DZPh, Berlin 57 (2009) 1, 73–95

Fehler im Haus der Vernunft*

Von MATTHIAS VOGEL (Basel)

Wenn man einigen gegenwärtig einflussreichen Theorien des Geistes folgt, dann stellt sich der Geist als ein Netz intentionaler Zustände und korrespondierender Äußerungen dar, die in Prozessen der rationalisierenden Interpretation oder in Praktiken des Kontoführens, die sich gleichfalls vor dem Hintergrund von Rationalitätsstandards ausweisen, Gehalte gewinnen. Weil Rationalität in diesen Theorien eine konstitutive Rolle für den Geist spielt, ist nicht ohne weiteres zu sehen, wie das Phänomen der Irrationalität in das Bild des Geistes integriert werden kann. Denn in dem Maße, in dem Geist als ein Phänomen beschrieben wird, das maßgeblich vom „Zement des Geistes", nämlich der Rationalität[1] zusammengehalten wird, wird fraglich, mit welchen theoretischen Ressourcen wir der Tatsache Rechnung tragen können, dass Menschen immer wieder auch irrational agieren.

Auch wenn die Instantiierung von Irrationalität keine unbezweifelbare Tatsache ist – nicht zuletzt, weil wir über einen stabilen Begriff der Irrationalität nicht unabhängig von einer Theorie des Geistes verfügen –, haben wir meines Erachtens gute Gründe für die Annahme, dass es über temporäre Irrtümer und nachvollziehbare Fehler hinaus eine bestimmte beharrliche Form der Irrationalität gibt, die zumindest die interpretationistische Theorie des Geistes vor anspruchsvolle Aufgaben stellt.

Da Davidson die Herausforderung erkannt hat, die das Phänomen der Irrationalität für eine interpretationistische Theorie des Geistes darstellt, und glaubt, ihr theoretisch begegnen zu können, werde ich in einem ersten Schritt zeigen, dass seine Theorie der Irrationalität, die im Wesentlichen eine Theorie der Spaltung oder Fragmentierung des Geistes ist, gemessen an seinen eigenen Ansprüchen im Wesentlichen unhaltbar ist. Meine These ist dabei nicht, dass Davidsons Analyse der Irrationalität falsch ist, weil er einen schlechten Tag hatte, als er etwa *Paradoxes of Irrationality* schrieb, oder allgemeiner: weil seine Theorie vermeidbare Fehler enthält. Meine These ist, dass seine Theorie falsch ist, weil sein Bild des Geistes *strukturell* keine hinreichenden Ressourcen bereitstellt, um das Phänomen der Irrationalität in seiner ganzen Breite angemessen zu rekonstruieren. Den Spielraum dafür, der Wirklichkeit irrationalen Handelns und Räsonierens Rechnung tragen zu können, so meine These, gewinnen wir nur dann, wenn wir die These, dass mentale Zustände ihre Gehalte *exklusiv* rationalisierenden Interpretationen verdanken, aufgeben, ohne dabei darauf verpflichtet zu sein, auch

* Der vorliegende Artikel geht auf Vorträge zurück, die ich in Basel und Potsdam gehalten habe, und hat von den dortigen Diskussionen, für die ich hier noch einmal danke, profitiert.

1 Vgl. Quante (1995).

zu behaupten, dass wir das Spezifikum unseres Geistes unabhängig von Prozessen der rationalisierenden (Selbst-)Interpretation erläutern können. Diese benötigten Ressourcen stehen uns – so jedenfalls meine These – erst dann zur Verfügung, wenn wir das etablierte Bild des Geistes um eine funktionalistische Theorie der Psyche ergänzen.

Um diese These plausibel zu machen, werde ich folgendermaßen vorgehen: Im ersten Abschnitt setze ich mich mit Davidsons Theorie der Irrationalität auseinander. Dabei versuche ich zu zeigen, dass seine Analyse sowohl an seinen eignen Ansprüchen scheitert als auch mit etwas unvereinbar ist, das man die Phänomenologie der Irrationalität nennen könnte.

Weil Davidson seine eigenen Überlegungen so versteht, dass sie jedenfalls einigen Vorstellungen der Psychoanalyse gerecht werden sollen, kontrastiere ich sie mit Fällen, die zwar ziemlich schematisch sind, aber der Phänomenologie solcher Formen von Irrationalität näher kommen, mit denen sich die Psychoanalyse auseinandersetzt, als jene Beispiele, mit denen Davidson operiert. Dass Davidson seine Überlegungen überhaupt zur Psychoanalyse in Beziehung setzt, ist aufschlussreich. Offenbar ist er der Auffassung, dass die Psychoanalyse als ein Bemühen verstanden werden kann, bestimmte Formen der Irrationalität theoretisch zu erläutern, sie in individuellen Fällen zu verstehen und zu überwinden. Die Tatsache, dass die Psychoanalyse auch eine verstehende Anstrengung ist, impliziert jedoch nicht, dass das Verstehen der Psychoanalyse eines ist, das sich allein auf alltagspsychologische Standards stützt. Obwohl ich Davidson darin zustimme, dass die Psychoanalyse ein, vielleicht das fortgeschrittenste Verfahren ist, bestimmte Formen irrationalen Handelns zu verstehen, sodass es der Philosophie des Geistes (entgegen dem modischen Freud-Bashing) gut ansteht, sich mit ihr auseinanderzusetzen, glaube ich, dass Davidson, aber auch vielen anderen Philosophen, die die Kontinuität zwischen alltagspsychologischem und psychoanalytischem Verstehen betonen, die Pointe psychoanalytischen Verstehens und Erklärens entgeht.

Im zweiten Abschnitt skizziere ich eine alternative Analyse, die in den vorangegangenen Überlegungen eher tentativ ins Spiel kommt, die Analyse nämlich, dass sich Irrationalität als Folge von Störungen des Geistes durch funktional verstandene psychische Mechanismen verstehen lässt. Dabei überlege ich etwas allgemeiner, wie die Rede von psychischen Zuständen und Mechanismen mit der Rede von intentionalen Zuständen zusammenhängt, und versuche zudem, deutlich zu machen, dass eine funktionalistische Analyse der Irrationalität keinerlei reduktionistische Implikationen haben muss.

Der dritte und letzte Abschnitt bemüht sich darum, das entworfene Bild von Irrationalität für ein Verständnis von Willensschwäche nutzbar zu machen.

I. Irrationalität – aber wie?

Das Phänomen der Irrationalität wirft im Kontext interpretationistischer Theorien des Geistes Probleme auf, weil der Interpretationismus geltend macht, dass zwischen intentionalen Beschreibungen einerseits und der Zuschreibung von Wahrheit und Rationalität andererseits eine enge *begriffliche* Beziehung besteht. Intentionale Zustände wie Absichten und Überzeugungen können wir nur zuschreiben, wenn wir uns dabei am Prinzip der Nachsichtigkeit orientieren, denn wenn wir das, was Sprecher sagen oder was Handelnde tun, nicht im Großen und Ganzen für wahr und rational halten, untergraben wir schlicht die Möglichkeit, Sprecher oder Handelnde zu verstehen. Verstehen und damit das Zuschreiben intentionaler Zustände muss sich an dem orientieren, was *wir* verstehen; dort wo es sich zu weit davon entfernt, zerfällt die Möglichkeit der Zuschreibung gehaltvoller Zustände. Im Rahmen einer Interpretation zu behaupten, ein Sprecher hielte eine ganze Reihe von Widersprüchen des Typs p und $\neg p$ für wahr, wirft

unmittelbar zwei Fragen auf: Wovon könnte jemand überzeugt sein, der diese Überzeugungen tatsächlich hat? Und, insofern wir darauf keine Antwort geben können, kann unsere Interpretation der Äußerungen des Sprechers zutreffend sein? Angesichts dessen sagt Davidson: „Die Nachsichtigkeit ist uns aufgezwungen, wenn wir andere verstehen wollen, müssen wir ihnen in den meisten Dingen recht geben, ob wir das mögen oder nicht."[2]

Wenn wir unsere Interpretation aber am Prinzip der Nachsichtigkeit orientieren müssen, um nicht die Bedingungen der Verstehbarkeit zu gefährden, taucht sogleich die nächste Frage auf: Wie schaffen wir dann Raum dafür, Äußerungen oder Handlungen zu Recht als irrational zu qualifizieren? Um diese Frage zu beantworten, ist es sinnvoll, das Phänomen der Irrationalität genauer zu bestimmen.

Irrationalität ist zunächst kein arationales Phänomen, das wir – wie bestimmte Aspekte der Wahrnehmung oder des Verhaltens – in blank kausalem Vokabular beschreiben können, sondern ein Phänomen, das inmitten des Intentionalen auftritt. Es ist, wie Davidson richtig sagt, „ein Versagen innerhalb des Hauses der Vernunft".[3] Irrational kann nur ein Wesen sein, das überhaupt im Raum der Gründe agiert, und dies gemäß unseren Interpretationen auch überwiegend erfolgreich tut. Von einer Fliege zu sagen, sie verhalte sich irrational, weil sie immer wieder gegen die Fensterscheibe fliegt, ergibt keinen Sinn, weil Fliegen die Bedingungen dafür, irrational zu sein, nicht erfüllen und wir ihr Verhalten daher gewöhnlich nicht im Rekurs auf ein inferentiell gegliedertes Netz intentionaler Zustände erklären.

In der alltagssprachlichen Kommunikation gehen wir mit dem Prädikat „irrational" relativ großzügig um. Wenn jemand etwas tut oder sagt, das *wir* gemessen an den Absichten einer Person für vollkommen untauglich oder inkohärent halten, das *wir* nicht verstehen, haben wir wenig Skrupel, von irrationalen Handlungen oder Äußerungen zu reden. Aber dass wir etwas nicht verstehen, dass wir eine Handlung für aussichtslos oder eine Äußerung für unvereinbar mit anderen Äußerungen halten, kann auch darauf zurückgehen, dass unsere Annahmen über die Überzeugungen, Wünsche und Absichten einer Person unvollständig oder falsch sind. Irrationalität liegt, wie Davidson zu Recht betont, *nicht* vor, wenn Leute etwas tun, das unseres Wissens keine vernünftige Aussicht auf Erfolg hat, denn selbst der „Versuch, den Kreis zu quadrieren, ist sinnvoll, wenn man nicht weiß, dass es nicht geht".[4] Und wenn Leute etwas sagen, das nach unserem Verständnis widersprüchlich ist, dann ist es schlicht möglich, dass ihnen der Widerspruch nicht bewusst ist, etwa weil sie die Implikationen dessen, was sie sagen, nicht überblicken.

Wenn wir Äußerungen oder Handlungen als irrational qualifizieren wollen, müssen wir unser Urteil relativ zu den intentionalen Zuständen treffen, die eine Person tatsächlich hat. Da unser Zugang zu diesen Zuständen aber immer interpretationsabhängig bleibt, bleiben auch unsere Urteile über die Irrationalität interpretationsabhängig:

(1) *P* sagt etwas, das gemäß *unserer Interpretation* mit anderen seiner Äußerungen nicht vereinbar ist.

(2) *P* tut etwas, das *unseres Wissens* nach ohne jede Aussicht auf Erfolg ist.

(3) *P* tut oder sagt etwas, für das *P* nach *unserem Verständnis* der Situation keine tragfähigen Gründe haben kann.

---

2 Davidson (1990a), 280.

3 Ders. (2006), 285.

4 Ebd., 287.

Verbunden sind diese Fälle durch die Tatsache, dass sie unser Bemühen, Äußerungen oder Handlungen zu verstehen, an eine Grenze stoßen lassen. Diese Grenze lässt sich jedoch leicht verschieben, indem wir die zugeschriebenen propositionalen Einstellungen modifizieren. So reicht es hin, wenn wir annehmen, dass *P* eine mit anderen Äußerungen unvereinbare Äußerung anders versteht als wir, dass *P* nicht weiß, dass eine Handlung relativ zu seinen Zielen ineffektiv ist, dass *P* eine Handlung ausführen möchte, die gerade ineffektiv sein soll, oder dass *P* eine Situation grundlegend anders beurteilt als wir.

Wir stehen also vor folgender Schwierigkeit: Wenn zwischen einem bestimmten Maß an Rationalität einer Person und dem Sachverhalt, dass diese Person überhaupt gehaltvolle intentionale Zustände hat, ein konstitutiver Zusammenhang besteht, dann wird fraglich, ob und in welchem Maße Irrationalität überhaupt möglich ist. Denn wenn eine Person das notwendige Maß an Rationalität nicht realisiert, dann hat sie keine gehaltvollen intentionalen Zustände, die die notwendige Grundlage dafür bilden, überhaupt irrational sein zu können. Die Aufgabe ist also folgende:

> „Um Irrationalität erklären zu können, müssen wir einen Weg finden, der es erlaubt, wesentliche Eigenschaften des Mentalen zu behalten – was die Bewahrung eines Rationalitätshintergrunds voraussetzt – und zugleich Formen von Kausalität zuzulassen, die von den Normen der Rationalität abweichen. Was benötigt wird, um Irrationalität zu erklären, ist eine mentale Ursache einer Einstellung, wobei die Ursache nicht der Grund für die Einstellung ist, die sie erklärt.“[5]

Aber welche Standards oder normativen Prinzipien muss ein Wesen erfüllen, damit der Hintergrund von Rationalität nicht zerreißt, vor dem wir Irrationalität diagnostizieren können? Sind es Standards wie:

(RPD)

(1) das *Prinzip des requirement of total evidence*: „Akzeptiere diejenige Hypothese, die durch alle verfügbaren relevanten Belege bestätigt wird!“[6];

(2) das *Prinzip der Selbstbeherrschung* oder *Kontinenzprinzip*: „Vollziehe diejenige Handlung, die du auf der Basis aller verfügbaren relevanten Gründe als die beste beurteilst!“[7];

(3) das *principle of conservation*: „Verändere so wenige Erwartungen wie möglich, wenn du dich unter Bedingungen der Konstanz einiger Dinge an die Eingliederung widerspenstiger Dinge machst!“[8];

(4) die elementaren *Prinzipien der Entscheidungstheorie* (zum Beispiel Transitivität der Präferenzen);

(5) die *Logik* des Satzkalküls?

---

5 Davidson (2006a), 318 (Übersetzung modifiziert).

6 Ders. (1990), 71.

7 Ebd.

8 Davidson (2006a), 320.

Selbst wenn wir uns vorstellen, dass Personen diese Prinzipien verletzen[9], so ist es in den Fällen (1), (3) und (4) möglich, dass sie glauben, sich an den ihnen verfügbaren Belegen zu orientieren, die für wahr gehaltenen Überzeugung nicht leichtsinnig zu revidieren oder den Prinzipien klugen Entscheidens zu folgen. Die Konditionierung des Irrationalitätsurteils auf subjektive Bedingungen des Handelnden kommt daher erst dort an ein Ende, wo Personen selbst einsehen, dass sie *ihren eigenen* Prinzipien nicht folgen.[10] Irrationalität ist daher daran gebunden, dass es dem Zusammenhang von Überzeugungen, Einstellungen, Gemütsregungen, Absichten und Handlungen im Inneren einer Einzelperson nach deren *eigenen* Maßstäben an Kohärenz oder Konsistenz mangelt.[11] Als Beispiele nennt Davidson „Wunschdenken, Handeln wider besseres Wissen, Selbsttäuschung und der Glaube an etwas, das nach *eigener* Meinung durch das Gewicht der Belege diskreditiert ist".[12] Um berechtigt Irrationalität zuschreiben zu können, bedarf es also einer relativ umfassenden rationalisierenden Interpretation, die dafür sorgt, dass eine Person überhaupt verständlich ist – einer Interpretation, in deren Lichte viele, wenn nicht die meisten Äußerungen der Person wahr und nicht widersprüchlich zueinander sind, sodass wir vor diesem Hintergrund *lokale* Defizite diagnostizieren können, die die Form von Widersprüchen, Wunschdenken, Handeln wider besseres eigenes Wissen, Selbsttäuschung und so weiter annehmen können. Versuchen wir zu erklären, wie Irrationalität *möglich* ist, müssen wir theoretischen Anforderungen gerecht werden, die in einem Spannungsverhältnis – Davidson redet von einer „Paradoxie" – zueinander stehen.

Wenn wir uns der Einfachheit halber zunächst einmal nur auf *Handlungen* beschränken, dann müssen wir dafür sorgen, dass eine Tätigkeit, die irrational ist, nicht den Status einer Handlung verliert. Da Handlungen gerade nur solche Tätigkeiten sind, die aus Gründen vollzogen wurden, für die wir also eine gründegestützte Erklärung geben können, würden Tätigkeiten, für die wir derartige Erklärungen nicht geben können, den Status einer Handlung verlieren. Sie wären bloßes Verhalten und als solches gar nichts, das überhaupt irrational sein könnte. Zur Beschreibung und Erläuterung dieses Verhaltens wären kausale Relationen schlicht hinreichend. Kausale Relationen evaluieren wir aber nicht vor dem Hintergrund normativer Rationalitätsstandards, sodass die Möglichkeit, eine Tätigkeit als irrationale Handlung zu beschreiben, verloren wäre. Wenn wir jedoch andererseits den Handlungsstatus einer Tätigkeit sichern wollen, dann müssen wir sie als rational beschreiben, und wären auf diese Weise der Möglichkeit beraubt, sie als irrational zu bewerten. Kurz: Handeln impliziert Rationalität, daher scheint irrationales Handeln nicht möglich. Kausal hinreichend erklärbares Verhalten hingegen ist arational und somit weder ein Kandidat für Rationalität noch für Irrationalität.

Davidsons Lösung für diese Paradoxie sieht folgendermaßen aus: Damit wir eine Tätigkeit als *Handlung* beschreiben können, müssen wir sie relativ zu intentionalen Zuständen eines Akteurs beschreiben, die sie rechtfertigen *und* verursachen. Irrational wird die Handlung genau dadurch, dass die mentalen Zustände, die sie begründen und verursachen, mit *anderen* mentalen Zuständen des Akteurs unvereinbar sind. Weil mentale Zustände auf Grund

---

9 Zwei etablierte Formen von Irrationalität implizieren unmittelbar die Verletzung dieser Prinzipien: Willensschwäche die Verletzung des Prinzips der Selbstbeherrschung und Selbsttäuschung die des Prinzips der verfügbaren Gesamtevidenz. Dass der Verletzung der anderen Prinzipien keine derartigen Syndrome korrespondieren, ist zumindest ein Hinweis darauf, dass sie einen anderen Status haben.

10 Vgl. Davidson (2006a), 317.

11 Vgl. ders. (2006), 287.

12 Ebd. (Hervorhebung von mir).

des Holismus des Mentalen nur insofern eine Identität und einen Gehalt haben, als sie in inferentiellen Beziehungen zu anderen mentalen Zuständen stehen, müssen wir – so Davidson – folgern, dass Akteure, die irrational handeln, über (mindestens) zwei Netze inferentiell organisierter intentionaler Zustände verfügen. Die Pointe seiner Lösung besteht, kurz gesagt, darin, dass eine irrationale Handlung *lokal* mit Rationalität ausgestattet wird, sodass ihr Status als Handlung gesichert wird, zugleich aber diese lokalen Bedingungen ihrer Rationalität mit mindestens einem weiteren gründeintegrierten Zusammenhang intentionaler Zustände des Akteurs nicht vereinbar sind.

Mit dieser Auflösung der Paradoxie der Irrationalität hat Davidson die Redeweise eingeholt, dass Irrationalität ein Fehler im Haus der Vernunft ist, und er hat das *begriffliche* Problem gelöst, wie Handlungen, für die Rationalität konstitutiv ist, dennoch irrational sein können. Allerdings halte ich Davidsons Lösung für alles andere als überzeugend, und zwar sowohl gemessen an seiner eigenen Theorie des Mentalen (1) als auch mit Blick auf die Phänomenologie der Irrationalität (2).

1. Lokale Rationalität und die Fragmentierung des Bewusstseins. In Bezug auf Davidsons Theorie des Mentalen stellt sich eine generelle Frage, deren Beantwortung eine ganze Serie miteinander zusammenhängender Probleme nach sich zieht: In welchem Verhältnis kann ein Akteur, der irrational handelt, zu den beiden intentionalen Netzen stehen, die er laut Davidson beherbergt? Da das Bewusstsein des Akteurs ein Produkt seiner Selbstinterpretation ist und eine Selbstbeschreibung impliziert, ist die Identität des Akteurs etwas, das sich nur im Rückgriff auf ein Netz intentionaler Zustände erläutern lässt. Da sich die beiden Netze, in denen die lokale Rationalität realisiert wird, unterscheiden müssen, scheint ein irrationaler Akteur zwei Identitäten haben zu müssen. Er müsste innerhalb jedes Netzes Selbstinterpretationen vollziehen, die sich auf die jeweils unterschiedlichen intentionalen Zustände der jeweiligen Netze stützen. Ein Bewusstsein von der Existenz dieser beiden Identitäten kann der Akteur jedoch nicht haben, denn dann müsste er ein Bewusstsein von der Existenz der beiden Netze haben, was nur möglich wäre, wenn es ein übergeordnetes Netz gäbe, das die beiden Netze als Subnetze integriert.

Wenn der Akteur jedoch zwischen den beiden Netzen nur *ohne* ein Bewusstsein von ihnen wechseln kann, dann ergäbe sich aus seiner Perspektive überhaupt kein Problem: Er könnte kein Bewusstsein von der Irrationalität seines Handelns gewinnen. Denn wenn er handelt, dann handelt er immer im Kontext eines Netzes intentionaler Zustände, die die Rationalität des Handelns gerade gewährleisten. Und selbst wenn wir annehmen, dass nach dem Übergang von einem Netz ins andere eine Erinnerung an die Überzeugungen übrig bliebe, die für das erste Netz spezifisch sind, müsste der Akteur an dieser Erinnerung zweifeln, weil sie mit seinen gegenwärtigen Überzeugungen überhaupt nicht harmoniert.[13] Davidson betont zu Recht, dass Irrationalität nur vorliegt, wenn einem Handelnden bewusst ist, dass seine Überzeugungen inkohärent sind, aber die Theorie der Spaltung in zwei inferentiell gegliederte Netze, die jeweils einige intentionale Zustände enthalten, die mit solchen des jeweils anderen Netzes unvereinbar sind, verhindert gerade, dass ein Handelnder dieses Bewusstsein erlangen kann, sodass die Möglichkeit irrationalen Handelns verschwindet.

Nun könnte man – anders als Davidson – denken, dass sich dieses Problem aus der Außenperspektive in den Griff bekommen ließe. Es wäre ja möglich, dass einer Beobachterin auffällt, dass eine Handlung des Akteurs, gemessen an solchen intentionalen Zuständen, die andere Hand-

[13] Die Alternative bestünde darin, daran zu zweifeln, ob der Akteur überhaupt eine Person ist. Aber wie könnte der Akteur auf diesen Zweifel reagieren?

lungen des Akteurs begründen, irrational ist. Warum sollte sie jedoch von Irrationalität reden? Sie könnte doch – gerade im Anschluss an die Theorie der Spaltung in zwei Netze – auch sagen, dass der Körper des Akteurs zwei im Großen und Ganzen rationale Personen beherbergt. Und genau das sollte sie sogar sagen, denn wenn sie, wiederum Davidson folgend, ihre Interpretationen am *Prinzip der Nachsichtigkeit* orientiert, kann sie alle Widersprüche beseitigen, indem sie annimmt, dass es sich um zwei Personen handelt, die im Körper des Akteurs hausen. Sie sollte mithin sagen, hier liege kein Fall von Irrationalität, sondern von multipler Persönlichkeit vor. Angesichts dieser Diagnose müsste sie nicht einmal mit Davidsons Kopfschütteln rechnen, denn Davidson wäre so ziemlich der letzte, der behaupten würde, dass Personen nur im Rückgriff auf solche Kriterien individuiert werden können, mit deren Hilfe man Körper individuiert. Die externe Diagnose der multiplen Persönlichkeit ist also teuer erkauft, denn sie nimmt uns die Möglichkeit, von Irrationalität zu reden. Um diesen Preis nicht zu zahlen, müssen wir annehmen, dass beide Subnetze durch die Selbstinterpretation *einer* Person aufgespannt werden. Dann aber wird vollkommen fraglich, wie es zu einer Verschiedenheit der Subnetze kommen könnte, die hinreichend groß ist, um Irrationalität zu ermöglichen.

Wenn man sich die Konsequenzen von Davidsons begrifflicher Lösung vergegenwärtigt, dann kommt man nicht umhin festzustellen, dass auch sie das Phänomen der Irrationalität zum Verschwinden bringt – und zwar entweder weil Akteure kein Bewusstsein von der Inkohärenz ihrer Netze gewinnen können oder weil sie in unterschiedliche, jeweils rationale Personen zerfallen, die nichts voneinander wissen.[14]

2. Die Wirklichkeit der Irrationalität. Kommen wir nun zum zweiten Grund, aus dem ich Davidsons „Lösung" des Problems der Irrationalität für unbefriedigend halte, nämlich die Phänomenologie oder die Wirklichkeit der Irrationalität.

(a) Es ist ohne Zweifel möglich, dass ein Akteur nach unseren Maßstäben irrational handelt, ohne dass ihm dies bewusst wäre. In diesen Fällen sind zwar wir als Beobachter und Interpreten des Akteurs vor Verständnisprobleme gestellt, nicht jedoch der Akteur selbst. Jemand tut etwas, ohne zunächst zu bemerken, dass seine Handlungen eindeutig dazu führen, eigene Präferenzen nicht zu erfüllen, sondern im Gegenteil zu frustrieren.

Ein anschauliches Beispiel hierfür ist das, was man neurotische Partnerwahl nennt: Eine Person hat den Wunsch nach einer harmonischen, erfüllenden Partnerschaft, sucht sich aber mit geradezu schlafwandlerischer Sicherheit stets Partner aus, die sich als Verhängnis herausstellen. Solange der Zusammenhang zwischen den eigenen Handlungen und der Frustration eigener Wünsche aber nicht bewusst ist, haben wir es hier *nicht* mit einem Fall von Irrationalität in Davidsons Sinne zu tun. Denn in Analogie zu Davidsons Bemerkung zur Quadratur des Kreises müssen wir einfach sagen: Die Person *weiß* eben nicht, dass die Art von Handlungen, die sie in Form von Partnerwahlen vollzieht, die Ursachen ihrer frustrierenden Schwierigkeiten sind.

Als Beobachter solcher Partnerwahlen mögen wir mitunter kopfschüttelnd – und in gewisser Weise auch hilflos – zusehen, wie sich Bekannte oder Freunde zum wiederholten Male in Beziehungen zu Personen stürzen, die ihnen das Leben zur Hölle machen werden.

---

[14] Eine umfassende Kritik an unterschiedlichen Formen der Spaltungsthese, wie sie sich bei Davidson und Pears finden, leistet Gardner (1993, Kap. 2 u. 3) im Rückgriff auf Sartres Kritik an der Psychoanalyse, die, wie Davidson, davon ausgeht, dass die Psychoanalyse von einer These über die Teilung von Personen in Subsysteme abhängt. Gardners eigene Überlegungen versuchen daher, das Unbewusste von der Vorstellung zu lösen, ein zweiter Geist zu sein. Zur Kritik der Teilungshypothese vgl. auch Henden (2004).

Ein Problem der Irrationalität liegt hier gemäß Davidsons Maßstäben jedoch nicht zwingend vor. Erst wenn die betreffende Person sagt: „Ich verstehe das nicht. Ich weiß, dass die Partner, die ich mir aussuche, schlecht für mich sind, aber ich mache es immer wieder, und ich weiß einfach nicht, warum", erst dann liegt ein interessanter und lebensnaher Fall von Irrationalität vor – und einer, den auch Davidson als solchen akzeptieren würde. Fälle dieses Typs möchte ich als genuin irrationales Handeln bezeichnen:

(G1) Eine Person *P* handelt genuin irrational, wenn

(1) *P* in gewissen Kontexten immer wieder so agiert, dass infolge ihres Agierens ihre Präferenzen frustriert werden;

(2) *P weiß*, dass ihre Präferenzen infolge ihres Agierens frustriert werden;

(3) *P* diese Aktivitäten nicht als Reflexe oder bloße Verhaltensweisen versteht, sondern als eigene *Handlungen*;

(4) *P weiß*, dass es bessere Alternativen zu diesen Handlungen gibt beziehungsweise gab;

(5) *P weiß*, dass sie diese Alternativen gleichwohl nicht realisiert;

(6) *P nicht weiß*, warum sie die Handlungen vollzieht, von denen sie weiß, dass sie ihr in der Konsequenz schaden.

Das läuft auf die These hinaus, dass Personen in praktisch relevanten Fällen von echter Irrationalität – wie Jonathan Lear bemerkt hat – so etwas erleiden wie einen *reflexive breakdown*.[15] Der These, dass man echte Irrationalität von den „kalten Fällen" der Irrationalität, in denen kognitive Fehler vorliegen, und gewöhnlichen Formen der Irrationalität, wie Wunschdenken, Selbsttäuschung und Willensschwäche, im Rekurs auf eine Störung der Selbstreflexion auszeichnen sollte, stimmt auch Sebastian Gardner zu, wenn er diese Formen von Irrationalität als *propositional transparent* bezeichnet und von anderen Formen unterscheidet, in denen Personen sich selbst undurchsichtig werden und deshalb an dem Versuch scheitern, sich rechtfertigende Selbsterklärungen zu geben oder sich dabei in Widersprüche verstricken.[16] Personen, die einen reflexiven Zusammenbruch erleiden und sich selbst keine Rechtfertigung für ihr Handeln geben können, handeln in Übereinstimmung mit Davidsons Charakterisierung gegen eigenes besseres Wissen, aber sie verstehen nicht, warum. Halten wir also fest:

(GI_2) Die genuine Irrationalität einer Person *P* impliziert, dass *P* ein Bewusstsein von der Inkohärenz ihrer intentionalen Zustände (beziehungsweise ihrer Handlungen mit ihren intentionalen Zuständen) hat. Dieses Bewusstsein impliziert wiederum, dass *P*s Fähigkeit zur Selbstrationalisierung an eine Grenze gekommen ist.

Nach (GI_2) liegt genuine Irrationalität vor, wenn Personen nicht verstehen, warum sie handeln, wie sie handeln; irrational sein heißt, sich selbst (lokal) unverständlich sein. Obwohl nun Davidson fordert, dass Irrationalität ein Bewusstsein der Inkohärenz impliziert, kann er eine Bestimmung im Sinne von (GI_2) nicht ins Spiel bringen, weil der lokale Zusammenbruch der Selbstinterpretation der Rede von Handlungen oder intentionalen Zuständen den Boden entziehen würde. Wenn zwischen einem Bewusstsein der Inkohärenz und Prozessen der Selbstinterpretation ein systematischer Zusammenhang besteht, dann liegt in den subjektrelativen

---

15 Vgl. Lear (1998), 81; dt. in diesem Band: 48.

16 Vgl. Gardner (1993), 15.

Klauseln, die Davidson in Formulierungen wie „Handeln wider *eigenes* besseres Wissen", „Glaube an etwas, das *nach eigener Meinung* durch das Gewicht der Belege diskreditiert" wird, verwendet, genau der Sprengstoff, der den lokalen reflexiven Zusammenbruch zur Folge hat. Aus Davidsons Perspektive gefährdet der lokale reflexive Zusammenbruch aber wiederum die Möglichkeit, Handlungen wider besseres eigenes Wissens überhaupt als Handlungen zu betrachten. Denn wenn man, um irrational zu sein, *wissen* muss, dass eine eigene Handlung gerade nicht durch die eigenen Absichten und Überzeugungen begründet werden kann, wie kann man dann das, was man tut, eine Handlung nennen?

Der Holismus der mentalen Sprache, in der sich die Konzepte Handlung, Überzeugung, Rationalität, Grund und Verstehen unauflöslich wechselseitig erläutern, scheint einfach keinen Raum für echte Irrationalität zu haben. Zwar lassen sich die propositional transparenten Formen von Irrationalität, etwa Wunschdenken, Selbsttäuschung und Willensschwäche, im Rahmen der Alltagspsychologie identifizieren oder diagnostizieren, aber eben nicht vollständig erklären. Den Raum für die Erklärung dieser Formen von Irrationalität, vor allem aber jener Formen, in denen die Selbstreflexion an ein Ende kommt – so meine Vermutung –, lässt sich nur gewinnen, wenn der konstitutive Zusammenhang zwischen der Eigenschaft, einen Gehalt zu haben, und der Tatsache, Gegenstand einer erfolgreichen rationalisierenden Interpretation zu sein, zu Gunsten anderer nichtrationalisierender gehaltsbestimmender Mechanismen geöffnet wird.

(b) Angesichts der genannten Probleme könnte man darauf verfallen, die Rede von Handlungen aufzugeben, und geltend machen, dass es sich bei den Aktivitäten, die wissentlich mit den eigenen Präferenzen unvereinbare Folgen zeitigen, nicht wirklich um Handlungen handelt. Denn wenn es sich um Handlungen handelte, müsste es gute Gründe für diese Handlungen geben, und das Wissen um die Folgen solcher Handlungen ist ja gerade ein Wissen darum, dass es diese guten Gründe nicht gibt. Mithin werden die Aktivitäten nicht durch Gründe verursacht und sind daher auch keine Handlungen. Man könnte also versuchen zu sagen, dass Personen, die in meinem Sinne genuin irrational handeln, ihre Aktivitäten als Handlungen *missverstehen*. Genau das ist jedoch aus Davidsons Perspektive nicht möglich, denn entweder hat eine Person Gründe für eine Aktivität, aus denen sie handelt, oder sie hat sie nicht, sodass nur die Alternative bliebe, die vermeintliche Handlung als *Verhalten* zu beschreiben, für das es zwar Ursachen, aber eben keine Gründe gibt. Dann aber, das haben wir ja gerade von Davidson gelernt, würde sich auch die Irrationalität in Luft auflösen. Die Situation wäre folgende: Anstelle der oben genannten Auskunft, die den reflexiven Zusammenbruch ausdrückt, würde unsere Person nun sagen: „Ich lebe immer wieder mit Partnern zusammen, die mir schaden und mit denen ich durch ein Verhalten zusammenkomme, das an mir wie ein Reflex abläuft und dessen Ursachen ich nicht kenne." Diese Auskunft aber wäre isomorph zu: „Immer wenn es einen lauten Knall gibt, erschrecke ich so, dass mir ganz übel wird, ich weiß aber nicht, warum das so ist." Hier mag es um Unkenntnis und manches Unangenehme gehen, um Irrationalität geht es nicht.

Fassen wir also den vorläufigen Stand der Dinge zusammen: Davidson sagt zu Recht, eine Handlung sei eine von Gründen verursachte Tätigkeit. Damit eine Tätigkeit irrational sein kann, muss sie zunächst einmal eine Handlung sein, sie muss also durch Gründe verursacht werden. Falls eine Tätigkeit nicht durch Gründe, sondern allein durch Ursachen verursacht wird, ist sie nicht Gegenstand einer Bewertung nach Standards der Rationalität; sie ist vielmehr ein Naturereignis. Diese Diagnose nimmt Davidson zum Anlass, das so genannte „Medea-Prinzip" als Modell für irrationales Handeln zurückzuweisen, denn wenn Medea ihre Hände anfleht, ihre Kinder nicht zu ermorden, ihr Wille jedoch überwältigt wird, dann liegt keine Irrationalität vor, sondern eine Ursache, die die Tötung verursacht, aber nicht begründet. Irrational kann ihre Tat daher nur dann sein, wenn die *Gründe*, die sie verursachen, mit

anderen intentionalen Zuständen Medeas konfligieren. Darüber hinaus muss dieser Konflikt der irrational handelnden Person bewusst sein. Denn eine Person, die inkohärente intentionale Zustände hat, sich der Inkohärenz aber nicht bewusst ist, erfüllt eine fundamentale Bestimmung der Irrationalität nicht, die Voraussetzung nämlich, dass jemand etwas tut, das *nach eigenen Standards* mit *eigenem besseren Wissen* konfligiert.

Ich will nicht verschweigen, dass es in Davidsons Texten Hinweise gibt, wonach die Rede von eigenen Standards nicht zu implizieren scheint, dass diejenigen, um deren Standards es geht, ein Bewusstsein (von der Verletzung) dieser Standards haben. Dies scheint etwa in folgenden Formulierungen der Fall zu sein: „Ich glaube nämlich, dass diese Prinzipien von jedem gutgeheißen werden – ob er's weiß oder nicht."[17] Oder noch deutlicher: „Innere Widersprüchlichkeit ist genau deshalb möglich, weil es Normen gibt, die keinem Akteur abgehen können. Die Widersprüchlichkeit braucht vom Akteur nicht bemerkt zu werden, aber es kann natürlich sein, dass er sich darüber im klaren ist; und das Vorhandensein einer Widersprüchlichkeit hängt nicht davon ab, dass der Akteur dazu imstande ist, die Prinzipien, gegen die er verstößt, zu formulieren."[18]

Diese Formulierungen werfen eine Reihe von Problemen auf: Unter anderem stehen sie in einem Spannungsverhältnis zu Formulierungen im gleichen Text, wenn es etwa über einen Mann, der seinen Nachbarn voreilig des Diebstahls verdächtigt, heißt: „Aber nehmen wir einmal an, er verfüge über weit bessere Belege gegen seine Überzeugung als für sie. Dann ist er dennoch nicht irrational, es sei denn, *er sieht ein*, dass seine Belege tatsächlich gegen seine Überzeugungen sprechen."[19] Hier und an anderen Stellen[20] wird Irrationalität klar an das Bewusstsein von der eigenen Widersprüchlichkeit beziehungsweise von der Verletzung eigener Standards gebunden. Die Spannung zwischen den genannten Formulierungen lässt sich jedoch folgendermaßen auflösen: Es ist zwar möglich, dass Akteuren ihre Widersprüchlichkeit nicht bewusst ist, aber diese Fälle sind keine Fälle, in denen Irrationalität instantiiert wird.

Wenn die Inkohärenz aber bewusst sein muss, und sowohl der Zusammenhang, den die rationalen intentionalen Zustände bilden, als auch der Zusammenhang, der die irrationale Handlung begründet und verursacht, inferentiell integriert sein müssen, bleibt nur das Modell der Spaltung in unterschiedliche Netze. Da, wie ich oben gezeigt habe, dieses Modell der Spaltung unhaltbare Konsequenzen hat, komme ich zu dem Schluss, dass Davidsons Überwindung der Paradoxien der Irrationalität eine *Ad-hoc*-Theorie ist.[21] Man könnte auch sagen: Sie führt zu einer degenerativen Problemverschiebung, weil mit ihrer Hilfe die tatsächlichen Probleme irrationalen Handelns nicht erläutert werden können, sondern nur begriffliche Spannungen innerhalb einer interpretationistischen Theorie des Geistes beseitigt werden. Mein Argument sieht etwas gerafft wie folgt aus:

(1) Es gibt Standards der Rationalität, die von denkenden Personen gutgeheißen werden, auch wenn sie diese Standards nicht formulieren können.

(2) Die Möglichkeit, dass Personen diese Standards nicht formulieren können, kann nicht implizieren, dass diese Standards den Personen kognitiv vollständig unzu-

---

17 Davidson (2006a), 326.

18 Ebd., 329.

19 Ebd., 317.

20 Siehe oben, wo vom dem *Wissen* die Rede ist, dass die Quadratur des Kreises unmöglich sei.

21 Eine ähnliche Einschätzung teilt auch Lear (1998, 83; dt. in diesem Band: 50), wenn er die Spaltungshypothese als Ergebnis konzeptueller Vorentscheidungen in der Theorie des Geistes beschreibt.

gänglich sind. Sie müssen wenigstens in der Lage sein, Fälle der Nichtbefolgung als Fehler zu identifizieren und angemessene Korrekturvorschläge zu machen, sich darüber hinaus aber auch als Adressaten dieser normativen Standards betrachten.

(3) Personen, die die in (2) genannten Fähigkeiten erfüllen, erfüllen im Großen und Ganzen auch die Rationalitätsstandards. Daher kann man sagen, dass es Standards dieser Personen sind.

(4) Fälle von Irrationalität implizieren, dass gegen mindestens einen (eigenen) Rationalitätsstandard verstoßen wurde.

(5) Der Verstoß gegen einen (eigenen) Rationalitätsstandard *R* impliziert kein Bewusstsein davon, dass gegen *R* verstoßen wurde.

(6) Gegen (eigene) Rationalitätsstandards zu verstoßen, ohne zu wissen, dass man gegen sie verstößt, ist nicht hinreichend dafür, irrational zu sein.

(7) Irrationalität impliziert, dass man ein Bewusstsein des eigenen Verstoßes gegen einen (eigenen) Rationalitätsstandard hat.

(8) Da das Bewusstsein von etwas im Interpretationismus nicht nach dem Modell der Introspektion rekonstruiert wird, sondern nach dem Modell einer Selbstinterpretation, deren Form isomorph zu rationalisierenden Fremdinterpretationen ist, ist ein Bewusstsein der eigenen Irrationalität unmöglich. Denn wenn der Zugang zu den Gehalten eigener Gedanken an aktualisierte Selbstinterpretationen gebunden ist, ist das, was dem Bewusstsein jeweils gegenwärtig ist, qua der Erfüllung von Rationalitätsstandards zugänglich.

Natürlich habe ich mit den vorstehenden Überlegungen weder gezeigt, dass es unmöglich ist, ein interpretationistisches Konzept genuiner Irrationalität zu entwickeln, noch habe ich gezeigt, dass weichere intentionalistische Programme der gleichen Kritik anheimfallen wie Davidsons Interpretationismus. So sieht es Robert Brandom, der wie Davidson ein auf die Normativität der Rationalität gestütztes Konzept des Geistes vertritt und dabei Rationalität in der Perspektive des Inferentialismus als eine bestimmte Form der Interpretierbarkeit betrachtet, als einen Vorzug seiner Konzeption an, kein derart rigides Verhältnis zwischen Intentionalität postulieren zu müssen wie die interpretationistische Position Davidsons. Brandom schreibt:

> „Rational zu sein heißt, Produzent und Konsument von Gründen zu sein, von Dingen, die sowohl die Rolle von Prämissen als auch von Konklusionen in Inferenzen spielen können. Solange wie jemand etwas behaupten (als Grund vorbringen) und etwas folgern (als Grund gebrauchen) kann, ist er rational. Die Details der einzelnen materialen inferentiellen Verbindungen, die jemand unterschreibt, affizieren den Gehalt der Sätze, die in jenen Relationen stehen, aber solange wie die Verbindungen genuin inferentiell sind, sind sie auch rational – in dem globalen Sinne, der mit lokalen Fehlern der Rationalität vereinbar ist, so dass jemand schlechte Inferenzen vollzieht oder gemessen an den inhaltskonstitutiven materialen inferentiellen Festlegungen, die für diese einzelnen Sätze gelten, unrichtig überlegt.“[22]

Doch auch wenn diese Formulierung mehr Spielraum für die Rekonstruktion kalter Formen von Irrationalität bereitstellt und sich nicht im Dickicht unterschiedlicher mentaler Netze verheddert, stellen Formen genuiner Irrationalität auch für Brandoms Position eine Heraus-

[22] Brandom (2002), 6 (meine Übersetzung).

forderung dar. Denn obgleich Personen, die genuin irrational agieren, Produzenten und Konsumenten von Gründen bleiben, sind die lokalen Fehler, die für diese Formen spezifisch sind, so beschaffen, dass die Fähigkeit mit Blick auf derartige Phänomene, Gründe zu geben oder sie als korrigierbare Fehler zu identifizieren, an eine Grenze kommt.

Was nun? Soweit ich sehe, bleibt der Hinweis auf die Möglichkeit, Aktivitäten als Handlungen *misszuverstehen*, dem ich nun genauer nachgehe, wobei ich mich bemühe, das Bild der genuinen Irrationalität etwas stärker zu konturieren.

Bisher habe ich die Möglichkeit, dass Personen ihre mit ihren Präferenzen unvereinbaren Aktivitäten als Handlungen verstehen, damit begründet, dass sie ein Bewusstsein von Alternativen zu diesen Handlungen haben. Ein anderer Zug, um die Rede von Handlungen zu rechtfertigen, besteht darin, dass die betroffenen Personen selbst solche Aktivitäten *während ihres Vollzugs* als Handlungen verstehen, weil sie vor und während des Vollzugs glauben, dass sie der Erreichung eigener Ziele dienen, *später* jedoch daran zweifeln. Verständlicher, so meine Vermutung, wird der Status irrationaler Handlungen, wenn wir das Bewusstsein von der eigenen Irrationalität in eine *prozessuale* Perspektive stellen. Etwa nach folgendem Modell:

(Phase 1) Paul tut etwas, sagen wir *T*, in der Überzeugung, dass *T* der Erfüllung eines seiner Wünsche dient. Paul erfüllt die Bedingung des Handelns, er verfügt über subjektive Gründe für *T* und sieht keine gewichtigen Gründe gegen *T*.

(Phase 2) Paul gewinnt retrospektiv die Einsicht, dass *T* regelmäßig dazu führt, dass der *T* (mit)begründende Wunsch frustriert wird.

(Phase 3) Paul weiß, dass es Alternativen zu *T* gibt.

(Phase 4) Paul gewinnt die Fähigkeit, während der Ausführung von *T* die schädlichen Folgen von *T* zu antizipieren, sodass er während der Ausführung weiß, dass *T* mit seinem Wunsch konfligiert. Dennoch führt Paul in gewissen Situationen *T* aus. Paul erfüllt die Irrationalitätsbedingung.

(Phase 5) Paul weiß nicht, warum er *T* ausführt. Paul erleidet einen reflexiven Zusammenbruch (mit Blick auf *T*).

Pauls Geist ist nicht gespalten, sondern ratlos. Er ist ratlos, weil er etwas, das er zunächst für begründet hielt, auch noch im Bewusstsein seiner Schädlichkeit und im Bewusstsein möglicher Alternativen praktiziert. Die Idee des Phasenmodells besteht darin, die subjektive Rationalitätsbedingung für Handeln und das Bewusstsein von der Inkohärenz *zeitlich* zu entkoppeln. Deshalb nehme ich an, dass es einen Zeitraum gibt, in dem Paul glaubt, Gründe für *T* zu haben. Und erst wenn das Bewusstsein der Inkohärenz auftritt, wird *retrospektiv* die Rationalitätsbedingung des Handelns in Zweifel gezogen.

Die Frage ist nun, wodurch *T* verursacht wird und warum das, was *T* verursacht hat, mit einem Grund verwechselt werden konnte. Meine Vermutung geht dahin, dass der Zustand, der *T* verursacht, retrospektiv als ein Zustand verstanden werden muss, der *T auf Grund eines Gehalts* verursacht hat, diesen Gehalt aber nicht inferentiellen Relationen, *sondern funktionalen psychischen Mechanismen verdankt*. Die Paradoxien der Irrationalität, so vermute ich weiter, tauchen nur deshalb auf, weil Davidson lediglich einen einzigen Mechanismus – nämlich rationalisierende Interpretation – kennt, der mentale Zustände mit einem Gehalt ausstatten kann. Wir können nun hoffen, die Paradoxien aufzulösen, indem wir einen weiteren derartigen Mechanismus ins Spiel bringen.

Ohne hier auf Details der Freud-Interpretation einzugehen, lässt sich auch ein psychoanalytisches Verständnis der Situation dadurch beschreiben, so jedenfalls meine These, dass

Pauls Tun durch einen Zustand verursacht wird, der aus Pauls Perspektive den Status eines Grundes hatte, in Wahrheit aber ein gehaltvoller Zustand ist, dessen Gehalt sich jedoch nicht der rationalen Selbstinterpretation, sondern funktionalen repräsentationalen Mechanismen verdankt. Unter einer funktionalen Beschreibung könnte sich Paul beispielsweise in einem Zustand befinden, solche Beziehungsmuster zu suchen, deren Struktur ihm – im Unterschied zu nichtmanipulativen Liebesbeziehungen – gut vertraut ist; zugleich könnte er diesen Beziehungswunsch gewissermaßen abstrakt als Wunsch nach einer *guten* Beziehung verstehen. Pauls Geist ist also nicht gespalten, sondern beherbergt unter einer intentionalistischen Beschreibung psychische Zustände, die vom rationalen Lernen im gewöhnlichen Sinne abgekoppelt sind und neurotisches Handeln verursachen können.

Irrationalität erweist sich jetzt als ein Durchgangsstadium zwischen einer Phase subjektiv intakter Selbstbeschreibung als ein Handelnder, der seine Handlungen zu begründen können glaubt, und einer Phase, in der diese Selbstbeschreibung partiell, mit Blick auf gewisse Tätigkeiten, zusammengebrochen ist. Innerhalb dieses Durchgangsstadiums, über dessen Dauer ich nichts sage, treffen die gründegestützte Beschreibung einer Handlung und das Bewusstsein ihrer faktischen Unbegründetheit aufeinander. Wird dieser Zustand überwunden, stellt sich die anfangs scheinbar intakte Rationalisierung als *Pseudo*rationalisierung heraus – und die vermeintliche Handlung als ein psychisch verursachtes Tun. In dieser Perspektive verwandelt sich die Frage, wie Irrationalität möglich ist, in die Frage, wie das Selbstmissverständnis, wie also Pseudorationalisierung möglich ist, und wie sich nicht nur dieses Selbstmissverständnis, sondern auch die Disposition, *T* auszuführen, überwinden ließe. Auf beide Fragen möchte ich eine Antwort geben, und zwar im Rückgriff auf ein funktionalistisches Verständnis der Psyche, das ich im folgenden Abschnitt wenigstens skizziere.

II. Von der Psyche gestört

Wie ist es möglich, dass ein mentaler Zustand als intentionaler Zustand, etwa als ein Grund, missverstanden wird? Um diese Idee zunächst einmal zu *plausibilisieren*, kann man auf folgendes Faktum hinweisen: Unser Geist ist nicht von Anfang an ein (alles in allem) rational strukturiertes Netz propositionaler Einstellungen. Dennoch können wir uns (wie ja auch Tiere), noch bevor wir einen solchen Geist entwickelt haben, in der Welt orientieren und etwa Bezugspersonen dazu veranlassen, Dinge für uns zu tun, die wir selbst nicht tun können. Ich will nun hier keine detaillierte Entwicklungsgeschichte unseres Geistes erzählen, sondern nur auf einen einzigen Aspekt dieser Geschichte hinweisen, der für die Lösung des Irrationalitätsproblems von Bedeutung ist.

Bevor wir rationale Wesen sind, orientieren wir uns in der Welt mithilfe repräsentationaler Zustände, die Produkte von Mechanismen sind, denen wir auf Grund ihrer evolutionären Geschichte die Funktion zuschreiben können, Weltzustände (und eigene Zustände) angemessen zu repräsentieren. Diese Annahme scheint nicht nur mit Blick auf die Fähigkeiten vorsprachlicher Kinder angemessen zu sein, sondern auch von grundlegender Bedeutung für die Möglichkeit, dass wir uns im Rahmen eines Lernprozesses überhaupt zu rationalen Selbstinterpreten entwickeln können. Befinden wir uns während dieser Phase in repräsentationalen Zuständen, und erfüllen die sie produzierenden Mechanismen ihre Funktion, dann haben diese Zustände Gehalte, die sich nicht rationalisierender Selbstinterpretation verdanken.

Wenn man annimmt, dass wir mit repräsentationalen Mechanismen ausgestattet sind, bevor wir uns zu Wesen entwickeln, die die Fähigkeit erwerben, Äußerungen hervorzubringen, die für sie selbst durch antizipierte Fremdinterpretation, also durch Selbstinterpretation Bedeu-

tung haben, warum sollte man dann annehmen, dass die repräsentationalen Mechanismen verschwinden, sobald sie ihren Beitrag zur Entfaltung unserer medialen und sprachlichen Fähigkeiten geleistet haben? Haben wir nicht vielmehr auch empirisch Grund zu der Annahme, dass in Situationen, in denen uns etwa infolge eines Schocks die so genannten höheren kognitiven Fähigkeiten abhanden gekommen sind, weiterhin Orientierungsfähigkeiten zur Verfügung stehen, die ganz offenbar auf grundlegende repräsentationale Mechanismen zurückgreifen? Wenn solche Annahmen begründet sind, dann geht es eher darum zu bestimmen, wie sich intentionale und repräsentationale Zustände als Aspekte unseres mentalen Lebens zueinander verhalten. Folgende Überlegung macht einen Anfang: Wenn wir in einer rationalisierenden oder alltagspsychologischen Perspektive davon reden, dass eine Person eine Überzeugung hat, dann können wir das schematisch so ausdrücken: „*S* ist überzeugt, dass *p*". Diese Formulierung dürfen wir sicher in folgende, etwas gestelzte Variante überführen:

(AP) „*S* befindet sich in einem Zustand des Überzeugtseins, dass *p* der Fall ist."

Einen Zustand hingegen, der seinen Gehalt *nicht* seiner inferentiellen Position, sondern repräsentationalen Mechanismen verdankt, würden wir in der Perspektive einer funktionalen Semantik so beschreiben[23]:

(FS) „*S* befindet sich in einem Zustand, der repräsentiert, dass *p* der Fall ist."

Es ist keine atemberaubende Annahme, dass ein Zustand, in dem sich eine Person befindet, sowohl mithilfe von (AP) als auch mithilfe von (FS) beschrieben werden kann. Allerdings bringen die beiden Beschreibungen natürlich unterschiedliche Voraussetzungen ins Spiel. (AP) sollte zum Beispiel ausschließen, dass *S* zugleich auch glaubt, dass $\neg p$ der Fall ist, beziehungsweise etwas glaubt, das mit *p* unvereinbar ist. (FS) wiederum lässt etwa zu, dass *S* kein Bewusstsein vom Gehalt der Repräsentation hat, während das Überzeugtsein sogar im Falle latenter Überzeugungen wenigstens die Bewusstseinsfähigkeit der Gehalte voraussetzt.

In welchem Verhältnis stehen AP- und FS-Formulierungen also zueinander? Lassen sich AP- und FS-Formulierungen ineinander übersetzen? Wenn ich nicht irre, ist genau das *nicht* der Fall. Ohne diese Auffassung hier zu begründen, gehe ich davon aus, dass jeder Zustand des Überzeugtseins als ein indikativischer repräsentationaler Zustand beschrieben werden kann, aber nicht jeder indikativische repräsentationale Zustand ist auch ein Zustand des Überzeugtseins. Damit ein indikativischer repräsentationaler Zustand einer des Überzeugtseins ist, müssen vielmehr Bedingungen erfüllt sein, die über die Bedingungen für repräsentationale Zustände *hinausgehen*. Und diese Bedingungen sind genau jene ihrer alltagspsychologischen Verstehbarkeit, also Bedingungen der inferentiellen Kohärenz. Mit anderen Worten und etwas allgemeiner: Die mentalen Zustände, von denen alltagspsychologische Formulierungen reden, bilden eine echte Teilmenge jener repräsentationalen Zustände, von denen eine funktionale Semantik handelt.

Wenn man das bisher skizzierte alternative Verständnis von Irrationalität mit einer theoretischen Grundlage ausstatten will, dann muss man die Möglichkeit in Betracht ziehen, dass sich erwachsene Menschen zu einem Zeitpunkt *t* in mentalen Zuständen befinden können, deren Gehalte durch unterschiedliche Mechanismen festgelegt werden. Die Einbeziehung funktional individuierter repräsentationaler Zustände ist natürlich geeignet, antinaturalistische Bedenken zu motivieren. Ich möchte daher wenigstens andeuten, warum ich glaube, dass diese Bedenken

[23] Eine ausgearbeitete funktionalistische Semantik liegt in Form von Ruth Millikans Teleosemantik vor; vgl. Millikan (1984).

nicht berechtigt sind. Dazu werde ich die Hinsicht, in der ein intentionaler Zustand spezifischer ist als ein repräsentationaler, als eine *historische* Hinsicht beschreiben. Die Idee ist einfach.

Als kleine Kinder sind wir weder rationale Selbstinterpreten, noch können wir anderen alltagspsychologische Zustände zuschreiben. Gleichwohl können wir uns in der Welt orientieren, haben Protoerwartungen und Protowünsche. Diese Fähigkeiten gehen auf repräsentationale Zustände zurück, die wir vorläufig etwa mit Searle über ihre Erfüllungsbedingungen individuieren können.[24] Nur dann, wenn Wesen, die auf Grund der Evolutionsgeschichte mit solchen funktionalen Mechanismen ausgestattet sind, in die Praktiken des rationalen Interpretierens oder des Verlangens und Gebens von Gründen sozialisiert werden, entwickeln sie sich zu Wesen, die ihr eigenes Tun und das anderer im Lichte von Gründen verstehen und prognostizieren. Der Lernprozess, den sie dabei absolvieren, ist ein individualgeschichtlicher Prozess, in dem Gehalte funktionaler Zustände an weitergehende, spezifischere Bedingungen geknüpft werden – Bedingungen nämlich, die durch die sozial etablierte Interpretationspraxis ins Spiel kommen.

Die Grundidee ist also, dass funktionale Zustände solche physikalischen Zustände sind, die im Kontext einer *Evolutionsgeschichte* stehen, während intentionale Zustände solche funktionalen Zustände sind, die im Kontext einer *Interpretationsgeschichte* stehen. Daraus folgt – wie ich glaube –, dass funktionale nicht auf physikalische und intentionale nicht auf funktionale Zustände reduziert werden können. Denn die spezifischen historischen Eigenschaften, die physikalische Zustände zu funktionalen oder intentionalen machen, sind mit Mitteln der Physik beziehungsweise der Biologie nicht zu erläutern.

Welches Bild ergibt sich nun, wenn wir diese Überlegungen auf das Problem der Irrationalität zurückbeziehen? Die These war ja, dass Paul einen mentalen Zustand, der faktisch ein repräsentationaler Zustand ist, als einen alltagspsychologischen oder intentionalen Zustand *miss*versteht. In der alltagspsychologischen Beschreibung ist das jener Zustand, in dem er davon überzeugt ist, dass seine Handlung *T* der Erfüllung seiner Wünsche dient. Wenn wir diesen Zustand als repräsentationalen Zustand beschreiben, haben wir mehr Spielraum, seinen Gehalt zu bestimmen. Denn es reicht hin, wenn wir ihn als einen beschreiben, der *T* als erstrebenswerte Handlung repräsentiert. Dass *T* als erstrebenswert erscheint, muss jedoch nicht im Rückgriff auf Wünsche und Überzeugungen erläutert werden, weil der inferentielle Zusammenhang für diesen repräsentationalen Gehalt nicht notwendig ist. Es reicht vielmehr aus, wenn *T* auf Grund funktionaler psychischer Bewertungsmechanismen als erstrebenswert erscheint.

Wie kommt es aber dazu, dass Menschen solche Zustände als intentionalistisch zu beschreibende Gedanken missverstehen? Und wie kommt es zu Pseudorationalisierungen? Ziehen wir, um Paul nicht über Gebühr zu strapazieren, den Fall von Gisela heran: Gisela plagt sich mit etwas herum, das in der Psychologie als Zwangshandlung beschrieben wird. Wenn Gisela ihre Wohnung verlässt, dann kontrolliert sie, ob der Herd und die letzte Zigarette aus sind, und schließt die Wohnungstür ab. Auf halbem Treppenabsatz macht sie jedoch kehrt, weil sie nicht sicher ist, ob sie Herd und Aschenbecher tatsächlich aufmerksam überprüft hat oder in ihren Gedanken nicht doch schon zu sehr bei dem Vortrag war, den sie in Berlin halten will. Sie schließt die Wohnung auf, prüft alles noch einmal und verschließt die Wohnung erneut. Da sie inzwischen mit der Angst kämpft, ihren Zug zu verpassen, zweifelt sie im Treppenhaus jedoch auch an dieser Prüfung, und so weiter und so fort.

Die erste Prüfung ist rational. Wenn man die Wohnung für längere Zeit verlässt, ist es klug, bestimmte Gefahren auszuschließen. Auch die zweite Prüfung kann man noch als rational beschreiben, denn abgelenkte Kontrolleure übersehen genauso leicht etwas, wie Autoren, die

[24] Vgl. etwa Searle (1983).

mit ihren Gedanken beschäftigt sind, Tippfehler übersehen. Das allerdings kann für jede weitere Prüfung auch geltend gemacht werden, sodass man kein Ende finden würde, insbesondere wenn die ablenkende Gefahr droht, den Zug nach Berlin zu verpassen. Wenn es auf Grund des Kontrollzwangs immer wieder dazu kommt, dass Gisela wichtige Absichten nicht realisieren kann, und sich herausstellt, dass sie eigentlich nie etwas übersieht, dann ist die zweite, mindestens aber jede weitere Prüfung jedoch irrational. Diese Prüfungen sind Handlungen, die dazu führen, dass wichtige Präferenzen Giselas immer wieder frustriert werden. Wenn man Gisela fragt, warum sie so oft kontrolliert, dann kann sie geltend machen, dass sie bei der ersten Prüfung abgelenkt war, bei der zweiten unter Stress stand und so weiter. Sie selbst muss sich erklären, was sie da tut, und sie wird solche Erklärungen nicht ohne Widerstand aufgeben, hängt doch für sie selbst von derartigen Erklärungen der Status ab, eine Person zu sein, *die sich selbst versteht*. Würden wir diese Perspektive leichtfertig aufgeben, dann würden wir uns selbst fremd werden. Um das zu vermeiden, sind Pseudorationalisierungen, jedenfalls auf den ersten Blick, ein probates Mittel. Sie sind keine Form, in der wir uns einfach in die Tasche lügen, denn die Erklärungen, die wir uns geben, stehen so lange ganz gut da, wie wir einfach keine anderen Erklärungen haben.

Auf *andere* Erklärungen könnte Gisela stoßen, wenn sie etwa im Rahmen einer Psychoanalyse herausfinden würde, unter welchen Bedingungen sie geneigt ist, sich selbst nicht zu vertrauen, auf welchen repräsentationalen Zustand ihre wiederholten Prüfungen reagieren und welche psychische Funktion die Zwangshandlung hat. Dabei könnte sich herausstellen, dass ein früher Fall von Zutrauen in die eigenen Fähigkeiten schlimme Folgen hatte, aber auch, dass es Gisela schwerfällt, sich von dem vertrauen Kontext ihrer Wohnung zu lösen, und der Zwang die Funktion hat, den damit verbundenen Schmerz nicht bewusst werden zu lassen. Einerlei, was dabei herauskommt: Giselas Irrationalität ist ein Fehler im Haus der Vernunft, seine Ursache jedoch liegt außerhalb dieses Hauses. Sie liegt in repräsentationalen Zuständen, die Gisela in ihrem inneren Monolog *fälschlich* als Gründe beschreibt.

Es ist interessant, dass Davidson seine eigenen Überlegungen so versteht, dass sie plausibel machen, warum sich Freud sowohl der physikalischen als auch der intentionalistischen Sprache bedient hat. Das Entscheidende an Freuds Ausdrucksweise ist ihm – wie übrigens auch Habermas, der bekanntermaßen von einem szientistischen Selbstmissverständnis der Psychoanalyse spricht[25] – meines Erachtens jedoch entgangen. Zweifelsohne stand Freud vor dem Problem, die Psychoanalyse mit einer Sprache auszustatten, die in seinen Ohren, also in denen eines gelernten Naturwissenschaftlers, einen wissenschaftlichen Klang hat. Doch seine Metaphern, die – wie Davidson richtig sagt – der Hydraulik, dem Elektromagnetismus, der Neurologie und der Mechanik entstammen, sind nicht Ausdruck der Tatsache, dass Freud anomaler Monist war und wie Davidson glaubte, dass Gedanken physikalisch realisiert werden, sodass die Rede von Gründen und die von Ursachen gleichermaßen im Spiel sein müssen.[26] Ich glaube vielmehr, dass wir Freud besser verstehen und seinen Absichten eher gerecht werden, wenn wir seine Metaphern als Ausdruck eines *funktionalistischen* Verständnisses der Psyche deuten, das noch nicht auf ein ausgereiftes funktionalistisches Vokabular zurückgreifen konnte, wie es inzwischen im Kontext der Biologie und der Philosophie der Biologie entwickelt worden ist.

Wenn Freud also vom *Sinn* eines Symptoms redet, dann hat er – so meine These – weder eine bloß physikalisch beschreibbare kausale Struktur vor Augen noch das Ergebnis einer Interpretation, das am Ende einer rationalisierenden oder hermeneutischen Anstrengung steht. Der Sinn eines Symptoms ist vielmehr seine *psychische Funktion*. Und in diesem Sinne sind

---

25 Vgl. Habermas (1981), 263 u. 300 ff.

26 Vgl. Davidson (2006), 289–292.

Ausdrücke wie „Verdrängung“, „Widerstand“, „Fixierung“ oder „Zensur“ weder intentionalistische noch physikalistische Begriffe. Sie sind vielmehr funktional bestimmt.[27]

Davidsons Überlegungen sind in einer weiteren Hinsicht phänomenologisch unangemessen. Denn im Gegensatz zu Pauls und Giselas Problemen sind die irrationalen Handlungen von Davidsons Akteuren merkwürdig episodisch. Ein Mann räumt im Park einen Ast aus dem Weg und entschließt sich später, den Ast dorthin zurückzulegen, weil von diesem auch in seiner neuen Position eine Gefahr ausgeht. Dann jedoch findet er, dass die Mühe des Zurücklegens nicht gerechtfertigt ist, geht gleichwohl aber in den Park und legt den Ast zurück.[28]

Interessant ist, dass Davidson in seiner Darstellung diesen Fall, den er sich bei Freud ausleiht, aller klinischen Konnotationen beraubt – zumindest aber kein Aufhebens davon macht, dass hier ein komplexes psychisches Problem vorliegen könnte. In Freuds Schilderung ist der Mann ein *Zwangsneurotiker*, und sein Bericht macht deutlich, dass die zweite Position des Astes deutlich ungefährlicher ist als die erste, sodass der Mann eine Situation wiederherstellt, die von jedermann als die gefährlichere betrachtet werden würde. Freud schreibt: „Die zweite feindselige Handlung [nämlich den Ast auf den Weg zurückzulegen], die sich als *Zwang* durchsetzte, hatte sich […] dem bewußten Denken [gegenüber] mit der Motivierung der ersten, menschenfreundlichen [Tat] geschmückt.“[29]

Dass Freuds Bericht den Zwang ins Spiel bringt, macht darauf aufmerksam, dass genuin irrationale Handlungen keine Handlungen sind, die uns, wie Irrtümer, gelegentlich unterlaufen, sondern Handlungen, die in gewissen Kontexten hartnäckig und systematisch auftreten. Wenn wir Paul oder Gisela auf die Probleme ihres Handelns hinweisen würden, dann sollten wir nicht damit rechnen, dass sie sagen: „Stimmt, du hast Recht, das sollte ich in Zukunft anders machen.“ Und selbst wenn sie es *sagen* würden, wird man gut daran tun, kein Vermögen darauf zu verwetten, dass sie zukünftig anders *handeln* werden.

Freuds Rede vom Zwang weckt natürlich neuerlich Zweifel daran, dass es sich bei den bewussten Aktivitäten überhaupt um *Handlungen* handelt. Man könnte vielmehr denken, dass eine „Zwangshandlung“ ein hölzernes Eisen ist, und fragen wollen, ob Fälle, in denen Zwang eine große Rolle spielt, nicht genau dem unterliegen, was Davidson das „Medea-Prinzip“ nennt, und daher gerade keine Fälle von Irrationalität darstellen. In solchen Fällen liegt Davidson zufolge kein irrationales Handeln vor, weil ein Ereignis x zwar als Ursache, nicht aber als ein (hinreichender) Grund für die Erklärung eines Tuns T auftritt. Weil x in solchen Erklärungen aber gar nicht als Grund für T betrachtet wird, hat das Medea-Modell gar keine begrifflichen Ressourcen dafür, x oder T als „irrational“ zu bewerten, die als bloße physikalische Ereignisse „arational“ sind.

Vielleicht lässt sich der Zweifel, dass Zwangshandlungen keine Handlungen sind, sondern dem Medea-Prinzip unterliegen, nicht endgültig und für alle Fälle beschwichtigen.[30] Immerhin aber könnte man geltend machen, dass die Handelnden auch in Davidsons Szenarien

---

27 Sartres Kritik am Begriff der Zensur macht deutlich, dass die Arbeit der intrapsychischen Zensur nur um den Preis von Paradoxien in der intentionalistischen Sprache beschrieben werden kann; vgl. Sartre (1976), 95–100.

28 Ein anderer Fall, den Davidson diskutiert, ist ein Fall von Wunschdenken hinsichtlich schöner Waden.

29 Freud (1999), 414 f., Fn. 3.

30 Einen weiten Begriff arationalen, gleichwohl intentionalen Handelns versucht Rosalind Hursthouse (1991) zu verteidigen. Wenn Handlungen aus Emotionen – zumal aus unmittelbar aktivitätssteuernden – als Handlungen zählen sollen, dann nicht, weil sie die Handlungen begründen, sondern weil sie in Kontexten stehen, die von Gehalten strukturiert werden, die sie artikulieren. Auch hier hängt viel davon ab, wie „Intentionalität“ erläutert wird.

überlegen, ob sie so oder anders handeln sollen, dass sie die Handlung also in den Kontext praktischen Räsonierens stellen, Alternativen und Begründungen erwägen und die Zwangshandlung nicht immer vollziehen – es gibt auch bei Gisela Tage, an denen sie ihre Wohnung nach einmaliger Kontrolle verlassen kann.

Ein Weg, auf dem man diesem Zweifel nachgehen kann, ist vielleicht der, dass man, auch angesichts der Tatsache, dass genuin irrationales Handeln oft hartnäckig ist, davon ausgeht, dass einige der mentalen Zustände, die an ihrer Verursachung beteiligt sind, einen anderen Status haben als gewöhnliche alltagspsychologisch beschreibbare Absichten, Wünsche oder Überzeugungen. Während wir nämlich gewöhnliche Überzeugungen im Nu modifizieren können, wenn wir dazu Gründe haben, scheinen die mentalen Zustände, die an der Formierung genuin irrationalen Handelns beteiligt sind, in bestimmten Hinsichten gegen die Kraft von Gründen immun zu sein. Wenn jemand zum Beispiel tatsächlich versteht, dass und warum es unmöglich ist, den Kreis zu quadrieren, dann wird es ihm, abgesehen vielleicht von einer gewissen Enttäuschung darüber, dieses Projekt nicht verwirklichen zu können, nicht schwerfallen, entsprechende Versuche aufzugeben. Aber selbst wenn Gisela einsieht, dass die Kontrolle der Kontrolle ab einem bestimmten Punkt sinnlos, ja sogar schädlich für sie ist, wird es ihr dennoch gerade nicht leichtfallen, auf diese Einsicht handelnd zu reagieren.

Eine nahe liegende Möglichkeit, den besonderen Status der repräsentationalen Zustände aufzuklären, die an der Formierung von Zwangshandlungen beteiligt sind, besteht darin, anzunehmen, dass ihr Gehalt den Handelnden *nicht bewusst* ist. Dann aber würden wir zugleich auch die Möglichkeit verlieren, zu erklären, wie Pseudorationalisierungen möglich sind. Wäre die Ursache des irrationalen Handelns nämlich vollständig unbewusst, verlören wir einen Gegenstand, auf den sich der Geist rationalisierend beziehen könnte. Im Rahmen dieser Analyse würden wir auf das zurückgeworfen werden, was Davidson als Aristoteles' Analyse der Irrationalität betrachtet[31]: Nehmen wir an, x sei ein Grund für T, y sei ein Grund gegen T und x und y seien unvereinbar. Wenn wir weiter annehmen, dass der Person P y nicht mehr gegenwärtig ist, dass sie y also vergisst oder kein Bewusstsein mehr von y hat, dann können wir erklären, dass P T vollzieht, obwohl T mit y unvereinbar ist. Doch gemäß dieser Analyse verlieren wir die Möglichkeit, P in dem Sinne als irrational zu beschreiben, dass P die Inkohärenz seines Handeln bewusst wird.

Sinnvoller erscheint mir folgende Überlegung: Der repräsentationale Zustand, den die irrationale Person als Grund missversteht, ist der Person *partiell* bewusst. Wir könnten beispielsweise annehmen, dass Gisela Szenarien im Kopf hat, die sie unbedingt vermeiden möchte: Bilder ihrer Wohnung in Flammen etwa. Diese Bilder interpretiert Gisela als Gründe für Handlungen, die die imaginierte Katastrophe vermeiden sollen. Dabei ist jedoch klar, dass die imaginierten Szenarien nicht in das Netz ihrer inferentiell gegliederten intentionalen Zustände integriert sind und daher nicht den Status von Gründen haben. Es ist nicht einmal klar, *warum* diese Szenarien in gewissen Situationen auftauchen. Ist man jedoch mit solchen Vorstellungen konfrontiert, kann es schwer sein, sie zu ignorieren, und es ist bei weitem attraktiver, handelnd auf sie zu reagieren, was sie jedoch in den Kontext der mentalen Zustände integriert, die die Handlung begründen. Gerade weil Gisela keine Einsicht in die psychische Funktion ihrer Phantasien hat, steht sie unter dem Druck, auf die Bilder in ihrem Kopf im Rahmen ihres Überlegens zu reagieren, sodass das Szenario in Form eines Grundes in ihr Räsonieren eingeht und sich dabei – wie Freud sagt – gegenüber „dem bewussten Denken" mit dem Status eines Grundes „schmücken" kann.

---

31 Vgl. Davidson (2006), 295.

Fälle wie diese scheinen mir die wirklich herausfordernden Fälle von Irrationalität zu sein, und ein philosophischer Kommentar muss sich an ihnen bewähren. Fassen wir den Stand der Überlegungen zusammen. Davidsons Modell für Irrationalität lässt sich folgendermaßen darstellen:

(DI)

(1) Sei *x* eine Ursache für *T*;

(2) sei *x* in N_1 zugleich ein Grund *für T*;

(3) sei *x* in N_2 zugleich ein Grund *gegen T*;

(4) wobei N_1 und N_2 „zwei halbautonome Abteilungen des Geistes" seien;

(5) und sei *T* rational relativ zu N_1;

(6) dann ist *T* irrational relativ zu N_2

(7) ergo: Irrationalität ist die Folge einer Spaltung des Geistes in halbautonome Netze.

Im Mittelpunkt meiner Kritik an diesem Modell stehen folgende Überlegungen:

(KD)

(1) Spaltung macht ein Bewusstsein von Inkohärenz unmöglich.

(2) Ein Bewusstsein von Inkohärenz ist laut Davidson jedoch notwendig für Irrationalität.

(3) Ein Bewusstsein der Inkohärenz setzt unter Bedingungen des Spaltungsmodells ein integratives Netz voraus, das N_1 und N_2 umfasst.

(4) Dann aber ist fraglich, wie N_1 und N_2 für Inkohärenz hinreichend verschieden und zugleich *Sub*netze eines durch rationale Selbstinterpretation aufgespannten integrativen Netzes sein können.

Mein bisher skizzierter Vorschlag liefe auf dieses Bild irrationalen Handelns hinaus:

(VI)

(1) Ein mentaler Zustand *x* ist *Ursache* für ein Tun *T* einer Person *P*;

(2) *x* wird von *P* als Grund für *T missverstanden*;

(3) tatsächlich ist *x* aber kein Grund für *T*, sondern ein Zustand, der die Ausführung von *T* auf Grund seines repräsentationalen Gehalts *motiviert*;

(4) in Situationen eines bestimmten Typs vollzieht *P T* immer wieder, obwohl *P* weiß, dass es bessere *Alternativen* zu *T* gibt, und es *P* manchmal gelingt, eine dieser Alternativen zu vollziehen;

(5) *T* ist irrational relativ zum Netz der intentionalen Zustände von *P*, mit anderen Worten: *P* versteht nicht, warum sie *T* vollzieht;

(6) der Gehalt von *x* verdankt sich *funktional* zu beschreibenden *repräsentationalen Mechanismen* und geht *partiell* in das Verständnis von *x* als Grund für *T* ein;

(7) ergo: Irrationalität ist die Folge einer Störung des Geistes durch repräsentationale Zustände, die als Gründe missverstanden werden.

(VI) stellt ein vorläufiges Verständnis von Irrationalität dar, das auf einer Verschränkung funktionalistischer und intentionalistischer Motive fußt. Es ist klar, dass diese Verschränkung vor der Aufgabe steht, die Ursache irrationalen Handelns nicht so weit der intentionalistischen Sprache zu entfremden, dass die irrationale Aktivität den Status einer Handlung

verliert. Zunächst kann man aber festhalten, dass genuin irrationale Handlungen, also Handlungen, die das Selbstverständnis der Akteure lokal unterlaufen, über folgende Eigenschaften verfügen, die sie von blankem Verhalten unterscheiden:

(E)

(1) Genuin irrationale Handlungen T^i treten in Kontexten auf, in denen die Akteure praktisch überlegen, wie sie handeln sollen, und dabei bilden sie Absichten des Typs ‚ich sollte (werde, muss) T^i tun'.

(2) Es ist möglich, dass die Akteure T^i im Rahmen des Überlegungsprozesses gegen Alternativen abwägen, die ihnen tatsächlich auch offenstehen.

(3) Wegen (1) betrachten die Akteure selbst die Ausführung von T^i nicht als Widerfahrnis, wie etwa einen Reflex, sondern als von ihnen selbst gewähltes Agieren.

(4) Wenn die Akteure ein (partielles) Bewusstsein von den gehaltvollen psychischen Zuständen haben, die an der Formierung der Absicht, T^i auszuführen, beteiligt sind, dann stehen ihnen diese Zustände nicht als fremde gegenüber, sondern werden als Teil der eigenen Person betrachtet.

(5) Der lokale reflexive Zusammenbruch wirft für die genuin irrational Agierenden nicht die Frage auf, ob sie überhaupt Personen mit inferentiell gegliederten Absichten und Überzeugungen sind; er setzt vielmehr voraus, dass die selbstinterpretativen reflexiven Kompetenzen soweit intakt bleiben, dass T^i (sowie die T^i motivierenden Zustände) auf das Netzwerk intentionaler Zustände bezogen werden können.

(6) Es ist möglich, die Gehalte der T^i motivierenden repräsentationalen Zustände einem möglicherweise aufwendigen Verstehensprozess (etwa einer Psychoanalyse) funktional verständlich zu machen und propositional zu artikulieren.

Das allgemeine Bild genuiner Irrationalität, das sich hier abzeichnet, teilt mit Davidson die Vorstellung, dass Irrationalität ein Fehler *im* Haus der Vernunft ist, dies allerdings nur in dem eingeschränkten Sinne, dass sich der Fehler im Haus der Vernunft *zeigt*. Anders als Davidson, der in diesem Haus zwei Zimmer einrichtet, in denen Zustände derselben Art wohnen, wobei Bewohner des einen Zimmers auf Bewohner des anderen Einfluss nehmen, ohne dabei Gründe in diesem anderen Zimmer zu sein, rechne ich nur mit einem Zimmer, in dem sich zwei Arten von Bewohnern aufhalten: repräsentationale Zustände, die inferentiell organisiert und propositional artikuliert sind, und solche, die es nicht sind, die aber gleichwohl Gehalte mit motivationaler Kraft haben. Da diese Bewohner zugleich die Kriterien von (E) erfüllen, sind sie Teil unseres funktionale Mechanismen umfassenden mentalen Lebens, und nicht bloß neuronale Zustände, die als physikalistische Eindringlinge im Haus der Vernunft ihr Unwesen treiben.[32]

---

[32] Der Unterschied zwischen propositional differenzierten und funktional individuierten mentalen Zuständen fällt dabei nicht mit dem Unterschied zwischen bewussten und unbewussten Zuständen zusammen. Thomas Nagels Überlegungen, die das Spezifische psychoanalytischen Verstehens darin sehen, dass irrationale Handlungen unter Heranziehung von mentalen Zuständen rationalisiert werden, die sich von gewöhnlichen nur durch ihre Unbewusstheit unterscheiden (vgl. dazu Nagel 1994, 42), machen deutlich, dass das Spezifische funktionaler psychischer Zustände auf diese Weise nicht erfasst werden kann. Denn verhielte es sich so, müsste Analysanden ihr ehedem irrationales Verhalten nach einer erfolgreichen Analyse als rational erscheinen. Eine übertriebene Kontrolle der Wohnung wird nicht rational dadurch, dass man unbewusste mentale Zustände ins Spiel bringt. Worum es in diesem Falle tatsächlich ginge, hätte unter anderem den Status einer funktionalen Einsicht in das irrationale Tun.

III. Irrationalität und die Schwächung des Willens

Während Wunschdenker ein aktives Verhältnis zu ihren mentalen Zuständen einnehmen müssen, um dafür zu sorgen, etwas zu glauben, das nicht durch alle verfügbaren Belege bestätigt wird oder mit dem unvereinbar ist, was durch alle verfügbaren Belege bestätigt wird, ist Akrasia eine passive Form der Irrationalität, für deren Erklärung wir nicht unterstellen müssen, dass der akratisch Agierende aktiv auf Formierungsprozesse mentaler Zustände einwirkt. Willensschwäche, so möchte ich sagen, widerfährt Handelnden. Nehmen wir den Fall von Fritz: Fritz muss am nächsten Tag einen wichtigen Vortrag über das Problem der mentalen Verursachung halten, vor dem ihm in Erwartung eines klugen Publikums ein wenig graut. Fritz ist der Überzeugung, dass das Beste, was er in seiner Lage tun kann, darin besteht, den Vortrag besonders gründlich vorzubereiten und alle verfügbare Zeit dafür zu nutzen. Dennoch verbringt Fritz einen beträchtlichen Teil der zur Verfügung stehenden Zeit damit, sich historische Bundestagsdebatten im Fernsehen anzuschauen. Welche Möglichkeiten haben wir, um zu verstehen, warum Fritz nicht das tut, was nach seiner eigenen Auffassung das Beste wäre? Folgende Möglichkeit scheidet aus: Fritz liegt tatsächlich nicht viel an dem Vortrag, und die Möglichkeit, sich mit der Debattengeschichte der Bundesrepublik vertraut zu machen, ist ihm eigentlich wichtiger. Allgemeiner: Das akratische Handeln kann nicht Ausdruck einer bewussten Präferenz sein, die als relevanter Faktor in Fritz' Abwägungsprozess eingegangen ist. Wäre dies der Fall, läge schlicht kein Fall von Willensschwäche vor. Fritz täte einfach das, was im Lichte seiner reflektierten Präferenzen auch zu tun wäre.

Akratisches Handeln kann daher nur zu Stande kommen, wenn es Faktoren gibt, die nicht auf den Prozess der Absichtsbildung einwirken, sondern auf die Fähigkeit, eine rational gebildete Absicht handelnd zu realisieren. Solche Faktoren können aber keine propositional differenzierten bewussten Zustände sein, weil diese als verfügbare Zustände jederzeit in einen umfassenden Prozess der Formierung der Absicht hätten eingehen können. Wir können also nicht sagen: *H* ist diejenige Handlung, die im gegenwärtigen Kontext *K* angesichts der verfügbaren relevanten Gründe die beste ist, aber Fritz vollzieht *H* nicht, weil er wichtige Gründe gegen die Ausführung von *H* hat. Wenn *H* in *K* nicht vollzogen wird, dann muss es dafür Ursachen geben, die gerade nicht die Form von Gründen haben. Potenzielle Faktoren sind: die Macht der Gewohnheit, Müdigkeit, Krankheit, sexuelle Attraktion, Heißhunger und Begierden.[33]

Derartige Faktoren werfen wiederum die Frage auf, inwiefern wir es hier überhaupt mit einem Fall von Irrationalität zu tun haben. Denn Davidson würde an diesem Punkt doch wohl einwenden, dass solche Fälle entweder dem Medea-Prinzip unterliegen oder unter die aristotelische Akrasia fielen. Ein willensschwaches Tun hätte somit im ersten Fall arationale Ursachen, weil es nicht von Gründen verursacht wäre, und im zweiten Fall, weil die eigentlichen Absichten zum Zeitpunkt des akratischen Tuns bereits in Vergessenheit geraten wären. Was kann die Rede von repräsentationalen Zuständen zur Lösung dieses Problems beitragen?

Nehmen wir an, Fritz' oben skizzierte Situation sei durch zusätzliche psychische Randbedingungen gekennzeichnet: Es ist der vierte Vortrag in zwei Wochen, den er zu halten hat, Fritz ist mit Blick auf seinen Ansatz von großen Zweifeln geplagt; und ein anstrengendes Semester hat ihn erschöpft. Kurz: Fritz ist die Arbeit an seinem Vortrag mindestens unangenehm. Wenn ihm diese Tatsachen und Einschätzungen bewusst sind, er aber im Lichte dieser Tatsachen dennoch die Absicht gebildet hat, die verfügbare Zeit zur Vorbereitung zu nutzen, dann sollte ihn nichts daran hindern, die Absicht zu realisieren. Falls er diese Tatsachen aber

[33] Vgl. auch Gardner (1993), 35.

nicht in sein Überlegen integriert hat, besteht die Möglichkeit, dass ihn die Unlust bei der Realisierung sabotiert. Entscheidend für die Möglichkeit, sein eigenes Verhalten als unverständlich zu betrachten, muss die Tatsache sein, dass er etwas tut, das wenigstens schlechter begründet ist als das, was gemäß seiner besten verfügbaren Gründe getan werden sollte. Doch in aller Regel geht willensschwaches Handeln nicht mit der Situation des reflexiven Zusammenbruchs einher. Vielmehr können willensschwache Akteure meist wenigstens retrospektiv angeben, von welchen Faktoren sie überwältigt wurden. Wenn wir den oben eingeführten strategischen Zug, Irrationalität in den Kontext einer Entwicklung zu stellen, auf die Willensschwäche anwenden, dann können wir Folgendes sagen:

(Phase 1) Fritz bildet auf Basis aller verfügbaren relevanten Gründe im Kontext *K* die Absicht, *H* in *K* zu tun.

(Phase 2) Fritz befindet sich in *K*, vollzieht aber nicht *H*, sondern *H'*.

(Phase 3) Fritz fragt sich, warum er *H'* und nicht vielmehr *H* vollzieht. Dabei fällt ihm auf, dass er müde ist, dem Vortrag mit Angst entgegensieht und an seinen Überlegungen zweifelt. Diese psychologischen Tatsachen *erklären*, warum Fritz *H* nicht vollzieht.

Nun ist offen, wie Fritz seinen in Phase 1 erwähnten Reflexionsprozess einschätzen soll. Denn gemessen an den neu ins Spiel gebrachten und nunmehr propositional artikulierten repräsentationalen Zuständen könnte er bezweifeln, ob er tatsächlich *alle* im Kontext *K* relevanten Gründe berücksichtig hat. Kommt er dabei zu dem Ergebnis, dass dies nicht der Fall war, weil wichtige seiner Präferenzen unberücksichtigt geblieben waren, muss ihm die akratische Handlung wie eine Handlung erscheinen, die eine Rationalität offenbart, welche sich sozusagen hinter seinem Rücken vollzogen hat. Kommt er aber zu dem Ergebnis, dass er auch im Lichte der retrospektiv entdeckten, zunächst nicht propositional artikulierten mentalen Zustände zu dem gleichen Ergebnis gekommen wäre, dann muss er sich Willensschwäche attestieren: Er müsste sich eingestehen, dass bei seinem Handeln Unfreiwilligkeit eine Rolle gespielt hat, und zwar auf Grund von zunächst nicht propositional artikulierten mentalen Zuständen, die – auf eine propositionale Form gebracht – die wider bessere Gründe ausgeführte Handlung *H'* zwar *lokal* rationalisieren, aber eben nicht vor dem Tribunal aller relevanten Gründe. Im Falle der Akrasia haben wir es in der Regel mit mentalen Zuständen zu tun, die unserer Fähigkeit, Absichten handelnd zu realisieren, Grenzen setzen. Im Unterschied zu Fällen genuiner Irrationalität, in denen wir uns selbst unverständlich werden, können wir den Gehalt dieser mentalen Zustände retrospektiv allerdings meist artikulieren und Einsicht in ihre lokale motivationale Kraft gewinnen. Insofern Akrasia gehaltvolle Zustände voraussetzt, die zum Zeitpunkt akratischen Handelns propositional opak sind, scheint eine am Fall genuin irrationalen Handelns entwickelte zweistufige Theorie des Mentalen genau die begrifflichen Ressourcen bereitzustellen, die auch die Rekonstruktion willensschwachen Handelns erlauben.

Dr. Matthias Vogel, Universität Basel, Philosophisches Seminar, Am Nadelberg 6–8, 4051 Basel, Schweiz

Literatur

Brandom, Robert (2002), Introduction: Five Conceptions of Rationality, in: ders., Tales of the Mighty Dead. Historical Essays in the Metaphysics of Intentionality, Cambridge/Mass., 1–17.

Davidson, Donald (1990), Wie ist Willensschwäche möglich? [1969], in: ders., Handlung und Ereignis, Frankfurt/M., 43–72.

Ders. (1990a), Was ist eigentlich ein Begriffsschema? [1974], in: ders., Wahrheit und Interpretation, Frankfurt/M., 261–282.

Ders. (2006), Paradoxien der Irrationalität [1982], in: ders., Probleme der Rationalität, Frankfurt/M., 285–315.

Ders. (2006a), Inhohärenz und Irrationalität [1985], in: ders., Probleme der Rationalität, Frankfurt/M., 316–331.

Freud, Sigmund (1999), Bemerkungen über einen Fall von Zwangsneurose [1909], in: ders., Gesammelte Werke, Bd. VII, Frankfurt/M., 381–438.

Gardner, Sebastian (1993), Irrationality and the Philosophy of Psychoanalysis, Cambridge.

Habermas, Jürgen (1981), Erkenntnis und Interesse [1968], Frankfurt/M.

Henden, Edmund (2004), Weakness of Will and Divisions of the Mind, in: European Journal of Philosophy, Bd. 12, Nr. 2, 199–213.

Hursthouse, Rosalind (1991), Arational Actions, in: The Journal of Philosophy, Bd. 88, Nr. 2, 57–68.

Lear, Jonathan (1998), Restlessness, Phantasy, and the Concept of Mind [1988], in: ders., Open Minded. Working Out the Logic of the Soul, Cambridge/Mass., 80–122.

Millikan, Ruth Garrett (1984), Language, Thought, and Other Biological Categories. New Foundations for Realism, Cambridge/Mass.

Nagel, Thomas (1995), Freud's Permanent Revolution [1994], in: ders., Other Minds. Critical Essays 1969–1994, New York 1995, 26–44.

Quante, Michael (1995), Rationalität – Zement des Geistes. Die pragmatische Rettung des Mentalen bei D. C. Dennett, in: ders., Pragmatische Rationalitätstheorien. Studies in Pragmatism, Idealism, and Philosophy of Mind, Würzburg, 223–268.

Sartre, Jean-Paul (1976), Das Sein und das Nichts [1943], Reinbek bei Hamburg.

Searle, John R. (1983), Intentionalität. Eine Abhandlung zur Philosophie des Geistes, Frankfurt/M.

Abstract

The first part of the essay tries to show that Davidson's explanation of irrationality in terms of a fragmentation of the mind is not compatible with interpretationist premises of his own theory. Instead of adopting the conception of two semi-autonomous departments of the mind, I argue for an explanation of strong forms of irrationality based on two kinds of contentful mental states: functionally individuated representational states and states whose content depends on a rationalizing interpretation. Akrasia – as a form of irrationality caused by mental states that are not propositionally transparent – seems to fit neatly into that picture.