

Sind Gedankenexperimente in der praktischen Philosophie besonders?

Are Thought Experiments in Practical Philosophy Special?

MARC ANDREE WEBER, MANNHEIM

Zusammenfassung: Dieser Text geht der Frage nach, ob und, wenn ja, inwieweit sich Gedankenexperimente in der praktischen Philosophie in ihrer Struktur und ihrer epistemischen Signifikanz von Gedankenexperimenten in der theoretischen Philosophie oder in den Naturwissenschaften unterscheiden. Anhand einer allgemeinen Strukturanalyse von Gedankenexperimenten wird dabei aufgezeigt, dass bei Gedankenexperimenten in der praktischen Philosophie zwar häufig die angemessene Bewertung eines zugrunde gelegten Szenarios im Zentrum steht und nicht, wie zum Beispiel in theoretischen Philosophie oft, dessen angemessene Beschreibung, dass dieser Unterschied aber kaum Auswirkungen hat auf die prinzipielle Anwendbarkeit – abgesehen von einer bisher übersehenen Konsequenz: Man darf in einem Gedankenexperiment nicht annehmen, dass moralische, ästhetische oder entscheidungstheoretische Prinzipien in unterschiedlichen möglichen Welten verschieden sind. Andernfalls käme dem Gedankenexperiment keinerlei epistemische Signifikanz zu.

Schlagwörter: Gedankenexperiment, Moralarteil, Vorstellungswiderstand, Möglichkeit

Abstract: This paper deals with the question whether and if yes, to what degree, thought experiments in practical philosophy differ in their structure and their epistemic significance from thought experiments in theoretical philosophy or in the sciences. Based on a general analysis of the structure of thought experiments, it is shown that on the one hand, thought experiments in practical philosophy often rely on an appropriate evaluation of a given scenario and not, as thought experiments in theoretical philosophy and elsewhere usually do, on an appropriate description thereof;

Alle Inhalte der Zeitschrift für Praktische Philosophie sind lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.



but that, on the other hand, nothing much follows from this difference for their applicability – with one exception that has hitherto been overlooked: One must not assume in a thought experiment that different moral, aesthetic, or decision-theoretical principles hold in different possible worlds. Otherwise the thought experiment would have no epistemic significance.

Keywords: Thought Experiment, Moral Judgment, Imaginative Resistance, Possibility

1. Ein Beispiel zur Einleitung

Gedankenexperimente in der praktischen Philosophie sind häufig nicht weniger schillernd und verblüffend als ihre oft bekannteren Verwandten aus theoretischer Philosophie oder Naturwissenschaft. Doch haben sie hier wie dort dieselbe epistemische Berechtigung? Funktionieren sie auf dieselbe Weise? Lassen sich Unterschiede feststellen – und wenn ja, welche?

Betrachten wir, um ein Beispiel zu haben, ein Szenario von Derek Parfit:

In *Hölle eins* besteht die letzte Generation von Menschen aus zehn unschuldigen Personen, die 50 Jahre lang große Qualen erleiden. Die Leben dieser Leute sind alles andere als lebenswert. Sie würden sich sofort umbringen, wenn sie nur könnten. In *Hölle zwei* besteht die letzte Generation von Menschen nicht aus zehn, sondern aus zehn Millionen unschuldiger Personen, die genauso große Qualen erleiden, allerdings nur für 50 Jahre minus einen Tag. (Parfit 1984, S. 393, meine Übersetzung.)

Es wird eine Situation beschrieben, die sehr wirklichkeitsfern und künstlich ist, die wir uns aber ohne Probleme vorstellen können. Doch das Szenario allein stellt noch kein Gedankenexperiment dar, sondern allenfalls den Versuchsaufbau. Parfit fährt fort:

Wenn wir annehmen, dass wir auf jeden Fall in einer dieser zwei Höllen existieren werden, wäre es aus Eigeninteresse heraus klarerweise rational, Hölle zwei zu bevorzugen, da wir dort einen Tag weniger lang leiden müssten. Sollten wir folgern, dass Hölle zwei besser wäre, in irgendeinem moralisch relevanten Sinne? Würde Satan weniger schlecht handeln, wenn er diese und nicht die andere Hölle herbeiführte?

Die Fragen, die Parfit stellt, sind rein rhetorisch; natürlich wäre aus moralischer Sicht Hölle zwei die offensichtlich schlechtere, weil es hier insgesamt deutlich mehr Leid gibt. Ein Satan, der seinen Namen verdient, sollte diese Hölle zu verwirklichen suchen und nicht die harmlosere erste.

Philosophisch spannend wird es, wenn wir anhand dieses Szenarios die zwei geläufigsten Arten des Utilitarismus miteinander vergleichen: den Gesamtnutzen-Utilitarismus und den Durchschnittsnutzen-Utilitarismus. Ersterer besagt (in einer seiner plausibelsten Varianten), dass diejenige Handlung die moralisch bessere ist, deren zu erwartende Folgen den größten Gesamtnutzen für die betrachtete Gesellschaft haben. Dieser Gesamtnutzen kann allerdings auch wachsen, wenn es jedem einzelnen Mitglied der Gesellschaft schlechter geht; dann nämlich, wenn die Bevölkerungszahl in ausreichendem Maße zunimmt. Als Alternative zur Maximierung des zu erwartenden Gesamtnutzens ist daher die Maximierung des zu erwartenden Pro-Kopf-Nutzens vorgeschlagen worden; die resultierende Theorie ist der Durchschnittsnutzen-Utilitarismus. Und genau gegen diesen richtet sich Parfits Höllenszenario.

Denn der Durchschnittsnutzen – oder besser: -schaden – ist in Hölle zwei geringer, da jeder Insasse einen Tag weniger leidet als in Hölle eins. Eine Durchschnittsnutzen-Utilitaristin muss daher Hölle zwei als die moralisch bessere ansehen und für den Fall, dass sie entscheiden kann, welche der beiden Höllen verwirklicht wird, die zweite wählen – trotz der unendlich viel größeren Qualen, die in dieser Hölle insgesamt, von allen dort vegetierenden Personen zusammen, erlitten werden müssen. Sie müsste genauso entscheiden, wie Satan es getan hätte, und träfe damit eine aus moralischer Sicht absurde Wahl.¹

Parfits Gedankenexperiment lässt sich somit wie folgt zusammenfassen: Der Durchschnittsnutzen-Utilitarismus kann nicht die beste Moraltheorie sein, da er, anders als der Gesamtnutzen-Utilitarismus, bezüglich des Höllenszenarios diejenige von zwei Situationen als besser identifiziert, die aus moralischer Sicht offensichtlich die deutlich schlechtere ist.²

1 Wäre die Durchschnittsnutzen-Utilitaristin eine der Personen in der Hölle, für die sie sich entscheidet, würde sich bei ihr immerhin das moralisch Gebotene mit dem aus Eigeninteresse Gebotenen decken.

2 Ich werde in diesem Aufsatz durchweg einen moralischen Realismus voraussetzen, weil viele derjenigen, die in der Ethik unter Zuhilfenahme von Gedankenexperimenten argumentieren, dies implizit oder explizit auch tun, und weil es die Darstellung des Folgenden an verschiedenen Stellen vereinfacht.

Typischerweise funktionieren Gedankenexperimente genau so wie Parfits Fall der zwei Höllen. Es wird ein Szenario beschrieben, bezüglich dessen eine bestimmte Einschätzung naheliegt; in unserem Fall die, dass Hölle eins der aus moralischer Sicht zu präferierende Ort ist. Diese Einschätzung entkräftet oder stützt eine bestimmte philosophische These; in unserem Fall wird der Durchschnittsnutzen-Utilitarismus widerlegt.

Ich werde im Folgenden genauer auf die Natur solcher Einschätzungen eingehen und vor allem auf die Besonderheiten hinweisen, die ihnen bei Gedankenexperimenten in der praktischen Philosophie eigen sind (Abschnitt 3). Vorher allerdings möchte ich noch einen Punkt hervorheben, der Gedankenexperimente von tatsächlichen Experimenten, in den Naturwissenschaften und anderswo, unterscheidet, nämlich die Irrelevanz empirischer Beobachtungen, und möchte zudem einiges zur allgemeinen Struktur von Gedankenexperimenten sagen (Abschnitt 2). Schließlich thematisiere ich eine bestimmte Vorgehensweise, die bei Gedankenexperimente in der praktischen Philosophie³ *nicht* funktioniert (Abschnitt 4).

2. Zur Struktur von Gedankenexperimenten

Mit der angesprochenen „Irrelevanz empirischer Beobachtungen“ ist nicht gemeint, dass kein Hintergrundwissen über unsere Welt vonnöten sein darf, um das Gedankenexperiment durchzuführen. Im Fall von Parfits Höllen etwa ist das Wissen unabdingbar, dass Menschen im Normalfall keine Masochisten sind, die Qualen als etwas Erstrebenswertes ansehen. Gemeint ist vielmehr, dass wir jenseits dessen, was normal informierte Personen über unsere Welt wissen, nichts an empirischen Beobachtungen voraussetzen. Insbesondere ist unwichtig, ob das zugrunde liegende Szenario tatsächlich schon einmal verwirklicht wurde, und darüber hinaus sogar, ob es in unserer Welt, vor dem Hintergrund unserer Naturgesetze, technisch realisiert werden könnte.

Der Grund für die Irrelevanz empirischer Beobachtungen liegt in der Natur der zu stützenden oder zu widerlegenden Thesen. Wenn zum Beispiel der Durchschnittsnutzen-Utilitarismus die beste Moraltheorie wäre, wäre er

3 Genau genommen geht es mir um Gedankenexperimente in Ethik, Ästhetik und Entscheidungstheorie. Nur der Einfachheit halber spreche ich von Gedankenexperimenten in der praktischen Philosophie; an Kategorisierungsthesen, etwa dass Ästhetik eher zur praktischen als zur theoretischen Philosophie zu zählen sei, liegt mir hier nichts.

dies nicht nur in unserer Welt, sondern auch in allen anderen möglichen Welten (oder zumindest denjenigen möglichen Welten, die unserer Welt hinreichend ähnlich sind; mehr dazu in Abschnitt 4). Es genügt zu seiner Widerlegung demnach, irgendeine Welt zu finden, in der er offensichtlich nicht die beste Moraltheorie ist. Anders gesagt: Wenn es eine Welt gibt, in der es moralisch besser ist, von zwei alternativen Situationen diejenige hervorzubringen, in der der Durchschnittsnutzen der Beteiligten geringer ist, muss der Durchschnittsnutzen-Utilitarismus falsch sein. Parfits Höllenszenario liefert, indem es eine solche Welt beschreibt, ein Gegenbeispiel zum Durchschnittsnutzen-Utilitarismus. Und da der Durchschnittsnutzen-Utilitarismus als verkappte Notwendigkeitsthese anzusehen ist – wenn er die beste Moraltheorie ist, ist er dies in allen möglichen Welten – genügt als Gegenbeispiel ein bloß mögliches Szenario.

Dies ist allgemein bei Gedankenexperimenten der Fall. Sie zeigen mögliche Umstände auf, in denen eine bestimmte Einschätzung korrekt erscheint, und können so Thesen entkräften, die für den Fall ihrer Geltung mit Notwendigkeit gelten würden und mit jener Einschätzung nicht in Einklang zu bringen sind. (Gelegentlich besteht die Einschätzung auch darin, dass ein bestimmter Sachverhalt unter keinen Umständen bestehen kann, und widerlegt dann Thesen, denen zufolge Sachverhalte, die nur gemeinsam mit jenem bestehen können, möglicherweise oder tatsächlich realisiert ist.) Gegen Thesen ohne modale Dimension können *Gedankenexperimente*, die ihrer Natur nach ohne konkrete empirische Beobachtungen auskommen, nichts ausrichten; hier wären tatsächliche Gegenbeispiele vonnöten. (So kann man zum Beispiel gegen die These, alle Schwäne seien weiß, nicht sinnvollerweise einwenden, die Existenz schwarzer Schwäne sei doch sicherlich möglich.)

Damit und mit der vorhin gemachten Beobachtung, dass in Gedankenexperimenten zunächst ein Szenario beschrieben und zu diesem dann eine Einschätzung vorgebracht wird, können wir Gedankenexperimente auf eine allgemeine Struktur bringen: Gedankenexperimente lassen sich ansehen als *modale Argumente mit zwei Prämissen*. Mit der ersten Prämisse wird dabei die metaphysische Möglichkeit des zugrunde liegenden Szenarios behauptet und mit der zweiten die Korrektheit einer bestimmten Einschätzung unter der Voraussetzung, dass das Szenario vorliegt. Die Konklusion des Arguments lautet dann, dass die Korrektheit einer bestimmten Einschätzung metaphysisch möglich ist; oder, in anderen Worten, dass es mögliche Welten gibt, in denen jene Einschätzung zutrifft.

Parfits Gedankenexperiment der zwei Höllen besteht demnach aus einer ersten Prämisse, derzufolge es metaphysisch möglich ist, dass jemand entscheiden kann, genau eine der beiden Höllen zu verwirklichen, aus einer zweiten Prämisse, derzufolge es, wenn man die Wahl hätte, entweder Hölle eins oder Hölle zwei zu realisieren, moralisch besser wäre, Hölle eins herbeizuführen statt Hölle zwei und damit nicht das zu tun, was der Durchschnittsnutzen-Utilitarismus verlangt, und einer Konklusion, derzufolge es metaphysisch möglich ist, dass die moralisch bessere Handlung nicht die vom Durchschnittsnutzen-Utilitarismus geforderte ist. Und weil der Durchschnittsnutzen-Utilitarismus, wie erwähnt, wenn er gilt, mit Notwendigkeit gilt, also in allen (relevanten) möglichen Welten, genügt die bloße Möglichkeit seiner Nichtgeltung, um ihn als falsch zu entlarven.

(Hier und im übrigen Text habe ich eine entscheidungstheoretische Ausdeutung von Parfits Gedankenexperiment gewählt: Jemand muss sich zwischen der Realisierung von Hölle eins und der von Hölle zwei unterscheiden, und die Frage ist, welche Handlung die moralisch bessere ist. Eine solche Auffassung klingt bei Parfit an, wenn er die Frage streift, welche Hölle Satan herbeiführen müsste, um möglichst schlecht zu handeln, wird aber dort nicht konsequent durchgehalten. Alternativ könnte man anstelle der moralischen Qualität von Handlungen die von Situationen vergleichen. Parfits Gedankenexperiment ließe sich dann beschreiben als bestehend aus einer ersten Prämisse, derzufolge es metaphysisch möglich ist, dass jede der zwei Höllen existiert, aus einer zweiten Prämisse, derzufolge, wenn diese Höllen existierten, Hölle eins, im Widerspruch zum Durchschnittsnutzen-Utilitarismus, moralisch besser wäre als Hölle zwei, und einer Konklusion, derzufolge es möglich ist, dass diejenige Situation die moralisch bessere ist, die gemäß dem Durchschnittsnutzen-Utilitarismus die schlechtere ist.)

Ich habe die These, dass Gedankenexperimente nichts anderes als modale Argumente mit zwei Prämissen sind, hier aus einem Einzelbeispiel heraus entwickelt und nur angedeutet, warum sich die wesentlichen strukturellen Züge dieses Einzelbeispiels auf alle Gedankenexperimente verallgemeinern lassen. Der zweischrittige Aufbau von Gedankenexperimenten – erst die Darstellung eines Szenarios, dann eine spezielle Einschätzung zu diesem – wird dabei durch das Vorliegen von genau zwei Prämissen wiedergegeben und die modale Komponente – Gedankenexperimente als Gegenbeispiele zu Notwendigkeits- oder Möglichkeitsthesen – durch die entsprechenden modalen Formulierungen der Prämissen und der Konklusion. Eine weitaus umfangreichere Besprechung von Fallstudien wäre nötig, um jene

These zu erhärten. Zudem müsste detaillierter ausgeführt werden, wie genau die logische Form insbesondere der zweiten Prämisse auszusehen hat, und die Gültigkeit des Arguments müsste bewiesen werden. All das brauche ich hier nicht zu tun, sondern kann auf die Literatur verweisen, insbesondere auf Sorensen 1992, Kapitel 6, wo für eine ganze Reihe von Gedankenexperimenten beispielhaft gezeigt wird, dass ihnen eine Struktur gemeinsam ist, die in allen wesentlichen Hinsichten der hier vorgestellten gleicht, und auf Williamson 2007, ebenfalls Kapitel 6, wo diese Struktur anhand eines anderen Beispiels deutlich ausführlicher diskutiert und die Gültigkeit des hier zitierten Arguments bewiesen wird.⁴ Gedankenexperimente haben epistemische Signifikanz, wenn sie nicht bloß gültige, sondern auch schlüssige Argumente sind, wenn also beide Prämissen wahr sind.

Mit der ersten Prämisse eines auf Argumentform gebrachten Gedankenexperiments wird, so habe ich gesagt, die metaphysische Möglichkeit eines Szenarios behauptet. Diese Möglichkeitsbehauptung ist unabdingbar, weil aus einer bestimmten Einschätzung bezüglich eines unmöglichen Szenarios nicht die mögliche Korrektheit der Einschätzung gefolgert werden kann. Die mit der ersten Prämisse vorgenommene Beschränkung für Szenarien von Gedankenexperimenten, mit denen der Anspruch auf epistemische Signifikanz verbunden ist, ist allerdings nicht sonderlich restriktiv; in aller Regel sind wir nicht versucht, Gedankenexperimente um metaphysisch unmögliche Szenarien herum zu bauen. Grob gesprochen sind dabei metaphysisch unmögliche Szenarien solche, die mit dem Wesen der Dinge nicht zu vereinbaren sind, in denen also zum Beispiel, wenn es zum Wesen von Gold gehört, aus Atomen mit 79 Protonen im Kern zu bestehen, Goldatome mit anderer Protonenzahl vorkommen (oder in denen gar begriffliche Unmöglichkeiten, runde Quadrate etwa, anzutreffen sind). Da wir uns in der Regel gar nicht erst vorstellen können, dass ein metaphysisch unmögliches Szenario realisiert ist, und es in den seltensten Fällen philosophischen Mehrwert zu versprechen scheint, doch zu versuchen, sich eines auszudenken, kommt uns die Beschränkung, Gedankenexperimenten keine metaphysisch unmöglichen Szenarien zugrunde zu legen, kaum einmal in die Quere – mit einer Ausnahme: Eine bestimmte Gruppe von Gedankenexperimenten in der

4 Siehe außerdem Häggqvist 2009, Ichikawa und Jarvis 2009, Malmgren 2011 und Grundmann und Horvath 2014 für ähnliche Strukturanalysen sowie Weber 2014, Kapitel 1, für eine Argumentation dafür, dass wir statt metaphysischer Möglichkeit bloß begriffliche Möglichkeit fordern sollten.

praktischen Philosophie verletzt diese Beschränkung auf subtile Weise, wie ich in Abschnitt 4 erläutern werde.

Zumeist ist die zweite Prämisse eines Gedankenexperiments die entscheidende, denn mit dieser Prämisse wird die Einschätzung vorgebracht, die eine bestimmte These ins Wanken bringen soll. Der Natur dieser Einschätzung wenden wir uns nun zu.

3. Drei Arten von Einschätzungen

Ich habe bisher recht vage von einer „Einschätzung“ gesprochen, die in einem Gedankenexperiment bezüglich eines Szenarios vorgenommen wird. Im Fall von Parfits Höllenszenario besteht diese Einschätzung in einem moralischen Urteil: Hölle eins zu realisieren ist weniger schlimm als Hölle zwei zu realisieren. In zahlreichen anderen Gedankenexperimenten in der Ethik fällen wir ebenfalls moralische Urteile, etwa dass es moralisch besser ist, die Weiche umzulegen und einen statt fünf Gleisarbeiter zu überfahren, oder es moralisch fragwürdig ist, von einer Person zu verlangen, neun Monate in einem Krankenhausbett zu verbringen, damit irgendein Musiker überleben kann, dessen Blutkreislauf illegalerweise mit dem jener Person verbunden wurde.

(Allerdings gibt es auch Gedankenexperimente in der Ethik, bei denen die Einschätzung nicht in einem moralischen Urteil besteht. Rawls' Urzustandsgedankenexperiment etwa geht vom Urzustandsszenario aus, in dem die involvierten Personen hinter dem Schleier des Nichtwissens eine Gesellschaftsordnung entwerfen; siehe Rawls 1975. Die relevante Einschätzung, für die Rawls lang und breit argumentiert, besagt, dass die Personen im Urzustand genau die von ihm, Rawls, propagierten zwei Grundprinzipien wählen würden. Dies ist eine Prognose darüber, was in einem Szenario passieren würde, und kein moralisches Urteil.)⁵

5 So dargestellt, ist es nicht Teil von Rawls' Gedankenexperiment, dass der Urzustand ein Szenario ist, in dem zwangsläufig die gerechtestmöglichen Prinzipien gewählt werden, welche auch immer das sein mögen; das Gedankenexperiment erstreckt sich nur darauf, welche konkreten Prinzipien im Urzustand gewählt würden. Man kann somit das Urzustandsgedankenexperiment als schlüssig ansehen und dennoch Rawls' Grundprinzipien ablehnen – etwa, weil man mit Nozick die Nichteinbeziehung historisch gewachsener Ansprüche für ungerecht hält (Nozick 1974, S. 198–204). Ich finde es sinnvoll, die Argumentation für den Urzustand als optimales Prinzipienwahlscenario mit der Argumentation für die Wahl von Rawls' Grundprinzipien nicht zu ver-

In anderen Bereichen der praktischen Philosophie wird jene Einschätzung in der Regel eher zum Beispiel die Rationalität einer Handlung betreffen („Handlung *X* ist die im Szenario für Subjekt *S* rational gebotene“, „Handlung *X* ist im Szenario für Subjekt *S* besser als Handlung *Y*“) oder auch den ästhetischen Charakter eines Kunstwerks („Das im Szenario geschilderte Kunstwerk *Z* ist ästhetisch gehaltvoll“). Allgemein wird die Einschätzung dort, wo ein qualitatives Element eine wesentliche Rolle spielt, die Form eines *Werturteils*, einer *Bewertung* annehmen.

Betrachten wir zum Kontrast ein Gedankenexperiment aus der theoretischen Philosophie, einen sehr simplen Gettierfall (siehe Russell 1948, S. 154):

STEHENGEBLIEBENE UHR. *Linus schaut auf seine Armbanduhr, die halb sieben anzeigt, und formt daraufhin die Überzeugung, dass es halb sieben ist. Die Armbanduhr, von der Linus weiß, dass sie bisher immer sehr zuverlässig war, ist allerdings stehengeblieben – und zwar vor exakt zwölf Stunden, sodass sie in dem Moment, in dem Linus auf sie schaut, die korrekte Uhrzeit anzeigt.*

Die relevante Einschätzung zu diesem Szenario besagt, dass Linus hier zwar eine gerechtfertigte, wahre Überzeugung bezüglich der Uhrzeit hat, aber kein Wissen. Dies widerlegt die These, Wissen könne gerechtfertigte, wahre Überzeugung sein. Die Einschätzung, dass bei Linus kein Wissen um die Uhrzeit vorliegt, ist hier lediglich eine Art und Weise zu beschreiben, was im Szenario der Fall ist. Auch diese Beobachtung lässt sich verallgemeinern, denn immer, wenn es sich bei der Einschätzung eines Szenario darum dreht, was im Szenario eigentlich der Fall ist, haben wir es mit einer *Beschreibung* wesentlicher Elemente des Szenarios zu tun.

Manchmal geht es in Gedankenexperimenten aber nicht darum, was der Fall *ist*, sondern darum, was der Fall *sein wird*. Galilei etwa wollte die aristotelische Auffassung widerlegen, derzufolge die Fallgeschwindigkeit von Körpern in einem gegebenen Medium proportional zu deren Gewicht ist und insbesondere schwerere Körper schneller fallen als leichtere.⁶ Zu diesem

mischen und nur Letztere als Urzustandsgedankenexperiment im engeren Sinne aufzufassen. Nebenbei: Das Problem mit der Bevölkerungsgröße, auf das Parfits Höllen-Gedankenexperiment hinweist, betrifft, wie man sich leicht klar machen kann, nicht nur den Durchschnittsnutzen-Utilitarismus, sondern auch Rawls' zweites Grundprinzip.

6 Siehe Galilei 1638/1890, S. 57–60, sowie die Rekonstruktionen des Gedankenexperiments in Norton 1996, S. 340–345, Gendler 1998 und Gendler 2000b.

Zweck sollten wir uns, so schlug Galilei vor, zwei unterschiedlich schwere Steine miteinander verbunden denken. Fällt dann das Konstrukt aus den zwei Steinen schneller als der schwerere der beiden Steine separat? Angenommen, die Fallgeschwindigkeit der Steine wäre proportional zu ihrem Gewicht und der schwerere fiel schneller. Dann müsste einerseits der langsamer fallende leichtere Stein den Fall des schneller fallenden schwereren, der mit ihm verbunden ist, abbremsen, sodass das Konstrukt langsamer fiel als der schwerere Stein allein. Andererseits ist aber das Gewicht des Konstrukts größer als das des schwereren Steins, sodass es schneller fallen müsste. Zusammengenommen müsste das Konstrukt aus den zwei Steinen also sowohl langsamer als auch schneller fallen als der schwerere Stein allein, was widersprüchlich ist. Die Annahme, die Fallgeschwindigkeit verändere sich proportional zum Gewicht, lässt sich somit nach Galilei nicht aufrechterhalten.

Könnte man jetzt, über Galilei hinausgehend, argumentieren, man müsse irgendwie einem Irrtum aufgesessen sein, wenn man glaubt, der langsamer fallende Stein bremse den mit ihm verbundenen schnelleren ab, und in Wirklichkeit falle das Konstrukt aus beiden Steinen schneller als der schwerere Stein allein? Falls ja, könnte man den Widerspruch vermeiden und an der Annahme, die Fallgeschwindigkeit eines Körpers sei proportional zu dessen Gewicht, festhalten.

Allerdings ist uns angesichts des Szenarios der zwei aneinander befestigten Steine, die aus einiger Höhe fallen gelassen werden, auf der Basis unseres empirischen Vorwissens sofort klar, dass dieses Konstrukt nicht einfach schneller fallen wird als der schwerere der beiden Steine separat. (Das wird besonders offensichtlich, wenn wir uns die beiden Steine nur eher lose miteinander verbunden denken, etwa durch einen Bindfaden.) Wir erkennen, dass der auftretende Widerspruch durch die Annahme, die Fallgeschwindigkeit sei proportional zum Gewicht, zustande kommen muss, und wir erkennen dies, weil wir einschätzen können, was passiert, wenn man zwei miteinander verbundene Steine fallen lässt. Anders gesagt, betrifft die relevante Einschätzung hier das, was im Szenario, den Naturgesetzen gemäß, passieren wird. Und aus dieser Einschätzung – zwei miteinander verbundene Steine fallen nicht schneller als der schwerere von beiden – folgt, wie Galilei gezeigt hat, dass die Fallgeschwindigkeit von Körpern nicht proportional von deren Gewicht abhängt.

Ganz allgemein geht es in Gedankenexperimenten in den Naturwissenschaften häufig um Einschätzungen genau dieser Art: Was wird, den Naturgesetzen gemäß, in einem gegebenen Szenario als Nächstes passieren?

Aus Einschätzungen dieser Art, die auf unseren empirischen Vorkenntnissen beruhen, diese aber in einem neuen Zusammenhang zur Anwendung bringen, wird auf die Unmöglichkeit bestimmter Kandidaten für Naturgesetze geschlossen.⁷

Halten wir fest: Gedankenexperimenten ist eine gemeinsame Struktur eigen. Sie bestehen jeweils aus einem Szenario, hinsichtlich dessen eine Behauptung darüber gemacht wird, wie bestimmte Aspekte dieses Szenarios (a) angemessen zu beschreiben oder (b) angemessen zu bewerten oder (c) angemessen weiterzudenken sind.⁸ Viele, aber nicht alle, Gedankenexperimente in der praktischen Philosophie unterscheiden sich von anderen Gedankenexperimenten dahingehend, dass es ihnen um eine Behauptung der Art (b) geht, also um eine angemessene Bewertung eines Szenarios. Gedankenexperimente in der theoretischen Philosophie beinhalten häufig eine Behauptung der Art (a), solche in den Naturwissenschaften häufig eine der Art (c).

Wodurch sind Behauptungen der Art (a), (b) oder (c) gerechtfertigt? Zwei Antworten liegen nahe, die, denke ich, im Kern dasselbe besagen. Man könnte bestreiten, dass Behauptungen der Art (a), (b) oder (c) überhaupt einer Rechtfertigung bedürfen. So leuchtet die Behauptung, Hölle eins herbeizuführen sei moralisch gesehen weniger schlimm als Hölle zwei herbeizuführen, unmittelbar ein, und Parfit kann ohne Weiteres davon ausgehen, dass seine Leser ihm zustimmen und keine weitere Begründung verlangen. Man könnte sich aber auch auf Intuitionen berufen und zum Beispiel sagen, dass wir bezüglich des Höllenszenarios die Intuition haben, Hölle eins herbeizuführen sei weniger schlimm als Hölle zwei herbeizuführen. In beiden

7 Man sieht hieran auch, welcher Unterschied zwischen Widerlegungen von Allaussagen wie *Alle Schwäne sind weiß* und Widerlegungen von Kandidaten für Naturgesetze besteht: Dass alle Schwäne weiß sind, ist solange mit allem, was wir über die Welt wissen, vereinbar, wie wir noch keinen nichtweißen Schwan entdeckt haben; wir benötigen daher ein tatsächliches Gegenbeispiel. Dass schwerere Körper schneller fallen als leichtere, ist hingegen nicht mit allem, was wir über die Welt wissen, vereinbar, wie uns in dem Moment klar wird, in dem wir unsere empirischen Vorkenntnisse auf ein bestimmtes, wohl-gewähltes Szenario anwenden. Solche vermeintlichen Naturgesetze, die auf subtile Weise unserem Weltwissen widersprechen, ohne dass uns dies bislang aufgefallen wäre, können wir mithilfe von Gedankenexperimenten, die den verborgenen Widerspruch sichtbar machen, als falsch entlarven. (Siehe hierzu auch Kuhn 1964/1977.)

8 Siehe auch Gendler 2000b, S. 25.

Fällen, dem der fehlenden Rechtfertigungspflicht und dem der Intuitionen, ist die Antwort nicht vollständig und bedürfte noch genaueren Ausführungen darüber, wann Behauptungen rechtfertigungspflichtig sind und wann nicht bzw. unter welchen Umständen wir epistemisch aussagekräftige Intuitionen haben und unter welchen nicht. Für das Gelingen eines Gedankenexperiment ist allerdings unwesentlich, wie diese genaueren Ausführungen ausfallen. Eine gute Gedankenexperimentatorin wird ihre Gedankenexperimente so entwerfen, dass ihre Leser die zentrale Einschätzung teilen, sei es, weil sie keiner Rechtfertigung bedarf, sei es, weil die Leser die Intuition haben, dass sie stimmt.

Umgekehrt kann man ein Gedankenexperiment dadurch zu entkräften versuchen, dass man die zentrale Einschätzung zurückweist, zum Beispiel mit der Begründung, sie sei durch bloße Berufung auf Intuitionen unzureichend gerechtfertigt, oder mit dem Hinweis, die fraglichen Intuitionen kämen auf fehlerhafte Weise zustande, würden etwa bloß durch irrelevante Details des zugrunde liegenden Szenarios evoziert oder würden sich verflüchtigen, wenn man mehr über, sagen wir, Logik oder Wahrscheinlichkeitstheorie lerne.⁹

Eine Variante dieser Kritik an der zentralen Einschätzung betrifft in besonderem Maße Gedankenexperimente in der praktischen Philosophie. Intuitionen bezüglich des moralischen Stellenwerts einer Handlung in einem bloß imaginierten Szenario, so der Einwand, seien unzuverlässig, weil die notwendige Unterbestimmtheit solcher Szenarios moralische Urteile erschwere.¹⁰ Die Szenarien, die Gedankenexperimenten zugrunde liegen, sind bloß imaginiert und können anderen daher nur durch Beschreibung zugänglich gemacht werden. Solche Beschreibungen lassen aber, anders als die Wirklichkeit, zwangsläufig vieles offen. In Parfits Höllenszenario kennen wir zum Beispiel weder Geschlecht noch Alter noch Augenfarben der involvierten Personen, wir wissen nichts über die Art der Qualen, die erlitten werden müssen, und auch nichts darüber, wie die übrige Welt, jenseits der realisierten Hölle, aussieht. Und selbst wenn uns all diese Informationen von Parfit mitgeliefert worden wären, gäbe es immer noch viele andere, über die wir

9 Ausführlicher gehe ich hierauf in Weber 2017 ein. Siehe auch Gendler 2007 für viele durch psychologische Erkenntnisse untermauerte Beispiele dafür, unter welchen Umständen uns Intuitionen bei Gedankenexperimenten in die Irre leiten können.

10 Ein Einwand dieser Art findet sich zum Beispiel in Dancy 1985, S.144f.

nicht verfügten. Szenarien, die wir bloß durch eine Beschreibung kennen und nicht dadurch, dass wir uns selbst in ihnen befinden, weisen notgedrungen Lücken auf. Das ist, dem Einwand zufolge, insofern problematisch, als dass von uns verlangt wird, bezüglich einer solchen unzureichend erfassten Situation den moralische Wert von Handlungen zu beurteilen. Doch gestaltet sich, dem Einwand zufolge, eine solche Beurteilung häufig umso schwieriger, je weniger ausgeprägt unser Überblick über die Gesamtsituation ist, da sehr wohl moralisch relevante Sachverhalte offen geblieben sein können. Im Höllenszenario etwa mag es durchaus relevant sein, ob einige oder alle der zu quälenden Personen Masochisten sind.

Es sei dahingestellt, wie plausibel der Gedanke ist, man benötige zur moralischen Bewertung einer Handlung eine umfassende Kenntnis des gesamten Kontexts. Der hier vertretenen Auffassung von Gedankenexperimenten zufolge hat man eine solche nämlich; die Beschreibung des Szenarios ist nicht lückenhaft. Oder, genauer gesagt, „wir überlassen es faktisch der Welt, die Details der Geschichte auszufüllen“ (Williamson 2007, S. 186, meine Übersetzung). Wir tun dies formal, indem wir die zentrale Einschätzung in ein kontrafaktisches Konditional einbetten, das grob gesprochen besagt: Wenn das Szenario *A* realisiert wäre, wäre Einschätzung *B* korrekt. Um zu überprüfen, ob ein solches Konditional wahr ist, müssen wir gemäß der Standardanalyse kontrafaktischer Konditionale die unserer Welt ähnlichsten möglichen Welten betrachten, in denen das Antezedens des Konditionals wahr und das Szenario somit realisiert ist. Das wiederum bedeutet, dass wir aufgefordert sind, die Lücken im Szenario auf genau die Weise zu füllen, die das Szenario unserer Welt möglichst ähnlich macht. Wir nehmen uns, mit anderen Worten, die Welt, wie wir sie kennen, zum Muster, und ändern nur das, was im Szenario explizit anders dargestellt wird (inklusive dem, was sich plausiblerweise aus diesen Änderungen ergibt). Und daher können wir davon ausgehen, dass zum Beispiel die Verteilung von Geschlecht, Alter und Augenfarben in Parfits Höllenszenario der in einer durchschnittlichen irdischen Gesellschaft entspricht und dass Masochisten bei Parfit nicht häufiger anzutreffen sind als üblich.

Auch wenn die hier präsentierte Auffassung von Struktur und Wirkweise von Gedankenexperimenten in der Literatur prominent vertreten wird, finden sich auch ganz andere Charakterisierungen. So schreibt zum Beispiel Cappelen nach einer Reihe von Fallstudien:

Gute philosophische Gedankenexperimente [cases] sind verwirrend: Sie sind schwierig und herausfordernd, sodass wir uns kaum einmal auf eine endgültige Antwort zu den verschiedenen von einem Gedankenexperiment aufgeworfenen Fragen einigen. (Cappelen 2012, S. 189; Cappelens Hervorhebungen, meine Übersetzung)¹¹

Das scheint meiner These zu widersprechen, dass die Einschätzungen, die wir bezüglich Szenarien abgeben, zumindest bei gelungenen Gedankenexperimenten entweder keiner Rechtfertigung bedürfen oder durch unsere entsprechenden Intuitionen ausreichend gerechtfertigt sind. Diese Einschätzungen – etwa, dass Hölle eins herbeizuführen weniger schlimm ist als Hölle zwei herbeizuführen – erscheinen erst einmal nicht schwierig. Was also meint Cappelen?

Ihm zufolge stehen Gedankenexperimente oft am Anfang intensiver philosophischer Debatten oder lösen sie gar erst aus. Verwirrend erscheinen sie, so lese ich Cappelen, insofern sie uns aus einer Komfortzone herausreißen und zwingen, über eine bestimmte Theorie neu nachzudenken. Ein Gettierfall wie STEHENGEBLIEBENE UHR zum Beispiel ist verwirrend und herausfordernd, weil er die naheliegende Identifikation von Wissen mit gerechtfertigter, wahrer Überzeugung konterkariert und uns veranlasst, den Wissensbegriff exakter zu fassen – ohne bereits eine exakte Fassung mitzuliefern. Parfits Höllen-Gedankenexperiment ist verwirrend und herausfordernd, weil es die zunächst plausibel anmutende Ersetzung des Gesamtnutzen- durch den Durchschnittsnutzen-Utilitarismus als zu kurz greifend entlarvt und uns auferlegt, den Durchschnittsnutzen-Utilitarismus entweder zu modifizieren oder ganz aufzugeben – ohne bereits eine Modifikation oder Alternativtheorie vorzuschlagen.

Zudem kann das Ergebnis der durch ein Gedankenexperiment erzwungenen Überlegungen im Einzelfall auch sein, die angegriffene These doch beizubehalten. Man kann dann entweder einräumen, dass man eben mit den unplausibel erscheinenden Konsequenzen, die das Gedankenexperiment aufzeige, leben müsse, oder dafür argumentieren, dass die Konsequenzen tatsächlich weniger unplausibel sind, als sie ursprünglich erschie-

¹¹ Cappelen stellt noch eine Reihe weiterer Charakteristika von Gedankenexperimenten zusammen, von denen manche ebenfalls auf den ersten Blick nicht recht zu der hier vorgestellten Auffassung von Gedankenexperimenten passen. Ich bespreche an dieser Stelle nur das Charakteristikum, das mir am bedenkenswertesten erscheint.

nen mochten. Eine Durchschnittsnutzen-Utilitaristin könnte etwa, eventuell unter Hinzuziehung zusätzlicher Argumente, die Kröte, die Parfit ihr hinhält, schlucken und behaupten, Hölle zwei herbeizuführen sei tatsächlich moralisch gesehen besser als Hölle eins herbeizuführen.

So verstanden, lässt sich Cappelens Beobachtung in Einklang bringen mit der hier vorgestellten Strukturanalyse von Gedankenexperimenten sowie der Behauptung, die zentrale Einschätzung bedürfe zunächst keiner Rechtfertigung oder nur der durch Intuitionen.

Die Tatsache, dass Cappelen gleichwohl die hier vorgestellte Strukturanalyse in Bausch und Bogen kritisiert, rührt daher, dass er sie als These darüber missversteht, wie Gedankenexperimente in der Literatur präsentiert werden bzw. präsentiert werden sollten (siehe Cappelen 2012, S. 197–99). Er stellt zunächst fest, dass die Urheberinnen von Gedankenexperimenten diese in aller Regel nicht als zweiprämisse Argumente der beschriebenen Form vorstellen und solche Argumente dabei wohl auch nicht im Sinn haben. Sodann überlegt er, ob Argumentationen, die Gedankenexperimente in Form solcher Argumente beinhalten, besser seien als Argumentationen, die die entsprechenden Gedankenexperimente in der von ihren Urheberinnen vorgebrachten Form beinhalten, und weist dies als unplausibel zurück, unter anderem mit dem Hinweis darauf, dass in der philosophischen Praxis oft gar keine Versuche unternommen werden, die betreffenden Argumentationen durch Formalisierungen zu verbessern.

Cappelen übersieht dabei, dass der Zweck der Strukturanalyse nicht darin besteht, eine deskriptive oder normative Aussage über die Verwendung von Gedankenexperimenten zu machen, sondern darin, zu *definieren*, was Gedankenexperimente sind: Etwas ist genau dann ein Gedankenexperiment, wenn es die besagte Form hat. Man vergleiche dies mit jener lehrbuchhaften Definition von *Argument* (die Details spielen keine Rolle), derzufolge ein Argument eine Folge von Aussagesätzen sei, mit der der Anspruch verbunden ist, dass ein Satz der Folge (die Konklusion) von den anderen Sätzen der Folge (den Prämissen) rational gestützt wird. Auch hier wird nicht beschrieben, wie Philosophen Argumente präsentieren oder präsentieren sollten. Es wird nur gesagt, was ein Argument ist; und alles, was nicht diese Form hat oder sich darauf bringen lässt, ist keines. Analoges gilt für Gedankenexperimente.

Dabei wird, ähnlich wie im Fall der Argumente, unter anderem in Kauf genommen, dass die ein oder andere schillernde Allegorie oder Illustration, die meist als Gedankenexperiment angesehen wird, streng genommen keines ist. Als Beispiel mag Platons Höhlengleichnis dienen. Hierin wird zu kei-

ner bestimmten modalen These Stellung genommen, sondern vielmehr eine Geschichte erzählt, die Platons Auffassung von Philosophie im Allgemeinen und seine Ideenlehre im Besonderen spiegelt. Wohl wegen des Einbezugs eines kontrafaktischen Szenarios zur Präsentation theoretischer Überlegungen wird das Höhlengleichnis oft als Gedankenexperiment bezeichnet – verwirrenderweise, weil ihm nicht Experimenthaftes eigen ist, weil damit keine neuen Daten oder Intuitionen gesucht werden, die zu einer bestimmten Problemstellung konkrete Lösungshinweise liefern sollen. Die Strukturanalyse von Gedankenexperimenten kann diese Verwirrung erklären, indem sie Gedankenexperimente als etwas ihrem Wesen nach Inferentielles, als eine besondere Art von Argument, charakterisiert.

Ferner wird auch bloßes kontrafaktisches Denken, das sich manifestiert in Fragen wie „Was würde passieren, wenn ich diese Serpentina doppelt so schnell hinunterführe?“ oder „Was täte ich, wenn ich Milliardärin oder Bundeskanzlerin wäre?“, korrekterweise von Gedankenexperimenten abgegrenzt. Bei bloßem kontrafaktischen Denken geht es nämlich nicht um modale Thesen, also darum, ob irgendetwas, eventuell unter Voraussetzung der Naturgesetze, möglich oder notwendig ist, sondern darum, was *das wahrscheinlichste Ergebnis* einer bestimmten Handlung oder *die wahrscheinlichste Folge* einer bestimmten Konstellation wäre. Wir benutzen solches kontrafaktische Denken, um uns in unserer Welt zu orientieren und hilfreiche Prognosen zu erhalten; wir benutzen es, anders als Gedankenexperimente, *nicht*, um etwas über naturgesetzliche, moralische oder metaphysische Zusammenhänge zu erfahren. Auch bloßes kontrafaktisches Denken wird also von der Strukturanalyse plausibel klassifiziert.

4. Die fehlende Weltrelativität von Werturteilen

Dem hier vertretenen Ansatz zufolge ist eine Art von Gedankenexperimenten problematisch, die auf den ersten Blick nicht sonderlich aufsehenerregend daherkommt: Gedankenexperimente, in denen wir um des Arguments willen annehmen, dass eine bestimmte moralische, entscheidungstheoretische oder ästhetische These korrekt ist.

Betrachten wir zunächst einen Fall in der Art, wie Jackson (1991) sie bespricht:

KASINO. *Justus hat eine Benefizgala veranstaltet und € 10.000 für wohltätige Zwecke eingesammelt. Als er das Geld zur Bank bringen will und auf dem Weg dahin ein Kasino sieht, kommt ihm der Gedan-*

ke, dass die € 10.000, beim Roulette auf der richtigen Zahl platziert, zu € 360.000 würden, einem Betrag, mit dem man noch viel mehr Gutes tun könnte. Justus hat 38 Handlungsalternativen: nicht ins Kasino gehen; das Geld auf die 0 setzen; das Geld auf die 1 setzen; das Geld auf die 2 setzen; ...; das Geld auf die 36 setzen.

Diejenige Handlung, die die besten Konsequenzen hat, ist, das Geld auf die Zahl zu setzen, die in der entsprechenden Spielrunde tatsächlich fällt. In einer möglichen Welt, in der das die 0 ist, wäre das also die Handlung, das Geld auf die 0 zu setzen; in einer möglichen Welt, in der das die 1 ist, wäre das die Handlung, das Geld auf die 1 zu setzen; und so weiter. In *keiner* relevanten möglichen Welt wäre die Handlung mit den besten Konsequenzen die, nicht ins Kasino zu gehen. Und doch ist es gerade diese Handlung, die moralisch geboten erscheint. Es ist somit, diesem Gedankenexperiment zufolge, möglich, dass die moralisch beste Handlung eine ist, von der wir wissen, dass sie weniger gute Konsequenzen hat als eine alternative Handlung.

Der Grund dafür liegt in der *epistemischen Unsicherheit* darüber, welche Zahl in der tatsächlichen Welt fallen wird. Da Justus die beste Alternative nicht kennt und die Wahl einer Alternative, die sich als die beste erweisen könnte, mit großer Wahrscheinlichkeit katastrophale Konsequenzen hätte, nämlich den Verlust des gesamten auf der Benefizgala eingesammelten Geldes sowie, für Justus selbst, einen hohen Prestigeschaden, sollte er sich für eine Alternative entscheiden, die in jedem Fall harmlose Konsequenzen hat, wenn auch in keinem Fall optimale.

Weatherson fragt sich, ob man solche Jackson-Fälle variieren könnte, indem man die epistemische Unsicherheit durch eine moralische ersetzt. Hier ist eines seiner Beispiele (Weatherson 2019, S. 43; ich habe unwesentliche Details des Szenarios geändert):

ABTREIBUNG. Kira und Leyla sind eng befreundet und leben in einem Staat, der Abtreibung mit mehreren Jahren Gefängnis bestraft. Kira ist in der zwölften Woche schwanger, möchte das Kind unbedingt abtreiben und bittet Leyla, die sehr viel besser in der Organisation solcher Dinge ist, um Hilfe.

Leyla hat nun, so können wir annehmen, drei Handlungsalternativen:

- 1. Sie kann Kira helfen, das Kind abzutreiben.*
- 2. Sie kann Kira an die Behörden übergeben, die die Abtreibung verhindern und Kira ins Gefängnis stecken werden.*

3. Sie kann weder das eine noch das andere tun, und ob Kira dann ohne ihre Hilfe die Abtreibung durchzieht oder das Kind austrägt, bleibt abzuwarten.

Weatherson betrachtet nun zwei mögliche Welten. In der ersten ist Abtreibung moralisch erlaubt. Da die Weigerung, einer engen Freundin in einer Notlage zu helfen, moralisch höchst fragwürdig ist, ist in dieser Welt die erste Handlungsoption die beste, die dritte etwas schlechter und die zweite ganz schlimm.

In der zweiten Welt ist es moralisch geboten, *nicht* abzutreiben; wir können annehmen, dass Abtreibung dort moralisch auf ähnlicher Stufe steht wie Mord. In dieser Welt würde die erste Handlungsoption einer Beihilfe zum Mord gleichkommen und müsste daher auf jeden Fall vermieden werden. Richtig wäre hier für Leyla, den Quasi-Mord zu verhindern und Kira den Behörden zu übergeben, auch wenn das wahrscheinlich den Bruch der Freundschaft bedeuten würde. Etwas schlechter als diese zweite Option, aber deutlich besser als die erste, wäre die dritte, die immerhin die Möglichkeit offenlässt, dass der Quasi-Mord nicht geschieht.

Die Frage, die Weatherson aufwirft, lautet nun, ob in Analogie zu den Jackson-Fällen auch im Abtreibungsgedankenexperiment diejenige Handlungsalternative als beste erscheint, die in keiner der betrachteten möglichen Welten die beste ist, nämlich die dritte. Dafür spricht, neben der Analogie, dass diese Handlungsalternative in keiner der betrachteten möglichen Welten aus moralischer Sicht katastrophal wäre, während die erste Option in Welt zwei und die zweite Option in Welt eins moralisch gesehen fürchterlich wären.

Weathersons Antwort hierauf ist, dass die dritte Option einen Auswuchs moralischer Feigheit darstellt in einer Situation, die nach moralischer Tapferkeit verlangt (Weatherson 2019, S. 44). Anders als in Fällen epistemischer Unsicherheit erscheine in Fällen moralischer Unsicherheit – in Fällen also, in denen man keine gerechtfertigten Überzeugungen hinsichtlich dessen hat, was moralisch geboten ist – die Wahl der überall bloß zweitbesten Option keineswegs sicher, klug und empfehlenswert.

Ich will an dieser Stelle nicht diskutieren, wie überzeugend Weathersons Antwort ist. Das Gedankenexperiment ist in mehrerlei Hinsicht schwierig und bietet Raum für Einwände. (Zum Beispiel ignoriert Weatherson, dass unabhängig vom moralischen Status von Abtreibungen auch das Brechen von Gesetzen wie denen, die Abtreibung verbieten, eine moralische Dimen-

sion hat, die berücksichtigt werden müsste. Auch über Gesetze, die etwas moralisch Neutrales verbieten, darf man sich nicht einfach hinwegsetzen.) Ich will stattdessen auf einen Aspekt hinweisen, den man leicht übersieht: Wir können nicht einfach zwei oder mehr verschiedene mögliche Welten betrachten, in denen unterschiedliche Moralgesetze herrschen.

Bereits im Zusammenhang mit dem Durchschnittsnutzen-Utilitarismus habe ich erwähnt, dass moralische Theorien, wenn sie gelten, nicht bloß in der tatsächlichen Welt gelten, sondern auch in allen anderen möglichen Welten (oder zumindest den unserer Welt hinreichend ähnlichen, und das sind in aller Regel die relevanten). Ich werde gleich noch mehr dazu sagen; setzen wir dies für den Moment voraus. Wenn Abtreibung dann moralisch erlaubt ist, wie in Weathersons erster Welt, muss sie auch in allen anderen (relevanten) möglichen Welten erlaubt sein. Moralgesetze können sich dann nicht von Welt zu Welt ändern; und daher kann man auch nicht wie Weather-son noch eine zweite Welt betrachten, in der Abtreibung aus moralischer Sicht verboten ist. Wenn Abtreibung moralisch erlaubt ist, dann ist die erste Option die moralisch gebotene; wenn Abtreibung moralisch verwerflich ist, dann ist die zweite Option die moralisch gebotene. Ob wir demnach die erste oder die zweite Handlungsoption ergreifen sollten, hängt vom moralischen Status ab, den Abtreibung hat – in unserer Welt, und damit auch in allen anderen.

Die dritte Option dagegen kann niemals die moralisch gebotene sein. Sie ist nicht in manchen Welten besser als die erste und schlechter als die zweite und in anderen Welten schlechter als die erste und besser als die zweite; vielmehr ist sie entweder in allen Welten besser als die erste und schlechter als die zweite, und dann sollten wir die zweite Option wählen, oder sie ist in allen Welten schlechter als die erste und besser als die zweite, und dann sollten wir die erste Option wählen. Die Analogie zu Jackson-Fällen kommt damit erst gar nicht auf, und die Frage, die Weatherson aufwirft, lässt sich gar nicht erst sinnvoll stellen. Abgesehen davon ist dieses Ergebnis in seiner Ablehnung der dritten Option Wasser auf Weathersons Mühlen, die ja die Analogie zwischen epistemischer und moralischer Unsicherheit zermahlen sollen.

Allgemeiner gesprochen, gibt uns bei Gedankenexperimenten in der praktischen Philosophie die Tatsache, dass sie oft auf Werturteilen beruhen, Raum für eine spezielle Art von Fehler: Da Werturteile nicht relativ zu möglichen Welten gefällt werden, sondern für den Fall ihrer Geltung weltübergreifende Geltung beanspruchen, können wir keine zwei möglichen Welten annehmen, die sich durch die Geltung oder Nichtgeltung eines Werturteils voneinander unterscheiden. Gedankenexperimente, die dies dennoch tun,

verletzen die erste Prämisse, derzufolge das zugrunde liegende Szenario metaphysisch möglich sein muss, es also, anders gesagt, mögliche Welten geben muss, in denen das Szenario realisiert ist.

Natürlich kann auch das zugrunde liegendes Szenario eines Gedankenexperiments in der theoretischen Philosophie oder in den Naturwissenschaften metaphysisch unmöglich sein. Man sollte zum Beispiel besser nicht den bekannten Zeitreisefall, in dem jemand in die Vergangenheit reist und dort seinen Vater tötet, bevor der ihn zeugen kann, als Ausgangsszenario eines solchen Gedankenexperiments verwenden. Nur ist in einem solchen Fall oft leicht zu sehen, dass etwas Unmögliches vorausgesetzt wird, und die Gefahr ist gering, daraus irrtümlicherweise philosophisches Kapital schlagen zu wollen. Die Annahme einer Welt, in der eine bestimmte moralische (oder ästhetische oder entscheidungstheoretische) These nicht gilt, erscheint jedoch erst einmal unverdächtig. Entsprechend größer ist die Gefahr bei Gedankenexperimenten in der praktischen Philosophie, die erste Prämisse zu verletzen. Um zu sehen, was hier schief laufen kann, muss man die fehlende Weltrelativität von Werturteilen mit der allgemeinen Struktur von Gedankenexperimenten zusammenbringen.

Zentral für die Argumentation in diesem Abschnitt ist die These, dass moralische Urteile bzw. Werturteile im Allgemeinen nicht weltrelativ sind, also im Falle ihrer Geltung in allen (relevanten) möglichen Welten gelten. Das ist, gerade auch mit Bezug auf Abtreibung, erklärungsbedürftig. Es gibt zum Beispiel mögliche Welten, in denen die physiologischen oder psychologischen Gegebenheiten bei Menschen völlig andere sind. Betrachten wir etwa eine Welt, die in allen relevanten Hinsichten wie unsere ist, außer dass Embryos schon relativ direkt nach der Zeugung nachweislich ein Bewusstsein haben, das dem eines 30 Wochen alten Fötuses in unserer Welt entspricht. Die Position, dass in dieser anderen möglichen Welt Abtreibung moralisch verboten ist, in unserer aber (zumindest in der ersten Schwangerschaftshälfte) erlaubt, erscheint dann zumindest nicht abwegig. Wer eine solche Position vertritt, setzt aber ein bestimmtes Verständnis von Abtreibung voraus, demzufolge sich der Begriff *Abtreibung* auch auf Schwangerschaftsabbrüche bei Wesen anwenden ließe, bei denen die embryonale Entwicklung grundlegend anders verläuft. Verstünde man unter *Abtreibung* hingegen so etwas wie *Schwangerschaftsabbruch bei Wesen, die physiologisch wie Menschen in unserer Welt funktionieren*, dann wäre Abtreibung, wenn moralisch erlaubt, auch in Welten mit superfrüh entwickelten Embryos erlaubt – nur wäre deren Tötung im Mutterleib dann eben keine Abtreibung.

Man kann Probleme wie das hinsichtlich des richtigen Verständnisses des Abtreibungsbegriffs umgehen, indem man die Formulierungen moralischer Prinzipien so allgemein und neutral wie möglich hält. Wenn man zum Beispiel nur darüber redet, unter welchen Umständen man Lebewesen, die Personen sind, töten darf, und unter welchen Umständen Lebewesen, die keine Personen sind, dürften die resultierenden Prinzipien unabhängig von den spezifischen physiologischen und psychologischen Gegebenheiten einer möglichen Welt gelten. Und wenn man, um ein zweites Beispiel zu geben, den Utilitarismus nicht mit Bezug auf Wohlergehen definiert, wie es manchmal geschieht, sondern mit Bezug auf einen neutral verstandenen Nutzen, dann gilt er, falls er gilt, auch in einer Welt voller Masochisten, in der der Nutzen dann am größten ist, wenn die Qualen am größten sind. Ausreichend allgemein und neutral formulierte moralische Prinzipien scheinen in allen möglichen Welten gültig; eingeschränkte Gültigkeit scheint bloß das Ergebnis zu spezifischer und zu wenig neutraler Formulierung zu sein.¹² Halten wir also fest, dass allgemein formulierte moralische Theorien, falls sie gelten, mit Notwendigkeit gelten, und spezifischer formulierte immerhin noch in allen unserer Welt hinreichend ähnlichen möglichen Welten – und dies sind so gut wie immer die relevanten.

In der Regel sind übergeordnete moralische Prinzipien deutlich allgemeiner gefasst als moralische Einzelfallurteile. Entsprechend klarer tritt daher bei solchen Prinzipien die fehlende Weltrelativität von Werturteilen zutage. Vergleichen wir zum Beispiel das Gesamtnutzen-Prinzip, wonach die moralisch gebotene Handlung stets jene ist, deren zu erwartende Folgen den größten Gesamtnutzen für die betrachtete Gesellschaft haben, mit dem ebenfalls bereits erwähnten Einzelfallurteil, dass es moralisch besser ist, die Weiche umzulegen und einen statt fünf Gleisarbeiter zu überfahren. Das Einzelfallurteil erscheint viel stärker in unserer Welt verhaftet; falls Weichen oder Gleisarbeiter in einer bestimmten anderen Welt nicht existieren oder vollkommen andere Eigenschaften haben als in unserer (etwa weil für Gleisarbeiter das Leben ähnlich schlimm wäre wie das der Insassen von Parfitts Höllen, sodass sie sich sofort umbringen würden, wenn sie könnten), würde das Einzelfallurteil anders ausfallen. Das übergeordnete moralische

12 Es gibt auch mögliche Welten, in denen keine Akteure existieren, die in mehr oder weniger hohem Maße moralisch handeln können; ja, die Existenz solcher moralischen Akteure mag sogar aufgrund der dort geltenden Naturgesetze unmöglich sein. Wenn Moralgesetze mit Notwendigkeit gelten, gelten sie auch in Welten ohne moralische Akteure.

Prinzip hingegen nimmt nicht verdeckt auf die tatsächliche Beschaffenheit der Welt Bezug; was zu erwartende Folgen, was Gesamtnutzen und was eine Gesellschaft ist, ist weltunabhängig oder könnte zumindest sinnvollerweise so aufgefasst werden, und ob beispielsweise so etwas wie eine Gesellschaft in der betrachteten Welt überhaupt existiert, ändert nichts daran, wie an eventuellen Gesellschaften zu handeln wäre.

Man könnte daher mit einiger Plausibilität sagen, die fehlende Weltrelativität von Werturteilen bestünde nur für übergeordnete moralische Prinzipien, nicht für moralische Einzelfallurteile. Das allerdings gilt nur insofern, als dass solche Prinzipien häufig bereits in der erforderlichen Allgemeinheit formuliert sind oder in ihrer Formulierung dieser wenigstens nahekommen, während es in der Formulierung von Einzelfallurteilen nicht selten von Ausdrücken wimmelt, die auf die spezifischen Gegebenheiten unserer Welt und unserer Kultur gemünzt sind. Packt man in die Formulierung eines Einzelfallurteils hingegen alle relevanten Fakten zu den darin erwähnten konkreten Gegenständen, Eigenschaften und Vorgängen hinein, wird alles zwar sehr viel unübersichtlicher, aber auch unabhängig von der zugrunde gelegten Welt. Das Entscheidende ist hier nicht die Unterscheidung zwischen allgemeinem Prinzip und spezifischem Moralurteil, sondern der Allgemeingrad der Formulierung. Ist dieser hinreichend hoch, besteht keine Weltrelativität mehr.

Als weiteren Beleg für die These, dass Werturteile nicht weltrelativ sind, möchte ich ein Phänomen anführen, das in der Literatur als Vorstellungswiderstand [*imaginative resistance*] bezeichnet wird.¹³ Wir finden es vergleichsweise schwierig, uns eine fiktionale Welt vorzustellen, die zum Beispiel in moralischer Hinsicht von unserer abweicht (Gendler 2000a, S. 56). Während wir kein Problem damit haben, uns Welten vorzustellen, in denen Mäuse sprechen und Menschen zaubern können, in denen Reisen in entlegene Orte des Universum, in andere Zeiten oder Parallelwelten an der Tagesordnung sind, können oder wollen wir uns nicht vorstellen, dass in einer fiktionalen Welt andere Moralgesetze herrschen und es dort zum Beispiel moralisch erlaubt ist, aus Spaß Babys zu foltern (unter Voraussetzungen wie der, dass es sich nicht um Masochistenbabys handelt). Es fällt uns zwar nicht schwer, uns vorzustellen, dass jemand aus Spaß Babys foltert, und auch nicht, dass in einer fiktionalen Welt niemand dies verwerflich fin-

13 Siehe hierzu zum Beispiel Walton 1994, Gendler 2000a und, als historische Quelle, Hume 1757/1964.

det, aber wir weigern uns zu akzeptieren, dass diese Handlung in irgendeiner Welt moralisch erlaubt ist. Ebenso bereitet uns die Vorstellung einer Welt keine Schwierigkeiten, in der Satan überlegt, welche von Parfits zwei Höllen er realisieren solle, wohl aber die einer Welt, in der die Realisierung von Hölle zwei die moralisch bessere Alternative ist. (Wie man sich leicht klar macht, gilt, was ich hier für moralische Urteile beschrieben habe, für andere Werturteile, etwa ästhetische, in gleichem Maße. Man versuche zum Beispiel probenhalber, sich eine Welt vorzustellen, die wie unsere ist, nur dass Goethes *Faust* darin kein Meisterwerk, sondern Schund wäre.)

Es ist umstritten, ob dieser Vorstellungswiderstand tatsächlich besteht, und falls ja, ob er auf einem Nichtvorstellenkönnen beruht oder auf einem Nichtvorstellenwollen (Tuna 2020). Darüber hinaus wird der genaue Zusammenhang zwischen Vorstellbarkeit und metaphysischer Möglichkeit kontrovers diskutiert und hierbei insbesondere, ob (eine hinreichend idealisierte Form von) Vorstellbarkeit metaphysische Möglichkeit impliziert oder umgekehrt oder beides oder nichts von beidem (Yablo 1993, Chalmers 2002). Ich bin daher nicht in der glücklichen Lage, unwidersprochen folgende Schlussfolgerungen aufstellen zu können: Das Phänomen des Vorstellungswiderstands zeige, dass wir uns nicht vorstellen *können*, die Moralgesetze seien andere als in unserer Welt, und die Unvorstellbarkeit eines Sachverhalts *impliziere* dessen metaphysische Unmöglichkeit. Aber wer beides nicht ganz unplausibel findet, dürfte im Phänomen des Vorstellungswiderstands dennoch einen Grund dafür finden, Werturteile für nicht weltrelativ zu halten.

Und das gilt auch für Werturteile, bei denen sich, wie in Weathersons Abtreibungsfall, nicht direkt ein Vorstellungswiderstand regt. Hätte Weatherson uns gebeten, eine Welt zu betrachten, in der es moralisch erlaubt ist, aus Spaß Babys zu foltern, hätten wir schnell gemerkt, dass mit dem Gedankenexperiment etwas nicht stimmen kann. Da er aber vorschlug, Welten zu betrachten, in denen etwas, dessen moralischer Status kontrovers diskutiert wird, einmal erlaubt und einmal verboten ist, ruft keine dieser beiden Welten auf Antrieb jenen Vorstellungswiderstand hervor. Erst wenn wir genauer über Abtreibungsmoral nachdenken und einen festen Standpunkt hierzu entwickeln, beginnen wir es problematisch zu finden, uns eine andere Abtreibungsmoral als richtig vorzustellen. Der Vorstellungswiderstand geht nur mit *eingesehener* moralischer Falschheit einher, nicht bereits mit moralischer Falschheit als solcher. Doch auch für einen auf eingesehene moralische Falschheit beschränkten Vorstellungswiderstand scheint mir die

beste Erklärung ebenjene fehlende Weltrelativität von Werturteilen zu sein, die wir bei Gedankenexperimenten in der praktischen Philosophie nicht aus dem Blick verlieren dürfen.

Insbesondere das Phänomen des Vorstellungswiderstands macht auch einen vielleicht anfänglich verwirrenden Aspekt verständlicher, nämlich dass es sich mit (hinreichend allgemein formulierten) Moralgesetzen grundlegend anders verhält als mit Naturgesetzen. Diese gelten in unserer Welt und den möglichen Welten, die unserer verhältnismäßig ähnlich sind, in vielen anderen dagegen nicht, während jene, wie ich gesagt habe, in ausnahmslos allen möglichen Welten gelten. Passend dazu können wir uns Welten mit ganz anderen Naturgesetzen – etwa solche ohne Gravitation oder mit Zeitreisen in die Vergangenheit – problemlos vorstellen, während uns Welten mit ganz anderen Moralgesetzen – etwa solche, in denen es generell moralisch gut ist, anderen zu schaden, und generell moralisch schlecht, anderen zu helfen – nicht in den Kopf wollen. Dies ist ein Hinweis darauf, dass Letztere nicht metaphysisch möglich sind.

Darüber hinaus gibt es keine Gründe zu denken, in Welten, in denen völlig andere Naturgesetze herrschen, könnten auch völlig andere Moralgesetze herrschen. Wissen über Moralgesetze ist Wissen *a priori*, kein Wissen *a posteriori*: Man muss nicht die spezifische Beschaffenheit der Welt vor Augen haben, um es zu erlangen. Man muss insbesondere weder die geltenden Naturgesetze noch die konkreten Ausprägungen und Organisationsformen intelligenten Lebens kennen. In Fällen, in denen es anders zu sein scheint – etwa, weil Fakten darüber, ob die Personen in einer Welt mehrheitlich masochistisch veranlagt sind oder nicht relevant erscheinen für die Beurteilung der moralischen Verwerflichkeit von Folter – haben wir es lediglich mit nicht hinreichend allgemein gefassten Prinzipien zu tun, in denen beispielsweise statt von einem neutral verstandenen Schaden die Rede ist von Folter oder dem Zufügen von Schmerzen, was nicht in jedem denkbaren Fall negativ sein muss.

Dass hinreichend allgemein formulierte Moralgesetze weltunabhängig gelten, wird im Übrigen nicht dadurch konterkariert, dass in der Ethik Dissens hinsichtlich der Frage herrscht, welche Moralgesetze denn nun bestehen. Ähnlich wie zum Beispiel Erkenntnistheoretikerinnen, die über die beste Analyse von Wissen streiten, haben auch Moralphilosophinnen noch keine Moraltheorie als die offenkundig korrekte etablieren können. Und genauso wenig, wie das Bestehen einer Meinungsverschiedenheit bezüglich des Wesens von Wissen ein Hinweis auf eine eingeschränkte Geltung einer Wissensanalyse ist – sicherlich ist Wissen beispielsweise nicht in manchen

möglichen Welten kontextualistisch und in anderen nicht –, sind unsere Moraldissense ein Hinweis auf eine eingeschränkte Geltung der Moralgesetze.

Ich habe mich in diesem Abschnitt weitgehend auf moralische Werturteile konzentriert und ästhetische und entscheidungstheoretische allenfalls gestreift. Meine Argumentation dafür, dass Gedankenexperimente, in denen die Geltung tatsächlich nicht geltender moralischer Prinzipien (oder die Nichtgeltung tatsächlich geltender) vorausgesetzt wird, fehlerhaft sind und somit als Argumente nicht funktionieren, lässt sich genau dann auf entscheidungstheoretische oder ästhetische Gedankenexperimente übertragen, wenn Prinzipien rationalen Handelns bzw. ästhetischen Gehalt genauso wenig weltrelativ sind wie Moralgesetze. Denn dann würde man nicht einfach nur etwas Falsches voraussetzen, wenn man kontrafaktischerweise die Geltung oder Nichtgeltung bestimmter entscheidungstheoretischer oder ästhetischer Prinzipien annähme – das wäre harmlos –, sondern etwas metaphysisch Unmögliches, und das darf in Gedankenexperimenten, der Strukturanalyse aus den Abschnitten 2 und 3 zufolge, nicht geschehen.

Die Gründe, die man dafür anführen könnte, entscheidungstheoretische und ästhetische Werturteile als weltunabhängig anzusehen, sind dieselben wie im Falle moralischer Werturteile: Durch eine hinreichend allgemeine Formulierung von Prinzipien oder auch Einzelfallurteilen verflüchtigt sich der Eindruck, das Werturteil gelte nur unter bestimmten Bedingungen; wir erfahren einen Vorstellungswiderstand bei erkanntermaßen falschen Werturteilen; und Wissen über die zugrunde liegenden Gesetze ist Wissen *a priori*. Ob diese Gründe im Falle entscheidungstheoretischer und ästhetischer Werturteile auch dieselbe Überzeugungskraft haben wie im Falle moralischer, ist nicht einfach zu beurteilen. So scheint mir zum Beispiel der nötige Allgemeinheitsgrad in den Formulierungen ästhetischer Prinzipien schwieriger zu erreichen als bei moralischen. Dennoch weist insbesondere das Phänomen des Vorstellungswiderstands auf die Weltunabhängigkeit entscheidungstheoretischer und ästhetischer Werturteile hin: Eine rationale Handlung können wir uns kaum irrational denken, etwas ästhetisch Gehaltvolles kaum ästhetisch belanglos – und wenn doch, dann ist offenkundig die fehlende Allgemeinheit des Werturteils der Grund. Ich habe beispielsweise eben davon gesprochen, dass wir uns Goethes *Faust* nicht als ästhetisch wertlos vorstellen können *in einer Welt, die wie unsere ist*. In einer Welt, in der es die Menschen und ihre Kulturgeschichte nicht gäbe, ginge das vielleicht. Aber das läge dann daran, dass es sich um ein ästhetisches Einzelfallurteil handelt, in dem von spezifischen kulturellen Gegebenheiten schlecht

abstrahiert werden kann. Allgemeine ästhetische Prinzipien, in denen solche Einzelfallurteile mutmaßlich gründen (wie auch immer diese Prinzipien aussehen mögen), können wir uns dagegen nicht als mal richtig und mal falsch vorstellen. Die fehlende Weltrelativität von Werturteilen scheint daher nicht auf moralische Werturteile beschränkt zu sein.

5. Ein Beispiel zum Ausklang

Angenommen, es ist moralisch unbedenklich, Fleisch zu essen (unter Bedingungen, wie sie in unserer Welt und zu unserer Zeit in Mitteleuropa herrschen). Würde jemand, der unsicher ist, ob der bewusste Verzehr von Fleisch moralisch bedenklich ist, Fleisch aber gleichwohl isst, dann etwas moralisch Kritikwürdiges tun? Falls ja – und wir scheinen dieser Antwort zuzuneigen –, spräche das dafür, dass für die moralische Bewertung einer Handlung nicht nur die moralischen Fakten relevant sind, sondern auch die moralischen Überzeugungen des Handelnden.

Nach dem im letzten Abschnitt Gesagten ist allerdings die Annahme, es sei moralisch unbedenklich, Fleisch zu essen, an sich schon problematisch. Denn falls es tatsächlich unter den genannten Bedingungen moralisch bedenklich wäre, Fleisch zu essen, wäre dies in allen möglichen Welten so, und dann ließe sich nicht einfach das Gegenteil annehmen. Zeigen meine Überlegungen also, dass die kurze Argumentation aus dem vorherigen Abschnitt, die ja nichts anderes ist als ein mit nur wenig Pomp daherkommen-des Gedankenexperiment, auf einer womöglich falschen Annahme beruht? Und wäre das nicht seltsam angesichts der offenkundigen metaphysischen Harmlosigkeit der Argumentation?

Beide Fragen sind zu bejahen. Ja, wir dürfen nicht einfach um des Arguments willen annehmen, dass es moralisch unbedenklich ist, Fleisch zu essen (es sei denn natürlich, wir hätten gute Gründe, dies zu glauben, und dann dürften wir nicht das Gegenteil annehmen). Diese Annahme könnte metaphysisch unmöglich sein, und es sind genau solche auf verdeckte Weise metaphysisch unmöglichen Annahmen, deren Gefahr für Gedankenexperimente in der praktischen Philosophie ich hier betonen will. Und ja, es wäre seltsam, wenn jene Argumentation wegen einer zugrunde liegenden metaphysischen Unmöglichkeit nicht aufginge. Glücklicherweise ist das aber nicht der Fall; dieses kleine Gedankenexperiment funktioniert (oder falls nicht, dann nicht wegen der metaphysischen Unmöglichkeit des zugrunde liegenden Szenarios).

In diesem speziellen Fall ist der wunde Punkt nämlich lediglich die Präsentation des Gedankenexperiments. Die Annahme, dass es moralisch unbedenklich sei, Fleisch zu essen, ist zwar problematisch, ist für das Argumentationsziel allerdings auch nicht vonnöten. Denn das Wesentliche ist hier nicht, ob es *de facto* moralisch unbedenklich ist, Fleisch zu essen, sondern ob nur der Hinweis auf moralische Fakten eine berechtigte Kritik darstellt oder bereits der Hinweis auf die Unsicherheit des epistemischen Subjekts hinsichtlich der moralischen Fakten.

Man kann also das Gedankenexperiment einfach folgendermaßen präsentieren: Wäre es berechtigt, jemanden, der unsicher ist, ob der bewusste Verzehr von Fleisch moralisch bedenklich ist, Fleisch aber gleichwohl isst, mit der Begründung zu kritisieren, dass er ja nicht ausschließen könne, etwas moralisches Falsches zu tun? Falls ja – und wir scheinen dieser Antwort zuzuneigen –, spräche das dafür, dass für die moralische Bewertung einer Handlung nicht nur die moralischen Fakten relevant sind, sondern auch die moralischen Überzeugungen des Handelnden.

Auch wenn man nicht einfach die Geltung umstrittener oder als falsch erkannter moralischer Prinzipien im Rahmen eines Gedankenexperiments annehmen darf, gefährdet nicht jede solche Annahme das eigentliche Argumentationsziel. (Außerhalb des Rahmens eines Gedankenexperiments kann es, nebenbei bemerkt, durchaus harmlos sein, die Geltung eines tatsächlich falschen moralischen Prinzips anzunehmen. Man kann sich zum Beispiel überlegen, was logisch aus einem solchen Prinzip folgt, genauso wie ein Mathematiker sich überlegen kann, was aus der Falschheit der Goldbach-Vermutung folgen würde, unabhängig davon, ob diese Vermutung wahr oder falsch ist. Kontrafaktisches Denken lässt sich auch dann durchführen, wenn die als wahr angenommenen Sachverhalte mit metaphysischer oder gar logischer Notwendigkeit falsch sind.)

Die fehlende Weltrelativität von Werturteilen wird bisweilen übersehen und mag dann dazu verleiten, irrigerweise die Koexistenz verschiedener möglicher Welten anzunehmen, in denen unterschiedliche moralische, ästhetische oder entscheidungstheoretische Prinzipien gelten. Gemäß der in Abschnitt 2 und 3 vorgestellten Strukturanalyse kommt Gedankenexperimenten, denen eine solche Koexistenzannahme zugrunde liegt, keine epistemische Signifikanz zu. Das Problem dürfte allerdings nur wenige Gedankenexperimente betreffen; ein Beispiel ist das auf dem Szenario ABTREIBUNG beruhende Gedankenexperiment in Abschnitt 4. Ein weiteres Beispiel habe ich gerade geschildert, habe aber angemerkt, dass in diesem Fall der Fehler

leicht behoben werden kann. Darüber hinaus besteht bei Gedankenexperimenten in der theoretischen Philosophie, in denen es zumeist nicht um die angemessene Beurteilung eines Szenarios, sondern um dessen angemessene Beschreibung geht, in aller Regel wenig Gefahr, ein metaphysisch unmögliches Szenario zugrunde zu legen, zum einen, weil in den hier relevanten Szenarien die Gültigkeit bestimmter Werturteile so gut wie nie angenommen wird und sich daher ein eventuelles Übersehen von Weltunabhängigkeit nicht auswirkt, und zum anderen, weil wir, wenn keine Werturteile involviert sind, meist schnell merken, ob etwas metaphysisch Unmögliches vorausgesetzt wird. Die hier identifizierte Fehlerquelle betrifft somit vor allem Gedankenexperimente in Moral, Ästhetik und Entscheidungstheorie und markiert dadurch einen Unterschied zwischen Gedankenexperimenten in diesen und solchen in anderen philosophischen Teilgebieten. Gleichwohl ist dieser Unterschied eher peripherer Natur; grundsätzlich, so können wir festhalten, funktionieren Gedankenexperimente in der praktischen Philosophie auf dieselbe Weise wie in der theoretischen und haben dieselbe epistemische Berechtigung.

Literatur

- [Cappelen 2012] Cappelen, Herman: *Philosophy Without Intuitions*. Oxford, 2012.
- [Chalmers 2002] Chalmers, David: „Does Conceivability Entail Possibility?“ In: Gendler, Tamar S., und John Hawthorne (Hrsg.): *Conceivability and Possibility*. Oxford, 2002, S. 145–200.
- [Dancy 1985] Dancy, Jonathan: „The Role of Imaginary Cases in Ethics.“ In: *Pacific Philosophical Quarterly* 66 (1985), Nr. 1–2, S. 141–53.
- [Galilei 1638/1890] Galilei, Galileo: *Unterredungen und mathematische Demonstrationen über zwei neue Wissenszweige, die Mechanik und die Fallgesetze betreffend*. Übersetzt von Arthur von Oettingen. Leipzig, 1890.
- [Gendler 1998] Gendler, Tamar Szabó: „Galileo and the Indispensability of Scientific Thought Experiment.“ *The British Journal for the Philosophy of Science* 49 (1998), Nr. 3, S. 397–424.
- [Gendler 2000a] Gendler, Tamar Szabó: „The Puzzle of Imaginative Resistance.“ *Journal of Philosophy* 97 (2000), Nr. 2, S. 55–81.
- [Gendler 2000b] Gendler, Tamar Szabó: *Thought Experiment: On the Powers and Limits of Imaginary Cases*. New York, 2000.
- [Gendler 2007] Gendler, Tamar Szabó: „Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.“ *Midwest Studies in Philosophy* 31 (2007), Nr. 1, S. 68–89.

- [Grundmann und Horvath 2014] Grundmann, Thomas, und Joachim Horvath: „Thought Experiments and the Problem of Deviant Realizations.“ *Philosophical Studies* 170 (2014), Nr. 3, S. 525–33.
- [Häggqvist 2009] Häggqvist, Sören: „A Model for Thought Experiments.“ *Canadian Journal of Philosophy* 39 (2009), Nr. 1, S. 55–76.
- [Hume 1964/1757] Hume, David: „Of the Standard of Taste.“ S. 266–84. In: T. H. Grose (Hrsg.): David Hume. *The Philosophical Works*. Vol. 3, Aalen, 1964/1757.
- [Ichikawa und Jarvis 2009] Ichikawa, Jonathan, und Benjamin Jarvis: „Thought-Experiment Intuitions and Truth in Fiction.“ *Philosophical Studies* 142 (2009), Nr. 2, S. 221–46.
- [Jackson 1991] Jackson, Frank: „Decision-Theoretic Consequentialism and the Nearest and Dearest Objection.“ *Ethics* 101 (1991), Nr. 3, S. 461–82.
- [Kuhn 1964/1977] Kuhn, Thomas: „Eine Funktion für das Gedankenexperiment.“ In: *Die Entstehung des Neuen*. Frankfurt/Main 1977, S. 327–56. Die Originalversion findet sich in *L’Aventure de la Science, Mélanges Alexandre Koyré*, Bd. 2 (Paris 1964, S. 307–34); sie wurde wiederabgedruckt in *The Essential Tension* (Chicago 1977, S. 240–65).
- [Malmgren 2011] Malmgren, Anna-Sara: „Rationalism and the Content of Intuitive Judgements.“ *Mind* 120 (2011), Nr. 478, S. 263–327.
- [Norton 1996] Norton, John: „Are Thought Experiments Just What You Thought?“ *Canadian Journal of Philosophy* 26 (1996), Nr. 3, S. 333–66.
- [Nozick 1974] Nozick, Robert: *Anarchy, State, and Utopia*. New York, 1974.
- [Parfit 1984] Parfit, Derek: *Reasons and Persons*. Oxford, 1984.
- [Rawls 1975] Rawls, John: *Eine Theorie der Gerechtigkeit*. Frankfurt/Main, 1975.
- [Russell 1948] Russell, Bertrand: *Human Knowledge: Its Scope and Limits*. New York, 1948.
- [Sorensen 1992] Sorensen, Roy: *Thought Experiments*. Oxford, 1992.
- [Tuna 2020] Tuna, Emine Hande: „Imaginative Resistance.“ In: Zalta, Edward N. (Hrsg.): *The Stanford Encyclopedia of Philosophy*. (Summer 2020 Edition.) URL = <https://plato.stanford.edu/archives/sum2020/entries/imaginative-resistance>
- [Walton 1994] Walton, Kendall: „Morals in Fiction and Fictional Morality, I.“ *Proceedings of the Aristotelian Society*, supp. vol. 68 (1994), S. 27–50.
- [Weatherson 2019] Weatherson, Brian: *Normative Externalism*. Oxford, 2019.
- [Weber 2014] Weber, Marc Andree: *Die Zerlegung des Ichs. Über die Grundlagen personaler Identität*. Münster, 2014.
- [Weber 2017] Weber, Marc Andree: „Die Aussagekraft wirklichkeitsferner Gedankenexperimente für Theorien personaler Identität.“ In: A. Oberprantacher und A. Siegetsleitner (Hrsg.): *Mensch sein – Fundament, Imperativ oder Floskel? Beiträge zum 10. Kongress der Österreichischen Gesellschaft für Philosophie*. Innsbruck, 2017, S. 493–503.

[Williamson 2007] Williamson, Timothy: *The Philosophy of Philosophy*. Oxford, 2007.

[Yablo 1993] Yablo, Stephen: „Is Conceivability a Guide to Possibility?“ *Philosophy and Phenomenological Research* 53 (1993), Nr. 1, S. 1–42.