

Zwölf Antworten auf Williams' Paradox

Marc Andree Weber

Zusammenfassung

Theorien personaler Identität müssen sich an einem auf Bernard Williams zurückgehenden Paradox messen lassen: Manche Situationen zeigen ganz offensichtlich, dass allein geistige Kontinuität für unser Überleben entscheidend ist; andere hingegen zeigen ebenso offensichtlich, dass einzig körperliche Kontinuität zählt. Verschiedene Autoren haben ganz unterschiedliche und einander teils widersprechende Antworten auf dieses Problem gefunden. Anhand aktueller Forschungen zur Struktur philosophischer Gedankenexperimente werden diese Antworten im vorliegenden Aufsatz im Rahmen einer Neubewertung erstmals übersichtlich klassifiziert. Vor allem aber werden mehrere bestehende Antwortansätze weiter entwickelt und zwei neue Antworten formuliert.

Theories of personal identity face a paradox, which traces back to Bernard Williams: some scenarios obviously show that mental continuity is what solely matters in survival; others, on the contrary, show with equal obviousness that it is bodily continuity. Different authors have produced diverging and partly conflicting answers in response to that problem. Based on recent research concerning the structure of philosophical thought experiment, this paper reevaluates and, for the first time, neatly classifies those answers. What is more, several existing approaches of how to answer the paradox are developed further, and two new answers are introduced.

Williams' Paradox

In seinem eleganten und viel beachteten Aufsatz *The Self and the Future* (1970) präsentiert Bernard Williams zwei auf demselben imaginierten Szenario beruhende Gedankenexperimente, deren zentrale Intuitionen allem Anschein nach jeweils zuverlässig sind und deren Aussagen einander dennoch widersprechen. Im ersten werden zwei Personen, schlicht *A* und *B* genannt,

einer Prozedur unterzogen, bei der sämtliche in *As* Gehirn gespeicherten Informationen auf *Bs* Gehirn überspielt werden und umgekehrt. Von den resultierenden Personen hat somit eine *As* Körper und *Bs* Erinnerungen, Überzeugungen und sonstige geistigen Eigenschaften und die andere *Bs* Körper und *As* Erinnerungen, Überzeugungen und sonstige geistigen Eigenschaften. Entsprechend heißen die resultierenden Personen „*A*-Körper-Person“ und „*B*-Körper-Person“. Williams ist vorsichtig genug zu betonen, dass möglicherweise eine nicht geringe Ähnlichkeit zwischen *A* und *B* vorausgesetzt werden muss, damit zum Beispiel der *B*-Körper *As* Charakterzüge und Verhaltensweisen überhaupt halbwegs gut widerspiegeln kann.

Angenommen nun, eine der beiden aus dieser Prozedur resultierenden Personen wird mit 100.000 € beschenkt, die andere gefoltert. Und angenommen, *vor* der Prozedur würden *A* und *B* gefragt, welcher Person welches Schicksal zuteil werden soll. Was sollte *A*, was *B* im eigenen Interesse wählen? Williams spielt die verschiedenen Alternativen ausführlich durch und liefert auf diese Weise ein starkes Argument dafür, dass physische Kontinuität noch nicht einmal eine notwendige Bedingung für personale Identität darstellt, sondern allein psychische Kontinuität ausschlaggebend ist. Denn wenn beispielsweise *A* wünscht, dass die *B*-Körper-Person die 100.000 € erhält, und *B* wünscht, dass die *A*-Körper-Person die 100.000 € erhält, und wenn Geld und Qualen gemäß *As* Wunsch verteilt werden, dann wird die *B*-Körper-Person nach dem Körpertausch echte oder vermeintliche Erinnerungen¹ daran haben, den nun realisierten Wunsch geäußert zu haben, und wird sich darüber freuen, die Wahl, die sie als ihre eigene empfindet, verwirklicht zu sehen. Die *A*-Körper-Person hingegen wird echte oder vermeintliche Erinnerungen daran haben, anders gewählt zu haben, und damit hadern, dass dem Wunsch, den sie als ihren eigenen empfindet, nicht entsprochen wurde. Wenn dagegen *A* der *B*-Körper-Person die Folter zuteil werden lassen möchte und *B* der *A*-Körper-Person und wieder *As* Wunsch verwirklicht wird, dann wird die *B*-Körper-Person der Meinung sein, um Folter für die *B*-Körper-Person gebeten zu haben, und wird sich sehr ärgern, dass dieser Bitte nachgekommen wurde. Die *A*-Körper-Person jedoch, die sich (tatsächlich oder vermeintlich)

¹Erinnerungen werden üblicherweise als faktiv angesehen: Man kann sich nur an das erinnern, was man selbst tatsächlich erlebt hat. Angenommen etwa, physische Kontinuität wäre das für personale Identität Ausschlaggebende. Dann wäre die *B*-Körper-Person nicht mit *A* identisch, und die „Erinnerungen“ der *B*-Körper-Person an das, was *A* dachte, wären keine echten, sondern bloß vermeintliche Erinnerungen. Um nicht schon durch tendenziöse Formulierungen bestimmte Weichen zu stellen, spreche ich daher vorsichtiger und umständlicher von „echten oder vermeintlichen Erinnerungen“. (Weiter unten, in der „Vierten Antwort“ und der „Zwölften Antwort“, wird das Thema der bloß vermeintlichen Erinnerungen noch vertieft.)

erinnert, für die A-Körper-Person das schlimme Schicksal gewählt zu haben, wird erleichtert sein, dass diese ihrer Auffassung nach von ihr getroffene Wahl nicht umgesetzt wurde. Beide sehen sich mit großer Selbstverständlichkeit als mit der Person identisch an, deren geistige Eigenschaften sie teilen.

So weit, so unspektakulär. Philosophen wie Locke oder Shoemaker haben ähnlich argumentiert (Locke 1700/1975; Shoemaker 1963). Im zweiten Gedankenexperiment aber fragt sich Williams, ob der blanke Horror, den wir angesichts der Aussicht, morgen gefoltert zu werden, empfinden, wohl komplett verschwände, wenn wir wüssten, dass (i) vorher unser Gedächtnis vollständig gelöscht wird. Eine Amnesie, so seine Antwort, würden wir als zusätzliches Übel empfinden, die Folter selbst aber weiterhin fürchten. Die Aussicht auf einen Gedächtnisverlust mit anschließender Marter erscheint uns deutlich schlimmer als die Aussicht auf einen Gedächtnisverlust ohne anschließende Marter. Und warum sollte sich an unserer Furcht vor den Schmerzen etwas ändern, wenn (ii) bei uns zusätzlich zum Gedächtnisverlust auch noch bestimmte Charaktermerkmale verändert werden, oder wenn (iii) wir darüber hinaus einige fiktive Erinnerungen induziert bekommen? Oder wenn (iv) illusionäre Erinnerungen und andere illusionäre geistige Eigenschaften so designt sind, dass sie denen einer tatsächlichen Person entsprechen, oder (v) gar von dem Gehirn dieser tatsächlichen Person auf unseres überspielt werden, ohne das Gehirn jener Person zu verändern? Williams' zweites Gedankenexperiment zeigt auf, dass wir die Folter, die einer Person mit unserem Körper, aber den geistigen Eigenschaften eines anderen bevorsteht, sehr wohl als *schlimm für uns* empfinden.

Kann es dann noch einen Unterschied machen, wenn (vi) nicht nur die geistigen Eigenschaften einer anderen Person auf unser Gehirn überspielt werden, sondern auch unsere geistigen Eigenschaften auf das Gehirn dieser anderen Person? Für die demnächst gefoltert werdende Person besteht zwischen (v) und (vi) kein spürbarer Unterschied; die einzige Veränderung betrifft lediglich jemand ganz anderen, der mit den bevorstehenden Qualen nichts zu tun hat. Es scheint daher irrational, in Situation (v) die Folter zu fürchten und in (vi) nicht. Und doch ist Situation (vi) genau die im ersten Gedankenexperiment beschriebene, und wenn die durch dieses Gedankenexperiment nahegelegte Schlussfolgerung korrekt ist, haben wir in (vi) keinen Grund, uns mit der demnächst gefoltert werdenden Person zu identifizieren. Wenn wir in (i) jedoch zurecht die Folter fürchten und zwischen (v) und (vi) kein relevanter Unterschied besteht, müsste nach dem ersten Gedankenexperiment irgendwo zwischen (i) und (v) der Punkt kommen, an dem diese Furcht unvernünftig wird. Keiner der Zwischenschritte erscheint allerdings dafür bedeutsam genug. Also scheint die Konklusion des ersten Gedankenexperiments, obgleich höchst überzeugend, falsch zu sein. Diese verwirrende

Situation werde ich im Folgenden als *Williams' Paradox* bezeichnen.

Die Frage, wie sich Williams' Paradox am besten auflösen lässt, hängt nicht von einer speziellen Theorie von Gedankenexperimenten ab. Der Einfachheit halber möchte ich dennoch im Folgenden eine bestimmte Auffassung philosophischer Gedankenexperimente voraussetzen. Dieser Auffassung zufolge sind Gedankenexperimente gültige Argumente mit zwei Prämissen. Die erste dieser Prämissen konstatiert die begriffliche Möglichkeit eines bestimmten imaginierten Szenarios; die zweite besagt, dass ein bestimmter in diesem Szenario geschilderter Sachverhalt auf eine spezifische Weise beschrieben oder bewertet werden sollte. Die Konklusion etabliert dann die Möglichkeit einer spezifischen Beschreibung oder Bewertung und greift auf diese Weise eine bestimmte Unmöglichkeitsthese an, gegen die das Gedankenexperiment gerichtet ist. (Auf ganz ähnliche Weise kann man Möglichkeitsthesen widerlegen: Man schließt aus der Abwegigkeit einer spezifischen Beschreibung oder Bewertung auf die Unmöglichkeit eines bestimmten imaginierten Szenarios, das sich nur so plausibel beschreiben oder bewerten lässt.)² Dieser Strukturanalyse von Gedankenexperimenten liegt die Einsicht zugrunde, dass Thesen, die sich mithilfe von Gedankenexperimenten untermauern oder angreifen lassen, modalen Natur sein müssen, also Möglichkeiten und Notwendigkeiten betreffen. Thesen über die tatsächliche, nichtmodale Beschaffenheit der Welt überprüft man sinnvollerweise durch empirische Beobachtungen; *Gedankenexperimente* dagegen klammern Empirie als Erkenntnisquelle explizit aus.³ Für die Auseinandersetzung mit Williams' Paradox besteht der explanative Gewinn der Charakterisierung philosophischer Gedankenexperimente als zweiprämisierte Argumente darin, dass sich dieser Charakterisierung zufolge Gedankenexperimente auf genau zwei Weisen kritisieren lassen, indem man nämlich entweder die erste oder die zweite Prämisse angreift. Das erleichtert uns die Klassifizierung von Vorschlägen, wie mit diesem Paradox umzugehen ist. Insoweit die im Folgenden vorgestellte Klassifizierung überzeugt, stellt sie umgekehrt ein Argument für die beschriebene Strukturanalyse philoso-

²In Weber 2014, Kap. 1, und Weber *im Erscheinen* erläutere ich diese Strukturanalyse philosophischer Gedankenexperimente genauer. Ähnliche Analysen finden sich in Sorensen 1992, Gendler 2000, Williamson 2007 und Häggqvist 2009. Insbesondere die Williamson's ist meiner sehr ähnlich; der Hauptunterschied besteht darin, dass Williamson fordert, das imaginierte Szenario müsse metaphysisch möglich sein (Williamson 2007, S. 205–6), während ich argumentiere, dass begriffliche Möglichkeit genügt (Weber 2014, S. 56–63). In der vorliegenden Arbeit hängt an diesem Unterschied allerdings kaum etwas; lediglich die erste der unten aufgeführten Antworten wäre im Falle bloß metaphysischer Möglichkeit noch etwas zwingender.

³Auf einem anderen Weg – Durchführung einer Vielzahl von Fallstudien – gelangt Sorensen ebenfalls zu dem Ergebnis, Gedankenexperimente zielten auf die Widerlegung modalen Thesen (Sorensen 1992, S. 135–160).

phischer Gedankenexperimente dar. Aber wie gesagt: Nichts von dem, was folgt, hängt an dieser speziellen Strukturanalyse. Dasselbe ließe sich auch im Rahmen anderer Auffassungen von Gedankenexperimenten vorbringen und begründen – nur eben häufig etwas umständlicher.

(Ebensowenig hängt etwas an der im Folgenden häufig verwendeten Redeweise von *Intuitionen*. Wenn ich schreibe, dass wir die *Intuition* haben, ein Szenario sei so-und-so zu beschreiben, meine ich schlicht, dass wir es plausibel finden, das Szenario auf genau diese Weise zu beschreiben, oder, anders gesagt, dass wir als kompetente Sprecher den Eindruck haben, unsere Begriffe müssten hier auf genau diese Weise verwendet werden. Die Rede von Intuitionen besagt hier *nicht*, dass wir über eine ominöse, von sinnlicher Wahrnehmung und begrifflicher Reflexion verschiedene Erkenntnisquelle namens *Intuitionsvermögen* verfügen, die uns vermittels Gedankenexperimenten Einblicke in tiefe metaphysische Zusammenhänge ermöglicht.⁴)

Wie also können wir mit Williams' Paradox umgehen? Zwölf Antworten auf diese Frage, die meisten davon der Literatur entnommen,⁵ werde ich im Folgenden exemplarisch präsentieren. Alle zwölf Antworten halte ich für sinnvoll (auch wenn ich eine favorisiere), alle haben in meinen Augen durchaus Überzeugungspotential. Und ohne Frage lassen sich weitere sinnvolle Varianten dieser Antworten finden – wenngleich meiner Auffassung nach alle zusätzlichen sinnvollen Antworten ebendies sein müssen: bloße Varianten einer meiner zwölf Antworten.

Dass es so viele gute Antworten auf Williams' Paradox gibt, spricht, so könnte man meinen, gegen die Paradoxität des Paradoxes. Das Gegenteil ist der Fall. Gibt es genau eine gute Antwort auf ein Paradox, so ist klar, wie sich das Paradox aus der Welt schaffen lassen muss, und wir können es beruhigt als gelöstes Problem betrachten. Bieten sich viele grundverschiedene und miteinander kaum kompatible Antworten an (zur Vereinbarkeit der verschiedenen Antworten werde ich gegen Ende des Aufsatzes noch mehr sagen), ist die Wahrscheinlichkeit, dass uns eine bestimmte Antwort in die Irre leitet, recht groß. Wir müssten zur Auflösung des Paradoxes dann nicht nur erklären, warum eine spezielle Antwort einleuchtend ist, sondern auch, warum andere es wider allem Anschein nicht sind. Auch wenn ich eine ganz bestimmte der zwölf Antworten für die richtige halte: Die Tatsache, dass man Williams' Paradox auch leicht ganz anders deuten kann, zeigt, dass Williams hier auf einen interessanten Punkt gestoßen ist, der ein gründliches Nachdenken auch heute noch verdient.

⁴Dies glaubt etwa Brown; siehe Brown 1991.

⁵Ausnahmen sind vor allem die Antworten eins und zwölf; aber auch wesentliche Teile der Antworten zwei, acht und zehn stammen von mir.

Erste Antwort

Prinzipiell, so habe ich gesagt, kann man ein philosophisches Gedankenexperiment auf zwei Weisen kritisieren: Man kann argumentieren, dass das zugrunde liegende imaginierte Szenario nicht begrifflich möglich ist, und man kann argumentieren, dass das imaginierte Szenario auf eine unangemessene Weise beschrieben oder bewertet wird. Beide Williams'schen Gedankenexperimente beruhen auf demselben imaginierten Szenario, das sich wie folgt darstellen lässt:

KÖRPERTAUSCH.⁶ Sämtliche in As Gehirn gespeicherten Informationen werden auf Bs Gehirn überspielt und umgekehrt; die resultierenden Personen haben den Körper von A bzw. B und die Erinnerungen, Überzeugungen, Charaktermerkmale etc. von B bzw. A. Nach diesem Vorgang wird die Person, die As Körper besitzt, gefoltert und die, die Bs Körper besitzt, nicht.

Wenn wir zeigen können, dass KÖRPERTAUSCH ein begrifflich unmögliches Szenario ist, brauchen wir uns um angemessene und unangemessene Beschreibungen sowie daraus entstehende Widersprüche gar nicht mehr zu kümmern, sondern können schlicht konstatieren, dass Williams' Gedankenexperimente nicht aussagekräftig sind und somit sein Paradox auf falschen Voraussetzungen beruht.

Um zu sehen, dass KÖRPERTAUSCH tatsächlich nicht begrifflich möglich ist, vergleiche man dieses Szenario mit einem Gehirntausch-Szenario, bei dem As Gehirn in Bs Körper und Bs Gehirn in As Körper transplantiert wird. Auch hier haben die resultierenden Personen die (überwiegende Mehrheit der) physischen Eigenschaften von A bzw. B und die psychischen von B bzw. A. Der Unterschied zu KÖRPERTAUSCH betrifft lediglich die Quantität physischer Kontinuität: Im einen Szenario wird das ganze Gehirn ausgetauscht, im anderen nur so viel davon, wie für den Austausch der auf bestimmten physischen Eigenschaften supervenierenden psychischen Eigenschaften erforderlich ist.⁷ Wegen der Supervenienz psychischer Eigenschaften auf physischen

⁶ Die Bezeichnung „Körpertausch“ ist tendenziös, insofern sie nahelegt, dass der Geist das Wesentliche ist und der Körper das, was getauscht werden kann. Präziser und neutraler wäre es, von einem Körper-oder-Geist-Tausch zu reden. Ähnliches gilt für „Gehirntausch“. Der Einfachheit halber verwende ich hier und im Folgenden die ungenaueren, aber weniger umständlichen Bezeichnungen.

⁷ Die Annahme der Supervenienz psychischer Eigenschaften auf physischen ist im Rahmen von KÖRPERTAUSCH unproblematisch, denn im Falle einer Nichtsupervenienz wäre völlig unklar, wie sich das Überspielen von geistigen Informationen von einem Gehirn auf ein anderes prinzipiell bewerkstelligen lassen könnte. Die psychischen Eigenschaften von A- und B-Körper-Person müssen sich ja in irgendeiner Weise kausal aus denen von B bzw. A ergeben, und um diesen kausalen Nexus gewährleisten zu können, bedarf es eines

sind nun aber die in einem Gehirn gespeicherten Informationen denkbar eng mit dem Aufbau des Gehirns verbunden, und weil die Erzeugung, Speicherung und Weiterverarbeitung dieser Informationen die entscheidende Aufgabe des Gehirns ist, hätte eine vollständige Änderung dieser Informationen derart tiefgreifende Auswirkungen auf die physische Beschaffenheit des Gehirns, dass man nicht mehr sinnvoll von demselben Gehirn sprechen könnte. Anders gesagt: Es ist es mit keinen plausiblen Kriterien von Gehirnidentität zu vereinbaren, dass sämtliche psychischen Eigenschaften einschließlich deren physischer Grundlage sich ändern können *und das Gehirn dasselbe bleibt*. Vielmehr verhält es sich mit Gehirnen wie mit Büchern: Wenn ich in einem Exemplar der *Kritik der reinen Vernunft* alle Seiten ausradiere (und zwar, im Unterschied zu vielen Palimpsesten, so, dass der Inhalt unwiederbringlich gelöscht ist) und den Text der *Phänomenologie des Geistes* darauf schreibe, so wird man kaum behaupten wollen, der Autor des resultierenden Buches sei Kant. Genausowenig kann man behaupten, ein Gehirn, das allein die geistigen Eigenschaften von *B* enthält, sei das Gehirn von *A*. Eine wechselseitige Übertragung psychischer Eigenschaften samt den zugrunde liegenden physischen ist daher *de facto* ein Gehirntausch.

(Einwand: Eine bessere Analogie als die zwischen Gehirn und Buch ist die zwischen Gehirn und Computer, wobei dem Gehirn die Rolle der Hardware zukommt und den darin gespeicherten Informationen die der Software. Dass wir nun unter der Identität eines Computers eindeutig eher die Identität der Hardware als die der Software verstehen würden, legt nahe, dass die Gehirn-Identität eben nicht von den im Gehirn gespeicherten Informationen abhängt. Erwiderung: Selbst wenn man zugesteht, dass die Computer-Analogie sinnvoll ist, sehen wir Computer-Identität nur deshalb als Hardware-Identität an, weil wir bei Computern an stupide Rechenmaschinen und nicht an Künstliche Intelligenzen denken. Vergewegen wir uns einen wirklich intelligenten Computer, wird schnell klar, dass bei ihm die Software fürs Fortbestehen entscheidend ist. Man denke an HAL, den Computer aus *2001 – Odyssee im Weltraum*: Würde er angesichts drohender Abschaltung oder Löschung den Gedanken, dass seine Hardware erhalten bleibt, ernsthaft tröstlich finden?)

Williams aber ist darauf angewiesen, dass seinen Gedankenexperimenten kein Gehirntausch-Szenario zugrunde liegt, sondern ein Szenario wie KÖRPERTAUSCH, weil andernfalls sein zweites Gedankenexperiment viel von seiner ursprünglichen Plausibilität einbüßen würde: Was würden wir sagen, wenn unser Gehirn unserem Körper entnommen, das Gehirn einer anderen Person in unseren Körper eingepflanzt und unser Körper anschließend gefoltert werden würde? Eine erschreckende Aussicht, fraglos, aber nicht wegen materiellen Ankers.

der Folter, sondern wegen des Schicksal unseres Gehirns (siehe Perry 2002, S. 116). Williams' Paradox entfaltet nur dann die suggestive Kraft, die es auf den ersten Blick hat, wenn es nicht auf einem Gehirn-, sondern nur auf einem Körpertausch beruht. Und ein solcher ist bei Licht besehen unmöglich.

Zweite Antwort

Im ersten Gedankenexperiment bilden wir die Intuition aus, dass, wenn KÖRPERTAUSCH realisiert wäre, *A* (im Gegensatz zu *B*) die Folter der *A*-Körper-Person nicht als schlimm für sich empfinden würde. Das Szenario wird uns dabei auf eine bestimmte Weise geschildert, und bezüglich dieser Darstellung erscheint uns *As* Empfindung höchst nachvollziehbar und plausibel. Im zweiten Gedankenexperiment bilden wir die Intuition aus, dass, wenn KÖRPERTAUSCH realisiert wäre, *A* die Folter der *A*-Körper-Person als schlimm für sich empfinden würde. Das Szenario wird uns dabei auf eine bestimmte, von der ersten Darstellung verschiedene Weise geschildert, und auch bezüglich dieser zweiten Darstellung erscheint uns *As* Empfindung höchst nachvollziehbar und plausibel.

Offensichtlich stehen beide Intuitionen im Widerspruch zueinander. Das bedeutet nicht nur, dass mindestens eine von ihnen uns in die Irre führt. Es bedeutet, dass *beide* uns bei der Suche nach einem geeigneten Begriff von personaler Identität nicht weiterhelfen können (und damit beide Gedankenexperimente nicht aussagekräftig sind). Denn jede der beiden Intuitionen ist weit davon entfernt, *darstellungsinvariant* zu sein: Eine abgewandelte Darstellung desselben Szenarios führt für jede der beiden Intuitionen dazu, dass sie nicht nur verschwindet (was allein schon für das Nichtvorliegen von Darstellungsinvarianz ausreichen würde), sondern sogar in die gegenteilige Intuition umschlägt. Dabei sei unter einer abgewandelten Darstellung eine Schilderung verstanden, die die durch die ursprüngliche Darstellung gegebenen Fakten übernimmt (und dabei in den wesentlichen Aspekten nicht weniger detailliert sein darf als jene) und das Szenario nur in anderen Worten, mit anderer Akzentuierung oder unter Hinzufügung unwesentlicher Details beschreibt.⁸ Eine Intuition, die nicht darstellungsinvariant ist, kann aber kein verlässlicher Beleg sein für die Wahrheit einer Proposition wie der, dass *A* in KÖRPERTAUSCH die Folter der *A*-Körper-Person als schlimm für sich empfinden würde, weil sie aufgrund von Details zustande gekommen sein muss, die

⁸Auch die Perspektive, aus der ein Szenario geschildert wird, darf keine Rolle spielen, wenn Intuitionen darstellungsinvariant sein sollen. Wenn etwa die Tatsache, dass Williams sein erstes Gedankenexperiment in der dritten Person präsentiert und sein zweites in der ersten, entscheidend ist für das Herausbilden der entsprechenden Intuitionen, spricht das gegen deren Zuverlässigkeit.

für die Wahrheit der Proposition nebensächlich sind. (Dies folgt direkt aus der Definition von Darstellungsinvarianz.) Also sind die Intuitionen, die wir in Williams' Gedankenexperimenten bezüglich KÖRPERTAUSCH ausbilden, beide unzuverlässig.

Dass die Intuitionen in Williams' Paradox nicht darstellungsinvariant sind, ist letztlich nicht allzu überraschend: Laut Shafir 1998 zeigen psychologische Studien nämlich, dass häufig dasselbe Szenario in Einzelevaluationen (wenn zum Beispiel nur die Sicht einer Person gezeigt wird) anders bewertet wird als in vergleichenden Evaluationen (wenn zum Beispiel die Sichtweisen verschiedener Personen gleichberechtigt nebeneinander stehen). Wir neigen somit ganz allgemein dazu, Situationen wie die von Williams geschilderten in Abhängigkeit der Art ihrer Darstellung zu bewerten; konsistente Intuitionen sind in derartigen Fällen also gar nicht erst zu erwarten. Williams' Paradox zeigt demnach lediglich eine Schwachstelle im menschlichen Denken auf.

(Shafir fügt hinzu, dass man die relative Wichtigkeit eines Ausgangs durch vergleichende Konstellationen besser erkennen kann als durch solche, in denen die Lage durch die Augen nur eines Beteiligten geschildert wird, weil Aspekte, die für die eigene Meinungsbildung und das eigene Handeln wichtig sind, in Fällen der zweiten Art mit größerer Wahrscheinlichkeit übersehen werden als in Fällen ersterer Art. Demnach wäre die im ersten Gedankenexperiment ausgebildete Intuition zuverlässiger. Jedoch kann dieses Prinzip hier nicht angewendet werden, da dadurch, dass dem Leser beide Gedankenexperimente nacheinander vor Augen geführt wurden, gewährleistet ist, dass ihm keine relevanten Aspekte entgehen.)

Dritte Antwort

Gedankenexperimente haben, wie ich vorausgesetzt habe, die Funktion, Notwendigkeits- oder Möglichkeitsbehauptungen zu widerlegen. Solche Behauptungen begegnen uns häufig in der auf den ersten Blick nichtmodalen Form *Dass dies-und-das der Fall ist, ist eine/keine notwendige/hinreichende Bedingung dafür, dass etwas ein/kein F ist.* Diese Form ist dann als striktes Konditional zu analysieren, also als eines mit der logischen Form $\Box(A \rightarrow B)$, denn die relevante Bedingung ist so aufzufassen, dass für ihr Erfüllt- oder Nichterfülltsein nicht nur tatsächlich bestehende Sachverhalte relevant sind, sondern auch bloß mögliche. Im Zusammenhang mit personaler Identität zum Beispiel interessieren wir uns für Bedingungen, die ganz allgemein, das heißt *unter allen möglichen Umständen* (und nicht bloß unter den zufälligerweise gerade realisierten), notwendig oder hinreichend für das Bestehen personaler Identität sind. So wollen wir etwa wissen, ob physische Kontinuität – eine Relation, die unseres Wissens in allen tatsächlichen Fällen personaler Identi-

tät besteht – auch notwendigerweise jedes Vorkommnis personaler Identität begleitet, und ob umgekehrt jedes Vorkommnis von physischer Kontinuität bei Personen zwangsläufig identitätsstiftend ist. Denn nur wenn beides der Fall ist, nur wenn physische Kontinuität *mit Notwendigkeit* eine notwendige und hinreichende Bedingung personaler Identität ist, ist es uns gelungen, eine wesentliche Eigenschaft der Identitätsrelation von Personen – und damit von Personen selbst – zu identifizieren. Der erkenntnistheoretische Sinn von Williams’ Gedankenexperimenten besteht demnach darin, zwei Relationen voneinander zu entkoppeln, die in der Realität normalerweise immer zusammen vorliegen, nämlich physische und psychische Kontinuität von Personen, um zu sehen, welche davon die für personale Identität wesentliche ist.⁹

Williams’ erstes Gedankenexperiment zeigt nun, dass physische Kontinuität weder notwendig noch hinreichend für personale Identität ist, während das zweite beweist, dass dasselbe für psychische Kontinuität gilt. Damit sind die beiden einzigen aussichtsreichen Kandidaten für solche Bedingungen aus dem Rennen, und es stellt sich die Frage, wie sinnvoll die Suche nach notwendigen und hinreichenden Bedingungen für personale Identität überhaupt ist.

Nehmen wir zum Beispiel an, alle uns bekannten und unter einen bestimmten Begriff B fallenden tatsächlichen oder bloß möglichen Entitäten haben die Eigenschaften F , G und H . Wenn wir nun, gleich ob in realen Szenarien oder imaginierten, auf eine Entität a treffen, die klarerweise unter B fällt, aber nur die Eigenschaften F und G hat, können wir einerseits vermuten, dass nur F und G wesentlich für die Zugehörigkeit zu B sind, oder andererseits, dass a nur deshalb unter B fällt, weil es den paradigmatischen Vertretern der durch B beschriebenen Art ausreichend ähnlich ist. Diese zweite Lesart ist oftmals spätestens dann angeraten, wenn wir Exemplaren jener Art begegnen sollten, die F oder G (oder beides) nicht aufweisen. Und nach dieser Lesart ist dann zum Beispiel ein Gedankenexperiment, demzufolge auch Entitäten unter B fallen, die nur H haben, deutlich weniger aufregend als nach der ersten, da es lediglich zeigt, wie stark sich eine Entität von den

⁹Man kann auf sehr allgemeiner Ebene gegen Gedankenexperimente, denen wirklichkeitsferne imaginierte Szenarien zugrunde liegen, einwenden, dass ein sinnvoller Begriff generell deutlich stärker an unserer Redeweise über alltägliche Fälle orientiert sein müsse als an unserer Redeweise über Fälle, die für hochphilosophische Zwecke konstruiert wurden, und dass daher unseren Intuitionen bezüglich der Verwendung von Begriffen in solchen Gedankenexperimenten keine oder nur eine sehr geringe Aussagekraft zukomme (siehe etwa Quante 2007, S. 101f.). Ich bespreche diese Art der Kritik an Gedankenexperimenten in Weber *im Erscheinen*, werde hier aber nicht weiter darauf eingehen, da die Annahme, philosophische Gedankenexperimente mit wirklichkeitsfernen Szenarien könnten prinzipiell epistemisch signifikant sein, unabdingbare Voraussetzung für Williams’ Paradox ist.

typischen Exemplaren ihrer Art unterscheiden kann.¹⁰

Wenn der Begriff der personalen Identität keiner ist, der sich durch notwendige und hinreichende Bedingungen explizieren lässt, sondern allenfalls durch Ähnlichkeiten zu paradigmatischen Fällen personaler Identität, wie wir sie aus unserem Alltag kennen, dann hat das zwei Konsequenzen. Erstens helfen uns in diesem Fall Gedankenexperimente bei der Untersuchung jener Relation nur sehr bedingt weiter, da sie letztlich bloß Beispiele von Einzelfällen liefern können, die sich nicht zu einheitlichen Auffassungen verallgemeinern lassen. Sie können sogar dadurch, dass man immer versucht ist, ihre Rolle zu überschätzen und sie nach der ersten Lesart zu deuten, derzufolge sie das Wesentliche vom Unwesentlichen zu trennen vermögen, den Blick auf die tatsächlichen und sehr viel diffuseren Sachverhalte trüben.

Die zweite Konsequenz ist, dass Ungereimtheiten wie Williams' Paradox erklärt werden können. Denn wenn Gedankenexperimente zu einem bestimmten Begriff nur dazu dienen können, den Grad an zulässiger Unähnlichkeit zu paradigmatischen Fällen, die wir der Alltagserfahrung entnehmen, zu bestimmen, dann zeigen Williams' Gedankenexperimente bloß, wie liberal und kontextabhängig unser Begriff von personaler Identität ist. Sie zeigen, dass wir in einer Situation, deren Darstellung interpersonalen Relationen die psychischen Aspekte des Personseins in den Fokus rückt, auch solche Relationen, in denen physische Kontinuität keine Rolle spielt, als Personale-Identitäts-Relation akzeptieren, und in einer Situation, deren Darstellung interpersonalen Relationen auf die physischen Aspekte des Personseins konzentriert ist, sogar Relationen, für die psychische Kontinuität irrelevant ist. Dass es mit Fall (vi) eine Situation gibt, die wir bei abgewandelter Darstellung anders bewerten, ist mit dieser Erklärung kompatibel, da der angesprochene Grad an Unähnlichkeit immer nur im Abgleich mit einer gegebenen Darstellung beurteilt werden kann und somit möglicherweise bezüglich verschiedener Schilderungen desselben imaginierten Szenarios variiert. Auf den ersten Blick widersprüchliche Gedankenexperimente zum Begriff der personalen Identität sind also nicht allzu ernst zu nehmen, weil eben diese Widersprüchlichkeiten zeigen, dass jener Begriff sich nicht durch Angabe notwendiger und hinreichender Bedingungen explizieren lässt und Gedankenexperimente daher höchstens die Besonderheiten seiner Verwendung beleuchten können. Gedankenexperimente zur personalen Identität sind wesentlich weniger aus-

¹⁰Zu diesen beiden Lesarten von Gedankenexperimenten siehe auch Gendler 1998, S. 607f. Nach Cohnitz 2003, S. 175, kann man eine Prototypentheorie personaler Identität, wie sie Gendler propagiert, nur vermeiden, indem man bei Williams unzulässig weitreichende Voraussetzungen nachweist oder zentrale Intuitionen wegerklärt. Cohnitz dürfte dabei vorschweben, was ich in einer Variante der Antwort acht bzw. in den Antworten fünf und zehn ausführe.

sagekräftig, als gemeinhin angenommen.¹¹

Vierte Antwort

Williams zufolge bilden wir im ersten Gedankenexperiment die Intuition aus, dass *A* die Folter der *B*-Körper-Person fürchtet, aber *B* nicht. Dazu macht er gewisse Vorhersagen bezüglich dessen, was *A*- und *B*-Körper-Person angesichts ihres Schicksals denken. Diese Vorhersagen, so argumentiert Noonan, treffen nicht zu (Noonan 2003, S. 181–85). In dem Moment, in dem uns dies klar wird, teilen wir aber auch die gewünschte Intuition nicht mehr und messen dem ersten Gedankenexperiment keine Aussagekraft mehr bei.

Angenommen, *A* und *B* vertreten beide eine Theorie physischer Kontinuität, also eine, derzufolge eine Kontinuität des Körpers notwendig und hinreichend für personale Identität ist. Konsequenterweise wünscht dann *A*, dass die *A*-Körper-Person das Geld erhält, und *B*, dass es an die *B*-Körper-Person geht. Angenommen, *A*s Wunsch wird in die Tat umgesetzt. Nach dem Körpertausch wird sich laut Williams die *A*-Körper-Person daran erinnern, den Wunsch geäußert zu haben, dass ihr die Folter zuteil werde, und wird erleichtert sein, dass dieser Wunsch nicht erfüllt wird. Aber erinnert sie sich wirklich daran, jenen Wunsch geäußert zu haben, oder meint sie nur, sich daran zu erinnern? Immerhin denkt auch die *A*-Körper-Person, dass physische Kontinuität das entscheidende Kriterium personaler Identität ist. (Diese Meinung hat sie von *B* übernommen, deren Überzeugungen sie ja teilt.) Demnach ist sie, die ja den *A*-Körper besitzt, auch der Auffassung, sie sei mit *A* identisch. Dass sie *B*s geistige Eigenschaften geerbt hat, ändert daran nichts, denn als Vertreter einer Theorie physischer Kontinuität sieht sie das als sekundär an. Und da sie sich als mit *A* identisch begreift, hat sie ihrer Meinung nach auch nie den Wunsch geäußert, dass die *B*-Körper-Person das Geld erhalten soll. Die „Erinnerung“, die sie an diese Wunschäußerung hat, ist insofern illusionär, als dass das vermeintlich Erinnerte nichts Selbsterlebtes ist, sondern ein phänomenal von einer wirklichen Erinnerung nicht zu unterscheidender Eindruck von einem Erlebnis eines anderen.¹² Vielmehr war es nach Ansicht

¹¹Auch Johnston schließt aus der Widersprüchlichkeit des Begriffs der personalen Identität in Gedankenexperimenten darauf, dass dieser Begriff nur sehr bedingt mittels Gedankenexperimenten untersucht werden kann (Johnston 1987, S. 69–75). Soweit ich seine bisweilen etwas enigmatischen Ausführungen verstehe, sehe ich darin keinen wesentlichen inhaltlichen Unterschied zu dieser dritten Antwort.

¹²Diese Art illusionärer Erinnerung könnte man mit Shoemaker als Quasi-Erinnerung bezeichnen (siehe Shoemaker 1970, S. 271.) Shoemakers Auffassung nach treten Quasi-Erinnerungen zwar nur in Fällen von Bewusstseinsteilung auf; das erklärt sich jedoch daraus, dass er kein Anhänger von Theorien physischer Kontinuität ist, sondern von solchen psychischer Kontinuität.

der *A*-Körper-Person sie selbst, die den Wunsch gehabt hat, die *A*-Körper-Person möge das Geld erhalten, auch wenn sie nachvollziehbarerweise keine Erinnerung an das Haben dieses Wunsches hat. Analog liegt der Fall bei der *B*-Körper-Person, die jetzt gefoltert werden soll. Dass sie sich vermeintlich daran erinnert, bestimmt zu haben, dass die *B*-Körper-Person gefoltert werden soll, betrachtet sie als nicht weniger illusionär als die *A*-Körper-Person ihre Erinnerung an ihre Wahl. Ihr schlimmes Schicksal wird die *B*-Körper-Person daher nicht einer eigenen törichten Wahl ankreiden. Vielmehr wird sie damit hadern, dass nicht dem Wunsch der ihrer Meinung nach mit ihr identischen Person *B* entsprochen wurde. Dass sie keine Erinnerung daran hat, diesen Wunsch geäußert zu haben, hält sie für nebensächlich.¹³

Williams übersieht also, dass auch die *A*- und die *B*-Körper-Person nach bestem Wissen und Gewissen Körpertheoretiker sein können. Ihnen – und damit auch dem Leser – wird mitnichten eine eventuelle Unhaltbarkeit eines Körperkriteriums bewusst gemacht; es wird lediglich eine gewisse Merkwürdigkeit vor Augen geführt, lebhaft und nicht eingebildete „Erinnerungen“ an tatsächliche Ereignisse als illusionär hinzustellen. Diese Merkwürdigkeit ist es auch, die Unvoreingenommene dazu tendieren lässt, bezüglich KÖRPERTAUSCH Identifizierungen eher anhand geistiger Gemeinsamkeiten vorzunehmen. Doch solche vorthoretischen Intuitionen sind bestenfalls ein Argument unter vielen in der Debatte zur personalen Identität und keinesfalls ein unumstößlicher Beweis für die Richtigkeit von Theorien psychischer Kontinuität.¹⁴

Dieser vierten Antwort zufolge zeigt sich demnach durch eingehendere Reflexion, dass wir in Wirklichkeit gar nicht dazu neigen, KÖRPERTAUSCH auf die vom ersten Gedankenexperiment nahe gelegte Weise zu beschreiben. Sie ähnelt damit der, die Williams selbst auf der letzten Seite von *The Self and the Future* andenkt. (Er präsentiert sie nur sehr zögerlich, mit Sympathie, aber ohne sie sich wirklich zu eigen zu machen.) Wenn man das erste Gedankenexperiment nur leicht abändere, so Williams, etwa indem man zwei resultierende Personen mit *As* geistigen Eigenschaften und Erinnerungen erschaffe statt einer, wären unsere Intuitionen hinsichtlich personaler Identität ganz andere. Die relevanten Intuitionen im zweiten Gedankenexperiment da-

¹³Die Fälle, in denen eine der beiden ursprünglichen Personen *A* und *B* einer Theorie physischer Kontinuität anhängt und die andere einer Theorie psychischer Kontinuität, bringen keine interessanten neuen Aspekte mehr ins Spiel und können daher hier übergangen werden. Für eine Besprechung dieser Fälle siehe Noonan 2003, S. 183f.

¹⁴Johnston berichtet, seiner Erfahrung nach hänge die Beurteilung von Williams' erstem Gedankenexperiment durch Nichtphilosophen stark davon ab, ob es vor oder nach Fällen, die die Bedeutung physischer Kontinuität betonen, präsentiert werde (Johnston 1987, S. 81). Wenn dies stimmt, unterstreicht das Noonans Diagnose, dass dieses Gedankenexperiment nicht geeignet ist, jemanden, der keine Theorie psychischer Kontinuität vertritt, von einer solchen zu überzeugen.

gegen seien abänderungsbeständiger. Das lege nahe, dass sie zuverlässiger seien.

Fünfte Antwort

Der vierten Antwort, so lässt sich einwenden, liegt eine irrige Vorstellung der epistemischen Rolle von Intuition bezüglich imaginierter Szenarien zugrunde. Diese Intuitionen sind Daten, auf deren Basis eine Theorie errichtet wird, und nicht die Prüfsteine für die Konsistenz einer Theorie. Wenn *A* intuitiv dazu tendiert, die Folter eher der *A*-Körper-Person zu wünschen als der *B*-Körper-Person, ist dies ein gewichtiges Datum dafür, dass psychische Kontinuität für unser Überleben relevanter ist als physische; und wenn die *A*-Körper-Person intuitiv dazu tendiert, sich mit eher mit *B*, deren Erinnerungen sie zu teilen scheint, zu identifizieren als mit *A*, ist dies ebenfalls ein Datum, das auf die vorrangige Rolle psychischer Kontinuität hindeutet. Diese Daten verlieren nichts an epistemischer Signifikanz, wenn *A* und *B* aus anderen Gründen Theorien physischer Kontinuität favorisieren. Dann nehmen sie die entsprechenden Identifizierungen im ersten Gedankenexperiment zwar, wie Noonan richtig ausführt, anders vor; das jedoch bedeutet nicht, dass wir die fraglichen Intuitionen gar nicht haben, sondern bloß, dass Vertreter von Theorien physischer Kontinuität durch ihre theoretischen Überzeugungen gezwungen sind, Identifizierungen *entgegen unseren Intuitionen* vorzunehmen. (Ähnliches lässt sich gegen Williams' Überlegung einwenden: Dass ein leicht abgeänderter Versuchsaufbau deutlich veränderte Daten produziert, enthebt uns nicht der Verpflichtung, die Daten des ursprünglichen Experiments in unserer Theoriebildung zu berücksichtigen.)¹⁵

In dieser Erwiderung schimmert allerdings ein pragmatisches Verhältnis zu Intuitionen durch. Intuitionen werden als Daten verstanden, die interpretiert und erklärt werden müssen. In diesen Erklärungen wird jedoch nicht vorausgesetzt, dass die Inhalte dieser Intuitionen wahr sind; vielmehr muss sich das Zustandekommen der Intuitionen plausibel explizieren lassen. Wenn wir beispielsweise, anders als in der vierten Antwort verfochten, tatsächlich die Intuition haben, dass im ersten Gedankenexperiment *A* die Folter der *B*-Körper-Person fürchtet, bedeutet das längst noch nicht, dass wir *A* und die *B*-Körper-Person für identisch halten müssen. Denn es könnte ja Gründe geben, warum wir jene Intuition auch dann haben, wenn sie uns in die Irre leitet. Solche Gründe legt beispielsweise Johnston dar (Johnston 1987, S. 80–83).

Weil in unseren Alltagsszenarien physische und psychische Kontinuität so

¹⁵Siehe Beck 1998 für eine ähnliche Kritik an Noonans und Williams' Einwänden.

gut wie immer miteinander einhergehen, sind wir nach Johnston sowohl gewohnt, Identifizierungen nach Körper-Kriterien vorzunehmen, als auch, Identifizierungen nach Geist-Kriterien vorzunehmen. Es ist demnach wenig verwunderlich, wenn wir uns in einem konkreten Fall, in dem diese Kriterien zu unterschiedlichen Identifizierungen führen und wir deshalb keine sichere Handhabe haben, relativ willkürlich für eines entscheiden. Johnston spricht in diesem Zusammenhang von einer „nachvollziehbaren Überverallgemeinerung“ (Johnston 1987, S. 80f., meine Übersetzung). Die Frage ist nur, warum wir im ersten Gedankenexperiment intuitiv eher das psychische Kriterium überverallgemeinern und nicht das physische. Ein Grund ist, dass wir aus praktischen Interessen die Fortexistenz einer Person gewährleisten wissen wollen, die lückenlos die komplexen und mannigfaltigen sozialen Rollen ausfüllen kann, die wir im Gefüge unserer Lebenswelt tagtäglich einnehmen. Um unsere Liebes-, Freundschafts- und Geschäftsbeziehungen zu pflegen, ist primär psychische Kontinuität vonnöten; eine bloß körperlich mit uns identische Person müsste in dieses gesellschaftliche Geflecht erst mühsam wieder hineinwachsen. Einen anderen Grund stellt die tradierte Seelenkonzeption dar, die immer noch zentral ist für das oft nur rudimentär reflektierte Alltagsverständnis von personaler Identität. Die Vertrautheit mit dieser Vorstellung vom Ich als nichtmaterieller Seele im Verbund mit der deutlich größeren Ähnlichkeit dieser Vorstellung zu psychischen als zu physischen Identitätskriterien verstärkt unsere Neigung, in vielen Zweifelsfällen eher auf psychische Kontinuität zu setzen.

All dies besagt noch nicht, dass es falsch ist, *A* mit der *B*-Körper-Person zu identifizieren. Aber wenn man mit Johnston oder Williams oder auch mit Animalisten wie Olson oder van Inwagen¹⁶ aufgrund anderweitiger theoretischer Überlegungen der Meinung ist, personale Identität bestehe in einer wie auch immer gearteten Identität oder Kontinuität des Körpers, kann man die gegenteilige Intuition, die wir in Williams' erstem Gedankenexperiment ausbilden, auf diese Weise erklären und damit aufzeigen, warum sie den jeweils eigenen Standpunkt weder widerlegt noch nennenswert schwächt.

Sechste und siebte Antwort

Die Strategien aus Antwort vier und fünf – erklären, warum wir die behaupteten Intuitionen in Wirklichkeit gar nicht haben bzw. warum wir diesen Intuitionen widersprechenden Theorien anhängen sollten – kann man auch gegen das zweite Gedankenexperiment fahren. Dieses Gedankenexperiment

¹⁶Animalisten sind der Auffassung, personale Identität bestehe in der Identität des lebensfähigen Organismus. Olson und van Inwagen entwickeln ihre jeweiligen animalistischen Theorien am ausführlichsten in Olson 1997 bzw. van Inwagen 1990.

ist allerdings etwas komplexer gebaut: Bezüglich (i) haben wir, dem Gedankenexperiment zufolge, die Intuition, dass die Folter schlimm für uns sei; bezüglich (m) haben wir, dem Gedankenexperiment zufolge, die Intuition, dass, wenn wir in (m – 1) die Folter aus rein egoistischen Gründen zurecht fürchten, wir auch in (m) Grund dazu haben ($m \in \{ii, iii, iv, v\}$); bezüglich (vi) haben wir, dem Gedankenexperiment zufolge, die Intuition, dass das, was mit einer dritten Person geschieht, keinen Unterschied machen kann in Hinblick darauf, ob wir die Folter zurecht fürchten. Die Aussagekraft all dieser Intuitionen kann man anzweifeln. Gute Gründe dafür präsentiere ich in den Antworten sechs bis elf.

Zäumen wir das Pferd von hinten auf. Laut Williams sind die Fälle (v) und (vi) gleich zu behandeln, weil die zu folternde Person in (v) sich weder qualitativ noch ihrem kausalen Ursprung nach von der in (vi) unterscheidet und daher nicht die eine mit uns identisch sein könnte und die andere nicht. (Zur Erinnerung: In (v) werden die geistigen Eigenschaften einer anderen Person auf mein Gehirn überspielt, ohne die Gehirnzustände dieser an deren Person zu verändern; in (vi) werden die geistigen Eigenschaften einer anderen Person auf mein Gehirn überspielt und meine geistigen Eigenschaften auf das jener anderen Person.) Meine Furcht, gefoltert zu werden, kann nämlich Williams zufolge ihre eventuelle Berechtigung nicht durch etwas verlieren, was einer ganz anderen Person an einem ganz anderen Ort widerfährt. Wenn meine Furcht in (v) berechtigt ist, ist sie es daher in (vi) auch. Und berechtigt kann sie nur sein, wenn es tatsächlich ich bin, der in (vi) gefoltert wird. Folglich kann etwas, das mit irgendjemand Anderem, an einem anderen Ort, geschieht, keine Auswirkungen haben auf etwas so Essentielles wie personale Identität; personale Identität darf, mit anderen Worten, nur von den intrinsischen Eigenschaften der beiden Personen, deren Identität infrage steht, und ihrer Beziehung zueinander abhängen. Könnte es eine dritte Person geben, von deren Eigenschaften jene Identität abhängt, besteht die Identität nicht, unabhängig davon, ob diese dritte Person tatsächlich existiert.¹⁷

Doch so einleuchtend dieses *Intrinsitätsprinzip* auch daherkommt, so schlecht passt es zu zu vielen Theorien psychischer Kontinuität, wobei ich unter Theorien psychischer Kontinuität allgemein solche verstehe, denen zufolge eine

¹⁷Dieses Prinzip, das Williams selbst nur in früheren Artikeln explizit formuliert (Williams 1956 und Williams 1960; siehe auch Nozick 1981, S. 31f.), erscheint etwas stärker als nötig, da wir für den Schritt von (v) zu (vi) im zweiten Gedankenexperiment nur benötigen, dass die Identität von keiner Tatsache über eine tatsächlich existierende dritte Person abhängt. Dass die Identität darüber hinaus auch von keiner Tatsache über eine *bloß mögliche* dritte Person abhängt, wird allerdings durch die Intuition nahegelegt, dass die Gerechtigkeit von so etwas wie Furcht vor Folter nur von den intrinsischen Eigenschaften der beteiligten Personen und ihrer Beziehung zueinander abhängt.

Kontinuität personenkonstituierender geistiger Eigenschaften notwendig und (außer in Fällen von Bewusstseinsverzweigung) auch hinreichend für personale Identität ist.¹⁸ Denn gerade in Gedankenexperimenten, die solche Theorien nahelegen, ist der explanative Mehrwert eng an die Aufgabe des Intransitivitätsprinzips geknüpft. Man betrachte etwa einen gängigen Gehirnteilungsfall: Mein Gehirn wird geteilt und jede Hälfte in einen anderen enthirnten Körper verpflanzt; da, wie angenommen wird, auch eine Gehirnhälfte ausreicht, um psychische Kontinuität zu gewährleisten, erben beide resultierenden Personen – wenn sie denn die Prozedur überleben – all meine geistigen Eigenschaften. Da das Szenario vollkommen symmetrisch ist, wäre es pure Willkür, eine der beiden resultierenden Personen als mit mir identisch anzusehen und die andere nicht. Weil aufgrund der Transitivität der Identität schlecht beide mit mir identisch sein können, ist es somit keine von beiden. Wenn allerdings nur eine der beiden Personen mit halbem Gehirn die Prozedur übersteht, liegt es nahe, diese wegen ihrer vollkommenen psychischen Kontinuität zu mir als mit mir identisch anzusehen. Insgesamt hängt dann allerdings die Identität einer aus der Gehirnteilung resultierenden Person mit mir davon ab, ob die Person, die die andere Gehirnhälfte erhalten hat, die Gehirnteilung überlebt hat – personale Identität wäre also keine intrinsische Beziehung. Aus diesem Grund stellen Gedankenexperimente wie dieses für Vertreter von Theorien psychischer Kontinuität eine starke Motivation dar, das Intransitivitätsprinzip rundweg abzulehnen. Die Intuitionen, dass in (i)–(v) die Schmerzen der Person, die unseren Körper besitzt, schlimm für uns sind, können sie dann immer noch getrost teilen und trotzdem jenen Widerspruch vermeiden, der Williams' Gedankenexperimente so paradox erscheinen lässt.

Ich möchte auf die zwei Weisen, das Intransitivitätsprinzip zurückzuweisen, etwas ausführlicher eingehen und sie am Beispiel von Robert Nozicks *Theorie des ähnlichsten Fortsetzers* [Closest Continuer Theory] erläutern. Nach Nozick besagt die Theorie des ähnlichsten Fortsetzers, „dass y zu t_2 nur dann dieselbe Person ist wie x zu t_1 , wenn erstens die Eigenschaften von y zu t_2 herrühren von, entwachsen aus, kausal abhängen von den Eigenschaften von x zu t_1 und es zweitens zu t_2 kein z gibt, das in einer engeren (oder genauso engen) Beziehung zu x zu t_1 steht als (bzw. wie) y zu t_2 “ (Nozick 1981, S. 36f., meine Übersetzung). Solch ein y ist der *ähnlichste Fortsetzer* [closest continuer] von x . Der ähnlichste Fortsetzer einer Person zu sein, ist nach Nozicks Theorie eine notwendige Bedingung für die Identität mit ihr. Wie genau dabei „in einer engeren Beziehung stehen“ zu verstehen ist, das heißt,

¹⁸Gut mit dem Intransitivitätsprinzip zu vereinbaren sind Theorien psychischer Kontinuität, die die Existenz zeitlicher Teile von Personen annehmen. Solche Theorien werden zum Beispiel in Lewis 1983, Perry 1972 und Noonan 2003, S. 228, entwickelt.

in wie hohem Maße physische und psychische Kontinuität jeweils relevant sein sollen, schreibt Nozick dabei, um größtmögliche Allgemeinheit bemüht, bewusst nicht vor. Seine Theorie ist daher eigentlich ein Theorienschema, das erst durch eine spezielle Bestimmung dessen, worin die Enge einer Beziehung zweier Personen zueinander besteht, zu einer waschechten Theorie wird.

Wenn man beispielsweise psychische Kontinuität als deutlich wichtiger erachtet als physische, diese allein aber im Falle des Fehlens von auch nur ansatzweise psychisch kontinuierlichen Fortsetzern auch noch als hinreichend für personale Identität ansieht, erhält man eine Theorie, die kein Problem mit den Fällen (i)–(vi) hat. Denn in (i)–(v) ist unser ähnlichster Fortsetzer dann die Person, die unseren Körper erbt, weil nach Eintritt des Gedächtnisverlusts niemand auch nur im Geringsten psychische Kontinuität zu uns aufweist und jene Person mit uns ja immerhin den Körper teilt. In (vi) hingegen gibt es eine Person, die, obwohl sie nicht unseren Körper hat, wegen der psychischen Kontinuität in einer wesentlich engeren Beziehung zu uns steht als die unseren Körper besitzende Person und die deshalb unser ähnlichster Fortsetzer ist. Während nach dem bisher Gesagten in (i)–(v) die zukünftige Person, die unseren Körper hat, durchaus mit uns identisch sein kann, ist dies in (vi) ausgeschlossen.

Doch wie gut ist Nozicks Theorie des ähnlichsten Fortsetzers selbst begründet? Wodurch ist sie motiviert? Nun, immerhin liefert sie einleuchtende Ergebnisse bei kritischen Gedankenexperimenten. Ihre Aussagen zum Beispiel zu Gehirnteilungsszenarien sind ebenso einfach wie überzeugend. Und nicht zuletzt spricht für sie die Tatsache, dass sie Williams' Paradox plausibel auflöst. Man kann daher einfach deshalb das Intrinsitätsprinzip ablehnen und sich eine Theorie wie Nozicks zu eigen machen, weil sie gut funktioniert. Dies ist die konventionalistische Variante: Weil unsere Intuitionen in Gedankenexperimenten zur personalen Identität widersprüchlich sind, wählen wir uns die Theorie, die diese Widersprüche am besten umgeht, und zahlen den erforderlichen Preis. Williams lehnt diese Position ab mit der Begründung, konventionale Entscheidbarkeit sei absurd, weil es doch keine Frage von Konventionen sein kann, ob wir in (i)–(vi) Grund haben, die Folter zu fürchten (Williams 1970, S. 89). Richtig; aber auch das ist Teil des Preises: Allgemein kann es angesichts inkompatibler Intuitionen passieren, dass wir der besten Theorie zufolge auch tief verwurzelte Intuitionen als falsch ansehen sollten.¹⁹ Dies ist hier der Fall mit der Intuition, personale Identität könne nicht auf bloßen Konventionen beruhen.

¹⁹Nida-Rümelin zufolge sind Intuitionen, die implizieren, dass personale Identität keine Frage von Konventionen ist, so tief in uns verwurzelt, dass wir sie gar nicht als falsch ansehen *können*. Sie leitet hieraus einen Anti-Reduktionismus bezüglich personaler Identität ab (siehe Nida-Rümelin 2006).

Konventionalistische Theorien setzen voraus, dass es keine tiefen Wahrheiten gibt, die wir erkennen oder verkennen können, und dass wir deshalb nichts gravierend falsch machen können, wenn wir uns (in gewissem Rahmen) *irgendwie* entscheiden – auch wenn uns das anders erscheinen mag. Eine Konvention darüber, wann Personen identisch sind, kann jenen Theorien zufolge also nie einen tiefen metaphysischen Irrtum darstellen, sie kann nur besser oder schlechter sein (und die Theorie des ähnlichsten Fortsetzers gehört sicherlich zu den überzeugenderen Varianten).

Eine solche konventionalistische Erwiderung stellt eine gute und plausible Antwort auf Williams' Paradox dar. Es ist allerdings nicht Nozicks Antwort. Seiner Auffassung nach fängt nämlich die Theorie des ähnlichsten Fortsetzers die Intuitionen, die wir bezüglich unseres Wohlergehens haben, tatsächlich exakt ein – dies ist die intuitionistische Variante der Keine-Intrinsität-Antwort.

Nozick zufolge orientieren sich unsere Intuitionen darüber, um welche zu einem zukünftigen Zeitpunkt existierende Person wir uns egoistischerweise ganz besonders sorgen, zwar am Ideal simpler transtemporaler Identität, bilden es aber nicht eins zu eins ab. So projiziert sich unsere Sorge in Szenarien, in denen mehr als einer unserer Fortsetzer genügend Ähnlichkeit zu uns aufweist, um sinnvollerweise als mit uns identisch gelten zu können, wir aber keinen ähnlichsten Fortsetzer haben, laut Nozick auf all diese hinreichend ähnlichen Fortsetzer (Nozick 1981, S. 62–70). Da auch in diesen Szenarien zu zukünftigen Zeitpunkten Personen existieren, auf deren Wohlergehen es uns ganz besonders ankommt, haben wir zudem keinen triftigen Grund, die Realisierung solcher Szenarien zu fürchten. Mit anderen Worten: Wir sorgen uns sehr um unseren ähnlichsten Fortsetzer, aber wir sorgen uns nicht so sehr darum, dass ein ähnlichster Fortsetzer existiert. Insbesondere richtet sich unsere Sorge nicht primär nach intrinsischen Faktoren, denn der qualitativ selbe Fortsetzer kann uns erstens, wenn er unser ähnlichster Fortsetzer ist, in ganz besonderer Weise am Herzen liegen, während er zweitens, wenn er zu einer Gruppe mehrerer uns ungefähr gleichähnlicher Fortsetzer gehört, lediglich eine von womöglich vielen verschiedenen Personen ist, um die wir uns in ganz besonderer Weise sorgen, und er uns drittens, wenn es einen uns deutlich ähnlicheren Fortsetzer gibt, sogar recht wenig bedeutet. Es stimmt demnach bei Licht besehen nicht, dass wir die von Williams behauptete Intuition, was mit einer dritten Person geschehe, könne für die gerechtfertigte Furcht vor Folter keinen Unterschied machen, tatsächlich haben.²⁰

²⁰Meine Auseinandersetzung mit dieser Motivation von Nozicks Theorie ist ungebührlich kurz und lückenhaft und verzichtet zudem auf eine Besprechung interessanter Einwände (siehe Johnston 1987, S. 67–69, und Noonan 2003, S. 227). Mir geht es an dieser Stelle nur darum, exemplarisch aufzuzeigen, dass es beachtenswerte Theorien gibt, die das Intrinsitätsprinzip ablehnen und mit denen dennoch der Anspruch verbunden ist, unsere

Achte und neunte Antwort

Der ähnlichste Fortsetzer einer Person zu sein, sieht Nozick, wie gesagt, als notwendige Bedingung für personale Identität an. Als weitere notwendige Bedingung – und natürliche Ergänzung zur Forderung nach einem ähnlichsten Fortsetzer – sieht er es an, *ähnlichster Vorgänger* [closest predecessor] zu sein: Es darf zu t_1 kein z geben, dass zu y zu t_2 in einer engeren (oder genauso engen) Beziehung steht als (bzw. wie) x zu t_1 . Diese zweite Bedingung hat jedoch eine von Nozick übersehene Konsequenz (siehe Noonan 2003, S. 223): Gemäß der obigen Ausbuchstabierung von Nozicks Theorie, wonach physische Kontinuität auch zählt, aber weniger wichtig ist als psychische, sind wir zwar in (i)–(iii) der ähnlichste Vorgänger der Person, die unseren Körper erbt, weil letztere keinerlei psychische Kontinuität mit einer früheren Person aufweist; in (vi) und auch in (v) teilt sich diese Person allerdings sämtliche geistigen Eigenschaften mit einer anderen, die deswegen dann ihr ähnlichster Vorgänger ist. (Komplizierter ist (iv), der Fall, in dem die auf unser Gehirn überspielten geistigen Eigenschaften so designt sind, dass sie denen einer tatsächlichen Person entsprechen. Die Bewertung dieses Falls hängt davon ab, ob man der Meinung ist, die geistigen Eigenschaften der Person, die unseren Körper erbt, entwachsen tatsächlich aus denen jener anderen Person, denen sie entsprechen.) Insgesamt sind also nur in den ersten drei Fällen die sich aus dieser Ausbuchstabierung von Nozicks Theorienschema ergebenden notwendigen Bedingungen für personale Identität zwischen uns und der Person, die unseren Körper erbt, klarerweise erfüllt; und in den letzten beiden Fällen ist mindestens eine der Bedingungen klarerweise nicht erfüllt.

Wenn wir uns also nicht einfach nur am Intransitätsprinzip stören, sondern darüber hinaus Identifizierungen von Personen auf Basis der gegebenen Ausbuchstabierung von Nozicks Theorienschema vornehmen möchten, sollten wir nicht erst beim Schritt von (v) nach (vi) einhaken, sondern mindestens einen Zwischenschritt früher. Und auch dieses Einhaken kann auf zweierlei Weisen begründet werden: als beste Konvention, deretwegen die Intuition, der Zwischenschritt von (iii) zu (iv) sei zu klein, um einen Unterschied zu machen (oder die, der Zwischenschritt von (iv) zu (v) sei dazu zu klein) als irreführend anzusehen ist (achte Antwort); oder als Einsicht, dass wir mindestens eine dieser Intuitionen bei Licht besehen gar nicht haben, weil sich unsere Intuitionen bezüglich personaler Identität gemäß Nozicks Position verhalten (neunte Antwort).

Varianten dieser beiden Antworten greifen andere Zwischenschritte an, oder sie greifen dieselben Zwischenschritte aus anderen Gründen an. So ist etwa Beck der Meinung, eingehendere Überlegungen machten uns klar, die Intuitionen bezüglich des eigenen Wohlergehens umfassend und korrekt wiederzugeben.

Kluft zwischen (ii) und (iii) sei in Wirklichkeit viel zu groß, um den betreffenden Schritt intuitiv akzeptabel erscheinen zu lassen (Beck 1998), und Noonan behauptet aus anderen Gründen dasselbe für den Übergang von (iv) zu (v) (Noonan 2003, S. 186–188). Mehrere Autoren kritisieren zudem, unsere Intuitionen bezüglich der Zwischenschritte würden verfälscht dadurch, dass Williams in seiner Präsentation des zweiten Gedankenexperiments sowohl auf die ursprüngliche als auch auf die resultierende Person mit „ich“ Bezug nimmt (Whiting 1999, S. 449; Oyowe 2013, S. 263–264; Cohnitz 2003, S. 175). In Sätzen wie „wenn der Moment der Folter kommt, werde ich nichts von dem erinnern, an das ich mich jetzt erinnern kann“ (Williams 1970, S. 81; meine Übersetzung) beispielsweise bezieht sich das erste „ich“ auf die resultierende und das zweite auf die ursprüngliche Person. Man kann dies allerdings auch als die plausibelste und natürlichste Weise ansehen, das Experiment zu schildern, sodass hier die Bezeichnungen den Intuitionen folgen und nicht umgekehrt (siehe Parfit 1994, S. 230). Wie dem auch sei, so selbstverständlich, wie Williams uns weismachen will, sind die Schritte von (i) bis (v) mitnichten.

Zehnte Antwort

Die achte und die neunte Antwort, so lässt sich einwenden, kleben zu eng an Williams' spezifischer Darstellung des Szenarios mit seinen Fällen (i) bis (vi). Denn wir können uns ohne Weiteres zu je zwei benachbarten Fällen sehr viele Zwischenfälle dergestalt vorstellen, dass sich benachbarte Fälle nur marginal unterschieden in Hinblick auf die Anzahl neu induzierter fiktiver oder von anderen Personen übernommener Erinnerungen und Charaktermerkmale. Es erscheint dann höchst unplausibel, bezüglich zweier solcher benachbarter Fälle zu behaupten, wir hätten nur in einem der beiden Grund, die Folter aus rein egoistischen Gründen zu fürchten. Die geradlinigere Lösung wäre dagegen, schon den Anfang nicht mitzumachen und zu bestreiten, dass in (i) die Folter schlimm für uns sei. Tatsächlich ist es für einen Verfechter einer Theorie psychischer Kontinuität naheliegend, einen vollständigem Gedächtnisverlust als einen derart drastischen Bruch der geistigen Kontinuität zu begreifen, dass personale Identität bereits in (i) nicht mehr attestiert werden kann.

(Parfit präsentiert ein ähnliches Spektrum von Fällen, in dem ich allerdings in den ersten Fällen nur nach und nach einzelne Erinnerungen verliere. Fall (i) liegt dann bereits recht weit in der Mitte des Spektrums, Fall (v) am entgegengesetzten Rand. Dieses Spektrum offenbart Parfit zufolge, wie unplausibel Williams' Annahmen, personale Identität könne nicht von Konventionen abhängen und liege entweder eindeutig vor oder eindeutig nicht

vor, sind. Ohne diese Annahmen bereite uns das zweite Gedankenexperiment aber keine Probleme. (Siehe Parfit 1984, S. 231–233.)

Die Intuition, dass in (i) die Folter schlimm für uns, kann man dann, unter Zuhilfenahme eines Begriffs aus Antwort fünf, als nachvollziehbare Überverallgemeinerung brandmarken. Diesmal ist dabei zu erklären, warum wir in den Fällen (i)–(v) fälschlicherweise dazu neigen, eher ein physisches Kriterium anlegen. (Fall (vi) klammere ich an dieser Stelle aus, denn in (vi) legen wir ein physisches Kriterium nur dann an, wenn wir zur Überzeugung gelangt sind, dass (v) und (vi) analog behandelt werden müssen.) Für die scheinbare Plausibilität eines physischen Kriteriums gibt es mehrere Gründe. So spielt sicherlich unsere Gewohnheit, Personen anhand von Körper-Kriterien zu identifizieren, eine wichtige Rolle. Schließlich sehen wir direkt, ob jemand allem Anschein nach physisch identisch mit einer uns bekannten Person ist; eine mutmaßliche geistige Identität hingegen lässt sich erst nach ausführlichem Gespräch erhärten. Besonders relevant wird dieser Unterschied in der forensischen Praxis, wo einzig und allein die körperliche Identität des Übeltäters mit dem Angeklagten zur Debatte steht. Selbst Locke, der Urvater von Theorien psychischer Kontinuität, betont, die irdische Gerichtsbarkeit habe sich an körperlicher Identität zu orientieren (Locke 1700/1975, II. 27. 22, S. 343–344). Ferner betrachten wir im Alltag meist auch Neugeborene und Komapatienten als mit einem normalen (zukünftigen bzw. vergangenen) Erwachsenen identische Personen, was sich allein mit Theorien psychischer Kontinuität nur schwer erklären lässt. Ein ganz anderer Grund ist außerdem Williams' Fokussierung auf zukünftige *physische* Schmerzen, die uns in den speziellen Situationen (i)–(v) eine besondere Sorge um unser körperliches Wohl einimpft und dieses dadurch über Gebühr ins Zentrum unserer Aufmerksamkeit rückt. Zusammengefasst erklären diese Gründe, warum wir im zweiten Gedankenexperiment die Intuition haben, die Folter rationalerweise fürchten zu müssen – auch wenn dies in Wirklichkeit gar nicht der Fall ist (wie Vertreter von Theorien psychischer Kontinuität mithilfe anderer, von Williams' Gedankenexperimenten unabhängiger Argumente zeigen können).

Elfte Antwort

In Antwort zehn wurden Gründe dafür genannt, die Intuition, in (i) sei die Folter schlimm für uns, als irreführend anzusehen. Das setzt immerhin voraus, dass wir diese Intuition überhaupt haben – was aus unterschiedlichen Gründen bei näherem Hinsehen höchst zweifelhaft erscheint.

Erstens beruht laut Perry die Plausibilität dieser Intuition darauf, dass gewisse relevante Details, die unseren Gedächtnisverlust in (i)–(v) betreffen, im Dunkeln belassen werden (Perry 2002, S. 114–118). Williams spricht von

„vollständigem Gedächtnisverlust“ („total amnesia“, S. 84), schreibt aber nicht explizit, ob er mit „vollständig“ meint, dass nicht eine einzige Erinnerung oder Überzeugung erhalten bleibt, oder dass der Gedächtnisverlust irreversibel ist. Man denke wieder wie in Antwort eins an ein Palimpsest: Das Ausradieren oder Abschaben der Seiten eines Buches und das anschließende Beschreiben mit anderen Inhalten kann die ursprünglich auf den Seiten enthaltene Information unwiderrufflich zerstören, wenn der Löschvorgang nur gründlich genug durchgeführt wurde, kann sie aber auch einfach nur dergestalt beschädigen, dass sie mit mehr oder minder großem technischen Aufwand wiederhergestellt werden kann. Dass ein Informationsverlust in diesem zweiten Sinn vorliegt, setzen wir normalerweise voraus, wenn wir über Gedächtnisverluste reden, denn zum einen kehren bei manchen realen Fällen von Amnesie vermeintlich vollständig verlorene Erinnerungen tatsächlich wieder, und zum anderen wissen wir, dass unser Gehirn sich mehr merkt, als an unser Bewusstsein dringt. Nach diesem Alltagsverständnis von Gedächtnisverlust erscheinen auch die von Williams geschilderten Situationen glaubhaft. Doch kann er diesen Sinn von Amnesie nicht meinen. Andernfalls nämlich hätte ein Vertreter einer Theorie psychischer Kontinuität keine Probleme mit den Fällen (i)–(vi), da Reversibilität ja impliziert, dass ein gewisses Maß an psychischer Kontinuität noch vorhanden sein muss, und könnte sogar darauf hinweisen, dass unsere Schwierigkeit, klar zu sagen, ab wann wir die Folter sinnvollerweise nicht mehr fürchten sollten, genau das ist, was seine Theorie voraussagt. Insbesondere wäre (vi) nichts anderes als ein verklausulierter Teilungsfall, da es hier zwei resultierende Personen gibt, die psychische Kontinuität zur ursprünglichen Person aufweisen. Wenn jedoch mit vollständigem Gedächtnisverlust eine (tatsächlich und theoretisch) irreversible Amnesie gemeint ist, dann gibt es wie erwähnt schon in (i) für einen Anhänger einer Theorie psychischer Kontinuität wenig Grund, etwas anderes als ebendiese Amnesie zu fürchten. Da eine irreversible Amnesie die personenspezifischen geistigen Eigenschaften wie Erinnerungen oder Überzeugungen zerstöre und nur allgemeine personale Fähigkeiten wie die zur Sprache intakt lasse, so könnte jener Anhänger argumentieren, verschwinde alles, was eine Person als Individuum ausmacht, unwiderrufflich. In diesem Sinne ist „Amnesie“ einfach ein „Euphemismus für ‚Tod‘“ (Perry 2002, S. 118). Sobald uns dies klar vor Augen steht, sehen wir ein, dass wir uns in (i) vor der Folter gar nicht fürchten, da sie nicht uns zuteil wird.

Ein zweiter Grund, dieselbe Intuition zu bestreiten, erwächst laut Coleman aus der merkwürdig zentralen Rolle, die den zukünftigen Schmerzen im zweiten Gedankenexperiment zukommt (siehe Coleman 2000). Wenn uns nicht Folter, sondern 100.000 € in Aussicht gestellt werden mit dem Hinweis, wir würden vor deren Erhalt unseres Gedächtnisses vollständig verlustig ge-

hen, ist alles andere als klar, ob wir uns über das Geld freuen. Vielmehr haben wir den Eindruck, dieses Geld nütze uns nach der Amnesie nichts mehr, weil alles, wofür wir es jetzt verwenden würden, uns danach nichts mehr bedeutet. Und das nicht, weil die 100.000 € angesichts der Schrecken des drohenden Gedächtnisverlusts nur ein kleiner Trost sind (auch über einen kleinen Trost freut man sich); vielmehr sie sind gar kein Trost und taugen nicht im Geringsten als Kompensation (und auch 100 Millionen € würden nichts taugen). Denn wir haben starke Zweifel daran, ob sie wirklich *uns* zugute kommen. Ähnliche Zweifel wären auch bei bevorstehender Folter angebracht, werden aber überdeckt von einigen anderen Faktoren. Zunächst erscheint uns Folter generell als so schlimm, dass uns selbst die kleinste Möglichkeit, dass es doch wir sein könnten, die gefoltert werden, einen kalten Schauer über den Rücken jagt. Dann sind wir eventuell nicht in der Lage, alle Hoffnung fahren zu lassen: Irgendwie, so sagen wir uns wider alle Gewissheit, ist es vielleicht doch möglich, dem Gedächtnisverlust zu entgehen. Dass uns in diesem Fall schreckliche Schmerzen erwarten, macht die gesamte Situation noch furchtbarer. Und drittens lässt sich intuitiv schlecht auseinanderdividieren, ob der Horror, mit dem wir angesichts dieser Aussichten in die Zukunft blicken, allein die Furcht vor der vollständigen Amnesie ist oder unter anderem auch die Furcht vor den zukünftigen Qualen einer mit uns mindestens körperidentischen Person. Ein solches Auseinanderdividieren wird erleichtert, wenn statt doppeltem Horror Furcht zum Beispiel mit Freude kombiniert wird. Dies sind die spezifischen Faktoren in Zusammenhang mit zukünftigen Schmerzen, die uns hier den Sand in die Augen streuen, der uns beim ersten Hinsehen dazu verleitet, Williams' Intuition, wir empfänden die Folter als schlimm für uns, zu teilen. In dem Moment, in dem wir von diesen Faktoren abstrahieren und uns beispielsweise überlegen, ob ein Geldgeschenk an die postamnesiale Person ein Grund zur Freude für uns wäre, teilen wir diese Intuition nicht mehr.

(Was die Rolle der Schmerzen angeht, kann man allerdings mit Unger zwar einerseits zugestehen, dass wir die fragliche Intuition nur haben, wenn im Szenario von Folter und nicht von Geldgeschenken die Rede ist, andererseits aber einwenden, dass verlässliche Intuitionen nur zu erhalten sind, wenn extrem negative Konsequenzen angedroht werden, da sich ansonsten zwangsläufig verschiedene Ziele vermischen (Unger 1990, S. 29–32). So möchte ich jetzt beispielsweise nicht, dass ich-in-10-Jahren viel Geld bekäme, wenn sich in der Zwischenzeit mein Charakter und meine Überzeugungen auf eine Weise wandelten, die ich-jetzt absolut nicht gutheiße; dies bedeutet jedoch nicht, dass ich in diesem Fall an meinem Überleben zweifle.)

Zwölfte Antwort

Unabhängig davon, ob es legitim ist, im zweiten Gedankenexperiment von Schmerzen statt zum Beispiel von Geldgeschenken zu sprechen, ist die Rede davon, zukünftige Schmerzen zu fürchten, hier alles andere als harmlos. Denn statt die Zuverlässigkeit oder die Existenz einer der relevanten Intuitionen anzugreifen, sollten wir besser eine Begriffsverwirrung diagnostizieren. Erinnern wir uns dazu zunächst an den Begriff der Quasi-Erinnerung, der in Fußnote 12 einen ersten Auftritt hatte: Im Falle einer Bewusstseinsverzweigung, resultierend zum Beispiel aus einer Gehirnteilung, teilen resultierende Personen die Erinnerungen der ursprünglichen Person, ohne mit ihr identisch zu sein. Diese intrinsisch von echten Erinnerungen nicht zu unterscheidenden geistigen Zustände sind Quasi-Erinnerungen. Es gibt nun eine wichtige Parallele zwischen den vergangenheitsgerichteten Erinnerungen und der zukunftsgerichteten (egoistischen) Sorge: Wahrhaft erinnern kann man sich nur an das, was man selbst zu einem vergangenen Zeitpunkt erlebt hat, wahrhaft und aus egoistischer Sicht sorgen nur um das, was einem selbst zu einem zukünftigen Zeitpunkt widerfahren wird. Beide, Erinnerungen wie Sorge, taugen daher als Indikatoren für personale Identität. Erinnerungen, insofern ihr Vorliegen eine identitätsbewahrende Verbindung zu einer früher existierenden Person anzeigt; Sorge, insofern ihr Vorliegen eine identitätsbewahrende Verbindung zu einer mutmaßlich zukünftig existierenden Person anzeigt. Und ebenso, wie wir von Erinnerungen intrinsisch nicht zu unterscheidende geistige Zustände haben können, die keine identitätsbewahrende Verbindung anzeigen, erscheint es plausibel, dass wir von egoistischer Sorge intrinsisch nicht zu unterscheidende Gefühle haben können, die keine solche Verbindung belegen. Im Falle einer Gehirnteilung etwa sorgt sich die ursprüngliche Person auf eine Weise um das Wohlergehen der beiden resultierenden Personen, die ganz anders ist als die Sorge ums Wohl von Freunden und Verwandten. Wir haben bereits in Antwort sieben auf Nozick zurückgehende Gründe dafür gesehen, sich nicht bloß um seinen ähnlichsten Fortsetzer zu sorgen; auf eine ähnliche Weise argumentiert Parfit, wenn er betont, personale Identität sei nicht das, was zähle (Parfit 1984, S. 255). Aus diesen Gründen sollten wir dem Begriff der Quasi-Erinnerung den der *Quasi-Sorge* beigesellen. Wir quasi-sorgen uns um das Wohlergehen einer zukünftigen Person, wenn wir ein von Sorge ums eigene Wohlergehen intrinsisch nicht zu unterscheidendes Gefühl haben, das auf eine bestimmte, aus begrifflichen Gründen nicht mit uns identische zukünftig existierende Person gerichtet ist. Der paradigmatische Fall ist dabei der, in dem wir uns um das Schicksal einer aus einer Teilung unseres Gehirns resultierenden Person quasi-sorgen. Doch ähnlich, wie ein Vertreter einer Theorie physischer Kontinuität von Quasi-Erinnerungen sprechen kann,

die die *B*-Körper-Person an *As* Leben hat (siehe Fußnote 12), kann auch ein Vertreter einer Theorie psychischer Kontinuität von der Quasi-Sorge sprechen, die wir im zweiten Gedankenexperiment für die in den Fällen (i)-(vi) resultierenden Personen empfinden.²¹

Ein wesentlicher Bestandteil egoistischer Sorge ist der Wunsch, das zukünftige Ich solle keine Schmerzen erleiden. Wenn diese egoistische Sorge allerdings intrinsisch ununterscheidbar ist von der Quasi-Sorge, die wir für eine nicht mit uns identische, aber hohe körperliche oder geistige Kontinuität zu uns aufweisende zukünftige Person empfinden können, können wir aus der bloßen Tatsache, dass wir die Schmerzen einer zukünftigen Person fürchten, nicht auf unsere Identität mit dieser Person schließen. Genauso, wie mir laut Nozick und Parfit aus rein egoistischen Gründen an der Weiterexistenz einer Person, die nicht mit mir identisch ist, ähnlich viel liegen kann wie am eigenen Überleben, kann ich, dieser zwölften Antwort zufolge, zukünftige Schmerzen, die nicht meine sind, *auf dieselbe Weise* fürchten wie eigene zukünftige Schmerzen (und diese Weise ist eine ganz andere als die, auf die ich zukünftige Schmerzen enger Freunde und Familienmitglieder fürchte). Dem zweiten Gedankenexperiment liegt demnach eine irrige Ansicht darüber, was es bedeutet, Schmerzen zu fürchten, zugrunde. Recht verstanden, erkennen wir, dass unsere Furcht vor Schmerzen hier kein Indikator für personale Identität sein kann.

*

Fassen wir zusammen: Während die erste Antwort das Paradox umgeht, indem sie ihm falsche Voraussetzungen nachweist, schlucken die zweite und die dritte Antwort seine Widersprüchlichkeit und schließen daraus auf die mangelnde Aussagekraft von Williams' Gedankenexperimenten (im Fall der zweiten Antwort) bzw. allen Gedankenexperimenten zur personalen Identität (im Fall der dritten Antwort). Antwort vier bestreitet, dass wir die entscheidende Intuition im ersten Gedankenexperiment tatsächlich haben. Antwort fünf räumt das Bestehen jener Intuition zwar ein, erläutert aber, dass sie auf die falsche Weise zustande kommt und somit unzuverlässig ist. Die Antworten sechs und sieben lehnen das für die Argumentation im zweiten Gedankenexperiment wesentliche Intrinsitätsprinzip ab, weil sich dadurch das

²¹Die Möglichkeit von dem, was ich hier Quasi-Sorge nenne, wird angedacht in Whiting 1999, S. 459, und Blackmore 2012. Whiting argumentiert dabei auf der Basis der strukturellen Ähnlichkeit von egoistischer Sorge und Sorge um enge Freunde, dass Sorge und Quasi-Sorge sich für Williams nicht klar auseinander halten lassen, und Blackmore zeigt auf, dass wir uns auch dann, wenn Personen keine transtemporale Existenz zukommt, um das Wohl bestimmter zukünftiger Personen ganz besonders sorgen.

Paradox am plausibelsten vermeiden lässt (Antwort sechs) bzw. weil es bei genauer Betrachtung nicht mit unseren Intuitionen zum eigenen Wohlergehen zu vereinbaren ist (Antwort sieben). Die Antworten acht und neun kritisieren wesentliche Zwischenschritte im zweiten Gedankenexperiment als für die Theoriebildung hinderlich (Antwort acht) bzw. durch unsere Intuitionen nicht gedeckt (Antwort neun). In Antwort zehn werden Gründe dafür aufgezeigt, die wichtige Ausgangsintuition im zweiten Gedankenexperiment als irreführend anzusehen; in Antwort elf wird argumentiert, dass wir diese Intuition gar nicht haben. Antwort zwölf schließlich besagt, das Paradox beruhe auf einem falschen Verständnis davon, was es heiße, zukünftige Schmerzen zu fürchten. Den Antworten eins und zwölf zufolge liegen demnach Williams' Paradox Denkfehler zugrunde; den Antworten vier, sieben, neun und elf zufolge stützt es sich auf Intuitionen, die wir bei Licht besehen gar nicht haben. Die übrigen Antworten räumen zunächst einmal ein, dass die paradoxe Situation tatsächlich besteht, und präsentieren unterschiedliche Wege, damit umzugehen.

Naturgemäß widersprechen manche dieser Antworten einander. Wie gerade gesagt, setzen einige Antworten zum Beispiel voraus, dass mit dem imaginierten Szenario und den behaupteten Intuitionen alles in Ordnung ist, während andere das Gegenteil behaupten oder zumindest implizit voraussetzen. Doch auch Antworten, die jeweils zugestehen, dass wir diese Intuitionen haben, können inkompatibel sein, wenn sie zum Beispiel wie zwei und drei aus der fehlenden Darstellungsinvarianz der Intuitionen konträre Schlüsse ziehen. Wichtiger noch als diese Unvereinbarkeiten ist allerdings, dass auch logisch konsistente Antworten häufig gegensätzlichen Geist atmen. Wer beispielsweise eine Theorie des ähnlichsten Fortsetzers vertritt, gleich ob mit konventionalistischer oder intuitionistischer Motivation, wird den Teufel tun, physische Kontinuität über psychische zu stellen und etwa Antwort fünf zu propagieren, da ein wesentlicher Vorzug, den Theorien physischer Kontinuität gegenüber solchen psychischer Kontinuität für sich in Anspruch nehmen, gerade darin besteht, das Intransitivitätsprinzip bewahren zu können. Ebenso wenig dürfte er sich für Gedankenexperiment-skeptische Antworten wie zwei oder drei erwärmen können, oder, falls er die im Rahmen der Antworten sechs und sieben angegebene Ausbuchstabierung favorisiert, für Antworten wie zehn oder elf, die seine Antwort überflüssig machen. Gleichzeitig Antwort fünf und Antwort zehn zu vertreten, erscheint ebenfalls wenig plausibel, zumindest dann, wenn man nicht die in Antwort drei gezogenen Schlüsse teilen möchte. Generell sollte man in fast allen Fällen als Befürworter einer bestimmten Antwort gute Gründe angeben können, warum wenigstens einige der übrigen Antworten falsch sind. Und das erscheint mir schwieriger, als unter diesen zwölf Antworten eine überzeugende zu finden.

Ich sprach davon, dass ich eine der Antworten favorisiere. Welche könnte das sein? Na ja, letztlich sind alle Alternativen auf dem Tisch, das Für und Wider ist dargelegt, ich habe keine mir bekannten wesentlichen Informationen zurückgehalten und habe versucht, allen Auffassungen so gerecht wie möglich zu werden. Mehr philosophisch Erhellendes habe ich nicht zu sagen. Die Argumente scheinen mir in eine bestimmte Richtung zu weisen, können aber natürlich auch anders bewertet werden. Und diese Bewertung muss jedem selbst überlassen bleiben. Meine Meinung tut nichts zur Sache.

Literatur

- Beck, S. (1998). Back to the self and the future. *South African Journal of Philosophy* 17, 211–225.
- Blackmore, S. (2012). She won't be me. *Journal of Consciousness Studies* 19, 16–19.
- Brown, J. R. (1991.). *The Laboratory of the Mind. Thought Experiments in the Natural Sciences*. London.
- Cohnitz, D. (2003). Personal identity and the methodology of imaginary cases. In K. Petrus (Hrsg.), *On Human Persons*, pp. 145–181. Frankfurt/Main.
- Coleman, S. (2000). Thought experiments and personal identity. *Philosophical Studies* 98, 53–69.
- Gendler, T. S. (1998). Exceptional persons: On the limits of imaginary cases. *Journal of Consciousness Studies* 5 5, 592–610.
- Gendler, T. S. (2000.). *Thought Experiment. On the Powers and Limits of Imaginary Cases*. New York.
- Häggqvist, S. (2009). A model for thought experiments. *Canadian Journal of Philosophy* 39, 55–76.
- Johnston, M. (1987). Human beings. *The Journal of Philosophy* 84, 59–83.
- Lewis, D. (1983). Survival and identity. In *Philosophical Papers. Vol. 1*, pp. 55–77. Oxford. Erweiterte Version; Original in Amélie Rorty (Hrsg.), *The Identity of Persons*, Berkeley 1976.
- Locke, J. (1700/1975.). *An Essay concerning Human Understanding* (4. Auflage ed.). Oxford.
- Nida-Rümelin, M. (2006). *Der Blick von innen*. Frankfurt/Main.

- Noonan, H. W. (2003.). *Personal Identity* (2. Auflage ed.). London.
- Nozick, R. (1981.). *Philosophical Explanations*. Oxford.
- Olson, E. (1997.). *The Human Animal. Personal Identity without Psychology*. Oxford.
- Oyowe, O. A. (2013). Physical continuity, self and the future. *Philosophia* 41, 257–269.
- Parfit, D. (1984.). *Reasons and Persons*. Oxford.
- Perry, J. (1972). Can the self divide? *The Journal of Philosophy* 69, 463–488. Wiederabgedruckt in John Perry, *Identity, Personal Identity, and the Self*, Indianapolis 2002.
- Perry, J. (2002). Williams on the self and the future. In *Identity, Personal Identity, and the Self*, pp. 103–118. Indianapolis. Überarbeitete Version; Original in *The Journal of Philosophy* 73 (1976), S. 416–428.
- Quante, M. (2007.). *Person*. Berlin.
- Shafir, E. (1998). Philosophical intuitions and cognitive mechanisms. In M. DePaul und W. Ramsay (Hrsg.), *Rethinking Intuition*, pp. 59–73. Lanham (MD).
- Shoemaker, S. (1963.). *Self-Knowledge and Self-Identity*. Ithaca (NY).
- Shoemaker, S. (1970). Persons and their pasts. *American Philosophical Quarterly* 7, 269–85.
- Sorensen, R. A. (1992.). *Thought Experiments*. Oxford.
- van Inwagen, P. (1990.). *Material Beings*. Ithaca (NY).
- Weber, M. A. (2014.). *Die Zerlegung des Ichs. Über die Grundlagen personaler Identität*. Münster.
- Weber, M. A. (im Erscheinen). Die Aussagekraft wirklichkeitsferner Gedankenexperimente für Theorien personaler Identität. In A. Oberprantacher (Hrsg.), *Mensch sein – Fundament, Imperativ oder Floskel? Beiträge zum 10. Kongress der Österreichischen Gesellschaft für Philosophie*. Innsbruck.
- Whiting, J. (1999). Back to „the self and the future“. *Philosophical Topics* 26, 441–477.
- Williams, B. (1956). Personal identity and individuation. *Proceedings of the Aristotelian Society* 57, 229–252. Wiederabgedruckt in Williams 1973..

Williams, B. (1960). Bodily continuity and personal identity. *Analysis* 21, 43–48.
Wiederabgedruckt in Williams 1973..

Williams, B. (1970). The self and the future. *Philosophical Review* 79, 161–180. Wiederabgedruckt in Williams 1973. Zitiert nach R. Martin und J. Barresi (Hrsg.), *Personal Identity*, Oxford 2003, S. 75–91.

Williams, B. (1973.). *Problems of the Self*. Cambridge.

Williamson, T. (2007.). *The Philosophy of Philosophy*. Oxford.