Selbsttäuschung

Anna Wehofsits

Druckfassung in: Vera Hoffmann-Kolss, Nicole Rathgeb (Hg.): Handbuch Philosophie des Geistes (2024), J. B. Metzler
https://link.springer.com/chapter/10.1007/978-3-476-05416-6_40

1. Einleitung

Selbsttäuschung scheint ein alltägliches Phänomen zu sein. Wir nehmen an Anderen wahr, wie sie mehr oder weniger bewusst einer Einsicht ausweichen, die sie nicht wahrhaben wollen, und die meisten von uns können sich an Situationen erinnern, in denen sie sich selbst etwas vorgemacht haben (wobei es charakteristisch für die eigene Selbsttäuschung ist, dass sie sich einer direkten Beobachtung im Vollzug entzieht). Wir können also Beispiele für Selbsttäuschung benennen, unser begriffliches Verständnis von Selbsttäuschung aber ist diffus, und der Versuch, das Phänomen aus philosophischer Perspektive begrifflich genau zu fassen, führt leicht zu starken Spannungen, für die bis heute keine allgemein akzeptierte Lösung gefunden wurde. Selbsttäuschung, so scheint es, ist weitverbreitet, philosophisch und psychologisch aber rätselhaft: Ist es überhaupt möglich, sich selbst zu täuschen? Wenn ja, wie und warum täuschen wir uns selbst? Inwieweit tragen wir aktiv oder gar strategisch zu Selbsttäuschung bei? Welche Rolle spielt der soziale Kontext? Wie ist Selbsttäuschung ethisch und moralisch zu beurteilen? Ist moralische Verantwortung für Selbsttäuschung möglich?

In einer ersten Annäherung lässt sich Selbsttäuschung als motivierte Irrationalität beschreiben: Eine Person glaubt, *weil* sie ein Motiv dafür hat, nicht das, was sie angesichts verfügbarer Belege glauben sollte. Zahlreiche psychologische Studien belegen, wie wir aufgrund von Motiven bzw. „directed goals" (Kunda 1990) unsere Fähigkeiten, Zukunftsaussichten und Kontrollmöglichkeiten überschätzen (für eine zusammenfassende Studie vgl. Schütz /Baumeister 2017). Wir schätzen uns und Nahestehende positiver ein, als eine akkurate Beurteilung der Sachlage rechtfertigen würde, und tendieren zu kognitiven Verzerrungen (‚biases', ‚distortions'), die entweder selbst als Selbsttäuschung zu verstehen sind oder aber als Mittel, die helfen,

Selbsttäuschung herbeizuführen und aufrechtzuerhalten (Kahnemann 2011). Trotz starker empirischer Belege für das breite Vorkommen motivierter Verzerrungen ist weiterhin ungeklärt, wie Motive kognitive Prozesse und insbesondere die Überzeugungsbildung beeinflussen – denn einfach glauben, was wir glauben wollen, können wir nicht. Tatsächlich scheinen wir die Mechanismen der Selbsttäuschung überhaupt nur deshalb zu nutzen, weil Rationalitätsstandards und Belege eine gewisse nötigende Kraft auf uns ausüben, der wir uns nur zu dem Maße entziehen können, zu dem es uns gelingt, unsere Empfänglichkeit für Gründe zu manipulieren. Wohl deshalb ist Selbsttäuschung eine Form der Irrationalität, die oft getarnt als Rationalität auftritt. Wir rationalisieren, überdehnen die Grenzen von Deutungsspielräumen, suchen selektiv nach Belegen, die unsere motivierten Fehleinschätzungen stützen, und vermeiden solche, die uns buchstäblich enttäuschen könnten.

Aus philosophischer Perspektive hilft die Auseinandersetzung mit Selbsttäuschung, eine zentrale Variante von Irrationalität zu verstehen und darüber Kernbegriffe philosophischer Theoriebildung wie Rationalität und Überzeugung zu erhellen sowie neue Einsichten über die Funktionsweise des menschlichen Geistes zu gewinnen. Philosophisch ist Selbsttäuschung aber auch deshalb interessant, weil sie dem Ideal rationaler (Selbst-)Erkenntnis widerspricht und auch die Rationalität der philosophischen Praxis in Frage stellt: Wie gut sind Philosophen in der Lage, zwischen guten Argumenten und bloßen (wenn auch vielleicht besonders geschickten) Rationalisierungen zu unterscheiden? Von großer Relevanz ist Selbsttäuschung außerdem für die praktische Philosophie. Selbsttäuschung kann zu Entfremdung führen, eine autonome Lebensführung behindern und dient häufig, etwa in Form von Selbstgerechtigkeit, der Vertuschung und Perpetuierung moralischen Unrechts. Sie kann soziale Ausgrenzung begünstigen und dazu beitragen, dass sich Verschwörungstheorien ausbreiten und festsetzen. Sie kann aber auch adaptive oder schützende Funktionen haben, etwa mit Blick auf psychische Resilienz, Selbstachtung und Wohlergehen. Im Dialog mit den empirischen Wissenschaften können wir ein zunehmend besseres Verständnis davon gewinnen, was Selbsttäuschung ist, wie sie wirkt und welche positiven und negativen Folgen sie hat. Fruchtbar ist auch die Auseinandersetzung mit literarischen Werken wie Flauberts *Madame Bovary*, Tolstois

Anna Karenina, Ibsens *Wildente*, Thomas Manns *Tod in Venedig* oder Frischs *Homo Faber*, da diese die existentielle Tragweite der Selbsttäuschung besonders differenziert und anschaulich zum Ausdruck bringen.

Im Mittelpunkt der neueren philosophischen Debatte um Selbsttäuschung steht die Debatte zwischen intentionalistischen und revisionistischen Lösungsansätzen für die begrifflichen Spannungen in traditionellen Definitionen von Selbsttäuschung (s. Abschn. 2–4). Historisch standen dagegen ethische, moralische und prudentielle Aspekte der Selbsttäuschung im Vordergrund; sie erfahren in der neuesten philosophischen Forschung wieder mehr Beachtung (s. Abschn. 5–6).

2. Begriffliche Spannungen

Im Fokus der aktuellen begrifflichen Diskussion steht die scheinbar paradoxe Struktur von Selbsttäuschung. Nach traditioneller Auffassung ist Selbsttäuschung in struktureller Analogie zu Fremdtäuschung (interpersonaler Täuschung) zu verstehen. Für typische Fälle von Fremdtäuschung scheinen jedoch zwei notwendige Bedingungen zu gelten, deren Übertragung auf Selbsttäuschung zu begrifflichen Spannungen führt.

Die erste Bedingung lautet, dass ein kommunikativer Akt nur dann eine Fremdtäuschung darstellt, wenn die täuschende Person eine andere Person von etwas überzeugt, das sie selbst für falsch hält bzw. von dessen Richtigkeit sie nicht überzeugt ist. Genauer ausgedrückt: Ein kommunikativer Akt stellt nur dann eine Fremdtäuschung dar, wenn die täuschende Person eine andere Person dazu bringt, eine Überzeugung anzunehmen (oder beizubehalten), die die täuschende Person für falsch hält bzw. von deren Richtigkeit sie nicht überzeugt ist, indem sie falsche oder irreführende Belege anführt oder relevante Belege unterschlägt. Diese Bedingung lässt sich als *Bedingung bewusster Diskrepanz* bezeichnen: Die täuschende Person bringt eine andere Person durch falsche, irreführende oder unterschlagene Belege dazu, eine Überzeugung anzunehmen oder beizubehalten, die von ihren eigenen Überzeugungen abweicht, und sie ist sich dieser Abweichung bewusst.

Die zweite Bedingung für Fremdtäuschung lautet, dass die täuschende Person die Absicht haben muss, zu täuschen. Die Täuschungsabsicht wird von den meisten Autorinnen als (zeitweise) bewusste Absicht verstanden. Die *Bedingung einer*

bewussten Täuschungsabsicht erlaubt es, Fremdtäuschung von unbeabsichtigter Irreführung abzugrenzen, zu der es beispielsweise kommen kann, wenn eine Person ohne Täuschungsabsicht ironisch spricht, ihr Gegenüber dies jedoch nicht versteht und deshalb falsche Überzeugungen annimmt. Auch die Weitergabe von falschen Überzeugungen, die man selbst für wahr hält, kann Andere unbeabsichtigt in die Irre führen. Die Abgrenzung von absichtlicher Fremdtäuschung und unbeabsichtigter Irreführung ist vor allem deshalb wichtig, weil es moralisch einen Unterschied macht, ob man mit oder ohne Täuschungsabsicht falsche Überzeugungen hervorruft. Unbeabsichtigte Irreführungen werden bei sonst gleichen Umständen moralisch weniger stark verurteilt als Fremdtäuschungen (obwohl bei unbeabsichtigter Irreführung der Vorwurf der Unachtsamkeit angebracht sein kann und es Fälle unbeabsichtigter Irreführung durch grobe Fahrlässigkeit geben kann, die moralisch problematischer sind als vergleichsweise unschuldige Fälle absichtlicher Fremdtäuschung).

Auf den ersten Blick gibt es gute Gründe, intrapersonale Täuschung analog zur interpersonalen Täuschung zu verstehen und davon auszugehen, dass die genannten Bedingungen auch für Selbsttäuschung gelten. Die strukturelle Engführung beider gewährleistet einen einheitlichen Täuschungsbegriff und wie bei Fremdtäuschung ist auch bei Selbsttäuschung die Abgrenzung von unbeabsichtigter Irreführung bzw. bloßen Fehlern in Wahrnehmungs- und Reflexionsprozessen relevant. Zudem ist die Rede von Selbsttäuschung häufig mit einem Vorwurf verbunden, der implizit unterstellt, dass Selbsttäuschung uns nicht passiv unterläuft, sondern ein aktiver Prozess ist, für den wir zumindest teilweise verantwortlich sind. Das Vorliegen einer bewussten Täuschungsabsicht könnte erklären, worin diese aktive Komponente besteht. Das Vorliegen einer unerwünschten Überzeugung p wiederum könnte erklären, warum die Absicht, sich selbst zu täuschen, überhaupt ausgebildet wird und im Fall erfolgreicher Täuschung zu der Überzeugung ¬p führt. Und die Spannung der widersprüchlichen Überzeugungen p und ¬p könnte erklären, warum die sich selbst täuschende Person die Absicht zur Selbsttäuschung auch bei Erfolg beibehält, weiterhin Maßnahmen zur Selbsttäuschung ergreift (sich also vor einer Ent-Täuschung schützt) und ihre Selbsttäuschung so über längere Zeit aufrechterhalten kann. Gegen eine Übertragung der Diskrepanz- und der Absichtsbedingung auf

Selbsttäuschung spricht allerdings, dass sie zu erheblichen begrifflichen Spannungen führt, weil bei Selbsttäuschung täuschende und getäuschte Instanz in einer Person zusammenfallen.

Statische Probleme: Wird die Bedingung bewusster Diskrepanz auf Selbsttäuschung übertragen, dann impliziert sie, dass die sich selber täuschende Person eine Überzeugung annimmt oder beibehält, von der ihr zugleich bewusst ist, dass sie mindestens einer ihrer weiteren Überzeugungen widerspricht. Sie müsste also urteilen können: ‚Ich halte hier und jetzt sowohl p als auch ¬p für wahr.' Ein solches Urteil aber ist unverständlich und lässt das Phänomen der Selbsttäuschung „statisch paradox" (Mele 2001) erscheinen. Die Paradoxie betrifft dabei nicht notwendig die Inhalte erster Ordnung der fraglichen Überzeugungen (etwa ‚Ich bin großzügig' und ‚Ich bin nicht großzügig'). Vielmehr kann sie auch durch widersprüchliche Überzeugungsgrade bezüglich dieser Inhalte entstehen (etwa: ‚Ich glaube fest daran, großzügig zu sein' und ‚Ich glaube nicht fest daran, großzügig zu sein').

Zu statischen Problemen – zu Problemen also, das Ergebnis bzw. den Zustand der Selbsttäuschung zu einer bestimmten Zeit zu beschreiben – kommt es auch, wenn man die Diskrepanzbedingung auf das Kriterium widersprüchlicher Überzeugungen beschränkt und auf das Kriterium der Bewusstheit verzichtet. Wird die Diskrepanz auf die Zuschreibung von Überzeugungen bezogen, führt dies zu einer logischen Paradoxie: (1) A glaubt, dass p, und (2) es ist nicht der Fall, dass A glaubt, dass p. Diese Zuschreibung von Überzeugungen ist selbstwidersprüchlich; sie verletzt den Satz vom Widerspruch und ist logisch ausgeschlossen. Logisch möglich ist dagegen, dass die sich selber täuschende Person zu einer bestimmten Zeit widersprüchliche Überzeugungen hat: (1) A glaubt, dass p, und (2) A glaubt, dass ¬p. Die Zuschreibung dieser beiden Überzeugungen ist logisch möglich, da nicht behauptet wird, dass A zu einer bestimmten Zeit ein und dieselbe Überzeugung hat und nicht hat (Davidson 2004, Funkhouser 2019). Kontrovers ist allerdings, ob widersprüchliche Überzeugungen in diesem Sinn psychologisch möglich sind und, wenn ja, wie (s. Abschn. 3). Unvereinbare Überzeugungen p und q, deren Unvereinbarkeit wir nicht erkennen, weil wir ihre inferentiellen Beziehungen nicht durchschauen, führen dagegen nicht zu statischen Problemen. Jedoch liegt hier auch keine Selbsttäuschung

vor – es sei denn, wir *wollen* die Unvereinbarkeit von p und q nicht erkennen, was erneut statische Probleme aufwirft.

Das dynamische Problem: Auch die Übertragung der Absichtsbedingung auf Selbsttäuschung gelingt nicht ohne Spannungen: Wie kann eine Person sich in bewusster Täuschungsabsicht dazu bringen, eine Überzeugung anzunehmen oder beizubehalten, von der sie glaubt, dass sie falsch ist? Bewusste Täuschungsabsichten, die sich auf die eigene Person beziehen, scheinen sich selbst zu untergraben, denn wenn wir als Adressatinnen der Täuschung um die Täuschungsabsicht wissen, wird sie kaum mehr erfolgreich sein. In der aktuellen Debatte wird diese Schwierigkeit als ‚dynamisches' (Mele 2001) oder ‚strategisches' Paradox bezeichnet.

3. Lösungsansätze

Die radikalste Lösung der aufgezeigten Spannungen besteht darin, die Möglichkeit von Selbsttäuschung zu bestreiten. So ist beispielsweise Steffen Borge (2003) der Auffassung, dass die Zusammenführung unserer alltagssprachlichen Begriffe ‚Überzeugung' (belief), ‚Täuschung' und ‚Selbst' im Begriffskonglomerat ‚Selbsttäuschung' zu unlösbaren Konflikten führt. Er schließt daraus, dass es Selbsttäuschung nicht geben kann und dass das Phänomen, das wir auf verworrene Weise so bezeichnen, auf andere Weise beschrieben werden muss. Die meisten Autoren aber verteidigen die Möglichkeit von Selbsttäuschung und vertreten, grob gesprochen, entweder eine intentionalistische oder eine revisionistische Auffassung, von denen es jeweils verschiedene Spielarten gibt.

Intentionalistische Lösungsansätze: Eine Gruppe von intentionalistischen Ansätzen hält sowohl an der Diskrepanz- als auch an der Absichtsbedingung fest und definiert Selbsttäuschung strikt analog zu Fremdtäuschung. Zur Vermeidung der aufgezeigten Paradoxien und Spannungen werden in der Tradition Freuds intrapsychische Teilungsmodelle vorgeschlagen, die einem Teil der Psyche die Rolle der täuschenden Instanz, einem anderen die Rolle der getäuschten Instanz zuschreiben. Es gibt stärkere und schwächere Teilungsmodelle. Stärkere Varianten, wie sie etwa Amélie Oksenberg Rorty in ihren frühen Schriften zu Selbsttäuschung vertritt, umgehen die statischen und dynamischen Probleme, werfen aber die Frage auf, ob täuschende und getäuschte Instanz trotz Teilung noch ein hinreichend einheitliches Selbst bilden, von

dem sich sagen lässt, dass es sich selber täuscht. Nach Mark Johnston (1988) laufen starke Teilungsmodelle Gefahr, Selbsttäuschung als Ausdruck einer dissoziativen Persönlichkeitsstörung darzustellen. Intentionalistische Ansätze könnten dadurch ihren vermeintlichen Vorzug verlieren, erklären zu können, wie Verantwortung für Selbsttäuschung möglich ist. Bei schwächeren Teilungsmodellen, wie sie etwa Donald Davidson (2004) (nach zuvor stärkeren Varianten) formuliert hat, treten diese Schwierigkeiten nicht auf, es ist jedoch umstritten, ob sie die statischen und dynamischen Probleme wirklich lösen oder lediglich verschieben.

Eine zweite Gruppe von intentionalistischen Ansätzen löst die statischen Probleme, indem sie Selbsttäuschung als einen zeitlich ausgedehnten Prozess beschreibt. Im Prozess der Selbsttäuschung kann eine Person demnach durchaus (bewusst) widersprüchliche Überzeugungen haben, wenn zwischen ihnen eine zeitliche Trennung besteht. Die Diskrepanzbedingung wird hier also deutlich abgeschwächt oder sogar ganz verworfen, was diese Gruppe intentionalistischer Ansätze bereits in eine gewisse Nähe zu revisionistischen Ansätzen rückt. Mit der ersten Gruppe intentionalistischer Ansätze hat die zweite Gruppe gemein, dass sie an der Absichtsbedingung festhält (in starker oder schwacher Form). Vertreterinnen einer starken Absichtsbedingung sind der Auffassung, dass eine zumindest zeitweise bewusste Absicht zur Täuschung notwendig für Selbsttäuschung ist. Gegen den dynamischen Einwand argumentieren sie, dass bewusste Täuschungsabsichten auf längere Sicht durchaus erfolgreich sein können dank unserer Vergesslichkeit und der Manipulierbarkeit unserer Überzeugungen durch mentale und praktische Habitualisierung. Ein beliebtes Beispiel hierfür ist eine bestimmte Lesart von Pascals berühmter Wette, der zufolge man sich selbst dsazu bringen kann, an Gott zu glauben, indem man ernsthaft so tut, als würde man an Gott glauben, und religiöse Praktiken ausübt.

Fraglich bleibt allerdings, ob es sich dabei um typische Fälle von Selbsttäuschung handelt. Durch eine Abschwächung der Absichtsbedingung rücken manche Intentionalistinnen noch einen Schritt näher an revisionistische Ansätze heran. So argumentiert etwa José Bermúdez (2003), dass ‚Kernepisoden' der Selbsttäuschung keine bewusste Absicht zur Täuschung voraussetzen, sondern lediglich die zeitweise bewusste Absicht, eine bestimmte Überzeugung zu erwerben, von der man weiß, dass

man sie ohne diese Absicht nicht erwerben würde, weil sie angesichts der verfügbaren Belege nicht gerechtfertigt ist. Die erforderliche Absicht zielt demnach auf eine nicht gerechtfertigte Überzeugung ab – etwa auf die Überzeugung, man verhalte sich kooperativ –, nicht auf Täuschung.

Revisionistische Lösungsansätze: Revisionistische Ansätze vermeiden die statischen und dynamischen Probleme, indem sie die Bedingung (bewusster) Diskrepanz und/oder die Bedingung einer Täuschungsabsicht zurückweisen. Sie verneinen in der Regel nicht, dass es die Phänomene, die (schwach) intentionalistische Ansätze beschreiben, geben kann, wohl aber, dass es sich dabei um typische, alltägliche Fälle von Selbsttäuschung handelt. Für eine Revision der Bedingungen für Selbsttäuschung (die je nach Ansatz als notwendig, hinreichend oder auch nur als charakteristisch angesehen werden) gibt es verschiedene Möglichkeiten:

(A) Zurückweisung der Diskrepanzbedingung: Viele Autorinnen weisen die Diskrepanzbedingung und damit die Annahme zurück, dass eine Koexistenz widersprüchlicher Überzeugungen p und ¬p notwendig ist. Sie verneinen entweder, dass die sich selbst täuschende Person (1) die gerechtfertigte (und je nach Ansatz wahre) Überzeugung ¬p oder (2) die ungerechtfertigte (und je nach Ansatz falsche) Überzeugung p oder (3) überhaupt Überzeugungen im vollen Sinne haben muss bezüglich des Sachverhalts, über den sie sich selber täuscht. Autoren, die die Diskrepanzbedingung auf dem ersten Wege (1) zurückweisen, bestreiten, dass eine Person, die sich selber täuscht, die angesichts der verfügbaren Belege gerechtfertigte Überzeugung ¬p (gehabt) haben muss. Selbsttäuschung, etwa in Bezug auf die eigenen Fähigkeiten, setzt demnach nicht voraus, dass man auch nur unbewusst eine gerechtfertigte (und wahre) Überzeugung bezüglich dieser Fähigkeiten hat oder hatte. Nach Alfred Mele (2001) beispielsweise reicht aus, dass die Person, die sich selber täuscht und deshalb fälschlich der Überzeugung p ist (etwa, sie sei eine überdurchschnittlich umsichtige Verkehrsteilnehmerin), Belege besitzt, die stärker für die gegenteilige Überzeugung ¬p sprechen. Mele widerspricht damit Davidson (2004) und anderen Verfechtern widersprüchlicher Überzeugungen, die annehmen, dass bei Selbsttäuschung die gerechtfertigte Überzeugung ¬p die ungerechtfertigte Überzeugung p hervorruft und oft auch aufrechterhält. Um trotz des Verzichts auf die Diskrepanzbedingung die für Selbsttäuschung charakteristische Spannung erfassen

zu können, unterscheidet Richard Holton zwei reaktive Varianten der Selbsttäuschung. Bei „running self-deception" und "up-front self-deception" ist "a triggering state [...] by the agent's own lights, *in tension* with the self-deceptive belief" (Holton 2022).

Autorinnen, die widersprüchliche Überzeugungen auf dem zweiten Wege (2) vermeiden, verneinen, dass eine Person, die sich selber täuscht, die angesichts der verfügbaren Belege ungerechtfertigte Überzeugung p erwerben muss. Im Falle von (1) gibt es keine Überzeugung, die den Ausgangspunkt eines Prozesses der Selbsttäuschung bildet (und später beibehalten oder verworfen wird), im Falle von (2) keine Überzeugung, die das Ergebnis dieses Prozesses darstellt. Nach Robert Audi (1982) weiß eine Person, die sich selber täuscht, unbewusst, dass ¬p, bekennt sich aber zu p; Ergebnis erfolgreicher Selbsttäuschung ist Audi zufolge keine Überzeugung, sondern ein Bekenntnis (‚avowal'). Auch Tamar Gendler (2007) widerspricht der Annahme, dass Selbsttäuschung zu ungerechtfertigten Überzeugungen führt, versteht das Ergebnis paradigmatischer Fälle von Selbsttäuschung aber als eine Art Vortäuschung (‚pretense'), die auf phänomenaler und motivationaler Ebene überzeugungsähnliche Folgen hat, anders als Überzeugungen aber nicht auf Wahrheit ausgerichtet ist. Gemeinsam ist Erklärungsansätzen des zweiten Typs, dass sie versuchen, die inhärente Spannung einzufangen, die gemeinhin als charakteristisch für Selbsttäuschung gilt, diese aber nicht als Paradoxie widersprüchlicher Überzeugungen beschreiben, sondern als Spannung zwischen Überzeugungen und Einstellungen anderer Art, etwa subdoxastischen oder überzeugungsähnlichen Einstellungen. Die sich selbst täuschende Verkehrsteilnehmerin weiß demnach (unbewusst), dass sie nicht überdurchschnittlich umsichtig ist, hat aber zudem Einstellungen anderer Art, die diesem (unbewussten) Wissen widersprechen und beeinflussen, wie sie denkt, fühlt und handelt.

Erklärungsansätze des dritten Typs (3) gehen noch einen Schritt weiter und versuchen zu zeigen, dass unter den konfligierenden Einstellungen überhaupt keine vollwertigen Überzeugungen sein müssen. Nach Eric Funkhouser (2019) lassen sich Fälle von Selbsttäuschung, die mit tiefgreifenden Konflikten einhergehen (‚deeply conflicted self-deception'), am besten durch konfligierende subdoxastische Einstellungen erklären. Mit Eric Schwitzgebel (2001; vgl. Funkhouser 2019) lassen sich diese

Konflikte als Konflikte dispositionaler Zustände beschreiben, die er ‚in-between believing' nennt, weil die sich selbst täuschende Person die propositionalen Gehalte dieser Zustände weder glaubt noch nicht glaubt.

(B) Zurückweisung der Absichtsbedingung: Die Zurückweisung der Diskrepanzbedingung löst das statische Problem und entschärft das dynamische. Eine Zurückweisung der Absichtsbedingung zielt direkt auf die Lösung des dynamischen Problems und wird in vielen Ansätzen mit einer Zurückweisung der Diskrepanzbedingung in einer der drei genannten Varianten kombiniert. Autoren, die die Absichtsbedingung zurückweisen, argumentieren, dass Täuschungsabsichten nicht notwendig sind, um Selbsttäuschung zu erklären. Sie gehen davon aus, dass Selbsttäuschung typischerweise nicht durch eine Täuschungsabsicht motiviert wird, sondern durch einen Wunsch, eine Emotion, ein Ziel oder eine Kombination solcher konativen Einstellungen. So kann man Max Frischs *Homo faber* als Veranschaulichung eines komplexen Falls von Selbsttäuschung lesen, der primär durch Angst motiviert wird: Die Hauptfigur Walter Faber täuscht sich selber über die Identität seiner Tochter und insbesondere über sich selbst, weil er Angst davor hat, entdecken zu müssen, der Geliebte seiner eigenen Tochter zu sein. Man kann aber auch annehmen, dass der Wunsch, nicht der Geliebte der eigenen Tochter zu sein, das zentrale Motiv von Fabers Selbsttäuschung ist – oder aber der Wunsch, zu *glauben*, dass man nicht der Geliebte der eigenen Tochter ist. Nach Annette Barnes (1997) dient Selbsttäuschung bezüglich p der Angstreduktion und wird durch den Wunsch, q möge der Fall sein, und die Angst, es sei nicht der Fall, dass q, motiviert (wobei möglich, aber nicht notwendig ist, dass p und q identisch sind). Mele (2001) erklärt Selbsttäuschung dagegen über unsere allgemeine psychologische Tendenz, motiviert-verzerrte Einstellungen (‚hot biases') auszubilden und lässt offen, ob ein Wunsch oder ein anderes Motiv die Verzerrung verursacht. Entscheidend ist nur, dass die verzerrte Auslegung (oder Vermeidung) vorliegender Belege zu einer ungerechtfertigten, falschen Überzeugung führt.

Es ist umstritten, ob Wünsche und andere Motive, die sich auf Sachverhalte in der Welt beziehen, erklären können, warum auch Fälle negativer Selbsttäuschung (‚twisted self-deception') vorkommen, in denen eine Person eine Überzeugung annimmt, die ihren Wünschen widerspricht. Ein beliebtes Beispiel ist der eifersüchtige Ehemann, der seiner Frau ungerechtfertigterweise eine Affäre unterstellt, obwohl er das

Gegenteil wünscht. Um auch negative Fälle von Selbsttäuschung erklären zu können, argumentiert u.a. Dana Nelkin (2012) für ein restriktiveres, spezifischeres Verständnis des Motivs: Selbsttäuschung werde motiviert durch den Wunsch zu *glauben*, dass ¬p. Die Ausrichtung des Motivs auf einen mentalen Zustand soll verständlich machen, warum eine Person, die sich beispielsweise vor einer Enttäuschung schützen möchte, motiviert sein kann, das Gegenteil von dem zu glauben, was sie sich wünscht. Einen weiteren Vorzug ihres ‚Desire to Believe'-Ansatzes sieht Nelkin in der Abgrenzung der Selbsttäuschung von anderen Formen motivierter Irrationalität wie zum Beispiel von Wunschdenken, die sich auf Sachverhalte in der Welt beziehen. Kontrovers bleibt allerdings, ob eine solche Abgrenzung sinnvoll ist: Ist Wunschdenken ein eigenständiges Phänomen oder eine Form von Selbsttäuschung?

4. Minimaldefinition

Bei allen Differenzen, die die Debatte zwischen intentionalistischen und revisionistischen Ansätzen in ihren verschiedenen Spielarten kennzeichnet, lässt sich doch eine Minimaldefinition von Selbsttäuschung angeben, die die meisten Autorinnen zumindest als Ausgangspunkt akzeptieren können: Eine Person täuscht sich selbst über p, wenn die verfügbaren Belege im Lichte von Rationalitätsstandards, die sie selbst üblicherweise anwendet, stärker für nicht-p sprechen, sie aber aufgrund eines verzerrenden Umgangs mit diesen Belegen zu einer ungerechtfertigten Einstellung bezüglich p kommt und die Verzerrung motiviert ist. Selbsttäuschung, minimal verstanden, ist eine kognitive Reaktion auf verfügbare Belege, die motiviert irrational ist. Das Vorliegen eines Motivs erlaubt es, Selbsttäuschung von unmotivierten kognitiven Fehlern (‚cold biases') wie dem Ankereffekt oder der Verfügbarkeitsverzerrung abzugrenzen, die bei der Anwendung von Urteilsheuristiken auftreten können, ohne dass ein Motiv diese Fehler verursacht. Uneinigkeit herrscht darüber, ob diese Definition zu minimal ist, um das zu erfassen, was Selbsttäuschung im Kern ausmacht, und sie abzugrenzen von verwandten Phänomenen wie Wunschdenken, gewolltem Unwissen, Konfabulation und solchen ‚hot biases', die keine Selbsttäuschung sind. Kontrovers bleibt auch, ob und gegebenenfalls wie solche Abgrenzungen erfolgen sollen. Dies zeigt, dass nicht nur

die Definition von Selbsttäuschung umstritten ist, sondern auch, was überhaupt paradigmatische Fälle von Selbsttäuschung sind.

5. Ethische, moralische und prudentielle Beurteilung

Die neuere philosophische Debatte konzentriert sich vor allem auf die Frage nach der Möglichkeit von Selbsttäuschung und die Auflösung begrifflicher Paradoxien. Die Auseinandersetzung mit ethischen, moralischen und prudentiellen Aspekten der Selbsttäuschung spielt bis in jüngste Zeit nur eine untergeordnete Rolle. In der Philosophiegeschichte stand die ethische und moralische Beurteilung von Selbsttäuschung dagegen im Vordergrund und das Urteil fiel überwiegend sehr negativ aus: Selbsttäuschung galt als große Gefahr für Selbsterkenntnis, Autonomie und moralische Integrität (Joseph Butler (1729), Kant (1797)), als Deckmantel für unmoralisches Handeln (Kant (1797)), als Ausdruck von Unaufrichtigkeit (Sartre (1943)) und uneingestandener Parteilichkeit und sogar als „the source of half the disorders of human life" (Smith (1759), TMS III. 4.6). Viele der historischen Kritikpunkte sind für eine moralphilosophische Beurteilung von Selbsttäuschung bis heute relevant: Selbsttäuschung kann zu einer massiven Einschränkung unserer Urteils- und Handlungsfähigkeit führen. Aus prudentieller Perspektive ist dies problematisch, weil durch Selbsttäuschung verzerrte Überzeugungen zweckrationales Handeln erschweren oder sogar verunmöglichen können. Aus ethischer und moralischer Perspektive lassen sich die Einschränkung autonomer Urteils- und Handlungsfähigkeit problematisieren, mangelnde Selbsterkenntnis und der Schaden, den Selbsttäuschung für Andere bedeuten kann. Neuere psychologische Studien deuten allerdings darauf hin, dass Selbsttäuschung unter bestimmten Bedingungen auch positive Funktionen erfüllen kann, indem sie zu psychischer Stabilität beiträgt, soziale Beziehungen fördert und im Sinne selbsterfüllender Prophezeiungen unsere Fähigkeiten mobilisiert (Taylor 1989; Newen/Vosgerau 2017; Schütz/Baumeister 2017). Diese Ergebnisse stellen die Einseitigkeit der historisch dominierenden Verurteilung von Selbsttäuschung in Frage und könnten mit Blick auf die moralische Handlungsfähigkeit konkreter Personen (im Unterschied zu idealen, rationalen Akteuren) sehr relevant sein.

Im Spannungsfeld dieser drei Diskussionsstränge – des begrifflichen, des philosophiegeschichtlichen und des psychologischen – sind noch viele Fragen ungeklärt, weshalb prudentielle, ethische und moralische Aspekte der Selbsttäuschung wieder zunehmend in den Fokus der Forschung rücken. Neil Van Leeuwen (2009) beispielsweise kritisiert, dass die Rede von ‚positiven Illusionen', die Shelley Taylor in die psychologische Debatte eingeführt hat, selbstwidersprüchlich sei, und zieht so eine Verbindung zwischen der psychologischen und der begrifflichen Debatte. Außerdem bezweifelt er, dass Selbsttäuschung ein geeignetes Mittel zu individuellem Wohlergehen ist, insbesondere auf längere Sicht, weil sie die Motivation zur Verbesserung der eigenen Fähigkeiten hemmt und typischerweise mit unterschwelligen psychischen Spannungen einhergeht, die Ängste und andere negative Emotionen schüren. Oksenberg Rorty (1988) verteidigt dagegen die These, dass Selbsttäuschung einen prudentiellen Nutzen haben kann: Sie sei ein irrationales, aber effektives Hilfsmittel gegen Melancholie, das uns helfen könne, die Energie aufzubringen, die wir für die Verwirklichung persönlicher Projekte benötigen. Auch nach Carla Bagnoli (2015) hat Selbsttäuschung als pragmatische Strategie der Selbststabilisierung einen prudentiellen Nutzen. Sie betont aber zugleich, dass dieser Nutzen mit erheblichen Autonomieverlusten einhergehe, und greift damit einen Schwerpunkt der historischen Kritik auf. Stephen Darwall (1988) untersucht ebenfalls das Verhältnis von Selbsttäuschung und Autonomie, allerdings in direkter Auseinandersetzung mit historischen Positionen. Seiner Auffassung nach ist Selbsttäuschung für konstitutionalistische Moraltheorien wie die von Butler (1727) und Kant (1797) besonders problematisch, weil für solche Theorien die moralische Qualität unserer Handlungen und unseres Charakters von bestimmten prozeduralen Tugenden des Urteilens abhängt. Aus konstitutionalistischer Perspektive ist Selbsttäuschung Darwall zufolge intrinsisch falsch, weil sie unsere autonome Urteilsfähigkeit untergräbt und damit genau die Fähigkeit, die die Moralität unserer Handlungen und unseres Charakters bestimmt. Anna Wehofsits (2020) argumentiert, dass Kants berüchtigte Kritik an Leidenschaften vor allem eine Kritik an rationalisierender Selbsttäuschung ist und die wichtige moralpsychologische Einsicht zum Ausdruck bringt, dass oft ausgerechnet ein Wunsch nach moralischer Integrität Selbsttäuschung motiviert und aufrechterhält. Kathi Beier (2010) entwickelt ihr Verständnis von

Selbsttäuschung als Privation vor allem im Rekurs auf Aristoteles und erläutert sie als Mangel unserer Fähigkeit zur Selbstbestimmung. Für Beier ist Selbsttäuschung der selbstverschuldete Grund, kein freies, autonomes und gutes Leben zu führen.

6. Moralische Verantwortung

Alltagssprachlich ist die Zuschreibung von Selbsttäuschung häufig mit dem Vorwurf verbunden, man hätte es besser wissen können und sollen. Wir machen uns und andere für Selbsttäuschung und ihre Folgen moralisch verantwortlich, häufig mit Verweis auf Belege, die der sich selbst täuschenden Person vorlagen oder aber leicht zugänglich gewesen wären. Aber ist die Zuschreibung moralischer Verantwortung berechtigt, wenn Selbsttäuschung ohne Täuschungsabsicht und (teils) unbewusst geschieht?

Die Annahme, dass Selbsttäuschung durch eine Absicht motiviert wird, macht es für intentionalistische Ansätze zunächst leichter als für nicht-intentionalistische, revisionistische Ansätze, zu erklären, wie moralische Verantwortung für Selbsttäuschung möglich ist: Absichtliche Handlungen sind typische Fälle verantwortlichen Handelns. Rekurrieren intentionalistische Ansätze allerdings auf starke psychische Teilungsmodelle, die die Einheit des Selbst gefährden, so ist fraglich, ob moralische Verantwortung zugeschrieben werden kann. Selbsttäuschung wäre dann möglicherweise ein pathologisches Phänomen.

Auch viele Vertreter nicht-intentionalistischer Ansätze (die oft auch als deflationistische oder motivationalistische Ansätze bezeichnet werden) gehen davon aus, dass wir für Selbsttäuschung moralisch verantwortlich sein können. Umstritten ist allerdings, ob ihre Ansätze die Ressourcen bereitstellen, um dies verständlich zu machen, da sie nicht nur annehmen, dass Selbsttäuschung in der Regel ohne Täuschungsabsicht erfolgt, sondern zudem, dass sie oft unbewusst geschieht. Aus diesem Grund argumentiert beispielsweise Neil Levy (2004), dass nicht-intentionalistische Ansätze weder Bedarf noch Raum haben, moralische Verantwortung für paradigmatische Fälle von Selbsttäuschung zuzuschreiben. Seiner Auffassung nach ist sich eine Person, die sich selber täuscht, der kognitiven Mechanismen, die ihre Überzeugungen verzerren, meist nicht bewusst, weshalb ihr die Kontrolle fehlt, die für moralische Verantwortung erforderlich ist. Bei fehlendem

Bewusstsein der verzerrenden Mechanismen fehle der Anlass zu einer sorgfältigen Überprüfung der fraglichen Überzeugungen. Moralische Verantwortung könne deshalb in der Regel weder für einzelne Episoden der Selbsttäuschung noch für eine allgemeine Disposition zur Selbsttäuschung zugeschrieben werden.

Da Selbsttäuschung schwerwiegende Folgen für Andere haben kann – sie kann dazu führen, dass das Ausmaß moralischen Unrechts unterschätzt oder übersehen wird, und so beispielsweise zu Gewalt- und Diskriminierungsverhältnissen beitragen – ist eine Zurückweisung der Möglichkeit moralischer Verantwortung mit hohen Kosten verbunden. Vor diesem Hintergrund argumentiert Ian DeWeese-Boyd (2007) gegen Levy, dass Personen, die sich selber täuschen, die Fähigkeit haben, verzerrenden Einflüssen von Motiven zu widerstehen, und damit über ausreichende Kontrolle verfügen, um moralische Verantwortung zu tragen, wenn sie nicht widerstehen. Schuldhaft sei Selbsttäuschung dann, wenn sie dazu dient, moralisches Unrecht zu ermöglichen, wobei der Grad der Schuldhaftigkeit von der Schwere des Unrechts abhänge und von den Anstrengungen, die die Person unternehmen muss, um Selbsttäuschung im fraglichen Fall zu verhindern. Ähnlich argumentiert Nelkin (2012) mit Verweis auf fahrlässige Handlungen, dass keine Absicht, kein Bewusstsein des verzerrenden Mechanismus und auch keine bewusste Entscheidung gegen Maßnahmen zur Vermeidung drohender kognitiver Verzerrungen notwendig ist, um moralische Verantwortung für Selbsttäuschung zuschreiben zu können. Sie teilt die verbreitete Auffassung, dass eine Person nur dann für ihr Handeln moralisch verantwortlich sein kann, wenn sie für Gründe empfänglich ist. Anders als beispielsweise Levy nimmt sie aber nicht an, dass der *Mechanismus*, durch den eine Person sich selber täuscht, selbst auf moralische und nicht-moralische Gründe reagieren muss, damit sie für den Prozess der Selbsttäuschung und ihre Folgen moralisch verantwortlich sein kann. Entscheidend sei vielmehr, ob die *Person*, die sich selber täuscht, in der Lage gewesen wäre, den fraglichen Mechanismus außer Kraft zu setzen oder einen anderen zu aktivieren.

Ein weiteres Argument zur Verteidigung der Möglichkeit von moralischer Verantwortung für Selbsttäuschung hat folgende Stoßrichtung: Wir können für Selbsttäuschung und ihre Folgen auch dann moralisch verantwortlich sein, wenn sie Folge unseres Versäumnisses ist, ausreichend moralische Sensibilität zu kultivieren,

um zu erkennen, wann es moralisch erforderlich ist, unsere Urteile und unser Verhalten besonders gründlich zu überprüfen (Darwall 1988) – und nicht schwerwiegende Umstände wie frühkindliche Vernachlässigung oder kognitive Beeinträchtigungen dieses Versäumnis entschuldigen. Problematisch bleibt allerdings, dass durch Selbsttäuschung erzeugte ‚Blindheit' und geschickte Rationalisierungen eine solche Überprüfung möglicherweise überdauern.

7. Offene Forschungsfragen

Die Debatte um intentionalistische und revisionistische Auffassungen von Selbsttäuschung ist bereits sehr ausdifferenziert. Mit Blick auf ein nuancierteres Verständnis der Motive für Selbsttäuschung, der Mechanismen, über die sie funktioniert, und der mentalen Zustände, die sie herbeiführt, sind dagegen noch viele Forschungsfragen offen, die sich wohl am besten im interdisziplinären Dialog beantworten lassen. Forschungsbedarf besteht außerdem aus moral- und sozialphilosophischer Perspektive: Wie lassen sich Grade der Kontrolle und Grade der Verantwortung in Bezug auf Selbsttäuschung und ihre Folgen unterscheiden? Wie trägt Selbsttäuschung zu moralischem Unrecht bei? Worin bestehen ihre sozialen Ermöglichungsbedingungen (Dietz 2017)? Kann die Erforschung von Selbsttäuschung uns dabei helfen, kollektive Prozesse verzerrter Meinungsbildung etwa bei der Verbreitung von ‚fake news' und Verschwörungstheorien besser zu verstehen?

Literatur

Audi, R.: Self-Deception, Action and Will. In: Erkenntnis 18/2 (1982), 133–158.

Bagnoli, C.: Self-Deception and Agential Authority. A Constitutivist Account. In: Humana. Mente Journal of Philosophical Studies 5/20 (2012) 93–116.

Barnes, A.: Seeing through Self-Deception. New York 1997.

Beier, K.: Selbsttäuschung. Berlin 2010.

Bermúdez, J.: Self-Deception, Intentions, and Contradictory Beliefs. In: Analysis 60/4 (2000), 309–319.

Borge, S.: The Myth of Self-Deception. In: Southern Journal of Philosophy 41/1 (2003), 1–28.

Butler, J.: Fifteen Sermons and Other Writings on Ethics [1729]. Hg. von D. McNaughton, Oxford 2017.

Darwall, S.: Self-Deception, Autonomy, and Moral Constitution. In: A. Oksenberg Rorty, B.P. McLaughlin (Hg.): Perspectives on Self-Deception. Berkeley 1988, 408–430.

Davidson, D.: Problems of Rationality. New York 2004.

DeWeese-Boyd, I.: Taking Care. Self-Deception, Culpability and Control. In: Teorema 26/3 (2007), 161–176.

Dietz, S.: Selbsttäuschung als sozialer Prozess. In: E. Angehrn, J. Küchenhoff (Hg.): Selbsttäuschung. Eine Herausforderung für Philosophie und Psychoanalyse. Weilerswist 2017, 223–239.

Funkhouser, E.: Self-Deception. London/New York 2019.

Gendler, T.: Self-Deception as Pretense. In: Philosophical Perspectives 21/1 (2007), 231–258.

Holton, R.: Self-Deception and the Moral Self. In M. Vargas; J. M. Doris (Hg.): The Oxford Handbook of Moral Psychology. Oxford 2022, 262-284.

Johnston, M.: Self-Deception and the Nature of Mind. In: A. Oksenberg Rorty; B.P. McLaughlin (Hg.): Perspectives on Self-Deception. Berkeley 1988, 63–91.

Kant, I.: Die Metaphysik der Sitten [1797]. In: Akademie-Ausgabe, Preußische Akademie der Wissenschaften. Berlin 1900ff., Bd. 6.

Kahneman, D.: Thinking, Fast and Slow. New York 2011.

Kunda, Z.: The Case for Motivated Reasoning. In: Psychological Bulletin 108/3 (1990), 480–498.

Levy, N.: Self-Deception and Moral Responsibility. In: Ratio 17/3 (2004), 294–311.

Mele, A.: Self-Deception Unmasked. Princeton 2001.

Nelkin, D.: Responsibility and Self-Deception. A Framework. In: Humana. Mente Journal of Philosophical Studies 5/20 (2012), 117–139.

Newen A., Vosgerau G.: Irren ist … sinnvoll! In: S. Ayan (Hg.): Rätsel Mensch - Expeditionen im Grenzbereich von Philosophie und Hirnforschung. Berlin 2017, 35–40.

Oksenberg Rorty, A.: The Deceptive Self. Liars, Layers, and Lairs. In: A. Oksenberg Rorty, B.P. McLaughlin (Hg.): Perspectives on Self-Deception. Berkeley 1988, S. 11–28.

Sartre, J.-P.: L' être et le néant. Essai d'ontologie phénoménologique [1943]. Paris 2017.

Schütz, A.; R.F. Baumeister: Positive Illusions and the Happy Mind. In: M.D. Robinson, Michael; M. Eid (Hg.): The Happy Mind. Cognitive Contributions to Well-Being. Heidelberg 2017, 177–193.

Schwitzgebel, E.: In-Between Believing. In: The Philosophical Quarterly 51/202 (2001), 76–82.

Smith, A.: The Theory of Moral Sentiments [1759]. Hg. von D.D. Raphael, A.L. Macfie, Oxford 1976 [= TMS].

Taylor, S. E.: Positive Illusions. Creative Self-Deception and the Healthy Mind. New York 1989.

Van Leeuwen, N.: Self-Deception Won't Make You Happy. In: Social Theory and Practice 35/1 (2009) 107–132.

Wehofsits, A.: Passions: Kant's psychology of self-deception. In: Inquiry, 66:6 (2023/online 2020) 1184-1208.