Studia philosophica 68/2009

Thomas Sturm

Selbsttäuschung: Wer ist hier (ir)rational und warum?*

I argue that both psychological and philosophical studies of self-deception suffer from serious weaknesses, albeit different ones. On the one hand, psychologists often use varying and unreflective conceptions of self-deception in their research. On the other hand, philosophers either ignore the necessity of paying attention to psychological research – or, if they do, they use empirical studies of human cognition and reasoning without realizing that theories and data are loaded with highly problematic assumptions. These weaknesses become centrally important in discussions about whether self-deception is a rational or irrational phenomenon. Both parties have tried to advance their views without clearly stating which normative theory of rationality they are committed to, and without explaining how this theory can be used to study or assess self-deception. More thorough interdisciplinary work is required to overcome naive conceptions and one-sided methodologies in the study of self-deception.

1. Einleitung

Während sowohl Philosophen als auch Psychologen am Phänomen der Selbsttäuschung interessiert sind, kommt es zu wenig Austausch; und wo er stattfindet, sprechen die Vertreter beider Disziplinen nicht selten aneinander vorbei. Viele derzeitige Analysen in der Philosophie sind zwar scharfsinnig, aber verlieren sich zu oft in fingierten Beispielen, von denen offen ist, ob ihnen reale Fälle von Selbsttäuschung entsprechen, und die auch zu weit von der empirischen Forschung über Voraussetzungen und Konsequenzen von Selbsttäuschung entfernt sind. Psychologen wiederum stellen häufig

* Dieser Beitrag baut auf einer erheblich überarbeiteten Übersetzung meines Beitrags *Self-Deception, Rationality, and the Self* (in *Teorema* 26 [2007] S. 73-95). Insbesondere habe ich das dort behandelte Thema der Rolle des Selbst in der Selbsttäuschung entfernt und beschäftige mich hier ausführlicher mit der Rationalität oder Irrationalität der Selbsttäuschung. Ich danke Gerd Gigerenzer für Diskussionen besonders zu «overconfidence»-Studien und Anton Hügli für verschiedene hilfreiche Fragen.

empirische Hypothesen über Selbsttäuschung auf, ohne die erforderliche Begriffsanalyse hinreichend durchgeführt zu haben. Manchmal werden so verschiedene Konzeptualisierungen von Selbsttäuschung miteinander vermengt und daher problematische Thesen darüber formuliert, was die empirischen Daten tatsächlich zeigen.[1] Dabei sollten Vertreter beider Disziplinen mehr zusammenarbeiten. Wie ich im Folgenden zeigen möchte, erfordert insbesondere die umstrittene Frage, ob Selbsttäuschung rational oder irrational ist, dringend mehr interdisziplinäre Kooperation.

Dazu möchte ich zunächst verdeutlichen, wie sich philosophische und psychologische Studien zur Selbsttäuschung in ihren Methoden und Hauptfragen unterscheiden (2). Dann werde ich argumentieren, dass die vertrauten philosophischen Methoden der Begriffsanalyse nicht ausreichen, um ein adäquates Verständnis von Selbsttäuschung zu erreichen (3). Die naheliegende Hoffnung ist, dass Psychologen hier weiterhelfen. Am Beispiel von einflussreichen Arbeiten zu der grundlegenden Frage, ob Selbsttäuschung überhaupt empirisch nachweisbar ist, wird jedoch deutlich werden, dass psychologische Studien nicht zuletzt gründlicherer begrifflicher Analysen bedürfen (4). Den gleichsam umgekehrten Punkt mache ich, indem ich philosophische Arbeiten diskutiere, die sich auf scheinbar stabile psychologische Forschungsergebnisse stützen. Ich denke hier besonders an den gründlichsten derartigen Ansatz: Alfred Meles nicht-intentionalistische oder deflationäre Konzeption der Selbsttäuschung. Mele zufolge sollte die Selbsttäuschung in den gewöhnlichen Fällen als eine Art Vorurteil oder verzerrte Meinung (*biased belief*) verstanden werden, in dem die Voreingenommenheit – und damit die behauptete Irrationalität der Selbsttäuschung – *motiviert* ist, aber keine Absicht im vollen Wortsinn erfordert. Meles Nicht-Intentionalismus vertraut wesentlich auf das «heuristics and biases»-Programm in der aktuellen psychologischen Forschung. Doch er unterschätzt dabei innerpsychologische Debatten über Theorien menschlichen Urteilens und Schließens dramatisch (5-7). Dass Selbsttäuschung irrational ist, darf nicht einfach vorausgesetzt werden. Denn zum einen liegt es nahe anzunehmen, dass sich praktisch alle Beispiele von Selbsttäuschung auch rationalisieren lassen. Zum anderen haben derartige Versuche jedoch zu wenig

[1] Das ist kein auf das Phänomen der Selbsttäuschung beschränktes Problem: Die Psychologie ist voll von nichtempirischen Annahmen, die oft auch notwendig sind, um den Gegenstand der Forschung genau zu bestimmen; doch häufig werden diese Annahmen nicht ausreichend geklärt. Vgl. Jochen Brandtstädter, Thomas Sturm: *Apriorität, Erfahrung und das Projekt der Psychologie*, in *Zeitschrift für Sozialpsychologie* 35 (2004) S. 15-32.

überdacht, was die Annahme von Rationalität erfordert, und zu oft unklare Begriffe von Selbsttäuschung verwendet. Zudem erfordert die Charakterisierung von Selbsttäuschung als (ir)rational substanzielle und umstrittene Rationalitätsnormen (8). Philosophen wie Psychologen werden sich offenbaren müssen: Entweder sie explizieren und begründen ihre Rationalitätsstandards – oder ihre Beispiele von Selbsttäuschung, ja sogar die verbreitete Annahme der Existenz von Selbsttäuschung, werden dubios bleiben. Selbst wenn wir am Ende bei der Auffassung bleiben sollten, dass Selbsttäuschung irrational ist, kann dies in seriöser Weise nur behauptet werden, wenn die Rationalitätsstandards reflektiert werden (9). Philosophen wie Psychologen haben sich also bislang ziemlich irrational verhalten, wenn sie meinten, sie könnten die jeweils andere Disziplin ignorieren.

2. *Ein philosophisches Paradoxon und ein Problem der Psychologie*

Plato hielt die Selbsttäuschung für eines der größten Übel, da der Betrüger in der Seele selbst sei, keinen Schritt weit entfernt vom Betrogenen (*Kratylos* 428d). Bischof Butler, Adam Smith, Kant, Sartre und andere haben darüber geschrieben. Seit der letzten Hälfte des 20. Jahrhunderts jedoch ist die Literatur hinsichtlich des Themas geradezu explodiert. Dabei lassen sich zwei neue Entwicklungen feststellen: eine philosophische und eine psychologische.

In früheren Jahrhunderten haben die Philosophen die Selbsttäuschung meist als ein ethisches Problem behandelt. In den letzten Jahrzehnten jedoch haben sie sich auf das theoretische Problem des «Paradoxons der Selbsttäuschung» konzentriert. Ein theoretisches Rätsel tritt nämlich auf, wenn man die Selbsttäuschung anhand des Vorbilds der absichtlichen Täuschung einer anderen Person analysiert. Eine solche Täuschung liegt etwa vor, wenn Cheney beabsichtigt, Powell vom Gegenteil dessen zu überzeugen, was seine (Cheneys) eigene beste Überzeugung ist, und dies mit verschiedenen vorsätzlichen Maßnahmen durchführt. Wenn man Selbsttäuschung analog konzeptualisiert, dann muss eine Person sich absichtlich dazu bringen, an etwas zu glauben, das sie im selben Moment nicht glaubt. Ist so etwas überhaupt möglich? Diskussionen über dieses Problem dominieren die philosophische Debatte über Selbsttäuschung seit der zweiten Hälfte des 20. Jahrhunderts.[2]

[2] Inzwischen werden sogar zwei Formen des Paradoxons unterschieden, ein «statisches» (Wie kann man gleichzeitig p und ¬p glauben?) und ein «dynamisches» (Wie kann man sich absichtlich dazu bringen, p zu glauben, während und

Psychologen wiederum haben 'Selbsttäuschung' und verwandte Begriffe ebenfalls schon länger verwendet. Neu in den letzten Jahrzehnten ist, dass sie sich für die Fragen interessieren, die sie mit empirischen Mitteln beantworten wollen. Kann man beispielsweise überhaupt nachweisen, dass es Fälle von Selbsttäuschung gibt? Was sind die Mechanismen und Funktionsweisen von Selbsttäuschung? Wie passt Selbsttäuschung zu empirischen Theorien über das Denken und Schließen des Menschen?

3. Ansätze zur Auflösung des Paradoxons

Zur begrifflichen Auflösung des Paradoxons der Selbsttäuschung gibt es verschiedene Ansätze. Einige Autoren wie etwa Raphael Demos in einem klassischen Artikel nehmen an, dass Selbsttäuschung möglich ist, weil die widersprüchlichen Meinungen auf verschiedenen Bewusstseinsebenen auftreten: Während die bevorzugte Meinung transparent ist und eine damit inkompatible Meinung nach außen hin auch abgelehnt wird, existieren sichere Indizien dafür, dass letztere Meinung unbewusst vorhanden ist.[3] Herbert Fingarette dagegen hat gefordert, die Rede von Meinungen und unbewussten Zuständen aufzugeben. Vielmehr sollten wir von verschiedenen «Verpflichtungen» (*commitments*) sprechen, die wir im Falle von Selbsttäuschung nicht fähig sind auszusprechen.[4] Robert Audi wiederum macht ein moderateres Angebot. Er stellt fest, dass eine der zwei inkompatiblen Propositionen nicht in Form einer Meinung akzeptiert wird, sondern als «aufrichtige Bekundung».[5] Donald Davidson schließlich hat behauptet, dass die Selbsttäuschung durch die Aufteilung des Geistes einer Person in unhabhängige Gruppen von Zuständen ermöglicht wird – unabhängig in dem

gerade weil man ¬p glaubt?) Vgl. Alfred Mele: *Recent Work on Self-Deception*, in *American Philosophical Quarterly* 24 (1987) S. 1-17; A. M.: *Self-Deception Unmasked* (Princeton: Princeton UP, 2001). Diese Unterscheidung benötige ich im Folgenden nicht.

3 Raphael Demos: *Lying to Oneself*, in *Journal of Philosophy* 57 (1960) S. 588-595; eine bis in die jüngste Zeit immer wieder vertretene Position: vgl. Baljinder Sahdra, Paul Thagard: *Self-Deception and Emotional Coherence*, in *Minds and Machines* 13 (2003) S. 213-231. Demos vertritt dabei nicht etwa ein starkes freudianisches Konzept des Unbewussten.

4 Herbert Fingarette: *Self-Deception* (London: Routledge & Kegan Paul, 1969).

5 Robert Audi: *Self-Deception, Action, and Will*, in *Erkenntnis* 18 (1982) S. 133-158.

Sinn, dass die gewöhnlichen logischen und epistemologischen Beziehungen zwischen ihnen zerrissen sind, obwohl die Zustände kausal miteinander verbunden bleiben, so dass die Annahme eines gespaltenen Selbst vermieden werden kann.[6]

Alle diese Explikationsversuche haben Probleme, die ich hier nicht diskutieren möchte. Wichtig ist vielmehr, den ihnen gemeinsamen Maßstab zu erfassen. Allen gemein ist nämlich die Suche nach der Lösung des Paradoxons der Selbsttäuschung. Und eine Explikation löst das Paradoxon so gut wie eine andere. Dabei habe ich nur eine begrenzte Anzahl der begrifflichen Optionen umrissen.[7] Die schiere Vielfalt von Möglichkeiten verdeutlicht, dass es mehr als fraglich ist, ob man durch begriffliche Auflösung des Paradoxons allein das Phänomen der Selbsttäuschung angemessen erfassen kann.

4. Existenzielle Erörterungen über Selbsttäuschung mittels eines Experiments – und seine Probleme

Gibt es nicht einen offensichtlichen Lösungsweg? Ist es nicht eine natürliche Aufgabe der Psychologie, dieses Phänomen angemessen zu beschreiben und zu erklären, die diese Disziplin auch besser erledigen kann als irgendeine andere? Leider wirkt ein näherer Blick in die entsprechende Literatur ernüchternd; Psychologen geraten hinsichtlich der Selbsttäuschung schnell in Schwierigkeiten.

Um dies zu zeigen, möchte ich mit einer vielleicht überraschenden Frage beginnen: Kommt Selbsttäuschung überhaupt jemals vor? Auch Psychologen und andere empirische Wissenschaftler sind oft fraglos von der Existenz dieses Phänomens ausgegangen. So ist vermutet worden, dass falsche Selbst-

6 Donald Davidson: *Deception and Division*, in *The Multiple Self*, hg. von John Elster (Cambridge: Cambridge University Press, 1986) S. 79-82. Die stärkere Annahme wird Davidson etwa von Alexander Bird (*Rationality and the Structure of Self-Deception*, in *European Review of Philosophy* 1 [1994] S. 19-38) zugeschrieben; tatsächlich behauptet wird sie z.B. von Amelie O. Rorty (*Akratic Believers*, in *American Philosophical Quarterly* 20 [1983] S. 175-183).

7 Auch etwa die in der Literatur oft mitbehandelte Frage, wie man Selbsttäuschung von Wunschdenken unterscheiden kann, schränkt die Suche nach dem besten Konzept nur geringfügig ein. Weitere Optionen s. unten in Abschnitten 5-6; umfassende Bibliographien finden sich im Internet: http://consc.net/mindpapers/5.11.5.13, http://philpapers.org/browse/self-deception/ und http://plato.stanford.edu/entries/self-deception/.

darstellungen in Persönlichkeitstests eher infolge von Selbsttäuschung als Fremdtäuschung auftreten.[8] Auch wurde mithilfe von Selbsttäuschung zu erklären versucht, warum Menschen an Hypothesen festhalten, selbst wenn deren Nichtbestätigung bereits erfolgt ist.[9] Selbsttäuschung wird auch als ein im Alltagsleben allseits präsenter Zustand angenommen. So wird sie im Verneinen von Krankheit, im fahrlässigen Verhalten von Autofahrern und im zuversichtlichen Optimismus von Arbeitslosen oder Soldaten im Kampfeinsatz gesehen.[10] Die Beispiele lassen sich vermehren.

Doch in keiner dieser Studien ist nachgewiesen worden, dass Selbsttäuschung überhaupt jemals auftritt. Haben die genannten Hypothesen also eine Basis? Auch angesichts des *prima facie* paradoxen Charakters des Phänomens sollte die Annahme ihrer Existenz nicht einfach hingenommen werden, so beliebt und alltäglich sie auch sein mag. Vielleicht werden Zuschreibungen von Selbsttäuschung nur vorgenommen, um vermeintlichen Selbstbetrügern irrationales oder gar unmoralisches Verhalten vorzuwerfen? Ist Selbsttäuschung lediglich ein Konstrukt unserer Kultur, von dem wir uns besser befreien sollten?

Vor diesem Hintergrund ist es keine bloß eitle Aufgabe, wenn Psychologen versuchen, die Existenz von Selbsttäuschung experimentell zu demonstrieren. Im frühesten und zugleich wohl auch letzten Ansatz einer derartigen Untersuchung sind Gur und Sackeim von der Idee Demos' ausgegangen, dass die widersprüchlichen Meinungen auf «verschiedenen Ebenen des Bewusstseins» gehalten werden. Genauer behaupten sie folgende Bedingungen als notwendig und hinreichend für Selbsttäuschung:

1. Das Individuum hat zwei einander ausschließende Meinungen (p und $\neg p$).
2. Diese zwei Meinungen werden gleichzeitig gehalten.
3. Das Individuum ist sich nur bei einer der Meinungen bewusst, dass es sie hat.
4. Der Akt, der bestimmt, welche Meinung unbewusst ist, ist ein motivationaler Akt.[11]

8 Paul E. Meehl, Starke R. Hathaway: *The K Factor As a Suppressor Variable in the Minnesota Multiphasic Personality Inventory*, in *Journal of Applied Psychology* 30 (1946) S. 525-564.

9 Peter C. Wason, Philip N. Johnson-Laird: *Psychology of Reasoning* (Cambridge, MA: Harvard University Press, 1972).

10 Vgl. Sahdra/Thagard, op. cit. (Fn. 3) S. 213.

11 Ruben C. Gur, Harold A. Sackeim: *Self-Deception: A Concept in Search of a Phenomenon*, in *Journal of Personality and Social Psychology* 37 (1979) S. 147-169, hier S. 149 (meine Übersetzung). – Vgl. Harold A. Sackeim, Ruben

Zum Nachweis von Selbsttäuschung werden dann Stimmenerkennungsexperimente eingesetzt. In einem typischen Experiment werden Einzelpersonen gefragt, ob sie eine auf Tonband aufgenommene Stimme für ihre eigene oder die einer anderen Person halten. Während die Probanden darüber berichten, werden Hautreflexe gemessen. Das soll Auskunft darüber geben, ob die Probanden dann, wenn sie ihre eigene Stimme nicht zu erkennen vorgeben, dies unterschwellig doch tun. Gur und Sackeim argumentieren zudem mit Hilfe von Persönlichkeitstests sowie einem Fragebogen speziell zur Selbsttäuschung, dass Menschen mit negativem Selbstwertgefühl oder mit diskrepanten Meinungen darüber, was sie selbst sind und was sie sein sollten, Konfrontation mit sich selbst – etwa mit der eigenen Stimme – unangenehm finden. Entsprechend sei die geäußerte «Nichterkennung» der eigenen Stimme motiviert. Umgekehrt finden Personen mit positivem Selbstwertgefühl an der Konfrontation mit der eigenen Stimme sogar Gefallen. So sei Selbsttäuschung tatsächlich nachweisbar.

Diese Studie ist in der psychologischen Forschung durchaus einflussreich geworden. Sie hat den Weg für weitere Untersuchungen von Mechanismen und Funktionen der Selbsttäuschung geöffnet, etwa für psychologische Erklärungen, Wirkungen oder den möglichen Nutzen dieses Phänomens.[12] Auf solche Studien komme ich später zurück.

Gegen Gurs und Sackeims Ansatz lassen sich zumindest drei Einwände vorbringen. Erstens ist argumentiert worden, dass die Aufgabe, die eigene

C. Gur: *Self-Deception, Self-Confrontation, and Consciousness*, in *Consciousness and Self-Regulation*, hg. von G. E. S. D. Shapiro (New York: Plenum, 1978) S. 139-197; G. E. S. D. S.: *Self-Deception, Other-Deception and Self-Reported Psychopathology*, in *Journal of Consulting and Clinical Psychology* 47 (1979) S. 213-215. Sie folgen Demos' Begriffsexplikation, weil sie Selbsttäuschung in Analogie zum Phänomen der «perceptual defense» sehen. Menschen neigen manchmal dazu, unangenehme Wahrnehmungen zu meiden, aber damit das überhaupt möglich ist, muss man einen bestimmten Reiz erst einmal erfassen. Das Konzept des Wahrnehmens eines Reizes bedeutet daher zum einen so viel wie 'vom sensorischen System erfasst werden' und zum anderen so viel wie 'bewusst erfahren werden'.

12 Vgl. etwa George A. Quattrone, Amos Tversky: *Self-Deception and the Voter's Illusion*, in *The Multiple Self*, hg. von John Elster (Cambridge: Cambridge University Press, 1986) S. 35-58; Joan S. Lockhard, Delroy L. Paulhus (Hg.): *Self-Deception: An Adaptive Mechanism?* (Englewood Cliffs, NJ: Prentice-Hall, 1988); Delroy L. Paulhus, Douglas B. Reid: *Enhancement and Denial in Socially Desirable Responding*, in *Journal of Personality and Social Psychology* 60 (1991) S. 307-317.

Stimme zu erkennen, problematisch ist, da ähnliche Ergebnisse erzielt werden, wenn die Testperson Stimmen von Bekannten erkennen soll. Dieser Einwand ist freilich schwach: Er zeigt nicht, dass Fehler bei der Erkennung anderer Stimmen nicht auch auf Selbsttäuschung beruhen können (weil sie etwa auch motiviert sind). Ein zweiter Einwand jedoch zieht stärker:[13] Wegen der gewählten Explikation des Begriffs von Selbsttäuschung, demzufolge es sich um eine Spannung zwischen gleichzeitig, aber auf verschiedenen Bewusstseinsebenen gehaltenen *Meinungen* handelt (Bedingungen 1-3 bei Gur und Sackeim), müssen die im Experiment gemessenen Hautreflexe Indikatoren von (unbewussten) Meinungen sein. Diese Annahme ist jedoch alles andere als zwingend. Beispielsweise mögen die Hautreflexe ja nur Unsicherheit darüber ausdrücken, was die Probanden öffentlich über die gehörten Stimmen vorgeben, oder auch eine Sensibilität für die gehörten Stimmen.[14] Drittens ist zweifelhaft, ob Gur und Sackeim wirklich Demos' Begriff der Selbsttäuschung verwendet haben. Ihre Beispiele zeigen nicht, dass aufseiten des Selbstbetrügers auch eine Intention vorliegen muss, sich selbst zu täuschen, wie Demos dies gefordert hat.[15] Bloß ein Motiv zu haben (Bedingung 4 bei Gur und Sackeim) ist schwächer: Es erfordert beispielsweise keine praktischen Überlegungen über Zwecke und Mittel, die zu einer gefestigten Absicht führen. Man sollte sich noch einmal die Parallele zur intentionalen Fremdtäuschung vor Augen halten. Wir würden nicht von einer Lüge sprechen, wenn z.B. Hans (der glaubt, dass p) unbeabsichtigt verursacht, dass Franz glaubt, dass $\neg p$. Wir würden dies selbst in dem Fall nicht tun, in dem Hans zwar auch wünscht oder ein Motiv dafür hat, dass Franz glaubt, dass $\neg p$, aber Hans etwa aus moralischen Gründen der Versuchung zum Lügen widersteht, während Franz aus irgendeinem anderen Grund heraus zu der Meinung gelangt, dass $\neg p$. Bei intentionaler

13 Der Einwand stammt von William Douglas, Keith Gibbins: *Inadequacy of Voice Recognition as a Demonstration of Self-Deception*, in *Journal of Personality and Social Psychology* 44 (1983) S. 589-592. Zur Erwiderung: Harold A. Sackeim, Ruben C. Gur: *Voice Recognition and the Ontological Status of Self-deception,* in *Journal of Personality and Social Psychology* 48 (1985) S. 1365-1368.

14 Vgl. Alfred Mele: *Recent Work on Self-Deception*, in *American Philosophical Quarterly* 24 (1987) S. 1-17. Später hat Sackeim dies eingeräumt (Harold A. Sackeim: *Self-deception: A Synthesis*, in Lockhard/Paulhus, op. cit. [Fn. 12] S. 146-165). Für weitere Kritik an vermeintlichen empirischen Demonstrationen des Haltens widersprüchlicher Meinungen vgl. Mele: *Self-Deception Unmasked*, op. cit. (Fn. 2) S. 76-93.

15 Demos, op. cit. (Fn. 3) S. 588.

Fremd- wie Selbsttäuschung muss der Wunsch auch bei höherstufiger Überlegung gewollt sein.

Die Probleme dieses experimentellen Existenzbeweises von Selbsttäuschung sind zum Teil methodologischer oder empirischer Natur (wie beim ersten vorgeführten Einwand), zum Teil begrifflicher Natur (wie bei den beiden letzteren). Daher ist es nicht leicht zu sagen, was zuerst zu tun wäre, um voranzukommen. Sollte man eine andere Begriffsexplikation wählen? Aber welche? Oder lieber das Experiment ändern? Oder gleich beides? Vermutlich würden ähnliche Probleme beim Gebrauch anderer Konzeptualisierungen auftreten. Wenn man etwa annimmt, dass die Selbsttäuschung z.B. eine noch tiefere Aufspaltung des Selbst erfordert als eine Verschiedenheit von Bewusstseinsebenen, dann wäre es womöglich erneut unnütz, den Existenzbeweis mittels Hautreaktionen durchzuführen.

Manche Autoren behaupten nun, empirische Studien zur Selbsttäuschung seien ohnehin überflüssig, da es sich nur um ein soziales oder kulturelles Konstrukt handele, wie der zuvor angesprochene skeptische Einwand schon angedeutet hatte.[16] Jedoch wäre es verfrüht, eine empirische Demonstration von Selbsttäuschung völlig auszuschließen, auf deren Basis man den Begriff dieses Phänomens gebrauchen könnte, um weitere Untersuchungen und Theorien zu entwickeln. Wegen der vielfältigen Möglichkeiten der Begriffsexplikation ist dies eine schwierige Aufgabe; nicht weniger, aber eben auch nicht unbedingt mehr. Diese brauche ich hier nicht zu verfolgen – das wäre schließlich nur in Zusammenarbeit mit empirischen Untersuchungen sinnvoll. Ziel war zunächst nur zu verdeutlichen, dass Psychologen sich nicht immer ausreichend um die notwendige konzeptuelle Analyse bemühen.

5. *Sparsam sein? Selbsttäuschung als «biased belief»*

Nun möchte ich gleichsam die umgekehrte Schwäche auf philosophischer Seite illustrieren: den Versuch, den Begriff der Selbsttäuschung mittels beliebter empirischer Theorien über Mechanismen und Funktionen menschlichen Denkens und Urteilens zu analysieren. So ein Ansatz entspricht schließ-

16 Kenneth J. Gergen: *The Ethnopsychology of Self-Deception*, in *Self-Deception and Self-Understanding*, hg. von Mike W. Martin (Lawrence, Kansas: Kansas University Press, 1985) S. 228-243; Brian L. Lewis: *Self-Deception: A Postmodern Reflection*, in *Journal of Theoretical and Philosophical Psychology* 16 (1996) S. 49-66.

lich der Hoffnung, nicht nur (aber auch) das genannte Paradoxon aufzulösen, sondern darüber hinaus ein sachlich adäquates Konzept von Selbsttäuschung zu gewinnen. Der bislang gründlichste derartige Ansatz findet sich zweifellos in den Publikationen Alfred Meles.

Mele behauptet, dass das Modell der intentionalen interpersonellen Täuschung unnötig ist, um Selbsttäuschung korrekt zu beschreiben. Vielmehr sollten wir Selbsttäuschung als eine Art von voreingenommener oder von aufgrund bestimmter kognitiver Mechanismen verzerrter Meinung (*biased belief*) auffassen. Damit bezieht Mele sich auf Studien des als «heuristics and biases» bezeichneten, höchst ernst zu nehmenden psychologischen Forschungsprogramms (dessen Hauptvertreter, Daniel Kahneman und Amos Tversky, dafür immerhin einen Nobelpreis für Ökonomie erhalten haben).[17] Diesem Programm zufolge ist unser Denken und Urteilen von kognitiven «Heuristiken» bestimmt, die passable Faustregeln darstellen, jedoch universell angewandt zu systematischen Fehlern führen. Die Fehler werden identifiziert, indem man gewisse Normen als Standards ansetzt – Regeln der Logik, der Wahrscheinlichkeitstheorie oder der Statistik etwa. Letztere Normen lassen sich auch als das «Standardmodell» von Rationalität bezeichnen.[18] Ich werde das Programm in den Abschnitten 7-8 noch näher diskutieren.

Meles Position zufolge sind nicht alle voreingenommenen oder verzerrten Meinungen Selbsttäuschungen. Für letztere ist es spezifisch, dass die Verzerrung oder Voreingenommenheit *motiviert* ist.[19] Es bedarf aber keiner vollen Absicht, keiner intentionalen Handlung und auch keines Festhaltens an sich widersprechenden Meinungen.[20] Meles Konzept ist daher sparsam oder deflationär. Er leugnet dabei nicht, dass intentionale Selbsttäuschungen möglich sind. Die von ihm formulierten Bedingungen sollen hinreichend, aber nicht notwendig sein. Allerdings nimmt er an, dass die gewöhnlichen Fälle von Selbsttäuschung eher seiner Analyse als den intentionalistischen

17 Alfred Mele: *Real Self-Deception,* in *Behavioral and Brain Sciences* 20 (1997) S. 91-102; Mele: *Self-Deception Unmasked*, op. cit. (Fn. 2). Für andere Ansätze, die dieses Forschungsprogramm aufgreifen, vgl. Ariela Lazar: *Deceiving Oneself or Self-Deceived?*, in *Mind* 108 (1999) S. 263-290; David Patten: *How Do We Deceive Ourselves?*, in *Philosophical Psychology* 16 (2003) S. 229-246.

18 Vgl. Edward Stein: *Without Good Reason* (Oxford: Clarendon Press, 1996).

19 Vgl. Ziva Kunda: *The Case for Motivated Reasoning*, in *Psychological Bulletin* 108 (1990) S. 480-498.

20 Der letzte Punkt wurde schon von Frederick A. Siegler (*Demos On Lying to Oneself*, in *Journal of Philosophy* 59 [1962] S. 469-475) gegen Demos, op. cit. (Fn. 3) vorgebracht.

Begriffsexplikationen entsprechen. Man bemerke auch einen wichtigen Vorteil seines Ansatzes: Mit einem derartigen moderaten Konzept von Selbsttäuschung, bei dem sich kein Paradoxon mehr stellt, und das zudem in eine Theorie eingebettet ist, die umfassend Verzerrungen oder Voreingenommenheiten in unseren Urteilen und Entscheidungen behandelt, stellt sich die Frage der Existenz der Selbsttäuschung nicht, oder jedenfalls nicht in der dramatischen Weise wie beim intentionalistischen Begriffsverständnis. Wenn wir häufig unter verzerrten Meinungen leiden und wenn Selbsttäuschungen oft ein Spezialfall hiervon sind, braucht man die Existenz von Selbsttäuschung kaum zu bezweifeln.

Nach dieser Auffassung ist Selbsttäuschung nach wie vor ein irrationales Phänomen, allerdings nicht mehr aus Gründen der Meinungsinkonsistenz oder wegen der Absurdität des Vorhabens, sich selbst absichtlich in die Irre zu führen – sondern vielmehr, weil sie irgendwelche Normen der Logik, der Wahrscheinlichkeitstheorie oder der Statistik verletzt. Man betrachte etwa den von Mele aufgegriffenen Befund, dass 94% der Universitätsprofessoren von sich glauben, besser in ihrem Beruf zu sein als ihre durchschnittlichen Kollegen.[21] Dies kann keinesfalls für alle diese Professoren gelten. Doch obwohl sie irren, müssen sie keine widersprüchlichen Meinungen besitzen oder sich selbst absichtlich getäuscht haben. Vielleicht haben sie ihre Ansicht aus den Feedbacks von Studenten gefolgert, die wiederum aufgrund des gewöhnlichen Wissensvorteils zwischen Lehrern und Schülern denken mögen, ihre Dozenten seien schlicht brillant. Die Professoren mögen daher Fakten über statistische Verteilungen übersehen. Wenn der Befund akzeptabel ist, dann erliegen die Professoren der Täuschung aus verständlichen Motiven und Prozessen – und täuschen sich daher in Meles Sinn selbst.

6. *Illusionen des Nicht-Intentionalismus*

Doch ist Meles Position überzeugend? Zwei Einwände von eher zwiespältiger Überzeugungskraft seien in diesem und ein weiterer und deutlich stärkerer in den nächsten zwei Abschnitten vorgeführt.

Erstens könnte man vermuten, dass in manchen Fällen von Selbsttäuschung kein Motiv vorliegt. So würde Meles zentrale, das Phänomen kennzeichnende Bedingung verfehlt sein. Dieser Einwand ist zumindest

21 Entnommen aus Thomas Gilovich: *How We Know What Isn't So* (New York: Free Press, 1991) S. 77.

aufgeworfen worden, um zu betonen, dass Meles Annahme, Selbsttäuschung sei stets motiviert, lediglich ein unbegründetes Postulat darstellen mag oder vielleicht nur aus einer Überverallgemeinerung von als typisch angesehenen Fällen herrührt.[22] Allerdings ist dieser Einwand nicht sonderlich zwingend. Man muss ja bei Begriffsexplikationen auch immer von Standardfällen ausgehen, und ohne die genannte Bedingung würde Selbsttäuschung sich überhaupt nicht mehr von anderen verzerrten Meinungen, Vorurteilen oder Täuschungen unterscheiden.

Eine zweite, plausiblere Frage geht in die umgekehrte Richtung: Können nicht einige von unseren Vorurteilen motiviert sein, ohne schon Selbsttäuschungen darzustellen? Wenn dies so wäre, wären Meles Bedingungen nicht nur nicht notwendig, sondern auch nicht hinreichend. In Richtung solcher Überlegungen hat Holton auf das Beispiel eines Rassisten verwiesen, der aus erkennbaren Motiven heraus bei seinen Überzeugungen bleibt, auch wenn die Argumente gegen den Rassismus deutlich vorgeführt worden sind.[23] Holton führt diesen Fall an, weil er denkt, dass Selbsttäuschung weniger durch die Person selbst *verursacht* ist als vielmehr eine Täuschung darstellt, die schlicht die Person selbst *betrifft*. Man irrt sich, anders gesagt, nicht selbst (schon gar nicht intentional nach Holtons Position), sondern *über* sich selbst. Der Rassist hingegen hat keine motivierte falsche Meinung über seine eigene Person, sondern über andere. Doch es ist überhaupt nicht einleuchtend, diese Einschränkung vorzunehmen. Man mag sich selbst über die Handlungen der eigenen Partner täuschen oder über die Talente und den Charakter der eigenen Kinder oder dass die eigene Regierung die Beteiligung an einem Krieg nicht aus den besten moralischen Gründen verweigert. Holton will mit seiner Explikation die ganzen Paradoxien der Selbsttäuschung erledigen. Doch auch wenn man gegenüber der übermäßigen philosophischen Vorliebe für die Beschäftigung mit Begriffsrätseln skeptisch sein darf, ist das zu handstreichartig.

Es gibt bessere Beispiele dafür, dass verzerrte Meinungen motiviert sein können, ohne schon Selbsttäuschungen darzustellen. Bereits Demos hat bestritten, dass der seine Fähigkeiten überschätzende Wissenschaftler sich selbst täuscht, da eben keine Intention für die Täuschung vorgelegen hat.[24] Ein Beispiel dieser Art wird von Mele freilich andersherum interpretiert, mit

22 Vgl. Patten, op. cit. (Fn. 17).

23 Richard Holton: *What is the Role of the Self in Self-Deception?*, in *Proceedings of the Aristotelian Society* 101 (2000) S. 53-69, hier S. 59.

24 Demos, op. cit. (Fn. 3) S. 588.

Verweis auf den oben erwähnten statistischen Befund.[25] So scheinen beide Positionen ihre Begriffsvorstellungen nur fruchtlos hin- und herzuschieben. Ich werde jedoch zeigen, dass Mele mit seiner Position nicht durchkommt (unten, Abschnitt 7).

Zudem kann der Intentionalist eine andere Route einschlagen und die Bedingung bestreiten, dass die beiden inkompatiblen Meinungen *gleichzeitig* gehalten werden müssen – also statt einer synchronen Trennung eine diachrone Trennung der relevanten geistigen Zustände verlangen. Dies kann die Position stärken, der zufolge Motivation allein nicht hinreichend ist. Beispiele hierfür sind leicht denkbar. Ein Mädchen, dessen Häschen gestorben ist und dessen boshafter Bruder es dauernd an den Trauerfall erinnern will, mag bei jedem Versuch des Bruders die Hände auf die Ohren legen, schreien und aus dem Zimmer laufen, um die schmerzhafte Information zu vermeiden. Durch seine intentionalen Handlungen bewahrt es geflissentlich seine Meinung, der Hase lebe noch.[26] Ich hörte vor einer Weile von einem philosophischen Vortrag, in dem der Redner energisch für Ansichten argumentiert hatte, die auf gar keinen Fall für wahr gehalten werden können, weil sie viel zu verworren sind (der Vortrag handelte über Hegel, den «Mythos des Gegebenen» und dergleichen). Ein klar denkender und uneitler Philosoph verließ den Saal und flüsterte zu einigen Kollegen: «Ich glaube, dies war eine sehr, sehr subtile Art von Selbsttäuschung.» Wie könnte man diese Bemerkung verstehen? Vielleicht hatte der Vortragende erfasst, dass ein Philosoph, der über die erwähnten Themen und Autoren redet, es zu Bekanntheit und zum Ruf eines tiefen Denkers bringen kann, und wollte dies auch erreichen. Er mag sich dazu einseitig auf entsprechende Artikel und Bücher gestürzt haben und irgendwann eben zu den fraglichen Meinungen gelangt sein, da ihm niemand klipp und klar gesagt hatte, dass das keine Garantie für die Vermeidung von Unsinn ist. So bildete er die verzerrten Meinungen und verdrängte immer wieder den aufkeimenden Gedanken, dass diese Meinungen gar nicht glaubwürdig sind. Ich will hier offenlassen, ob derartige Fälle typisch sind. Dass die Selbsttäuschung als «sehr, sehr subtil» charakterisiert wurde, zeigt vielleicht, dass sie eher untypisch ist. Allerdings gibt es Autoren, die das anders sehen.[27]

25 Mele: *Self-Deception Unmasked*, op. cit. (Fn. 2) S. 3, 11.

26 Das Beispiel stammt von Christian Perring (*Direct, Fully Intentional Self-Deception Is Also Real*, in *Behavioral and Brain Sciences* 20 [1997] S. 123-124).

27 So José L. Bermudez: *Self-Deception, Intentions, and Contradictory Beliefs*, in *Analysis* 60 (2000) S. 309-319; etwas anders Anna Nicholson: *Cognitive Bias, Intentionality, and Self-Deception*, in *Teorema* 26 (2007) S. 45-58.

Sicher würde Mele wieder antworten, dass solche Fälle eine schlechte Explikation des Konzeptes der Selbsttäuschung darstellen. Er verweist zum Beispiel darauf, dass man bei den vorgeführten intentionalen Handlungen zwischen solchen unterscheiden sollte, die aus irgendwelchen Absichten heraus geschehen, und solchen, die direkt aus einer Absicht der Selbsttäuschung heraus erfolgen. Doch selbst wenn das Mädchen oder der Philosoph intentionale Handlungen vollzogen, die zur Selbsttäuschung führten, ohne ursprünglich dieses Ziel gehabt zu haben, kann Mele nicht ausschließen, dass solche Fälle vorkommen. Damit jedoch stehen wir vor einer Sackgasse: Anhänger des Nicht-Intentionalismus und des Intentionalismus sprechen offenbar von verschiedenen Phänomenen. Ihre Konzeptualisierungen bezwecken, das Paradox der Selbsttäuschung zu lösen, und beide sind darin erfolgreich. Doch wenn man von diesem oder ähnlichen begrifflichen Problemen besessen ist, wird man über die Phänomene keine wirkliche Einsicht erlangen. Insofern ist Meles Ansatz inkonsequent durchgeführt: Obwohl er sich stärker als andere Philosophen auf psychologische Forschungen stützen will, bleibt er vorwiegend im üblichen philosophischen Verfahren des Gebrauchs von ausgedachten Beispielen und Begriffsanalysen befangen. Er bietet zwar einige empirische Argumente für die Existenz der «gewöhnlichen» Fälle von Selbsttäuschung, aber keinen Beleg dafür, dass diese den Großteil von Selbsttäuschung ausmachen. Nun möchte ich zeigen, dass auch seine empirischen Argumente auf schwachen Füßen stehen.

7. Das «heuristics and biases»-Programm und seine Grenzen

Ich will weitere in der Literatur erhobene Einwände nicht erörtern, da sie an vergleichbaren Mängeln leiden wie die eben vorgeführten. Es gibt ohnehin einen dritten und destruktiveren Einwand, der noch nicht klar genug betrachtet worden ist. Er betrifft das von Mele genutzte psychologische Forschungsprogramm von «heuristics and biases» (HB-Programm).

Dem HB-Programm zufolge ist unser Denken und Urteilen von kognitiven «Heuristiken» bestimmt, die passable Faustregeln darstellen, jedoch universell angewandt zu systematischen Fehlern führen.[28] Das Programm

[28] Vgl. zur Übersicht etwa Amos Tversky, Daniel Kahneman: *Judgment Under Uncertainty: Heuristics and Biases*, in *Science* 185 (1974) S. 1124-1131; Richard E. Nisbett, Lee Ross: *Human Inference* (Englewood Cliffs, NJ: Prentice Hall, 1980). Thomas Gilovich, Dale W. Griffin, Daniel Kahneman (Hg.): *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge: Cambridge University Press, 2002).

benötigt für jeden empirischen Versuch eine zuvor bestimmte Norm, an der dann das Verhalten des Probanden gemessen wird – wie in Abschnitt 5 angesprochen, sind dies in aller Regel Normen des Standardmodells von Rationalität. Hier sind vier Beispiele für das HB-Programm:

(1) Mittels der Wason-Kartenwahlaufgabe hat man geprüft, ob Versuchspersonen die Wahrheitsbedingungen der *materialen Implikation* der Logik befolgen. Wenn die Probanden aufgefordert werden, ihnen vorgelegte Karten (mit Buchstaben und Zahlen oder anderen Inhalten) als Instanzen der Verknüpfung $p \rightarrow q$ zu prüfen, so übersehen sie dabei regelmäßig die für die Falsifizierung wichtigen $\neg q$-Karten. Dies ist oft als ein übermäßiger Hang zu Bestätigungen («confirmation bias») angesehen worden – so, als ob wir für gewöhnlich Anti-Popperianer wären.[29]

(2) Im «Linda problem» wird den Probanden eine Person beschrieben: Linda sei eine aufgeweckte, intelligente und an politischen Themen und Aktivitäten interessierte ehemalige Philosophiestudentin. Dann ist gefragt worden, ob es wahrscheinlicher sei, dass Linda *(A)* eine Bankangestellte ist oder *(A&B)* eine Bankangestellte und Feministin ist. In Kahnemans und Tverskys Studien hierzu haben die Versuchspersonen ganz überwiegend Antworten gegeben, wonach *(A&B)* wahrscheinlicher sei als *(A)*. Das widerspricht jedoch der *Konjunktionsregel* der Wahrscheinlichkeitstheorie – der Regel, der zufolge ein Ereignis *A* niemals weniger wahrscheinlich sein kann als die Konjunktion der (unabhängigen) Ereignisse *A* und *B* (formal: $p(A) \geq p(A\&B)$). Erklärt wurde dies so, dass die Probanden ihr Urteil aufgrund der «Repräsentativität» der Informationen über Linda bilden («representativeness heuristics»).[30]

(3) *Basisraten* sind ein wichtiger Bestandteil zur Bildung von Urteilen über Wahrscheinlichkeiten, werden jedoch häufig ignoriert («base rate neglect»). Probanden wurden in Tests etwa nach dem wahrscheinlichen Beruf einer Person gefragt. Dabei erhielten sie neben der Basisrate (es seien z.B. 30 Anwälte und 70 Ingenieure in einer Population von 100) nur Informationen über die Person, die in keine Richtung Hinweise enthielten. Die Probanden urteilten zumeist, dass es gleich wahrscheinlich

[29] Peter Wason: *Reasoning About a Rule*, in *Quarterly Journal of Experimental Psychology* 20 (1966) S. 273-281.

[30] Amos Tversky, Daniel Kahneman: *Extensional Versus Intuitive Reasoning: Conjunction Fallacy in Probability Judgment*, in *Psychological Review* 90 (1983) S. 293-315.

sei, ob die Person Anwalt oder Ingenieur sei (50%). Trotz der erhaltenen Basisrateninformation schienen sie also erneut die inhaltliche Beschreibung stärker heranzuziehen.[31] Sogar statistisch ausgebildete Angehörige der medizinischen Fakultät von Harvard (!) neigen dazu, Basisraten zu ignorieren, zumal bei der Schätzung bedingter Wahrscheinlichkeiten – wie bei der Frage, wie wahrscheinlich eine Krankheit ist, gegeben das Vorliegen bestimmter Symptome oder Testergebnisse.[32] Die fehlerhaften Urteile werden zum Beispiel dadurch erklärt, dass die Probanden sich eher auf leicht verfügbare Informationen stützen («availability heuristic»).

(4) Wenn Personen eine Reihe von Sachfragen erhalten, diese beantworten und hinterher befragt werden, für wie wahrscheinlich sie es halten, dass diese oder jene ihrer Antworten korrekt ist, überschätzen sich viele Versuchspersonen («overconfidence bias»).[33]

Die Vertreter des HB-Programms behaupten regelmäßig, dass diese und viele andere Fehler stabil sind. Die Subjekte wurden beispielsweise nach der ersten Versuchsreihe aufgeklärt – sie wurden auf die relevante Norm hingewiesen. Dennoch begingen sie die Fehler danach aufs Neue. Man spricht in dieser Tradition daher von «kognitiven Illusionen», um darauf hinzuweisen, dass weder das Wissen, dass eine bestimmte Meinung falsch ist, noch das Wissen, warum sie es ist, zur Korrektur der Meinung führt. Da viele der Studien in diesem Programm so unerfreulich ausgefallen sind, ist nicht selten geurteilt worden, dass wir Menschen allgemein wohl ziemlich irrational sind. Obwohl Kahneman und Tversky hier stets eher zurückhaltend formuliert haben, haben andere Vertreter des HB-Programms, und zumal manche ihrer Leser die Schlussfolgerung gezogen, die Resultate hätten «bleak implications for human rationality».[34]

31 Tversky/Kahneman, op. cit. (Fn. 28).

32 Das «Harvard medical school problem» (Ward Casscells, Arno Schoenberger, Thomas B. Grayboys: *Interpretation by Physicians of Clinical Laboratory Results*, in *New England Journal of Medicine* 299 [1978] S. 999-1001). Die Berechnung hat gemäß der Bayes'schen Regel zu erfolgen, was auch gewisse Schwierigkeiten auslöst: $p(A/B) = (p(B/A) \times p(A))/p(B)$.

33 Sarah Lichtenstein, Baruch Fischhoff, L. D. Phillips: *Calibration of Probabilities*: *The State of the Art to 1980*, in *Judgment under Uncertainty: Heuristics and Biases*, hg. von Daniel Kahneman, Paul Slovic, Amos Tversky (Cambridge, Cambridge University Press, 1982) S. 306-334.

34 Richard B. Nisbett, Eugene Borgida: *Attribution and the Psychology of Prediction*, in *Journal of Personal and Social Psychology* 32 (1975) S. 932-943, hier S. 935; vgl. u.a. Massimo Piattelli-Palmarini: *Inevitable Illusions: How Mistakes of Reason Rule our Minds* (New York: Wiley, 1994); Stephen Stich: *Could Man Be an Irrational Animal?*, in *Synthese* 64 (1980) S. 115-135.

Mele bezieht sich beispielsweise auf das Vorziehen von bestätigenden Evidenzen («confirmation bias»), leicht verfügbaren Informationen («availability heuristic») oder auch Fälle des Überschätzens der Korrektheit eigener Meinungen («overconfidence bias»). Diese und andere Mechanismen könnten durch Motive angeregt werden und so Selbsttäuschung erzeugen.[35] Kann Mele sich auf das HB-Programm verlassen, um Selbsttäuschung und ihre Irrationalität zu erklären?

Eine erste Merkwürdigkeit ist, dass Wason und Johnson-Laird die «confirmation bias» beim Testen von Hypothesen dadurch erklärt haben, dass die Versuchspersonen einer Selbsttäuschung anheimfallen. Der problematische Status dieses Konzepts wurde dabei nicht bemerkt.[36] Mele dürfte sich natürlich nicht auf diese Erklärung des Hangs zu Bestätigungen stützen. Wollte man die Erklärung zudem auf andere fehlerhafte Urteile oder Schlussfolgerungen ausdehnen, dann ließe sich das HB-Programm noch weniger zur Erklärung von Selbsttäuschung heranziehen – es sei denn, man hat eine Schwäche für zirkuläre Erklärungen. Konsequenter im Rahmen des Programms ist es, das Phänomen selbst als eine psychologisch zu erklärende Fehlleistung anzusehen. So haben Quattrone und Tversky (mit Übernahme der Begriffsexplikation von Gur und Sackeim) Selbsttäuschung der Verwechslung von kausalen und «diagnostischen» Zusammenhängen zugeordnet, einem Mechanismus, der auch andere Urteile und Entscheidungen in die Irre leite.[37]

Natürlich verfällt Mele nicht dem Fehler von Wason und Johnson-Laird. Doch, was wichtiger ist, sein Ansatz ignoriert die erheblichen Grenzen des HB-Programms. Dieses Programm wird schon seit einer Weile mit guten Gründen angegriffen.

Zum einen ist gezeigt worden, dass viele vom HB-Programm erzeugte Daten experimentelle Artefakte darstellen. So beruht die Missachtung der wahrscheinlichkeitstheoretischen Konjunktionsregel (Beispiel [2] oben) wohl auf Mehrdeutigkeiten der entscheidenden Termini des Materials. Im alltäglichen Sprachgebrauch besitzen 'und' und 'wahrscheinlicher als' noch andere legitime Bedeutungen als die in Logik und Wahrscheinlichkeitstheorie gebrauchten. Wenn das beachtet wird, kann man die scheinbar

35 Mele: *Self-Deception Unmasked*, op. cit. (Fn. 2) S. 3, 11, 28-31.

36 Wason/Johnson-Laird, op. cit. (Fn. 9).

37 Quattrone/Tversky, op. cit. (Fn. 12). Vgl. dazu unten Fn. 41.

fehlerhaften Antworten auch als in gewisser Hinsicht rational interpretieren. Den Versuchssubjekten wird Linda als aufgeweckte, an politischen Themen interessierte ehemalige Philosophiestudentin beschrieben; und dann wird gefragt, ob es wahrscheinlicher ist, dass Linda *(A)* eine Bankangestellte ist oder *(A&B)* eine Bankangestellte und Feministin ist. Wenn man von einem Experimentator vor eine so scharfe Alternative gestellt wird, mag man die erste Alternative *(A)* leicht so verstehen, dass bei ihr *(B)* ausgeschlossen sein soll – als ob *(A)* so viel bedeutet wie *(A&¬B)*. Das aber macht (*A*) in der Tat zur unplausiblen Antwort; also liegt *(A&B)* nicht so fern. Diese Wahl der Antwort mag noch dadurch verstärkt werden, dass die Probanden annehmen, dass die *Beschreibung* von Linda doch für die Aufgabenlösung relevant sein muss – warum sollte sie sonst angeführt werden? Gegeben die Beschreibung, scheint es doch plausibel, dass Lindas Bankangestelltendasein keineswegs ausschließt, dass sie gleichwohl Feministin ist. Zudem lassen sich durch bestimmte Testvarianten die fehlerhaften Antworten drastisch verringern. Wählt man statt Fragen nach subjektiven Wahrscheinlichkeiten Fragen nach objektiven Häufigkeiten («Von 100 Personen mit Lindas Charakteristika, sind wie viele *(A)* Bankangestellte oder *(A&B)* feministische Bankangestellte?»), verschwinden die Verzerrungen drastisch: Es kann schließlich nicht mehr feministische Bankangestellte als Bankangestellte geben. Hinter dieser Option steckt vermutlich mehr: Auch andere scheinbar unvermeidliche Fehlleistungen lassen sich nämlich beseitigen, wenn man statt nach subjektiven Wahrscheinlichkeiten nach natürlichen Häufigkeiten fragt.[38] Davidsons bekannter Auffassung zufolge sollten wir das *principle of charity* befolgen, also menschliches Verhalten so weit wie möglich als rational interpretieren. Wie diese Überlegungen zeigen, können wir das in sehr konkreter Weise.

Ähnliche Punkte ließen sich auch für die HB-Programm-Beispiele (I) und (III) vorbringen, aber aus Platzgründen verzichte ich hier darauf. Es sei nur noch bemerkt, dass dies keinesfalls Argumente sind, die aus dem

38 Klaus Fiedler: *The Dependence of the Conjunction Fallacy on Subtle Linguistic Factors*, in *Psychological Research* 50 (1988) S. 123-129; Lola L. Lopes: *The Rhetoric of Irrationality*, in *Theory & Psychology* 1 (1991) S. 65-82; Gerd Gigerenzer: *How to Make Cognitive Illusions Disappear: Beyond Heuristics and Biases*, in *European Review of Social Psychology* 2 (1991) S. 83-115; Ralph Hertwig, Gerd Gigerenzer: *The «Conjunction fallacy» Revisited: How Intelligent Inferences Look Like Reasoning Errors*, in *Journal of Behavioral Decision Making* 12 (1999) S. 275-305; Gerd Gigerenzer, Ulrich Hoffrage: *How to Improve Bayesian Reasoning Without Instruction: Frequency Formats*, in *Psychological Review* 102 (1995) S. 684-704.

berüchtigten philosophischen Lehnstuhl gegen die empirische Forschung vorgebracht werden. Viele der Punkte sind von Psychologen gesichtet und getestet worden – freilich teils von sprachphilosophisch informierten Psychologen: Die Reinterpretation des Verhaltens der Probanden im Linda-Test mittels der Annahme, die Informationen über Linda müssten doch relevant für die Aufgabenlösung sein, stützt sich auf Grice' Arbeiten zu Maximen der Konversation.[39]

Ich will lieber den Fall der sich selbst überschätzenden Wissenschaftler, den Mele als Beispiel von Selbsttäuschung ansieht, in ähnlicher Weise angreifen – was teils auf der Kritik am obigen Beispiel (IV) beruhen wird. Drei Punkte seien genannt:

(1) Es müsste auch hier geprüft werden, ob die Testfragen fair und ohne Mehrdeutigkeiten in den entscheidenden Fragestellungen formuliert worden sind. Was soll eine Frage wie «Halten Sie sich für kompetenter als den durchschnittlichen Wissenschaftler?» schon bedeuten? Kompetenter im Bereich der Lehre oder der Forschung? Und in welchen Forschungsbereichen? Welcher Durchschnitt sollte hier gemeint sein? Wenn Wissenschaftler nach ihren Fähigkeiten gefragt werden, mögen sie natürlich zuerst an das denken, worin sie wirklich Leistungen vollbracht haben. Der Blick auf diese eingegrenzten Kompetenzbereiche mag es sein, der sie meinen lässt, dass sie ihre durchschnittlichen Kollegen überragen.

(2) Es ist empirisch gezeigt worden, dass sich der scheinbar starre Hang zur Überschätzung («overconfidence bias») der Korrektheit eigener Meinungen sehr wohl beseitigen lässt. Wenn man die Versuchspersonen nämlich nicht fragt, für wie wahrscheinlich sie es halten, dass diese oder jene ihrer Antworten korrekt ist, sondern danach, wie hoch die Zahl der für korrekt gehaltenen Testantworten insgesamt ist (also wieder nach Häufigkeiten statt nach subjektiven Wahrscheinlichkeiten fragt), verschwinden die Fehlleistungen in dramatischem Umfang.[40]

(3) Sofern man die Urteile der Wissenschaftler als relativ zu ihren herausragenden Leistungen interpretieren kann, muss bei einem solchen Urteil kein Motiv vorliegen. Daher ist es fraglich – und bedürfte einer empirischen

39 Hertwig/Gigerenzer, *The «Conjunction fallacy» Revisited*, op. cit., beziehen sich etwa auf Paul Grice: *Logic and Conversation*, in *Syntax and Semantics 3: Speech Acts*, hg. von Peter Cole, Jerry L. Morgan (New York: Academic Press, 1975) S. 41-58.

40 Gerd Gigerenzer, Ulrich Hoffrage, H. Kleinbölting: *Probabilistic Mental Models: A Brunswikian Theory of Confidence*, in *Psychological Review* 98 (1991) S. 506-528.

statt einer begrifflichen Untersuchung –, ob es sich in Meles Sinn um Selbsttäuschung oder «nur» um eine Art von durch kognitive Heuristiken erzeugte verzerrte Meinung handelt.

Zweifellos ist es offen, ob sich alle von Mele angesprochenen Kandidaten für empirisch nachgewiesene Selbsttäuschungen so bestreiten lassen. Ich muss es an dieser Stelle dem Leser überlassen, hier selbst weiter nachzudenken.[41] Aber klar sollte sein, dass Mele keineswegs sicher sein kann, dass

41 Einen von Vertretern des HB-Programms als Selbsttäuschung beschriebenen Fall bestreitet Mele selbst, aber aus ganz anderen Gründen als den oben vorgeführten. Es handelt sich um die Studie von Quattrone/Tversky, op. cit. (Fn. 12): Probanden wurde hier erzählt, sie nähmen an einem Test über Typen von Herzen teil. Ihnen wurde gesagt, dass es Herzen zweierlei Typs gebe: Typ 1 neige zu mehr Krankheiten, während Typ 2 davon frei sei. Die Versuchspersonen wurden dann aufgefordert, ihren Arm so lange wie möglich in eiskaltes Wasser zu halten. Danach wurden sie auf ein Trainingsrad gesetzt, und während einer Hälfte der Gruppe gegenüber behauptet wurde, dass solch ein Training bei Personen mit Typ-2-Herz die Schmerztoleranz erhöhe, erklärte man der anderen Gruppe, das Training verringere die Schmerztoleranz von Typ-2-Herzen. In einer zweiten Versuchsrunde sollten die Probanden wieder den Arm in kaltes Wasser halten: bei fast allen verlängerte bzw. verkürzte sich die Ausdauer. Doch sie leugneten, dass sie versucht hätten, ihre Schmerztoleranz auszudehnen oder zu verringern (Quattrone und Tversky haben Mittel angewendet, um Lügen auszuschließen). Mele (*Self-Deception Unmasked*, op. cit. [Fn. 2] S. 84-93) bestreitet, dass die Probanden widersprüchliche Meinungen haben müssen: Die erste Gruppe muss gar nicht unbewusst glauben, dass sie versucht hätten, ihre Schmerztoleranz auszudehnen. Freilich könnten Quattrone und Tversky hier erwidern, dass es sich gleichwohl um Selbsttäuschung handeln könnte – zwar nicht nach Gurs und Sackeims Konzeption, sondern etwa nach Meles, die ja keine widersprüchlichen Meinungen erfordert. Eine andere Strategie der Kritik an Quattrone und Tversky wäre die, die behauptete Irrationalität der Probanden zu bezweifeln. Quattrone und Tversky sehen das Verhalten der Subjekte als irrational an, weil sich darin eine Verwechslung von kausalen und diagnostischen Zusammenhängen ausdrücke. Das heißt, das HB-Programm unterstellt hier die Richtigkeit der sog. kausalen Entscheidungstheorie: Wir sollen Handlungen nur wählen, um damit gewünschte Ziele zu verursachen, nicht aber, um Evidenzen von Zuständen zu erhalten, auf die wir keinen Einfluss haben (hier etwa: darauf, welches Herz man hat, hat man keinen Einfluss, und daher ist der Versuch, die Schmerztoleranz zu erhöhen (bei Gruppe 1) der Versuch, ein günstigeres Bild der eigenen Gesundheit zu gewinnen). Man mag nun die Grundannahme der kausalen Entscheidungstheorie bestreiten und an der sogenannten evidenzialen Entscheidungstheorie festhalten, die einfach fordert, eine Maximierung des konditionalen erwarteten Nutzenwertes zu verfolgen, ohne Rücksicht auf Kausalzusammenhänge. Dies ist jedoch nicht leicht, schon gar nicht im Fall der Selbsttäuschung. Jean-Pierre Dupuy (*Rationality and Self-Deception,* in *Self-Deception and Paradoxes of Rationality*, hg. von Jean-Pierre Dupuy [Stanford: CSLI Publications, 1998] S. 113-150) behauptet so eine Rationalisierung der Selbsttäuschung zu liefern, doch er baut letztlich nur darauf auf, mittels des Evidenzialismus Entscheidungen zu rationalisieren. Es bleibt dunkel, ob sich das auf Meinungen oder ähnliche kognitive Einstellungen übertragen lässt. Mir ist keine klare Diskussion von Quattrones und Tverskys Studie in dieser Richtung bekannt.

das HB-Programm eine gesicherte Basis für seine Konzeption und für die Existenz von Selbsttäuschungen darstellt. Angesichts des problematischen Charakters von Selbsttäuschung gilt erneut die Maxime: Die Beweispflicht liegt aufseiten dessen, der die Existenz des Phänomens behauptet.

8. Wie (weit) lässt Selbsttäuschung sich rationalisieren?

Die Debatte über das HB-Programm geht noch tiefer. Sie betrifft nämlich sogar die Basis von Rationalitätsstandards selbst. Wann – bei welchen Aufgaben – dürfen wir welche Regeln als normativ gültig einsetzen? Was ist eigentlich das Fundament von Rationalitätsstandards? Dies kann ich hier nicht weiter ausführen, da die Debatte zu komplex und in vielen Hinsichten offen ist.[42] Klar ist allerdings, dass es dubios ist, Selbsttäuschung nach dem HB-Programm als motivierte Art von verzerrten Meinungen zu verstehen und *deshalb* als irrational zu bewerten. Umgekehrt formuliert: Es lässt sich zumindest in Grenzen plausibel machen, dass Selbsttäuschung aus Gründen rational sein kann, die von anderer Art sind als die vom Standardmodell vorgegebenen Prinzipien des Urteilens und Entscheidens.

Eine erste Konkretisierung dieser Idee ist die, dass Selbsttäuschung auf praktischer (im Sinne von instrumenteller) Rationalität beruht: dass

42 Vgl. etwa Michael Bishop: *Reflections on a Normative Psychology*, in *Philosophie: Grundlagen und Anwendungen*, hg. von Ansgar Beckermann, Holm Tetens, Sven Walter (Paderborn: Mentis, 2008) S. 249-262; L. Jonathan Cohen: *Can Human Irrationality Be Experimentally Demonstrated?*, in *Behavioral and Brain Sciences* 4 (1981) S. 317-331; Alvin Goldman: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986); A. G.: *Human Rationality: Epistemological and Psychological Perspectives*, in *Philosophie: Grundlagen und Anwendungen*, hg. von Ansgar Beckermann, Holm Tetens, Sven Walter (Paderborn: Mentis, 2008) S. 230-247; Stein, op. cit. (Fn. 18); Thomas Sturm: *What Is the Foundation of Norms of Rationality?*, in *Philosophie: Grundlagen und Anwendungen*, hg. von Ansgar Beckermann, Holm Tetens, Sven Walter (Paderborn: Mentis, 2008) S. 189-201.

sie ein Resultat der Verbindung von subjektiven Zielen mit angemessenen praktischen Überlegungen ist.[43] Dies ist allerdings eine starke Art des Intentionalismus, die allenfalls für einige Fälle geeignet ist. Einen moderateren Ansatz bietet die psychologische Konzeption von Rationalität an, die den Hauptkonkurrenten zum HB-Programm darstellt. Dieser moderate Ansatz lehnt die Annahme des HB-Programms ab, dass Rationalitätsnormen nur die Regeln des «Standardmodells» sein können. Stattdessen sollten wir die Idee einer «beschränkten Rationalität» *(bounded rationality)* zulassen: Regeln sind nur insofern rational oder normativ gültig, als sie eingegrenzten Aufgaben und Kontexten angepasst sind.[44] Dieser Ansatz geht zudem davon aus, dass Normen eine evolutionäre Anpassung aufweisen müssen, um erfolgreiches Räsonnieren zu ermöglichen. Dies konvergiert mit einer Tendenz in empirischen Studien über Selbsttäuschung. Anstatt die Irrationalität der Selbsttäuschung hervorzuheben, wird zunehmend betont, welche Vorteile die Selbsttäuschung haben mag: etwa die Verminderung von Stress oder die Erhaltung von Selbstachtung und Wohlbefinden[45] oder das effektivere Verstecken der wahren Absichten vor anderen Personen: Täuschung hat einen adaptiven Nutzen, und Selbsttäuschung verstärkt die Fähigkeit zur Täuschung.[46]

Doch können alle Fälle von Selbsttäuschung so rationalisiert werden? Obwohl es verführerisch ist, dagegen Gedankenexperimente oder empirische Gegenbeispiele anzuführen – etwa Studien, die die Selbsteinschätzungen von Autofahrern als Selbsttäuschungen beurteilen, die keinesfalls rational sind – werde ich mich solcher Einwände hier enthalten. Ich möchte lieber drei andere Schwierigkeiten hervorheben:

43 So etwa Amelia O. Rorty: *Belief and Self-Deception*, in *Inquiry* 15 (1972) S. 387-410; Davidson, op. cit. (Fn. 6).

44 Vgl. Gerd Gigerenzer: *Adaptive Thinking* (New York: Oxford University Press, 2000).

45 Larry J. Jamner, Gary E. Schwarz: *Self-Deception Predicts Self-Report and Endurance of Pain*, in *Psychosomatic Medic*ine 48 (1986) S. 211-223; James F. Welles: *Self-Deception as a Positive Feedback Mechanism*, in *American Psychologist* 41 (1986) S. 325-326; Lockhard/Paulhus, op. cit. (Fn. 12); Sahdra/Thagard, op. cit. (Fn. 3).

46 Robert Trivers: *Social Evolution* (Menlo Park, CA: Benjamin/Cummings, 1985); R. T.: *The Elements of a Scientific Theory of Self-Deception*, in *Annals of the New York Academy of Sciences* 907 (2000) S. 114-131; Christopher C. Byrne, Jeffrey A. Kurland: *Self-deception in an Evolutionary Game*, in *Journal of Theoretical Biology* 212 (2000) S. 457-480.

(1) Zunächst sollte klar sein, dass es sich um Rationalisierungen handelt, die aus externer oder Beobachterperspektive gebildet werden. Angenommen, Selbsttäuschung erfüllt die erwähnten Funktionen und liefert diese oder andere nützliche Ergebnisse, dann kann sie das – zumindest typischerweise – nur, insofern sie den Subjekten nicht bewusst ist und von ihnen nicht intentional geplant wurde (lassen wir hier die Grenzfälle intentionaler Selbsttäuschung außer acht, bei denen auf die Forderung der Gleichzeitigkeit der inkompatiblen Meinungen verzichtet wird). Wir verlangen aber von Rationalität nicht bloß, dass sie zu vorteilhaften Resultaten führt. Wir verlangen auch, dass sich die Subjekte im Prinzip bewusst sein können, *was sie tun und warum sie es tun*. Wenn eine Gruppe von Affen zufällig korrekte Kalkulationen oder Schlussfolgerungen auf Papier kritzelt, ist das noch kein hinreichendes Zeichen von Rationalität. Man muss auch auf in gewissem Umfang reflektierte Weise zu den Resultaten gelangt sein. Darauf baut schließlich das Lernen, Prüfen und Verbessern von Strategien und Lösungen im Urteilen und Entscheiden auf. Sobald diese Forderung aber auf Selbsttäuschung angewendet wird, ist nicht zu sehen, wie sie im gleichen Sinn eine rationale Strategie sein kann, wie wir es von anderen Problemlösungsmethoden meinen. Das gilt übrigens unabhängig davon, ob man in normativer Hinsicht das Standardmodell oder das adaptive Modell von Rationalität bevorzugt.

(2) Ein methodologisches Problem der angedeuteten Rationalisierungen besteht erneut darin, dass manche Psychologen nur ungenaue Vorstellungen von Selbsttäuschung haben. So untersucht Whittaker-Bleuler den Zusammenhang zwischen Täuschung, Selbsttäuschung und sozialer Dominanz im Tennis.[47] Sie argumentiert wie folgt: Indem ein Tennisspieler sich über seine Leistungsfähigkeit, den allgemeinen Stand des Spiels und ähnliches täuscht, kann er seine Unsicherheit besser gegenüber dem Gegner verbergen. So mag er den Kopf aufrecht halten, statt ihn zaudernd zu schütteln, oder es vermeiden, immer wieder wie zur Übung Trockenübungen mit dem Tennisschläger zu machen – kurz, er mag *cool* bleiben. Der Grad der Selbsttäuschung muss Whittaker-Bleuler zufolge dann besonders hoch sein, wenn der sich selbst täuschende Spieler in einen Punkterückstand gerät und sich dennoch besonders dominant verhält. Egal, wie das Spiel verlief,

47 Sharon Whittaker-Bleuler: *Deception and Self-Deception: A Dominance Strategy in Competitive Sport*, in Lockhard/Paulhus, op. cit. (Fn. 12) S. 212-228; für Schwimmer: Joanna E., Carolina F. Keating: *Self-Deception and Its Relationship to Success in Competition*, in *Basic and Applied Social Psychology* 12 (1991) S. 145-155.

Pete Sampras hat bekanntlich fast nie Unsicherheit oder Schwäche gezeigt. Jedoch macht Whittaker-Bleuler nicht klar, warum man das als Selbsttäuschung verstehen muss. So müssen in diesem Fall ja keine inkompatiblen Meinungen vorliegen. Sampras hat vielleicht die Hinweise auf den aktuellen Punktestand oder auf die letzten Punkte nie so wichtig genommen wie seine Fitness, seine exzellente Technik, seine Fähigkeit, sich vollkommen auf den nächsten Ballwechsel zu konzentrieren. Es mag auch bezweifelt werden, dass jemand, der mehr Grand-Slam-Titel gewonnen hat als Rod Laver, dies wegen (oder auch nur teils wegen) selbsttäuscherischer Strategien erreicht hat. Er mag einfach ein berechtigtes Vertrauen in sich selbst haben. Auch der Versuch, ihm eine wenn auch noch so kleine Überschätzung der eigenen Fähigkeiten zuzuschreiben, wäre absurd, weil man dann auch den Besten Selbstüberschätzung unterstellen würde – was es wiederum unmöglich macht, klare Begriffe von berechtigtem Vertrauen in sich selbst und unberechtigter Selbstüberschätzung zu bilden. Daher kann nicht einfach auf ein anderes Konzept von Rationalitätsstandards zurückgegriffen werden, um kurzerhand zu zeigen, dass Selbsttäuschung rational ist.

Sahdra und Thagard[48] haben in einem Artikel auf empirische Studien verwiesen, in denen Selbsttäuschung zur Erklärung von verbreiteten Verhaltensweisen herangezogen werde: den «positiven Illusionen» von Soldaten im Kampfeinsatz,[49] dem Verneinen von Krankheit,[50] dem zuversichtlichen Optimismus von Selbständigen[51] und dem fahrlässigen Verhalten von Berufskraftfahrern.[52] Doch nur in der letzten dieser vier Studien herrscht begriffliche Klarheit. Auch diese Autoren folgen Sackeim und Gur. In der ersten Studie wird der Begriff nicht genau analysiert und manchmal mit «self-serving bias», manchmal mit dem Konzept von positiven Illusionen gleichgesetzt. Im Grunde könnte dieser Autor (Wrangham) auch einfach von Irrtümern sprechen. In der zweiten und dritten Publikationen kommt der

48 Sahdra/Thagard, op. cit. (Fn. 3).

49 Richard Wrangham: *Is Military Incompetence Adaptive?*, in *Evolution and Human Behavior* 20 (1999) S. 3-17.

50 Rainer Goldbeck: *Denial in Physical Illness*, in *Journal of Psychosomatic Research* 43 (1997) S. 575-593.

51 Gholamreza Arabsheibani et al.: *And a Vision Appeared Unto Them of a Great Profit: Evidence of Self-Deception Among the Self-Employed*, in *Economics Letters* 67 (2000) S. 35-41.

52 Timo Lajunen et al.: *Impression Management and Self-Deception in Traffic Behavior Inventories*, in *Personality and Individual Differences* 22 (1996) S. 341-353.

Begriff der Selbsttäuschung gar nicht oder so gut wie gar nicht vor, weder als explanatorisches Konstrukt noch als etwas, dessen Existenz demonstriert oder erklärt wird. Umso weniger kann in diesen Studien gezeigt worden sein, dass das, was hier jeweils rationalisiert wird, wirklich ein Fall von Selbsttäuschung ist.

(3) Schließlich sollten wir nicht übersehen, dass zahlreiche der oft als «gewöhnlich» bezeichneten Fälle von Selbsttäuschung Beispiele von Lebenssituationen voller Unklarheiten und Ungewissheiten sind. Diese erlauben schwerlich eindeutige Zuschreibungen von Meinungen oder anderen für Selbsttäuschung relevanten propositionalen Einstellungen. Ein Fall wie der des sich selbst täuschenden Ehemanns mag dabei noch lösbar sein. Wenn er seine Frau verfolgen und am Ende das untreue Handeln beobachten kann, so reduziert sich seine Unsicherheit auf Null, und er mag sich wohl sogar selbst eingestehen können, zuvor einer Selbsttäuschung erlegen zu sein. Aber unzählige andere Fälle sind ganz anders. Kann man sich jemals im Klaren darüber sein, die beste Partnerin oder den besten Partner geheiratet zu haben, den optimalen Job gefunden zu haben? Kann man sich sicher sein, wie die Aussichten für die eigenen künftigen Unternehmungen stehen, sei es als Sportler oder Soldat, Geschäftsmann oder Glücksritter? Sofern die Antworten hier negativ ausfallen, steht auch die Zuschreibung von propositionalen Einstellungen, insbesondere von Meinungen und Motiven, auf schwankendem Grund – oft zu schwankend, um fairerweise eine schwerwiegende Zuschreibung wie die der Selbsttäuschung oder der Irrationalität zu erlauben.

9. Was gefordert wäre

Was kann man aus dem Gesagten für die weitere Erforschung der Selbsttäuschung lernen? Zumindest zweierlei: Erstens benötigen empirische Studien und Theorien über die Selbsttäuschung eine klare Analyse des Phänomens, aber genauso eine bewusst gemachte und reflektierte normative Theorie der Rationalität. Zweitens wären empirische Nachweise von Selbsttäuschung die notwendige Basis für jegliches weitere Forschen und Theoretisieren über das Phänomen. Solche Nachweise erfordern nicht nur, dass man im konkreten Fall nachweist, dass die begrifflichen Bedingungen für Selbsttäuschung erfüllt werden. Zusätzlich ist auch folgendes notwendig:

(1) die Wahl einer speziellen Regel;
(2) der Nachweis, dass diese Regel für gegebene Testaufgaben die normativ angemessene ist;

(3) der Nachweis, dass die Regel von den Probanden in der Test-Situation erfasst werden kann; und
(4) der Ausschluss der Möglichkeit, dass angebliche Fälle von Selbsttäuschung wohlwollend als etwas Anderes interpretiert werden können.

Es sollte klar sein, wie anspruchsvoll solche Forderungen sind. Ich behaupte keinesfalls, dass man ihnen nicht nachkommen kann. Wir sollten jedoch philosophische Konzeptualisierungen ablehnen, die ungeprüft gewisse empirische Befunde und Theorien aus der Psychologie übernehmen. Ebenso wenig sollten wir glauben, wir wüssten schon, dass Selbsttäuschung irrational (oder rational) ist. Es bedarf erheblich verbesserter Anstrengungen zu einer Zusammenarbeit von Philosophen und Psychologen, bevor wir bei diesem Thema weiterkommen können.