

Goldman über Selbsttäuschung, Gedächtnis und Irrationalität

Matthias Schaedler

1. Die leitende Frage meiner nachfolgenden Betrachtungen, die vorwiegend kritisch sein werden, lautet: Wie kommt Goldman im Rahmen seiner naturalistischen Konzeption von Erkenntnistheorie mit dem Phänomen der Selbsttäuschung zurecht? Insonderheit: Wie fruchtbar ist seine Unterscheidung zwischen aktivierten und nicht-aktivierten Überzeugungen für eine angemessene Beschreibung und Erklärung dieses Phänomenbereichs?

2. Die genannte Unterscheidung steht in unmittelbarem Zusammenhang mit seiner grundlegenden meta-epistemologischen These, die das Kernstück seines Projekts einer naturalisierten Erkenntnistheorie bildet und besagt, dass es einen engen systematischen Zusammenhang zwischen Erkenntnistheorie und empirischen Theorien der Kognition gibt (vgl. insb. Goldman, 1986). Im uns interessierenden Zusammenhang ist es die These, wonach empirisch ausgerichtete Untersuchungen zum Gedächtnis eine Revision des gewöhnlichen, in der Alltagspsychologie verankerten Überzeugungsbegriffs verlangen und dass die Erkenntnistheorie bei der Verfeinerung ihrer deskriptiven Ressourcen zwecks psychologisch adäquater Begrifflichkeit auf die Hilfe der Kognitionswissenschaften angewiesen ist (vgl. Goldman, 1986: S. 199–200).

3. Im Anschluss an Andersons *The Architecture of Cognition*, worin aktive Erinnerungen (*active memories*) als aktivierte Zustände des sog. Langzeitgedächtnisses aufgefasst werden, unterscheidet Goldman – eingeständnermaßen vereinfachend – zwischen aktivierten und nicht-aktivierten Zuständen, die er mit zwei Arten von Glaubenszuständen identifiziert: *aktivierten* und *nicht-aktivierten Überzeugungen* (vgl. Goldman, 1986: S. 200–1). Die Bedeutsamkeit dieser Unterscheidung zeigt sich ihm zufolge in doppelter Weise: Zum einen im Hinblick auf deskriptive Zwecke, d. h. im Hinblick auf die Erklärung und Voraussage von (menschlichem) Verhalten, zum anderen im Hinblick auf die normative Dimension epistemischer Prinzipien (vgl. Goldman, 1986: S. 202).

4. Im 1. Teil meiner Untersuchung werde ich mich kritisch mit der deskriptiven Relevanz dieser Unterscheidung auseinandersetzen, im 2. Teil mit deren normativer Relevanz.

1. Selbsttäuschung und Gedächtnis

1. Die deskriptive Relevanz erläutert Goldman anhand zweier Beispiele. Im ersten – auf das ich mich hier beschränke – geht es um eine Person, die sich über die Öffnungszeiten einer Bibliothek täuscht, obwohl sie eigentlich über die nötigen und korrekten Informationen verfügt.

2. Eines Sonntagmorgens ist Melanie am Überlegen, ob sie in die Bibliothek arbeiten gehen soll. Sie arbeitet am Sonntag zwar selten dort, aber heute muss sie unbedingt einen für sie wichtigen Artikel bearbeiten. Nun weiss sie, dass die Bibliothek normalerweise um 7 Uhr öffnet, und ihr ist auch bei verschiedenen Gelegenheiten aufgefallen, dass sie sonntags erst um 13 Uhr offen ist. Würde sie sich dessen erinnern, ginge sie sicher diesen Morgen nicht zur Bibliothek, denn dies wäre verlorene Zeit und Mühe. Doch gerade heute vergisst sie, dass für Sonntag andere Öffnungszeiten gelten und entscheidet sich zu gehen. Als bald steht sie vor geschlossenen Türen und tadelt sich selbst, indem sie zu sich selbst sagt: „War das nicht dumm von mir; ich wusste doch, dass die Bibliothek sonntags erst um 13 Uhr öffnet.“ (Vgl. Goldman, 1986: S. 202–3.)

3. Nach Goldmans Auffassung würde ein Modell, das menschliche Handlungen nur auf der Basis von einer Art von Überzeugungen bzw. Wünschen und den entsprechenden Überzeugungs/Wunsch-Paaren erklärt, (wahrscheinlich) voraussagen, dass sich Melanie dafür entscheidet, *nicht* zu gehen, denn – so sein Argument – die Gesamtheit (der Korpus) ihrer Überzeugungen schliesst die Überzeugungen mitein, dass heute Sonntag ist und dass sonntags die Bibliothek erst um 13 Uhr öffnet. Ein Modell hingegen, so seine These, das den Unterscheid zwischen aktivierten und nicht-aktivierten Überzeugungen macht und dabei in Rechnung stellt, dass nur die aktivierten Überzeugungen handlungs- bzw. entscheidungswirksam sind, könnte Melanies Verhalten korrekt voraussagen. Die Erklärung, die er gibt, ist denkbar einfach: Im Moment der Entscheidung versäumt es Melanie, sich entweder die Überzeugung, dass heute Sonntag ist oder aber die Überzeugung, dass die Bibliothek sonntags erst um 13 Uhr öffnet, bewusst zu machen. Die entscheidungswirksame Überzeugung ist vielmehr die momentan aktivierte Überzeugung, dass die Bibliothek um 7 Uhr öffnet. Er schreibt:

It is a familiar occurrence that people often do what Melanie does – they fail to *access* part of what they know, and therefore perform silly actions. Failure to distinguish U-beliefs [unactivated beliefs, M. S.] and A-beliefs [activated beliefs, M. S.] makes it impossible to account for these occurrences. (Goldman, 1986: S. 203)

Welche Entscheidungen eine Person fällt oder welche Schlüsse sie zieht (seien es praktische oder theoretische), so hält er abschliessend fest, ist nicht durch die Überzeugungen determiniert, die im Langzeitgedächtnis gespeichert sind, sondern allein durch jene, die zum fraglichen Zeitpunkt aktiviert sind (ebd.).

4. Eine wichtige Präzisierung sei hier noch angefügt. Für Goldman heisst ‚vergessen‘ nicht, Informationen unwiederruflich zu verlieren, sondern lediglich, dass man den *Zugang* zu der im Langzeitgedächtnis gespeicherten Information verloren hat (vgl. Goldman, 1986: S. 223). Dabei erachtet er zwei Faktoren als massgebend: mit der Zeit schwächer werdende Gedächtnisspuren (*decay of information*) bzw. deren Überlagerung durch andere Informationen (*interference of information*) (vgl. Goldman, 1986: S. 222–5 sowie Goldman, 1993: S. 11–12). Entsprechend unterscheidet er zwischen zwei Formen des Vergessens. Die erste Form ist ihm zufolge schlicht ein Fehlschlag, den Zugang zu einer einmal gespeicherten Information wiederherzustellen. In diese Form des Vergessens, so betont er, ist nicht das Hervorbringen einer *falschen* aktivierten Überzeugung involviert, sondern vielmehr das *Ausbleiben* einer solchen Überzeugung. Unter der zweiten Form des Vergessens versteht er hingegen eine *fehlerhafte* Wiederherstellung, die er sich anhand sog. Interferenzeffekte erklärt.

5. Goldmans Erklärungsansatz wirft eine Reihe von Fragen auf, die ich hier nicht im Detail behandeln kann. So wäre etwa zu fragen, was es heisst zu sagen, menschliches Verhalten sei durch die zu einem bestimmten Zeitpunkt aktivierten Überzeugungen *determiniert*. Ist mit ‚determiniert‘ hier bloss *Abhängigkeit* gemeint oder die stärkere Idee verknüpft, unser Verhalten sei durch die fraglichen Überzeugungen *eindeutig bestimmt*? Damit zusammenhängend wäre zweitens zu fragen, ob unser Verhalten durch *alle* aktivierten Überzeugungen bestimmt ist oder nur durch die für das jeweilige Verhalten *relevanten* Überzeugungen. Falls letzteres zutrifft, erhebt sich die weitere Frage, wie dann die relevanten herausgefiltert werden können. Zu prüfen wäre ferner, was genau es heissen soll, den gewöhnlichen Begriff der Überzeugung (in deskriptiver Hinsicht) zu verfeinern. Goldman hält, wie wir gesehen haben, diesen Begriff für psychologisch inadäquat und schlägt daher vor, zwischen aktivierten und nicht-aktivierten Überzeugungen zu unterscheiden. Das ‚eigentliche‘ Motiv für die vorgenommene Unterscheidung ist das Bestreben, intentionale Phänomene zu *naturalisieren*, und zwar entlang der Unterscheidung zwischen Glaubenszuständen, die kausal wirksam sind, und solchen, die es nicht sind (vgl. Goldman, 1986: S. 207). Zugestanden, ein kausal wirksamer und in diesem Sinne aktivierter Glaubenszustand ist ein intentionales Phänomen. Aber hat ein solcher Zustand überhaupt noch etwas mit den unzähligen, langandauernden intentionalen Phänomenen zu tun, die mit dem gewöhnlichen Begriff der Überzeugung erfasst werden?

Müsste für ihn nicht zumindest gelten, sein Inhalt werde von *S* zu *t* geglaubt, was nur heissen würde, dass solche Zustände Überzeugungen voraussetzen (vgl. Kemmerling, 1998)?

6. Goldman behauptet, dass nur das von ihm favorisierte Modell mit Fällen wie dem im Beispiel beschriebenen zu Rande kommen kann. Ich lasse hier ausser Acht, ob dieser Ausschliesslichkeitsanspruch zu Recht erhoben wird. Mich interessiert vielmehr die Frage, ob sein Modell mit Blick auf das Phänomen der Selbsttäuschung wirklich funktioniert.

7. Ausgehend von diesem Problemhorizont haben wir uns zunächst zu fragen, ob es sich bei dem im Beispiel zur Sprache gebrachten Fall überhaupt um einen ‚echten‘ Fall von Selbsttäuschung handelt oder nicht. In der einschlägigen Literatur und namentlich in der im angelsächsischen Bereich geführten Diskussion zum Phänomen der Selbsttäuschung herrscht alles andere als Einigkeit, was als echter Fall von Selbsttäuschung gilt. Eine für meine Zwecke nützliche Sichtweise ist die, zwischen Selbsttäuschung in einem *schwachen* und in einem *starken Sinne* zu unterscheiden. Gemäss Martin, auf den ich mich bei dieser Unterscheidung stütze, handelt es sich bei Selbsttäuschung in einem schwachen Sinne – er spricht von *nonwillful self-deception* – um ein Phänomen, bei dem Unachtsamkeit und möglicherweise vermeidbare Unwissenheit eine zentrale Rolle spielen (vgl. Martin, 1985: S. 1). Selbsttäuschung in einem starken Sinne – *willful self-deception*, wie er sie auch nennt – setzt demgegenüber einen aktiveren, zweckvoll und intentional agierenden Täuscher voraus, der seine Emotionen und propositionalen Einstellungen manipuliert, um sich selbst etwas glauben zu machen, von dem er ‚tief drinnen‘ weiss, dass es falsch ist (ebd.).¹

8. Wenn nun Goldmans Unterscheidung zwischen aktivierten und nicht-aktivierten Überzeugungen und sein Konzept der Zugänglichkeit mit Blick auf Selbsttäuschung in einem umfassenden Sinne tragfähig sein sollen, dann müssten sie dies auch bezüglich Fällen von Selbsttäuschung im starken Sinne sein. Betrachten wir zu diesem Zweck das folgende Beispiel, das ich Davidson entlehnt habe (vgl. Davidson, 1985: S. 145–6):

9. Carlos hat gute Gründe zu glauben, dass er die Fahrprüfung, auf die er sich seit Wochen intensiv vorbereitet, nicht bestehen wird. Nicht nur hat er bisher schon zwei Mal bei der Prüfung versagt, sondern sein jetziger Fahrlehrer hat ihn auch mehrmals darauf hingewiesen, dass seine Chancen, die Prüfung zu bestehen, äusserst gering sind. Kurz: er ist sich bewusst, dass alles – d.h. die Gesamtheit der Belege – auf ein erneutes Scheitern hindeutet. Der Gedanke, erneut zu scheitern, ist aber für Carlos absolut unerträglich. Daher hat er gleichsam ein natürliches Motiv zu glauben, er werde re-

üssieren. Das heisst, er hat ein Motiv, sich selbst glauben zu machen, er werde die Prüfung bestehen. Carlos tut nun alles Mögliche, sich in diesem Glauben zu stärken; sei es, dass er die ‚negativen‘ Belege in den Hintergrund zu drängen versucht, sei es, dass er entweder nach neuen, ‚positiven‘ Belegen sucht oder die alten akzentuiert (etwa, dass er den Examinator persönlich kennt oder dass er auf seinen Charme setzt).

10. Was den eben beschriebenen Fall von Selbsttäuschung zu einer echten Herausforderung macht, ist zu erklären, wie es möglich ist, dass eine Person – in unserem Fall Carlos – gleichzeitig zwei sich widersprechende Überzeugungen hegen kann. Um diesen Fall angemessen zu beschreiben, scheinen wir nämlich folgendes sagen zu müssen: Carlos glaubt, dass er die Prüfung nicht bestehen wird, und glaubt zugleich, dass er sie bestehen wird.

11. Bemerkenswert ist zunächst, dass in diesem Fall das primäre Interesse nicht einer Voraussage von Carlos Verhalten gilt. Eine solche dürfte überaus schwierig sein. Das primäre Interesse ist vielmehr auf die Frage zu richten, wie ein solcher Fall zu *verstehen* ist.² Kurz: ich möchte prüfen, wie Goldman mit einem solchen Fall zurecht kommen *könnte*. Kann hier seine Unterscheidung zwischen aktivierten und nicht-aktivierten Überzeugungen Abhilfe schaffen?

12. Die entscheidende Aufgaben besteht nach meiner Auffassung darin zu klären, wie Carlos die zwei sich widersprechenden Überzeugungen zugleich hegen kann und wie er es zugleich versäumen kann, die beiden inkonsistenten Überzeugungen ‚zusammenzubringen‘. Die Aufgabe so zu formulieren, beinhaltet die folgenden wichtigen Präzisierungen: (1) Es geht um die Erklärung von *synchroner*, nicht aber *diachroner* Inkonsistenz. (2) Carlos werden zwei miteinander unverträgliche Überzeugungen zugeschrieben, nämlich *p* und nicht-*p*; nicht aber die Überzeugung [dass *p* und nicht-*p*].³

13. Aus der Sicht Goldmans scheint es auf die zu klärende Frage eine einfache Antwort zu geben. Carlos versäumt es die beiden Überzeugungen zur gleichen Zeit zu aktivieren, sie nebeneinander zu stellen, was natürlicherweise zur Folge hätte, dass die eine oder die andere Überzeugung ‚getilgt‘ würde. Ihre Koexistenz verdankt sich demnach einfach dem Umstand, dass die eine Überzeugung schlummert, während die andere aktiviert ist, und umgekehrt (vgl. Goldman, 1986: S. 204). Diese Erklärung mag im Fall von Melanie greifen; im Fall von Carlos scheint sie mir jedoch nicht zu funktionieren. Der Grund ist der, dass wir eine Erklärung für synchrone Inkonsistenz suchen und nicht eine für diachrone. Im Fall synchroner Inkonsistenz müssen wir aber gerade annehmen, dass die Überzeugungen, Motive,

Wünsche und Prinzipien, die den Konflikt erzeugen, zu ein und demselben Zeitpunkt aktive Kräfte sind (vgl. Davidson, 1985a: S. 353).

14. Erinnern wir uns: Gemäss Goldmans Auffassung enthält eine nicht-aktivierte Überzeugung Informationen, zu der die fragliche Person den Zugang verloren hat. Könnte es sein, dass im Begriff der Zugänglichkeit ein Schlüssel liegt, um zu einer befriedigenderen Klärung der Frage zu gelangen, wie es möglich sein soll, dass Carlos die beiden sich widersprechenden Überzeugungen *zugleich* hegt? Goldman selbst hat diesen Lösungsansatz, so weit ich sehe, nicht explizit weiterverfolgt. Von McLaughlin stammt jedoch ein Vorschlag, der in meinen Augen hinreichend deutlich macht, in welche Richtung Goldmans Erklärungsansatz weitergedacht werden könnte (vgl. McLaughlin, 1988: S. 48–51).

15. McLaughlin zufolge ist es *unmöglich*, zur selben Zeit zu glauben oder zu denken, dass p und dass nicht- p , sofern man sich beider Überzeugungen bzw. Gedanken ‚bewusst‘ ist. Daher schlägt er vor, zwischen zwei Arten von Überzeugungen zu unterscheiden: zwischen *zugänglichen* und *nicht-zugänglichen* Überzeugungen. Seine These lautet nun wie folgt:

It is impossible to hold accessible contradictory beliefs simultaneously. [...] However, it is possible [...] simultaneously to hold contradictory beliefs if at least one of the beliefs is inaccessible. One can in this case believe that p and at the same time believe that not- p , without the beliefs' clashing in conscious thought when one thinks of p . (McLaughlin, 1988: S. 50)

16. An dieser These scheint mir folgendes richtig zu sein: Hätte Carlos vollen oder integralen Zugang zu all seinen geistigen Operationen bzw. Zuständen, dann wäre kaum zu erklären, wie er es versäumen kann, eine der sich widerstreitenden Überzeugungen zu tilgen. Nach einem solchen Modell wäre Selbsttäuschung *in einem starken Sinn* ein Ding der Unmöglichkeit. Auf der anderen Seite steht die These im Widerstreit mit den folgenden Überlegungen: Carlos mag sich noch so bemühen, neue Belege zu suchen, die ihn in der Überzeugung stützen könnten, dass er die Prüfung bestehen wird. Aber dies ändert nichts daran, dass er sich zugleich jener Belege bewusst bleibt, ja bleiben muss, die seine Überzeugung stützen, dass er die Prüfung nicht bestehen wird. Es ist ja gerade das ‚Bewusstsein‘ ebendieser Sachlage, das seine Anstrengungen in Gang bringt, sich von der Angst, er werde versagen, zu befreien (vgl. Davidson, 1985: S. 146). Diese Überlegung lässt sich auch mit Blick auf das Phänomen des Vergessens verdeutlichen. Carlos würde wohl liebend gerne vergessen, wie es dazu kam zu glauben, was er eben tut. Da er aber zu keinem Zeitpunkt die Gewähr hat, dass

die Realität oder sein eigenes Gedächtnis ihn nicht doch noch eines besseren belehrt, kann er es sich gar nicht leisten, die Belege zu vergessen, die letztlich seine Selbsttäuschung veranlasst haben. Der Grund ist wie gesagt der, dass gerade das Gewährsein dieser Belege es ist, was ihn antreibt, sich der durch die Belege gestützten Überzeugung zu entledigen.

17. Dieser Widerstreit, den ich hier nur benennen, nicht aber lösen kann, findet sich auch in Erklärungsansätzen, die eine Art von ‚Aufteilung‘ des Geistes (*compartmentalization of the mind*) postulieren. Gemäss diesem Ansatz – am prominentesten von Davidson vertreten – können zwei sich offensichtlich entgegengesetzte Überzeugungen nur dann koexistieren, „if they were somehow kept separate, not allowed to be contemplated in a single glance“ (Davidson, 1998: S. 8). Dabei meint die Rede von einem ‚geteilten Geist‘ (*of the mind as being partitioned*) nicht mehr, als „that a metaphorical wall separated the beliefs which, allowed into conciousness together, would destroy at least one“ (ebd.). Auf der anderen Seite macht er jedoch geltend, dass kein Grund zur Annahme besteht, einer dieser ‚Teile‘ müsse für das Bewusstsein unzugänglich sein (vgl. Davidson, 1985: S. 147).

18. Wie gesagt, auch dieser Erklärungsansatz ist nicht frei von Widersprüchen und es bedürfte weiterer Erklärungen um deutlich zu machen, wie ein einzelner Geist (*a single mind*) aus nur lose integrierten, relativ unabhängigen Teilen oder Subsystemen bestehen kann. Einen unbestreitbaren Vorteil scheint dieser Erklärungsansatz aber zu haben. Er erlaubt es, Selbsttäuschung als ein *irrationales Phänomen*, genauer: als einen *irrationalen Zustand* zu verstehen und zu erklären. Auf die Frage, an welcher Stelle genau Irrationalität eintritt, gibt es diesem Ansatz zufolge eine klare Antwort: Der irrationale Schritt ist das Getrennthalten der konfligierenden Überzeugungen, gleichsam das Ziehen einer Grenze, mit dem die eine oder andere der Überzeugungen zwar ausserhalb der Jurisdiktion der Vernunft zu liegen kommt, deshalb aber nicht etwa ‚machtlos‘ wird (vgl. Davidson, 1985: S. 148).

19. Die eben vorgebrachten Überlegungen haben in *normativer* Hinsicht weitreichende Konsequenzen, denn Irrationalität ist, wie Rationalität, ein im wesentlichen normativer Begriff. Bemerkenswert ist nun folgendes: Würde der von Goldman vorgeschlagene und mit McLaughlin weitergedachte Erklärungsansatz ausreichen, wäre es möglich, das Phänomen der Selbsttäuschung weitgehend von der Frage zu entkoppeln, wie weit wir es hier mit einem *irrationalen* Phänomen zu tun haben. Gemäss diesem Erklärungsansatz ist es möglich, eine sich selbsttäuschende Person so zu beschreiben, dass sie sich in keinem Moment in dem irrationalen Zustand befindet, etwas zu glauben, das sie aufgrund des Gewichts der verfügbaren Belege selbst gleichzeitig – und in gewisser Weise auch bewusst – für diskreditiert hält.

20. Werfen wir noch einmal einen Blick auf Carlos. Seine Angst vor einem erneuten Scheitern mag erklären, warum er sich selbst zu überzeugen versucht, dass er die Prüfung bestehen werde. Aber seine Angst oder der durch sie veranlasste Wunsch ist kein rationaler Grund für das, was er sich glauben machen möchte; und es gibt auch keinen Grund zur Annahme, er erachte denselben als guten Grund zu glauben, er werde die Prüfung bestehen (vgl. Davidson, 1998: S. 7). Trifft dies zu, dann haben wir zumindest im Fall von Carlos eine Person vor uns, die gegen ihre eigene Auffassung dessen, was vernünftig ist, denkt, fühlt und möglicherweise handelt; ein Fall also, bei dem eine bestimmte Art von innerer Inkonsistenz am Werk ist. Oder anders ausgedrückt: Insofern Carlos nicht in Übereinstimmung mit dem von ihm selbst akzeptierten Standard (d. i. das *Prinzip der Gesamtheit der Belege*) überlegt und Schlüsse zieht, sind wir zu dem normativen Urteil berechtigt, sein Denken sei *irrational* (vgl. Davidson, 1985a: S. 346).

21. Nachdem ich in diesem 1. Teil die deskriptive Relevanz von Goldmans Unterscheidung zwischen aktivierten und nicht-aktivierten Überzeugungen kritisch beleuchtet und einige grundsätzliche Schwierigkeiten dieses Erklärungsansatzes zur Sprache gebracht habe, werde ich im nun folgenden 2. Teil die normative Relevanz dieser Unterscheidung auf den Prüfstand stellen.

2. Irrationalität

1. Die normative Relevanz der fraglichen Unterscheidung versucht Goldman mit Blick auf das sog. *Prinzip der Gesamtheit der Belege* auszuweisen. Nach seiner Auffassung ist in die ursprüngliche Formulierung dieses Prinzip die unrealistische, mithin unplausible Rationalitätsforderung inkorporiert, nur derjenigen Proposition Glauben zu schenken, die durch *alle* verfügbaren Belege, bestätigt wird, miteingeschlossen alle nicht-aktivierten Überzeugungen (vgl. Goldman, 1986: S. 204). Dagegen führt er an, dass, wenn sich jemand wie Melanie im 1. Beispiel aus Vergesslichkeit, aus Acht- oder Gedankenlosigkeit in einem Datum täuscht, nichts darauf hindeutet, dass im Schliessen der betreffenden Person ein Fehler aufgetreten ist oder sie sich irgendeiner Irrationalität schuldig gemacht hat (ebd.). Melanies Versäumnis ist ihm zufolge lediglich eine Art von Panne (*breakdown*), die sich mit Hilfe der Unterscheidung zwischen aktivierten und nicht-aktivierten Überzeugungen präzise und korrekt wie folgt identifizieren lässt:

Melanie's problem is a failure to *activate* all the LTM [long-term memory, M. S.] stored facts relevant to her decision. This may be a fault of sorts, but it is not a flaw in reasoning. (Goldman, 1986: S. 204)

2. Was ihn angesichts solcher ‚Pannen‘ beschäftigt, ist jedoch nicht die Frage nach der begrifflichen Dimension, die einer solchen Charakterisierung zugrunde liegt, sondern die Frage, *warum* es einer Person misslingt, relevante Daten aus dem Langzeitgedächtnis wiederherzustellen, d. h. die im Langzeitgedächtnis gespeicherten Informationen zu *aktivieren* (vgl. S. 205–6). Gemäss seiner naturalistischen Konzeption (von Erkenntnistheorie) ist Schliessen, ob praktisches, induktives oder deduktives, ein *kausaler Prozess*, bei welchem in der Erzeugung einer neuen Überzeugung alle Prämissen *kausal wirksam* sind. Da jedoch nur *aktivierte* Überzeugungen kausal wirksam sind und da kein Wesen mit unserer Natur in der Lage ist, alle Überzeugungen (gleichzeitig) zu aktivieren, ist jedes Rationalitätsprinzip, so sein Argument, das die Unterscheidung zwischen aktivierten und nicht-aktivierten Überzeugungen ignoriert, unplausibel (vgl. S. 207).⁴

3. Zwei Dinge möchte ich hier zunächst festhalten. Erstens: Bislang hat Goldman, wie er selbst einräumt, erst einen negativen Punkt geltend gemacht; eben den, dass das *Prinzip der Gesamtheit der Belege* ohne Bezugnahme auf psychologische Fakten nicht angemessen spezifiziert werden kann und jeder Plausibilität entbehrt. Trifft meine Kritik, die ich im 1. Teil vorgetragen habe, zu, dann ist sein Versuch, die normative Relevanz an die von ihm ins Auge gefasste deskriptive Basis zurückzubinden, jedoch in einem hohen Masse problematisch. Zweitens: Die von ihm betrachteten Fälle von Selbsttäuschung scheinen nicht wirklich ein normatives Urteil zu verlangen. Damit entfällt gewissermassen auch die Notwendigkeit zu klären, was genau es heisst, normative Fragen – vorsichtig ausgedrückt – an deskriptive Fragen zurückzubinden.

4. Mein erster Einwand ist nun folgender: Schon in der Wiedergabe und erst recht in seiner Kritik verengt Goldman in meinen Augen das fragliche Prinzip auf mechanisch-kausale Aspekte, die es in der ursprünglichen, sicher aber in den modifizierten Formulierungen Hempels gar nicht aufweist (vgl. etwa Hempel, 1965: S. 397–403). Dass Goldman ausschliesslich diese Aspekte betont, hängt vielmehr mit seinem Versuch zusammen, diesem ursprünglich als handlungsleitend gedachten, auf Praxis bezogenen Prinzip eine Form zu geben, die mit seinem naturalisierten Begriff der Überzeugung kompatibel ist. Dabei wird gleichsam ‚unter Hand‘ der Bezug auf den normalen Überzeugungsbegriff zum Verschwinden gebracht, obwohl er sich bei der Frage, was es heisst, ein Stück Evidenz – einen Beleg – zu *besitzen*, an folgender Antwort orientiert: *S* ist im Besitz des Belegs *E* nur dann, wenn *S* glaubt, dass *E* (vgl. Goldman, 1986: S. 204).

5. Zweitens haben wir zu fragen, ob sich Goldman nicht in ein Dilemma verstrickt. Einerseits will er Erkenntnistheorie auch in naturalisierter Form explizit als eine normative Disziplin verstanden wissen. Andererseits deutet,

wie eben angesprochen, seine Behandlung des Überzeugungsbegriffs daraufhin, dass sich sein naturalistischer Erklärungsansatz der begrifflichen Grundlagen beraubt, die erforderlich sind, um sich als eine normative Disziplin ausweisen zu können. Trifft dies zu, dann wäre es kein Zufall, dass Selbsttäuschung im strengen Sinne innerhalb seines Erklärungsansatzes keinen Platz findet. Wenn wir in solchen Fällen von Selbsttäuschung eine Überzeugung als irrational bezeichnen, so fällen wir, wie bereits gesagt, ein normatives Urteil. Die kritische Frage lautet demnach: Ist er in *begrifflicher* Hinsicht überhaupt in der Lage, ein solches Urteil zu fällen?

6. Goldman hat sowohl für den Wissens- als auch für den Rechtfertigungsbegriff eingehende Analysen vorlegt. Es fehlt jedoch eine vergleichbare Analyse des Rationalitätsbegriffs; und dies, obwohl er die Frage, was Rationalität sei, in seinem Buch *Philosophical Applications of Cognitive Science* als eine der Leitfragen der Erkenntnistheorie kennzeichnet (vgl. Goldman, 1993: S. 1).⁵ In *Epistemology and Cognition* heisst es, von einer positiven Analyse des Rationalitätsbegriffs werde deshalb Abstand genommen, weil dieser Begriff in der Philosophie äusserst vage verwendet werde (vgl. Goldman 1986: S. 27, 318). Diese Aussage halte ich aus verschiedenen Gründen für irreführend. (1) Wenn es nichts zu klären gibt, warum setzt er sich dann überhaupt mit Rationalitätsprinzipien auseinander? (2) Gilt gleiches nicht auch für den Wissens- bzw. Rechtfertigungsbegriff? Und (3): Kann der Rechtfertigungs-, mithin der Wissensbegriff ohne einen zugrunde gelegten Rationalitätsbegriff überhaupt angemessen verstanden werden?

7. Betrachten wir zum Schluss eine Passage, wo er ausdrücklich die Frage stellt, was denn eigentlich als eindeutiger Fall von Irrationalität zählen würde. Irrationalität, so lautet seine Antwort, ist ein Defekt in grundlegenden überzeugungsbildenden Prozessen, genauer: ein Defekt, der etwas mit den grundlegenden ‚Schlussprozessen‘ (*reasoning processes*) zu tun hat (vgl. S. 318). Dass hier der Begriff des Schliessens, der ja aufs engste mit dem Begriff der Rationalität zusammenhängt, in Anführungszeichen gesetzt ist, scheint mir kein Zufall. Es ist durchaus legitim und im allgemeinen auch unproblematisch, intentionale Phänomene – mentale Zustände oder Ereignisse – unter Berufung auf neurophysiologische oder gar physikalische Ursachen zu erklären (etwas, was beispielsweise für die Analyse des Gedächtnisses zentral ist). Meine Bedenken gelten vielmehr dem folgenden Umstand: Wenn wir die Ursachen in nicht-intentionalem oder einem ‚hybriden‘ Vokabular beschreiben – als solches verstehe ich die Rede von aktivierten und nicht-aktivierten Überzeugungen –, verlieren wir notwendigerweise den Kontakt zu dem, was erforderlich ist, um den Schritt vom rein Deskriptiven zum Normativen, mithin zum Rationalen, miteingeschlossen das Irrationale machen und erklären zu können (vgl. Davidson, 1982: S. 299). Aus eben diesem Grund erachte ich seine Charakterisierung von Irrationalität als ei-

nen Defekt in grundlegenden überzeugungsbildenden Prozessen (vgl. Goldman, 1986: S. 318) als nicht angemessen.

8. Geht man vom gewöhnlichen Überzeugungsbegriff aus, so mutet Goldmans strikte Unterscheidung zwischen deskriptiven und normativen Belangen krude an, denn in diesem Begriff sind deskriptive und normative Elemente unlöslich miteinander verklammert. Goldman, so mein Einwand, löst diese Verklammerung auf, ohne dass ersichtlich wird, wie die einmal aufgerissene Kluft wieder geschlossen werden kann. Der Schritt vom Deskriptiven zum Normativen scheint mir jedoch nur dann kein blinder Schritt zu sein, wenn die in unserem gewöhnlichen Begriffen verankerten Standards der Normativität bei diesem (Erklärungs-)Schritt Pate stehen.⁶ Dass Goldman sich diesbezüglich keiner Selbsttäuschung hingibt, dafür sprechen in meinen Augen die besagten Führungszeichen.

Anmerkungen

- ¹ Wie gesagt, diese Unterscheidung ist für meine Zwecke hilfreich und ich denke auch ausreichend. Sie ist aber, dies sei hier wenigstens erwähnt, nicht unproblematisch. So haben etwa McLaughlin und Jonston nachzuweisen versucht, dass Selbsttäuschung zwar motiviert und insofern zweckgerichtet ist, deswegen aber nicht intentional zu sein braucht (vgl. McLaughlin (1988) und Jonston (1988)). In eine ähnliche Richtung zielen auch Autoren wie Bach und Mele sowie jüngst Lazar (vgl. Bach (1992), Mele (1999) sowie Lazar (1999)). Der Hauptpunkt der genannten Autoren scheint mir in deren Bestreben zu liegen, der Tendenz entgegenzuwirken, mentale Prozesse im allgemeinen und Selbsttäuschung im besonderen zu ‚überrationalisieren‘. Ich für meinen Teil sehe hier eine gewisse Gefahr, diese Phänomene zu ‚unterrationalisieren‘.
- ² Auf diesen Punkt, wonach der Begriff der Selbsttäuschung mehr interpretativen oder diagnostischen Zwecken dient als prediktiven, hat insb. Audi hingewiesen (vgl. Audi, 1985).
- ³ Vgl. zu dieser wichtigen Unterscheidung insb. Davidson, 1985: S. 138–9, 147, ders., 1985a: S. 353, ders., 1998: S. 5 sowie McLaughlin, 1988: S. 48–9. Diese Unterscheidung verdankt sich weniger einem empirischen Faktum, sondern dem Umstand, dass die Zuschreibung der Überzeugung [dass p und nicht- p] – also das Glauben einer Kontradiktion – es fraglich macht, ob wir denjenigen, dem wir eine solche Überzeugung zuschreiben, überhaupt noch als ein rationales Wesen betrachten.
- ⁴ In einem Buch neueren Datums greift Goldman diese Überlegungen wieder auf (vgl. Goldman, 1993), ohne in der Substanz jedoch etwas wesentlich Neues hinzuzufügen.
- ⁵ Die Fragen, die er behandelt, sind vielmehr Fragen nach den *Aussichten* und insbesondere den *Quellen* (d. h. den kausalen Ursprüngen) von Rationalität (vgl. Goldman, 1993: S. 9, 13).
- ⁶ In einem Aufsatz neueren Datums wird eben dies auch von Goldman behauptet, wenn er schreibt: „It may well be desirable to reform or tran-

scend our epistemic folkways [commonsense epistemic concepts and norms, M. S.] [...] But it is essential to preserve continuity; and continuity can only be recognized if we have a satisfactory characterization of our epistemic folkways.“ (Goldman, 1992: S. 155f.)

Literatur

- Anderson, J.; 1983: *The Architecture of Cognition*, Cambridge, Mass.
- Audi, R.; 1985: Self-Deception and Rationality, in: *Self-Deception and Self-Understanding. New Essays in Philosophy and Psychology*, hrsg. v. M. W. Martin, Lawrence 1985, S. 169–94.
- Bach, K.; 1992: Reviews, in: *Noûs*, Vol. 26, 1992, S. 495–504.
- Davidson, D.; 1982: Paradoxes of irrationality, in: *Philosophical essays on Freud*, hrsg. v. R. Wollheim u. J. Hopkins, London 1982, S. 289–305.
- ; 1985: Deception and Division, in: *Actions and Events. Perspectives on the Philosophy of Donald Davidson*, hrsg. v. E. LePore u. B. McLaughlin, Oxford 1985, S. 138–48.
- ; 1985a: Incoherence and Irrationality, in: *Dialectica*, Vol. 39, 1985, S. 345–54.
- ; 1998: Who Is Fooled?, in: *Self-Deception and Paradoxes of Rationality*, hrsg. v. J.-P. Dupuy, Stanford 1998, S. 1–18.
- Goldman, A. I.; 1986: *Epistemology and Cognition*, Cambridge, Mass.
- ; 1992: Epistemic Folkways and Scientific Epistemology, in: ders.; *Philosophy Meets the Cognitive and Social Science*, Cambridge 1992, S. 155–75.
- ; 1993: *Philosophical Applications of Cognitive Science*, Boulder.
- Hempel, C. G.; 1965: *Aspects of Scientific Explanation*, New York.
- Jonston, M.; 1988: Self-Deception and the Nature of Mind, in: *Perspectives on Self-Deception*, hrsg. v. B. P. McLaughlin u. A. O. Rorty, Berkeley 1988, S. 63–91.
- Kemmerling, A.; 1998: Selbstkenntnis als ein Test für den naturalistischen Repräsentationalismus, Manuskript, abgedr. in: *Naturalismus. Philosophische Beiträge*, hrsg. v. G. Keil u. H. Schnädelbach, Frankfurt a. M. 1998.
- Lazar, A.; 1999: Deceiving Oneself or Self-Deceived? On the Formation of Beliefs „Under the Influence“, in: *Mind*, Vol. 108, 1999, S. 265–90.
- Martin, M. W.; 1985: General Introduction, in: *Self-Deception and Self-Understanding. New Essays in Philosophy and Psychology*, hrsg. v. M. W. Martin, Lawrence 1985, S. 1–27.
- McLaughlin, B. P.; 1988: Exploring the Possibility of Self-Deception in Belief, in: *Perspectives on Self-Deception*, hrsg. v. B. P. McLaughlin u. A. O. Rorty, Berkeley 1988, S. 29–62.
- Mele, A. R.; 1999: Twisted self-deception, in: *Philosophical Psychology*, Vol. 12, 1999, S. 117–37.