Internalisierung und Rationalität – Der Akteur als Mehrebenen-Prozess

Johannes Huinink

Vortrag im Kolloquium des Instituts für Soziologie der Universität Leipzig am 27.1.2021

1. Vorwort

Wie oft haben wir in den 1990er Jahren, als ich in Leipzig war, einige Male über das so genannte Opportunismus-Problem oder das Problem der Kooperation diskutiert. Dabei ging es um die Frage, wie Kooperation zwischen rationalen Akteuren entsteht und aufrechterhalten wird, wenn sie doch durch Detektion eines oder mehrerer Kooperationspartner einseitig ausgenutzt und damit untergraben werden kann. Ich hatte damals schon die Vermutung, dass wir mit "unseren" Mitteln, mit denen wir als Soziologen die RC-Diskussion führen, das Problem nicht klären können. Man muss und kann, so meine heutige Überzeugung viel von anderen Forschungsgebieten lernen, die sich ja ebenfalls intensiv mit diesem Problem beschäftigen. Ein interessantes frühes Beispiel ist die Zusammenarbeit von dem Politologen Axelrod mit dem Biologen Hamilton (Axelrod/Hamilton 1981).

Für befriedigende Antworten braucht man meines Erachtens die evolutionär-anthropologische, psychologische und neurowissenschaftliche Forschung. Das Problem der Kooperation ist auch dort immer noch ein "heißes" Thema gewesen und immer noch umstritten. Es sollte nicht verschwiegen werden, dass in diesen Disziplinen noch keine Einigkeit herrscht. Jüngste Darstellung der Kontroverse findet sich McCollough (2020). McCullough und Koautor melden auch Zweifel bezüglich der externen Validität der in diesem Zusammenhang wichtigen empirischen 'public-good games' – insbesondere was den ersten Zug angeht – an (McAuliffe/ McCullough 2017, Frey 2017). Solche Auseinandersetzungen sind für sich schon interessant. Sie bedeuten auch nicht, dass man sich mit den Thesen und Studien, die dort verhandelt werden, in der Soziologie nicht auseinandersetzen muss.
Mehr und mehr verstärkt sich daher seit einiger Zeit bei mir das Ärgernis, dass die Soziologie sich diesbezüglich weitgehend abstinent zeigt – im Gegensatz zur Ökonomie etwa. Ich schließe mich da Daniel Dennett an, der in dem Buch „Den Bann brechen. Religion als natürliches Phänomen“ (im englischen Original publiziert in 2006) schreibt: „Es ist eine gängige Annahme unter den Sozial- und Geisteswissenschaften, die es oft als »reduktionistisch« (und als schlechten Stil) erachten, wenn die Frage nach der biologischen Basis solch wunderbarer und bedeutender Phänomene [hier: die Religion, JH] nur *gestellt* wird. Ich sehe schon die Kulturanthropologen und Soziologen verächtlich mit den Augen rollen – »O nein! Jetzt kommt wieder Darwin daher und mischt sich ein, wo er nicht gebraucht wird!« –, ...“ (Dennett 2016: 100f).

Herbert Gintis schreibt in einem seiner Bücher zur Spieltheorie: "The self-conceptions and dividing lines among the behavioral disciplines make no scientific sense. How can there be three separate fields, sociology, anthropology, and social psychology, for instance, studying

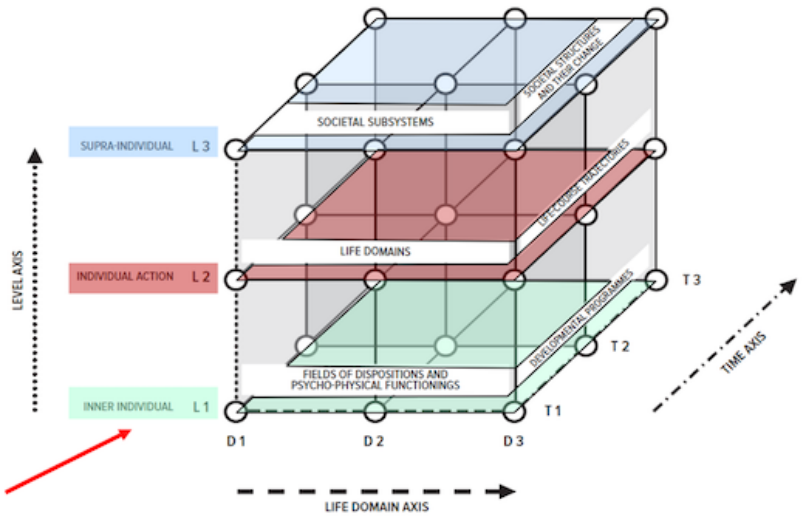
social behavior and organization? How can the basic conceptual frameworks for the three fields, as outlined by their respective Great Masters and as taught to Ph.D. candidates, have almost nothing in common? In the name of science, these arbitraries must be abolished" (Gintis 2009: xv).

In diesem Beitrag beschäftige ich mich beispielhaft mit Themen, die neben und im Zusammenhang mit der Forschung zum Problem der Kooperation interessante Gegenstände interdisziplinärer Forschung darstellen. Das tue ich an Hand der Konzepte der Rationalität und Internalisierung. Das Problem der Kooperation, wie man es auch in der Biologie nennt, ist dabei immer zentral mitbedacht, wird aber nur unter dem Gesichtspunkt der Evolution der Internalisierung prosozialer Verhaltensnormen (als genetisch ermöglichte Disposition) behandelt. Ich beginne mit ein paar Bemerkungen zum Akteur als Mehrebenen-Prozess, daran anschließend, zum Rationalitätskonzept. Danach werde auf Beispiele evolutionstheoretischer Forschung zur Internalisierung von Normen, Verhaltensregeln oder Werte eingehen. Sie hängt, wie wir sehen werden, eng mit dem Problem der Kooperation und Altruismus zusammen. So schreibt Gintis, der sich ausführlich mit dieser Frage beschäftigt hat: „In effect, altruism 'hitchhikes' on the personal fitness-enhancing capacity of norm internalization" (Gintis 2016a: 3).

2. Der Life Course Cube und die "innerindividuelle" Mehrebenen-Hierarchie

Im Zuge der Überlegungen zu einem einfachen Modell für die theoretische und empirische Lebenslaufforschung habe ich mir vor einigen Jahren ein Schema überlegt, das ich den Life-Course Cube genannt habe (publiziert in Bernardi/Huinink/Settersten 2019). Ich gehe darauf hier nicht im Detail ein, sondern will einen besonderen Aspekt hervorheben (siehe Abb. 1).

Abb 1: Der Life Course Cube; die äußere und „innerindividuelle Mehrebenen-Hierarchie" (und die Zeit und die Lebensbereiche) (nach Bernardi/Huinink/ Settersten 2019)

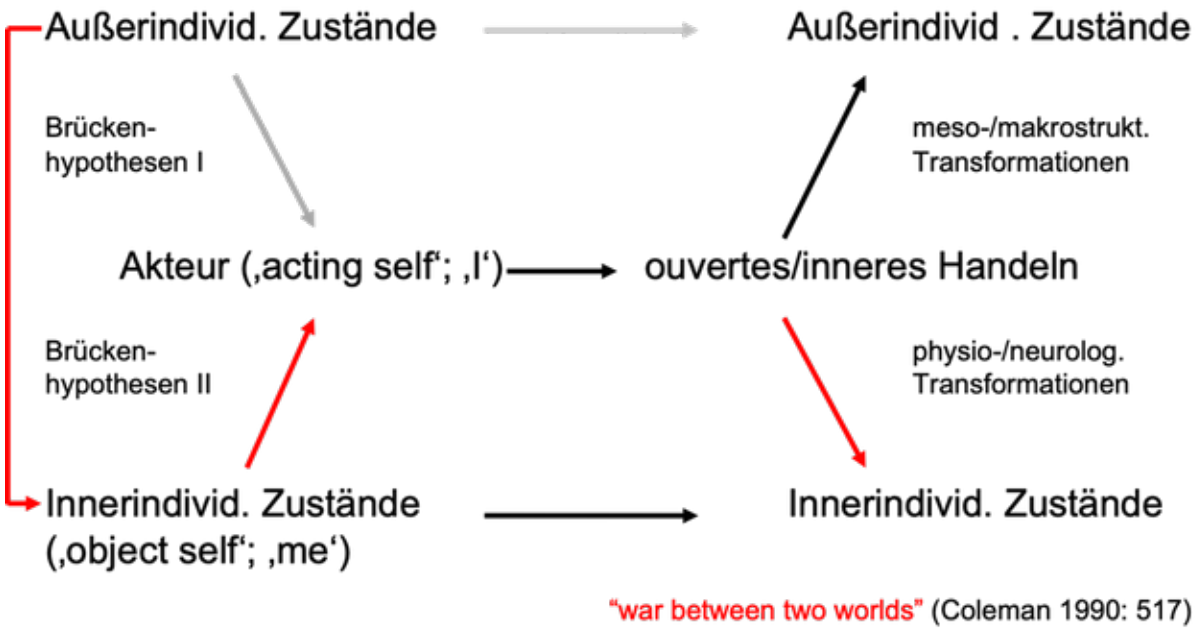


Eine Dimension in dem Cube ist die Mehrebenen-Hierarchie von Prozessebenen, Bekannt ist die interdependente Beziehung individueller Handlungsprozesse zu unterschiedlich stabilen bzw. veränderlichen, relevanten supraindividuellen Prozessen, wie sie unter anderem von Coleman (1990) eingeführt wurde. Man könnte diesbezüglich auch von der von mir hier so genannten „außerindividuellen Mehrebenen-Hierarchie“ sprechen, die hier in dem Cube nur auf eine supra-individuelle Prozessebene verkürzt angedeutet ist.

Hinzu kommt, und das wird bisher in der Soziologie zu wenig mitgedacht, eine "innerindividuelle Mehrebenen-Hierarchie“, die für den individuellen Handlungsprozess ebenso wichtig und sogar wirkmächtiger ist, als die supra-individuelle Einbettung (vgl. Abb. 2). Auch hier ist in dem Cube eine Mehrebenen-Struktur nur durch eine intra-individuelle Ebene verkürzt angedeutet. Diese innerindividuellen Ebenen – wir auch die außerindividuellen – könnte man, wie ich in einem früheren Beitrag in definitorischer Absicht vorgeschlagen habe, durch unterschiedliche Veränderungsgeschwindigkeiten der Prozessgrößen unterscheiden (Huinink 1989).

Das erweiterte Modell deutet sich durchaus auch bei Coleman an wenn er in seinem interessanten Kapitel zum „Self“ in den „Foundations“ zum dem, was er den „elementaren Akteur“ nennt schreibt: "Because the actor both acts and is an object of others' actions, it appears useful to conceive the self as consisting of at least two parts: an object self, which experiences satisfaction or the lack of it; and an acting self, which is in the service of the object self, attempting to bring it satisfaction. " (Coleman 1990: 507).

Abb. 2: Erweitertes Mehrebenen-Model**l**



Wir verweisen weitergehend auf die innerindividuellen Prozesse, die zum Teil nicht für einzelne Individuen, ja nicht einmal für Menschen spezifisch sind, zum Teil aber nur beim Menschen vorzufinden sind und teilweise bei einzelnen Individuen unterschiedlich ausgeprägt sind. Wir haben es Prozessen zu tun, deren Parametrisierung unterschiedlich stabil bzw. veränderbar ist. Letztendlich steuern sie das individuelle Handeln und Verhalten (des 'agent' im Colemanschen Sinne).

Die Prozesse auf den verschiedenen innerindividuellen (Sub-)Ebenen sind interdependent dem Akteurshandeln verknüpft. Dessen Folgen schlagen sich in Form individueller Erfahrungen und Lernprozessen darin nieder. Darüber hinaus sind sogar die innerindividuelle und außerindividuelle Mehrebenen-Hierarchie über das Scharnier des individuellen Handlungsprozesses, der ja selbst bewusst nach innen (inneres Handeln) und prinzipiell beobachtbar nach außen (ouvertes Handeln) gerichtet sein kann, miteinander interdependent verknüpft.

Das führt zu einer Reihe von Typen von Brücken- und Transformationshypothesen. Mit der Einführung der innerindividuellen Mehrebenen-Hierarchie brauchen wir zu eine modifizierte Klasse von Top-Down-Brückenhypothesen, von denen wir annehmen, dass sie die Handlungsebene überspringen und direkt auf die innerindividuellen Zustände abzielen, die man mit dem Colemanschen 'object self' in Beziehung setzen könnte, (Brückenhypothesen I). Hinzu kommen dann die Bottom-Up-Hypothesen zu dem Einfluss der innerindividuellen Zustände auf den Akteur als handlungsausführende Einheit ('acting self' nach Coleman). Hier werden Hypothesen zu Verbindung der innerindividuellen Ebene zur Handlungsebene (Brückenhypothesen II) formuliert, die Gesetzmäßigkeiten der innerindividuellen Prozesse im Hinblick auf die Handlungswahl des Akteurs berücksichtigen.
Neben der meso- und makrostrukturellen Transformation individuellen Handeln auf die außerindividuellen Zustände sind auch die Auswirkungen von ouverten und inneren Handlungen auf Zustände auf den verschiedenen Ebenen der innerindividuellen Mehrebenen-Hierarchie (des 'object self') zu beachten (physio-/neurologische Transformationen). Handlungen des 'active self' verändern physischer Zustände und psychische Dispositionen, wie die Bewertungsstrukturen auf das 'object self', allein schon durch die selektive Anreicherung des in der Zeit Erfahrenen (Gedächtnis). Auf der innerindividuellen Ebene sind auch unwillkürlich stattfindende physische und psychisch/neurologische Transformationsprozesse zu beachten sein.
Auch Coleman ist sich dieser doppelten Struktur von Einflüssen auf das Handeln bewusst, wenn er vom "war between two worlds", d.h. zwischen den sich verändernden äußeren und inneren Rahmenbedingungen des Akteurhandelns, schreibt (Coleman 1990: 517). Er diskutiert auch verschiedene Möglichkeiten, "change inside the actor" (Identifikation, Balancetheorie, Pico-Ökonomie) zu erklären, zu denen auch Internalisierungsprozesse zählen. Aus interdisziplinärer Perspektive sind seine Ausführungen, so inspirierend sie sein mögen, aber unzureichend.

Wenn man die Interdependenz wischen der innerindividuellen und außerindividuellen Mehrebenen-Hierarchie akzeptiert, so wird einem einleuchten, dass sozialwissenschaftliche, neurowissenschaftliche, psychologische und evolutionäre-anthropologische Kompetenz kooperieren muss, um individuelles Handeln und seine innerindividuellen und außerindividuellen Konsequenzen zu erklären.
Das heißt nicht unbedingt, dass man in Analysen Ebenen-Hierarchien nach unten oder nach oben nicht gezielt abschneiden kann. Ein Grund dafür ist die Tatsache, dass ebenspezifische Prozessgrößen oder -phänomene als emergent betrachtet werden können – oder gar müssen–, da deren Veränderung sich aufgrund der hohen Interaktionsdichte der sie erzeugenden Elemente auf der niedrigeren Ebene nicht mehr analytisch auf die einzelnen "Aktivitäten" dieser Elemente auf der niedrigeren Ebene zurückführen lassen. Sie sind gleichsam (nur) als integriertes Ergebnis des dynamischen Interaktionszusammenhangs dieser Elemente beobachtbar. Man denke an die Temperatur eines Gases oder den Gasdruck in der Physik. Das Bewusstsein wird von Pfützner zum Beispiel auch als Emergenzphänomen angesehen, dass ab einer bestimmten Interaktion- oder Komplexitätsdichte in einem neuronalen Netz entsteht (Pfützner 2014). Auch Dehaenes Beobachtungen zum Entstehen von bewusster Wahrnehmung könnte meines Erachtens so gedeutet werden (Dehaene 2014).

Als eine Konkretisierung einer innerindividuellen Mehrebenen-Hierarchie kann man das Vier-Ebenen-Modell der Persönlichkeit (des Gehirns als Steuerungsorgan) von Gerhard Roth anführen (Roth 2015, 145ff; vgl. auch Strüber/Roth 2020; siehe Abb. 3). Er unterscheidet die vegetativ-affektive Ebene (untere limbische Ebene, stammesgeschichtliches Erbe, unbewusst), die Ebene der emotionalen Konditionierung (mittlere limbische Ebene, Belohnungs- und Motivationssystem, unbewusst), die Ebene des individuell-sozialen Ich (obere limbische Ebene, soziales Lernen, bewusst) und die Ebene des kognitiv-kommunikativen Ich (kognitiv-sprachliche Ebene, Kommunikation und verstandesgelenktes Denken, bewusst). Ich kann hier auf die neurologischen Einzelheiten und die experimentellen Befunde, die dahinterstehen, nicht eingehen. Meine Intention, das zu zeigen ist, ist durch einen wichtigen Sachverhalt begründet, der sich auch so deutlich machen lässt.

Abb. 3: Vier-Ebenen-Modell der Persönlichkeit von Strüber/Roth (2020, 136f)

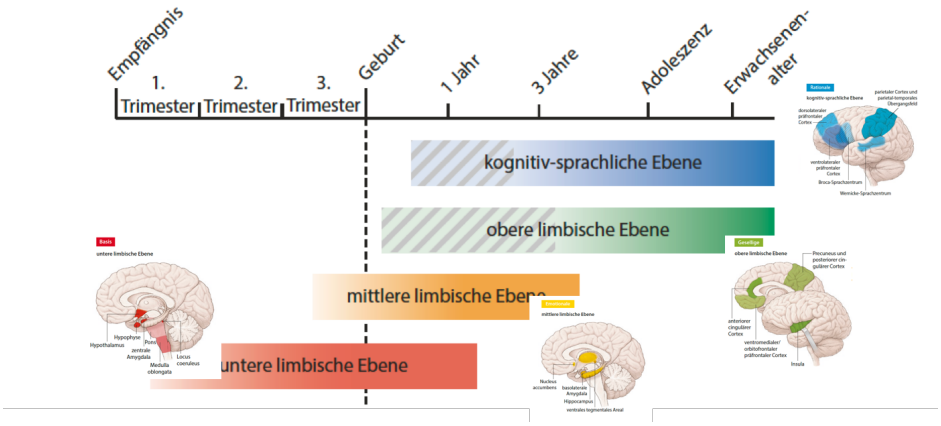
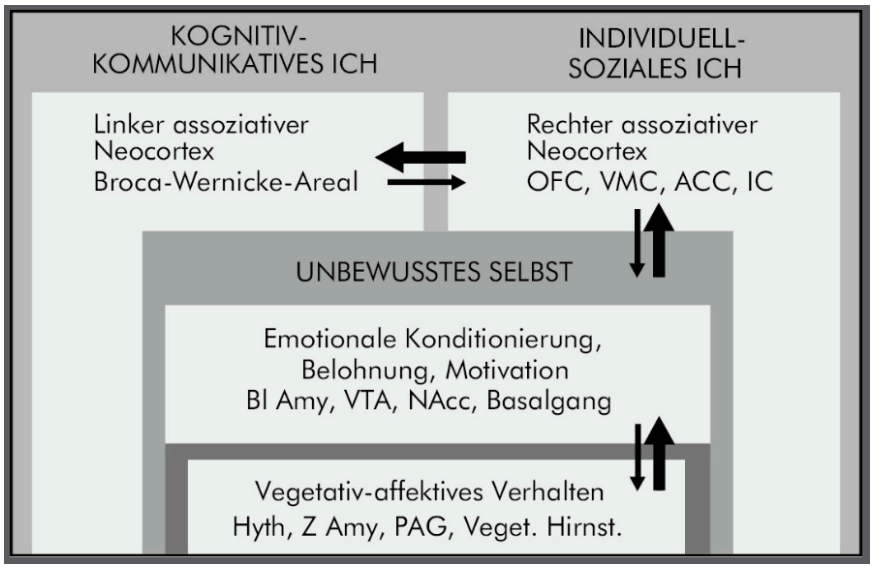


Abb. 4: Vier-Ebenen-Modell der Persönlichkeit von Roth (2015. 147)



In der Abbildung 4 wird angenommen, dass die Einflussmöglichkeiten der Prozessebenen aufeinander asymmetrisch geartet sind. Das heißt, die Prozesse auf der jeweils "niedrigeren" Ebene beeinflussen die Prozesse auf der nächsthöheren Ebene stärker als umgekehrt. Doch es gibt umgekehrte Effekte. Das bedeutet zum Beispiel, dass auf den Ebenen der vegetativ-affektive oder emotionalen Konditionierung verankerte Dispositionen durch erlernte Überzeugungen und schließlich auch durch situationales, verstandesmäßiges Denken (Ebene des kognitiv-kommunikativen Ich), nicht gelöscht, aber in ihrer Wirkung überschrieben oder modifiziert werden können. Beispielsweise können negative Auswirkungen eines des tief verwurzelten "ingroup-favouritism" durch die "Intervention" von internalisierten Gleichheitsnormen oder situational instrumentellen Zweckmäßigkeitsüberlegungen reduziert werden. Wichtig ist aber – und deshalb müssen Soziologen sich damit auseinandersetzen –, dass es sich um Überschreibungsvorgänge handelt, die auf Grund der Veränderung von situationalen Rahmenbedingungen oder die Abschwächung von Internalisierung auch abbrechen oder an Wirkung verlieren können.
Entsprechungen zu dem hier vorgestellten Modell findet man in der Unterscheidung zwischen schnellem und langsamem Denken etwa bei Kahneman (Kahneman 2011) oder zur Dual Process Theory (Fazio u.a. 1999, Vaisay 2009). Bezüge zu neuronalen und physiologischen Prozessen, zu Erkenntnissen über die Bedeutung von Emotionen und zu vielen anderen Aspekten lassen sich daran anschießen. Siehe zum Beispiel Damasios bemerkenswerte Abhandlung über “The Strange Order of Things: Life, Feeling, and the Making of Cultures” (Damasio 2018)

Soziologen müssen m.E. den Forschungsstand dazu kennen und im "Hinterkopf" behalten, wie im Folgenden zu plausibilisieren versucht. Das Abschneiden von Analyse-Ebenen ist daher grundsätzlich immer zu begründen. Coleman zum Beispiel tut das in seinen "Foundation" ansatzweise in seiner Rechtfertigung, sich auf ökonomisch rationales Verhalten zu beschränken und die komplexen innerindividuellen Prozesse radikal zu vereinfachen. So

kann man es im Kapitel 1 und dem außerordentlich interessanten Kapitel 19 (zum Selbst) nachlesen. Dabei geht es nicht um eine Relativierung der Zielgröße der "Befriedigungsoptimierung" für das individuelle Selbst, sondern um die Einschränkung der Komplexität betrachteter Bedingungsfaktoren auf den unterschiedlichen inner- und außerindividuellen Prozessebenen.

3. Rationalität und der neuronale Entscheidungsalgorithmus

Das Handeln (und Verhalten; sowie alle physiologischen Vorgänge) des Menschen wird durch einen hoch nichtlinearen inneren Mehrebenen-Prozess gesteuert, der letztendlich auf physikalisch-chemischen Vorgängen beruht. So schreibt zum Beispiel der amerikanische Philosoph Searle: „At the system level we have consciousness, intentionality, decisions, and intentions. At the micro level we have neurons, synapses, and neurotransmitters“ (Searle 2001: 503).
Das kann man als eine materialistische Grundlegung menschlichen Denkens und Verhaltens verstehen. Was bedeutet das für die Erklärung individueller Entscheidungen und Handlungen? Konzepte dazu sind in der Soziologie diskutiert worden. Dabei ist sie dafür eigentlich gar nicht prädestiniert.

Was dabei herauskommt, wenn man berechtigter Weise ein den empirischen Erkenntnissen über menschliches Verhalten angemessenes RC-Modell vorschlägt, sieht man beispielsweise bei Boudon (2013). Er hat mit seinem Modell der kognitiven und axiologischen Rationalität versucht, neben den außerindividuellen auch die innerindividuellen Faktoren zu entscheidend zu berücksichtigen. Boudon geht davon aus, dass individuelles Handeln auf individuellen – meist aber sozial geteilten (common sense) – Überzeugungen basiert. Er spricht von einem System {S} wechselseitig miteinander vereinbarer und für sich einzeln akzeptierter Gründe oder Überzeugungen, die ein Akteur in einer Handlungssituation hat. Die Ursachen für individuelles Handeln liegen demnach in dem System akzeptierter Gründe {S}, die ein Akteur in einer bestimmten Handlungssituation hat. {S} ist immer vorläufig und kann revidiert werden, wenn darin enthaltene Überzeugungen neuer empirischer Evidenz und besserem Wissen nicht standhalten (Boudon 2013, 56ff).
Boudon geht soweit, das Konsequentialismus-Postulat aufzugeben. Er erlaubt die Annahme, dass Akteure als rational zu bewertenden Handlungen durchführen können, die für sie ohne Konsequenzen sind. Das ist – nicht nur meines Erachtens – nicht akzeptabel. Meine These ist, dass dieses Postulat auch vermeidbar wäre, wenn sein handlungstheoretisches Modell einem materialistischen Ansatz folgte. Materialistische Soziologie, wie es nennt, lehnt er aber ab, was in der für mich rührenden Frage gipfelt: "Müssen die Geisteswissenschaften das Menschliche wirklich über Bord werfen, um wissenschaftlich zu sein?" (Boudon 2013, XXI).
Boudons Ansatz verbleibt in einer großen Beliebigkeit, so verstehe ich auch die Kritik von Opp (2014).

Das Postulat des Konsequentialismus führt zu den Letztgründen des Handelns, die laut Hume am Ende nicht mehr hinterfragbar, weil "existenziell" gesetzt sind (Hume 1902). Was ist der aber Prozess, der diese Setzung generiert hat: die Evolution, im Zuge derer sich Merkmale der sich erfolgreich reproduzierenden Entitäten im Nachhinein für eine gewisse Zeit als funktional für die hinreichend erfolgreiche Reproduktion erweisen (Dennett 2017). Man ist versucht zu sagen, sie findet in der gegebenen Umwelt durch Versuch und Irrtum eine "rationale" Lösungen für überlebensförderliches Verhalten.

Gintis, formuliert einen Ansatz zu einem Rationalitätsbegriff, der dem klassischen SEU-Modell folgt, aber schon sehr stark von evolutionstheoretischen Überlegungen beeinflusst ist. (Gintis 2016b, 2017). Sein Verständnis des Rationalitätskonzepts scheint dem Biuodonschen gar nicht so fern, wobei er dennoch explizit dem Nutzenmaximierungsprinzip des RC-Modells folgt. Das wird in dem folgenden Zitat deutlich:
"Choice behavior can generally be best modeled using the rational actor model, according to which individuals have a time-, state-, and social context-dependent preference function over outcomes, and beliefs concerning the probability that particular actions lead to particular outcomes. Individuals of course value outcomes besides the material goods and services depicted in economic theory. Moreover, actions may be valued for their own sake" (Gintis 2017: 1). Gintis lässt auch “character virtues, including honesty, loyalty, and trustworthiness, that have intrinsic moral value, in addition to their effect on others or on their own reputation” als Einflussfaktoren zu. Er nimmt an, dass Akteure nicht nur “self-regarding payoffs such as personal income and leisure, but also other-regarding payoffs, such as the welfare of others, environmental integrity, fairness, reciprocity, and conformance with social norms” schätzen können (Gintis 2017: 1). Er verweist schließlich darauf, dass "the social actor’s preference function will generally depend on his current motivational state, his previous experience and future plans, and the social situation that he faces" (Gintis 2017: 1).
Dabei gilt: "Effective choice must be a function of the organism’s state of knowledge, which consists of the information supplied by the sensory inputs that monitor the organism’s internal states and its external environment" (Gintis 2009: 3).

Zum methodologischen Verständnis einer Theorie rationalen Handelns merkt Gintis zwei Prinzipien an. Das erst finde ich überzeugend: „The rational choice model expresses but does not explain individual preferences“ (Gintis 2017, 1). Das eigentlich interessante, und komplexe Erklärungsproblem ist damit verschoben. Um Präferenzen zu verstehen (oder zu erklären) braucht es eine intensive Beschäftigung der "psychology of goal-directed and intentional behavior", evolutionary theory, and Theorien der "problem-solving heuristics". Weniger überzeugend scheint mir die Aussage, dass man das Maximierungsprinzip als "analytical convenience" zu verstehen habe. Gintis nimmt an, dass Menschen nicht bestimmte Letztziele (Fitness oder Wohlfahrtsinteressen) verfolgen - genauso, wie Licht nicht versucht die Laufzeit zu minimieren: "The standard conditions for rationality, for instance, do not imply that rational Alice chooses what is in her best interest or even what gives her pleasure. There are simply no utilitarian or instrumental implications of these axioms. If a rational actor values giving to charity, for instance, this does not imply that he gives to charity in order to

increase his happiness. ... Finally, the rationality assumption does not suggest that Alice is "trying" to maximize utility or anything else." (Gintis 2017: 2). Das könnte man das ebenfalls als die Aufgabe des Postulats des Konsequentialismus interpretieren, was aus meiner Sicht problematisch wäre.

Wir auch immer, es wird gemäß dem Ansatz von Gintis die Handlung gewählt, für die sich im subjektiven Bewertungsprozess situational (gerade) der höchsten erwarteten Nutzen ergibt. Gintis' bayesianisches Rationalitätsmodell basiert auf klassischen Annahmen zur Präferenzfunktion (Eigenschaften der Vollständigkeit, Transitivität, Gelten der IIA), entsprechender Nutzenfunktion und unter Annahme von vier weiteren Axiomen. Es ist also eine Version des Erwartungsnutzenmodells (kopiert aus Gintis 2017: 6, 8), wobei eine Nutzenfunktion auf der Basis einer Präferenzfunktion über Kombinationen von Handlungen a, b, ... und Lotterien π, ρ, ... und einer subjektive priors p (Wahrscheinlichkeitsverteilung über den Zustandsraum) definiert wird (Gintis 2017: 7).

$$\mathbf{E}_\pi[u|a;p] = \sum_{\omega\in\Omega} p(\omega)u(\pi(\omega),a), \qquad (1)$$

then for any $\pi, \rho \in \mathcal{L}$ and any $a, b \in A$,

$$(\pi, a) \succ (\rho, b) \Longleftrightarrow \mathbf{E}_\pi[u|a;p] > \mathbf{E}_\rho[u|b;p]. \qquad (2)$$

Unter wenig restriktiven Annahmen kann er die Existenz von p und u (expected utility theorem) beweisen. Dabei wird wenig empirisch beobachtbares Verhalten als nicht-rational ausgeschlossen, was weiterer Anlass zur Kritik (Immunisierung) sein könnte. Ginits schließt von seinem Rationalitätsbegriff aus, dass die Entscheidung auf "wishful thinking" beruht: "That is, the probability that Alice implicitly attaches to a particular outcome by her preference function over lotteries does not depend on how much she stands to gain or lose should that outcome occur" (Gintis 2017: 3). Außerdem wird ausgeschlossen, dass die Wahl einer Lotterie durch den Akteur nicht den 'state of nature' ändert. Gintis führt weitere Beispiele für irrationales Entscheiden anführen (Allais Paradox, Ellsberg Paradox) und diskutiert Bewertungs- oder Beurteilungsfehler, die aber durchaus im Sinne des Modells interpretiert werden könnten (Gintis 2017: 13ff), auch wenn sie von einem allwissenden Beobachter als nicht rational bewertet werden.

Obschon es mathematisch und formal ausgearbeiteter ist als Modell von Boudon, hängt auch das Modell von Gintis letztendlich in der Luft. Im fehlt, so meine Behauptung, eine materialistische Basis und kommt eher als reines Konstrukt daher, dessen Nutzen man in Zweifel ziehen kann. Eine konsequente und auf Vollständigkeit bedachte Konzeption eines ist meines Erachtens ein eben ein radikaler, materialistischer Ansatz, der auch explizit das Postulat des Konsequentialismus aufrecht erhält: Handeln und (unbewusstes) Verhalten ist Ergebnis eines algorithmisch–gesteuerten Bewertungsprozesses, der das, was als nächstes zu tun ist, nach dem "proximalen" Prinzip des optimalen Wohlbefindens (als Indikator für die

Herstellung und Aufrechterhaltung von Homöostase; Utility) unter veränderlichen inner- und außerindividuellen Nebenbedingungen, bestimmt. Ultimates Ziel ist individuelles Gedeihen und Reproduktion (Fitness).

Interessanter Weise kann man bei Coleman schon lesen. "In a theory of purposeful action, the actor is a kind of homeostatic or goal-seeking entity" (Colemann 1990: 504) - wobei das "or" logisch nicht zwingend ist. Ausführlich befasst sich Damasio damit: "Homeostasis is the powerful, unthought, unspoken imperative, whose discharge implies, for every living organism, small or large, nothing less than enduring and prevailing" (Damasio 2018: 25).

Wo wird das Handeln und Verhalten gesteuert? Im (Wesentlichen im) Gehirn! Einmal recht naiv gedacht, muss es einen neuronal verankerten, wenn auch hoch komplexen Algorithmus geben, der zu Handlungs- oder Verhaltensentscheidungen führt und bestimmt, was als Nächstes zu tun ist. Als selbst-bewusste Wesen schreiben wir "uns" diese Entscheidung, soweit sie bewusst realisiert wird, als willentlich zu.
Der Algorithmus der Handlungs- und Verhaltenssteuerung hat eine durch biologisch-genetische und kulturelle Evolution geprägte hierarchische, funktionale Struktur, die unter schiedlich stabil parametrisiert ist mit:

- neuro-biologisch basierten Parametern (Ebenen 1 und 2 bei Roth; z.B. ingroup-outgroup-Differenz)
- durch Sozialisation und Lebenserfahrung geprägten und „Räsonnement" modifizierbaren Parametern (Ebenen 3 und 4 bei Roth; z.B. internalisierte Normen, Einstellungen, Wissen, Bewusstsein, Reflexionsvermögen)
- Handlungssituation bezogenen Parametern (Opportunitäten, Ressourcen)

Der Algorithmus basiert auf physiologischen und neuronalen Prozessen. Diese Prozesse sind situations–sensibel. Sie nutzen Wahrnehmungsfähigkeiten und ein Erfahrungsgedächtnis und darüber wahrscheinlichkeitsbasierte Annahmen über die Folgen des Handelns in einer jeweiligen Handlungssituation. Annahmen dazu werden in Brückenhypothesen II und III zu formulieren.
Updating des Parameterapparats ist an der Tagesordnung. Im Prozess des Verhaltens, Handelns und Erlebens des Akteurs erweist sich der Algorithmus auf Grund der (mehr oder weniger großen) Veränderbarkeit der Parameter (inkl. der priors) als lernfähig. Die Plastizität geht soweit, dass auch physiologische Veränderungen, die Modifikation genetischer Funktionen (Epigenetik) und – auf sehr lange Sicht – die Veränderung von Genen selbst möglich ist. Diese Anpassungsfähigkeit der Parameterstruktur ist offensichtlich Menschen besonders groß, womit sein evolutionärer Erfolgt begründet wird.
Die „Entscheidung" eines Akteurs als „ausführendes Organ" dazu, was als Nächstes zu tun ist, ist Ergebnis dieses Algorithmus A. Handlungsalternativen ($V_i(t)$) werden durch A auf der Basis der Individuen spezifisch und situationsbezogenen, im Prinzip neurologisch abbildbaren Parametrisierung unter Unsicherheit bewertet ($W(V_i(t))$. Der Ordnung der Bewertungen liegt eine latente „Präferenzdimension", zugrunde, die den geforderten Eigenschaften einer Präferenzfunktion (Vollständigkeit, Transitivität, IAA-Annahme) gehorcht, da der

Algorithmus sonst nicht zu einer eindeutigen Entscheidung käme. Spontane "Fluktuationen" im physiologischen "Entscheidungsprozess sind allerdings nicht ausgeschlossen. Dennoch, dDie Entscheidung erfolgt grundsätzlichgemäß:

–

$$E_A(W_{t+>t}(V_i(t)) = \max_i$$

Das „Ergebnis“ muss für den Akteur aus der Sicht eines vollinformierten Beobachters nicht optimal sein – im Hinblick auf die proximalen Ziele (Wohlbefinden) und weniger noch im Hinblick auf die ultimaten Ziel (Fitness). Diese Aussage ist nicht gleichzusetzen mit der Aussage von Gintis, dass das Optimierungsprinzip als "analytical convenience" zu betrachten sei. Ihm liegen materiell verankerte, neuronale Mechanismen zugrunde, die gleichwohl nur ansatzweise verstanden sind. Das Modell ist deshalb aber konsequentialistisch. Das durch A bestimmte Handeln oder Verhalten ist gemäß der momentan relevanten Parametrisierung von A instrumentell und sogar effizient, beides Grundprinzipien erfolgreicher evolutionärer Entwicklung. Der Algorithmus berücksichtigt im Übrigen auch die Effizient bezogen auf Zeit, die ein durch das Bewusstsein ermöglichtes, reflektierendes Innehalten dauern darf. Für den Fall, dass bei zwei Alternativen gleich bewertet werden, entscheidet der Zufall.

Um ein möglichst erwartungsgemäßes Ergebnis zu erreichen, ist eine möglichst "valide" Wahrnehmung der Rahmenbedingungen und ein der Situation angemessenes Überzeugungssystem wichtig. Das verweist auf die „objektiv-rationale" Bewertung des Handelns oder Verhaltens – vom allwissenden Standpunkt aus betrachtet. Das Gehirn macht aber "Wahrnehmungs- oder Bewertungsfehler" aus der Sicht eines solchen allwissenden Beobachters. Die Problematik gilt besonders für die ultimaten oder "distalen" Ziele des Handelns. Für die Fitness oder langfristige Sicherung von Wohlfahrt ist vor allem eine evolutionär erfolgreiche Parametrisierung mit einer Gewichtung von Bewertungen "zuständig", welche das mit "im Blick haben". Wie wir sehen werden, sind dafür Internalisierung basierte Parameter des Algorithmus A auf der Ebene 3 nach Roth wichtig.

Andererseits, falls ein Akteur an maladaptiven " Überzeugungen festhält, führt sein Handeln zu für den Akteur schädlichen Entscheidungen. Um solche Fehler korrigieren zu können, tritt die Fähigkeit des Gehirns auf den Plan, Bewusstsein und Reflexion zu erzeugen, so Dehaene (2014). Das Bewusstsein erlaubt uns zu denken, Zeit zugewinnen (oder zu verlieren - auch ein Optimierungsproblem), bis zu einem gewissen Grad zu "überschreiben", was uns festverdrahtete, vielleicht genetisch prädisponierte perimetrische Strukturen vorgeben "wollen". dazu kann man auf die vierte Ebene im Schalenmodell von Roth verweisen, um das zu konkretisieren.

Die Erforschung des postulierten Algorithmus liegt außerhalb der Reichweite soziologischer (und wohl auch rein psychologischer) Analyse. Wir sind weit davon entfernt ihn zu verstehen, doch Einiges weiß man schon. Man siehe dazu Beispielsweise Damasio (2018), Dörner

(1999), Pfützner (2014) und viele andere. Hier sei ein Zitat von dem Gehirnforscher Dehaene genannt, welches das Bild von dem Menschen als einer Bayesianischen Maschine aufgreift:

" Indeed, the human brain increasingly fits the bill for a superb Bayesian machine that makes massively parallel inferences and micro-decisions at every stage. Many of us think that our sense of confidence, stability and even conscious awareness may result from such higher-order cerebral "decisions" and will ultimately fall prey to the same mathematical model. Valuation is also a key ingredient ... it demonstrably plays a crucial role in weighing our decisions. Finally, the system is ripe with a prioris, biases and time pressures and other top evaluations that draw it away from strict mathematical optimality" (Dehaene 2012 in edge.org).

Es stellt sich die Frage, ob der Begriff der Rationalität hier noch Sinn macht? Opp hat zu der Frage eine interessante Abhandlung geschrieben (Opp 2018). Vielleicht sollten wir den Begriff nur für die "Allwissenheit" unterstellende, ökonomische RC-Version reservieren.

3. Internalisierung

Das Algorithmus-Modell weist uns den Weg zur Relevanz der Internalisierung von Normen und Verhaltensregeln als Teil der Parametrisierung des Entscheidungsalgorithmus. Sie wird im Zuge der Ontogenese durch die Transmission kulturellen Wissens und andere Lebenserfahrungen (Roths Ebene 3) implementiert und handlungsrelevant: “Humans internalize norms through socialization by parents (vertical transmission), by extraparental conspecifics who control educational and religious practices (oblique transmission), and informal organizations of friends and neighbors (horizontal transmission) (Gintis 2016a: 6).

Das Phänomen der Internalisierung spielt in der Soziologie auch eine große Rolle. Betrachten wir noch einmal Coleman, der sich mit diesem Phänomen befasst hat:
"To examine the process by which norms are internalized is to enter waters that are treacherous for a theory grounded in rational choice!", schreibt er (Coleman 1990: 292). Coleman spricht denn auch von einer "deficiency" der RC-Theorie und gibt indirekt zu, dass man Internalisierung nicht auf einer Basis dieser Theorie gut erklären kann. Internalisierung einer Norm heißt bei ihm: " that an individual comes to have an internal sanctioning system which provides punishment when he carries out an action proscribed by the norm or fails to carry out an action prescribed by the norm" (Coleman 1990: 293).
Das könnte man als eine Kosten-zentrierte Definition bezeichnen. Coleman untersucht dann im Kapitel 11 der "Foundations", warum und wann jemand daran interessiert ist, bei jemandem anderen eine Norm zu internalisierten. Seine Antwort: Es spart Kontrollkosten. Wenn Aufwand eine Internalisierung zu bewirken kleiner ist als diese Ersparnis, lohnt es sich einen solchen Versuch zu unternehmen.
Im Kapitel 19 untersucht er Konstellationen, in denen jemand bereit sein könnte, eine Norm zu internalisieren, also sein von Coleman so genanntes 'object self' zu verändern. Hier

beschäftigt er sich im Wesentlichen mit: 'identification' und der 'balance theory and divestment' oder, interessanter Weise, ‚pico-economics'.
Ich gehe auf Colemans Darstellungen zur Internalisierung hier nicht weiter ein. Es wäre sicher interessant, das zu tun, doch ich will die Frage hier aus der Sicht evolutionstheoretischer Ansätze beleuchten, die mir geboten scheint, und die Coleman außer Acht lässt.

Ein anderer prominenter Soziologe, der im Zusammenhang interessant sein könnte, ist Elias. Seine Soziogenese und Psychogenese, die interdependent miteinander verbunden sind, verbindet dieser Begriff sozusagen (Elias 1977: 312 ff). Sehr grob zusammengefasst kann man sagen, dass das Wachstum und Verdichtung bzw. Ausdifferenzierung von Verflechtungszusammenhängen zunehmende Verhaltenskontrolle qua internalisierter Verhaltensregeln verlangen. Das kann durch Fremdzwänge geschehen, doch, bei Elias ohne Benutzung dieses Begriffs ausformuliert, die Internalisierung von Verhaltensnormen (Selbstzwang) erst ermöglichen zunehmend größere Verflechtungszusammenhänge (Figurationen) und nicht allein das von Elias besonders angeführte Gewaltmonopol als Mittel der inneren Befriedung des Verflechtungszusammenhangs Staat. Letzteres wird als Voraussetzung der zivilisierenden Psychogenese angesehen.
Ein wesentliches Moment der Interdependenz zwischen Fremd- und Selbstzwang wird von Elias nicht eingehender betrachtet. Die Kosten der die Kooperation sichernden Kontrolle wachsen proportional zur Größe der Figuration kooperierender Individuen, sind aber umgekehrt-proportional zum Ausmaß der Internalisierung. So fördert der Selbstzwang das Wachstum von auf Kooperation beruhenden sozialen Strukturen. Das wird bei dem Beispiel, das weiter unten ausgeführt wird eine gewissen Rolle spielen.
Ein weiterer soziologischer Theoriestrang, in dem das Phänomen der Internalisierung eine zentrale Rolle spielt, ist die phänomenologische Soziologie, insbesondere á la Berger und Luckmann (1980). Die Begrifflichkeit ist dort weiter gefasst als diejenige, mit der wir uns hier beschäftigen. Sie meint hier geradezu die Einverleibung von Sinn-Wissen um die Gesellschaft generell, die dem Menschen die Teilhabe an Gesellschaft ermöglicht (139). Daher gehe ich darauf hier nicht weiter darauf ein.

Evolutionstheoretiker argumentieren, dass die menschliche Fähigkeit Normen und Regeln zu internalisieren, nicht zu beobachten wäre, hätte sie sich nicht – zumindest in einem früheren, hinreichen langen Zeitraum als vorteilhaft und Fitness fördernd – man könnte mit Dennett sagen, im Nachhinein als funktional – erwiesen (Dennett 2017). Die Voraussetzung für Internalisierung dürfte die Fähigkeit der kumulativen Kulturproduktion sein. Die Fähigkeit von Normeninternalisierung ist genetisch verankert, ihre Entwicklung wird als "Produkte" einer 'gene-culture coevolution' zu erklären versucht.

Internalisierung einer Norm wird in evolutionstheoretischen Ansätzen nutzenorientiert definiert. Die Befolgung einer internalisierten Norm wird selbst zu einem (belohnenden) Ziel, neben möglichen instrumentellen Vorteilen, die dessen Befolgung haben mag, und trotz nennenswerter Kosten im instrumentellen Sinn. Sie kann, so könnte man hinzufügen, auch ein konsistentes Selbstbild des Menschen in seinem Handeln und verhalten sichern helfen.

Internalisierung bewirkt also eine starke Form von – in der Regel erlernten, ansozialisierten – psychologischen ‚biases' oder Dispositionen und ist nicht Mittel zum Zweck.
Internalisierungen sind also Bestandteil der Parametrisierung des Algorithmus A, die in den neuronalen Strukturen festgelegt sind und unter der Voraussetzung, dass äußere Parameter bestimmte Ausprägungen haben, im Entscheidungsalgorithmus eine Handlung determinierende Bedeutung haben. Sie blockieren bis auf Weiteres keinen Rückverweis auf Reflexionszentren (Neocortex) und führen zum missachten Unsicherheiten für den instrumentellen Handlungserfolg. Der Akteur spart Reflektions- bzw. Entscheidungskosten, kann aber das Risiko einer Fehlentscheidung im Sinne „objektiver" Rationalität erhöhen.
Es muss eine starke „Erschütterung" einer internalisierten Norm geben, damit deren Determinationsstärke beeinträchtigt wird, sie Gegenstand bewusster Reflektion wird und deren Einfluss im Entscheidungsalgorithmus reduziert oder gar eliminiert wird. Letzteres ist für den Fall "starker Überzeugungen" gar nicht so unwahrscheinlich, wenn man Gesetzen der nichtlinearen Systemtheorie folgt (Hysterese-Phänomen).

Eine Erklärung, warum die Fähigkeit zu der Internalisierung entstanden ist, hat der schon erwähnte Gintis einige Überlegungen dargelegt, die dem allgemein anerkannten Prinzip der Gene-Culture-Evolution folgen (Richerson/Boyd 2005). Das heißt insbesondere, dass ohne kumulative Kultur basierend auf entsprechenden kognitiven Fähigkeit Internalisierung nicht denkbar wäre. Nur Imitation wäre unter diesen Umständen wohl möglich und auch im Tierreich beobachtbar. Die genannte Voraussetzung ist insofern einleuchtend, als kognitiven Fähigkeiten zu sozialem Lernen vorhanden sein müssen (Ebene 3 bei Roth). Soll die Befolgung einer Norm als solcher belohnend sein, wäre es vorteilhaft, so könnte man annehmen, wenn vielleicht auch nicht zwingend, wenn Individuen die Vorstellung eines Selbst entwickelt haben.
Aber warum Internalisierung und nicht nur Imitation? Ein egoistisch orientier Akteur müsste nicht die Internalisierungskosten auf sich nehmen und ggf. einen Normenbefolger lediglich imitieren, wenn es für ihn nützlich ist (Gintis 2003). Gintis argumentiert: "But this assumes that agents maximize fitness. In general, how- ever, in any species, individuals do not maximize fitness but rather a objective function that has evolved to reflect biological fitness more or less accurately for a given environment. If the Homo sapiens objective function were perfectly adapted, internalization would not be fitness- enhancing. But the rapid cultural evolution and highly variable environments (...) that characterized the period in which Homo sapiens developed doubtless led to a situation in which the unsocialized human objective function deviated strongly from fitness maximization" (Gintis 2003: 408).
Das gilt, weil die genetische Anpassung zu langsam wäre, um die Akteure in der "Fitness-Spur" zu halten."Imitation (the replicator dynamic) will not correct this failure, because agents copy objective-function-successful, not fitness-successful, strategies. In this situation, there are large fitness payoffs to the development of a non-genetic mechanism for altering the agent's objective function, together with a genetic mechanism for rendering the individual susceptible to such alteration. Internalization of norms, which may be an elaboration upon imprinting and imitation mechanisms in non-human animals, doubtless emerged by virtue of

its ability to alter the human objective function in a direction conducive to higher fitness (Gintis 2003: 417).

Gintis sieht die Fähigkeit zu Internalisierung interdependent mit der evolutionären Entwicklung des Altruismus ("prosocial emotions") verknüpft: "... Since culture is very important in the fitness of humans, internal norms become the proximate cause of complex behaviors in humans, but the ultimate cause of the capacity to internalize, and the content of internal norms themselves, in the first instance remains the same: the capacity to enhance the fitness of the individuals who express them. However, we have seen that there is a second instance in this case: altruistic norms can hitchhike on personally fitness-enhancing norms" (Gintis 2003: 417).

Die Überlegungen von Gintis sind durchaus konventionell und seine – durchaus auch beachtenswerten – formalen Arbeiten zum Thema scheinen mir nicht zum Kern des Problems zu führen. Gavrilets und Richerson (2017) haben eine, wie ich finde, aufschlussreiche, weitergehende Analyse vorgelegt, die die Folgenden etwas ausführlicher vorgestellt werden soll. Sie definieren Internalisierung so : "Certain norms are internalized (i.e., acting according to a norm becomes an end in itself rather than merely a tool in achieving certain goals or avoiding social sanctions)" (2017: 6068). Auch sie unterstellen, dass sich eine genetische Disposition entwickelt haben muss, damit Individuen Normen internalisieren können. Diese Fähigkeit Normen ist ein Ergebnis der Evolution des Menschen, was bedeutet, dass sie zu einem Fitness-Vorteil bei Individuen und in Gruppen geführt haben muss. Dabei hat in der Evolution grundsätzlich der individuelle Fitnessvorteil das letzte Wort. Allerdings – und das ist nicht unumstritten – kann bei der Internalisierung prosozialer Normen auch eine Fitness-begründete Selektion auf der Ebene der sozialen Gruppe wirken. Das behaupten die Anhänger der Multi-Level-Selektion. Mit einer einfachen Gleichung (Price equation) kann angeben, wann Gruppen-Selektion möglich ist (Price 1972). Diese Gleichung ist nicht unumstritten. Umstritten ist, ob die Bedingungen, die zu einer Gruppenselektion führen empirische je nachweisbar sind. Dieses ist, so argumentieren einige, zumindest in der 'cultural evolution' aber sehr wahrscheinlich.
Die zentrale These bezogen auf die Internalisierung prosozialer Normen ist:
„Internalizing a norm has two significant effects upon human behavior: People who have internalized a norm follow it even when doing so is personally costly, and they will tend to criticize or punish norm violators (...). Norm internalization allows individuals to reduce the costs associated with information gathering, processing, and decision making (...) and the costs of monitoring, punishments, or conditional rewards that would otherwise be necessary to ensure cooperation“ (Gavrilets/Richerson 2017: 6068; Bullets von JH)

Das soll in einem Modell abgebildet werden. Es ist ein durch entsprechend theoretisch begründete Gleichungen gesteuertes Simulationsmodell der evolutionären Entwicklung von Internalisierung von Kooperations- und Bestrafungsnormen (prosozialen Normen) in Gruppen unterschiedlicher Größe, das als Agent-Based Model durchgerechnet wird. Gavrilets und Richerson modellieren die Situation des kollektiven Handelns als ‚volunteer-dilemma‘ Spiel:

"The collective action models considered above belong to a general class of the volunteer dilemmas (...), where individuals prefer to free ride on the effort of their groupmates but if nobody else is willing to contribute it may become advantageous to volunteer despite the costs involved" (Gavrilets/Richerson 2017, 6071).

Es werden pro Generation Q = 40 Abfolgen eines kollektiven Handlungsaktes, nachfolgender Bestrafung und anschießender, erneuter Strategiewahl gerechnet und langfristig die Entwicklung von 500 Gruppen, deren Größe sehr klein angesetzt wird (n = 8, 16, 24), verfolgt. Über 10000 Generation wird die Evolution von Internalisierung fortgeschrieben. Es werden zwei Formen kollektiven Handelns unterschieden: 'us-vs.-nature' and 'us-vs.-them' games. Bei dem ersten Typ geht es um die kollektive Bearbeitung/Ausbeutung von Natur und Schutz vor natürlichen Bedrohungen in einer Gruppe; bei den zweiten um kollektives Handeln bei Konflikten oder Konkurrenz mit anderen Gruppen.

Es wird von einer Situation ausgegangen, in die (profitable) Kooperation und insbesondere die Praxis Trittbrettfahrer zu bestrafen schon existiert. Auch die kulturelle Transmission entsprechender Handlungserwartungen existiert. Dann spezifizieren sie folgendes Modell:

Der individuelle Payoff ohne Bestrafung ist:

$$\pi_{CA} = bP - cx,$$

wobei b und c konstante Benefit- und Kostenparameter sind, x gleich 1 bei Kooperation und x=0 bei Defektion. Im Fall des 'us-vs.-nature' game ist $P=X/(X+X_0)$, im Fall des 'us-vs.-them' game ist $P=X/X'$. X ist der gesamt Gruppeneinsatz ($\sum x$) und X_0 ist so gewählt, dass wenn $X=X_0$ ist P=1/2). "The larger X_0, the more group effort is required to secure the reward" (Gavrilets/Richerson 2017, 6072). X' ist das durchschnittliche Einsatzniveau über alle Gruppen.

Kommt die Bestrafung hinzu (y=1 heißt, man bestraft, und y=0 heißt, man bestraft nicht), dann ist der Payoff wie folgt:

$$\pi(x,y) = \pi_{CA} - y[(1-p')\delta + c_{mon}] - (1-x)\kappa q',$$

wobei p' und q' die Häufigkeiten von kooperierenden und bestrafenden Akteuren unter den anderen n-1 Gruppenmitgliedern sind, δ sind die Kosten, n-1 Trittbrettfahrer zu bestrafen, κ sind die Kosten von n-1 Gruppenmitgliedern bestraft zu werden und c_{mon} sind Kosten n-1 Gruppenmitglieder zu überwachen. Man sieht, dass die Kosten für Bestrafen und Monitoring mit der Größe der Gruppe zunehmen.

Die Internalisierungseigenschaft η, $0 \leq \eta \leq 1$ ist genetisch angelegt und bleibt lebenslang konstant, ändert sich aber durch zufällige Mutation bei der Reproduktion. η=0 heißt, es gibt

keine Internalisierung ("undersocialized"), η=1 heißt vollkommene Intenalisierung, das heißt die materiellen Payoffs sind gleichgültig ("oversocialized") . Es gibt alle Werte dazwischen, was bedeutet, dass kein Entweder-Oder angenommen wird – im Unterschied zu X und Y.

Die individuelle Nutzenfunktion unter Berücksichtigung der Internalisierung ist:

$$u_\eta(x,y) = (1- \eta)\, \pi(x,y) + \eta\, (v_1 x + v_2 y),$$

wobei v_1 und v_2 den maximalen Wert unbedingter Normenbefolgung im Hinblick auf Kooperation und Bestrafung angeben. Sie sind nach den Autoren ein exogen vorgegebenes Maß für die soziale Erwartung, eine Norm auszuüben. Das ist zum einen nur eine realistische Annahme, wenn Kooperations- und Bestrafungsnormen schon existieren, worauf auch immer das beruht. Außerdem könnte damit implizit unterstellt sein, dass das Belohnende der Internalisierung doch soziale Anerkennung ist oder Entgehen von Bestrafung durch verweigerte Anerkennung. Das setzt wiederum voraus, dass soziale Anerkennung als belohnend empfunden wird. Diese Modellierung ist meines Erachtens zu unscharf und Zusatztheorien (etwa: die Domestizierung-Theorie: Henrich 2016, Wrangham 2019, Hare/Woods 2020) sind notwendig.

Schließlich gilt für die Fitness w:

$$w = 1 + \pi' - c_{opt}\,(1-\eta) - c_{int}\,\eta,$$

wobei π' der durchschnittliche Payoff über die Q Runden und c_{opt} bzw. c_{int} die Kosten einer Strategiewahl bzw. genetisch/physiologischen Kosten der Internalisierung sind.
Nach jeder Spielrunde wählt jedes Gruppenmitglied neu, ob es kooperiert und bestraft oder nicht gemäß der Nutzenwerte für die vier Alternativen für $u_\eta(x,y)$ und unter den Annahme, dass alle anderen bei ihrer Strategie bleiben. Mit Wahrscheinlichkeit 1-e wählt es dann die Strategie mit dem höchsten Nutzen.

"Group selection is captured by making each group in the new generation independently descend from a group in the previous generation with probability proportional to their average success in collective actions P across Q rounds. Individual selection within each group is implemented by first independently choosing n parents from the group members with probabilities proportional to biological fitness w and then producing n offspring subject to random mutation. Offspring production is followed by random dispersal of half of the offspring (interpreted as females, ref. 42)" (Gavrilets/Richerson 2017, 5).

Zu den Ergebnissen, die in den folgenden Grafiken illustriert sind. Ich beschränke mich auf das 'us-vs.-nature' game.

Abb. 5 (aus Gavrilets und Richerson 2017, 6069)

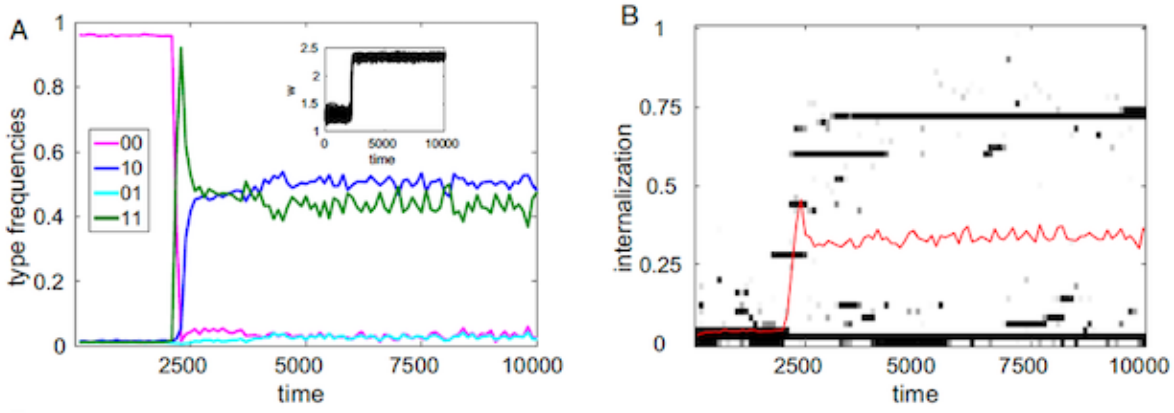


Fig. 1. : Examples of evolutionary dynamics. Us-vs.-nature game with n=16, b=4, v_x =0, v_y =0:5, X_0 =8, δ =0:50, k =4. (A) Frequencies of individuals using different combinations of strategies (x, y). (Inset) The average fitness. (B) The dynamics of the distribution of the internalization trait η.The intensity of the black color is proportional to the number of individuals with the corresponding trait values present at a given time. The red line shows the mean value of η.

Abb. 6 (aus Gavrilets und Richerson 2017, 6070)

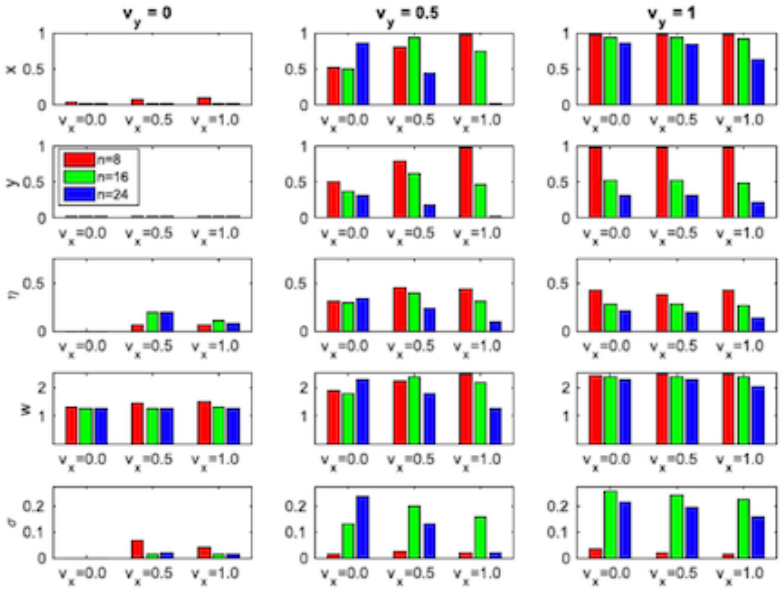


Fig. 3. Summary graphs for us-vs.-nature games: efforts *x*, punishment *y*, internalization η, fitness *w*, and SD σ in internalization trait η for different normative values of production v_x and punishment v_y, and group size *n*. Other parameters: $X_0 = n/2$, $\delta = 0.5$, $b = 4$, $K = 3$. Shown are averages based on 10 runs for each parameter combination.

"Overall, simulations show that if the norm internalization trait η remains small, individuals make no effort (x =y =0). If norm internalization trait η instead evolves to a large value, individuals both contribute to collective action and monitor and punish occasional free riders

(x =y =1). ... If norm internalization evolves, it often emerges quite rapidly after some waiting time, suggesting a transition between alternative quasi-stable states. Occasionally, rapid reverse transitions to low-internalization states are also observed" (Gavrilets/Richerson 2017: 6070).

Differenzierte Befunde zeigt Abbildung 3 für das 'us vs. nature'-Spiel mit unterschiedlichen, vorgegebenen Parametern. Ohne soziale Erwartungen passiert nichts ($v_1 = v_2 = 0$) Es gibt keine Kooperation und keine Internalisierung.
In erster Line relevant ist interessanter Weise dafür der soziale Druck bezüglich der Bestrafung (v_2=0), weniger der soziale Druck zu kooperieren (v_1=0). Allerdings: "Increasing v_y typically increases η, x and y, but increasing v_x can decrease η, x and y, especially in large groups" (Gavrilets/Richerson 2017: 6070).

"If [In us-vs.-nature game, JH] the cost of being punished is small, larger groups can evolve higher η (because the total effect of punishment is higher). If this cost is moderate or large, smaller groups evolve higher η. ... Unexpectedly, simulations often show high genetic variation in η and the emergence of different clusters of individuals with high and low η values similar to those in Fig. 1 (last row of graphs in Figs. 3 ...) (Gavrilets/Richerson 2017: 6070f).
Des Weiteren (in us-vs.-nature games)

- "cooperation and punishment come hand in hand"
- „difficulty of us-vs.-nature games“ fördert „norm internalization“
- relativ kleine Zahl von „oversocialized individuals –‘true believers‘ or ‚heroes‘ – willing to make sacrifices whereas the masses express only a limited norm internalization“
- „effect [on material payoff, JH] ist positiv in 'us-vs.-nature games' – Normeninternalisierung hat weniger Defektion und reduzierte „costs of optimization“ zur Folge.
- große Unterschiede bei unterschiedlichen Parameterkonstellationen.
-

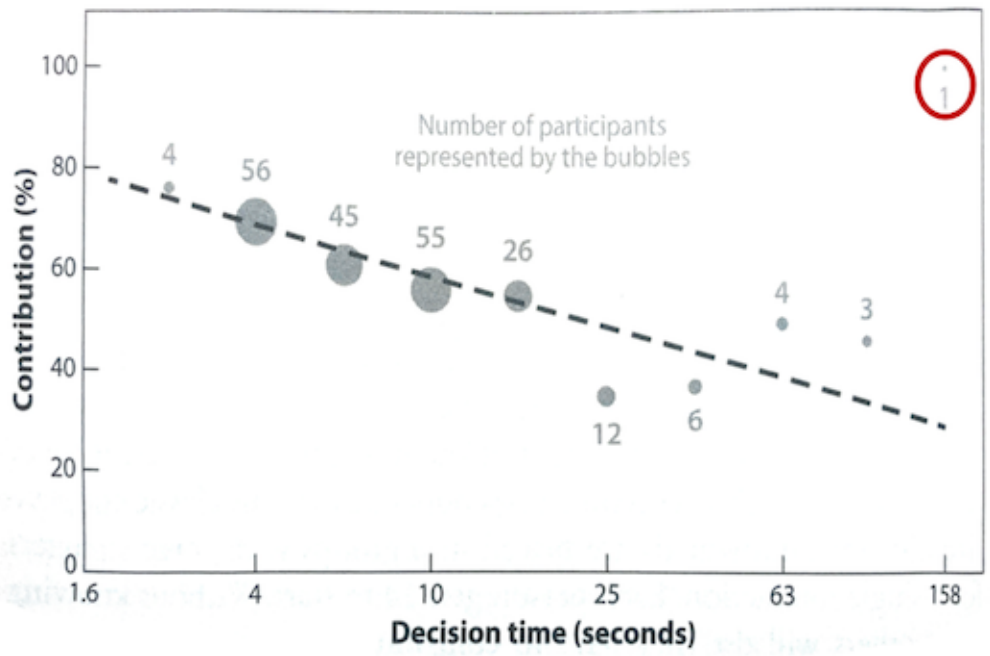
Viele Varianten sind in dem Beitrag nicht genauer ausgeführt – etwas mehr gibt es in einem Anhang zum Beitrag.
Es gibt auch reichlich Anlass zu Kritik.

1. Das Argument der mit der Gruppengröße gesteigerten Kontrollkosten wird nicht hinreichend untersucht. Auch die Monitoringkosten sind m.E. nicht angemessen modelliert.
2. Die Konzeptualisierung der Belohnungswerte der Internalisierung (die v's) ist unklar. Sozialer Druck als Argument ist nicht sehr überzeugend, ich würde eher die Güte der intergenerationalen Transmission hier ins Spiel bringen.
3. Es ist unklar was Werte von η zwischen 0 und 1 bedeuten. Dabei wird auch intragenerationale Dynamik von Internalisierungen nicht abgebildet, die umso höher sein könnte, je kleiner η ist.

4. Die angenommene Gruppengröße ist sehr gering – auch gemessen an den Gruppengrößen zu der Zeit, in der die Evolution stattfand. Wir wissen nicht, was mit der Internalisierung in sehr großen Gruppen passiert. Gavrilets und Richerson meinen jedoch: "Evolving the ability to internalize norms was likely a crucial step on the path to large-scale human cooperation" (Gavrilets und Richerson 2017: 6068). Kooperation wird als "instinctive" verstanden und das ermöglicht eine Vergrößerung der Kooperationsnetzwerke. Auch wird auf die Relevanz von kultureller Gruppenselektion verwiesen.

Dafür, dass die Kooperationsbereitschaft als Norm internalisiert ist, spricht Einiges, wenn man beispielsweise Henrich (2016) folgt. Er verweist auf neurologische Studien, die zeigen, dass die Befolgung von Kooperationsnormen oder die Bestrafung von Nichtkooperation in Gehirn mit einer Aktivität des Belohnungsmechanismus einhergehen - in diesem Sinne also, wenn ich auf die Algorithmus-Überlegungen zurückkomme, folgerichtig ist. Egoistisches Räsonnement kommt dann ins Spiel, wenn bewusste Kognition im Neokortex "anspringt". Ein Studie, die Henrich erwähnt, untersucht, wie die Bereitschaft zur Kooperation in einem 'public-good game' von der Zeit abhängt, die Akteure bis zu ihrer Entscheidung brauchten bzw. davon abhängt, ob sie unter Zeitdruck zu entscheiden sollten oder nicht (Henrich 2016: 194).
Danach scheint eine internalisierte Kooperationsbereitschaft im schnellen Denken verankert zu sein. Reflexion erlaubt sie zu überschreiben oder fehlende Internalisierung begünstigt "rationale" Reflexion. Dazu die folgende Abbildung, die dem Band von Henrich entnommen worden ist.

Abb. 7 (aus Henrich 2016, 194**):** Beteiligung einem 'public-good game' in Abhängigkeit von der Zeit für die Entscheidung.



4. Allgemeinere Schlussbetrachtungen

Nachdem ich hier Beispiel aus der Evolutionsforschung zur Entstehung von Internalisierung und ihren zu einem frühen Zeitpunkt in der Entwicklung von homo sapiens offensichtlich Fitness steigernden Effekt vorgestellt habe, einige abschließende Bemerkungen zu Fragen, die uns heute bewegen. Was ist die Relevanz von Internalisierung für kollektives Handeln heute? Wie wirkt sich der moderne gesellschaftliche Wandel auf Internalisierungsprozesse aus? Auch wenn die Fähigkeit zu internalisieren nicht verloren gehen dürfte, wie gut sind die Transmissionsmechanismen noch, damit Internalisierung erfolgt?

Gintis macht eine Anmerkung , was die sich möglicherweise verändernde Fitnessrelevanz der Internalisierung von Normen angeht: “Social change since the agricultural revolution some 10,000 years ago has been far too swift to permit even the internalization of norms to produce a close fit between utility and fitness. Indeed, with the advent of modern societies, the internalization of norms has been systematically diverted from fitness (expected number of offspring) to welfare (net degree of contentment) maximization” (Gintis 2016: 8).

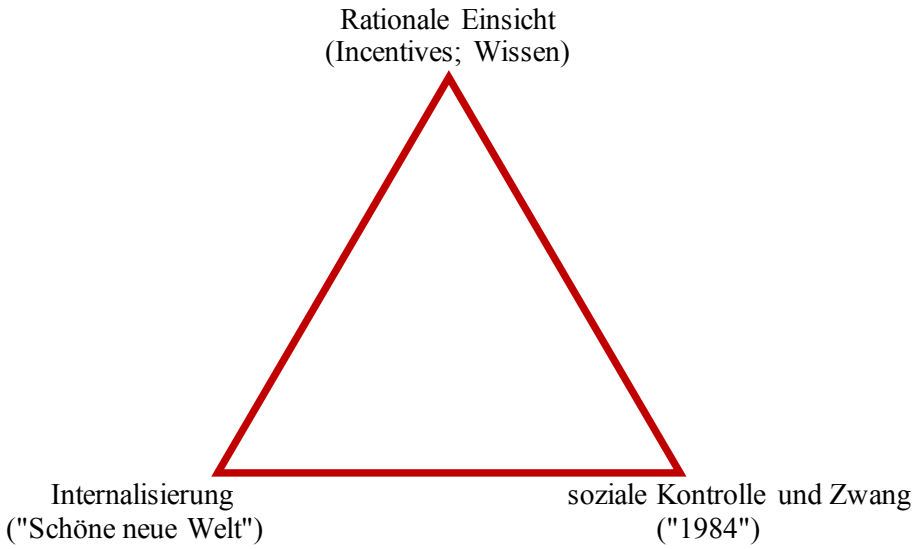
Man kann das auf zweierlei Weise interpretieren. Zum einen folgen die situational relevanten 'objective functions' einer eigensinnigen, pfadabhängigen (kulturellen) Evolution, in der biologische Fitness keine Rolle mehr spielt. Sie könnten sogar zu "fehlgeleiteten", weil nicht prosozialen oder gar antisozialen, Internalisierungen führen. Zum andern, auch wenn Internalisierung von Normen und Verhaltensregeln, die einen Belohnungswert für sich haben, gegen über genetischer Prädisponierung (die auch so unflexibel und situationsunabhängig gar nicht ist) den Vorteil einer größeren Plastizität hat, ist sie in der heutigen Zeit doch zu unflexibel. Das heißt, sie zu verändern, braucht einen inter- oder intragenerational relativ hohen internen oder von außen aufgebauten Wandlungsdruck. Beide Aspekte können zeitweilig zu starken Mismatches oder fehladaptiven Verhalten führen.
So ist die Frage, ob die Plastizität des kulturellen Transformationsmechanismus ausreicht, um in einer sich schnell wandelnden Welt stabile Parametrisierungen in Form von prosozialen Internalisierungen zu erreichen und anzupassen.
Ein weiterer Aspekt betrifft die Frage der Güte der Transmission bzw. der Mechanismen, die Internalisierungen bewirken. Auch hierzu gibt es Forschung in den soziologischen Nachbardisziplinen, etwa zur Frage des sozialen Lernens, die hier zentral ist. Auch die Motivation der Internalisierungsagenten, dürfte wichtig sein – etwas, was Coleman ja schon angesprochen hatte. In den Beispielen, die ich besprochen habe, wird das nicht thematisiert. Wenn etwa für zentral gehaltene prosoziale Normen nicht mehr internalisiert werden, dürfte das in der weiteren Entwicklung des Menschen – biologisch gesprochen – zu Fitnessverlusten führen. Es sei denn,

- Kontrollmechanismen werden so verbessert, dass sie relativ kostengünstig ein adäquates, d.h. vollständiges Monitoring der Mitglieder der Gesellschaft erlauben. Das wird zurzeit in China ja versucht.

- rationale Einsicht (und Weitsicht), also die vierte Ebene bei Roth, gewinnt die Oberhand, was wiederum auf zwei Weisen erfolgen kann:
  - durch persönliche (selektive) Incentives, welche die bei Internalisierung fälligen "Eigenbelohnungen" ersetzen oder sogar übersteigen. Das wäre die 'object function'-Lösung, die eher kurzfristige Effekte haben dürfte.
  - durch "belastbare" vertrauenswürdige Informationen, das heißt Vermittlung von Wissen um Wirkungsmechanismen, die den "Schatten der Zukunft" von kollektivem Handeln, also dessen kurz-, mittel- und langfristige Kosten-Nutzen-Bilanz – oder auch Fitness-Bilanz – als vorteilhaft erweisen.

Das führt mich zu dem einfachen Diagramm in Abb. 8

Abb. 8: Varianten der Sicherung sozialer Kooperation



Und man kann nun die Frage stellen, wohin sich unsere Gesellschaften in dem Spannungsfeld dieser drei Varianten bewegen wird, die nicht im Widerspruch zueinanderstehen und heute schon immer in Kombination vorkommen.
Huxley präsentierte in seinem Buch "Schöne neue Welt" eine (Wohlfühl-) Gesellschaft, in der (außer durch Soma) eine umfassende Internalisierung – je nach Kategorie unterschiedlicher – Verhaltensnormen einen unbedingten Konformismus im individuellen Handeln garantieren soll. Orwell dagegen setzt in "1984" auf totale externe Kontrolle und den Zwang zu normenkonformen Verhalten. In beiden Fällen handelt es sich allerdings um durch Machteliten erzwungene Ordnungen, die sich nur unterschiedlicher Mittel bedienen, diese auch durchzusetzen. Schauen wir heutige Gesellschaften an, so finden wir eine Mischung von beidem, wobei das Schwergewicht mal eher auf dem einen mal eher auf dem anderen liegt.

Man kann diese Unterscheidungen sozial- wie gesellschaftstheoretisch intensiv diskutieren. Das will ich hier nicht tun, sondern meinem Anliegen gemäß deutlich machen, dass wir soziale Kooperation in einem kompletteren Sinne erklären und deren Existenzbedingungen

nur erfassen können, wenn evolutionstheoretische Forschung zentral einbezogen wird. Man könnte aus den bisherigen Befunden der soziobiologischen und anthropologischen Evolutionsforschung auch Schlussfolgerungen dazu ziehen, was für die heutige Pandemie-Zeiten zu erwarten war und sein wird. Daraus uns aus weiteren Erkenntnissen lassen sich einige, zum Teil durchaus problematische Schlussfolgerungen ziehen:

1. Vornehmlich auf die Eigenverantwortlichkeit von Menschen zu zählen, wenn das erwartete Handeln möglicherweise internalisierten Regeln oder Gewohnheiten widerspricht, ist problematisch. es bedarf daher glaubwürdige Sanktionsdrohungen bei "Nichtkooperation". Appelle reichen nicht.
2. Rationale Einsicht braucht in der Pandemie glaubwürdige Informationen sowie Zeit und soziale Interaktion diese zu reflektieren ('bayesianisches updating').
3. Da 'non-reciprocial altruism', wenn es ihn den gibt, vornehmlich parochial orientiert ist, ist zu erwarten, dass sich soziale 'ingroup-outgroup'-Konflikte verschärfen. Unbearbeitet drohen sie "überschriebene" Gruppenresentiments zu (re)aktivieren. Soziale Ungleichheit spielt dabei eine große Rolle.
4. Es gibt die übersozialisierten Pandemie-Gegner, die schon gar nicht mit vernünftigen Argumenten zu überzeugen sind. Sie sind "altruistisch" motiviert Solidargemeinschaften organisieren, innerhalb derer – im Protest gegen die Corona-Maßnahmen – eine Quelle der Befriedigung von Bedürfnissen nach gegenseitiger Unterstützung und Anerkennung sowie narzistischer Selbstdarstellung geschaffen wird.

Alle Aspekte sind in der einen oder anderen Form den aktuellen Entwicklungen der Pandemie zu erkennen. Die Empirie bestätigt offensichtlich Einiges. Meines Erachtens wir das oft übersehen, wenn die Begleitkosten einer restriktiven Politik, deren Größenordnung und langfristigen Folgen noch genau zu untersuchen wären (Bildungsausfall, Psychische Traumata etc.), im Vordergrund der Argumentation stehen. Es tut aber auch Not, die kulturabhängig unterschiedlich ausgeprägten (durch kulturelle Evolution entstandenen), "neuralgischen Stellen" in der sozialen und individuellen Krisenbewältigung zu identifizieren und eine angemessene professionelle Unterstützung zu bieten,

Zu all diesen Fragen haben die Soziologie meines Erachtens viel zu sagen. Weil sie aber nicht den anthropologischen Hintergrund berücksichtigt, bleibt ihr Beitrag substantiell begrenzt und essayistisch.

Aus diesen Überlegungen ergeben sich aber eine Reihe von Fragen an die Soziologie, die wesentliche Beiträge zu dieser Forschung liefern kann, da sie die Interdependenz zwischen sozio-kulturellem Kontext und individuellen Handlungs- und Verhaltensprozessen einerseits und die Folgen dieser Prozesse für den sozio-kulturellem Kontext andererseits zum Thema hat. Wo ist aber, wenn man die Erforschung von Internalisierung und deren potentiellen Abschwächung im Zuge der postmodernen Gesellschaften denkt, die soziologische Sozialisationsforschung. Wo ist die Forschung zum sozialen Lernen, die für soziobiologische, neurologische und evolutionär anthropologische und 'cultural evolution'-Forschung offen ist? Was kann man zum Beispiel diesbezüglich auf den konkreten Fall der Bereitschaft, sich an die Corona-Regeln zu halten, belastbar aussagen? Wo ist die ebenfalls durch

soziobiologische, neurologische und evolutionär anthropologische und 'cultural evolution'-Forschung informierte soziologische Forschung zu den Bedingungen der Herstellung einer sozialen Struktur, die die Kooperationsdisposition stützt und Prozesse der Dehumanisierung vermeidet. Welche Rolle spielen dabei die Definitionen von Ingroup und Outgroup und welcher Dynamik unterliegen diese Definitionen warum?

Im Soziologie-Studium bräuchte es ein Propädeutikum zur evolutionären Humanbiologie bzw. Anthropologie und zur Soziopsychologie menschlichen Verhaltens. Es gibt hervorragende Literatur dafür (z.B. Sapolsky 2017).

Literatur

Berger, Peter L. und Thomas Luckmann. 1980. Die gesellschaftliche Konstruktion der Wirklichkeit. Eine Theorie der Wissenssoziologie. Fischer.

Bernardi, Laura, Johannes Huinink und Rick Settersten. 2019. The life course cube: A tool for studying lives. In: Advances in Life Course Research, 41 (Special Volume), Article 100258, Seite 1-13.

Boudon, Raymond. 2013. Beiträge zur allgemeinen Theorie der Rationalität. Mohr Siebeck.

Coleman, James S. 1990. Foundations of Social Theory. Harvard University Press.

Damasio, Antonio. 2018. The Strange Order of Things: Life, Feeling, and the Making of Cultures. Pantheon.

Dehaene, Stanislas. 2012**.** The Universal Algorithm For Human Decisions. https://www.edge.org/response-detail/10260

Dehaene, Stanislas. 2014. Consciousness and the brain: deciphering how the brain codes our thoughts, Penguin.

Dennett, C. 2016. Den Bann brechen: Religion als natürliches Phänomen. Suhrkamp.

Dennett, C. 2017. From From Bacteria to Bach and Back. The Evolution of Minds. Norton.

Dörner, Dietrich. (1999). Bauplan für eine Seele. Rowohlt.

Elias, Norbert. 1977. Über den Prozess der Zivilisation. Soziogenetische und psychogenetische Untersuchungen. Zweiter Band. Wandlungen der Gesellschaft. Entwurf zu einer Theorie der Zivilisation. Suhrkamp.

Fazio, R. H., und T. Towles-Schwen. 1999. The MODE model of attitude-behavior processes. In: S. Chaiken & Y. Trope (Eds.), Dual-process theories in social psychology. The Guilford Press, 97–116

Frey, Ulrich J. 2017. Assessing the generalizability of behavior and motives across different settings. Evolution and Human Behavior, 38 (3) 281-282.

Gavrilets, Sergey und Peter J. Richerson. 2017 Collective Action and the Evolution of Social Norm Internalization. PNAS, 114(23):6068-6073.

Gintis, Herbert . 2003. The Hitchhiker's Guide to Altruism: Gene-culture Coevolution, and the Internalization of Norms. Journal of theoretical Biology (2003) 220, 407–418.

Gintis, Herbert. 2009. The Bounds of Reason: Game Theory and the Unification of the

Behavioral Sciences. Princeton University Press.

Gintis, Herbert. 2016a. The Genetic Side of Gene-Culture Coevolution: Internalization of Norms and Prosocial Emotions. Manuskript.

Gintis, Herbert. 2016b. Individuality and Entanglement. The Moral and Material Basis of Social Life. Princeton University Press.

Gintis, Herbert . 2017. Rational Choice Explained and Defended. Manuscript.

Hare, Brian und Vanessa Woods 2020. Survival of the Friendliest: Understanding Our Origins and Rediscovering Our Common Humanity. Random House.

Henrich, Joseph 2016. The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter. Princeton University Press.

Huinink, Johannes. 1989. Mehrebenensystem-Modelle in den Sozialwissenschaften. Deutscher Universitätsverlag.

Hume, David. 1902. Enquiries Concerning the Human Understanding and Concerning the Principles of Morals, edited by L. A. Selby-Bigge, 2nd ed. Clarendon Press.

Kahneman, Daniel. 2011. Thinking, Fast and Slow. Penguin**.**

McCullough, Michael E. 2020. The Kindness of Strangers: How a Selfish Ape Invented a New Moral Code. Basis Books.

McAuliffe, William H.B. und Michael E. McCullough. 2017. Validation is a Galilean enterprise. Evolution and Human Behavior, 38 (3) 279–280.

Opp, Karl-Dieter. 2014. The Explanation of Everything. A Critical Assessment of Raymond Boudon's Theory Explaining Descriptive and Normative Beliefs, Attitudes, Preferences and Behavior. Revista de Sociologia, 99(4), 481-514.

Opp, Karl-Dieter. 2018. Do the Social Sciences Need the Concept of "Rationality"? Notes on the Obsession with a Concept. In: Di Iorio, Francesco und Gérald Bronner (eds.) The Mystery of Rationality. Mind, Beliefs and the Social Sciences. VS Springer, 191-217.

Pfützner, Helmut. 2014. Bewusstsein und optimierter Wille. Springer.

Price, George R. 1970. Selection and Covariance. Nature 227, 20-521.

Richerson, Peter J. und Robert Boyd. 2005. Not by Genes Alone: How Culture Transformed Human Evolution. University of Chicago Press.

Roth, Gerhard. 2015. Persönlichkeit, Entscheidung und Verhalten. Warum es so schwierig ist, sich und andere zu ändern. Klett-Cotta.

Sapolsky, Robert M. 2017. Behavior. The Biology of Humans at Our Best and Worst. Penguin.

Searle, John R. 2001. Free Will as a Problem in Neurobiology.Philosophy 76)298), 491-514.

Strüber, Nicole und Gerhard Roth. 2020. Entwicklungsneurobiologie. In: Roth, Gerhard; Andreas Heinz und Henrik Walter (Hrsg.) Psychoneurowissenschaften, Springer, 119-146.

Trivers, Robert L. 1971. The Evolution of Reciprocal Altruism. The Quarterly Review of Biology, 46,(1) pp. 35-57.

Vaisay, Stephen, 2009. Motivation and Justification: A Dual-Process Model of Culture in Action. American Journal of Sociology, 114 (6), 1675–1715

Wrangham, Richard. 2019. The Goodness Paradox: The Strange Relationship Between Virtue and Violence in Human Evolution. Random House.