Ising-modeller af sociale systemer?

Eivind Tøstesen
CONNECT
Niels Bohr Institutet
tostesen@connect.nbi.dk

Kan fysikken bidrage til sociologi, samfundsvidenskab, økonomi og statskundskab? Eller mere specifikt; kan den statistiske mekanik bruges til at beskrive et system af mennesker eller andre sociale væsener?

Umiddelbart lyder det som den gamle reduktionistiske drøm om at statistisk mekanik en dag kan udstrækkes fra at beskrive en gas af partikler til at beskrive et "samfund". De fleste forkaster straks ideen fordi "mennesker er for komplekse til at sætte på formel". Men denne form for vanetænkning overser at bl.a. økonomerne længe har "sat mennesker på formel".

Jeg vil beskrive en dynamisk model af *sociale dilemmaer* som handler om hvornår man samarbejder. Erfaringerne med Ising-modeller indenfor den statistiske mekanik kan bidrage til forståelsen af denne models dynamik.

1 Sociale dilemmaer

Et *socialt dilemma* er det følgende problem: I en gruppe af individer kan hvert individ vælge frivilligt hvorvidt han/hun vil hjælpe de andre i gruppen. At *hjælpe de andre*, eller at være social og samarbejde, vil sige at yde et bidrag til et fælles bedste. Et eksempel er frivilligt arbejde og støtte til græsrodsbevægelser, politiske organisationer og lignende, der kommer *alle* til gode, både de solidariske der har støttet og de andre, egoisterne. Et andet eksempel er forbedring af miljøet. Hver især kan man begrænse sin forurening og spare på ressourcerne. Det koster noget for en selv, men kommer alle til gode.

Der er en lang tradition for at undersøge den slags problemer matematisk indenfor bl.a. biologi og økonomi, ved hjælp af *spilteori*. Her antager man at individerne (spillerne) hver især søger at maksimere en tal-størrelse som angiver deres egen "gevinst" eller *payoff*. Dette gør de ved at vælge deres "træk" på en mere eller mindre optimal eller rationel måde. I spilteori udstyrer man derfor et individ i med en *nyttefunktion* (utility function) u_i der giver i's payoff som funktion af alle spillernes valg. Nyttefunktionen skal vælges så den afspejler hvad der er godt for individet i de situationer man modellerer.

Et socialt dilemma beskriver man med følgende spil. Der er N spillere (individer) i gruppen. Hvert individ kan vælge mellem to tilstande C og D, som står for hhv. *cooperate* (hjælpe til) og *defect* ("slappe af"). Lad $n_c(t)$ være antallet der samarbejder (tilstand C) til tiden t, og lad $k_i(t)$ indikere om i hjælper til ($k_i(t) = 1$) eller ej ($k_i(t) = 0$) til tiden t. Nyttefunktionen for hvert individ i er da

$$u_i(t) = \frac{b}{N} n_c(t) - ck_i(t) \tag{1}$$

hvor b er en positiv konstant der angiver udbyttet til den enkelte når alle samarbejder, og c er en positiv konstant der angiver omkostningen for i ved at hjælpe til. Hvis konstanterne vælges så

$$b > c > \frac{b}{N}, \tag{2}$$

betyder det, at det er mere ønskværdigt for et individ at alle samarbejder end at ingen samarbejder, samt at omkostningen ved at bidrage til samarbejdet er større end hvad individet selv får ud af sit bidrag. Så uanset hvor mange der samarbejder er D umiddelbart det valg der giver højest u_i, men det er ikke godt hvis alle vælger D. Heri ligger dilemmaet for gruppen.

Dette spil kaldes et *N-personers prisoners dilemma*. Det er kendt af mange i 2-personers versionen som bare kaldes *prisoners dilemma*[1]. Der er f.eks. blevet lavet en del computer simulationer af 2-personers spillet, hvor en population af spillere mødes parvis gennem en "turnering" på en eller anden måde. Men det er ikke det der interesserer os i denne artikel.

I spilteori arbejder man med de enkelte spilleres *strategier*, der angiver hvilke træk de skal foretage som funktion af de andres træk (spillets historie). I vores N-personers spil er antallet n_c, der samarbejder den eneste information en spiller har at beslutte sig udfra. De kan ikke se hvem der samarbejder, kun hvor mange.

Hvad er så løsningen på dilemmaet? En løsning er jo at indføre en "bussemand" der kommer og straffer dem der ikke hjælper til. Men det vil jo gå ud over spillets regler. Spilteorien har mange forskellige løsninger. Generelt kan man sige at de går ud på at *lade individerne spille spillet længe nok*. Man kan f.eks. lade dem spille i diskret tid med synkrone træk eller i kontinuert tid med asynkrone træk. Hvert individ skal ikke kun kortsigtet maksimere sin nuværende payoff, men også tænke på fremtidige payoffs. Dermed kan gensidigt samarbejde opretholdes: Hvis man skulker fra samarbejdet vil man måske for en stund nyde godt af de andres indsats, men snart vil de andre også holde op med at samarbejde og dermed straffe den oprindelige synder. Dermed bliver C det bedste valg for hver enkelt i det lange løb. Dette er et centralt aspekt når vi kommer til modellen af sociale dilemmaers dynamik. Men først henter vi lidt inspiration fra Ising-modeller.

---

[1] Navnet hentyder til to fanger i hver deres celle: De mistænkes for at have begået en forbrydelse sammen. De tilbydes hver at tilstå. Hvis begge tilstår straffes de hårdt. Hvis ingen tilstår (dvs. de samarbejder) får begge en mild dom. Hvis kun den ene tilstår løslades han med en dusør mens den anden straffes meget hårdt.

2 Ising-modeller

Ising modellen blev opfundet af W. Lenz i 1920 som en simpel model af en *ferromagnet*. Et Ising-system består af mange *spin* som kan opfattes som små magneter, se fig. 1. Spinnene har to mulige retninger, op og ned, og de an-

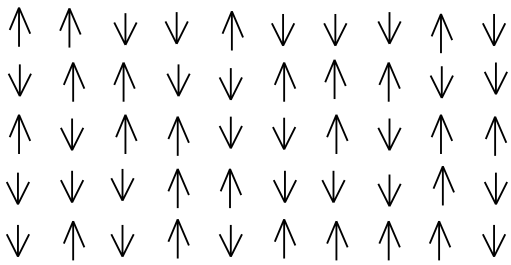

Figure 1: Ising-model.

gives med variablerne s_i som er $+1$ for "op" og -1 for "ned", og i nummererer spinnet. To spin kan enten have en ferromagnetisk vekselvirkning som søger at ensrette dem, eller en antiferromagnetisk vekselvirkning der søger at modsatrette dem. Desuden kan der være et ydre felt h som spinnene forsøger at have samme fortegn som. I et to-dimensionelt gitter som i fig. 1 kan man f.eks. lade et spin vekselvirke med sine fire nærmeste naboer. Systemet kobles til et varmereservoir med temperatur T, hvorefter spinnene flipper op og ned "tilfældigt".

Ising modeller bruges i statistisk mekanik til at modellere *faseovergange*. En ferromagnet har ferromagnetiske vekselvirkninger og udviser en faseovergang. Systemet har en kritisk temperatur T_c. Når $T > T_c$ er varmen så stærk, at spinnene i gennemsnit peger lige meget op som ned når $h = 0$. Men når $T < T_c$ er varmen svag nok til at de ferromagnetiske vekselvirkninger får overtaget, og spinnene får en tendens til at være ensrettede. Systemets magnetisering $m = \frac{1}{N}\sum_i s_i$ antager da i gennemsnit en enten positiv eller negativ permanent værdi, selv uden ydre felt.

Ising systemets Hamilton er

$$H = -\frac{1}{2}\sum_{i \neq j} J_{ij} s_i s_j - h \sum_i s_i \tag{3}$$

hvor J_{ij} er vekselvirkningen mellem spin i og j. Den er positiv for ferromagnetiske og negativ for antiferromagnetiske. Hamiltonfunktionen giver systemets energi som funktion af konfigurationen af spin. Konfigurationer med lave energier er mest sandsynlige.

Ising modeller er ret simple modeller af magnetiske systemer. Man kan næsten kalde dem ufysiske legetøjsmodeller. Men meningen med en forsimplet abstrakt model af et system med faseovergang er at den skal fange den *essentielle* fysik bagved, samtidig med at man kan regne på den. Ising modeller er så abstrakte at de kan finde anvendelse i mange forskellige sammenhænge. F.eks. bruges Ising systemer der har både positive og negative J_{ij} som modeller af spin glasser. Teorien om spin glasser finder stor anvendelse i f.eks. neurale netværk og proteiners dynamik.

Kunne det tænkes at en ferromagnetisk Ising model kan bruges som model af et socialt system? Handler mennesker i et socialt dilemma som om de var spin med ferromagnetiske vekselvirkninger, (hvis man lader op-spin (+1) være C og ned-spin (-1) være D)? I så fald har man en model hvor hver spiller ser på sine "naboer" og vælger det samme som dem, så hvis flertallet er C så bliver spilleren selv C, og D tilsvarende. Det kunne godt lyde som en rationel strategi at man kun samarbejder hvis de andre også gør det. Man får altså et billede af sociale dilemmaers dynamik hvor hvert individ står og fluktuerer lidt tilfældigt frem og tilbage mellem C og D, men med en tendens til at følge de andre.

Vi skal se at vi får en social model der svarer til følgende *mean field* Ising model af en ferromagnet med Glauber dynamik: Der er N spin og de vekselvirker alle med hinanden (ikke kun nærmeste naboer), $J_{ij} = 1$ for alle i og j:

$$H = -\frac{1}{2}\sum_{i \neq j} s_i s_j - h \sum_i s_i \tag{4}$$

Glauber dynamik bruges i Monte Carlo simulationer til at simulere dynamikken af systemet ved temperatur T, og går ud på følgende: Spinnene opdateres hver for sig spontant (som radioaktive henfald) med rate α. Når et spin i opdateres sker det i et effektivt lokalt felt h_i der er den samlede vekselvirkning fra de andre spin og det ydre felt

$$h_i = h + \sum_{j \neq i} s_j = h + 2n - N \tag{5}$$

hvor n er antal op spin. Sandsynligheden ρ_{+1} for at spinnet bliver "op" er så defineret ved

$$\begin{aligned} \rho_{+1} &= \frac{1}{2} + \frac{1}{2}\tanh(\beta h_i) \\ &= \frac{1}{2} + \frac{1}{2}\tanh\{2\beta N f + \beta(h - N)\} \end{aligned} \tag{6}$$

hvor $f = n/N$ og $\beta = 1/T$. Denne sandsynlighed for "op" er større jo større h_i er. Den går fra 0 til 1 når h_i går fra $-\infty$ til $+\infty$. Ved $h_i = 0$ er $\rho_{+1} = 1/2$. Et spin har altså størst sandsynlighed for at ensrette sig h_i, og denne sandsynlighed bliver større jo lavere temperaturen er. Ved $T = 0$ fås en tærskel-opførsel

$$\rho_{+1} = \begin{cases} 1 & for \quad h_i > 0 \\ 0 & for \quad h_i < 0. \end{cases} \tag{7}$$

3 Sociale dilemmaers dynamik

Glance og Huberman har opfundet en model af sociale dilemmaers dynamik [1]. Det er en stokastisk model. Som beskrevet er der N individer med hver to mulige tilstande C og D. Vi har asynkron dynamik i kontinuert tid: Hvert individ opdaterer sin tilstand spontant med rate α.

Lad os se på et individ i der opdaterer på tidspunktet t_0. Han/hun tager sin egen indsats op til overvejelse ud fra det opfattede samarbejdsniveau $n_c(t)$: "Skal jeg vælge C eller D?" Denne overvejelse skal også tage hensyn til fremtidige payoffs, og det gøres ved at maksimere den akkumulerede payoff

$$U_i = \int_{t_0}^{\infty} u_i(t) e^{-\frac{(t-t_0)}{H}} dt \tag{8}$$

I dette integral lægges der mindre vægt på payoffs, jo længere ud i fremtiden de er, v.hj.a. *discountfaktoren* $e^{-\frac{(t-t_0)}{H}}$. H er *tidshorisonten* og den er hvor langt ind i fremtiden individet kigger. Den kan fortolkes som spillets resterende middellevetid.

For at kunne se på fremtiden må individet have nogle forventninger om hvor mange der samarbejder i fremtiden. Modellens centrale antagelser er at individet har følgende forventninger: (1) Den handling individet vælger nu vil tilskynde andre til at gøre det samme i fremtiden, og (2) denne indflydelse på de andre er mindre jo større gruppestørrelsen N er, og (3) denne indflydelse stiger jo flere $n_c(t_0)$ der samarbejder. I fig. 2. er vist hvordan et individ forudsiger tidsudviklin-

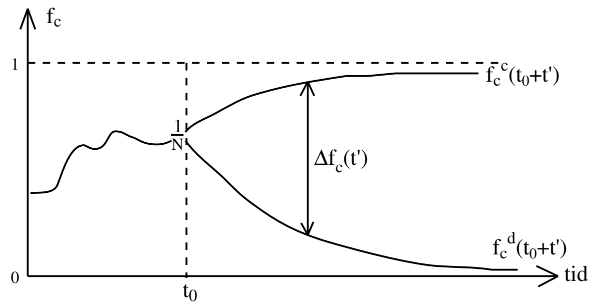

Figure 2: Hvordan de extrapolerer brøkdelen af samarbejdende ind i fremtiden hvis hhv. C og D vælges.

gen af brøkdelen $f_c = n_c/N$ der samarbejder i de to tilfælde, hvor det vælger C hhv D til tiden t_0. Med de to forudsigelser kan individet let udregne de forventede konsekvenser af sin handling, ved at udregne den akkumulerede payoff i de to tilfælde. Derefter kan spilleren vælge den handling der giver det største integral.

Det antages at forskellen mellem de to forudsigelser er

$$\Delta f_c(t') \quad = \quad f_c^c(t_0 + t') - f_c^d(t_0 + t')$$

$$= \quad 1 - (1 - \frac{1}{N}) \exp\left(-\frac{\alpha f_c(t_0)}{N} t'\right) \tag{9}$$

som vist i fig. 3. Hældningen i starten angiver hvor hurtigt de to kurver fjerner

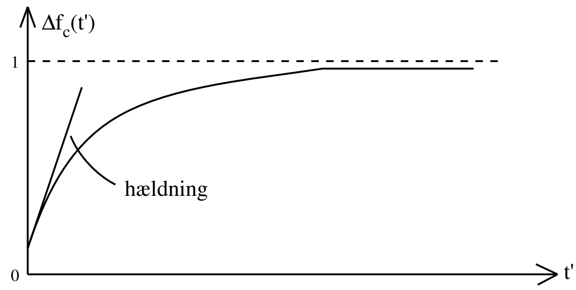

Figure 3:

sig fra hinanden og er proportional med $\frac{f_c(t_0)}{N}$. Herved indgår antagelserne (2) og (3).

Med de opstillede antagelser kan vi nu spørge hvornår individet vælger C, med andre ord hvornår det "øverste" integral er større end det "nederste". Betingelsen herfor bliver

$$0 < \int_0^{\infty} [b\Delta f_c(t') - c] e^{-t'/H} dt' \tag{10}$$

som reducerer til

$$f_c(t_0) > \frac{1}{H\alpha} \left(\frac{Nc - b}{b - c}\right) \equiv f_{crit} \tag{11}$$

Det viser sig altså at C er et optimalt valg hvis der er mange nok der samarbejder til tiden t_0, og D er optimal hvis der ikke er mange nok, og skillelinien går ved en kritisk brøkdel f_{crit} af samarbejdende individer.

Denne opførsel er jo stort set hvordan spinnene i vores mean field Ising model opfører sig, ihvertfald ved temperaturen $T = 0$! Man kan på forskellige måder indføre støj i den sociale model, svarende til at individerne begår fejl, har begrænset rationalitet, fejlbedømmer samarbejdsniveauet $f_c(t_0)$ og lign. En måde er f.eks. at sige at individet i opfatter individ j's tilstand korrekt med sandsynlighed p, og som den forkerte tilstand med sandsynlighed $1-p$. Resultatet bliver at individet, når det skal opdatere sig, vælger C med sandsynligheden

$$\rho_C = \frac{1}{2} + \frac{1}{2} \mathrm{erf} \left\{ \left(\frac{2p-1}{\sigma\sqrt{2}}\right) f_c(t_0) + \left(\frac{1 - p - f_{crit}}{\sigma\sqrt{2}}\right) \right\} \tag{12}$$

Sammenlign med ligning 6. Hvis vi bruger approksimationen erf=tanh og sætter de to udtryk ρ_C og ρ_{+1} lig hinanden, så fås at spinnene og individerne i de to modeller *har samme dynamik* når

$$T = \frac{\sqrt{8Np(1-p)}}{2p-1} \tag{13}$$

$$h = \frac{N(1-2f_{crit})}{2p-1} \tag{14}$$

4 Fysikkens bidrag

Vi har altså at dynamikken af den sociale model er ækvivalent (på nær ubetydelige forskelle) med den tilsvarende Ising models Glauber dynamik givet med ligningerne 13 og 14. Dermed kan vi nu overføre alle de teoretiske resultater man er kommet frem til i statistisk mekanik til den sociale model. Vi kan anvende vores begreber om faseovergange.

Det sociale system udviser altså en faseovergang. Det har et kritisk støjniveau svarende til den kritiske temperatur T_c, som kan findes af sammenhængen mellem T og p i ligning 13. For $p < p_c$ (støjen er under det kritiske niveau) er der to mulige ligevægte for systemet; en hvor de fleste samarbejder og en anden hvor de fleste "slapper af". Antallet n_c der samarbejder fluktuerer i begge tilfælde omkring en bestemt værdi. Disse to værdier svarer til de to minima A og B af den *fri energi* i fig. 4. B er den ligevægt hvor de fleste samarbejder. Den

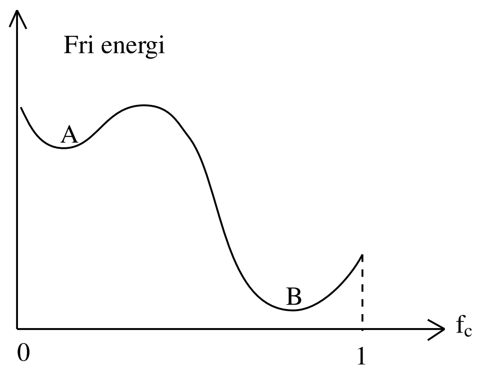

Figure 4:

fri energi viser stabilitetsforholdene for de to ligevægte. I figuren er minimaet B lavest og denne ligevægt er derfor stabil. A er kun en *metastabil* ligevægt. Systemet kan godt befinde sig i denne ligevægt i nogen tid, men pludselig kan fluktuationer skubbe systemet over fri energi barrieren og det havner nede i den anden ligevægt B. I ligevægt A er de fleste individer i tilstand D og de holder hinanden fast, ingen har lyst til at begynde at hjælpe de andre, men en fluktuation kan skabe et tilstrækkeligt stort antal samarbejdende som de andre begynder at slutte sig til. Dette svarer til når underafkølet vand pludselig kan blive til is.

Hvilket minimum der er dybest afhænger af det ydre felt h. Hvis $h = 0$ er de to minima lige dybe. Hvis $h > 0$ er ligevægten med positiv magnetisering stabil, og den svarer til ligevægten med samarbejde. Af ligning 14 får vi betingelsen $f_{crit} < 1/2$ for at samarbejde er en stabil ligevægt i gruppen. Af ligning 11 fås at denne betingelse er

$$\frac{1}{H\alpha}\left(\frac{Nc-b}{b-c}\right) < \frac{1}{2} \tag{15}$$

Det ses at samarbejde kun kan opretholdes hvis deres tidshorisont H er stor nok og hvis gruppen er lille nok.

Disse resultater svarer jo til vores intuitive forestillinger om samarbejde. Men de er til dels indeholdt i antagelserne. Det vigtigste bidrag denne "termodynamiske" tilgang til sociale dilemmaer kan give, er billedet af fluktuationer der kan skabe en overgang fra den metastabile ligevægt til den stabile. I forhold til nogle af de simple modeller af dynamik som spilteorien ellers har arbejdet med, nærmer denne model sig noget realistisk.

Vi har set et eksempel på hvad der er typisk for den statistiske mekanik: Den beskriver nogle overordnede træk ved systemerne, f.eks. faseovergange. Det er ikke de simple antagelser om systemets enkelte dele (f.eks. at mennesker opfører sig som Ising spin) man skal hæfte sig ved, dem lærer vi ikke noget af, men derimod systemets overordnede træk, f.eks. dets opførsel i middel, som ikke afhænger så stærkt af detaljerne i delenes vekselvirkninger. De overordnede træk er *universelle*.

5 Skal fysikere lave ikke-fysik?

Jeg indledte denne artikel med at spørge om fysikken kan bidrage til sociologi osv. Det har været min hensigt at give et eksempel hvor dette er muligt. Der er *ikke* tale om at vi vil forklare fænomener indenfor andre fag udfra underliggende (mikroskopiske) fysiske love. Mit eksempel skal vise hvordan fysikere kan bruge deres metoder, værktøjer og begreber til at løse en matematisk model opstillet i et andet fag.

Det lyder meget uskyldigt, men her vil læseren måske protestere: "Den sociale model er ikke opstillet i et andet fag, den er kraftigt inspireret af fysikkens Ising-model!" Ja, det er rigtigt at jeg har opbygget denne artikel som om målet var at lave en Ising model af sociale dilemmaer. Men det er ikke indbygget i Glance og Hubermans model at den er så Ising-centreret som jeg giver indtryk af. De har også andre versioner af modellen. F.eks. får de en model ved at ændre på antagelse (3), hvor individerne ikke opfører sig som spin, men vælger D både hvis f_c bliver lav nok og hvis f_c bliver høj nok og vælger C derimellem [4]. Glance og Huberman har selv en fysikbaggrund, men de er ikke så fysikorienterede i deres artikler som jeg er. Jeg er kun Ising-centreret fordi jeg skriver til fysikere. Det er kun for fysikere at det er underholdende og interessant at se *hvor* eksakt man kan etablere en ækvivalens mellem den sociale model og vores Ising model. Men set fra et tværfagligt synspunkt er Ising-modellen ikke noget særligt, og det er dårlig videnskab at ville presse sine Ising antagelser ned over modellen, hvis de er urealistiske.

Dette er en svær balancegang: Skal videnskaben lave nye modeller med skræddersyede antagelser og begreber eller genbruge det gamle med de erfaringer og løsninger vi allerede har fundet? Glance og Huberman forsøger at lave modeller hvor de kan anvende deres fysik-metoder, men de tager hovedsagligt udgangspunkt i spilteoriens modeller og begreber og er mest interesserede i at lære noget om sociale systemer. Modellerne bliver måske sværere at anvende statistisk mekanik direkte på, men det er stadig muligt at anvende nogle værktøjer til at undersøge dynamikken. Og det er dette budskab jeg vil bringe.

Dermed er jeg en del af en hel bevægelse indenfor fysikken: Mange fysikere har en tro på at de kan bevæge sig ind på andre fags domæner (biologi, økonomi m.v.) og beskrive *komplekse systemer*. Ideen er at vi har nogle universelle begreber og nogle metoder, f.eks. til beskrivelse af dynamiske systemer (kaos) herunder computersimulationer, som gælder bredt og ikke begrænser sig til et enkelt fagområde, som vi kan anvende. Måske skyldes denne bevægelse for at finde nye græsgange, at jorden så at sige er udpint indenfor fysikken, at der er "arbejdsløshed".

Men det kan føre til problemer. Fysikken har haft så stor succes med matematiske metoder, og får æren for hele vores omfattende teknologi, at den står som forbilledet for god videnskab. Det har ført til en "ulandshjælp"-situation, hvor man forestiller sig at fysikere tager "ned" til de andre fag (ulandene) for at lære dem sine overlegne matematiske metoder. Nogle fysikere møder måske op med en arrogant indstilling overfor "de dumme biologer" som ikke kan regne. Det er rigtigt at vores fysiske formalismer ofte ikke forstås. Men så må de lægges på hylden. Man må dog ikke forfalde til at tro at fysikere er de eneste der behersker matematiske metoder. Fysikeren, på den anden side, mødes måske med benovelse og autoritetstro som repræsentant for videnskabernes konge (fysik). Men fysikeren mødes også tit med skepsis, p.g.a. dårlige erfaringer med arrogante fysikere. Fysikeren beskyldes f.eks. for hellere at ville lege med sine modeller end interessere sig for den virkelighed han modellerer. Problemet er, at fysikeren selvfølgelig må overlade det til de virkelige eksperter på feltet at vurdere om modellen er god, og derfor også let fralægger sig ansvaret overhovedet for om det han laver er god videnskab. Og måske gider han ikke gå hen på det andet institut. Men man kommer ikke langt med tværfagligt arbejde uden feedback fra og tæt samarbejde med hinanden.

På trods af sådanne samarbejdsproblemer tror jeg på at matematiske metoder har et bidrag at give i tværfaglig videnskab.

References

[1] Natalie S. Glance and Bernardo A. Huberman. Dynamics of social dilemmas. *Scientific American*, March 1994.

[2] Natalie S. Glance and Bernardo A. Huberman. The Outbreak of Cooperation. *Journal of Mathematical Sociology*, 17(4):281-302, 1993.

[3] Bernardo A. Huberman and Natalie S. Glance. Diversity and collective action. In H. Haken and A. Mikhailov, editors, *Interdisciplinary Approaches to Complex Nonlinear Phenomena*. Springer, New York, 1993.

[4] Bernardo A. Huberman and Natalie S. Glance. Beliefs and Cooperation. In *Modelling Rational and Moral Agents*, P. Danielson ed., Oxford University Press, 1995.

[5] Eivind Tøstesen. *Dynamics of hierarchically clustered cooperating agents.* Speciale i fysik, Niels Bohr Insitutet, 1995.