

Selbsttäuschung leicht gemacht
Bemerkungen zu Alfred R. Meles deflationärer Position in *Self-Deception
Unmasked*, Princeton Univ. Pr., Princeton and Oxford 2001.

Von Guido Löhrer (Bern)

Abstract: This article deals with Mele's antiagency view of self-deception, which takes falsity of the biased belief, motivationally biased treatment of data being also a nondeviant cause of the false belief, together with greater warrant for its falsity, to be jointly sufficient conditions for entering self-deception. Mainly concerned with a critique of the first condition, I also challenge Mele's suggestion of a test for self-deception, involving an impartial control group. In the face of the phenomenon's frequency it becomes questionable whether such a group might be found. The more common self-deception is, the more difficult it becomes to detect it.

I. Eine deflationäre Theorie der Selbsttäuschung

Philosophische Auffassungen gewinnen in Auseinandersetzung mit ihren Opponenten oftmals deutlichere Konturen. Diesen Umstand macht sich Mele zunutze, wenn er seine Theorie der Selbsttäuschung mit solchen Auffassungen konfrontiert, die das Phänomen – lexikalisch orientiert – analog zur Fremdtäuschung behandeln und Selbsttäuschung als denjenigen Sonderfall absichtlicher Fremdtäuschung ansehen, bei dem die täuschende und die getäuschte Person dieselbe ist.

Mele markiert in Kapitel 1 zwei Probleme. Erstens müsste eine Person, die sich im Glauben, dass $\neg p$, mit Absicht dazu bringt, gleichsam wider besseres Wissen p zu glauben, im Zustand der Selbsttäuschung zugleich von $\neg p$ und von p überzeugt sein (*dual-belief requirement* (76)).¹ Dies beschwört ein „static puzzle“ (6 f., 59 f.) herauf, zu dem sich ein „dynamic puzzle“ (8 ff., 60 ff.) gesellt. Denn zweitens fragt sich, ob die Kenntnis dieses Widerspruchs und das Bewusstsein von der Täuschungsabsicht nicht jeden Versuch einer absichtlichen Selbsttäuschung unterlaufen oder vereiteln müssten. Diese Schwierigkeit könnte vielleicht nur mittels der Annahme vermieden werden, sich selbst

¹ Dies in dem Sinne, dass eine sich selbst täuschende Person in diesem Moment sowohl glaubt, dass p , als auch glaubt, dass $\neg p$ ($Bp \ \& \ B\neg p$), nicht aber in dem Sinn, dass sie glaubt, dass $p \ \& \ \neg p$ ($B(p \ \& \ \neg p)$) (76); vgl. Donald Davidson, „Who Is Fooled?“ (1997), in: ders., *Problems of Irrationality*, Oxford 2004, 213-230, hier: 217. Die Möglichkeit von Überzeugungszuständen des ersten Typs hatte am Beispiel von Pierre, der – nicht wissend, dass es sich um dieselbe Stadt handelt – das ihm vom Hörensagen bekannte *Londre* für *jolie* hält, das dann tatsächlich von ihm besuchte *London* dagegen *not pretty* findet, seinerzeit Saul A. Kripke, „A Puzzle about Belief“, in: Avishai Margalit (Hrsg.), *Meaning and Use*, Dordrecht 1979, 239-283, bes. 254-267, herausgestellt, dessen Argumentation zu übergehen Mele sich (125 Anm. 5) explizit gestattet.

täuschende Personen seien in ein aktiv täuschendes und ein getäuschtes Selbst fragmentiert und diese Relata träten, obgleich in zentralen Punkten überlappend, in dissonanten Details relativ selbständig (semi-autonom) gegeneinander auf.² Selbsttäuschung wäre jedoch in diesem Fall durch zwei Bedingungen charakterisiert:

- (i) Eine Person *A* täuscht eine Person *B*, welche dieselbe sein mag wie *A*, nur dann dahingehend, dass *B* *p* glaubt, wenn *A* glaubt, dass $\neg p$, und wenn *B*s Überzeugung, dass *p*, von *A* verursacht ist.
- (ii) Selbsttäuschung ist wie Fremdtäuschung eine intentionale Tätigkeit (6).

Gegen diese theoretische Option bringt Mele ein Arsenal von Argumenten vor, das eine eigene, von ihm als deflationär charakterisierte Sicht (4) profilieren, verteidigen und stärken (76) hilft. Zum einen kommt sie ohne Intentionalität und Agentivität aus, so dass eher von Selbsttäuschungsprozessen (27, 102) als von Handlungen der Selbsttäuschung die Rede sein sollte. Zum anderen weist sie das Erklärungsmuster der personalen Spaltung als mental exotisch zurück (4). Stattdessen legt Mele den Akzent seiner Theorie auf die Einflüsse und Prozesse, die uns dazu verleiten, Daten und Informationen verzerrt oder befangen (*biased*) zur Kenntnis nehmen bzw. gewichten, und uns auf diese Weise zu verzerrten Überzeugungen gelangen lassen, ohne dass wir zur Erklärung solcher Vorgänge Absichtlichkeit ins Spiel bringen müssten. Verstehen Davidson und Pears und mit ihnen diejenigen Autoren, die Mele zu den Lexikalistern rechnet, das „selbst“ (*self*) in „Selbsttäuschung“ (*self-deception*) in erster Linie reflexiv, so hängt Mele einer eher possessiven Deutung an. Gegen deren „Ich täusche mich selber“ steht sein „Ich selbst werde von mir eigenen, aber kognitiv weitgehend unverfügbaren Mechanismen getäuscht“. Doch zieht dieser Ansatz womöglich neue Schwierigkeiten auf sich. Denn hat der lexikalisch orientierte intentionale Ansatz verständlich zu machen, wie Selbsttäuschung psychologisch überhaupt möglich ist, so droht bei Meles Theorie die Demarkationslinie zwischen Selbsttäuschung und Irrtum unscharf zu werden,³ und es fragt sich, was Selbsttäuschung von unkontrollierbaren Prozessen unterscheidet, die uns beherrschen (101 f.).

² Vgl. Davidson, 1997 (Anm. 1), 220 f., u. ders., „Paradoxes of Irrationality“ (1982), in: ders., *Problems of Rationality*, Oxford 2004, 169-187, hier: 181-185.

³ Mele scheint dies hinsichtlich der eigenen Theorie nicht zuzufürchten, erhebt aber einen vergleichbaren Einwand (105 f.) gegen den rein kognitiven Ansatz von Martha Knight, „Cognitive and Motivational Bases of Self-Deception: Commentary on Mele’s *Irrationality*“, *Philosophical Psychology* 1 (1988), 179-188, bes. 183.

II. Die Kontrarität der Auffassungen und Meles Argumentationsstrategie

Mit Blick auf motivational verzerre Überzeugungen kontrastiert Mele zwei Sichtweisen miteinander:

„1. The agency view: all motivationally biased beliefs are intentionally produced or protected. In every instance of motivationally biased belief that p , we try to bring it about that we acquire or retain the belief that p , or at least try to make it easier for ourselves to acquire or retain the belief.

2. The antiagency view: no motivationally biased beliefs are intentionally produced or protected. In no instance of motivationally biased belief that p does one try to bring it about that one acquires or retains the belief or try to make it easier for oneself to acquire or retain the belief.“ (13)

Agency view und *antiagency view* stehen nach Mele in konträrem Verhältnis zueinander. Sie können also nicht beide wahr sein. Doch können beide falsch sein. Man vermutet, wie Mele (13) sagt, dass die Wahrheit irgendwo in der Mitte zwischen den beiden Polen liege. Doch frage sich, welche der beiden der Wahrheit mit größerer Wahrscheinlichkeit näher komme. Die Überlegungen, die Mele daraufhin in Kapitel 2 anstellt, betreffen jedoch keineswegs die damit angesprochene graduelle Abwägung. Vielmehr zielen sie auf den Nachweis, dass sich als erklärungsstark und haltbar ausschließlich eine Auffassung erweist, die in allen Punkten und für sämtliche diskutierten Beispiele mit dem *antiagency view* und in keinem Punkt mit dem *agency view* zusammenstimmt (49).

Dieses Vorgehen birgt Probleme, die auf den Stil des Buchs abfärben.

Die konträren Positionen *agency view* und *antiagency view* sind durch Allaussagen bestimmt und können folglich durch ein einziges einschlägiges Gegenbeispiel widerlegt werden. Um die Wahrheit des von ihm favorisierten *antiagency view* aufweisen zu können, muss Mele zeigen, dass die dazu kontradiktorische Position – „Es gibt mindestens eine absichtlich herbeigeführte oder bewahrte (abgesicherte) motivational verzerre Überzeugung“ – falsch ist. Da es dafür, wie es aussieht, keinen einfachen, eleganten Beweis gibt, bleibt Mele nichts übrig, als mögliche Typen von Einwänden anhand von Musterbeispielen zu behandeln und die mit ihnen verbundenen Ansprüche Mal für Mal zurückzuweisen.⁴ Dies verleiht dem Buch generell und Kapitel 2 im besonderen die Form eines Rondeaus: Auf die varianten- und variationsreiche Diskussion möglicher Gegenbeispiele folgt jeweils wie ein Refrain die Versicherung, dass zur Erläuterung des jeweils in Rede stehenden Falls wieder keines der den *agency view* charakterisierenden Momente beigezogen müsse (33, 38, 41 f., 42, 44, 46, 48, 49, vgl. auch 59, 69 f., 73 f., 90).

⁴ „Obviously, it is impossible to examine an endless series of cases of self-deception that are allegedly isomorphic in the relevant respects with stereotypical interpersonal deception. [...] I argue against the leading attempted empirical demonstrations of such self-deception.“ (75)

Auf diese Weise könnte es Mele allerdings gelingen, eine kritische Masse an Punkten zusammenzutragen, die es nahe legen, die Beweislast umzukehren. Statt eine deflationäre Position weiterhin verteidigen müssen, wäre dann der Opponent genötigt zu zeigen, welche Beispiele und welche Erklärungsdesiderate den Gebrauch agentiver Momente zwingend erforderlich machen, weil Fälle von Selbsttäuschung andernfalls nicht adäquat beschrieben und hinreichend erklärt werden könnten. Zumindest aber müsste er zeigen können, dass die Intentionalitätshypothese nicht etwa weniger plausibel ist als Meles Vorschlag (59).

III. Nichtagentive Sichtweise und mechanistische Wissenschaftsmetaphorik

Mele illustriert seine nichtagentive Auffassung geradliniger Selbsttäuschung (*straight self-deception*) mit vier Beispielen (26 f.). 1. Dons hoffnungsmotivierte negative Fehlinterpretation (*negative misinterpretation*) kritischer Gutachterkommentare, die ihn annehmen lässt, sein Aufsatz wäre fälschlich abgelehnt worden. 2. Sids von Leidenschaft getragene positive Fehlinterpretation (*positive misinterpretation*) der Zurückweisung seiner Liebe zu Roz, die er als Liebesbeweis und Prüfung der Ernsthaftigkeit seiner Ambitionen deutet. 3. Beths selektive Fokussierung (*selective focusing/attending*) auf Hinweise, die ihrem Wunsch entsprechen, Liebbling des unterdessen verstorbenen Vaters gewesen zu sein, wobei Zeugnisse, die dagegen sprechen könnten, unterdrückt werden. Und 4. die selektive Evidenzsuche (*selective evidence-gathering*) der Wahlhelferin Betty, die von der Hoffnung motiviert und verzerrt wird, ihr Kandidat möge kein Sexist sein, was er in Wahrheit – wie der autoritative Erzähler des Beispiels besser weiß – jedoch ist.

In keinem der vier Fälle glaube, so Mele, die betroffene Person zunächst $\neg p$, um sodann mit Selbsttäuschungsabsicht (*intentionally*) herbeizuführen, dass sie p glaubte. Statt Selbsttäuschungen als Handlungen nach einem *belief/desire*-Modell agentiv zu deuten, sei es angemessen, sie als Prozesse (27, 61) zu begreifen, die einerseits von Wünschen ausgelöst und getragen (*trigger/prompt/prime and sustain* (27-30, 32, 43, 53, 62)) sein können, andererseits aber auch in Gestalt kalter bzw. nichtmotivierter Verzerrung von Überzeugungen auftreten. Für letztere unterscheidet Mele drei Fälle (28 f.): 1. Verzerrungen aufgrund der Lebendigkeit der Information in Abhängigkeit von individuellen Interessen, 2. solche aufgrund der Verfügbarkeit von Information in Abhängigkeit von individuellen Zugangsmöglichkeiten und dem Aufwand, der für den Zugang geleistet werden müsste, und 3. Bestätigungsverzerrungen, die dadurch entstehen, dass wir, aufgefordert eine beliebige Hypothese zu prüfen, dazu neigen, die verfügbaren Daten eher

als Bestätigungen denn als Einwände gegenüber dieser Hypothese zu deuten. Die Motivation tritt in diesen Fällen nicht als Auslöser, sondern lediglich als Verstärker kalter Überzeugungsverzerrung mit Rücksicht auf die Lebendigkeit und Zugänglichkeit von Informationen auf (29 f.). Doch zeigt die Beobachtung, dass Hüttenwirte eher dazu neigen, Verstöße gegen die Hüttenordnung zu konstatieren, als Hüttenbesucher, die dieselbe Hypothese testen (33), dass Bestätigungsverzerrung noch nicht die ganze Erklärung sein kann.

Mele profitiert von dem von Friedrich, Trope und Liberman entwickelten Modell (FTL bzw. PEDMIN („primary error detection and minimization“)), nach dem pragmatisch Denkende eher mit dem Aufspüren und Minimieren kostspieliger Irrtümer und Fehler als mit Wahrheitsprüfung beschäftigt sind. Dabei kommen nicht eigentlich Minimierungsintentionen, sondern automatische Teststrategien zum Zuge (31 f.), die den Aufwand für einen Informationsgewinn, die Kosten für die fälschliche Akzeptanz einer Hypothese und die Kosten für deren fälschliche Ablehnung zueinander in Beziehung setzen und entsprechende Akzeptanzschwellen gleichsam mechanisch festlegen. Nun treibt der Wunsch, p möge wahr sein, die Kosten dafür, fälschlicherweise $\neg p$ zu glauben, stärker in die Höhe als die Kosten dafür, fälschlich das gewünschte p zu glauben. Und da erstens mit der Abneigung dagegen, etwas fälschlich zu glauben, auch die Schwelle, es überhaupt zu glauben, ansteigt (42), und zweitens eine geringere Akzeptanzschwelle für p als für $\neg p$ die Wahrscheinlichkeit, dass man p glauben wird, erhöht, so wird man im Fall, dass man p wünscht, wahrscheinlich auch dazu kommen, p zu glauben.

Da das FTL-Modell die motivationale Verzerrung von Überzeugungen zu erklären erlaubt, ohne Selbsttäuschungsabsichten ins Spiel zu bringen, darf Mele insistieren, dass Selbsttäuschung nicht eine Handlung ist, mit der man sich absichtlich wider besseres Wissen wunschgemäße Überzeugungen zulegt. Die Beschreibung des Phänomens kann demnach auf ein intentionalistisches Vokabular verzichten. Statt seiner favorisiert Mele eine mechanistische Metaphorik (30, 46), die *automatische Strategien* (32, 43, vgl. 60) der *Vermeidung* von Fehlern kennt, welche ihrerseits durch *Aversion* (42 ff., 48) gegen psychische Unannehmlichkeiten *ausgelöst* und in Aktion gehalten werden.⁵ Selbsttäuscher handeln nicht, sondern reagieren ausweichend. Wird Selbsttäuschung so weitgehend von einer motivational verzerrten Haltung zu alternativen Hypothesen p und $\neg p$ und einem entsprechenden Erwerb der kostenärmsten Überzeugung p getragen, dann

⁵ Methodisch problematisch scheint, dass Mele hier, statt durchgängig von Mechanismen zu reden, mit „strategy“ einen Ausdruck verwendet, der im Normalfall intentional gebraucht wird und entsprechende Konnotationen weckt, um seine Intentionalität für den eigenen Ansatz jedoch sogleich zurückzuweisen. Dies weckt den Verdacht, Mele könne den Purismus seiner Theorie nur durchhalten, weil er derlei Anklänge und Ungenauigkeiten zulässt. Vgl. auch den Gebrauch des Verbs „seek“ (43), das zu einem intentionalen Vokabular zu gehören scheint.

scheint eine aktuelle Überzeugung $\neg p$ für eine Selbsttäuschung nicht nötig. Denn der aversionsgesteuerte Mechanismus kommt selbst dann in Gang, wenn die Person zunächst weder einen Grund hat, p zu glauben, noch einen, von $\neg p$ überzeugt zu sein, und lediglich wünscht, p möge wahr sein.

Allerdings könnte dies den Bedingungen widerstreiten, unter denen etwas sinnvoll gewünscht werden kann. Beths Bemühung, sich im nachhinein zu Vaters Liebling zu machen, scheint die Überzeugung vorauszusetzen, eben dies nicht gewesen zu sein. Denn nur das kann herbeigeführt, was nicht bereits der Fall ist, und etwas herbeiführen zu wollen ist nur dann rational, wenn man annimmt, es sei nicht bereits der Fall.⁶ Doch dürfte, was für Wünsche im Rahmen eines *belief/desire*-Modells des Handelns gilt, nicht auf Wünsche aller Art zutreffen.⁷ So gibt es neben Wünschen, die von der Überzeugung begleitet werden, der erwünschte Zustand stehe noch aus und könne entweder absichtlich herbeigeführt werden oder sich womöglich von selbst einstellen,⁸ solche, bei denen der p Wünschende den aktuellen Zustand (p oder $\neg p$) nicht kennt und p nicht herbeiführen kann, sondern hofft, das p der Fall sei oder werde, sobald sich der Schleier seines Nichtwissens hebt.⁹ Hier könnten allein der ängstliche Wunsch (*anxious desire*), dass p ,¹⁰ die Angst vor $\neg p$ (100) und die unterschiedliche psychische Belastung, die mit einer möglicherweise irrigen Überzeugung von der einen oder der anderen Hypothese verbunden sind, eben den Verzerrungsmechanismus aktivieren, der eine Selbsttäuschung produziert. So würde Paul, der beim alpinen Alleingang in eine Gletscherspalte gestürzt ist und sich nicht selber befreien kann, allein aufgrund seines ängstlichen Wunsches, gerettet zu werden, und der Aversion gegenüber psychisch kostspieligen Annahmen und Irrtümern auch dann zu der verzerrten Überzeugung kommen, eine Rettungsmannschaft sei zu ihm unterwegs, wenn er über deren Aufbruch oder Nichtaufbruch gänzlich im Ungewissen ist.

⁶ Vgl. Aristoteles, *Nikomachische Ethik* VI 2, 1139b 5-7.

⁷ Hier gilt zudem, dass zu keinem Moment der Handlung zugleich p und $\neg p$ gewünscht oder geglaubt wird. Zwar ist der Wunsch nach p nur dann sinnvoll, wenn man $\neg p$ glaubt. Doch hat man p willentlich und wissentlich herbeigeführt, so verschwindet damit, während man nun weiss, dass p , sowohl der Wunsch, dass p , als auch die Überzeugung dass $\neg p$.

⁸ Letzteres entspricht in etwa Kants Unterscheidung von Wollen und Wünschen in der *Grundlegung zur Metaphysik der Sitten*, Akad.-Ausgabe Bd. 4, Berlin 1968, 394.

⁹ Fälle dieser Art würden zeigen, dass Davidsons Bedingung, wer sich dahingehend täusche, p zu glauben, müsse sich zugleich der starken, ja größeren Evidenz für $\neg p$ bewusst sein, nicht notwendig ist, nämlich nicht erfüllt sein muss, damit es zur ((angst-)motivierten) Selbsttäuschung kommt. Vgl. Davidson, „Deception and Division“ (1986), in: ders., *Problems of Irrationality*, Oxford 2004, 199-212, hier: 209: „[C]ore cases of self-deception demand that Carlos remain aware that his evidence favours the belief that he will fail, for it is awareness of this fact that motivates his efforts to rid himself of the fear that he will fail.“

¹⁰ Vgl. Annette Barnes, *Seeing through Self-Deception*, Cambridge 1997, 37-39 u. 67, und die Diskussion bei Mele (54-56).

IV. Meles Bedingungen für Selbsttäuschung

In eigener Sache gefordert, bestimmt Mele Selbsttäuschung in Kapitel 3 anhand des folgenden Bedingungskatalogs (50 f., 120).

1. Die Überzeugung, dass p , zu der S durch Selbsttäuschung gelangt, ist falsch.
2. S behandelt die für die Bestimmung des Wahrheitswerts von p relevanten oder zumindest dem Anschein nach relevanten Daten auf motivational verzernte Weise.
3. Die verzernte Behandlung ist eine nichtabweichende Ursache dafür, dass S zur Überzeugung, dass p , gelangt.
4. Das Korpus aus Daten, die S zur Zeit besitzt, stellt stärkere Belege für $\neg p$ als für p zur Verfügung.

Die Erfüllung dieses Ensembles von Bedingungen hält Mele für hinreichend für Selbsttäuschung, wobei er (1.) bis (3.) als notwendige Bedingung erachtet, (4.) dagegen nicht (51).¹¹ Letzteres ist nach Mele dem Umstand geschuldet, dass jemand Evidenz für etwas haben kann, ohne aktuell über sie zu verfügen. In diesem Fall handelt es sich um potentielle Evidenz, die S leicht aktualisieren könnte. Doch tut S dies nicht und trägt stattdessen aktuelle Evidenz nur für die gegenteilige, aber falsche Überzeugung zusammen.

Allerdings könnte man sich Fälle wie den von Paul denken, die Bedingung 4 als nicht notwendig ausweisen, weil S weder für $\neg p$ noch für p überwiegende aktuelle oder potentielle Evidenz besitzt, so dass seine verzernte Überzeugung davon abhängt, wie groß seine Aversion dagegen ist, $\neg p$ oder p zu Unrecht für falsch zu halten. Verhält sich das aber so, dann könnte dies weiterführend ebenfalls die Bedingung 1 tangieren, nach der eine Person in ihrem Glauben, dass p , *per definitionem* nur dann (selbst-)getäuscht ist, wenn p falsch ist (51). Das Beispiel scheint geeignet, ihren Status als notwendige Bedingung für Selbsttäuschung in Zweifel zu ziehen.¹² Denn es wäre angesichts derselben motivational induzierten Datenverzerrung merkwürdig, nur dann von Selbsttäuschung zu reden, wenn sich niemand auf die Suche nach Paul gemacht hat, während der glückliche Umstand des Aufbruchs einer Rettungsmannschaft Paul vom Vorwurf der Selbsttäuschung freisprechen würde, ganz gleich, wie aberwitzig seine wunsch- und aversionsgelenkten Raisonnements gewesen sein mögen und welche Rechtfertigungsgründe auch immer ihn glauben ließen, man sei zur Suche aufgebrochen.

¹¹ Allerdings betont er: „I have not offered a statement of individually *necessary* and jointly sufficient conditions“ (119). Vgl. auch Mele, „Emotion and Desire in Self-Deception“, in: Anthony Hatzimoyis (Hrsg.), *Philosophy and the Emotions*, Cambridge 2003, 163-179, hier: 163 Anm. 2.

¹² Es unterscheidet sich von Robert Audis Fall einer abweichenden kausalen Verkettung, den Mele (122 f.) diskutiert.

Nun scheint es denkbar, dass Mele dieses Beispiel erst gar nicht als Fall von Selbsttäuschung akzeptieren würde,¹³ da dabei nicht einmal eine potentielle Evidenz Pauls im Spiel ist, gegenüber der sich eine verzerrte Überzeugung als Selbsttäuschung abheben könnte. Doch könnte sich auch unabhängig davon Kondition 1 als nicht notwendig erweisen, die besagt, dass die verzerrte Überzeugung, dass p , und die Falschheit von p zwingend zusammen vorliegen. Zwar würde diese Bedingung immerhin verständlich machen, warum, im Lichte von Meles Theorie betrachtet, die Unterscheidung von Selbsttäuschung und Irrtum Schwierigkeiten bereitet. Denn ist auch womöglich nicht jeder Fall von Irrtum durch Selbsttäuschung verursacht, so liegt nach Bedingung 1 doch in jedem Fall von Selbsttäuschung ein Irrtum vor. (Es sei denn, man reservierte die Bezeichnung „Irrtum“ für den Fall, in dem man ohne Verzerrung der Daten eine falsche Proposition für wahr oder eine wahre Proposition für falsch hält.) Nur fragt sich, ob bei Wahrheit von p und der Überzeugung, dass p , in der Tat keine Selbsttäuschung vorliegen kann oder ob nicht im Gegenteil, so gut es auf der einen Seite Falschheit ohne Befangenheit und Selbsttäuschung gibt, auf der anderen Seite ein Einklang mit propositionaler Wahrheit auch durch Verzerrung und Selbsttäuschung zustande kommen kann – so wie es eben möglich ist, durch mehrmaliges Verrechnen zufällig das richtige Ergebnis einer arithmetischen Aufgabe zu treffen.

Die Annahme, dass ein motivational befangener bzw. verzerrter Umgang mit Daten auch in dem Fall nichtabweichende Ursache für die Überzeugung, dass p , sein kann, in dem p wahr ist, würde die Erwägung nahe legen, bei Selbsttäuschungen hätten wir es mit einem verzerrenden Umgang mit epistemischen Größen – wie z.B. Rechtfertigungsprozeduren – zu tun, vor denen die Bedeutung nichtepistemischer Größen – wie z.B. die Wahrheitswerte bestimmter Propositionen – zurücktreten. Entscheidend für Selbsttäuschung wäre dann nicht, ob ein bestimmter Wahrheitswert getroffen oder verfehlt wird, sondern mit welchen Gründen und mittels welcher Prozeduren dies geschieht.¹⁴ Im Sinne einer Akt/Objekt- bzw. Akt/Resultat-Unterscheidung müsste demnach bei einer Untersuchung des Selbsttäuschungsphänomens das Hauptaugenmerk auf der Selbsttäu-

¹³ Eine Grundschwierigkeit der Debatte über Selbsttäuschung besteht offenbar darin, dass sie vornehmlich anhand von Beispielen geführt wird, deren Mustergültigkeit oftmals nur schwer eingesehen werden kann. Zudem werden in Bezug auf Einzelfälle und Exempla mitunter so subtile Unterscheidungen wie die zwischen Wunschenken (*wishful thinking*) und Wunschglauben (*wishful believing*) eingeführt, deren Unterscheidungsmodus dann selbst noch einmal in einen stipulativen, einerseits, und einen substanziellen, andererseits, ausdifferenziert wird (73 f.), ohne dass von diesem filigranen Instrument im weiteren Verlauf der Debatte noch irgendwelcher analytischer Gebrauch gemacht würde. Stattdessen handelt sich die Diskussion weiter von Beispiel zu Beispiel.

¹⁴ In Pauls Fall liegen solche Gründe wahrscheinlich gar nicht vor, weil ihm keine akzeptable Prozedur für ihre Gewinnung zur Verfügung steht, so dass – rein epistemisch betrachtet – Urteilsenthaltung am Platz gewesen wäre. Er besitzt aber am Ende die falschen Gründe, weil ein inkorrekt Mechanismus ihn damit versorgt hat.

schung als Akt, nicht auf ihr als Resultat liegen, bzw. wenn auf ihr als Resultat, dann als Resultat dieses Akts und in Abhängigkeit von diesem Akt.¹⁵ Das meint: Um auf Selbsttäuschung (Resultat) erkennen zu dürfen, müssen wir prüfen, wie sie zustande gekommen ist.

V. Kein Bedarf an nicht-deflationären Erklärungen?

Die oben verfolgte Strategie gleicht derjenigen, von der Mele selber reichlich Gebrauch macht. Eine Theorie scheint angreifbar, wenn sie nicht-notwendige Bedingungen aufnimmt und wenn das Ensemble ihrer Kriterien nicht hinreicht, um Standardfälle von Selbsttäuschung zuverlässig zu identifizieren. Nebenher verspricht sich Mele von der Schwächung seiner Gegenpositionen eine Stärkung der eigenen (76). In dieser Absicht untersucht er in Kapitel 4, ob (i) verzerrter bzw. befangener Umgang mit Daten bewusst und absichtlich zu erfolgen hat und ob (ii) eine Überzeugung p und eine Überzeugung $\neg p$ zugleich vorliegen müssen, damit wir von Selbsttäuschung sprechen dürfen (*dual-belief requirement* (76)).

(i) scheidet am FTL-Modell, das die in Rede stehenden Fälle auch nichtagentiv und somit deflationär zu erklären vermag (90 f.). (ii) kommt dadurch zu Fall, dass die Gleichzeitigkeit solcher Überzeugungen zwar nicht prinzipiell ausgeschlossen wird, ihr Auftreten jedoch etwas gänzlich anderes bezeugen würde als ein Vorkommnis des Standardtyps von Selbsttäuschung (77). Für die Erläuterung der vermeintlichen oder tatsächlichen Fälle von Selbsttäuschung ist es nach Mele dagegen nicht erforderlich, auf die „dual-belief condition“ ($Bp \ \& \ B\neg p$ (76)) zurückzugreifen. Es besteht kein Erklärungsbedarf, den nur nicht-deflationäre Annahmen decken könnten („[T]here is no explanatory need to postulate such beliefs“ (92, vgl. 77)). Denn alle diskutierten Fälle lassen sich im Sinne einer deflationären Explikation uminterpretieren (80 f., 82 f., 85 f.). Dazu werden zum einen die Gehalte p und $\neg p$ unterschiedlich kontextualisiert, so dass sie nicht länger als miteinander unverträglich erscheinen. Zum anderen sieht sich Mele berechtigt, die Vorkommnisse der Einstellung B in $Bp \ \& \ B\neg p$ mit unterschiedlicher epistemischer Kraft zu versehen, wodurch Überzeugung *in sensu stricto* (durch *Autopsie*, Beweis oder starker Bestätigung) ggf. gegen Glauben (vom Hörensagen, einigermaßen belegt oder blind), Verdacht (*suspicion* (79)) oder Vermutung stehen kann, ohne Unvereinbarkeiten zu erzeugen.

¹⁵ Um hier „Akt“ nicht gegen Mele im Sinne von „Agentivität“ zu verstehen, könnten wir auch etwas umständlicher von einer *Mechanismus in Aktion/Resultat*-Ambiguität des Ausdrucks „Selbsttäuschung“ sprechen.

Ist dieses intervenierende Vorgehen statthaft, so könnte sich eine nicht-deflationäre Bedingung für Selbsttäuschung nur dann als notwendig erweisen, wenn die einschlägigen Exempla allein unter einer Beschreibung aufzutreten vermöchten, die diese Bedingung interpretationsresistent instantiiert. Ob solche Fälle vorkommen, scheint fragwürdig, da es womöglich überhaupt keine Zustände, Prozesse oder Handlungen gibt, die nur eine einzige akzeptable Beschreibung zulassen. Dies könnte auf Meles eigenen Bedingungskatalog zurückschlagen. Doch scheint er – wahrscheinlich zu Unrecht – davon auszugehen, dass zur deflationären Lesart hin eine Art natürliches Gefälle besteht, weshalb diese, da bereits durch Erklärungsökonomie nimbiert, lediglich auf Plausibilität („Does anything block the supposition that ...?“ (85)), jede *lectio difficilior* aber auf den Nachweis ihrer Unabdingbarkeit („explanatory need“) verpflichtet wäre.

VI. Verdrehte Fälle

Neben Fällen geradlinger (*straight*) Selbsttäuschung, bei denen die verzerrte Überzeugung, dass p , mit dem Wunsch, p möge wahr sein, bzw. der Befürchtung, p könne falsch sein, harmoniert, diskutiert Mele in Kapitel 5 verdrehte Fälle (*twisted cases*), bei denen jemand aufgrund fadenscheiniger Anhaltspunkte etwas für wahr hält, dessen Falschheit (nicht aber Wahrheit) er zugleich wünscht. Zur Bestimmung dieses Typs von Selbsttäuschung gebraucht Mele abwechselnd zwei voneinander abweichende Formulierungen:

(A) „In twisted cases, people are self-deceived in believing something that they want to be false“ (4, 94): $Bp \ \& \ D\neg p$, und

(B) „By definition, in *twisted* instances, people who are self-deceived in believing that p do not desire that p “ (95, 4): $Bp \ \& \ \neg Dp$.

Während sowohl (A) als auch (B) zeigen würden, dass für die Selbsttäuschung, dass p , ein Wunsch, dass p , nicht zwingend erforderlich ist – und tatsächlich nimmt Mele Wünsche in seinen Bedingungskatalog nicht auf (50 f., 120), sondern behandelt sie als kontingente Verzerrungsverstärker –, so ist (B) ohne eine Ergänzung streng genommen nicht hinreichend, um Fälle dieser Art als verdrehte auszuweisen. Denn „not wanting it to be the case that [p]“ (4) wäre – als voluntative Indifferenz ($\neg Dp$) gegenüber p gedeutet – nichts, was zur Überzeugung, dass p , in einer verdrehten Beziehung stehen könnte. Dazu wäre vielmehr die stärkere Variante (A), „wanting it to be false“ (94) bzw. „wanting it to be the case that she is innocent of the charge“ (95) ($D\neg p$), nötig.

Nun könnte einerseits Meles unschuldig alternierender Umgang mit beiden Formulierungen dafür sprechen, dass er sie umgangssprachlich gleichbedeutend im Sinne von (A) verwendet wissen möchte. Andererseits dürfte (B) um eine emotionale Erläuterung

so ergänzt werden können, dass Emotionen die systematische Stelle von Wünschen einnehmen, und insofern mit Meles Hypothese E^{16} kompatibel sein, von der sich der Autor überdies zumindest vorübergehend eine bessere Explikation verdrehter Fälle erwartet. – Zwei Schwierigkeiten stehen im Raum: Erstens müssen die in Rede stehenden Fälle zweifelsfrei als verdrehte ausgewiesen werden, und zweitens sollte eine angemessene Erläuterung für sie gefunden werden können.

Mele betrachtet zunächst Pears' Beispiel eines unsicheren, eifersüchtigen Ehemanns, der seine Frau für untreu hält (Bp), während er zugleich wünscht, dies sei unzutreffend ($D\rightarrow p$).¹⁷ Nun ist es aber, wie das Beispiel es will, erst der Wunsch des Eifersüchtigen, die Treue seiner Frau durch Beseitigung sämtlicher Rivalen sicherzustellen (Dx), dasjenige, was ihn motiviert, ihr aufgrund vager Anzeichen Untreue zu unterstellen. Denn diese verzerrte Überzeugung ist seinem Beseitigungsvorhaben förderlich und würde es ggf. rechtfertigen (95 f.).¹⁸

Doch weckt eine nähere Betrachtung den Verdacht, hier sei allein die Beschreibung verdreht. (i) Zum einen scheint es, das Beispiel verwende „wünschen/desire“ in Abweichung von Fällen geradliniger Selbsttäuschung im Sinne des *belief/desire*-Modells für Handlungen. So kann der Wunsch, die Rivalen aus dem Weg zu schaffen, als deliberative Präzisierung des Wunschs nach Treue durch ein hinreichendes, wenngleich drastisches Mittel zu seiner Verwirklichung betrachtet werden, wogegen der Wunsch, etwas Unverfügbares möge eingetroffen oder ausgeblieben sein, keiner Deliberation offen steht. Was aber deliberativ präzisierbar ist, muss selbst von der Art einer Handlungsabsicht sein. (ii) Ist dieser Befund stichhaltig,¹⁹ so dürfte der Fall zum anderen kaum mehr als verdreht und vielleicht nicht einmal mehr als Selbsttäuschung angesprochen werden

¹⁶ „Perhaps emotions play roles in twisted self-deception that parallel some of the roles I attributed to desire in straight self-deception. I dub the idea that they do this and that twisted self-deception has no desires as significant causes ‚hypothesis E.‘“ (99).

¹⁷ Vgl. David Pears, *Motivated Irrationality*, Oxford 1984, 42-44.

¹⁸ Ein solches Beispiel mutet vielleicht nicht ganz aus der Luft gegriffen an, wenn man sich parallel dazu Fälle jenes paranoid erscheinenden gegenwärtigen Bellizismus vor Augen führt, der seinem latenten *anxious desire* nach einer bedrohungsfreien Weltlage ($D\rightarrow p$) präventiv durch Ausschaltung aller Feinde nachkommen möchte (Dx), wozu ihm jede noch so haltlose Überzeugung von einer akuten Bedrohung (Bp) als Motivations- und Rechtfertigungsgrund willkommen ist. Dies schließt selbstverständlich nicht aus, dass verzerrte Überzeugungen dieser Art zudem in kausalem Zusammenhang mit Fremdtäuschungsversuchen auftreten. Vgl. infra Abschnitt VII.

¹⁹ Allerdings ist nicht klar, wie bewusst Absichten gefasst und entsprechende Überlegungen angestellt werden und welche Handlungen der eifersüchtige Ehemann ausführt, um Rivalen auszuschalten. Deflationisten haben ein Interesse, den Anteil bewussten Planens zu marginalisieren, zumindest zu minimalisieren (cf. 95). Pears, 1984 (Anm. 17), 44, teilt dieses Interesse in diesem Fall und garantiert unbewusstes Tun durch Einführung eines bewusstseinslosen Agenten: „There is, presumably, a wish for the ulterior goal, safety or the elimination of a rival, but nature takes over at this point and sets up an emotional programme that ensures its achievement. The plan is nature's and not the person's, and that is why the formation of the intrinsically belief is not felt to be the object of the wish.“

können. Andernfalls setzte etwa jedes Durchsetzen einer Norm eine Selbsttäuschung des verdrehten Typs voraus, was ungereimt anmutet. Wer wünscht, dass bestimmte Handlungen unterlassen werden ($D\neg p$), und zudem eine Sanktionierung dieser Norm wünscht (Dx), glaubt, dass es Normverletzungen – wie unter Menschen üblich – geben kann (Bp), auch wenn ihm dafür keine konkreten Anhaltspunkte vorliegen. Sonst nämlich wären Norm und Sanktionierung unnötig. Betonen wir daher den wunschbestimmten Handlungsaspekt des Eifersuchtsexempels, so schwinden die Anhaltspunkte für Selbsttäuschung. Nehmen wir dagegen an, die Unsicherheit des Ehemanns veranlasse dessen Untreueverdacht, dann erscheint der Fall kaum mehr verdreht, sondern geradlinig. Daher könnten wir in Meles Stil fragen, ob hier stets die explanatorische Notwendigkeit besteht, von einer verdrehten Selbsttäuschung auszugehen, und dürften mit einem knappen „No“ antworten.

Schwierigkeiten der skizzierten Art und eine gewisse Unübersichtlichkeit der Explikation nehmen eher zu, wenn Mele das Beispiel im Lichte der FTL-Strategie interpretiert. Und dies – unabhängig davon, ob Eifersucht als Trennungsangst (97, 114) phänomenologisch korrekt erfasst wird – zumal dann, wenn Mele die motivationale Richtung der Erklärung im Verlauf seiner Erörterung offenbar umkehrt. War bei Pears die verzerrte Überzeugung von einem Wunsch, Nebenbuhler zu beseitigen, motiviert, der wiederum als Mittel zur Realisierung eines Treuwunschs verstanden werden konnte, so behauptet Mele nun (97), das unterschiedliche Gewicht, das der Vermeidung irrtümlicher Untreuebezeichnungen oder irrtümlicher Treueannahmen je nach psychischer Disposition beigemessen wird und entsprechende Propositionen bereitwilliger affirmieren lässt (vgl. 35 ff., 41 f.), mindere oder verstärke die Bereitschaft, gegen mögliche Störenfriede oder Rivalen vorzugehen.

Doch ist dies nicht Meles letztes und vielleicht nicht einmal ein entscheidendes Wort, wenn er für verdrehte Fälle sodann einen Ansatz erprobt, in dem Wünsche – wenn überhaupt – eine nebengeordnete Rolle spielen. So erweisen sich seine Ausführungen dort als fruchtbarer, wo er den verdrehten Fall unter Verzicht auf auxiliäre Handlungsabsichten (Dx) so beschreibt, dass er als Konflikt zwischen (Untreuevermutungen veranlassender) Eifersucht und (Treueerwartungen erleichterndem) Treuwunsch erscheint (100). An dieser Stelle fragt er nun, inwieweit hier Emotionen für Wünsche (vielleicht die Angst, dass p , für $D\neg p$) zu substituieren vermögen und ob sie gemäß „hypothesis E “ in verdrehten Fällen die Rolle spielen können, die bei geradliniger Selbsttäuschung den Wünschen zukomme (101 f.). Zweierlei gilt dabei als garantiert. Erstens besteht empirisch gestützte Gewissheit darüber, dass (i) Emotionen wie Wünsche verfügbare Daten unterschiedlich lebendig erscheinen lassen, (ii) die Aufmerksamkeit selektiv fokussieren und (iii) Bestätigungsverzerrungen verstärken, wenn nicht sogar auslösen (98 f., vgl.

supra S. 4 f.). Zweitens geht Mele davon aus, dass das Phänomen der Selbsttäuschung durch die Beschreibung der Verzerrungsmechanismen und -prozesse bereits korrekt erfasst ist, so dass keine Theorierevision, sondern lediglich die Prüfung ansteht, ob Emotionen Auslöser und Unterstützer solcher Prozesse sind. Daher kann auch dem Vorwurf, die Beteiligung von Emotionen mache Selbsttäuschung zu einem Prozess, der sich unserer Kontrolle entziehe, mit dem Nachweis begegnet werden, dass Emotionen bzw. deren Intensität von uns steuerbar und die Auslöser verzerrter Überzeugung somit zumindest partiell kontrollierbar sind. Wir sind ihnen nicht machtlos ausgeliefert. Selbsttäuschung kann im Prinzip entgegengewirkt werden.

Ist Hypothese *E* unter diesen Auspizien – selbst wenn ein *Modus operandi* Desiderat bleibt – nicht aus dem Rennen, so wird sie nun von einem rein kognitiven Ansatz herausgefordert, der bei der Erklärung verdrehter Selbsttäuschung sowohl auf Emotionen als auch auf Wünsche zu verzichten können meint und den erwähnten Kontrollaspekt zu einem entscheidenden Faktor macht.²⁰ Denn Selbsttäuschung werde durch Persönlichkeitsprofile wie Schuldkomplexe oder chronische Selbstunterschätzung ausgelöst (104 f.). Zu Selbsttäuschungen im eigentlichen Sinn komme es jedoch erst dann, wenn Personen, die sich einer Neigung zu verzerrter Wahrnehmung bewusst seien, nicht gegen ihre perzeptiven und kognitiven Verzerrungen vorgehen. Selbsttäuschung würde demnach auf Unterlassung hinauslaufen – auf die selbstverschuldete Unterlassung, kognitiven Verzerrungen bzw. Befangenheit erster Stufe auf einer zweiten Stufe korrigierend zu begegnen –, und damit in etwa dem Fall eines Arithmetikschülers entsprechen, der es versäumt, seine Lösungen nachzuprüfen, obwohl er weiß, dass er mitunter Fehler macht (105 f., 109).

Erst an dieser Stelle führt Mele für Selbsttäuschung einen konditionalen Unvoreingenommen-Beobachter-Test (*impartial-observer test*) ein, den zu passieren *conditio sine qua non* ist und der hilft, die Beispiele für rein kognitive Befangenheit als Fälle von Selbsttäuschung zu diskreditieren.²¹ Wenn *S* darin selbstgetäuscht ist, dass *p*, und über einen potentiellen Satz relevanter Daten verfügt, den unvoreingenommene, kognitiv gleichartige Personen (*cognitive peers*) aktualisieren könnten, so würde die Zahl derer, die daraus die Falschheit von *p* folgern würden, die der *p* Folgernden signifikant übersteigen (106, 121). Sind nun aber die Unvoreingenommenen, die weder Wünsche in Bezug auf *p* oder $\neg p$ hegen noch Vermeidungspräferenzen im Sinne einer FTL-Strategie aufweisen, dem vermeintlich Selbstgetäuschten kognitiv äquivalent (indem sie z.B.

²⁰ Vgl. Martha Knight, 1988 (Anm. 3), bes. 186.

²¹ Die Position eines unparteiischen bzw. unvoreingenommenen Zuschauers (*impartial spectator*) ist bereits bei Adam Smith, *The Theory of Moral Sentiments*, ed. Knud Haakonssen, Cambridge 2002, 182-184 (III 4, §§ 1-6) u.ö., Bestandteil einer Erörterung von Selbsttäuschung (*self-deceit*).

ihre Fähigkeiten chronisch unterbewerten) und steht für dessen verzerrte Überzeugung kein weiterer Auslöser als eben diese kognitive Ausstattung zur Verfügung, so müssten die unvoreingenommenen Beobachter zum selben Resultat kommen wie der Beobachtete. Die notwendige Bedingung wäre damit verfehlt und jedem begründeten Erkennen auf Selbsttäuschung auf rein perzeptiver oder kognitiver Basis der Boden entzogen (107-109). Eine gewisse Voreingenommenheit ist demnach zwingend Bestandteil jeder Selbsttäuschung (109 f.). Sie ist mehr als nur selbstverschuldeter Irrtum.

Dieser Schachzug scheint schlagend. Doch könnte er sich als Pyrrhussieg erweisen. Abgesehen von der Frage, was eine Mehrheit zur qualifizierten macht (121), verlangt der *impartial-observer test* nicht allein ein numerisches Übergewicht der gegenteiligen Überzeugung, sondern zudem die Unvoreingenommenheit ihrer Vertreter. Angesichts der von Mele angenommenen Häufigkeit von Selbsttäuschung (124) und der Ubiquität verzerrender oder nicht verzerrender Motivation,²² die das FTL-Modell in Anschlag bringt (110), wird fraglich, ob eben die Unvoreingenommenheit einer Kontrollgruppe garantiert sein kann, die es wiederum erst erlauben würde, von Selbsttäuschung zu reden. Je kommuner Selbsttäuschung, umso rarer die Unvoreingenommenen, deren abweichendes Urteil in entsprechender Häufigkeit Selbsttäuschung als communes Phänomen ausweisen könnte. Hier scheint sich ein Einfallstor für skeptische Einlassungen aufzutun. Doch ließe sich einwenden, die notwendige Bedingung verlange allein eine hypothetische Kontrollgruppe kognitiv gleichartiger, motivationsneutraler und insoweit unvoreingenommener Personen, die hypothetisch ein nicht-befangenes Verhalten an den Tag legt, und der Test könne jeweils auf eine einzige relevante Fragestellung eingeschränkt werden. Insofern würde der Test kein skeptisches Definitionsproblem, sondern nur ein skeptisches Anwendungsproblem aufwerfen.

Entscheidend ist, dass emotionale und desiderative Motivation durch den *impartial-observer test* als Auslöser und Verstärker verdrehter Selbsttäuschung zurück ins Spiel gebracht werden. Mit ihr jedoch auch jene auxiliären Wünsche, die, wie gesehen, Schwierigkeiten bereiten, sofern sie als Handlungsabsichten gedeutet werden können. So mag es unproblematisch erscheinen, dass Dolores' verzerrte Überzeugung, den Tod ihrer Tochter verschuldet zu haben, von der alles überwiegenden Aversion gegen die Annahme, das Schicksal habe blind zu geschlagen, motiviert ist. Dagegen trägt Susans Selbstunterschätzung womöglich wieder Züge eines Mittels, wenn sie damit ihre Fallhöhe in peinlichen Situationen senken möchte (110, 113).

Sodann gelingt es auch im zweiten Anlauf nicht, die Hypothese *E* zu stützen, nach der Emotionen Überzeugungen auch im Alleingang verzerren können. Zwar scheint es nach einem bestimmten Verständnis von Eifersucht plausibel, dass sie u.a. vom Wunsch nach

²² Zu deren Unterscheidung müsste der Test ebenfalls in Anspruch genommen werden.

einer engen Beziehung verursacht sein kann, woraufhin sie Lebendigkeits-, selektive Aufmerksamkeits- und Bestätigungsverzerrungen direkt auslöst (*direct emotion hypothesis*). Doch wird es wahrscheinlich selbst der subtilsten Kasuistik misslingen, einen Fall emotionaler Alleinverursachung zu etablieren, so dass Hypothese *E* weiterhin in der Luft hängt.²³

VII. Konklusion

Self-Deception Unmasked ist das Buch eines Tüftlers für Tüftler. Statt einen generalisierenden Überblick über eine abgeschlossene Diskussion zu geben, verstrickt es den Leser in die philosophische Debatte, zwingt ihn zur argumentativen Mitarbeit und provoziert es seine Einwände. Zu erheblichen Teilen aus früheren Aufsätzen kompiliert, gibt es Einblick in das Atelier eines Philosophen, der weniger fertige Denksteine zeigt als sehen lässt, wie er sie erstellt. Ist dies auch vielleicht nicht authentisch (und warum sollte man das verlangen?), so ist es doch gut inszeniert.

Meles Arbeit zeigt, wie bereits Julie Kirsch unterstrichen hat, dass wir gerade über diejenigen Überzeugungen, die wir am stärksten gewichten, eher wenig Kontrolle haben.²⁴ Während wir unser Nachdenken für vernünftig, reiflich überlegt und wohlberaten halten, trägt es wie von selbst, doch bedingt selbstverschuldet, allenthalben auch Züge von Wunschdenken, ja von Wunderglauben und Selbstbetrug. Doch könnte der deflationäre Ansatz über der Leichtigkeit, mit der es nach ihm zu Selbsttäuschung in inflationärem Ausmaß kommen kann, wie bereits erwähnt, selbst in Schiefelage geraten.

Nun mag einerseits verwundern, dass es Mele, der das Feld immerhin seit 1982 einigermaßen monoman beackert,²⁵ bis dato nicht gelungen ist, eine befriedigende Begriffsanalyse vorzulegen. Dies mag u.a. daher rühren, dass er weniger mit der strukturellen Analyse des Zustands der Selbsttäuschung als mit dessen Genese oder Verursachung beschäftigt ist. Doch steht auch die Bestimmung einer kanonischen Form für Selbsttäuschung (Prozess) aus, die „a wide variety of ‚normal‘ routes to self-deception“ (123) von abwegigen (*deviant*) unterscheiden und so einen entsprechenden Begriff finden helfen könnte. Ist der Einwand berechtigt, dass die Falschheit von *p* dafür nicht notwendig ist (vgl. supra S. 7 f.##), dann wird auch seine Protoanalyse eines „focal concept“ – „*S* enters self-deception in acquiring a belief that *p* if and only if *p* is false and *S*

²³ Dies ist auch noch Stand der Dinge in Mele, 2003 (Anm. 11), bes. 172 f. u. 176 f.

²⁴ Vgl. Julie E. Kirsch, „Mele, Alfred R. *Self-Deception Unmasked*“, *Ethics* 112 (2002), 628-631, hier: 631.

²⁵ Vgl. „Self-Deception, Action, and Will: Comments“, *Erkenntnis* 18 (1982), 159-164.

acquires that belief that p in „a suitably biased way“ (123)²⁶ – die Sache nicht richtig treffen. Zudem laufen Meles Bemühungen um ein Brennpunkt-konzept Gefahr, in eine *petitio principii* zu münden, d.h. stipulativ nur solches als Fall von Selbsttäuschung zuzulassen, was von eben diesem Konzept erfasst wird.

Doch könnte sich andererseits herausstellen, dass das, was unter dem Ausdruck „Selbsttäuschung“ firmiert, selbst wenn sein Gebrauch, wie Mele anstrebt, sinnvoll eingeschränkt wird, lediglich eine Familienähnlichkeit aufweist.²⁷ Dann bleibt der Ansatz mit systematischer Notwendigkeit, was er ist: eine Basteltheorie, die auf Einwände und neue Beispiele reagiert, indem sie sie zurückweist oder integriert. So könnte vielleicht auch die nachfolgende Erwägung eine Revision des Ansatzes nach sich ziehen.

Wenn Meles Argumentation triftig ist, müssen sich Fälle von Fremdtäuschung von solchen der Selbsttäuschung strikt unterscheiden lassen. Sie sind Gegenstände verschiedener Beschreibungsweisen, die sich eines strikt geschiedenen Vokabulars bedienen (vgl. jedoch Fn. 5). Doch könnte dagegen sprechen, dass praktisch interessante Fälle eine scharfe Unterscheidung von Phänomen, Beschreibung und Vokabular womöglich nicht zulassen. Solche Fälle von Selbsttäuschung würden nicht nur eine einzelne Person involvieren. Entsprechend müsste auch weniger die Frage geklärt werden, wie ein einzelnes Subjekt dazu kommt, sich selbst zu täuschen, sondern wie Selbsttäuschungsmechanismen und -strategien sozial wirksam werden können und wie sie sozial gestützt werden.

Wird z.B. die Überzeugung, die deutsche Wehrmacht habe während des zweiten Weltkriegs niemals Kriegsverbrechen verübt, von ehemaligen Angehörigen nicht bloß als Schutzbehauptung vorgebracht, sondern, wie angenommen werden darf, von vielen Veteranen wider mögliches besseres Wissen geglaubt, so lässt sich das vielleicht als Fall kollektiver Selbsttäuschung verständlich machen. Hier scheint es fragwürdig anzunehmen, Selbsttäuscher machten sich (nichtagentiv) jeweils so etwas vor, dass an ihnen *für sich isoliert*, doch irgendwie auch noch parallel zu anderen für sich isolierten Selbsttäuschern, FTL-Strategien der Leidvermeidung wirksam würden, weil Kriegsverbrechen der eigenen Streitkräfte zu Unrecht bestreiten für Kombattanten psychisch weniger kostspielig ist als sie zu Unrecht anerkennen. Stattdessen dürften hier Selbsttäuschungen im Stil Meles gemeinsam mit Fremdtäuschungen eine Gemengelage erzeugen, die kollektive Selbsttäuschung ermöglicht und durch wechselseitige Bestätigung und Sanktionierung sozial aufrecht hält. Selbsttäuschung wird ausgelöst und produziert, weil man zu einer Gruppe gehören will. Diese übt sozialen Druck auf ihre Mitglieder aus, etwas von ihr als Gruppe zu glauben, was die Mitglieder der Gruppe von

²⁶ Vgl. auch Mele, 2003 (Anm. 11), 163.

²⁷ Siehe auch Davidson, 1997 (Anm. 1), 221: „I am pretty sure that no one scheme will fit all examples.“ Vgl. ebd., 230.

individueller Verantwortung und Schuld entlastet. Um die Gruppe darin zu stabilisieren, werden aus der Gruppe heraus historische Fälschungen vorgenommen (Fremdtäuschung), die von den Gruppenmitgliedern bereitwillig geglaubt werden (Selbsttäuschung), am Ende sogar von den Fremdtäuschern.²⁸

Nicht Fehlervermeidung (PEDMIN-Strategie) ist hier wichtiger als Wahrheitssuche, sondern ein Wunsch nach Gruppenzugehörigkeit, der wahrscheinlich nicht auf Aversionen gegen Sanktionen (i.S. einer FTL-Strategie) reduzierbar ist. Erfolgreiche Fremdtäuschung wäre somit auf Selbsttäuschung ebenso angewiesen wie sie ihrerseits Selbsttäuschung stützt. In Fällen dieser Art bezeichnen die Ausdrücke „Selbsttäuschung“ und „Fremdtäuschung“ nicht kategorial verschiedene Vorgänge, als vielmehr unterschiedliche Aspekte eines kollektiven Prozesses.

In ganz anders gelagert Fällen könnte man indes auch fragen, ob Selbsttäuschung zwingend durchweg als Fall von Irrationalität angesehen werden muss und ob es nicht, eben so wie sie zum Schlechten ausfallen kann, auch Selbsttäuschung zum lebenspraktisch Guten gibt, die nicht als Wunschdenken oder Lebenslüge angesprochen werden muss.²⁹ Doch wie immer wir diese Fragen bescheiden wollen, haben wir es hier mit einem bedeutsamen Phänomen zu tun, ohne dessen Erklärung wir – wie Davidson in Bezug auf interne Inkonsistenzen im allgemeinen meinte³⁰ – auch unsere selbstkritischen Anstrengungen und den Wandel unserer Meinungen und Einstellungen nicht verständlich machen können. Dies dürfte Revisionen theoretischer Überzeugungen (*belief revision*) wie die praktischer Ansichten (moralisches Umdenken) gleichermaßen betreffen. Meles Arbeit leistet hierzu einen enorm wichtigen Betrag.³¹

²⁸ Gleiches dürfte für esoterische und hermetische Zirkel gelten, wie Umberto Eco sie in den Romanen *Il pendolo di Foucault* (Milano 1988) und *Baudolino* (Milano 2000) und in der theoretischen Arbeit *I limiti dell'interpretazione* (Milano 1990, 41-99) beschreibt, aber auch für Tätergruppen, die sich in Opfer, Opportunisten, die sich in Widerstandskämpfer umdeuten.

²⁹ So könnte das Wort „Wenn du aber Almosen gibst, soll deine linke Hand nicht wissen, was deine rechte tut“ (*Matthäusevangelium* 6,3) geradezu als Aufforderung zu lebensdienlicher Selbsttäuschung gelesen werden. Dass Selbsttäuschung nicht jederzeit irrational sein muss, zeigt überdies Reschers Diskussion eines fundamentalen Problems rationaler Entscheidung (*Buridans Esel*). Hier erscheint eine ungerechtfertigte und verzerrte *ad hoc* vorgenommene Ausbildung von Präferenzen vernünftiger als ein Verharren im Patt wohlabgewogener Wünsche und Überzeugungen. Vgl. Nicholas Rescher, „Choice Without Preference: A Study of the Logic and of the History of the Problem of ‚Buridan’s Ass‘“, *Kant-Studien* 51 (1959/60), 142-175, bes. 167-170. Siehe auch Meles „*all-things-considered* judgment“ in ders., *Irrationality. An Essay on Akrasia, Self-deception, and Self-Control*, New York 1987, 110 f. – Ob im poetischen Spiel der Liebe (vgl. Mele (26)) statt restloser Selbstaufklärung nicht eine kalkuliert eingegangene Selbsttäuschung (als deren Salz) gefragt sein kann, wäre gleichfalls zu diskutieren.

³⁰ Vgl. Davidson, 1982 (Anm. 2), 186 f.

³¹ Für Kritik und Hinweise zu früheren Fassungen danke ich Andreas Graeser, Wolfgang Gessner und den TeilnehmerInnen ihres Kolloquiums zu Meles Buch.