

Meinungsbildung, Kommunikation und Kooperation aus physikalischer Perspektive

Modelle und Analogien der Physik helfen, gesellschaftliche Phänomene und soziale Interaktion zu beschreiben.

Frank Schweitzer

Die individuellen und kollektiven Formen des menschlichen Verhaltens sind komplex genug, um neben den Geistes- und Sozialwissenschaften auch die Naturwissenschaften zu Erklärungs- und Modellierungsversuchen herauszufordern. So wird in der Soziobiologie die These vertreten, Sozialverhalten auf biologische und erbliche Ursachen zurückzuführen. Die kontroversen Diskussionen über diese Art des Reduktionismus sind noch nicht beendet, da etabliert sich bereits ein weiterer naturwissenschaftlicher Ansatz, die Soziophysik. Hier geht es um ein Verständnis der kollektiven Dynamik sozialer Akteure unter Zuhilfenahme von Methoden der statistischen Physik.

Das Bestreben, naturwissenschaftliche, quantitative Methoden in den Sozialwissenschaften zu verankern, zieht sich durch die gesamte neuzeitliche Wissenschaftsgeschichte. Hervorzuheben ist hier vor allem der französische Mathematiker Adolphe Quetelet (1796–1874). Ihm verdanken wir nicht nur den Quetelet-Index (heute „body mass index“, BMI, genannt), um unser Übergewicht zu quantifizieren, sondern auch grundlegende Konzepte zur Sozialstatistik. In seinem 1835 erschienenen Buch „Sur l'homme et le développement de ses facultés, essai d'une physique sociale“ taucht erstmals die Idee einer Soziophysik auf, um den *homme moyen*, den Durchschnittsmenschen Otto N., wissenschaftlich zu charakterisieren.

Wahrscheinlichkeitsrechnung und Statistik bilden inzwischen die quantitativen Säulen der Soziologie. Eine konsistente Theorie für die dynamische Modellierung sozialer Phänomene ist dagegen noch in der Entwicklung begriffen – was angesichts der Komplexität dieser Systeme keineswegs überrascht. Anders als in der Physik, wo zumeist Teilchen mit identischen Eigenschaften auf die stets gleiche Weise interagieren, weisen die mikroskopischen „Elemente“ sozialer Systeme, also die (menschlichen) Individuen, ein erhebliches Maß an interner Komplexität auf. Diese befähigt sie in einer gegebenen Situation zu durchaus unterschiedlichen Reaktionen in Abhängigkeit von ihren Zielen, Erwartungen oder Wertvorstellungen. Auch entwickeln sich die Wechselwirkungen in sozialen Systemen im Verlaufe der Zeit, wobei das individuelle ebenso wie das kollektive Gedächtnis eine Rolle spielen.

Keine Chance also für eine Physik sozialer Systeme?

Sicher nicht, wenn es um ein genaues Verständnis des einzelnen sozialen Akteurs mit seiner vielschichtigen Weltanschauung geht. Vielmehr modelliert die heutige Soziophysik die kollektive Dynamik einer Vielzahl derartiger Akteure, wobei Analogien zu physikalischen Vielteilchensystemen eine tragende Rolle spielen. Zu den Anwendungsbeispielen zählen vor allem Meinungsbildungs- und Entscheidungsprozesse, räumliche Strukturbildung in sozialen Systemen und Migrationsdynamik [1–8]. In der Soziophysik werden aber auch Fragen der Strategieentwicklung untersucht, man denke an das sehr erfolgreiche *Minority Game* [9, 10] oder Modelle der evolutionären Spieltheorie, die mit Methoden der statistischen Physik gut behandelt werden können [11–15].

„Every theory, whether in the physical or biological or social sciences, distorts reality in that it oversimplifies. But if it is a good theory, what is omitted is outweighed by the beam of light and understanding thrown over the diverse facts“.

Paul A. Samuelson (Nobelpreisträger für Ökonomie, 1970)

Dabei ist jedoch ein interdisziplinärer Diskurs unverzichtbar. Dies reflektiert auch die internationale Wissenschaftsentwicklung, in der Sozialwissenschaften, Informatik und Naturwissenschaften neue Verbindungen eingehen – Econophysics, Socionics, Computational Economics sind nur einige Beispiele. Zwar haben Sozialwissenschaftler eine eigene Tradition bei der Computersimulation sozialer Systeme [16], aber bei der Analyse und der Vorhersage der kollektiven Dynamik übt auch die statistische Physik eine große methodische Vorbildwirkung aus [17]. Schließlich besteht der Sinn der „simplifizierten“ Soziophysik gerade darin, die grundlegende Dynamik des Systemverhaltens tatsächlich aus den Modellannahmen heraus zu verstehen – anstatt in der Komplexität von scheinbar so attraktiven „intelligenten“ *Agenten* unterzugehen. Mit diesem Terminus *technicus* werden die mikroskopischen Einheiten sozialer Systeme in der Modellierung allgemein bezeichnet. Das ist weniger als Hommage an James-Bond-Abenteuer gemeint – vielmehr hat sich der Begriff „Agent“ in der Ökonomie und Informatik durchgesetzt, um (autonome) Systemelemente mit einer „gewissen“ internen Komplexität zu bezeichnen,

Priv.-Doz. Dr. Dr. Frank Schweitzer, Fraunhofer Institut für Autonome Intelligente Systeme, Schloss Birlinghoven, 53754 Sankt Augustin und Institut für Physik der Humboldt-Universität, Unter den Linden 6, 10099 Berlin

welche natürlich von dem zu beschreibenden System abhängt. Multi-Agenten-Systeme (MAS) – also Systeme aus zahlreichen Agenten unterschiedlicher Art und Wechselwirkung – werden heute sehr erfolgreich eingesetzt, um die Entstehung von kollektiven Phänomenen bei nicht-physikalischen Problemstellungen zu modellieren.

Ising-Modelle für soziale Interaktion

Die Herausbildung einer kollektiven Meinung in sozialen Systemen zählt zu denjenigen Problemen, die seit langem auch von Physikern untersucht werden [1, 2]. Auf der Suche nach *einfachen* Modellen haben sie sich unter anderem von Ising-artigen Systemen inspirieren lassen. Auf einem regulären, zumeist quadratischen Gitter mit periodischen Randbedingungen sind gleichmäßig $i = 1, \dots, N$ Agenten verteilt, die jeweils durch einen inneren Freiheitsgrad $\theta_i \in M$, ihre „Meinung“, charakterisiert sind. Das Meinungsspektrum M wird als diskret angenommen, zumeist beschränkt man sich auch nur auf zwei (entgegengesetzte) Meinungen, $\theta_i \in \{-1, +1\}$, die einem nach oben oder nach unten gerichteten Spin entsprechen. In Abhängigkeit von der Art der lokalen oder globalen Spinkopplung lässt sich für diese Systeme eine Energiefunktion aufstellen. Die Minimierung von Konflikten mit Agenten anderer Meinung entspricht dem Übergang in einen möglichen Gleichgewichtszustand des Gesamtsystems. Von „frustrierten“ Systemen spricht man, wenn widerstrebende lokale Anforderungen existieren, die von den Agenten nicht gleichzeitig erfüllt werden können. In diesem Fall gibt es sehr viele solcher Gleichgewichtskonfigurationen, die unterschiedlichen Konfliktlösungen entsprechen. Auch Entmischungseffekte, also die räumliche Separation von Meinungen, oder

Phasenübergänge lassen sich simulieren und analytisch behandeln. Beispielsweise wird die in der Physik gebräuchliche Mean-Field-Näherung, in welcher der Einfluss aller anderen Agenten j auf Agent i summarisch durch ein gemitteltes Feld ersetzt wird, benutzt, um eine Art „Antizipationseffekt“ für die Gruppenmeinung zu berücksichtigen [5]. Der Erkenntnisgewinn für die *soziale* Dynamik begrenzt sich hier allerdings von selbst auf die bloße Konstruktion von Analogien – sodass Kritik von Sozialwissenschaftlern an einer solchen simplifizierenden Modellierung durchaus berechtigt ist.

Eine andere Art von lokaler Wechselwirkung wird in den so genannten Voter-Modellen untersucht. Hier ändert ein Agent seine Meinung in Abhängigkeit von der Meinung der Agenten in seiner unmittelbaren Umgebung. Beim „majority voting“ passt er sich der Meinungsmehrheit an, sodass sich häufige Meinungen eher durchsetzen. Beim „minority voting“ ist es gerade umgekehrt. In den nichtlinearen Voter-Modellen muss die Übergangswahrscheinlichkeit für die Meinungsänderung nicht proportional zur Zahl der entsprechenden Agenten sein. Dies ermöglicht auch ein Wahlverhalten gegen den Trend – zum Beispiel, indem Meinungen unterstützt werden, so lange sie weder einer Mehrheit

noch einer Minderheit zuzuordnen sind. Voter-Modelle werden unter anderem verwendet, um die Frage nach der räumlichen Verteilung von Mehrheiten und Minderheiten in sozialen Systemen zu beschreiben.

Einfache Computersimulationen mit stochastischen zellulären Automaten zeigen bereits die interessante Dynamik solcher Modelle (Abb. 1). So findet man für bestimmte Übergangswahrscheinlichkeiten die Herausbildung von *räumlichen Domänen* mit Agenten gleicher Meinung, die unter Umständen auf einer langen Zeitskala zum Verschwinden der Minderheit führt. Viel interessanter sind allerdings Parameterkonstellationen, die auch für lange Zeit zu einer *Koexistenz* beider Meinungen führen, wobei die Mehrheits- und Minderheitsverhältnisse sich ständig ändern. „Opposition“ und „Regierungspartei“ wechseln sich also ständig ab. Dabei zeigt die räumliche Verteilung, dass die „globale“ Minderheit lokal durchaus die Mehrheit bilden kann – was keineswegs im Widerspruch zur Erfahrung steht.

Voter-Modelle haben den Vorteil, dass sie auch analytisch behandelt werden können; so lassen sich Gleichungen für die raum-zeitlichen Korrelationen herleiten und Aussagen über die zeitliche Entwicklung von Meinungshäufigkeiten treffen.

Ein anderes Ising-artiges Modell der kollektiven Meinungsbildung ist das von Stauffer propagierte Sznajd-Modell [18], benannt nach den polnischen Autoren [8]. Anders als bei den Voter-Modellen, wo die Agenten ihre Meinung der Umgebung anpassen, bringen im Sznajd-Modell die Agenten ihre Umgebung dazu, die von ihnen vertretene Meinung anzunehmen; es wird also eine Art Missionierung beschrieben. Gemäß dem alten Gewerkschaftsmotto „United we stand, divided we fall“, können zwei benachbarte Agenten *gleicher* Meinung ihre jeweils nächsten Nachbarn ebenfalls

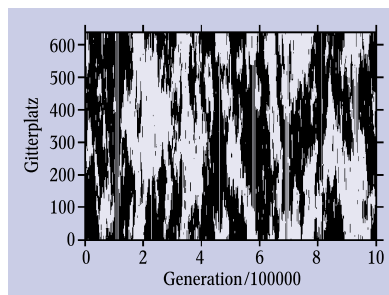


Abb. 1: Simulation des nichtlinearen Voter-Modells mit einem eindimensionalen stochastischen zellulären Automaten (Gittergröße $L = 640$). Die vertikale Achse gibt die räumliche Verteilung von Agenten mit der Meinung $\{-1\}$ (schwarz) und $\{+1\}$ (weiß) an. Die Zeitentwicklung ist über der horizontalen Achse aufgetragen. Die Perioden, in denen jeweils eine Meinung vorherrscht, sind deutlich zu erkennen, ebenso die Nichtstationarität und die Koexistenz auf langer Zeitskala.

von dieser Meinung überzeugen.

Auch hier stellt sich die Frage, ob ein globaler „Konsens“ erzielt wird, sodass eine der beiden Meinungen verschwindet, oder ob eine dauerhafte Koexistenz entstehen kann. Das ursprüngliche Modell zeigt zwar auf einer mittleren Zeitskala eine interessante Dynamik (zum Beispiel hinsichtlich der Korrelationen für eine Meinungsänderung), asymptotisch allerdings nur ein triviales Verhalten (50/50-Koexistenz oder 100prozentige Majorität). Jedoch gelang es mit modifizierten Sznajd-Modellen (Berücksichtigung von Bias, Interaktion auf einem Barabasi-Albert Netzwerk anstelle lokaler Wechselwirkung auf dem Gitter, erweitertes Meinungsspektrum u.a.), durchaus soziologisch interpretierbare Ergebnisse zu erzielen [19].

Eine wesentliche Erweiterung von Ising-artigen Modellen fand durch Einbeziehung der so genannten „social impact theory“ statt, die zunächst von Psychologen und Soziologen entwickelt wurde [20]. Die Wirkung auf Agent i durch eine Gruppe wird hier beschrieben durch eine Art „soziales Feld“, das von der Zahl der Agenten j , deren Meinung θ_j und ihrem Einfluss abhängt. Dabei unterscheidet man zwischen der Fähigkeit, andere zu überzeugen, p_j , und andere zu unter-

stützen, s_j . Derartige Einflüsse lassen sich empirisch validieren und quantifizieren, sodass letztlich durch die Mitwirkung von Physikern ein formales dynamisches Modell aufgestellt wurde [4, 21], das auch einer weiteren analytischen Behandlung zugänglich ist. So findet man für den sozialen Einfluss I_i , der auf Agent i ausgeübt wird:

$$I_i = \sum_{j=1}^N \beta_{ij} [p_j(1 - \theta_i \theta_j) - s_j(1 + \theta_i \theta_j)] \quad (1)$$

Der erste Term, der die Überzeugungskraft der Gruppe beschreibt, verschwindet immer dann, wenn die Agenten i und j bereits die gleiche Meinung haben. Die Unterstützung im zweiten Term verschwindet dagegen, wenn die Agenten verschiedener Meinung sind. Die Größe β_{ij} berücksichtigt hierbei, dass der Einfluss zweier Agenten aufeinander mit der Distanz abnimmt. Ist der soziale Impact I_i negativ, also die Unterstützung für Agent i groß, dann hat er nur geringe Veranlassung, seine Meinung zu ändern. Ein möglicher Ansatz für die Übergangswahrscheinlichkeit zur entgegengesetzten Meinung lautet: $w(-\theta_i|\theta_i) \sim \exp\{I_i/T\}$. Hier tritt als weiterer Parameter die so genannte *soziale Temperatur* auf (die ihre formale Verwandtschaft mit der physikalischen Temperatur nicht leugnet). Sie beschreibt die Tatsache, daß Agenten aufgrund von Störungen in der sozialen Interaktion auch „falsche“ Entscheidungen treffen können. Abbildung 2 zeigt eine Simulation des Social-Impact-Modells mit einem zellulären Automaten, bei der die charakteristische Dynamik deutlich wird. Ausgehend von einer Zufallsverteilung bildet sich im System sehr schnell eine Mehrheit bzw. Minderheit heraus, wobei Agenten der gleichen Meinung die Tendenz zur Bildung von Domänen haben. Im deterministischen Modell ($T=0$) bleiben diese separaten Domänen stabil, während sie bei Berücksichtigung von Fluktuationen ($T>0$) zu zwei getrennten Phasengebieten verschmelzen.

Der Einfluss der Kommunikation

Für Soziophysiker haben Ising-artige Modelle den unbestreitbaren Vorteil, dass ihnen sowohl die formale Beschreibung als auch die zu erwartenden Resultate „vertraut“ vorkommen. Dies täuscht jedoch nicht über die grundlegenden Defizite derartiger Ansätze hinweg. So ist die Interaktion von Agenten in der Regel auf nächste Nachbarn oder auf Mean-Field-Wechselwirkung beschränkt. Außerdem werden Änderungen im System ohne Verzögerung wirksam (was einer Kommunikation mit unendlicher Geschwindigkeit entsprechen würde), und Gedächtniseffekte werden nicht berücksichtigt. Falls Migration modelliert wird, macht es sich ebenfalls nachteilig bemerkbar, dass die Eigenschaften der Agenten mit den Gitterplätzen verknüpft sind. Von Vorteil ist hier eine objekt-orientierte Beschreibung, wie sie zum Beispiel durch Modelle Brownscher Agenten realisiert ist [17], bei denen auch die Kommunikation untereinander explizit modelliert wird.

Dazu nimmt man eine Art raum-zeitliches Kommunikationsfeld $h_\theta(\mathbf{r}, t)$ für den Informationsaustausch an, das aus unterschiedlichen Komponenten $\theta \in M$ bestehen kann. Agenten mit der Meinung θ_i tragen mit einem individuellen Beitrag s_i zum Aufbau dieses Feldes bei, indem sie an ihrem jeweiligen Ort \mathbf{r}_i Information generieren. Diese Information hat jedoch nur eine endliche Lebensdauer, modelliert durch eine „Halbwertszeit“ $1/k_\theta$, das heißt, es wird ein *Gedächtniseffekt*

berücksichtigt. Außerdem kann die Information mit endlicher Geschwindigkeit im Gesamtsystem propagieren, was durch einen räumlichen Diffusionsprozess (D_θ) beschrieben wird. Die Dynamik des Kommunikationsfeldes lässt sich dann in Analogie zu Reaktions-Diffusions-Gleichungen modellieren, die aus der Theorie der Strukturbildung gut bekannt sind:

$$\frac{\partial}{\partial t} h_\theta(\mathbf{r}, t) = \sum_{i=1}^N s_i \delta_{\theta, \theta_i} \delta(\mathbf{r} - \mathbf{r}_i) - k_\theta h_\theta(\mathbf{r}, t) + D_\theta \Delta h_\theta(\mathbf{r}, t) \quad (2)$$

Für die Dynamik der Meinungsbildung wird vereinfacht angenommen, dass ein Agent mit der Meinung θ_i die Informationen über die Meinungen seiner Gegner und Befürworter miteinander vergleicht. Ähnlich wie beim Herdenverhalten wird er dann seine eigene Meinung entsprechend der verfügbaren Information anpassen:

$$w(-\theta_i|\theta_i) \sim \exp\left\{\frac{h_{-\theta}(\mathbf{r}_i, t) - h_\theta(\mathbf{r}_i, t)}{T}\right\} \quad (3)$$

Der soziale Impact I_i , der instantan und ohne Gedächtniseffekte wirksam wird, ist hier also durch die unterschiedlichen Informationen ersetzt, die dem Agenten an seinem Platz \mathbf{r}_i zugänglich sind – das heißt, die Informationsausbreitung wird implizit mit berücksichtigt. Prinzipiell können die Werte für k_θ und D_θ für die beiden Komponenten $\theta \in \{-1, +1\}$ unterschiedlich sein. Wenn man allerdings den Beitrag $s_i = s_\theta$ konstant setzt und die Gleichung für $h_\theta(\mathbf{r}, t)$ so skaliert, dass die Verhältnisse $\beta = k_\theta/s_\theta$ und $\gamma = D_\theta/s_\theta$ für beide Komponenten gleich sind, dann ergibt sich für die beiden Teilpopulationen eine Kommunikation auf unterschiedlicher Zeitskala, $\tau = t D_\theta/\gamma$. Das heißt, eine Erhöhung der Diffusionskonstanten D_θ modelliert in diesem Fall einen schnelleren Informationsaustausch bei sonst gleichen Bedingungen.

Wie wird sich nun eine unterschiedliche Effizienz der Kommunikation innerhalb der beiden Teilpopulationen auf die Herausbildung von Mehrheiten und Minderheiten auswirken? Dazu wurde das Verhältnis $d = D_{+1}/D_{-1}$ in Computersimulationen variiert. Abbildung 3 zeigt für zwei verschiedene Werte von d den

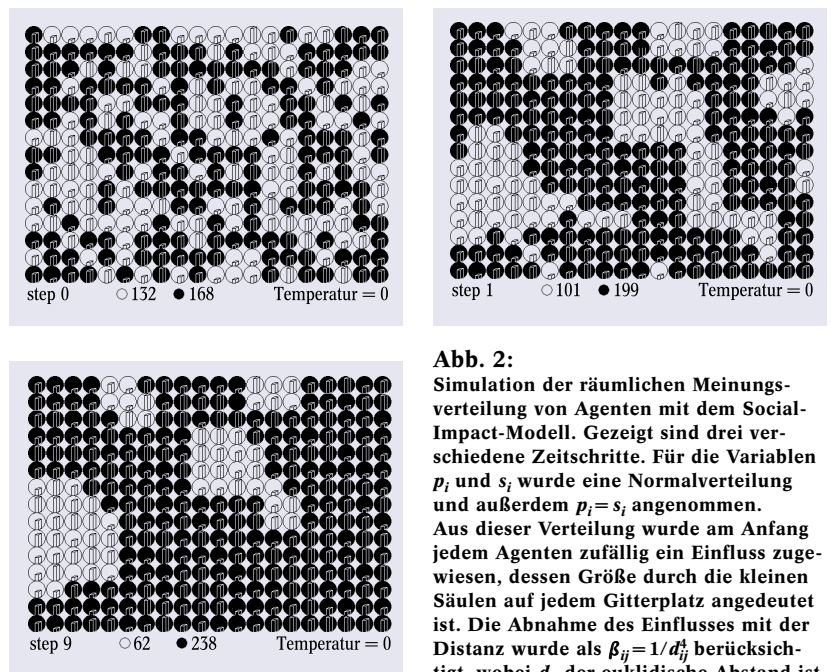


Abb. 2: Simulation der räumlichen Meinungsverteilung von Agenten mit dem Social-Impact-Modell. Gezeigt sind drei verschiedene Zeitschritte. Für die Variablen p_i und s_i wurde eine Normalverteilung und außerdem $p_i = s_i$ angenommen. Aus dieser Verteilung wurde am Anfang jedem Agenten zufällig ein Einfluss zugewiesen, dessen Größe durch die kleinen Säulen auf jedem Gitterplatz angedeutet ist. Die Abnahme des Einflusses mit der Distanz wurde als $\beta_{ij} = 1/d_{ij}^2$ berücksichtigt, wobei d_{ij} der euklidische Abstand ist.

Mittelwert des Populationsanteils, der die Meinung $\{-1\}$ vertritt, sowie die Minimum-Maximum-Streuung. Für $d \rightarrow 1$ kann die Teilpopulation $\{-1\}$ mit gleicher Wahrscheinlichkeit Majorität oder Minorität werden ($\langle x_{-1} \rangle \rightarrow 0,5$). Die große Streuung der Werte lässt ferner erkennen, dass in den verschiedenen Simulationen fast jede beliebige Teilpopulationsgröße realisiert wurde.

Mit anwachsendem d , also effizienterer Kommunikation der *anderen* Teilpopulation, sinken jedoch für die Agenten mit der Meinung $\{-1\}$ die Chancen auf eine Mehrheit im System. Bereits bei $d = 1,5$, also schon bei einem relativ geringen Unterschied in der Informationsausbreitung, bleibt nur noch ein Minderheitenstatus übrig. Dieses Ergebnis lässt sich durchaus in einem soziologischen Kontext interpretieren, wo Kommunikationseffizienz durch den Zugang zu Medien, aber auch durch die Möglichkeiten zum Spracherwerb bestimmt wird. Wer aufgrund mangelnder Sprachkenntnisse schlechter kommunizieren oder seine Meinung nicht adäquat über Medien verbreiten kann, wird in der Tat

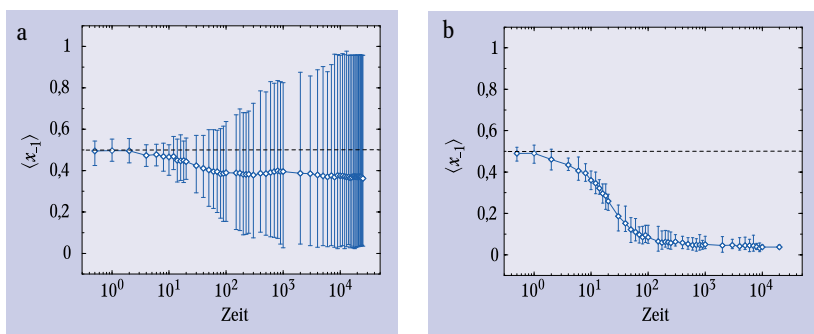


Abb. 3: Anteil der Teilpopulation von Agenten mit der Meinung $\{-1\}$, gemittelt über 20 Simulationen, sowie Minimum-Maximum-Streuung der erzielten Werte.

- ▶ a) Multi-Attraktor-Regime ($d=1,1$),
- ▶ b) Single-Attraktor-Regime ($d=1,5$) [6].

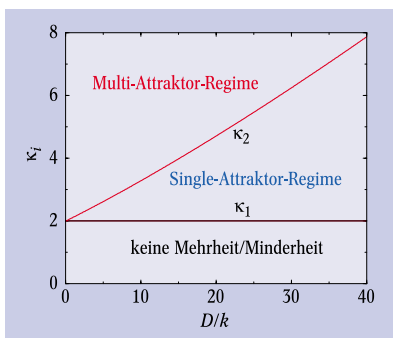


Abb. 4: Einfluss der Kommunikationsparameter D (Diffusion) und k (Lebenszeit der Information) auf den Ausgang des kollektiven Entscheidungsprozesses. $\kappa_1 = 2sN/ATk$ und $\kappa_2(D/K)$ bezeichnen zwei Bifurkationsparameter, die von den Systemgrößen abhängen. Erst oberhalb einer kritischen Populationsdichte tritt eine Spaltung in Mehrheiten und Minderheiten auf. Im Multi-Attraktor-Regime sind vielfältige Mehrheitsverhältnisse möglich, der Ausgang des Entscheidungsprozesses ist also „offen“, im Single-Attraktor-Regime gibt es statt dessen nur ein fixes (vorhersagbares) Mehrheits-/Minderheits-Verhältnis.

häufiger zu Minderheiten und Randgruppen gehören.

Die Tatsache, dass sich in der Simulation für $d \rightarrow 1$ beliebige Mehrheitsverhältnisse zwischen den Teilpopulationen etablieren können, deutet aus physikalischer Sicht auf ein *Multi-Attraktor-Regime* hin. Wie eine genauere Analyse zeigt, existieren in diesem Modell zwei Bifurkationsparameter (Abb. 4). Diese hängen einerseits von der mittleren räumlichen Dichte der Agenten N/A sowie der Stärke ihrer Meinung s ab, andererseits von Kommunikationsparametern wie der Diffusionsgeschwindigkeit D und der mittleren Lebenszeit k der Information. Ein Multi-Attraktor-Regime, das für beschränkte Vorhersagbarkeit und damit letztlich für Offenheit im kollektiven Entscheidungsprozess steht, wird nur erreicht, wenn das System *nicht* durch einen

schnellen Informationsaustausch „globalisiert“ wird und wenn die Agenten ihre Meinung mit einem entsprechenden „Nachdruck“ vertreten. Wird dagegen der Gedächtniseffekt verringert (zum Beispiel auch durch geschickte „Ablenkungsmanöver“) oder die Interaktion der Agenten gestört (etwa durch Erhöhung der sozialen Temperatur), dann erreicht das Agentensystem ein *Single-Attraktor-Regime*, in dem nur *ein* vorhersagbares Mehrheits- und Minderheitsverhältnis existiert.

Trittbrettfahrer und Kooperativeure

In den bisher beschriebenen Ansätzen sind die Agenten *reaktiv*; sie passen also ihre Meinung einer auf sie wirkenden Kraft an, die entweder „von außen“ oder durch die Einflüsse anderer Agenten zustande kommt. In einem sozio-ökonomischen Kontext jedoch verfolgen die Agenten sehr wohl eigene Ziele, denen sie ihre Meinung und ihr Verhalten unterordnen. Dies lässt sich durch eine individuelle Nutzensfunktion u_i beschreiben, die sowohl von äußeren Bedingungen (z. B. verfügbaren Ressourcen) als auch von der Interaktion mit anderen Agenten abhängt. Jeder Agent wird dann bestrebt sein, seinen eigenen Nutzen nach Möglichkeit zu maximieren.

Von einer solchen utilitaristischen Perspektive aus gesehen, wäre unsere Welt von lauter Egoisten bevölkert, die nur auf ihren eigenen Vorteil und die Maximierung ihres Gewinns bedacht sind – ein Zustand, den man zu keiner Zeit als wünschenswert empfand. Deshalb bildete die Frage, wie in einer solchen Welt überhaupt Kooperation zum Wohle des Ganzen möglich sein könnte, schon den Ausgangspunkt der Staatstheorien des 17. und 18. Jahrhunderts, die von Hobbes, Hume oder Smith entworfen wurden. Aber das Problem hat natürlich auch eine aktuelle Dimension, wenn es zum Beispiel um die Nutzung begrenzter Ressourcen oder öffentlicher Güter geht. In vielen Fällen ist es für den einzelnen Agenten lukrativ, seinen privaten Vorteil auf Gemeinkosten zu maximieren; das zeigt die Diskussion um „Trittbrettfahrer“ im öffentlichen Personennahverkehr, bei der gesetzlichen Krankenversicherung, bei der Begrenzung des CO_2 -Ausstoßes, usw.

Wie stellt sich dieses Problem aus der Sicht der Soziophysik dar? Der innere Freiheitsgrad des einzelnen Agenten beschreibt offensichtlich zwei verschiedene Verhaltensweisen, eine *egoistische*, in denen der Agent nicht kooperativ ist ($\theta_i = -1$ bzw. D , „to defect“) und eine *kooperative* ($\theta_i = +1$ bzw. C , „to cooperate“). Vom Zustand der Kooperation profitieren *alle* Agenten, auch die Trittbrettfahrer. Andererseits ist die Bereitschaft zur Kooperation, also der Verzicht auf den eigenen Vorteil, einem Kostenfaktor c vergleichbar, der nur für die kooperierenden Agenten auftritt. Für den einzelnen Agenten ergibt sich damit die Nutzensfunktion

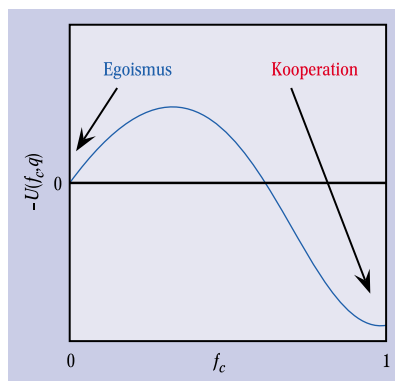
$$u_i = b f_c - c \delta_{\theta_i, +1}; \quad b > c > b/N \quad (4)$$

wobei $f_c = N_c/N$ der Anteil kooperierender Agenten in der Population ist und b der Nutzen pro Kooperativeur. Die Kosten der Kooperation werden für den einzelnen Agenten also erst dann kompensiert, wenn eine kritische Zahl von Agenten kooperiert. Da aber jeder darauf bedacht ist, seinen Nutzen zu maximieren, wird er im ersten Schritt keine Veranlassung zur Kooperation haben, sodass der Anfangszustand des Gesamtsystems lauter Egoisten enthält. Der Übergang in einen globa-

len Zustand der Kooperation wäre dann einem *Phasenübergang erster Ordnung* vergleichbar, wie er zum Beispiel durch die Keimbildungstheorie beschrieben wird. Entsprechend den Gesetzen der statistischen Physik wird ein solcher Phasenübergang nur dann eine relevante Wahrscheinlichkeit haben, wenn das thermodynamische Potential des Gesamtsystems im Zustand der Kooperation ein tieferes Minimum hat als im Zustand des Egoismus und wenn die Energiebarriere, die die beiden Zustände trennt, nicht zu hoch ist (Abb. 5).

Ein thermodynamisches Potential mit diesen Eigen-

Abb. 5: Gesamtnutzen eines Agentensystems in Abhängigkeit vom Anteil kooperierender Agenten f_c . q beschreibt einen Satz von Kontrollparametern (z. B. den Zeithorizont H der Erwartungen), die den Verlauf der Energiefunktion wesentlich mit bestimmen.



schaften lässt sich in der Tat aufstellen, wenn man folgende Annahmen macht [13]: (i) jeder Agent hat einen Zeithorizont H , um die Zukunft zu beurteilen, wobei für $t' > t$ zu erwartende Gewinne mit einem Faktor $e^{-t'/H}$ gewichtet werden, (ii) eine Entscheidung des Agenten, sein Verhalten zu ändern, $\theta_i \rightarrow -\theta_i$, stimuliert andere Agenten, dasselbe zu tun – eine Art Herdeneffekt, der auf Imitation beruht. Das thermodynamische Potential hängt neben dem Parameter H stark von der Populationsgröße N ab. Man findet, dass unterhalb einer kritischen Größe $N_c^{(1)}$ der „egoistische“ Ausgangszustand instabil ist, in kleinen Populationen setzt sich Kooperation also immer durch (monostabiles System). Dann gibt es einen Bereich mittlerer Populationsgröße, $N_c^{(1)} < N < N_c^{(2)}$, in dem der Übergang vom Egoismus zur Kooperation *möglich* ist (bistabiles System), allerdings hängt die Übergangszeit exponentiell von der Populationsgröße ab. Oberhalb einer zweiten kritischen Populationsgröße $N_c^{(2)}$ wird dieser Übergang dann unmöglich (monostabiles System), d. h. *große* Populationen kooperieren *nicht* mehr. Die kritischen Werte $N_c^{(1)}$, $N_c^{(2)}$ hängen außerdem noch vom Erwartungshorizont H ab – nur für $H \rightarrow \infty$ wird auch in großen Populationen noch Kooperation möglich sein. Dieses Modell eines so genannten *public good games* (PGG) basiert natürlich auf sehr vereinfachten Annahmen. Gleichwohl erlaubt es interessante Schlussfolgerungen vor dem Hintergrund der Globalisierungsdebatten – etwa, wenn es um die Frage geht, ob in einer „größer“ werdenden Welt tatsächlich auch ein „besserer“ Zustand für alle erreicht werden kann.

Evolutionäre Spieltheorie

Von besonderem Interesse in sozialen und vor allem auch in ökonomischen Systemen sind Situationen, in denen ein Agent *direkt* mit anderen Agenten, beispielsweise aus seiner unmittelbaren Umgebung, interagiert. Auch hier hat er meist die Auswahl zwischen zwei Handlungen, zu kooperieren (C) oder nicht (D), wobei er nicht weiß, wie sich sein jeweiliger Gegenspieler im gleichen Zug entscheiden wird. Allerdings kann der Agent (mit Ausnahme der ersten Interaktion) mögli-

cherweise auf die Erfahrungen zurückgreifen, die er mit seinem Gegenspieler gemacht hat. Bei einem Ein-Schritt-Gedächtnis, $n_m = 1$, erinnert sich Agent i also daran, ob Agent j sich beim letzten Treffen kooperativ oder als Schmarotzer verhalten hat und wird in Abhängigkeit davon sein eigenes Verhalten festlegen. Aus dieser Konstellation resultieren für $n_m = 1$ prinzipiell acht verschiedene Strategien, die in einem 3-Bit-Binärstring, $[I_0|I_c I_d]$, kodiert werden können. Der Wert 1 für I bezeichnet dabei C, 0 bezeichnet D, wobei I_0 die Anfangshaltung von Agent i ist. Die anderen beiden Werte, I_c und I_d , kodieren die Antwort von Agent i auf die vorherige Handlung von Agent j . Also $I_c = 1$, falls Agent j im letzten Zeitschritt kooperativ war und Agent i deshalb beschließt, im jetzigen Zeitschritt ebenfalls kooperativ zu sein. Wenn er stattdessen nicht kooperieren will, gilt $I_c = 0$. Hat Agent j im letzten Zeitschritt nicht kooperiert und Agent i beschließt dennoch, im jetzigen Zeitschritt zu kooperieren, dann gilt $I_d = 1$, usw.

Die bekannteste der acht möglichen Strategien ist [110], TFT (tit for tat), oder „Wie du mir, so ich dir“. Agent i startet kooperativ ($I_0 = 1$) und bleibt kooperativ, solange Agent j ebenfalls kooperativ ist ($I_c = 1$). Falls aber Agent j unkooperativ wird, passt Agent i sich dem an und kooperiert ebenfalls nicht mehr ($I_d = 0$). Natürlich hängt die Entscheidung, ob Agenten kooperieren oder nicht, grundsätzlich von dem dabei zu erwartenden Gewinn (Payoff) ab. In einer als „Prisoner's Dilemma“ (PD) bekannten Klasse von Spielen sieht die Payoff-Matrix¹⁾ für Zwei-Personen-Spiele folgendermaßen aus:

	C	D
C	R = 3	S = 0
D	T = 5	P = 1

Ein Agent wird also den größten Payoff haben, wenn er gegen einen kooperativen Agenten (C) spielt, den er ausnutzt (D). Dann erhält er $T = 5$ Punkte, während der kooperative Agent $S = 0$ Punkte bekommt. Der *Gesamt*-Payoff wäre in diesem Fall $T + S = 5$. Er würde dann maximal werden, wenn *beide* kooperieren (C, C), also jeder $R = 3$ Punkte bekommt. Daraus ergibt sich das Dilemma zwischen dem Eigennutz auf Kosten anderer ($T > S$) und dem Gemeinwohl, bei dem jeder ein wenig zurückstecken muss ($R < T$), damit es allen besser geht ($2R > T + S$). Die Strategie, kooperative Agenten auszunutzen, ist natürlich besonders erfolgreich, wenn man diese nur einmal trifft, denn bereits beim zweiten Mal kann der andere Agent seine Strategie auf dieses Schmarotzer-Verhalten einstellen. Die Zahl der Spiele zwischen je zwei Agenten, n_g , spielt also eine besondere Rolle. Für $n_g \geq 2$ spricht man auch vom *iterated Prisoner's Dilemma* (IPD).

Wie sieht die Dynamik eines solchen Spiels aus, wenn man die Agenten wie bei einem Ising-System auf einem Gitter verteilt? Der innere Freiheitsgrad θ_i des Agenten soll dabei eine der acht möglichen Strategien bezeichnen. In jeder „Generation“ (die diskrete Zeiteinheit) spielt jeder Agent entsprechend seiner Strategie $n_g = 2$ mal gegen seine vier nächsten Nachbarn. Danach hat er die Möglichkeit, seinen Payoff mit dem seiner Nachbarn zu vergleichen und in der nächsten Generation die Strategie zu übernehmen, die in seiner Nachbarschaft am erfolgreichsten war. Dieser Adaptationsprozess wird in einem räumlichen System im Laufe der Zeit zur lokalen Dominanz bestimmter Strategi-

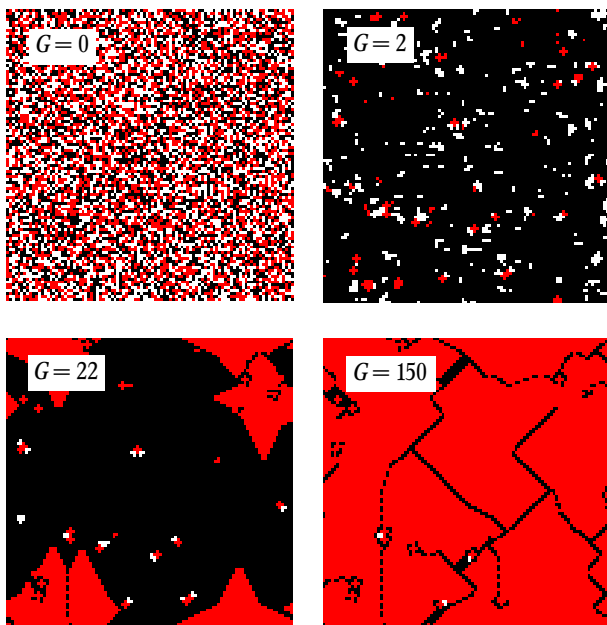
1) Die Zahlenwerte sind nur Beispiele. Allgemein müssen für die Payoffs des PD folgende Ungleichungen erfüllt sein: $T > R > P > S$, $2R > S + T$.

en führen – und es ist von Interesse, welche davon sich letztlich auf einer globalen Ebene durchsetzen werden.

Abbildung 6 zeigt eine solche Dynamik für den Fall dreier Strategien. Agenten mit der Strategie TFT [110] sind rot gekennzeichnet, Agenten, die sich ausschließlich unkooperativ verhalten, [000], sind schwarz markiert. Die dritte Strategie, [001] (anti-TFT), ist von Interesse, weil die Agenten (weiß markiert) gewissermaßen antizyklisch agieren. Sie starten nichtkooperativ, machen dann aber im nächsten Zug immer genau das Gegenteil von dem, was ihr Gegenüber gerade getan hat. Einem Schmarotzer begegnen sie kooperativ, einen kooperativen Agenten dagegen nutzen sie aus.

Wie man an der Computersimulation sehr gut sehen kann, setzten sich in den ersten Generationen diejenigen Agenten durch, die ihren Gewinn allein auf Kosten vorhandener Kooperationspartner machen. Agenten mit den Strategien [110] und [001] überleben dagegen nur noch in kleinen Clustern, deren Ende vorprogrammiert

Abb. 6: Verteilung von Agenten mit den Strategien [000] (schwarz), [001] (weiß) und [110] (rot) auf einem Gitter der Größe 100×100 . G (die Generation) gibt den jeweiligen Zeitschritt an, die Zahl der Interaktionen zwischen zwei benachbarten Agenten pro Generation ist $n_g = 2$.



scheint. Dann aber wandelt sich das Bild grundlegend: Ausgehend von diesen kleinen Clustern gewinnen Agenten mit der kooperativen Strategie TFT die Oberhand und breiten sich in Fronten über das gesamte System aus – die phänomenologischen Ähnlichkeiten zum Domänenwachstum, beispielsweise auf Oberflächen, sind unverkennbar. Getrieben wird diese Wachstumsdynamik natürlich von dem Payoff-Gefälle an der jeweiligen Front, wobei die lokalen Konfigurationen eine besondere Rolle spielen: Agenten mit der Strategie TFT können sich anfänglich nicht durchsetzen, solange Agenten mit der Strategie anti-TFT in ihrer Nachbarschaft sind – erst wenn diese von den „ewigen“ Schmarotzern dezimiert worden sind, haben die kooperativen Agenten wieder eine Chance. Warum? Weil in einer Nachbarschaft, in der Agenten mehrmals kooperieren, alle einen größeren Payoff erzielen als ein Schmarotzer, der nach dem ersten Mal bereits „enttarnt“ ist und danach leer ausgeht. Insofern setzt sich die kooperative Strategie letztlich auf dem globalen Level durch. An den Domänenrändern können Schmarotzer nur für $n_g = 2$ überleben, für $n_g = 3$ verschwinden auch sie vollständig. Gleiches gilt für Agenten mit der Strategie anti-TFT, die nur für $n_g = 2$ vereinzelt überleben können.

Dieser überwältigende Sieg der Kooperation freut

natürlich den Soziophysiker, selbst wenn ihm bewusst bleibt, dass die Realität komplexer ist als seine Modelle und Schmarotzer oftmals noch die Oberhand behalten. Vielleicht liegt ja hier einer der Gründe für die Attraktivität dieses neuen Forschungsgebietes: Durch das stückweise Verstehen sozialer Interaktion erhält nicht nur die Hoffnung auf eine bessere Welt neuen Auftrieb, auch der Weg dorthin wird am Modell schon einmal deutlich.

*

Für die Erstellung der Abbildungen dankt der Autor Laxmidhar Behera (Abb. 1), Krzysztof Kacperski (Abb. 2) und Robert Mach (Abb. 6).

Literatur

- [1] W. Weidlich, Physics Reports **204**, 1 (1991)
- [2] D. Helbing, Quantitative Sociodynamics. Stochastic Methods and Models of Social Interaction Processes, Kluwer Academic (1995)
- [3] S. Solomon et al., Physica A **277**, 239 (2000)
- [4] J. Holyst, K. Kacperski und F. Schweitzer, Physica A **285**, 199 (2000)
- [5] S. Galam und J.-D. Zucker, Physica A **287**, 644 (2000)
- [6] F. Schweitzer, J. Zimmermann und H. Mühlenbein, Physica A **303**, 189 (2002)
- [7] J. Mimkes, Journal of Thermal Analysis and Calorimetry **60**, 1055 (2000)
- [8] K. Sznajd-Weron und J. Sznajd, Int. Journal of Modern Physics C **11**, 1157 (2000)
- [9] F. Slanina, Physica A **286**, 367 (2000)
- [10] D. Challet und Y.-C. Zhang, Physica A **256**, S. 14 (1998)
- [11] H. Ebel und S. Bornholdt, Phys. Rev. E **66**, 1 (2002)
- [12] G. Szabo und C. Hauert, Phys. Rev. Lett. **89**, 118101 (2002)
- [13] N. S. Glance und B. A. Huberman, Journal of Mathematical Sociology **17**, 281 (1993)
- [14] J. Berg und A. Engel, Phys. Rev. Lett. **81**, 4999 (1998)
- [15] K. Klemm et al., Phys. Rev. E **67**, 026120 (2003)
- [16] G. Nigel Gilbert und K. Troitzsch, Simulation for the Social Scientist, Open University Press (1999)
- [17] F. Schweitzer, Brownian Agents and Active Particles. Collective Dynamics in the Natural and Social Sciences, Springer Verlag, Heidelberg (2003)
- [18] D. Stauffer, Comp. Phys. Commu. **146**, 93 (2002)
- [19] B. Schlechter, New Scientist **175**, 2357, S. 42 (2002)
- [20] A. Nowak, J. Szamrej und B. Latané, Psychological Rev. **97**, 362 (1990)
- [21] M. Lewenstein, A. Nowak und B. Latané, Phys. Rev. A **45**, 703 (1992)

Der Autor

Frank Schweitzer hat in theoretischer Physik und Wissenschaftsphilosophie promoviert. Seit seiner Habilitation in Physik ist er am Fraunhofer-Institut für Autonome Intelligente Systeme in Sankt Augustin sowie als Privatdozent an der Humboldt-Universität Berlin tätig. Gegenstand seiner jetzigen Forschung ist die Agenten-basierte Modellierung und Computersimulation von komplexen Systemen mit besonderem Schwerpunkt auf biologischen und sozio-ökonomischen Fragestellungen. Frank Schweitzer ist Mitglied des Steering Committees des EU-Network of Excellence „Complex Systems“ (EXYSTENCE) und Vorsitzender des DPG-Arbeitskreises „Physik sozio-ökonomischer Systeme“ AKSOE (www.ais.fhg.de/~frank/AKSOE).

