i retning av den stabile likevekten  $p = 1$  (alle samarbeider). Igjen har vi en situasjon hvor den evolusjonære prosessen er historiebetinget. Hvor samfunnet ender opp avhenger av hvor det starter.

Dersom  $\gamma > a$ , det vil si dersom straffen ved å velge 'ikke samarbeid'

er stor, så vil  $U(S) > U(O)$  for alle  $p \in (0, 1)$ . Dette kommer tydeligere fram ved å betrakte figur 5. Når  $\gamma > a$  vil kurven for  $U(O)$  ligge under kurven for  $U(S)$  i hele intervallet for  $p$ . Det betyr at dersom vi starter ut med et samfunn som både inneholder samarbeidere og opportunister, så vil samarbeiderne gjøre det bedre enn opportunistene, og samfunnet vil i det lange løp bestå kun av samarbeidere.

Dersom  $\gamma < a - b$ , det vil si dersom straffen ved å velge ‘ikke samarbeid’ er liten, så vil  $U(S) < U(O)$  for alle  $p \in (0, 1)$ . Det betyr at dersom vi starter ut med et samfunn som både inneholder samarbeidere og opportunister, så vil samfunnet i det lange løp bestå kun av ikke samarbeidende opportunister.

## 5.2 Normbrytere blir av og til straffet ( $0 \leq \sigma \leq 1$ )

I dette tilfellet vil den evolusjonære dynamikken avhenge av både størrelsen på straffen ved å bryte normen om samarbeid ( $\gamma$ ) og sannsynligheten for å bli straffet ( $\sigma$ ). For å kunne si noe mer presist om den evolusjonære prosessen, og hvilket adferdsmønster som til slutt vil prege samfunnet, så er det nyttig å skille mellom følgende tre tilfeller.

**Tilfelle 1:** Når  $\gamma > a$  blir opportunistene straffet hardt dersom de bryter normen om samarbeid. Hvorvidt det lønner seg å bryte normen framfor å samarbeide avhenger av sannsynligheten for å bli straffet ( $\sigma$ ). Dersom denne sannsynligheten er stor, det vil si dersom  $\sigma \geq a/\gamma$ , så er  $U(S) > U(O)$  for alle  $p \in (0, 1)$ . Dette fører til at  $p = 1$  (alle samarbeider) blir den eneste likevekten, og denne er stabil. Dersom sannsynligheten for å bli straffet ved å bryte normen om samarbeid er lav, det vil si dersom  $\sigma \leq (a - b)/\gamma$ , så er  $U(S) < U(O)$  for alle  $p \in (0, 1)$ . Dette fører til at  $p = 0$  (ingen samarbeider) blir den eneste likevekten, der også denne er stabil. Dersom sannsynligheten for å bli straffet verken er høy eller lav, det vil si dersom  $(a - b)/\gamma < \sigma < a/b$ , får vi en ustabil indre likevekt  $p = p^*$ . Dersom den initiale andelen samarbeidere i samfunnet er større enn  $p = p^*$ , så er  $U(S) > U(O)$ .  $p = 1$  (alle samarbeider) blir den eneste likevekten, og denne er stabil. Dersom den initiale andelen samarbeidere i samfunnet er mindre enn  $p = p^*$ , så er  $U(S) < U(O)$ .  $p = 0$  (ingen samarbeider) blir den eneste likevekten, og denne er også stabil.

**Tilfelle 2:** Når  $a - b < \gamma < a$  blir opportunistene straffet mindre hardt dersom de bryter normen om samarbeid, sammenlignet med tilfelle 1 ovenfor. Dette gir mindre rom for samarbeid siden det forventede utbyttet ved å velge ‘ikke samarbeid’ blir høyere, selv med en stor sannsynlighet for å bli straffet. Dersom  $\sigma > (a - b)/\gamma$  får vi en ustabil indre likevekt  $p = p^*$ . Samarbeid overlever bare dersom den initiale andelen samarbeidere

er større enn  $p^*$ , der  $p^*$  er mindre jo større  $\sigma$  er. Når  $\sigma \leq (a - b)/\gamma$ , så er  $U(S) < U(O)$  for alle  $p \in (0, 1)$ .  $p = 0$  blir den eneste stabile evolusjonære likevekten.

**Tilfelle 3:** Når  $\gamma < a - b$  er straffen for å bryte normen om samarbeid lav. Dette fører til at  $U(S) < U(O)$  for alle  $p \in (0, 1)$ , uansett størrelsen på sannsynligheten for å bli straffet ( $\sigma$ ). Samarbeiderne, som i utgangspunktet valgte å samarbeide, vil i det lange løp endre strategi til ‘ikke samarbeid’ – noe som fører til at  $p = 0$  (ingen samarbeider) blir eneste stabile evolusjonære likevekt.

Konklusjonen er altså at, som i likhet med tilfellet med en internalisert norm om samarbeid, så kan ekstern håndhevelse av en norm om samarbeid lede til en situasjon hvor noen medlemmer av samfunnet samarbeider, alle samarbeider eller ingen samarbeider. I hvilken situasjon samfunnet ender opp avhenger av samfunnsmedlemmenes tilbøyelighet til å straffe avvik fra normen om samarbeid, kostnaden ved å bli straffet samt den initiale sammensetningen av samarbeidere og opportunisten i samfunnet.

## 6 Avslutning

Dersom en tilstrekkelig stor andel av medlemmene i et samfunn er motivert av en norm om samarbeid, så kan samarbeid etableres som en stabil likevekt i en sosial dilemmasituasjon. I denne artikkelen er dette resultatet utledet ut i fra ulike antakelser om den underliggende sosiale interaksjonen. Et fellestrekk ved de ulike situasjonene er at den evolusjonære prosessen kan ende opp i en indre likevekt der en andel av medlemmene i samfunnet samarbeider, og hvor disse samarbeiderne lever side om side med en andel ikke-samarbeidende opportunisten. Dette er et trekk vi finner igjen i mange samfunn og organisasjoner. Flertallet overholder etablerte lover, regler og normer, mens en liten gruppe handler på tvers av dem. Et annet gjennomgående resultat er at den indre likevekten er ustabil. Dette betyr at et samfunn lett kan “tippe” ut av den indre likevekten og ende opp i en situasjon der enten alle samarbeider eller der ingen samarbeider. Begge situasjoner er mulig, det vil si at den evolusjonære prosessen er historiebetinget. Hvor samfunnet ender opp avhenger av hvor det starter.

Ovenfor har jeg studert ulike forhold som kan forklare hvordan en norm om samarbeid biter seg fast og sprer seg i et samfunn. Disse forholdene er studert hver for seg. Mer realistisk er det imidlertid at flere eller av disse forholdene er i sving samtidig. Det vil da være mange krefter som trekker i retning av at en norm om samarbeid vil bite seg fast og spre seg i et samfunn



når den først har fått fotfeste hos en liten andel av befolkningen.

## References

- [1] Axelrod, R. 1984. *The Possibility of Cooperation*. New York: Basic Books.
- [2] Basu, K. 1995. Civil Institutions and Evolution: Concepts, Critiques and Models, *Journal of Development Economics*, 43, 19-33.
- [3] Bendor, J. and P. Swistak. 2001. The Evolution of Norms, *American Journal of Sociology*, 106(6), 1493-1545.
- [4] Bowles, S. 2001. Individual Interactions, Group Conflicts, and the Evolution of Preferences. I S. Durlauf og P. Young (red.) *Social Dynamics*. Cambridge: The MIT Press.
- [5] Bowles, S. 2004. *Microeconomics: Behavior, Institutions, and Evolution*. New York: Princeton University Press.
- [6] Bowles, S. og H. Gintis. 2001. *The Economics of Shame and Punishment*. Working paper, Santa Fe Institute.
- [7] Bowles, S. og H. Gintis. 1998. The Moral Economy of Community: Structured Populations and the Evolution of Prosocial Norms, *Evolution & Human Behavior*, 19(1), 3-25.
- [8] Boyd, R. og P. Richerson. 1985. *Culture and the Evolutionary Process*. Chicago: Chicago University Press.
- [9] Borgstede, von C. 2002. *The Impact of Norms in Social Dilemmas*. Doktorgradsavhandling. Universitetet i Gøteborg.
- [10] Coleman, J. 1990. *Foundations of Social Theory*. Cambridge, Mass.: Harvard University Press.
- [11] Elster, J. 1983. *Explaining Technical Change*. Cambridge: Cambridge University Press.
- [12] Elster, J. 1989. Social Norms and Economic Theory, *Journal of Economic Perspective*, 3(4), 99-117.

- [13] Fehr, E., U. Fischbacher. 2004. Social Norms and Human Cooperation. *Trends in Cognitive Sciences*, 8(4), 185-188.
- [14] Fehr, E., U. Fischbacher og S. Gächter. 2002. Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms, *Human Nature*, 13, 1-25.
- [15] Fehr, E. og S. Gächter. 2002. Altruistic Punishment in Humans. *Nature*, 415(10), 137-140.
- [16] Gintis, H. 2003. Hitchhikers Guide to Altruism: Genes, Culture and the Internalization of Norms, *Journal of Theoretical Biology*, 206, 169-179.
- [17] Gintis, H, S. Bowles, R. Boyd og E. Fehr. 2003. Explaining Altruistic Behavior in Humans, *Evolution and Human Behavior*, 24, 153-173.
- [18] Güth, W. og H. Kliemt. 1998. The Indirect Evolutionary Approach: Bridging the Gap Between Rationality and Adaption, *Rationality and Society*, 10(3), 377-399.
- [19] Hechter, M. and K.D. Opp. 2001. What Have We Learned About the Emergence of Norms? I M. Hechter og K.D. Opp (red.) *Social Norms*. New York: Russel Sage Foundation.
- [20] Horne, C. 2001. Sociological Perspectives on the Emergence of Norms. I M. Hechter og K.D. Opp (red.) 2001. *Social Norms*. New York: Russel Sage Foundation.
- [21] Marx, K. 1963[1852]. *The Eighteenth Brumaire of Louis Bonaparte*. New York: International Publishers.
- [22] North, D.C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge MA: Cambridge University Press.
- [23] Opp, K.D. 2001. Social Networks and the Emergence of Protest Norms. I M. Hechter og K.D. Opp (red.) 2001. *Social Norms*. New York: Russel Sage Foundation.
- [24] Platteau, J-P. 2000. *Institutions, Social Norms, and Economic Development*. Amsterdam: Harwood Academic Publishers.
- [25] Putnam, R.D. 1993. *Making Democracy Work: Civic Tradition in Modern Italy*. Princeton NJ: Princeton University Press.

- [26] Richerson, P., R. Boyd og J. Henrich. 2003. Cultural Evolution og Human Cooperation. I P. Hammerstein (red.) *Genetic and Cultural Evolution of Cooperation*. Cambridge: The MIT Press.
- [27] Rothstein, B. 2000. Trust, Social Dilemmas and Collective Memories, *Journal of Theoretical Politics*, 12(4), 477-501.
- [28] Sacco, P.L. 1997. On the Dynamics of Social Norms. I C. Bicchieri, R. Jeffrey and B. Skyrms (red.) *The Dynamics of Norms*. Cambridge: Cambridge University Press.
- [29] Sethi, R. 1996. Evolutionary Stability and Social Norms, *Journal of Economic Behavior and Organization*, 29, 113-140.
- [30] Young, P. 1998. *Individual Strategy and Social Structure*. Princeton: Princeton University Press.