



meta(φ)

Methode und Wirklichkeit

Was ist Leben?

Marco R. Schori

Gemeinsamer Nenner? Die Konzept-Konzeptionen-Unterscheidung bei John Rawls und Ronald Dworkin

Fabio G. Briante

A Trilemma in the Philosophy of Scientific Explanation

Andreas Lüchinger

Ein bayesianischer Lösungsansatz für das Raben-Paradox

Simon Richmond

Thought Experiments in Ethics

Georg Brun

Moral Philosophers in Wonderland: How Outlandish May Ethical Thought Experiments Be?

Maurice Grüter

Zwischen Eigentlichkeit und Uneigentlichkeit: Nachweis einer graduellen Unterscheidung anhand von Heideggers Befindlichkeitstheorie

Omar Ibrahim

Herausgeberin

Fachschaft Philosophie der Universität Bern,
Länggassstrasse 49a, 3012 Bern, Schweiz
mit Unterstützung des Instituts für Philosophie der Universität Bern.

Redaktion

meta(φ)
Fachschaft Philosophie, Universität Bern
metaphi.philo@lists.unibe.ch

Chefredaktion

Céline Hoog

Lektorat

Joannes B. Campell, Michael Müller, Leona Sager,
Xenia Schmidli

Produktion und Gestaltung

Michael Müller

Coverbild: Timo Junger

ISSN

Gedruckte Ausg.: meta(φ) ISSN:2297-9948
Online-Ausg.: meta(φ) [Elektronische Ressource] ISSN:2297-9956

Text-Einreichungen

Einreichungen zur Veröffentlichung sind erwünscht an metaphi.philo@lists.unibe.ch. Die Auswahl wird vom Redaktions-Team entsprechend dem Themenschwerpunkt, der Relevanz und des Schreibstils vorgenommen. Es besteht kein Anrecht auf Veröffentlichung, weder bei erfolgreicher Einreichung noch nach Anfrage durch ein Redaktions-Mitglied. In diesem Magazin veröffentlichte Texte gelten als wissenschaftliche Publikation und sind somit zitierfähig.

Erste Worte

Liebe Leserin, lieber Leser

Die Philosophie ist eine besondere Wissenschaft. Dies lässt sich bereits daran erkennen, dass es im Vergleich zu anderen Wissenschaften äusserst schwierig scheint, ihr einen Gegenstandsbereich zuzuschreiben. Es gibt nichts, was nicht Gegenstand von philosophischem Fragen und Denken sein könnte. Insofern kennt die Philosophie keine Grenzen. Vielmehr überschreitet sie Grenzen, indem sie beispielsweise in der Metaphysik danach fragt, was es über das Seiende hinaus gibt. Dies bedeutet jedoch keineswegs, dass die Philosophie wirklichkeitsfern ist. Philosophinnen und Philosophen setzen sich mit der Wirklichkeit auseinander, sie hinterfragen scheinbar Selbstverständliches, leisten Begriffsarbeit und konstruieren Gedankenexperimente, um unser Verhalten und Denken in Bezug auf die Wirklichkeit besser zu verstehen.

Diese Ausgabe handelt von Wirklichkeit und der Methode des philosophischen Tuns. Um Dasein und Wirklichkeit geht es bei Omar Ibrahims Arbeit «Zwischen Eigentlichkeit und Uneigentlichkeit», in der er, anhand Heideggers Befindlichkeitstheorie, die Eigentlichkeit als graduelle Eigenschaft nachweist. Einen Einblick in den philosophischen Umgang mit Begriffen liefern zum einen Marco Schori, der sich in seinem Essay mit dem Begriff des Lebens und damit, wie dieser in der Biologie aufgefasst werden kann, auseinandersetzt, und zum anderen Fabio Briante, der eine Untersuchung der Konzept-Konzeptionen-Unterscheidung bei John Rawls und Ronald Dworkin vor-

nimmt und somit zwischen einem Begriff und dessen Dimensionen differenziert.

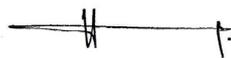
Mit Erklärungen der Wirklichkeit setzen sich Andreas Lüchinger und Simon Richmond auseinander. In «A Trilemma in the Philosophy of Scientific Explanation» erläutert Andreas Lüchinger die Unvereinbarkeit von genuin mathematischen Erklärungen mit wissenschaftlichen Erklärungen, wenn man letztere als Kausalerklärungen auffasst. Simon Richmond zeigt in seiner Arbeit auf, wie das Raben-Paradox mittels eines bayesianischen Lösungsansatzes gelöst werden kann. Und schliesslich widmet sich der dritte Teil dieser Ausgabe einer, insbesondere in der Ethik zentralen Methode philosophischen Arbeitens: dem Gedankenexperiment. Georg Brun erläutert als Einführung in seinem Essay die Struktur und Funktion von Gedankenexperimenten sowie verschiedene Einwände gegen den Gebrauch von jenen in der Ethik. Anschliessend diskutiert Maurice Grüter in «Moral Philosophers in Wonderland» zwei dieser Einwände.

Nun wünschen wir Ihnen eine vielfältige, inspirierende sowie spannende Lektüre und hoffen, Sie erhalten einen Einblick in Gegenstand und Methode des philosophischen Tuns am Institut für Philosophie der Universität Bern.

Herzlichst, Céline Hoog
im Namen des Redaktionsteams.



Leona Sager



Céline Hoog



Michael Andreas Müller



Joannes B. Campell



Xenia Schmidli

«Die Philosophie ist eine Art Rache an der
Wirklichkeit.»

Friedrich Nietzsche
Der Wille zur Macht, aus dem Nachlass 1884/88

«Die Wirklichkeit von der wir sprechen können, ist
nie die Wirklichkeit an sich.»

Werner K. Heisenberg
Ordnung der Wirklichkeit, 1989

Inhaltsverzeichnis

Begriff und Wirklichkeit

Was ist Leben? – Marco R. Schori	6
Gemeinsamer Nenner? Die Konzept-Konzeptionen-Unterscheidung bei John Rawls und Ronald Dworkin – Fabio G. Briante	9

Erklärung und Wirklichkeit

A Trilemma in the Philosophy of Scientific Explanation – Andreas Lüchinger	22
Ein bayesianischer Lösungsansatz für das Raben-Paradox – Simon Richmond	31

Gedankenexperimente und Wirklichkeit

Thought Experiments in Ethics – Georg Brun	40
Moral Philosophers in Wonderland: How Outlandish May Ethical Thought Experiments Be? – Maurice Grüter	51

Dasein und Wirklichkeit

Zwischen Eigentlichkeit und Uneigentlichkeit: Nachweis einer graduellen Unterscheidung anhand von Heideggers Befindlichkeitstheorie – Omar Ibrahim	58
--	----

Was ist Leben?

Die Definition des Lebensbegriffs ist im wissenschaftlichen Diskurs in mehrfacher Weise interessant. Naturwissenschaften wie die Biologie beziehen sich dabei stärker auf die deskriptiven Komponenten des Begriffs. In Disziplinen wie der Ethik wird hingegen eher der normative Aspekt des Lebensbegriffs hervorgehoben.¹ Daher ist nicht nur die Biologie an einer allgemeinen Definition, welche gleichzeitig eine brauchbare Unterscheidung zwischen belebten und unbelebten Gegenständen ermöglicht, interessiert. Im vorliegenden Essay wird ein Definitionsvorschlag des Lebensbegriffs von Georg Toepfer aus *Philosophie der Biologie. Eine Einführung*² diskutiert. Zudem werden nur die deskriptiven Aspekte des Begriffs behandelt, da diese für die Biologie von zentraler Bedeutung sind und auch im Referenztext eingehender behandelt werden. In einem ersten Schritt wird die Kritik von Toepfer an vorangehenden Definitionen erläutert. Danach wird der Vorschlag des Referenztextes ausgeführt. Nach der Darstellung der Argumente von Toepfer wird der Definitionsvorschlag von Toepfer zuerst im Hinblick auf seine Kritik an vorangehenden Überlegungen betrachtet. In einem vierten Schritt wird untersucht, ob der Lebensbegriff von Toepfer zu eng oder zu weit gefasst ist. Zum Schluss werden die Ergebnisse in einem Fazit zusammengefasst.

Nach G. Toepfer werde im Verlauf der Geschichte der Wissenschaften der Begriff des Lebens an einzelnen oder mehreren Merkmalen festgemacht. Zudem sei man von Eigenschaften von lebendigen Gegenständen ausgegangen. Prominente Merkmale von Lebendigkeit seien zum Beispiel die Selbstbewegung, Selbsterhaltung und Organisation gewesen.³ Eine Definition mit einem bestimmenden Merkmal oder Prinzip sei jedoch aus der Sicht der modernen Biologie nicht adäquat. Das Prinzip der Selbstbewegung würde demnach auch Himmelskörper oder fließende Gewässer als lebendige

Gegenstände miteinbeziehen.⁴ In der Thermodynamik sei das Prinzip der Selbsterhaltung auch anzutreffen. Thermodynamische Prozesse seien jedoch auch bei unbelebten Gegenständen vorhanden.⁵ Die Organisation als leitendes Prinzip des Lebensbegriffs würde auch Maschinen umfassen. Kant habe die Organisation in die Selbstorganisation umgedeutet, um dieser Kritik gerecht zu werden. Demnach sei ein Lebewesen „ein organisiertes und sich selbst organisierendes Wesen“⁶. Doch diesem Kriterium würde zum Beispiel auch das Ökosystem gerecht werden, da es ein System ist, welches in kleinere, voneinander abhängige Einheiten unterteilt werden könne. Nach diesen Überlegungen sei die Angabe von einem Prinzip oder Merkmal für die Definition des Lebensbegriffs zu weit und würde auch unbelebte Gegenstände miteinschließen.⁷

Ein weiterer Kritikpunkt von Toepfer an bestehenden Definitionen des Lebensbegriffs ist, dass das Auflisten von mehreren Merkmalen oder Prinzipien zum einen willkürlich und nicht systematisch ist, und zum anderen nicht von einem Ursprung des Lebens gesprochen werden kann.⁸ Wenn eine Liste von Merkmalen den Begriff des Lebens definiere, müsse erläutert werden, weshalb gerade diese Merkmale aufgelistet werden, um den Vorwurf der Willkür zu vermeiden. Zudem sei es auch nötig, zu belegen, weshalb diese Prinzipien miteinander bestünden und welches Verhältnis sie zueinander hätten.⁹ Da die aufgelisteten Merkmale nicht zeitgleich in Erscheinung getreten seien, wäre es zudem nicht möglich, von einem Ursprung des Lebens zu sprechen, vielmehr müsste man von mehreren Ursprüngen sprechen.¹⁰

Georg Toepfer definiert den biologischen Lebensbegriff als „[...] eine Seinsweise von (Natur-)Gegenständen, die sich durch Organisation, Regulation und Evo-

1 G. Toepfer, 2005, Der Begriff des Lebens, in: Krohs, Ulrich / G. Toepfer (Hrsg.): *Philosophie der Biologie. Eine Einführung*, Frankfurt am Main, Suhrkamp Verlag, S. 170.

2 ib., S. 157 – 174.

3 ib., S. 160f.

4 ib., S. 160.

5 ib., S. 162.

6 ib., S. 161.

7 ib., S. 162.

8 ib., S. 163f. und S. 166.

9 ib., S. 165.

10 ib., S. 163f.

lution auszeichnet [...]“¹¹. Die Reihenfolge der Begriffe Organisation, Regulation und Evolution sei von Bedeutung. Die Organisation sei demnach das erste Prinzip, da dies eine Voraussetzung für die Regulation und auch für die Evolution sei.¹² Ein Gegenstand, der sich selbst erhält, sollte als eine Einheit gedacht werden, da ansonsten die Organisation der interdependenten Einzelteile nicht erkennbar würde. Zudem beschreibe die Regulation das Verhältnis der einzelnen Bestandteile eines Gegenstandes, sowie das Verhältnis zu seiner Umgebung.¹³ Die Regulation stehe daher an zweiter Stelle. Da die Evolution, bzw. Fortpflanzung, von einem bestehenden Verständnis von Organismen (organisierte Einheit) und deren Fähigkeit zur Selbsterhaltung abhängig sei, komme die Evolution an dritter Stelle.¹⁴ Die Definition des biologischen Lebensbegriffs stütze sich demnach auf drei weitere Begriffe, welche durch Merkmale definiert sind. Nach Toepfer ist die Organisation die Einteilung eines Gegenstandes in mehrere Untereinheiten und Abläufe, welche voneinander abhängig sind. Die Regulation umfasse die Gesamtheit der notwendigen Vorgänge, welche für die Selbsterhaltung des (Natur-)Gegenstandes unabdingbar seien. Darunter versteht er zum Beispiel die Ernährung, den Schutz und die Koordination der Untereinheiten des Gegenstandes. Mit Evolution meint Toepfer die Fähigkeit Nachkommen zu generieren, welche über mehrere Generationen hinweg zu komplexeren Variationen des ursprünglichen Gegenstandes führen könne.¹⁵ Hält nun der Lebensbegriff von Toepfer seinen eigenen Kritikpunkten stand? Der erste seiner Kritikpunkte an vorangehenden Definitionsvorschlägen ist, dass der Gegenstandsbereich bei einer Definition mit nur einem Merkmal zu gross ist.¹⁶ Der Lebensbegriff von Toepfer stützt sich hingegen, wie oben dargestellt, auf drei Merkmale. Obwohl die Organisation, Regulation und Evolution Begriffe sind, die voneinander abhängig sind, umschreiben sie unterschiedliche Kriterien. Dadurch wird der Gegenstandsbereich stärker eingeschränkt als bei Definitionen, welche nur ein Prinzip als zentrales Unterscheidungsmerkmal verwenden. Der zweite Kritikpunkt von Toepfer betrifft die unterschiedlichen Definitionen des Begriffs mit mehreren Prinzipien. Dabei würden verschiedene Eigenschaften

aufgelistet, ohne einen systematischen Zusammenhang zu erläutern. Dies habe auch zur Folge, dass die Auswahl der Merkmale als willkürlich erscheine.¹⁷ Im Gegensatz dazu folgt der Definitionsvorschlag von Toepfer einem systematischen Aufbau. Dieser Aufbau leitet sich zudem auch von der begrifflichen Abhängigkeit der Prinzipien, wie auch von deren historischen Auftreten ab. Am Anfang steht die Organisation. Der Begriff der Organisation ermöglicht das Konzept des Organismus als Einheit. Danach kommt die Regulation. Diese setzt voraus, dass schon ein Gegenstand besteht (Organismus) welcher sich selbst erhalten muss. Zum Schluss gibt es noch das Prinzip der Evolution. Die Evolution und damit verbundene Fortpflanzung muss von einem bestehenden Gegenstand, welcher sich selbst reguliert, ausgehen.¹⁸ Demnach ist der Kriterienkatalog von Toepfer systematisch aufgebaut und entgeht der Kritik der Willkür. Der letzte Kritikpunkt, welcher aufgeführt wird, ist der, dass mit einer Liste von Kriterien nicht von einem Ursprung des Lebens gesprochen werden könne.¹⁹ Wird der Ursprung des Lebens als ein Moment in der Geschichte verstanden, ist dies auch mit dem Definitionsvorschlag von Toepfer der Fall. Es müsste demnach von drei Ursprüngen des Lebens gesprochen werden. Der erste wäre, als der erste Organismus entstand. Der zweite Ursprung wäre dann der Moment, als der erste Organismus zur Regulation imstande war. Und der dritte und letzte, sobald dieser Organismus, welcher zur Regulation imstande ist, Nachkommen generieren kann. Diese Auffassung ist jedoch nicht zutreffend, da der Ursprung des Lebens eher als ein Prozess gedacht werden muss. In diesem Fall würde auch der Definitionsvorschlag von Toepfer einen Ursprung des Lebens zulassen.

Wird der Ursprung des Lebens als Prozess verstanden, hält der Definitionsvorschlag von Toepfer den drei geschilderten Kritikpunkten stand. Nun wird noch untersucht, ob der Gegenstandsbereich des biologischen Lebensbegriffs von Toepfer zu eng oder zu weit gefasst ist. Da es sich um die Definition des biologischen Lebensbegriffs handelt, ist an dieser Stelle ein Beispiel aus der modernen Biologie zur Veranschaulichung angebracht: Viren werden in der modernen Biologie meist nicht als Lebewesen klassifiziert. Demnach sollten sie nicht unter den Begriff des Lebens von Toepfer fallen. Viren sind Partikel, welche ausserhalb der Zelle von einer Proteinhülle umgeben sind und im Innern liegen RNA- oder DNS-Moleküle vor. Dies ist zwar

11 ib., S. 169.

12 ib., S. 168.

13 ib., S. 167.

14 ib., S. 168.

15 ib., S. 169f., die Definitionen von Organisation, Regulation und Evolution sind sprachlich nahe am Referenztext, da dies für den weiteren Verlauf des Essays bedeutend ist.

16 ib., S. 162.

17 ib., S. 165.

18 ib., S. 169.

19 ib., S. 163f.

eine sehr einfache, aber doch eine Art von Organisation. Das erste Kriterium ist demnach erfüllt. Dringt der Virus in eine Wirtszelle ein, versorgt er sich mit den nötigen Stoffen zu seiner eigenen Replikation und seinem Erhalt. Solange der Virus ausserhalb der Zelle ist, schützt ihn die Proteinschicht. Der Übergang von ausserhalb der Zelle in das Innere erfordert eine Koordination der Proteinschicht des Virus. Demnach ist auch das Kriterium der Regulation erfüllt. Der Virus kann sich nicht selbstständig fortpflanzen. Er benötigt die Wirtszelle dazu. Trotzdem ist er in der Lage, ähnliche oder mutierte Nachkommen zu generieren. Auch Parasiten, welche gemeinhin als Lebewesen charakterisiert werden, können sich zum Teil nicht ohne einen entsprechenden Endwirt fortpflanzen. Aufgrund dessen ist der Gegenstandsbereich des Definitionsvorschlags von Toepfer zu weit. Nach diesen Kriterien würden auch Viren unter den Begriff des Lebens fallen. Um den biologischen Lebensbegriff weiter einzuengen, ist es möglich, den Metabolismus als abgrenzendes Merkmal miteinzubeziehen. Viren würden damit nicht mehr erfasst, da sie keinen Stoffwechsel machen. Ein Beispiel für einen (Natur-)Gegenstand, welcher in der Biologie als Lebewesen betrachtet wird und keinen Stoffwechsel aufweist, ist mir unbekannt. Es gibt zwei Möglichkeiten, um diesen Prozess in die Definition miteinzubeziehen. Einerseits in der Definition des biologischen Lebensbegriffs und andererseits in der Definition des Regulationsprinzips. Da das Regulationsprinzip die Eigenschaft des Selbsterhalts mit Hilfe von Schutz, Ernährung (als Stoffaufnahme aus der Umwelt) und Koordination und Integration umschreibt, wäre es sinnvoll den Metabolismus in diese Definition einzufügen. Toepfer definiert die Ernährung als „[...] Versorgung des Systems mit notwendigen Stoffen aus der Umwelt [...]“²⁰. Dies trifft auch auf Viren zu, da sie die Stoffe aus der umliegenden Zelle für sich und ihre Nachkommen verwenden. Weiter schliesst diese Definition des Regulationsprinzips den Metabolismus nicht aus, es können jedoch auch andere Prozesse als eine solche Versorgung mit Stoffen betrachtet werden. Georg Toepfer hat in seinem Text zum Begriff des Lebens zuerst die bestehenden Definitionsversuche in drei Punkten kritisiert. Der Gegenstandsbereich war zu weit gefasst, wenn nur ein Merkmal angegeben wurde. Listen von Kriterien für eine Definition waren unsystematisch und erschienen daher willkürlich. Ein Ursprung für das Leben kann nicht ausgemacht werden, vielmehr gäbe es dann mehrere Ursprünge. Der

20 ib., S. 169.

Definitionsvorschlag von Toepfer nimmt diese Kritik auf und bestimmt den biologischen Lebensbegriff als „[...] eine Seinsweise von (Natur-) Gegenständen, die sich durch Organisation, Regulation und Evolution auszeichnet [...]“²¹. Um dem letzten Kritikpunkt gerecht zu werden, muss der Ursprung des Lebens jedoch als Prozess gedacht werden. Die drei Begriffe, welche in der Definition des Lebensbegriffes vorkommen, werden von Toepfer definiert. Der Gegenstandsbereich des Definitionsvorschlags ist enger gefasst als die von ihm kritisierten Definitionen, jedoch umfasst er noch Gegenstände, denen die Biologie keine Lebendigkeit zuspricht. Dies kann umgangen werden, indem das Regulationsprinzip um das Merkmal des Metabolismus erweitert wird. Dieser ergänzte Definitionsvorschlag wird der allgemeinen Praxis in der Biologie gerecht.

21 ibd.

Marco Raphael Schori (29) schliesst gegenwärtig sein Bachelorstudium in Philosophie mit Nebenfach Sozialwissenschaften, Schwerpunkt Politikwissenschaften, ab. Er interessiert sich für antike Philosophie, Wissenschaftsphilosophie und Philosophie der Logik.

Gemeinsamer Nenner?

Die Konzept-Konzeptionen-Unterscheidung bei John Rawls und Ronald Dworkin

1. Einleitung

In dieser Arbeit verfolge ich die Frage, wie sich die Unterscheidung zwischen einem Konzept auf der einen und dessen Konzeptionen auf der anderen Seite, i.e. hier synonym verwendet, zwischen einem Begriff und dessen Auffassungen, verstehen lässt. Dabei werde ich auf eine bestimmte Version dieser Begriff-Begriffsauffassungs-Unterscheidung (BAU) fokussieren, welche sich primär auf deren Verwendung und Beschreibung in den Werken von John Rawls (1999 [1971]) und Ronald Dworkin (1977, 1986) stützt.¹

Insbesondere in der Rechtsphilosophie und der politischen Philosophie, wird die BAU meist im Zusammenhang mit Begriffsstreitigkeiten verwendet, in welchen darüber gestritten wird, was mit einem bestimmten normativen Begriff, wie z.B. GERECHTIGKEIT, RECHT oder FREIHEIT, genau gemeint ist (Kapitalchen zeigen im Folgenden stets die Rede von Begriffen an). Nun wird diese Unterscheidung zuweilen als ein nützliches Werkzeug ausgewiesen (vgl. z.B. Dworkin 1986: 71, Criley 2007, Lalumera 2014, Solum 2004). So meint Elizabeta Lalumera, dass die BAU eine angemessene Beschreibung von Phänomenen wie genuiner Uneinigkeit und Begriffsstreitigkeiten liefere (2014: 73) und Dworkin verknüpft mit der Unterscheidung einen strukturierenden Nutzen. So beschreibt er diese Unterscheidung als „imposing a certain structure on [a] community’s practice such that particular substantive theories can be identified and understood as subinterpretations of a more abstract idea“ (1986: 71).

Ungeachtet der Frage nach dem Nutzen, den eine solche Strukturierung stiften soll, werde ich in dieser Untersuchung zeigen, dass die BAU so wie sie von Rawls und Dworkin beschrieben wird, nicht zur rein deskriptiven Strukturierung von Begriffsstreitigkeiten genutzt werden kann. Hierzu werde ich im dritten Teil

dieser Arbeit die charakteristischen Behauptungen im Zusammenhang mit der BAU herausarbeiten, um im dritten Teil zu zeigen, dass die zentrale Annahme eines von allen Streitparteien geteilten Begriffskerns, auf zwei Arten verstanden werden kann. Wenn wir den Begriffskern als normatives Postulat verstehen, dann mischen wir uns beim Strukturieren eines entsprechenden Begriffsstreits inhaltlich in diesen ein. Wollen wir hingegen hinsichtlich der zu strukturierenden Debatte neutral bleiben, i.e. die BAU *deskriptiv* verwenden, dann können wir nicht an der bestehenden Beschreibung der BAU festhalten. Denn, wie ich im vierten Teil zeigen werde, droht die Tatsache, dass sich in realen Begriffsstreitigkeiten nicht zwingend nur *ein* Begriffskern identifizieren lässt, die Unterscheidung zum Verschwinden zu bringen – zumindest solange wir an der Grundannahme festhalten, wonach umstrittene Begriffe nur eine Bedeutung haben. Obwohl Dworkin dieser Annahme ohnehin skeptisch gegenüberzustehen scheint, scheint sie für viele, wie ich nun im folgenden ersten Teil erläutern werde, die BAU überhaupt erst zu motivieren.

2. Der konstante Bezug in Begriffsstreitigkeiten

Gemäss der Darstellung von Mark E. Criley (2007) wird in Begriffsstreitigkeiten um die Anwendungsregeln von Begriffen gerungen: „the issue that is under dispute is what the correct criteria of application are“ (2007: 8). Auch wenn ich hier Begriffsstreitigkeiten als Uneinigheiten hinsichtlich der *Bedeutung* von Begriffen beschreibe, nehme ich an, dass sich diese zwei Darstellungen grundsätzlich ineinander übersetzen lassen. Dies wird deutlich im ersten Abschnitt dieses Teils, wo mein Augenmerk zunächst jenen Annahmen über Begriffe gilt, von denen Rawls und Dworkin implizit auszugehen scheinen. Im Anschluss daran werde ich zeigen, welche Diagnose zu Begriffsstreitigkeiten sich aus einer bestimmten Auffassung zur Individuierung von Begriffen ergibt. Im letzten Abschnitt wird klar, dass sich aus der Ablehnung dieser Diagnose ein

¹ Der Grund für diese Fokussierung ist, dass oft auf die genannten Autoren verwiesen wird, wenn die BAU in der praktischen Philosophie verwendet, charakterisiert oder diskutiert wird (vgl. z.B. Swanton 1992, Criley 2007, Lalumera 2014, List & Valentini 2016). Im Folgenden werde ich auf Rawls’ Werk mit der originalen Jahreszahl verweisen, während sich die Seitenzahlen auf die Neuauflage von 1999 beziehen.

Problem ergibt, auf welches die BAU eine dringliche Antwort zu liefern scheint.

2.1. Grundannahmen über Begriffe

Was wir unter *concepts* genau verstehen sollen, ist Gegenstand reger Debatten in der Sprachphilosophie, der Philosophie des Geistes sowie im Zusammenhang mit Begriffsanalyse und es ist mir hier nicht möglich, die verschiedenen Auffassungen zu diskutieren. Daher möchte ich unter Bezugnahme auf Margolis & Laurence (2014) lediglich jene Annahmen über Begriffe herausstreichen, von denen Dworkin und Rawls auszugehen scheinen.

Begriffe werden hier als Bestandteile von Propositionen verstanden und können von unterschiedlicher Komplexität sein. Das heisst sie umfassen sowohl zusammengesetzte, wie z.B. AN EINEM FREITAG GESPIELTE SCHACHTURNIERE als auch einfache Begriffe wie z.B. TISCH. Nun wird oft angenommen, dass Begriffe einen Inhalt haben, i.e. aus einfacheren, untereinander logisch verknüpften Begriffen zusammengesetzt sind, wobei diese Bestandteile nicht zwingend als basale oder atomare Bestandteile verstanden werden müssen (vgl. Margolis & Laurence 2014: Kap.2). So könnte bspw. TISCHLÄUFER die Begriffe BAND und TISCHDEKORATION beinhalten.

Obwohl der Inhalt, i.e. der konzeptuelle Gehalt von Begriffen zuweilen mittels Definitionen angegeben wird, möchte ich hier offenlassen, ob Konzepte definitoren Charakter haben, i.e. ob deren konstituierende Bestandteile notwendige und zusammen hinreichende Bedingungen für die Klassifizierung von Objekten angeben. Dennoch ermöglicht der Inhalt eines Konzeptes, Gegenstände und Eigenschaften – egal ob abstrakte oder konkrete – zu klassifizieren und schränkt somit die Fälle gültiger Begriffs-Anwendungen ein (vgl. ebd., Criley 2007: 49). So können wir bspw. TISCHLÄUFER deshalb nicht auf eine Tischdecke anwenden, weil (oder wenn) diese kein Band ist. In diesem Sinne liefert uns der Inhalt von Konzepten Kriterien für deren Anwendung, i.e. Anwendungsregeln, die unsere Verwendung des Wortes anleiten.

In Begriffsstreitigkeiten geht es nun aber nicht lediglich um Streitigkeiten darüber, ob ein Wort auf ein bestimmtes, konkretes oder abstraktes Objekt anzuwenden ist oder nicht, sondern um die Uneinigkeiten darüber, *warum* gewisse Anwendungen richtig und andere falsch sind. Dies kann als Uneinigkeit hinsichtlich des konzeptuellen Gehaltes oder – wenn man diese Übersetzung zulassen will – als Streit um die

Bedeutung eines Begriffes verstanden werden.² Dabei gehe ich hier lediglich davon aus, dass wenn wir die Bedeutung eines Wortes propositional erfassen, dies auch als Angabe dessen konzeptuellen Gehaltes verstanden werden kann.

2.2. Begriffsstreitigkeiten und Begriffsindividuation

Die Freunde Anna, Bea, Chris und Dana befinden sich auf einem gemeinsamen Spaziergang im Park und diskutieren darüber, welche Tiere als Hunde gelten sollen und welche nicht. Sie streiten also nicht darum, wie HUND gemeinhin verwendet *wird*, sondern wie man diesen Begriff eigentlich verwenden *sollte*. Bei den Labradoren, Border Collies und Schäferhunden waren sich alle einig, dass dies Hunde sind. Bezüglich Gaia, einer Füchsin, die von Yanik als Haustier gehalten wird, waren nur Anna und Bea der Meinung, dass sie zu den Hunden gezählt werden sollte. Teacup-Hunde und Chihuahuas bezeichnen alle ausser Dana als Hunde und den Marder von Zora will niemand einen Hund nennen. Auf dem Rückweg läuft ihnen ein wilder Fuchs über die Strasse und Anna ruft Bea grinsend zu: „Schau Bea, ein Hund!“, woraufhin aber Bea meint: „Nein, *das* ist ein Fuchs.“

Offensichtlich verwenden die vier Freunde HUND mit einer je anderen Bedeutung. Dabei halten sie ihre unterschiedlichen Begriffsanwendungen aufgrund ihrer jeweils assoziierten Bedeutung für richtig. Da HUND für jede und jeden einen anderen Inhalt hat, teilen die vier Freunde nicht denselben Begriff. Wenn wir nämlich feststellen, dass sich der Inhalt zweier Begriffe unterscheidet, so müssten wir aufgrund von Identitäts-Überlegungen davon ausgehen, dass zwei verschiedene Begriffe vorliegen. Wenn z.B. für Bea HAUSTIER notwendigerweise zum Begriff HUND gehört, für Anna aber nicht, dann haben Anna und Bea nicht denselben Begriff davon, was Hunde sind, sondern haben die Begriffe HUND_(Anna) respektive HUND_(Bea) – und ihre Uneinigkeit bezüglich des wilden Fuchses lässt genau dies vermuten. Folglich sind im obigen Beispiel unterschiedliche Begriffe im Spiel und es stellt sich die Frage, inwiefern überhaupt vom selben gesprochen wird. Diese Überlegungen entsprechen in etwa folgender Position zur Begriffsindividuation, i.e. zur Frage, wann zwei Begriffe dieselben sind und wann sie zwei unterschiedliche Begriffe darstellen:

2 Damit will ich mitnichten behaupten, dass der Inhalt von Begriffen propositionaler Natur sein muss, sondern nur, dass deren Gehalt oder deren Intension oftmals propositional erfasst wird. Wenn die Struktur und die klassifikatorische Leistung von Konzepten bspw. mittels Prototypen verstanden wird (vgl. Margolis & Laurence 2014: Kap.2), dann bedürfte eine solche Übersetzung freilich der genaueren Überprüfung.

P1: Begriffe sind über deren Inhalt zu individualisieren, wobei für zwei Begriffe gilt: wenn sie einen Unterschied im begrifflichen Gehalt aufweisen, stellen sie unterschiedliche Begriffe dar.

Vor diesem Hintergrund müsste zugegeben werden, dass HUND für die vier Freunde genau genommen ein mehrdeutiger Begriff ist und durch eine Disambiguierung ihr Streit beigelegt werden kann. Dann würden die vier Freunde künftig zwischen folgenden vier Begriffen unterscheiden: HUND_(A), HUND_(B), HUND_(C) und HUND_(D). Tun sie dies nicht, laufen sie Gefahr, aneinander vorbeizureden, da sie mit ihren jeweiligen Begriffen über je andere Dinge sprechen.

Eine solche Diagnose für Begriffsstreitigkeiten lehnen nun aber verschiedene Autoren ab, da sie der Meinung sind, dass sich theoretische Begriffsstreitigkeiten nicht als ein Aneinander-Vorbeireden aufgrund unterschiedlicher Begriffe verstehen lassen (vgl. z.B. Dworkin 1986: 46; Criley 2007: 8; Lalumera 2014: 78). So sagt bspw. Lalumera zu einem Streit bezüglich der Bedeutung von SOFA folgendes:

“[O]ur disagreement with A could not be put to an end by disambiguation, that is, by simply admitting that we are using the term ‘sofa’ differently, or we are employing different concepts. Not at all: A wants to persuade us that he is right and we are wrong on the same topic, and our intuition is just that he is wrong.” (2014: 78)

Der Ausschluss von Ambiguität scheint im Kontext von Begriffsstreitigkeiten zentral zu sein. So formuliert Criley eine Eindeutigkeits-Bedingung („univocality condition“) für (echte) Begriffsstreitigkeiten: „A contested concept is a single concept, used univocally across a community over time, rather than a collection of several distinct or shifting concepts“ (2007: 8). Seiner Meinung nach gehen im Falle von Begriffsstreitigkeiten alle beteiligten Streitparteien davon aus, dass sie den umstrittenen Begriff teilen und kompetent im Umgang mit dem selbigen sind (vgl. ebd.). Das heisst, dass unsere vier Freunde davon ausgehen, dass sie denselben Begriff von HUND teilen, sich aber darüber streiten, was er genau bedeutet. Dies entspricht im Wesentlichen der folgenden Annahme:

P2: Umstrittene Begriffe sind (a) nicht mehrdeutig und (b) auch dann intersubjektiv geteilt, wenn um deren Bedeutung gestritten wird.

Nun unterscheiden sich aber trotzdem die Erklärungen dafür, *warum* gewisse Begriffsanwendungen richtig sein sollen. Und diese Erklärungen liefern die jeweiligen, mit dem Begriff assoziierten Bedeutungen. So hält Bea die Anwendung von HUND auf Gaia u.a. deshalb für richtig, weil Gaia ein Haustier ist. Dadurch ergibt sich das im Folgenden beschriebene Problem.

2.3. Dringlichkeit der Begriff-Begriffsauffassungs-Unterscheidung

Obwohl verschiedene Personen zwar durchaus mehr oder weniger ähnliche Begriffe haben mögen, teilen sie gemäss P1 nur dann den *gleichen* Begriff, wenn sie mit diesem auch den gleichen Inhalt verknüpfen. Wenn wir aber zudem an P2 festhalten wollen, ergibt sich der folgende Widerspruch: Es müsste behauptet werden, dass die Parteien in Begriffsstreitigkeiten zugleich denselben wie auch einen je unterschiedlichen Begriff von denselben Dingen haben.

Eine Lösung scheint nun die folgende Unterscheidung anzubieten. Wenn wir mit „Begriff“ das bezeichnen, was alle Streitparteien teilen, während wir die unterschiedlichen, damit verknüpften Inhalte „Auffassungen“ nennen, können wir sagen, dass Unterschiede in den *Begriffsauffassungen* nicht unterschiedlichen *Begriffen* entsprechen, sondern eben bloss unterschiedlichen *Auffassungen zu ein und demselben* Begriff. Das ist vermutlich des Rätsels Lösung, von der Lalumera implizit spricht, wenn sie sagt: „phenomena such as genuine disagreement and concept contestation can be better described in terms of both concepts and conceptions, and would otherwise remain implausibly puzzling“ (2014: 73). Damit teilen alle Parteien in einem Begriffsstreit denselben Begriff und reden in diesem Sinne vom selben. Aber wie können wir uns unterschiedliche Begriffsauffassungen erklären, wenn ein Begriff intersubjektiv geteilt ist?

Eine Antwort darauf liefert eine Position aus der Sprachphilosophie und linguistischen Theorie, der zufolge die Bedeutung von Begriffen und damit die Kriterien für deren korrekte Anwendung *objektiv* sind (vgl. bspw. Peacocke 1992, Higginbotham 1998^{a,b}, Lalumera 2014: 76). Damit ergibt sich die folgende Erklärung für P2:

P3: Die Bedeutung von Begriffen ist objektiv gegeben und intersubjektiv zugänglich.

Die unterschiedlichen Begriffsauffassungen entsprechen dabei verschiedenen, falsifizierbaren und mehr oder weniger vollständigen Hypothesen über die

(tatsächliche) Bedeutung eines Wortes (vgl. Higginbotham 1998^b: 153). Demnach würden die vier Freunde den selben Begriff teilen, weil die Bedeutung von HUND unabhängig von ihren jeweiligen Auffassungen existiert. Ihre unterschiedlichen Hypothesen über die tatsächliche Bedeutung, i.e. ihre jeweiligen Konzeptionen von HUND, können mehr oder weniger zutreffend sein – je nach Übereinstimmungsgrad mit der objektiven Bedeutung des Begriffes. Damit kann die BAU sowohl erklären warum Konzepte intersubjektiv geteilt sind – weil für alle Konzept-Nutzer eigentlich dieselbe Begriffsbedeutung gilt – als auch warum es trotzdem zu Begriffsstreitigkeiten kommt. So resümiert Maite Ezcurdia:

“An advantage to drawing the concept-conception distinction lies in its allowing for accounts of (a) the public or intersubjective character of concepts and (b) a certain normative aspect involved in possessing a concept, viz. the possibility of misapplying a concept one possesses.” (1998: 188)

Auch wenn diese Antwort auf die durch die BAU aufgeworfene Frage klar und simpel ist, so handelt man sich mit P3 unter anderem epistemische Fragen nach unserem Zugang zur tatsächlichen Bedeutung ein.³ Zudem gelangen wir zu einem Resultat, dem nicht alle zustimmen würden: Konzeptionen können wahr oder falsch sein. Das hiesse dann, dass wann immer sich zwei Personen hinsichtlich der Bedeutung eines Begriffes uneinig sind, mindestens eine von beiden falsch liegen muss. Obwohl viele mit diesem Ergebnis leben könnten, werde ich nun zeigen, dass der Ansatz von Rawls und Dworkin eine Alternative zu P3 darzustellen scheint. Die nun im Folgenden vorgestellte und diskutierte Version der BAU beinhaltet ebenfalls eine Antwort auf die Frage, wie wir P2 verstehen können. Das heisst, sie erklärt sowohl, wie es zu Begriffsstreitigkeiten kommen kann, wenn wir Begriffe doch teilen, als auch inwiefern wir Begriffe als von allen geteilt verstehen können, auch wenn um deren Bedeutung gestritten wird.

3. Charakteristika der Konzept-Konzeption-Unterscheidung

In diesem Teil wird das Augenmerk darauf gerichtet sein, wie sich die in den Beiträgen von Rawls (1971) und Dworkin (1977, 1986) verwendete und charak-

³ Higginbotham bspw. meint, dass wir grundsätzlich in der Lage sind Wortbedeutungen zu erfassen (vgl. 1998a) und Lalumera schreibt hierbei Experten – vermutlich auch Philosophen – eine besondere Rolle zu: „The idea would be that experts fix concepts, and possess the correct conception” (2014: 78).

terisierte Unterscheidung verstehen lässt. Da Rawls die BAU nur verwendend einführt und nicht gesondert betrachtet, werde ich dessen Ausführungen mit Dworkins Bemerkungen ergänzen und dabei fünf Behauptungen herausarbeiten, mittels derer sich das von dieser Version der BAU skizzierte Bild von Begriffsstreitigkeiten charakterisieren lässt.

3.1. Die grundsätzliche Einigkeit

Im Rahmen seines Einführungskapitels benutzt Rawls die BAU dazu, die Funktion der Prinzipien sozialer Gerechtigkeit zu bestimmen (1971: 6). Dabei meint er, dass gegenwärtigen Gesellschaften eine „public conception” von Gerechtigkeit fehle (Rawls 1971: 5). Rawls zufolge erfüllt eine solche in wohlgeordneten Gesellschaften die Funktion, die Wahl und/oder Ausgestaltung und fortlaufende Anpassung einer friedlichen Gesellschaftsordnung mittels allgemein anerkannter Gerechtigkeitsprinzipien anzuleiten. Zudem garantiere sie ein von allen Mitgliedern einer Gesellschaft geteiltes Einverständnis mit der durch diese Prinzipien bestimmten Verteilung von Vor- und Nachteilen sozialer Kooperation (vgl. Rawls 1971: 4-5, 12).

Obwohl nun die Streitparteien gegenwärtig ein je eigenes Set von Prinzipien, i.e. ihre je eigenen Konzeption von Gerechtigkeit vertreten, würden aber alle einsehen, dass es irgendeines solchen Sets bedarf um zu bestimmen, ob bspw. eine Institution oder eine soziale Praxis gerecht ist oder nicht (Rawls 1971: 5). Nun sind sich aufgrund der Funktion des Gerechtigkeitsbegriffes alle hinsichtlich des ungefähren Inhaltes von GERECHTIGKEIT einig:

“Those who hold different conceptions of justice can, then, still agree that institutions are just when no arbitrary distinctions are made between persons in the assigning of basic rights and duties and when the rules determine a proper balance between competing claims to the advantages of social life.” (Rawls 1971: 5)

GERECHTIGKEIT wird für Rawls dabei grundsätzlich durch die Funktion der Prinzipien definiert: eine nicht willkürliche Verteilung von Grundrechten und ein angemessenes Gleichgewicht zwischen widerstreitenden Forderungen sicherzustellen (1971: 9). Da er diese Ideen als vom Begriff beinhaltet („are included in”; 1971: 5) beschreibt, gehe ich hier davon aus, dass seiner Meinung nach die Bedeutung von GERECHTIGKEIT u.a. durch NICHT-VERTEILUNGS-WILLKÜR und ANGEMESSENES FORDERUNGS-GLEICHGEWICHT propositional er-

fasst werden kann.

Auch in Dworkins Beispiel, in welchem sich eine fiktive Gemeinschaft darum streitet, was Höflichkeit sei, gibt es eine Einigung auf einen begrifflichen Gehalt. Bei der Suche nach einer möglichst allgemeinen Beschreibung, der alle zustimmen können, werden sich die Mitglieder dieser Gemeinschaft einig, dass Höflichkeit etwas mit Respekt zu tun hätte. Dworkin meint dazu: „It would then be natural for people to regard [the tie between courtesy and respect] as special and in the way of conceptual, to say, for example, that respect is part of the «very meaning» of courtesy” (1986: 70).

In diesen Darstellungen gibt es also eine Überlappung zwischen den Konzeptionen, die sich auf gewisse Bedeutungsbestandteile eines umstrittenen Begriffes beziehen. Diese inhaltliche Überlappung wird zuweilen auch „Begriffskern“ genannt (vgl. z.B. Swanton 1992: 2-3) – eine Idee, die m.E. auch durch „Bedeutungskern“ ausgedrückt werden kann. Diese ersten Annahmen entsprechen folgender Behauptung:

B1: Ein umstrittenes Konzept enthält einen Bedeutungskern, i.e. konzeptuelle Bestandteile, die von allen Konzeptionen geteilt werden.

Damit lässt sich die Frage, inwiefern Konzepte intersubjektiv geteilt sind, vorerst wie folgt beantworten: verschiedene Personen teilen auch in Streitfällen ein und denselben Begriff, weil sie sich hinsichtlich eines Begriffskerns einig sind. Doch wie kommt es nun trotz dieser Einigkeit zu unterschiedlichen Begriffsauffassungen?

3.2. Gründe und Art der Uneinigkeit

Rawls' wie auch Dworkins Darstellungen scheinen von Herbert L. A. Hart (1961) inspiriert zu sein. So gibt Rawls bei seiner Verwendung der BAU explizit an, diesem zu folgen (vgl. Rawls 1971: 5, Ftn.1). Hart identifiziert bei seinen Untersuchungen auf den von Rawls angegebenen Seiten eine Behauptung, welche bei jedem Versuch den Rechtsbegriff zu analysieren, die zentrale Rolle spielen sollte (vgl. Hart 1961: 155). Und zwar, dass RECHT eine notwendige Verbindung mit MORAL, oder genauer: der Idee der Gerechtigkeit aufweise (Hart 1961: 156-57). Nun stellt sich aber die Frage, was denn diese konzeptuellen Bestandteile NOTWENDIGE VERKNÜPFUNG und MORAL – von Hart Schlüsselbegriffe genannt – *ihrerseits* genau bedeuten sollen. Denn: „There are many possible interpretations

of the key terms ‘necessary’ and ‘morality’” (Hart 1961: 156).

Durch die Tatsache, dass unterschiedliche Interpretationen der konzeptuellen Bestandteile möglich sind, ergibt sich laut Dworkin folgende baumartige Struktur von Begriffsstreitigkeiten:

“People by and large agree about the most general and abstract propositions about courtesy, which form the trunk of the tree, but they disagree about more concrete refinements or subinterpretations of these abstract propositions, about the branches of the tree.” (1986: 70)

Dabei stellen diese verschiedenen (Sub-)Interpretationen unterschiedliche Konzeptionen des Begriffes dar, wobei letzterer durch den Begriffskern – gemäss Dworkins Beispiel im Falle von HÖFLICHKEIT also RESPEKT – erfasst wird. So meint Dworkin: „respect provides the concept of courtesy and [...] competing positions about what respect really requires are conceptions of that concept” (1986: 71). Demnach lässt sich die Uneinigkeit wie folgt beschreiben:

B2: Die Uneinigkeit in Begriffsstreitigkeiten besteht lediglich darin, wie der Bedeutungskern eines Konzeptes zu interpretieren ist.

Nun wird sich zeigen, dass die Autoren Rawls, Dworkin und Hart davon ausgehen, dass die Bedeutung der konzeptuellen Bestandteile eines umstrittenen Begriffes offen ist und dass genau dies unterschiedliche Konzeptionen zum Begriff ermöglicht.

3.2.1. Bedeutungsoffenheit

Dass die Bedeutung des Begriffskerns offen gelassen wird, scheint sowohl für die Einigkeit, wie auch für die Uneinigkeit in Begriffsstreitigkeiten verantwortlich zu sein:

“Men can agree to this description of just institutions since the notions of an arbitrary distinction and of a proper balance, which are included in the concept of justice, are left open for each to interpret according to the principles of justice that he accepts.” (Rawls 1971: 5; Kursivierungen hinzugefügt)

Laut Dworkin muss ein Begriff genauer bestimmt werden, damit er auf einen gegebenen, strittigen oder schwierigen zu entscheidenden Fall angewendet werden

kann (vgl. 1977: 102-3). Denn das Konzept besitzt eine offene Textur („open texture“; Dworkin 1977: 102) und enthält lediglich unpräzise Anwendungsregeln („imprecise rules“; ebd.).⁴ Das heisst, sie lassen hinsichtlich nicht-paradigmatischer Fälle kein klares Urteil zu. Deshalb muss sich in einem anderen, als dem bereits betrachteten Beispiel Dworkins, ein Schiedsrichter beim Beurteilen eines schwierigen Falles in einem Schachspiel, für eine der möglichen Auslegungen einer bestimmten Regel entscheiden (vgl. Dworkin 1977: 102-3). Diese Auslegung muss dabei aber mit den Konventionen, i.e. den weithin geteilten Meinungen hinsichtlich der *Natur* des Schachspiels im Einklang stehen. Dies kann so verstanden werden, dass die von ihm „abstrakt“ genannten Konventionen zwar einen Begriffskern zum Ausdruck bringen, aber zu schwierigen Fällen schweigen. So meint Dworkin:

“[The conventions] are not incomplete, like a book whose last page is missing, but abstract, so that their full force can be captured in a concept that admits of different conceptions; that is, in a *contested* concept. [...] The referee must select one or another of these conceptions, not to supplement the convention but to enforce it.” (Dworkin 1977: 103)

Nun spricht Dworkin sowohl von „abstrakt“ wie auch von „unpräzise“, was die Frage aufkommen lässt, wie wir die Unterbestimmtheit der Schlüsselbegriffe und damit auch des Begriffes selbst zu verstehen haben. Dworkin weist jedenfalls fehlende Präzision, i.e. *Vagheit*, als eine angemessene Beschreibung des Grundes für Begriffsstreitigkeiten zurück (vgl. 1986: 39-43). Daher müssten umstrittene Konzepte in einem anderen Sinne unterbestimmt sein, als z.B. GLATZE oder HAUFEN. So kann die Bedeutung eines Begriffes auch *unklar* sein. Dann ist dessen Bedeutung bspw. unterbestimmt, weil die Schlüsselbegriffe abstrakt oder zu *allgemein* sind, um in bestimmten Kontexten noch ausreichend informativ zu sein (vgl. Lanius 2016: 39).⁵ Nun kann ich im Kontext dieser Arbeit keine Bestimmung jener Art begrifflicher Unterbestimmtheit vornehmen, die für Einigkeit und Uneinigkeit verantwortlich sein soll, weswegen ich im Folgenden lediglich

⁴ Die Terminologie folge dabei Hart (1961: 121-32; vgl. Dworkin 1977: 102, Ftn.1).

⁵ Gerade umstrittene Begriffe scheinen auf verschiedene Arten gleichzeitig unterbestimmt zu sein (vgl. Lanius 2016: z.B. 34f). Zum Problem der weitläufigen Unterlassung einer Unterscheidung verschiedener Arten von Unterbestimmtheit vgl. ebd.: S.5f; zum Unterschied zwischen Klarheit und Präzision ebd.: S. 23ff.

„Bedeutungsoffenheit“ als Platzhalter für die Unterbestimmtheit umstrittener Begriffe verwende. Dennoch lässt sich folgende Behauptung festhalten:

B3: Umstrittene Konzepte lassen verschiedene Interpretationen und damit Konzeptionen zu, weil der Begriffskern bedeutungsoffen ist.

Gemäss der vorliegenden Darstellung, werden im Zuge einer Interpretation die Schlüsselbegriffe genauer bestimmt, wodurch auch in umstrittenen Fällen klar wird, ob und wie der Begriff anzuwenden ist. Dabei kommen aber unterschiedliche Interpretationen zu je unterschiedlichen Urteilen hinsichtlich umstrittener Fälle zustande und so gibt es trotz B1 weder Einigkeit hinsichtlich der Frage, wie der Begriff genau zu verstehen ist, noch wie nicht-paradigmatische Fälle zu beurteilen sind. Damit liefert die in B3 festgehaltene Bedeutungsoffenheit die Antwort auf die Frage, wie es trotz der intersubjektiven Geteiltheit von Begriffen, i.e. einem geteilten Begriffskern (vgl. B1), zu Streitigkeiten kommen kann. Im folgenden Abschnitt werde ich nun das wechselseitige Verhältnis zwischen den verschiedenen Konzeptionen genauer betrachten.

3.2.2. Der Wettstreit der Konzeptionen

Gerade in der Einführung bei Rawls scheint klar zu sein, dass sich die Streitparteien auf *eine* Konzeption einigen müssen, um eine wohlgeordnete Gesellschaft zu ermöglichen. Auch im Kontext der Rechtsprechung – Dworkins Fokus – scheint es wichtig, dass Einigkeit bzgl. des Rechtsbegriffes besteht, da ansonsten der Vorwurf der Justizwillkür droht. Damit stellt sich die Frage, auf *welche* Konzeption sich eine entsprechende Streitgemeinschaft festlegen soll, wobei jede Streitpartei versucht ihre eigene Konzeption gegenüber den anderen Positionen als die beste auszuweisen und sie zur allgemeinen Annahme zu empfehlen. Angestrebt wird also eine von allen geteilte Begriffsauffassung, um welche die verschiedenen Konzeptionen miteinander rivalisieren:

B4: Die Konzeptionen stehen in direkter Rivalität zueinander, i.e. jede versucht sich gegen die anderen durchzusetzen, so dass sie zur einzigen gemeinhin geteilten Auffassung des Begriffes wird.

Damit stellt sich freilich die Frage, nach welchen Kriterien eine Konzeption ausgewählt werden soll. Das

Problem dabei ist aber, dass allfällige Bemessungskriterien ihrerseits unterschiedlich gewichtet werden können und es daher kein Set von solchen gibt, das den Widerstreit entscheiden könnte und welches alle Streitparteien teilen würden.⁶

Klar ist hingegen, dass höchstens eine der sich im Streit befindenden Konzeptionen zur gemeinhin geteilten Auffassung werden kann. Dies rührt daher, dass die Streitparteien, wie wir im vorangehenden Teil bereits gesehen haben, davon ausgehen, dass sie um die Bedeutung eines *einzigsten, nicht-mehrdeutigen* Begriffes streiten. Würde man nicht von einer solchen Begriffssingularität ausgehen, i.e. mehrere Begriffe z.B. von Gerechtigkeit zulassen, dann könnte für jeden von diesen eine Kandidatin unter den Konzeptionen gesucht werden und es würde in der Folge nicht nur um *eine* geteilte Begriffsauffassung, sondern um *mehrere* rivalisiert. Zudem würde dies bedeuten, dass es nicht *einen*, sondern *mehrere* Begriffskerne gäbe.

3.3. Schärfung der Unterscheidung

Durch die in diesem Teil extrahierten Behauptungen werden Begriffsstreitigkeiten also wie folgt beschrieben: Umstrittene Konzepte beinhalten einen unumstrittenen Bedeutungskern (B1), um dessen Interpretation gerungen wird (B2), weil dessen begriffliche Bestandteile bedeutungsoffen sind und daher verschiedene Konzeptionen zulassen (B3), welche miteinander rivalisieren, i.e. darum ringen, zur von allen akzeptierten Begriffsauffassung zu werden (B4), wobei sich alle einig sind, dass es nur *eine* gemeinhin geteilte Auffassung des umstrittenen Begriffes geben kann. Gemäss diesem Bild ist nun also der Unterschied zwischen einem Konzept und einer Konzeption der, dass letztere umstritten ist, ersteres jedoch nicht. Es scheint aber noch einen weiteren Unterschied zu geben. So suggeriert diese Darstellung, dass alle Begriffsauffassungen Interpretationen von ein und demselben Begriffskern darstellen. Bevor ich diese Vermutung im nächsten Teil einer kritischen Prüfung unterziehe, möchte ich das hier skizzierte Bild und die darin enthaltene Rede etwas schärfen.

Bislang wurde lediglich von einem Begriff und dessen Auffassungen gesprochen. Genaugenommen müsste aber von einem „Konzept von Gerechtigkeit“ und von den „Konzeptionen von Gerechtigkeit“ gesprochen werden. Wenn wir nun in dieser Rede x als Platzhal-

ter für einen beliebigen Terminus einsetzen, erhalten wir „Konzept von x “ und „Konzeptionen von x “, was ich im Folgenden mit K_x bzw. $k_{x,i}$ abkürzen werde, wobei i die Konzeptionen individuieren soll.⁷ Obwohl ich hier die Frage, worauf x genau referieren soll, nicht zu untersuchen vermag, möchte ich auf eine mögliche Antwort hinweisen, ohne für diese zu argumentieren. Für Dworkin wird ein geteilter Begriff durch jene Propositionen angezeigt, denen alle Streitparteien im Grossen und Ganzen zustimmen (vgl. 1986: 70). So meint er, dass eine gegebene Gemeinschaft z.B. einen Begriff von Unfairness hat, wenn „they agree on a great number of standard cases of unfairness and use these as benchmarks against which to test other, more controversial cases“ (Dworkin 1977: 134; vgl. auch 1986: 72). Nun geben aber Anwendungsbeispiele lediglich implizite Hinweise auf die Bedeutung eines Begriffes, welche das Konzept im Allgemeinen und die Konzeptionen im Spezifischen anzugeben versuchen. Da sich in diesem Sinne sowohl Konzepte, wie auch Konzeptionen auf Anwendungsbeispiele *als die zu erklärenden Objekte* beziehen, scheint es mir sinnvoll mindestens die entsprechenden Wortverwendungen als deren gemeinsames Bezugsobjekt x zu verstehen.⁸ Dabei stünde ein sprachlicher Ausdruck wie „Gerechtigkeit“ elliptisch für x .

Nun wurde bislang davon gesprochen, dass sich die Konzeptionen als Interpretationen auf den Begriffskern beziehen, was suggeriert, dass Konzeptionen – entgegen der soeben gemachten Erläuterung – nicht dasselbe Bezugsobjekt haben wie das Konzept. Diese beiden Redeweisen könnten aber kompatibel gemacht werden, wenn wir annehmen, dass sich die Konzeptionen indirekt über deren Begriffskern auf x beziehen. So lässt sich das Verhältnis R zwischen $k_{x,i}$ und K_x , i.e. $R(k_{x,i}, K_x)$ trotzdem erfassen als ‘ist eine Interpretation von’ und, konvers, $R(K_x, k_{x,i})$ als ‘ist das Interpretationsobjekt von’. Aus allem bisher gesagten lässt sich somit die folgende These formulieren:

T1: Etwas ist nur dann eine Konzeption k_x von einem Konzept K_x , wenn es (a) eine Interpretation ist, die (b) die konzeptuellen Bestandteile von K_x zum Objekt hat, wobei sich, in der Folge, alle $k_{x,i}$ auf dieselben konzeptuellen Bestandteile von K_x beziehen.

Obwohl nun die Methode der Interpretation in Dwo-

6 Für eine Übersicht über mögliche Beurteilungskriterien siehe List & Valentini (2016: 533-35). Für die Uneinigkeit hinsichtlich der Gewichtung von Kriterien vgl. Walter B. Gallie (1956: bspw. 173).

7 Dies ähnelt der Redeweise, wie sie z.B. Swanton verwendet (vgl. 1992: 1)

8 Für einen ausführlicheren Vorschlag vgl. Swanton (1992: 6-7).

rkins Analyse des Rechtsbegriffes eine zentrale Rolle spielt und er sie deshalb auch eingehend diskutiert (vgl. 1986: 45-86), kann ich hier nicht auf seine Bestimmung derselbigen eingehen. Dennoch stellt sich die grundsätzliche Frage, ob eine Begriffsauffassung nur dann als eine Konzeption bei der Strukturierung eines Begriffsstreits berücksichtigt werden soll, wenn sie nach bestimmten methodischen Kriterien gewonnen wurde. Gilt eine Auffassung z.B. nur dann als eine Konzeption von x , wenn sich deren Entstehung als einen interpretativen Akt verstehen lässt oder sind auch andere Methoden qualifizierend? Wenn mit der BAU methodische Einschränkungen verbunden sein sollten, müsste geklärt werden welche dies sind und warum sie gelten. Ich kann hier freilich weder auf diese Frage antworten noch entscheiden, ob die BAU überhaupt allen möglichen Techniken zur Gewinnung von Begriffsauffassungen gerecht werden muss. Daher lasse ich 'Interpretation' hier als Platzhalter für eine genauer zu bestimmte Menge von qualifizierenden Methoden stehen.

Nun möchte ich mich der Frage zuwenden, ob T1 ein geeignetes Werkzeug darstellt um Begriffsstreitigkeiten zu strukturieren indem wir Konzeptionen als unterschiedliche Interpretationen eines Begriffskerns identifizieren. Da (b) voraussetzt, dass ein Bedeutungskern zur Interpretation vorliegt, werde ich im nächsten Teil untersuchen, inwiefern diese Annahme zutrifft. Wie sich zeigen wird, hängt die Antwort davon ab, ob wir die BAU normativ oder deskriptiv verstehen wollen.

4. Zwei Lesarten der Einigkeitsbehauptung

Wenn das Verhältnis zwischen K_x und $k_{x,i}$ ein Interpretationsverhältnis ist, dann ergeben sich für die Formulierung von Konzeptionen inhaltliche Einschränkungen aus dem Begriffskern. Das heisst, dass K_x einen substanziellen Rahmen für alle $k_{x,i}$ vorgibt und somit die Menge möglicher, mit dem Begriff assoziierter Inhalte beschränkt. Damit entscheidet das Konzept inhaltlich darüber, welche in einem Begriffsstreit verhandelten Konzeptionen wir bei dessen Strukturierung zu berücksichtigen haben. Wenn z.B. in unserem Beispiel mit den vier Freunden, HAUSHUND den Bedeutungskern von HUND darstellen würde, dann stellt bspw. Chris' Auffassung eine Konzeption dieses Begriffes dar, Annas Auffassung hingegen nicht. Der Grund dafür wäre, dass Annas Auffassung einen grösseren Allgemeinheitsgrad aufweist als HAUSHUND und damit nicht als eine Spezifikation des Begriffskerns verstan-

den werden kann. Denn sie scheint HUND im zoologischen Sinne zu verstehen, i.e. $HUNDE_{(CANIDAE)}$ mit diesem zu verknüpfen. Damit weist Annas Begriff von Hunden eine grössere Extension auf als HAUSHUND. Das heisst, dass $HUNDE_{(CANIDAE)}$ als Konzeption von „Hund“ in diesem Fall ausscheidet und zwar auch dann, wenn man, wie Bea, die Einschränkung HAUSTIER hinzufügen würde. Denn Füchse würde man auch dann nicht zu den Haushunden zählen, wenn sie als Haustiere gehalten werden. Chris hingegen interpretiert die Wortverwendungen zwar auch zoologisch, aber eben als $HAUSHUND_{(C.L. FAMILIARIS)}$ und Dana versteht den Begriff primär im züchterischen Sinne, fügt aber zu HAUSHUND gewisse phänotypische Einschränkungen hinzu. Demnach hätten also nur Chris und Dana rivalisierende Konzeptionen von HUND, Anna und Bea aber nicht. Wenn wir bereit sind diesen Ausschluss über den Allgemeinheitsgrad als eine inhaltliche Beschränkung durch HAUSHUND zu verstehen, so zeigt dieses Beispiel, inwiefern ein Begriffskern inhaltliche Restriktionen generiert und somit bestimmte Konzeptionen von der Berücksichtigung bei der Strukturierung eines Begriffsstreites ausschliesst. Nun ist aber die Frage, inwiefern wir einen Bedeutungskern als gegeben zu verstehen haben und wie wir *diesen* identifizieren können. Die Antwort darauf hängt nun davon ab, wie wir die BAU grundsätzlich verstehen wollen. So gibt es bei Rawls zwei Passagen, die m.E. verschiedene Lesarten ermöglichen, welche das Zustandekommen der Einigkeit in B1 je unterschiedlich erklären. Zum einen wird für Rawls (i) das Konzept der Gerechtigkeit durch jene Rolle spezifiziert „which [...] these different conceptions, have in common“ (Rawls 1971: 5; Kursivierung hinzugefügt). Zum anderen meint er, dass (ii) die von ihm angegebenen Bedeutungsbestandteile *vom Konzept beinhaltet* würden (vgl. ebd.). Damit lassen sich, wie ich nun zeigen werde, unterschiedliche Vorgehensweisen bei der Identifikation des Begriffskerns assoziieren, die je unterschiedliche Konsequenzen für das Verhältnis zwischen Begriff und Begriffsauffassungen und damit für T1(b) haben.

4.1. Den Begriffskern postulieren?

Die erste, normative, Lesart mit Fokus auf (ii) wird durch die Betrachtung Rawls' Vorgehen in seiner Einführung nahegelegt, wo er zunächst die zentrale Rolle der Prinzipien sozialer Gerechtigkeit skizziert und daraufhin behauptet, dass es *diese* Rolle ist, die von allen Konzeptionen interpretiert würde (vgl. 1971: 4-5). Das heisst, dass das Konzept unabhängig von allfälligen

Konzeptionen gewisse Charakteristika zu besitzen scheint, an denen sich Konzeptionen entweder aus Einsicht orientieren oder aber sich orientieren müssen um als Konzeptionen des entsprechenden Konzeptes zu gelten bzw. berücksichtigt werden zu können. Aus diesem Grund weisen alle Konzeptionen jene Elemente auf, welche den *postulierten* Charakteristika des Konzeptes entsprechen. Dies entspräche im Minimum folgender Annahme:

L1: Konzepte enthalten ungeachtet ihrer Konzeptionen gewisse Bedeutungsbestandteile.

Um einen Begriffsstreit zu strukturieren, müssten wir demnach zunächst den Begriffskern, i.e. die grundlegenden Charakteristika von K_x identifizieren. Daraufhin können wir entschieden, welche Konzeptionen diesem Begriff K_x zugeordnet werden können und welche nicht. Damit verhalten wir uns aber nicht mehr neutral zur betrachteten Diskussion, sondern beziehen inhaltlich Stellung indem wir mitunter gewisse der darin vertretenen Positionen aus inhaltlichen Gründen, vorgegeben durch den postulierten Begriffskern, disqualifizieren. Damit liefert uns diese Version der BAU keine nicht-diskriminierende Alternative zu P3.

Es bleibt hier allerdings unklar wie die Identifikation der als elementar angenommenen Bedeutungsbestandteile genau erfolgen müsste. So könnte sich L1 durchaus wie in P3 aus einer angenommenen objektiven Bedeutung des Begriffes ergeben, womit wir nur eine andere Version von P3 hätten. Oder aber sie ergeben sich, wie von Rawls suggeriert, aus einem urtümlichen Verwendungszweck des Begriffes, womit aber gezeigt werden müsste, dass der Begriff tatsächlich nur diesen Verwendungszweck hat und dass folglich alle, die diesem nicht zustimmen, den Begriff falsch, oder zumindest nicht im *eigentlichen* Sinne verstehen und verwenden. Welche Art von Argumenten es hierzu bedürfte bleibt hier freilich offen.

Obwohl nun L1 genau der Aussage von T1(b) entspricht, so ist diese Lesart für unsere Belange wenig informativ. Wir wissen zwar, dass der Begriffskern ein Postulat darstellt aber es bleibt die Frage offen wie wir den Begriffskern identifizieren können – hierauf bedarf es einer Antwort, selbst wenn wir diese Version der BAU zur parteilichen Strukturierung eines Begriffsstreits verwenden wollen. Wie wir nun sehen werden, stellt sich dieses Problem nicht, wenn wir die Einigkeit anders erklären.

4.2. Den Begriffskern extrahieren?

Die zweite Lesart versteht Rawls' Aussagen mit Fokus auf (i) faktisch: wenn alle Konzeptionen zu einem gegebenen Konzept *tatsächlich* ein und dasselbe Element beinhalten, dann lässt sich der Begriffskern aus den Konzeptionen extrahieren. Dies würde bedeuten, dass wir, um den Begriffskern zu bestimmen, alle Konzeptionen betrachten und auf deren Gemeinsamkeiten hin überprüfen müssten. Hinter diesem Vorgehen stünde dann die folgende Annahme:

L2: Die inhaltliche Bestimmung eines Konzeptes ergibt sich durch die von allen Konzeptionen tatsächlich beinhalteten Bedeutungsbestandteile.

Wenngleich ich nicht ausschliessen will, dass Rawls bei seiner Bestimmung des Begriffskerns von GERECHTIGKEIT hintergründig ähnlich verfahren ist, so wird diese Lesart klar durch Dworkins Erläuterungen nahegelegt (vgl. Dworkin 1986: 70-71). So beschreibt er in seinem Höflichkeits-Beispiel den beauftragten Philosophen als die Struktur *aufdeckend* („[e]xposing this structure“, Dworkin 1986: 71) und die Richtigkeit oder den Erfolg dessen Analyse als abhängig davon, dass sich die Leute auch *tatsächlich* einig sind („people are by and large agreed that courtesy is a matter of respect“, *ibid.*, Kursivierung hinzugefügt). Zudem soll der gelieferte Ansatz weniger inhaltlich („more conceptual and less substantive“, Dworkin 1986: 70) sein als die bereits bestehenden Theorien. Gemäss meiner Deutung sprechen diese Passagen dafür, dass mit dem Strukturierungs-Unterfangen ein deskriptiver Anspruch verbunden ist. Zumindest ist unter der Annahme von L2 das Konzept nicht-diskriminierend gegenüber diesen Konzeptionen und wir bleiben hinsichtlich der zu strukturierenden Debatte inhaltlich neutral.

Obwohl L2 auf erste Sicht eine einfache Erklärung für die Einigkeit liefert, so stellt sich die Frage, worauf sich die Konzeptionen nun beziehen. Da der Begriffskern aus bereits vorhandenen Konzeptionen extrahiert werden muss, liegt er letzteren ja nicht zur Interpretation vor – und dies widerspricht T1 (b) und (c). Auch scheint bei näherer Betrachtung erklärungsbedürftig, wie es dann dazu kommt, dass jede Konzeption gewisse, über alle $k_{x,i}$ identische Bedeutungselemente beinhaltet – doch bevor ich näher auf die Probleme dieser Lesart eingehe, bietet sich eine zusammenfassende Gegenüberstellung der beiden Lesarten an.

Während sich also die in B1 beschriebene Einigkeit gemäss der ersten Lesart durch die Übereinstimmung mit einem Postulat ergibt, wird sie bei der zweiten als *de facto* gegeben angenommen. Daraus hat sich ergeben, dass sich unter L1 alle $k_{x,i}$ auf das Konzept als Interpretationsobjekt beziehen, während der Bezug der $k_{x,i}$ unter der zweiten Lesart ungeklärt ist. Wie ich gezeigt habe, ist die normative Lesart somit mit der hier skizzierten Version der BAU kompatibel, stellt aber keine deskriptive Alternative der gewünschten Art zu P3 dar. Im Gegensatz dazu liefert uns die zweite, deskriptive Lesart eine nicht-diskriminierende Version der BAU. Da diese aber inkompatibel mit T1 ist, müssen wir entweder L2 aufgeben oder die Charakterisierung der BAU anpassen. Im Folgenden werde ich zeigen, welchen Herausforderungen wir zu begegnen haben, wenn wir letzteres versuchen und welche Ressourcen in Dworkins Werken uns dabei zur Verfügung stehen.

5. Uneinigkeit in realen Begriffsstreitigkeiten

Mir scheint, dass auch die deskriptive Lesart davon ausgeht, dass die grundsätzliche Bedeutung eines Begriffes den Konzeptionen irgendwie zugänglich ist, da es ansonsten rätselhaft bleibt, wie angenommen werden kann, dass eine Einigkeit hierzu zustande kommt. Im Unterschied zu L1 darf die grundsätzliche Bedeutung aber nicht durch den Begriffskern vorgegeben sein, da sonst diese Lesart auf eine problematische Weise zirkulär wird. Nun besteht die Möglichkeit, dass sich die Konzeptionen direkt auf x beziehen, i.e. auf das, was wir durch die Angabe der Bedeutung eines Begriffes zu erklären versuchen. Das heisst, dass das allen $k_{x,i}$ gemeinsame Interpretationsobjekt eine übereinstimmende Menge von Begriffsverwendungen ist, für welche x elliptisch steht. Dementsprechend könnte T1 folgendermassen neu formuliert werden:

T2: Etwas ist nur dann eine Konzeption $k_{x,i}$ von einem Konzept K_x , wenn es (a) eine Interpretation ist, die (b) eine variable Menge von Begriffsverwendungen x zum Objekt hat, wobei es (c) für alle $k_{x,i}$ eine Übereinstimmung in x gibt.

Dieser gemeinsame Bezug würde zumindest die Wahrscheinlichkeit steigern, dass alle Parteien ungefähr dieselben Bedeutungselemente als zentral identifizieren. Doch wenn wir davon ausgehen, dass zu einem Begriff über verschiedene Epochen und in unterschiedlichen kulturellen Kontexten und unter unter-

schiedlichen doxastischen Systemen eine Vielzahl von Begriffsauffassungen entworfen wurden, scheint mir weder die Identifikation identischer oder ähnlicher konzeptueller Bestandteile – i.e. die Garantie für einen Begriffskern –, noch die Annahme eines geteilten Interpretationsobjektes – Annahme T2(c) – plausibel. Aber wie könnten wir die BAU dann charakterisieren? Und können wir unter L2 überhaupt an ihr festhalten? An dieser Stelle scheint es mir sinnvoll, einen Blick auf die Gegebenheiten realer Begriffsstreitigkeiten zu werfen um die Tragweite des hier aufgeworfenen Problems genauer aufzuzeigen.

5.1. Der verlorene Unterschied

Swanton beobachtet in Bezug auf den Streit um den Begriff FREIHEIT, dass selbst auf höchstem Abstraktions-Level gestritten wird und nicht einmal Einigkeit über die formale Struktur des Begriffes herrscht (1992: viii). Ihre Diagnose ist: „There is a multitude of conceptions of freedom that defy even attempts at classification; [...] no agreed upon conceptual core” (Swanton 1992: ix). Auch Dworkin selbst gesteht ein, dass es nicht möglich sei, ein Bedeutungselement zu identifizieren, das zu allen Zeiten, von allen Gemeinschaften, die an der sozialen Praxis der Höflichkeit partizipieren, geteilt wird: „there is no such feature: courtesy is at one stage regarded as a matter of respect, and at another as something very different” (Dworkin 1986: 69). Diese Aussage Dworkins steht nun im starken Kontrast zu der Charakterisierung der BAU, wie sie auch durch seine eigenen Ausführungen nahegelegt wurde. Wie ist das zu erklären und was bedeutet das allfällige Fehlen eines Begriffskerns für die BAU unter L1?

Da die Beobachtung einer fehlenden inhaltlichen Überlappung zwischen den Konzeptionen, i.e. ein fehlender Begriffskern, äquivalent zu $\neg B1$ ist, ist B1 in der vorliegenden Formulierung – zumindest für gewisse Begriffe – nicht aufrechtzuerhalten. Dasselbe gilt für B2, da ja auch der Begriffskern selbst umstritten ist, i.e. sich der Streit eben nicht nur darum dreht, wie die durch den Begriffskern gegebenen konzeptuellen Bestandteile zu interpretieren sind. Auch B3 scheint abgelehnt werden zu müssen, zumal die Tatsache, dass umstrittene Konzepte mehrere Konzeptionen ermöglichen, nicht mehr lediglich darauf zurückgeführt werden kann, dass der Begriffskern bedeutungs offen ist, sondern die Bedeutungs offenheit umstrittener Begriffe anders erklärt werden müsste. Da also bereits ein Wettstreit darum besteht, welches die *grundlegenden* Bedeutungsbestandteile eines umstrittenen Begriffes sind, besteht Rivalität nicht mehr lediglich zwischen

Konzeptionen. Vielmehr besteht sie auch zwischen den verschiedenen, vorgeschlagenen oder identifizierbaren Begriffskernen womit B4 nicht mehr lediglich für Konzeptionen, sondern auch für die verschiedenen Bedeutungskerne bzw. Konzepte gilt.

In der Folge scheinen wir nun aber *jede* Position, die irgendeine Erklärung für x vorschlägt als eine Auffassung (oder einen Begriff) von x ansehen zu müssen. Und zwar, erstens, unabhängig von deren Abstraktionsgrad und, zweitens, ungeachtet dessen, ob einige Positionen einen gemeinsamen Bedeutungskern aufweisen. Denn es kann nicht mehr daran festgehalten werden, dass es genau ein K_x gibt, das für *alle* $k_{x,i}$ identisch ist. Das heisst, dass die verschiedenen $K_{x,j}$ ebenfalls lediglich Auffassungen von x darstellen, wenn gleich von mehreren Parteien geteilt – doch diese Eigenschaft ist m.E. auch für Konzeptionen nicht ausgeschlossen. Damit würde aber die Unterscheidung zwischen Begriff und Begriffsauffassung in sich zusammenfallen, da die ursprünglich behaupteten Unterschiede nicht mehr gegeben wären.

5.2. Die Relativierung der Einigkeit

Nun können wir versuchen die Unterscheidung zu retten, indem wir B1 kompatibel mit der allfällig fehlenden Einigkeit in realen Begriffsstreitigkeiten machen. Dazu muss aber neu bestimmt werden worauf „Konzept“ referiert, wenn nicht auf einen *allen* Konzeptionen gemeinsamen Bedeutungskern. Das Beispiel von Dworkin suggeriert, dass es mitunter möglich ist, für eine *gegebene Gemeinschaft* von Konzept-Nutzern, zu *einem gegebenen Zeitpunkt* einen Begriffskern zu extrahieren und somit *deren* Begriff von x zu charakterisieren (vgl. 1986: 69-72). Somit könnten wir B1 folgendermassen relativieren:

B1*: Ein Konzept K_x existiert für eine gegebene Gemeinschaft G zu einem Zeitpunkt t gdw. sich K_x durch einen Bedeutungskern mit Geltung für G zu t konstituiert, i.e. gdw. $K_{x,G,t}$ konzeptuelle Bestandteile enthält, die von allen Konzeptionen $k_{x,i}$ in G zu t geteilt werden.

Die Frage ist nur: Lässt sich nun $K_{x,G,t}$ aus den $k_{x,G,t,i}$ einer gegebenen, raum-zeitlich lokalisierten Gemeinschaft G extrahieren oder konstituiert sich die Gemeinschaft G erst, in abstrakter Weise, durch die Extraktion eines von *einigen* $k_{x,t,i}$ geteilten Begriffskerns? Die Idee hinter B1* könnte durchaus auch so verstanden werden, dass eben nur jene, welche tatsächlich einen

gemeinsamen Bedeutungskern von x aufweisen, auch einen Begriff von x teilen und sich G erst durch diese Einigkeit konstituiert und nicht umgekehrt.

Eine weitere offene Frage ist, wie wir möglichen, untergeordneten Streitigkeiten terminologisch begegnen. Wenn wir bspw. innerhalb einer gegebenen Gemeinschaft feststellen, dass es verschiedene Vorschläge dazu gibt, wie die Bedeutungsbestandteile einer bestimmten Konzeption, bspw. $k_{x,G,t,1}$, zu spezifizieren sind und wir nun *diesen* Streit erneut auf dieselbe, in B1* beschriebene Art strukturieren können, dann hätte $k_{x,G,t,1}$ doch eine grundsätzliche Ähnlichkeit mit einem Konzept – zumindest wäre es der spezifische Begriff dieser Untergruppe.⁹ Obwohl dies eine detailliertere Strukturierung von Begriffsstreitigkeiten eröffnet, so wird diesbezüglich eine terminologische Frage aufgeworfen: Wie sollen wir diese nennen?

Abgesehen von diesen offenen Fragen stellt sich ein grundsätzliches Problem. Wenn wir gemäss B1* für ein bestimmtes G und t einen Begriffskern extrahieren können, dann entspricht das extrahierte Konzept lediglich jenem Begriff von x , um welchen *die entsprechende Gemeinschaft* streitet, nicht aber einem *einigen* Begriff von x , um den *alle* Gemeinschaften streiten würden. Nun haben wir ja soeben bestritten, dass die verschiedenen Konzepte von x , *im Kontext der gesamten Diskussion* ebenfalls nur Konzeptionen zu x darstellen, sondern eben von den entsprechenden Gemeinschaften zu bestimmten Zeitpunkten geteilte *Begriffe* repräsentieren. Damit gestehen wir aber ein, dass ein sprachlicher Ausdruck eben doch für verschiedene Begriffe stehen kann, von denen aber keiner die Eigenschaft hat intersubjektiv geteilt zu sein – zumindest nicht im Kontext der gesamten Diskussion. Damit lehnen wir die Eindeutigkeits- und Intersubjektivitätsbedingung für umstrittene Begriffe, i.e. P2 ab. Wenn wir aber dies tun, dann brauchen wir nicht nur eine andere Motivation für den terminologischen Rahmen der BAU, sondern müssen erneut der Frage begegnen, wie denn nun sichergestellt werden kann, dass *alle* Streitparteien vom selben sprechen und nicht einige einfach aneinander vorbeireden. Ich will freilich nicht ausschliessen, dass hierauf keine erhellende Antwort gegeben werden kann (vgl. z.B. Swanton 1992: Kap.1). Es reicht mir gezeigt zu haben, dass die BAU unter L2 keine solche bereithält.

An dieser Stelle bietet es sich an, die Ergebnisse die-

⁹ Der hier drohende Regress scheint mir höchstens durch Verweis auf die praktische Unwahrscheinlichkeit, nicht aber durch Verweis auf die theoretische Unmöglichkeit ausgeschlossen werden zu können.

ses Teils kurz zusammenzutragen. Ein Blick auf reale Streitigkeiten hat gezeigt, dass das Konzept nicht mit einem Begriffskern verknüpft werden kann, der tatsächlich von allen Streitparteien geteilt wird, da ein solcher nicht immer vorzufinden ist. Mit Dworkins Ressourcen liesse sich die BAU zwar grundsätzlich so relativieren, dass sie die Realität von Begriffsstreitigkeiten zu erfassen vermag. Dann identifizieren wir aber bei der Strukturierung einer Debatte mitunter mehrere Begriffe und es stellt sich die Frage nach deren gemeinsamen Bezugsobjekt. Zudem scheint es keinen zwingenden Anlass mehr dazu zu geben, zwischen einem Begriff und verschiedenen Auffassungen zu unterscheiden. Denn wir könnten m.E. genauso gut von verschiedenen Begriffen (oder Auffassungen) mit unterschiedlichen inhaltlichen Überlappungsbereichen sprechen.

5.3. Geht es nicht auch ohne?

Nun ist es in der Philosophie gebräuchlich, bei der Betrachtung oder Darstellung eines Begriffsstreits von unterschiedlichen Auffassungen eines Konzeptes zu sprechen – eine Praxis, der ich hier mitnichten die Berechtigung absprechen möchte. Dabei muss aber m.E. mit „Konzept“ nicht zwingend die Idee eines Begriffskerns verbunden werden. Damit kann auch lediglich gemeint sein, dass ein entsprechender Ausdruck von verschiedenen Theorien mit unterschiedlichem Gehalt versehen wird. Dies lässt aber die Möglichkeit offen, umstrittene Begriffe so zu verstehen, dass mit ihnen (unzählige) verschiedene Bedeutungen verbunden sein können. Dann können sprachliche Ausdrücke, wie z.B. „Gerechtigkeit“ oder „Freiheit“, für mehrere Begriffe stehen und es scheint nicht viel daran zu hängen, ob wir von unterschiedlichen Auffassungen von x, oder von unterschiedlichen Begriffen von x sprechen.

Mir scheint, dass auch ohne die Verwendung der Unterscheidung zwischen Konzept und Konzeptionen, Begriffsstreitigkeiten strukturiert werden können. So beschreibt bspw. Swanton zweierlei Arten der Uneinigkeit in Begriffsstreitigkeiten: Kern- und Randstreitigkeiten. Während erstere sich um grundlegende Fragen im Zusammenhang mit der Bedeutung des Begriffes drehen, beziehen sich letztere auf mindestens eine der in Kernstreitigkeiten gegebenen Antworten (vgl. Swanton 1992: 63). Mit diesem Ansatz müssten wir zur Strukturierung von Begriffsstreitigkeiten nicht die BAU bemühen. Wenn wir dies dennoch tun wollen, so habe ich nun einige der Fragen aufgeworfen, die zuvor beantwortet werden müssen.

Natürlich können wir eine normative Lesart der BAU

wählen und somit auch an der gegenwärtigen Charakterisierung festhalten. Dann müssen wir diese Unterscheidung aber als ein Mittel ansehen, um uns inhaltlich in eine Debatte einzumischen, da wir mit einem postulierten Begriffskern gewisse Konzeptionen disqualifizieren. Vor diesem Hintergrund bezweifle ich aber Rawls' folgende Aussage: „Clearly, this distinction between the concept and the various conceptions of justice settles no important questions“ (1971: 5).

6. Fazit

In dieser Arbeit habe ich gezeigt, dass wir die Aussagen, mittels derer Rawls und Dworkin die BAU charakterisieren, auf zwei Arten verstehen können – eine normative und eine deskriptive. Wenn wir die BAU zur Strukturierung von Begriffsstreitigkeiten gemäss der normativen Lesart verwenden, dann können wir zwar an der gegenwärtigen Charakterisierung festhalten, müssen aber die Frage beantworten, wie wir den Begriffskern bestimmen können. Zudem bleiben bei diesem Vorgehen all jene Konzeptionen, die den postulierten Begriffskern nicht teilen, unberücksichtigt.

Unter einer deskriptiven Lesart wird hingegen der Begriffskern aus den in einem Begriffsstreit vertretenen Konzeptionen extrahiert. Wenn es aber für eine gegebene Debatte nicht möglich ist, einen von allen Positionen geteilten Begriffskern zu identifizieren, so fällt entweder die Unterscheidung in sich zusammen oder wir formulieren die Annahmen, die die BAU charakterisieren so, dass sie die realen Gegebenheiten solcher Begriffsstreitigkeiten einzufangen vermögen. Dann bestätigen wir aber die Annahme, dass umstrittene Begriffe mehrdeutig sind und müssen eine Antwort auf die Frage finden, inwiefern die Streitparteien vom selben sprechen – eine Frage freilich, auf die die BAU ursprünglich eine Antwort bereitzuhalten schien.

Literatur

- Criley, Mark E. 2007. Contested concepts and competing conceptions. Dissertation, University of Pittsburgh, <<https://core.ac.uk/download/pdf/12209654.pdf>>.
- Dworkin, Ronald. 1977. Taking Rights Seriously. Cambridge, Mass.: Harvard University Press.
- - 1986. Law's Empire. Cambridge, Mass.: Harvard University Press.
- Ezcurdia, Maite. 1998. "The Concept-Conception distinction". Philosophical Issues 9, 187-192.
- Gallie, Walter B. 1956. "Essentially Contested Concepts". Proceedings of the Aristotelian Society 56, 167-198.
- Hart, Herbert L.A. 1961. The Concept of Law. Oxford: Clarendon Press.
- Higginbotham, James. 1998a. "Conceptual Competence". Philosophical Issues 9, 149-162.
- - 1998b. "On Knowing One's Own Language". In C. Wright, B.C. Smith und C. Macdonald (Hg.). Knowing Our Own Minds. Oxford: Clarendon Press. 429-441.
- Lalumera, Elisabetta. 2014. "On the explanatory value of the concept-conception distinction". Rivista Italiana di Filosofia del Linguaggio 8(3), 73-81.
- Lanius, David. 2016. Strategic Indeterminacy in the Law. Dissertation, Humboldt-Universität zu Berlin. (Anstehende Veröffentlichung: Oxford University Press).
- List, Christian und Laura Valentini. 2016. "The Methodology of Political Theory". In H. Cappelen, T.S. Gendler und J. Hawthorne (Hg.). The Oxford Handbook of Philosophical Methodology. Oxford: Oxford University Press. 525-53.
- Margolis, Eric und Stephen Laurence, "Concepts". In E. N. Zalta (Hg.). The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), <<https://plato.stanford.edu/archives/spr2014/entries/concepts/>>. Zuletzt besucht am: 13.03.2018.
- Peacocke, Christopher. 1992. A study of concepts. New York: The MIT Press.
- Rawls, John. 1999 [1971]. A Theory of Justice, 2nd Ed. (1999), Cambridge, Mass.: Harvard University Press.
- Solum, Lawrence B. 2004 (Version Jan. 2018). "Legal Theory Lexicon 028: Concepts and Conceptions". <<http://lsolum.typepad.com/legal-theory-lexicon/2004/03/legal-theory-le-1.html>>. Zuletzt besucht am: 07.03.2018.
- Swanton, Christine. 1992. Freedom. A Coherence Theory. Indianapolis/Cambridge: Hackett Publishing Company.

Fabio Giovanni Briante (35) schliesst gegenwärtig sein Masterstudium in Philosophie und Soziologie an der Universität Bern ab. Sein Interesse gilt den Debatten um den Autonomiebegriff und die Natur philosophischer Begriffsstreitigkeiten sowie Themen der Sozialphilosophie, Handlungstheorie, Ethik und Metaethik.

A Trilemma in the Philosophy of Scientific Explanation

1. Introduction

Science has accomplished a lot. Not only has it improved our lives substantially by providing us with technological advances, but it has also offered us a way of empirically investigating the world over the past centuries. As a result, our understanding of the world has been influenced dramatically by scientific investigation and reasoning: Science has been able to provide us with explanations for the phenomena we observe and will continue to do so in the future. Philosophers have developed theories of scientific explanation, i.e. systematic descriptions of what characteristic of science enables its explanatory capacities.¹ Michael Strevens is one of those philosophers. He developed the so-called *kairetic account* of scientific explanation, a theory that I will defend in this essay.²

An important distinction I will introduce is the distinction between the communicative sense and the ontological sense of the term “explanation”: The communicative sense regards explanations as a communicative act of conveying scientific knowledge, whereas the ontological sense regards explanations as a part of the stock of scientific knowledge itself. To make this distinction clearer, imagine yourself sitting on a train. You are gazing at the majestic Swiss Alps. While you are about to get carried away by a daydream, a thought suddenly strikes you – why is it possible to see through solid glass? Why is it transparent? At this moment, you are looking for an explanation in the ontological sense of the word, that is, some scientific facts which explain the transparency of window glass. If you had a physicist sitting next to you who happened to give you the explanation, the way in which they would present the scientific facts constitute the communicative sense of an explanation. Although both the ontological as well as the communicative sense of

the word “explanation” are legitimate, this essay will take the ontological sense as more fundamental and the communicative sense as secondary. The reason for this lies in the nature of the kairetic account: it is a causal theory of explanation.³

Causal theories of scientific explanation have been very prominent where the explanandum (the proposition to be explained) is an empirical phenomenon or regularity. These theories generally claim that to explain an empirical explanandum means to cite causes and mechanisms implying the explanandum. Rival theories of explanation include the pragmatic account (an explanation is an answer to a “why”-question) and the unification account (an explanation constitutes a step towards a unified picture of the world). So, what makes causality more attractive than other accounts of explanation? I will answer this question and discuss some causal theories in section two of this essay.

Unfortunately for causal theories, philosophers have recently drawn attention to a special type of explanation which seems to be at odds with causal theories of scientific explanation – genuine mathematical explanations. Those are scientific explanations which heavily rely on mathematical propositions in the explanans (the propositions evoked to explain the explanandum). Several examples have been presented, so the validity of such explanations seems to be confirmed. On the one hand, we have an attractive theory of scientific explanation, i.e. the kairetic account or, more generally, causal theories of scientific explanation. On the other hand, there exist scientific explanations which appear to reach beyond the remit of causal theories of scientific explanation. We are therefore left with a trilemma:

1 I speak of scientific explanation in the singular because most philosophical theories assume that there is one overarching characteristic about all explanations which they all have in common. Furthermore, scientific explanation and “ordinary” explanation are not categorically separated. This is why some examples of explanations will not be strictly speaking scientific.

2 Strevens, 2008

3 Strevens, 2008: p. 6.

1. All scientific explanations are causal explanations.
2. There are genuine mathematical explanations.
3. Genuine mathematical explanations are non-causal explanations.

The aim of this essay is to resolve this trilemma by rejecting one of the presented propositions.

I will argue against the second, the existence of genuine mathematical explanations. I believe that causal theories of scientific explanations are promising, especially in the context of the kairetic account of scientific explanation. Regarding the third proposition of the trilemma, I accept that a genuine mathematical explanation would be explanatory in so far as it does not make reference to causal factors and causal processes. Firstly, I will introduce the kairetic account of difference-making for the scientific explanation of empirical events. Secondly, genuine mathematical explanation will be characterised and two arguments will be raised against genuine mathematical explanation. Thirdly, some consequences and remaining questions will be considered.

In the next section, I will present a part of the kairetic account of scientific explanation. It should be noted that there are many different forms of scientific explanation. However, due to the limited scope of this paper and the irrelevance of some types of explanations to the issues under discussion, they will not all be considered here. The main focus will be restricted to event explanation, especially because regularity explanation can be reduced to event explanation in causal theories of explanation.

2. The Kairetic Account of Scientific Explanation

2.1. Causal Theories of Explanation

The fundamental question in need for an answer is the following: What constitutes a scientific explanation? A plausible way to approach this question is to show how the explanandum came to be and how it fits in the bigger picture of the world. Causality meets this requirement. It shows what the event's causes are and which processes gave rise to it.⁴ For this reason, it is

⁴ Explanations may contain events and states of affairs as their explananda. In this essay, I will treat them interchangeably where explanations are concerned. For reasons of convenience, I will simply use events as explananda, although states of affairs may also be at stake in explanations.

not surprising that philosophers have proposed causal theories of scientific explanation. Take, for example, David Lewis. He claimed that to explain an event means to cite some of its causal history, i.e. some events and processes that demonstrate how the explanandum event came about.⁵ To explain why I bruised my knee is to tell the story of how I tripped and fell playing tennis, for example. A lot more could be said about why I went to play tennis, about details of the physiological processes giving rise to bruises and so on. On one hand, it might seem that a full causal history would constitute the ideal scientific explanation. On the other hand, established practice of scientific explanation does not conform to this ideal. This indicates that a complete reconstruction is not required to arrive at a full explanation as long as our philosophical endeavour is descriptive rather than normative.

The kairetic account of causal difference-making follows this idea: While maintaining that mentioning only a part of the causal history of the explanandum event constitutes a full explanation, the account also manages to identify *which* causes and causal processes are needed for a sufficient explanation. Although the kairetic account builds heavily on the notion of causality and difference-making, the central conception of the nature of scientific explanation is also inspired by the deductive-nomological model of explanation (which claims that a scientific explanation is a sound deductive argument where the premises need to include at least one scientific law), whereby an explanation is a deductive argument with the explanans as its premises and the explanandum as its conclusion. The premises should contain a causal law having the form of a conditional and the required difference-makers which constitute the antecedens of the causal conditional. Those terms will be defined in sections below because they have a technical meaning in the context of the kairetic account of explanation. An explanation takes the following abstract form in propositional logic:

$$\frac{c_1 \wedge c_2 \wedge \dots \wedge c_n}{(c_1 \wedge c_2 \wedge \dots \wedge c_n) \rightarrow e} \quad e$$

The first premise is the conjunction of all difference-making propositions. The second premise is the causal conditional by use of which conclusion *e* is obtained. The causal law constituting the second premise

⁵ See for example Lewis, 1986: p. 217.

represents a causal process taking place in the world. Nevertheless, no specific theory of causality has been invoked. What is required is that some scientific laws reflect causality. e is entailed by the difference-making propositions and the causal law. This special type of entailment, causal entailment, is thus required to deduce the explanandum.

It should be noted again that I will not present the complete kairetic account of scientific explanation. Rather, I will focus on those parts which will be applied to genuine mathematical explanation later in this essay.

2.2. Event Explanation

Let us first consider the explanation of an arbitrary event/state of affairs e , for instance my having a bruised knee. What is its explanation? According to the kairetic account, answering this question more or less equals answering the question “What states of affairs made a difference to e ’s occurring?”. If this premise is accepted, the kairetic account will follow rather smoothly. Essentially, the causal relation between the difference-makers and e becomes the explanatory relation.⁶ This, I would argue, makes intuitive sense. Both causality and explanations concern events and states of affairs which bring about change. Additionally, both causality and explanation have a directionality associated with them: The direction of movement from cause to effect parallels that from explanans to explanandum. Three central concepts are now introduced, each building on the other (causal models, explanatory kernels and standalone explanations).

The first important concept is a *veridical deterministic atomic causal model* for an explanandum event e (I sometimes use the term “causal model” as a short form in order to simplify the sentences). This is a set of propositions describing the influence of some causal factors and a causal law on the explanandum event e which causally entail that e occurs. In short, there are some causal factors c_1, c_2, \dots, c_n and a causal law l . The law l tells you how e follows from c_1, c_2, \dots, c_n because it represents a particular causal mechanism or process which takes place in the world. l has to be a causal law and as deep as possible in the sense that physical laws are the most elemental, most general and most abstract. The account embraces a certain discrimination against higher order terms in this regard for reasons which will become clear later in the discussion. There is some concern that this may result in reductionism. I agree

partially, but do not believe this to be a problem in itself. The claim is not that every explanandum can be formulated in purely physical terms. However, where a reduction to a lower level is possible, the explanation on the lower level should be preferred over the explanation on the higher level.⁷

A causal model takes the form of a deductive argument with the explanandum event as its conclusion. The model can be mapped onto the causal network of the world, which is the main difference between the kairetic account and the deductive-nomological model. Nevertheless, some resemblance to the deductive-nomological model of scientific explanation remains.

The “veridical” stands for the truth of the propositions in the model, the “deterministic” stands for logical entailment of e under the propositions of the model and the “atomic” stands for a singular causal process that does not specify any details about the causal process other than the premises for and occurrence of e . If one were to map the causal history of the world, an atomic model would represent just one line connecting the causal factors and e . This does not mean that the causal relation cannot be individuated further, just that the causal model is as simple as possible (i.e. some causal factors, one law, one explanandum event). Atomic causal models can be stacked together. A causal factor in an atomic causal model can be made the explanandum event of another atomic causal model and so forth. The result is a compound causal model. Arbitrarily complex causal models can be built this way.⁸ Note that the models I have described so far are deterministic since e is logically entailed by the model. If e were connected to a causal factor via a probabilistic model, the probability of e occurring would be logically entailed by the causal model.⁹ For simplicity’s sake, the remaining discussion here will focus solely on deterministic models. However, every principle can of course be adjusted so that it also holds for probabilistic cases.

The question remains: What are the factors that really make a difference to the occurrence of an explanandum event? In order to answer this question, the notion of an *explanatory kernel* is introduced. An explanatory kernel is a veridical deterministic atomic causal model for e which contains only difference-makers as causal factors, where a difference-maker is defined as follows:

⁶ The explanatory relation is the relation between explanans and explanandum.

⁷ Levels are a metaphor for the relations between the different scientific disciplines. Physics is usually considered the fundamental level science, followed by chemistry, biology and other special sciences.

⁸ Strevens, 2008: pp. 71-74.

⁹ Strevens, 2008: p. 361.

A difference-maker c is a causal factor whose removal from at least one causal model for an explanandum event e results in the model failing to entail e .¹⁰

“Removal” does not mean replacing c with $\neg c$ in the model – it means that the model remains silent about whether c occurs or not. Simply put, if there exists a causal model for e containing c and then the model is adjusted so that nothing is said about c and subsequently the model ceases to entail e , then c is a difference-maker for e .

There are some additional requirements for explanatory kernels once difference-makers are identified. Let us say that an explanatory kernel is denoted as K and a corresponding causal model as M . K must fulfil the following two desiderata:

1. K must maximise the level of abstraction.
2. K must be as cohesive as possible.¹¹

Where do the two requirements emanate from? Starting with the causal model M , when a causal factor is identified with a difference-maker, the question suggests itself what property exactly made a difference in this particular case. To obtain only the relevant factors means removing all irrelevant ones, i.e. abstraction.¹² This requirement equips the kairetic account with the feature which makes the unification theory of scientific explanation attractive: It bundles multiple explananda under the same abstract conditions.¹³ In a more formal manner we might say:

Let A be the conjunction of all propositions in the causal model A^* and B be the conjunction of all propositions in the causal model B^* , where A^* and B^* are causal models for the same explanandum event. B^* is more abstract than A^* iff $A \rightarrow B$ holds but $B \rightarrow A$ does not.¹⁴

The “ \rightarrow ” in this definition does not stand for a causal law, but for a material conditional. With this conception of abstraction a problem arises instantly: $c \rightarrow (c \vee d)$ always holds, but the converse does not. As an example, assume you have a headache. An ato-

mic model for it might be a bike accident in which you hurt your head and a physiological law describing how a headache is a result of that injury. Another atomic model for your headache might be stress with quite different causal processes. Let us suppose you really have chronic stress and the second model is veridical. Then, according to the abstraction criterion, the disjunction of the stress model and the accident model is a better explanation than the stress model alone because it is more abstract.

The problem can be solved by the second desideratum of explanatory kernels, namely cohesion. Cohesion ought to keep abstraction under control, so that causal contiguity is not lost. Fleshing out the meaning of cohesion is not straightforward and going into a possible solution would be irrelevant for the aim of this essay. It is best, I suggest, to picture cohesion as the “causal glue” which prevents absurdities such as a disjunctive model.

Now I will turn to the third step in event explanation: The creation of standalone explanations, that is, how actual explanations can be created. So far, I have illustrated what veridical deterministic atomic causal models are. Then I have explained how explanatory kernels are obtained from these models by means of abstraction. Explanatory kernels contain nothing more than the most abstract difference-makers for the explanandum event. The definition of a standalone explanation is as follows:

A standalone explanation for an event e is an agglomeration of explanatory kernels which are defined recursively:

Let K_1 be an explanatory kernel for e . Then K_1 is a standalone explanation for e .

Take any difference-maker in a standalone explanation and add an explanatory kernel K_2 with said difference-maker as its target. The result will also be a standalone explanation. This process can be repeated ad infinitum – there is no upper bound to the number of reiterations.¹⁵

Let me justify this definition. To give an explanation means to point to some difference-makers and causal processes which resulted in the event occurring. The given definition fully satisfies this desideratum. An explanatory kernel only contains difference-makers for the target event. Combining explanatory kernels will

¹⁰ Strevens, 2008: p. 87.

¹¹ Strevens, 2008: p. 110.

¹² Abstraction is also responsible for the preference of lower-level over higher-level explanation, as lower-level explanations increase the level of abstraction.

¹³ Kitcher, 1981: p. 508.

¹⁴ Strevens, 2008: p. 102.

¹⁵ Strevens, 2008: pp. 118-119.

also result in a model which contains only difference-makers. Logically speaking, however, some difference-makers will become redundant in this process, namely those which become the target of a new explanatory kernel.

One consequence of the definition is that there are many, possibly an infinite number of standalone explanations for *e*. Although this is true, it is not a problem for the definition. Some explanations might be preferred over others. I will not go into this discussion here.¹⁶

All necessary concepts from the kairetic account for the following discussion of genuine mathematical explanation are now given. In a next step, the notion of genuine mathematical explanation will be clarified. Consequently, two arguments will challenge the possibility of genuine mathematical explanation.

3. Genuine Mathematical Explanation

3.1. Characterising the Category

As the name already suggests, genuine mathematical explanations are explanations which heavily rely on mathematics. The mathematical premises do most of the explanatory work, so to speak. At first glance, this phenomenon seems to be at odds with invoking causal mechanisms or processes. Mathematical propositions do not refer to anything with causal power after all. This is why the trilemma mentioned in the introduction to this paper is pressing. Some proponents of genuine mathematical explanation have put forward the so-called *indispensability argument*¹⁷ – an argument supporting the existence of mathematical objects. Since issues with this argument have been raised (especially with confirmational holism, a necessary premise for the indispensability argument) the discussion has focused on specific examples of mathematical explanations in science. These examples should satisfy the following requirements:

1. The explanandum is external to mathematics.
2. The phenomenon in question is truly in need of an explanation and not just pure coincidence.

16 A brief summary: Extending a standalone explanation by adding explanatory kernels might only improve an explanation if the added kernels present more details about the causal past which were previously unknown. Adding other kernels never improves a standalone explanation. For a detailed argument on what improves an explanation and what does not see: Strevens 2008, pp. 123-138.

17 For a detailed defence of the argument see: Colyvan, 2001.

3. The phenomenon is genuinely explained by the mathematical propositions in the explanans, meaning that the mathematical propositions do not merely represent concrete objects and their structural relations.¹⁸

The first requirement is necessary because otherwise the explanation would be a purely mathematical explanation of a mathematical explanandum. The second requirement is necessary for all explanations since without it the explanandum would be lacking an explanation, i.e. it would be pure chance. The third requirement adds that the explanation must rely on mathematical propositions in their own right in order to be genuinely mathematical. If the mathematical proposition is only explanatory insofar as it mirrors concrete objects and their structural relations, then the explanatory power of the proposition stems from its representational capacities and is fundamentally non-mathematical.¹⁹

Two important conclusions are drawn from the phenomenon of genuine mathematical explanation. The first is the discussion concerning the existence of mathematical objects.

An argument for their existence can be made as follows: There are genuine mathematical explanations. Those explanations make reference to mathematical objects. We are justified in believing in the existence of any object indispensable for our best scientific explanations. Therefore, we are justified in believing in the existence of some mathematical objects. This conclusion implies the existence of abstract objects.

The second conclusion that has been drawn is that there is some kind of explanation which is to a significant degree non-causal. Recalling the trilemma, this means that the proposition “All scientific explanations are causal explanations” must be rejected. This is a possible solution to the trilemma which I, however, want to argue is the wrong move. The kairetic account of scientific explanation does rely on some causal notions. Difference-makers collectively imply the explanandum event by use of a causal process.²⁰ Difference-makers are abstract causes/causal factors. What I take to be the most imminent issue with genuine mathemati-

18 Baker, 2005: pp. 233-234.

19 I use the term “concrete objects and structural relations” to denote non-abstract objects with causal capacities and causal relations.

20 The kairetic account does not itself define the notion of causality, cause and effect, and so on. Nevertheless, causal processes are invoked in the account itself. This is not a problem, I think, because they are characterised as the fundamental physical laws ideally and by higher order laws in other cases.

cal explanation is the following: The kairetic account of explanation must invoke difference-makers (causal factors) and causal processes for any explanation and most of the explanatory power comes from showing how the given difference-makers together with the causal process imply the explanandum event. Genuine mathematical explanation relies for the most part on other explanatory factors, whatever these may be. Hence, there is a conflict waiting for a solution. I propose that there are no genuine mathematical explanations and will present my argument in support of this conclusion below.

3.2. Two Arguments against Genuine Mathematical Explanation

An important methodological question ought to be raised at the beginning of this discussion: How do philosophers identify genuine mathematical explanations? It usually works in the form of a discourse where specific examples are put forward. When the explanations are considered, they seem to rely on mathematical premises for the most part and the conclusion is best grasped through understanding the relevant mathematical premises. As I have already proposed earlier, a scientific explanation should not focus on the communicative sense of the term “explanation” but on its ontological reading. An analysis of the explanatory relations in the kairetic account reveals an important point: Mathematics helps us understand the phenomenon through an adequate representation of the difference-makers and causal processes. The fact that mathematics is able to represent concrete objects and structures is also accepted by proponents of genuine mathematical explanation. The views start to differ when it comes to the question if there is more to mathematics than this representational role. If there is not, then the application of mathematics will be perfectly compatible with causal theories of explanation. I suggest the role of mathematics is representational only and will present two arguments to defend this position. These two arguments are categorised as follows: The first argument concerns explanations which require the explanans to be true (veridical explanations). Although it might seem intuitive at first, not all explanations are veridical, that is, some explanations justifiedly contain false propositions in the explanans. I use the term “justifiedly” to indicate that sometimes the falsehood of a proposition invalidates an explanation. As an example, I recall my hypothetical account of an accident during a game of tennis and my bruised knee. My accident and the physiological

processes involved are an explanation for my having a bruised knee as long as the explanans is actually true. Were one to find out that I have never played tennis in my entire life, then the explanation would surely no longer be legitimate.

Idealisations are different, however. They provide examples of explanations in which part of the explanans is blatantly false, we know that part of it is false and yet the explanation does not lose its validity. This is why I will present two arguments, one for veridical explanations and one for idealised explanations.

Although the propositions are stated generally, the scope of the following propositions has to be confined to veridical explanations.²¹

1. The mathematical propositions p_1, p_2, \dots, p_n are explanatory for empirical explananda only if the mathematical objects and structures in p_1, p_2, \dots, p_n represent concrete objects and their structural relations.
2. If the mathematical objects and structures in p_1, p_2, \dots, p_n represent concrete objects and their structural relations, then p_1, p_2, \dots, p_n are true iff they merely represent concrete objects and their structural relations.
3. p_1, p_2, \dots, p_n are explanatory for empirical explananda.
4. Therefore: p_1, p_2, \dots, p_n are true iff they merely represent concrete objects and their structural relations.
5. An explanation for an empirical explanandum is genuinely mathematical only if p_1, p_2, \dots, p_n do not merely represent concrete objects and their relations.
6. p_1, p_2, \dots, p_n are explanatory for empirical explananda only if said propositions are true.
7. Therefore: An explanation for an empirical explanandum cannot be genuinely mathematical.

²¹ In the argument I use p_1, p_2, \dots, p_n in order to speak of all mathematical propositions featuring in veridical explanations for empirical explananda. This allows me to present the argument in a universal form.

Obviously, this argument is not convincing where idealisations are concerned. Before I turn to idealisations, let me justify the truth of the premises in the above argument.

The first premise holds because mathematics is an abstract discipline whose content does not depict the world itself. Mathematical objects can be seen as elements with specific properties. The properties are abstract and prima facie not connected to any other discipline. The meaning of mathematical propositions is therefore very abstract. In this state, mathematical concepts are thought of as different from any concrete objects or structures. However, when mathematics is applied to the world, the mathematical structures are taken to represent concrete ones. The representation creates relevance. Without such a representation mathematical concepts are irrelevant in the empirical context. Only if such a relation is established does mathematics become relevant for empirical phenomena and, notably, become explanatorily relevant. In this way a representational relation between mathematical structures and structures between concrete objects is required to render the mathematics explanatory for an empirical explanandum. The representation serves the purpose of modelling. In this way it helps us to understand some relevant abstract information necessary to grasp the explanandum.

The second premise relies on the principle that applied mathematical propositions are true iff they only represent structures between concrete objects. This condition is rather different from a general theory of mathematical truth, because the latter concerns the truth of mathematical propositions in their abstract form. The proposition “ $7 + 5 = 12$ ” may or may not be true in the abstract, yet when it is applied, its truth conditions change accordingly. For example, “Seven apples plus five apples is equal to twelve apples” is true iff the structural relations between individual apples is such that “1” denotes an individual apple and the structures between apples are such that addition can be applied. The number “7” then denotes one bunch of apples, the number “5” another one and the number “12” their total. In this context, only concrete objects and their relations are required for the truth of the proposition “Seven apples plus five apples is equal to twelve apples”. The same holds for p_1, p_2, \dots, p_n .²²

The third premise refers to the fact that mathematics can be explanatory for empirical explananda. This takes examples of genuine mathematical explanation

seriously and treats them as examples of explanation. What is concluded from the first three premises is that the mathematical propositions in an explanation are true iff they merely represent concrete objects and their relations.

The fourth premise (proposition number five) recalls that a genuinely mathematical proposition ought to provide more for the explanation than a mere indexing of concrete objects and their structural relations.

The fifth premise indicates that the propositions in the explanans of a good explanation must be true or approximately true. Of course there are historical examples of scientific explanations which are considered false by today’s standards, for example the existence of planet “Vulcan” required to explain Mercury’s peculiar orbit in classical mechanics (as we now know, there is no planet between Mercury and the sun).²³ A response to that counterargument is that we also no longer regard those historical explanations as real explanations – the existence of Vulcan is not a good explanation for Mercury’s orbit because Vulcan does not exist.

Another problem arises when idealisations are used in an explanation, i.e. propositions which are deliberately used even though their falsehood is well known. This is why I propose another argument for idealisations here:²⁴

1. There are idealised explanations for empirical explananda in science with the false mathematical propositions f_1, f_2, \dots, f_n .
2. If f_1, f_2, \dots, f_n feature in an idealised explanation, then they are explanatory only insofar as they help to draw attention to difference-makers.
3. If f_1, f_2, \dots, f_n are explanatory only insofar as they help to draw attention to difference-makers, then the explanation is not genuinely mathematical.
4. Therefore: An idealised explanation for an empirical explanandum cannot be genuinely mathematical.

²³ Yagisawa, 2018.

²⁴ Analogously to above, I use f_1, f_2, \dots, f_n in order to make reference to all possible mathematical propositions in idealised explanations. While the first argument against genuine mathematical explanation remains neutral on the trilemma mentioned in the introduction, the second already presupposes that the kairetic account of scientific explanation is correct. Consequently, the second argument should be treated as an argument which can be made from a defender of causal theories of scientific explanation.

²² A version of this so-called mapping account can be found in Pincock, 2004: p. 70.

Following the same structure of argument as used before, I will now have a closer look at each premise to demonstrate its truth.

The first premise encapsulates the fact that there are idealised explanations containing mathematical propositions. The examples are numerous: maximally rational humans in economics, infinite populations in biology, non-existent air-resistance in physics, etc.²⁵

The second premise is justified by some general characteristics of idealisations: The mathematical propositions obviously fail to represent structures of concrete objects and are false in this regard, they are non-difference-makers according to the kairetic account, consequently not specified by any standalone explanation and therefore open for any specification consistent with the required difference-makers. Consider, for example, an electron and a proton attracting each other. The trajectories can best be evaluated by only considering the electromagnetic force and neglecting gravity since the latter is about $2 \cdot 10^{39}$ times smaller than the first. For many situations gravity will not be a difference-maker and may therefore be assigned the null vector in classical or a non-existing gravitational potential in quantum mechanics respectively. The ability of mathematics to easily assign extreme or default values such as 0, ∞ or others to highlight the value's insignificance is particularly beneficial for an explanation.²⁶

The third premise concludes from this that an explanation with false mathematical propositions cannot be genuinely mathematical – the mathematical propositions do not explain in their own right. Their role is to draw attention away from irrelevancies to difference-makers and add the pragmatic benefit of simplifying mathematical expressions. This is of purely instrumental benefit because what really explains the explanandum are the difference-makers, not the false mathematical propositions.

4. Conclusion

Let me review the main arguments I have made in this essay:

First of all, the use of mathematics can be justified in causal theories of scientific explanation. Additionally, the role of mathematics in veridical models or, more generally, veridical explanations is confined to its ability to mirror the world because this is the most sensible way in which an applied mathematical proposition

can be said to be true. Some issues may arise in the context of my overall argument. I would like to address a number of them. A first counterargument that might be made is that the kairetic account is wrong. If the kairetic account is wrong, then I hope that the characteristics necessary for the general argument are also exemplified in other causal theories of explanation.²⁷ Consequently, the argument does not rely specifically on the kairetic account and instead on more general features that causal theories of explanation exemplify. A second problem might arise in connection with my definition of genuine mathematical explanation. While I believe that to define genuine mathematical explanation is not an easy task, I have attempted to include the most important characteristics, for example that the explanation relies on mathematical premises for the most part and that these mathematical premises are explanatory in their own right, i.e. mathematics is not just explanatory in so far as it mirrors concrete objects and structures.

A third challenge might arise when the applicability of mathematics is considered. How can mathematics be so useful to science if its explanatory faculty is only instrumental? It achieves this goal by improving the communicative sense of an explanation. It provides a clear formalism, deductive rigour and generally accepted manipulation rules. Consequently, the deductive force in an explanation (which is a type of argument) is massively increased. Furthermore, stating scientific laws in mathematical terms allows for a relatively easy identification of difference-makers.

Whatever the results of future research may be, the puzzle of mathematics and its explanatory faculty in science is certain to be solved. Whether the solution I have suggested here proves to be the correct approach remains open to discussion and further exploration.

25 Strevens, 2008: pp. 322-323.

26 Strevens, 2008: p. 318.

27 It would be even more interesting to construct a more general argument against genuine mathematical explanation for idealisations that does not rely on causal theories of explanation.

References

- Baker, Alan: Are there genuine Mathematical Explanations of Physical Phenomena?, in: *Mind, New Series*, Vol. 114, No. 454 (Apr., 2005), pp. 223-238.
- — : Mathematical Explanation in Science, in: *The British Journal for the Philosophy of Science*, Vol. 60, No. 3 (September, 2009), pp.611-633.
- Colyvan, Mark: *The Indispensability of Mathematics*, Oxford University Press, New York, 2001.
- Kitcher, Philip: Explanatory Unification, *Philosophy of Science*, Vol. 48, No. 4 (December, 1981), pp. 507-531.
- Lange, Marc: *Because without Cause*, Oxford University Press, New York, 2017.
- Lewis, David: *Causal Explanation*, in: *Philosophical Papers*, Oxford University Press, New York, 1986.
- Pincock, Christopher: A Revealing Flaw in Colyvan's Indispensability Argument, in: *Philosophy of Science*, Vol. 71, No. 1 (January, 2004), pp. 61-79.
- — : *Mathematics and Scientific Representation*, Oxford University Press, Oxford, 2011.
- Salmon, Wesley C.: Causality and Explanation, in: *Philosophy of Science*, 64 (September, 1997) pp. 461 -477.
- Strevens, Michael: *Depth, An Account of Scientific Explanation*, Harvard University Press, Cambridge, 2008.
- — : The Mathematical Route to Causal Explanation, To appear in *Explanation Beyond Causation*, edited by Alexander Reutlinger and Juha Saatsi, Oxford University Press. URL = <http://www.strevens.org/research/onlinepapers.shtml>. Accessed on August 7, 2018.
- Ráz, Tim: Euler's Königsberg: The Explanatory Power of Mathematics. in: *European Journal for Philosophy of Science*, ISSN 1879-4912, 2017. URL = <http://philsciarchive.pitt.edu/14442/>. Accessed on July 18, 2018.
- Yagisawa, Takashi: Possible Objects, *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/possible-objects/>. Accessed on December 10, 2018.

Andreas Lüchinger (22) schliesst gegenwärtig sein Bachelorstudium in Philosophie mit Nebenfach Physik ab. Er interessiert sich für Wissenschaftsphilosophie, Logik, Philosophie der Logik, Philosophie der Mathematik, Erkenntnistheorie und Metaphysik.

Ein bayesianischer Lösungsansatz für das Raben-Paradox

1. Ausgangslage

Zu Gunsten des Bayesianismus wird manchmal angeführt, er ermögliche eine Lösung des von Carl Gustav Hempel formulierten Raben-Paradoxes. Das Raben-Paradox leitet aus einigen grundsätzlich plausiblen Prämissen die unbequeme Konklusion ab, dass jede Beobachtung eines nicht-schwarzen Objekts, das kein Rabe sei – beispielsweise eines weissen Schuhs –, die Hypothese, alle Raben seien schwarz, stütze.¹

Ein bayesianischer Lösungsvorschlag für das Raben-Paradox, den Colin Howson und Peter Urbach stark machen, läuft im Kern darauf hinaus, die auf den ersten Blick unannehmbar scheinende Konklusion für annehmbar zu erklären. Ausgehend vom bayesianischen Relevanzkriterium wird argumentiert, die Beobachtung eines nicht-schwarzen Nicht-Raben stütze zwar die Annahme, dass alle Raben schwarz seien, die Stützung sei aber derart minimal, dass sie akzeptiert werden könne.²

Mark Siebel wendet in „Der Rabe und der Bayesianist“³ gegen diesen Lösungsversuch ein, es sei grundsätzlich nicht einsehbar, weshalb eine geringe Stützung annehmbar sein soll, und wenn, dann ergebe sich die befremdliche Konsequenz, dass bloss genügend nicht-schwarze Nicht-Raben beobachtet werden müssten, um eine substantielle Stützung für die Annahme, alle Raben seien schwarz, zu erhalten. Damit stelle sich das Raben-Paradox erneut, denn auch die wiederholte Beobachtung beliebiger nicht-schwarzer Nicht-Raben sei irrelevant für die Frage, ob alle Raben schwarz seien.⁴

In dieser Arbeit sollen Siebels Einwände gegen den Lösungsversuch von Howson und Urbach kritisiert werden, indem gezeigt wird, dass es durchaus plausibel ist, im Einzelfall eine geringe und in hinreichend vielen Beobachtungen von nicht-schwarzen Nicht-Raben eine substantielle Stützung für die Raben-Hy-

pothese zu sehen. Dazu werden zunächst Hempels Raben-Paradox, der Lösungsversuch von Howson und Urbach und Siebels Einwände skizziert. Anschliessend wird der Versuch unternommen, Siebels Einwände zu entkräften.

2. Das Raben-Paradox

Das Raben-Paradox ist eingebettet in die Debatte darüber, unter welchen Umständen experimentelle Daten für oder gegen eine empirische Hypothese sprechen – eine Hypothese stützen oder widerlegen – und wann Daten für eine Hypothese irrelevant sind. Das Paradox besteht aus drei Prämissen und einer Konklusion, die aus den Prämissen folgt. Um ein Paradox handelt es sich deshalb, weil zwar gute Gründe für die Prämissen sprechen, die Konklusion aber intuitiv wenig plausibel ist. Die Prämissen sind:⁵

- (a) Hypothesen der Form „alle A sind B“ werden gestützt durch jede Beobachtung von B in einer Instantiierung von A. Umgekehrt spricht die Beobachtung einer Instantiierung von A ohne B gegen die Hypothese.⁶
- (b) Hypothesen der Form „alle A sind B“ lassen sich als universale Konditionalsätze verstehen und mit prädikatenlogischen Mitteln fassen.⁷
- (c) Sind zwei Hypothesen äquivalent, dann gilt: Daten, die eine der beiden Hypothesen stützen, stützen ebenso die andere.⁸

Gemäss (b) können Hypothesen der Form „alle A sind B“ als universale Konditionalsätze verstanden und im Sinn der Prädikatenlogik als „ $\forall x(A(x) \rightarrow B(x))$ “ interpretiert werden.

1 Vgl. Colin Howson/Peter Urbach. 1989. *Scientific Reasoning. The Bayesian Approach*. Chicago/Lasalle Ill.: Open Court. Hier: 127-130.

2 Vgl. Ebd.

3 Mark Siebel. 2004. „Der Rabe und der Bayesianist“. In: *Zeitschrift für allgemeine Wissenschaftstheorie* 35, 313-329.

4 Vgl. Ebd. 318-320.

5 Eine Darstellung des Raben-Paradoxes findet sich in: Carl Gustav Hempel. 1965. „Studies in the Logic of Confirmation“. In: Hempel, Carl Gustav (Hg.). *Aspects of Scientific Explanation. And Other Essays in the Philosophy of Science*. New York: The Free Press. 3-46. Hier: 14-20.

6 Vgl. Ebd. 10.

7 Vgl. Ebd. 15-16.

8 Vgl. Ebd. 13.

Die Hypothese „Alle Raben sind schwarz“ lässt sich somit formalisieren als „ $\forall x(\text{Rabe}(x) \rightarrow \text{Schwarz}(x))$ “, wobei „Rabe“ für das Prädikat „...ist ein Rabe“ und „Schwarz“ für „... ist schwarz“ steht.

Nach (a) stützt die Beobachtung eines Objekts, das sowohl ein Rabe als auch schwarz ist, die Hypothese, alle Raben seien schwarz (im Folgenden Hypothese h genannt). Ein Objekt, das weder schwarz, noch ein Rabe ist, stützt die Hypothese, was nicht schwarz ist, sei kein Rabe (Hypothese h'). Formal ausgedrückt: ein schwarzer Rabe stützt h „ $\forall x(\text{Rabe}(x) \rightarrow \text{Schwarz}(x))$ “ und ein nicht-schwarzer Nicht-Rabe stützt h' „ $\forall x(\neg \text{Schwarz}(x) \rightarrow \neg \text{Rabe}(x))$ “. Das folgt aus (a), und ist soweit plausibel.

Nun zeigt Hempel aber, dass in Kombination mit der Äquivalenzbedingung (c) daraus Sonderbares folgt. Denn die beiden Hypothesen h und h' sind prädikatenlogisch äquivalent. Da nun nach (c), was eine von zwei äquivalenten Hypothesen stützt, auch die andere äquivalente Hypothese stützt, und da die Beobachtung eines nicht-schwarzen Nicht-Raben die zu h äquivalente Hypothese h' stützt, folgt: Die Beobachtung irgendeines nicht-schwarzen Objekts, das kein Rabe ist, stützt die Hypothese „Alle Raben sind schwarz“. Diese Konsequenz, die sich aus den drei eigentlich überzeugenden Prämissen des Paradoxes ergibt, sei aber, so Hempel, intuitiv wenig attraktiv. Denn die Beobachtung beispielsweise eines weissen Schuhs, einer blauen Baseballmütze oder eines roten Ruderbootes scheinere irrelevant zu sein für die Hypothese, dass alle Raben schwarz seien. Wer etwas über die Farbe von Raben wissen wolle, müsse sich Raben und nicht sein Schuhregal anschauen.⁹

Will man ein Paradox auflösen, bieten sich allgemein drei Möglichkeiten. Man kann zeigen, dass in der Herleitung der Konklusion aus den Prämissen ein Fehler unterlaufen ist und die Prämissen nicht auf die Konklusion verpflichtet. Man kann überdies eine der Prämissen zurückweisen oder ferner die Konklusion akzeptieren als durchaus überraschende und zunächst unbequeme Konsequenz dreier überzeugender Prämissen, die unter näherer Betrachtung aber an Plausibilität gewinnt.

Die zweite Strategie verwirft Hempel. Zwar räumt er ein, dass die Prämisse (a) kein umfassendes Kriterium liefere, um zu entscheiden, ob Daten eine Hypothese stützen würden, weil sich das in (a) formulierte Kriterium ausschliesslich auf Hypothesen in der Form universaler Konditionalsätze anwenden lasse. Eine

notwendige Bedingung für Stützung liefere die Prämisse (a) daher nicht. Dennoch scheinere es nicht sinnvoll, die Überzeugung fallenzulassen, dass jedes Objekt, das sowohl A als auch B ist, die Hypothese „alle A sind B“ bestätige,¹⁰ weshalb (a) weiterhin als *hinreichende* Bedingung für Stützung gesehen werden dürfe – wenngleich eingeschränkt auf Hypothesen, die als universale Konditionalsätze gefasst werden könnten.¹¹ (a) als hinreichendes Kriterium für diese Hypothesen zu akzeptieren, reicht bereits, damit in Verbindung mit (b) und (c) das Paradox resultiert. Die Prämisse (a) zurückzuweisen, bietet daher keine angemessene Lösung des Paradoxes.

Verzichte man auf die Äquivalenzbedingung (c), hänge es von der Formulierung der Hypothese ab, ob Daten diese stützen würden. Das sei, so argumentiert Hempel, abwegig und werde darüber hinaus der Weise nicht gerecht, wie äquivalente Hypothesen in theoretischen wissenschaftlichen Überlegungen behandelt würden: Eine gut bestätigte Hypothese könne etwa als Prämisse in einem deduktiven Argument dienen. Ein solches Argument sollte den Regeln der formalen Logik genügen, nach denen die Ersetzung einer Prämisse durch eine äquivalente Aussage die Gültigkeit eines Schlusses nicht beeinträchtigt.¹²

Die Prämissen (a) und (c) lassen sich demnach schwerlich aufgeben. Ob das gleichermassen für die Prämisse (b) gilt, ist indessen weniger augenfällig. Tatsächlich könnte man gegen (b) einwenden, Aussagen der Form „Alle As sind Bs“ als „ $\forall x(A(x) \rightarrow B(x))$ “ zu formalisieren, werde dem Umstand nicht gerecht, dass wenn in alltäglichen Situationen eine solche Hypothese vorgebracht werde, darin oft die zusätzliche Behauptung mitschwinge, es existiere ein solches A. Diese Intuition liesse sich etwa berücksichtigen, indem allgemeine empirische Hypothesen als Kombination aus einem universalen Konditional und einer Existenzbehauptung verstanden würden. Die Äquivalenz zwischen h und h' wäre dann nicht mehr gegeben.

Hempel argumentiert allerdings dafür, an dieser Stelle nicht (b) zurückzuweisen oder abzuwandeln, weil auch in vielen Fällen mit einer allgemeinen empirischen

⁹ Vgl. Hempel. 1965. 14-15.

¹⁰ Howson und Urbach führen hier Fälle an, in denen (a) zwar erfüllt ist, es aber dennoch nicht unmittelbar plausibel scheint, eine Stützung anzunehmen. Sollten diese Gegenbeispiele adäquat sein, wäre (a) möglicherweise auch als hinreichende Bedingung für Stützung ungeeignet und die zweite Strategie, eine der Prämissen abzulehnen, erfolgsversprechend. In dieser Arbeit soll allerdings der Lösungsversuch von Howson und Urbach diskutiert werden, der auf der dritten Strategie aufbaut, weshalb die Frage, inwiefern (a) zurückzuweisen ist, im Folgenden nicht weiter diskutiert wird. Vgl. Howson/Urbach. 1989. 90-91.

¹¹ Vgl. Hempel. 1965. 11-15.

¹² Vgl. Ebd. 13.

Hypothese offensichtlich keine Existenz impliziert werde.¹³ Stattdessen schlägt er vor, die dritte Strategie zu wählen, wonach die Konklusion des Paradoxes bloss *prima facie* unhaltbar sei. Die Intuition, die Konklusion sei schlicht abzulehnen, sei eine Täuschung und teilweise zurückzuführen auf die beschriebene alltagssprachliche Unschärfe von allgemeinen Hypothesen, nicht aber auf Mängel in den Prämissen des scheinbaren Paradoxes.¹⁴

Der hier untersuchte bayesianische Lösungsansatz von Howson und Urbach verfolgt ebenfalls die dritte Strategie, ist allerdings theoretisch umfassender ausgearbeitet. Einige Elemente von Hempels Lösungsversuch für das Paradox werden später hilfreich sein, um Siebels Einwänden gegen den bayesianischen Lösungsansatz zu begegnen.

3. Bayesianismus

Bevor nun der Lösungsansatz von Howson und Urbach dargestellt wird, sollten zuerst einige bayesianische Grundideen skizziert werden. Der folgende kurze Abriss verläuft weitgehend entlang der Darstellung, die Howson und Urbach in *Scientific Reasoning* geben.¹⁵

Die Menge der möglichen Fälle, über die eine Hypothese etwas aussagt, ist oft um ein Vielfaches grösser, als die Menge jener Einzelfälle, die effektiv beobachtet wurden. Es stellt sich daher die Frage, wie anhand von Daten entschieden werden kann, ob eine Hypothese plausibel ist, wenn doch verglichen mit der Reichweite der Hypothese bloss wenige Einzelfälle als Daten vorliegen.¹⁶ Die bayesianische Idee dazu ist, die Plausibilität einer Hypothese als die Wahrscheinlichkeit, mit der die Hypothese wahr ist, zu interpretieren. Zu sagen, Daten würden eine Hypothese stützen, bedeutet dann nicht, dass sie die Wahrheit der Hypothese garantieren, sondern dass die Daten die Wahrscheinlichkeit der Hypothese erhöhen.¹⁷

Die Stützungsrelation zwischen einer Hypothese und empirischen Daten wird demnach durch Wahrscheinlichkeiten beschrieben und auf diese Wahrscheinlichkeiten finden die Prinzipien der Wahrscheinlichkeitsrechnung Anwendung. Ist eine nicht-leere Menge S

von Sätzen a, b, c, ..., gegeben, denen eine Wahrscheinlichkeitsfunktion P reelle Zahlen zwischen 0 und 1 zuordnet, gelten folgende Axiome:¹⁸

$$(1) \quad P(a) \geq 0, \text{ für alle } a \text{ in } S$$

$$(2) \quad P(t) = 1, \text{ wenn } t \text{ tautologisch ist.}$$

$$(3) \quad P(a \vee b) = P(a) + P(b), \text{ wenn } a, b \text{ und } a \vee b \text{ in } S \text{ sind und } a \text{ und } b \text{ zusammen inkonsistent.}$$

$$(4) \quad P(a | b) = \frac{P(a \wedge b)}{P(b)}, \text{ wenn } a, b \text{ und } a \wedge b \text{ in } S \text{ sind, und wenn } P(b) \neq 0.$$

$P(a | b)$ ist die bedingte Wahrscheinlichkeit von a, gegeben b. Für bedingte Wahrscheinlichkeiten gilt Bayes' Theorem:¹⁹

$$P(a | b) = \frac{P(b | a) \cdot P(a)}{P(b)}$$

Um die Frage zu klären, ob Daten d eine Hypothese h stützen, ist Bayes' Theorem hilfreich, weil es ermöglicht, die bedingte Wahrscheinlichkeit $P(h | d)$ – die Nachher-Wahrscheinlichkeit der Hypothese –²¹ zu berechnen, wenn die Ausgangs-Wahrscheinlichkeit der Hypothese $P(h)$, $P(d)$ und die bedingte Wahrscheinlichkeit $P(d | h)$ bekannt sind.²² Ob und wie sich ein Set von Daten auf die Plausibilität einer Hypothese auswirkt, hängt alsdann davon ab, ob und wie sich die Nachher-Wahrscheinlichkeit von der Ausgangs-Wahrscheinlichkeit der Hypothese unterscheidet:²³

$$P(h | d) > P(h) - d \text{ stützt } h$$

$$P(h | d) = P(h) - d \text{ ist neutral zu } h$$

$$P(h | d) < P(h) - d \text{ spricht gegen } h$$

13 Eine ausführlichere Argumentation, weshalb die Interpretation von allgemeinen empirischen Hypothesen als universale Konditionale nicht zurückgewiesen werden sollte, gibt Hempel in: Hempel. 1965. 16-17.

14 Vgl. Ebd. 18-20.

15 Vgl. Howson/Urbach. 1989. 15-16, 26, 79-81.

16 Vgl. Ebd. 3.

17 Die Debatte darüber, was als ideales Stützungsmaß gilt, wird hier ausgeklammert. Für das vorliegende Unterfangen reicht es, festzuhalten, dass der Stützungsgrad mit dem Ausmass der Wahrscheinlichkeitserhöhung korrespondiert. Vgl. Howson/Urbach. 1989. 8-10, 79-80.

18 Vgl. Ebd. 15-16.

19 Vgl. Ebd. 26.

20 Zusätzlich zu den Daten wird die Nachher-Wahrscheinlichkeit der Hypothese gewöhnlich noch von Hintergrundinformationen abhängig gemacht. Wenngleich der Hintergrund in der folgenden Argumentation eine gewisse Rolle spielt, wird er aus den Formeln weggelassen um eine bessere Übersichtlichkeit zu gewährleisten.

21 Die Identifikation der Nachher-Wahrscheinlichkeit mit $P(d | h)$ folgt aus dem Prinzip der bayesianischen Konditionalisierung, welches seinerseits umstritten ist. Auf die Debatte, ob es als Rationalitätsforderung anerkannt werden sollte, kann hier allerdings nicht eingegangen werden.

22 Freilich dürfte es in vielen Fällen schwierig sein, diese Wahrscheinlichkeiten konkret zu bestimmen. Der hier behandelte Lösungsansatz für das Raben-Paradox setzt allerdings nicht voraus, die einzelnen Wahrscheinlichkeiten exakt zu kennen.

23 Vgl. Howson/Urbach. 1989. 79-81.

4. Bayesianische Lösung

Auf dieser theoretischen Grundlage lasse sich, so Howson und Urbach, das Raben-Paradox lösen: Was die Konklusion des Paradoxes so befremdlich erscheinen lasse, sei, dass sie nahezu legen scheine, beliebige nicht-schwarze Objekte seien ebenso gut geeignet wie schwarze Raben, um die Hypothese „Alle Raben sind schwarz“ zu stützen. Es wäre aber trügerisch zu glauben, bloss weil die Beobachtung eines weissen Schuhs die Raben-Hypothese stütze, brauche man keine Raben mehr zu untersuchen, sondern könne stattdessen bequem sein Schuhregal etwas genauer betrachten und schon finde man Daten, um die Hypothese substantiell zu stützen. Denn auch wenn sowohl schwarze Raben als auch weisse Schuhe geeignet seien, die Hypothese zu stützen, schwarze Raben und weisse Schuhe würden sie nicht gleichermassen stützen. Tatsächlich sei der Stützungsgrad, den ein weisser Schuh für die Raben-Hypothese generiere, äusserst gering, und damit, argumentieren Howson und Urbach, liefere der Bayesianismus die richtige Antwort auf das Paradox. Der Stützungsgrad sei nämlich dermassen klein, dass er vernachlässigt werden könne und die Konklusion ihren absonderlichen Anschein verliere.²⁴

Diesem Lösungsansatz liegt die Überlegung zugrunde, dass die Beobachtung eines nicht-schwarzen Objekts, das kein Rabe sei, die Raben-Hypothese nur marginal stütze. Sie bildet Ausgangspunkt der Argumentation und sie lässt sich rechnerisch überprüfen. Siebel führt das folgendermassen vor:²⁵

Die Nachher-Wahrscheinlichkeit der Hypothese h „Alle Raben sind schwarz“, angesichts der Beobachtung eines Objekts, das nicht schwarz und kein Rabe ist – hier abgekürzt durch „ $\neg S \wedge \neg R$ “ – lässt sich gemäss Bayes' Theorem berechnen:

$$P(h \mid \neg S \wedge \neg R) = \frac{P(h) \cdot P(\neg S \wedge \neg R \mid h)}{P(\neg S \wedge \neg R)}$$

$$= \frac{P(h) \cdot P(\neg S \mid h) \cdot P(\neg R \mid h \wedge \neg S)}{P(\neg S) \cdot P(\neg R \mid \neg S)}$$

Die Hypothese h „ $\forall x(\text{Rabe}(x) \rightarrow \text{Schwarz}(x))$ “ ist äquivalent mit der Hypothese h' „ $\forall x(\neg \text{Schwarz}(x) \rightarrow \neg \text{Rabe}(x))$ “. $\neg R$ folgt daher aus h und $\neg S$. Somit gilt:

$$P(\neg R \mid h \wedge \neg S) = 1$$

²⁴ Umgekehrt liesse sich zeigen, dass ein Schwarzer Rabe die Hypothese deutlich stärker stützt, als ein nicht-schwarzes Objekt, insbesondere dann, wenn die Hypothese noch vergleichsweise schlecht bestätigt ist und ihre Wahrscheinlichkeit entsprechend gering. Vgl. Ebd. 1989. 89-90, 101.

²⁵ Vgl. Siebel. 2004. 316-317.

Ferner wird die Wahrscheinlichkeit, dass ein beliebiges Objekt schwarz ist, von der Hypothese, alle Raben seien schwarz, nicht beeinflusst: $\neg S$ ist unabhängig von h . Die bedingte Wahrscheinlichkeit von $\neg S$, gegeben h , darf daher vereinfacht werden zur Wahrscheinlichkeit von $\neg S$, womit für die Nachher-Wahrscheinlichkeit von h resultiert:

$$P(h \mid \neg S \wedge \neg R) = \frac{P(h) \cdot P(\neg S) \cdot 1}{P(\neg S) \cdot P(\neg R \mid \neg S)}$$

$$= \frac{P(h)}{P(\neg R \mid \neg S)}$$

Die Höhe des Stützungsgrads, der sich aus der Beobachtung eines nicht-schwarzen Nicht-Raben für h ergibt, entscheidet sich daran, um wie viel die Nachher-Wahrscheinlichkeit die Ausgangs-Wahrscheinlichkeit von h übersteigt. Das wiederum hängt von der Wahrscheinlichkeit $P(\neg R \mid \neg S)$ ab und diese dürfte ziemlich hoch sein. Denn es gibt viel mehr Objekte, die nicht schwarz sind, als es Raben gibt, was selbst dann noch zuträfe, wären nicht alle Raben schwarz. Wenn aus der Gruppe der nicht-schwarzen Objekte ein beliebiges ausgewählt wird, ist die Wahrscheinlichkeit verschwindend klein, dass es sich dabei um einen Raben handelt. Umgekehrt liegt $P(\neg R \mid \neg S)$ – in anderen Worten, die Wahrscheinlichkeit, einen Nicht-Raben aus der Gruppe der nicht-schwarzen Objekte auszuwählen – sehr nahe bei 1. Diese Überlegungen sind Teil der Hintergrundinformationen.

Die Nachher-Wahrscheinlichkeit der Hypothese ist also grösser als ihre Ausgangs-Wahrscheinlichkeit, der Unterschied aber ist, weil $P(\neg R \mid \neg S)$ so nahe an 1 liegt, unbedeutend klein. Der weisse Schuh stützt die Raben-Hypothese zwar, aber er stützt sie kaum.²⁶

5. Siebels Einwände gegen die bayesianische Lösung

Diese Berechnung scheint die Argumentation von Howson und Urbach prinzipiell zu unterstützen. Mindestens bestätigt sie die grundlegende Annahme des ungleichen Stützungsgrads. Siebel konstatiert allerdings, das Paradox sei damit nur dann gelöst, wenn die verfeinerte Konklusion, ein weisser Schuh stütze die Raben-Hypothese minimal, wirklich unproblematisch sei. Und das zieht er aus zwei Gründen in Zweifel.²⁷

Erstens sage die Hypothese etwas über die Färbung von Raben aus, und diese gelte es zu untersuchen, um die Hypothese auf ihre Plausibilität hin zu prüfen. Gäbe es nun irgendeinen Zusammenhang zwischen der Farbe

²⁶ Vgl. Siebel. 2004. 316-317.

²⁷ Vgl. Ebd. 318.

von Raben und der Existenz von weissen Schuhen, der nahelegen würde, dass alle Raben schwarz seien, falls es weisse Schuhe gebe, und dieser Zusammenhang wäre ausserdem Teil der Hintergrundinformationen, dann liesse sich glaubwürdig eine Stützung der Raben-Hypothese durch weisse Schuhe attestieren. Bloss gehöre zum Hintergrundwissen zwar, dass die Zahl der nicht-schwarzen Objekte jene der Raben übersteige, eine Verbindung der beschriebenen Art zwischen weissen Schuhen und der Farbe von Raben sei darin dennoch nicht enthalten.

Ohne einen solchen Zusammenhang als Teil des Hintergrundwissens verrate uns die Farbe jeglicher Nicht-Raben nichts über die Farbe von Raben. Wenn uns aber die Beobachtung eines Objekts nichts über die Farbe von Raben sage, gebe es keinen Grund, anzunehmen, sie beeinflusse die Plausibilität der Raben-Hypothese. Dem widerspreche grundsätzlich jede diesem Objekt attestierte Stützung, sei sie noch so gering. Selbst eine geringe Stützung zuzuschreiben, sei daher vorschnell.²⁸

Zweitens würde darüber hinaus, laut Siebel, der Lösungsversuch selbst dann dem Paradox nicht entgegen, wenn man mit der Argumentation Howson und Urbachs soweit mitginge, anzunehmen, eine geringe Stützung sei unproblematisch. Angesichts der Behauptung, die kleine Stützung, die *ein einzelner* weisser Schuh mit sich bringe, sei vernachlässigbar, dränge sich die Frage auf, wie die Sache ausschaue, wenn die Daten sehr viele weisse Schuhe umfassen würden. Dann wäre womöglich die Stützung irgendwann zu gross, um vernachlässigt zu werden. Wie Siebel vorführt, wird diese Ahnung bestätigt, wenn man sich derselben Methode bedient, mit der Howson und Urbach ihre Lösung herleiten:²⁹

Gesucht ist die Nachher-Wahrscheinlichkeit der Hypothese h , gegeben n nicht-schwarze Nicht-Raben $\neg S_1 \wedge \neg R_1, \neg S_2 \wedge \neg R_2, \dots, \neg S_n \wedge \neg R_n$, die sich wie im Lösungsversuch Howson und Urbachs mit Bayes' Theorem errechnen lässt:

$$P(h \mid \neg S_1 \wedge \neg R_1 \wedge \neg S_2 \wedge \neg R_2 \wedge \dots \wedge \neg S_n \wedge \neg R_n) = \frac{P(h) \cdot P(\neg S_1 \wedge \neg R_1 \wedge \neg S_2 \wedge \neg R_2 \wedge \dots \wedge \neg S_n \wedge \neg R_n \mid h)}{P(\neg S_1 \wedge \neg R_1 \wedge \neg S_2 \wedge \neg R_2 \wedge \dots \wedge \neg S_n \wedge \neg R_n)}$$

Zu den Hintergrundinformationen gehört, dass es ausserordentlich viele Dinge gibt, die keine Raben und nicht schwarz sind. Daher verändere sich die Wahrscheinlichkeit, einen nicht-schwarzen Nicht-Ra-

ben als nächstes Objekt zu finden, wenn die Daten bereits eine grosse Zahl solcher Beobachtungen enthalten, nur unbedeutend. Der Unterschied zwischen $P(\neg S_i \wedge \neg R_i \mid \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ und $P(\neg S_{i+1} \wedge \neg R_{i+1} \mid \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_i \wedge \neg R_i)$ dürfe entsprechend vernachlässigt werden:

$$P(\neg S_i \wedge \neg R_i \mid \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1}) = P(\neg S_i \wedge \neg R_i)$$

Auf Basis dieser Überlegungen lässt sich, wie Siebel im Detail zeigt, die Nachher-Wahrscheinlichkeit von h vereinfachen zu:

$$P(h \mid \neg S_1 \wedge \neg R_1, \neg S_2 \wedge \neg R_2, \dots, \neg S_n \wedge \neg R_n) = \frac{P(h) \cdot P(\neg S_1 \mid h) \cdot P(\neg R_1 \mid h \wedge \neg S_1) \cdot \dots \cdot P(\neg S_n \mid h) \cdot P(\neg R_n \mid h \wedge \neg S_n)}{P(\neg S_1) \cdot P(\neg R_1 \mid \neg S_1) \cdot \dots \cdot P(\neg S_n) \cdot P(\neg R_n \mid \neg S_n)}$$

Wie zuvor wird nun $\neg S_i$ als unabhängig von h angenommen, womit $P(\neg S_i \mid h) = P(\neg S_i)$, und aus h und $\neg S_i$ folgt weiterhin $\neg R_i$, weshalb gilt: $P(\neg R_i \mid h \wedge \neg S_i) = 1$. Demzufolge ergibt sich für h im Lichte von n untersuchten Dingen, die nicht schwarz und keine Raben sind, die Nachher-Wahrscheinlichkeit:

$$P(h \mid \neg S_1 \wedge \neg R_1, \neg S_2 \wedge \neg R_2, \dots, \neg S_n \wedge \neg R_n) = \frac{P(h)}{P(\neg R_1 \mid \neg S_1) \cdot \dots \cdot P(\neg R_n \mid \neg S_n)}$$

Wiederum beinhalten die Hintergrundinformationen, dass es sehr viel weniger Raben als nicht-schwarze Dinge gibt, weshalb $P(\neg S_i \wedge \neg R_i)$ beinahe 1 ist. Dann aber gilt, je grösser n wird – je mehr nicht-schwarze Nicht-Raben gefunden werden –, desto mehr nähert sich der Nenner des Bruchs 0 an und die Nachher-Wahrscheinlichkeit steigt entsprechend.³⁰

Siebel verdeutlicht diesen Punkt, indem er versuchsweise als Wahrscheinlichkeit für $P(\neg R_i \mid \neg S_i)$ 0.999 annimmt. Unter dieser Voraussetzung, rechnet er, würde bereits mit der Sammlung von 5000 Datensätzen der Sorte „ $\neg S \wedge \neg R$ “ die Wahrscheinlichkeit der Hypothese 140-fach gesteigert. Das zeige, dass man mit einer Untersuchung grösseren Ausmasses, bei der genügend Daten geringen Stützungsgrads für die Hypothese zusammengetragen würden, weiterhin wertvolle Beiträge zur Stützung der Raben-Hypothese leisten könnte, ohne dazu einen einzigen Raben zu untersuchen. Diese Konsequenz enthalte der Bayesianismus selbst dann, wenn eine geringe Stützung vernachlässigbar wäre. Plausibel zu machen, weshalb die Beobachtung von

28 Vgl. Siebel. 2004. 318.

29 Vgl. Ebd. 318-320.

30 Vgl. Siebel. 2004. 318-320.

Objekten, die mit der Farbe von Raben nichts zu tun hätten, eine Hypothese substantiell stützen sollte, die sich mit der Farbe von Raben beschäftige, gelinge indes nicht.

Folglich verschiebe der Lösungsversuch das Paradox bloss geringfügig. Die unangenehme Konklusion, die Beobachtung *eines* weissen Schuhs stütze die Raben-Hypothese, werde ersetzt durch jene, dass *viele* weisse Schuhe die Hypothese wesentlich stützen würden. Siebel folgert daraus, der bayesianische Lösungsversuch vermeide das Paradox nicht, er münde darin.³¹

6. Kritik an Siebels Einwänden

Siebel kritisiert die bayesianische Lösung des Raben-Paradoxes aus zwei Gründen. Erstens argumentiert er grundsätzlich gegen ein Stützungspotential eines Nicht-Raben für die Hypothese und zweitens zweifelt er an der Strategie, eine geringe Stützung als unproblematisch zu akzeptieren, weil die Stützung mit dem Datensatz wachse und irgendwann nicht mehr vernachlässigbar sei.

Der erste Einwand baut auf die Annahme, die Raben-Hypothese beschreibe ausschliesslich eine Eigenschaft von Raben, nicht aber von Objekten anderer Sorte. Diese Annahme verbucht zunächst viel intuitive Plausibilität für sich. Zu behaupten, die Raben-Hypothese treffe auch eine Aussage über Hauskatzen und Zierpflanzen, wirkt abenteuerlich, besonders wenn wir uns daran orientieren, welche Rolle allgemeinen Hypothesen in alltäglichen Situationen zukommt. Niemand würde etwa ernsthaft die Warnung „Alle Herdplatten sind möglicherweise heiss!“ als Hinweis darauf verstehen, dass die Butter im Gefrierfach keine Herdplatte ist. Wer diese Warnung äussert, will verhindern, dass sich jemand die Finger verbrennt, und bezieht sich ganz offensichtlich nur auf Herdplatten.

Hempel stellt in diesem Zusammenhang allerdings fest, wir würden an dieser Stelle praktische Überlegungen mit logischen verwechseln: Eine Hypothese der Form „alle A sind B“ formuliere für alle Objekte die Bedingung, entweder beide Eigenschaften A und B zu haben oder nicht zur Gruppe der Objekte mit der Eigenschaft A zu gehören. Jedes einzelne Objekt erfülle diese Bedingung oder widerspreche ihr, weshalb alle Objekte – und nicht nur jene mit der Eigenschaft A – Gegenstände der Hypothese „alle A sind B“ seien.³² Damit wird klar, dass „Alle Raben sind schwarz“ sehr wohl eine Aussage über die Eigenschaften *aller* Dinge

und nicht nur über die Eigenschaften von Raben ist. Enthalten ist darin nämlich auch die Einschränkung, rote, gelbe und grüne Objekte seien keine Raben. Darauf legt sich fest, wer die Prämisse (b) des Paradoxes und damit die Äquivalenz von h und h' akzeptiert.

An diesen Überlegungen können wir anknüpfen, wenn wir die gängige Methode heranziehen, Hypothesen anhand ihrer logischen Konsequenzen zu prüfen. Lässt sich aus einer allgemeinen Hypothese eine Annahme logisch ableiten, kann man die Hypothese prüfen, indem man die abgeleitete Annahme untersucht. Erweist sich diese als wahr, so wird das oft als Stützung für die Hypothese gewertet, erweist sie sich dagegen als falsch, spricht das gegen die Hypothese.³³

Verbinden wir dies nun mit Hempels Feststellung, nicht nur As seien relevant für Hypothesen der Form „alle A sind B“, können wir Siebels Verdikt, der Bayesianismus müsse zu schnell von Stützung sprechen, zurückweisen: Die Hypothese h „Alle Raben sind schwarz“ beschreibt, wie Hempel zeigt, nicht nur Eigenschaften von Raben, sondern auch von allerlei nicht-schwarzen Objekten. Eine Hypothese, aus der sich Aussagen über nicht-schwarze Objekte ableiten lassen, lässt sich prüfen, indem ebendiese Aussagen über nicht-schwarze Objekte auf ihre Wahrheit hin untersucht werden. Findet sich nun ein Objekt, das nicht schwarz ist, und es stellt sich zusätzlich heraus, dass es kein Rabe ist, dann stützt dieser Fund eine Aussage über nicht-schwarze Objekte, die sich aus h ableiten lässt – namentlich h' – und in diesem Sinn die Hypothese h selbst.

Es zeigt sich: Die Annahme, bloss schwarze Raben seien geeignet, die Raben-Hypothese zu stärken, ist näher betrachtet keineswegs so offensichtlich, wie sie auf den ersten Blick scheint. Mindestens eine geringe Stützung durch Nicht-Raben scheint vielmehr gut verträglich mit der Rolle, die beim Überprüfen einer Hypothese ihren logischen Konsequenzen zukommt. Siebel meint in seinem zweiten Einwand, der Preis, den man dafür zahle, eine schwache Stützung der Raben-Hypothese durch *einen* weissen Schuh doch nicht von vornherein als kontraintuitiv abzulehnen, sei, akzeptieren zu müssen, dass sich aus der Beobachtung *vieler* weisser Schuhe eine starke Stützung ergebe. Um zu unterstreichen, wie unattraktiv diese Konsequenz des bayesianischen Lösungsversuchs sei, führt er beispielhalber eine Berechnung an, die nahelegt, man brauche bloss einige, vielleicht 5000, weisse Schuhe zu kaufen, um sein Schuhregal erneut für die Untersu-

31 Vgl. Siebel. 2004. 320.

32 Vgl. Hempel. 1965. 18-19.

33 Howson und Urbach stellen die bayesianische Interpretation dieses Vorgehens vor, in: Howson/Urbach. 2004. 81-82.

chung beliebiger allgemeiner Hypothesen verwenden zu können. Eine Position, aus der Solches folgt, wäre in der Tat befremdlich.

Allerdings ist dieser zweite Einwand Siebels aus zwei Gründen kritisierbar: Erstens operiert Siebel in seinem Beispiel mit irreführenden Wahrscheinlichkeitswerten und zweitens trifft er die zweifelhafte Annahme, die Wahrscheinlichkeit $P(\neg S_i \wedge \neg R_i)$ dürfe stets als die gleiche betrachtet werden, unabhängig davon wie viele Daten des Typs „ $\neg S \wedge \neg R$ “ zuvor bereits ausgewertet wurden.

Die Wahrscheinlichkeit $P(\neg R_i | \neg S_i)$ setzt Siebel probenhalber bei 0,999 an. Das wäre allenfalls dann akkurat, wenn wir Grund hätten, als Hintergrundinformationen anzunehmen, von 1000 nicht-schwarzen Objekten seien 999 keine Raben. Selbst wenn alle Raben eine andere Farbe als schwarz hätten, dürfte der Anteil der Raben an den nicht-schwarzen Objekten aber deutlich tiefer sein und folglich $P(\neg R_i | \neg S_i)$ grösser als 0,999. Viel ist damit freilich noch nicht erreicht, einzig dass wesentlich mehr nicht-schwarze Nicht-Raben nötig wären, um dieselbe Stützung zu erreichen, die Siebel für 5000 nicht-schwarze Nicht-Raben errechnet.

In dieser Hinsicht folgenschwerer ist Siebels Annahme, die Wahrscheinlichkeit $P(\neg S_i \wedge \neg R_i | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ dürfe zu $P(\neg S_i \wedge \neg R_i)$ vereinfacht werden, bloss weil sich $P(\neg S_i \wedge \neg R_i | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ und $P(\neg S_{i+1} \wedge \neg R_{i+1} | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_i \wedge \neg R_i)$ nur geringfügig unterscheiden. Es trifft sicherlich zu, dass $P(\neg S_1 \wedge \neg R_1)$ und $P(\neg S_2 \wedge \neg R_2 | \neg S_1 \wedge \neg R_1)$ kaum voneinander abweichen und entsprechend keine grosse Verzerrung in Kauf genommen wird, wenn $P(\neg S_2 \wedge \neg R_2 | \neg S_1 \wedge \neg R_1)$ und $P(\neg S_2 \wedge \neg R_2)$ gleichgesetzt werden. Allerdings gilt dies immer weniger, je grösser i wird, denn:

$$P(\neg S_1 \wedge \neg R_1) < P(\neg S_2 \wedge \neg R_2 | \neg S_1 \wedge \neg R_1) < P(\neg S_n \wedge \neg R_n | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{n-1} \wedge \neg R_{n-1}).$$

Warum das so ist, erklärt sich daraus, dass die Wahrscheinlichkeit eines Datums, das sich aus einer Hypothese ableiten lässt, umso grösser ist, je höher die Wahrscheinlichkeit der Hypothese ist. Letztere wächst mit jedem neuen Objekt, das die Hypothese stützt. Entsprechend tendiert $P(\neg S_n \wedge \neg R_n | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{n-1} \wedge \neg R_{n-1})$ gegen 1.³⁴

Der Unterschied zwischen $P(\neg S_i \wedge \neg R_i | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ und $P(\neg S_{i+1} \wedge \neg R_{i+1} | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_i \wedge \neg R_i)$ mag gewiss vernachlässigbar sein. Daraus folgt

aber nicht, dass dasselbe auch für den Unterschied zwischen $P(\neg S_i \wedge \neg R_i | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ und $P(\neg S_i \wedge \neg R_i)$ gilt. Letztere setzt Siebel aber gleich, indem er für alle Daten des Typs „ $\neg S_i \wedge \neg R_i$ “ einfach $P(\neg S_i \wedge \neg R_i)$ als Wahrscheinlichkeit annimmt und so den Einfluss des vorab erhobenen Datensatzes auf die Wahrscheinlichkeit neuer Daten ignoriert.

Solange wir bloss die Stützung der Raben-Hypothese durch wenige nicht-schwarze Nicht-Raben diskutieren, bleibt der Unterschied zwischen $P(\neg S_i \wedge \neg R_i | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ und $P(\neg S_i \wedge \neg R_i)$ unwesentlich. Sobald wir uns aber mit Datensätzen beschäftigen, die gross genug sind, um eine stabile Stützung für die Hypothese bereitzustellen, dann darf der Unterschied nicht mehr vernachlässigt werden. Die Folge davon ist wiederum, dass für eine substantielle Stützung deutlich mehr nicht-schwarze Nicht-Raben observiert werden müssen, als von Siebel angenommen. Denn der Stützungsgrad eines jeden neuen Untersuchungsobjekts ist etwas kleiner, als der seines Vorgängers, da $P(\neg S_i \wedge \neg R_i | \neg S_1 \wedge \neg R_1 \wedge \dots \wedge \neg S_{i-1} \wedge \neg R_{i-1})$ mit der Höhe von i korrespondiert. In meinem Schuhregal müsste daher schon ein ziemlich grosser Teil der Welt verstaubt werden, damit sich dieses für die Untersuchung der Färbung von Raben eignen würde.

Die Stützung für die Raben-Hypothese, die in einem einzelnen weissen Schuh steckt, ist also nicht nur kleiner als Siebel unterstellt, sie verringert sich noch dazu mit jedem weiteren weissen Schuh. Nun gut, könnte man hier einwenden, dann braucht man eben noch etwas mehr nicht-schwarze Nicht-Raben zu sammeln – der Preis der bayesianischen Lösung des Raben-Paradoxes bleibt: Sie legt einen darauf fest, dass hinreichend viele weisse Schuhe die Plausibilität der Raben-Hypothese stark erhöhen können, ohne dass klar wird, weshalb Nicht-Raben überhaupt die Möglichkeit eingeräumt werden soll, diese Hypothese in nicht zu vernachlässigender Weise zu stützen. Die zwei Optionen, die an dieser Stelle offenstehen, sind entweder den Preis zu bezahlen, und dafür Gründe anzugeben, oder von der hier diskutierten bayesianischen Lösungsstrategie für das Raben-Paradox aufgrund ihrer abwegigen Konsequenzen Abstand zu nehmen.

Mir scheint, der Preis ist nicht so hoch, wie er zunächst wirken mag, vor allen Dingen, wenn wir uns vor Augen führen, welche Menge an Nicht-Raben wir untersuchen müssten, bis wir eine ernstzunehmende Stützung beisammen hätten, wollten wir die Raben-Hypothese stärken, ohne uns mit Raben zu beschäftigen. Bereits das erste Untersuchungsobjekt würde die Hypothese kaum stützen. Mit jedem wei-

34 Vgl. Howson/Urbach. 1989. 82.

teren Datum der Sorte „ $\neg S \wedge \neg R$ “ schwände die zuzätzlich gewonnene Stützung, bis wir irgendwann an den Punkt gelangen würden, an dem die Gruppe der noch nicht erfassten Nicht-Raben vielleicht nicht gerade erschöpft, aber doch mindestens stark verkleinert wäre. Warum das für die Option spricht, die scheinbar absurde Konsequenz des Paradoxes anzunehmen, lässt sich anhand eines Beispiels verdeutlichen:

Stellen wir uns eine Welt vor, die aus 100 Objekten besteht. Wir wissen, 20 der 100 Objekte sind schwarz und ein halbes Dutzend sind Raben. Wollen wir in dieser Welt belegen, dass alle Raben schwarz sind, können wir jeden einzelnen Raben untersuchen, um seine Farbe festzustellen. Ist jeder der sechs untersuchten Raben schwarz, liefert dies gemeinsam mit dem Hintergrundwissen maximale Stützung für die Hypothese, alle Raben seien schwarz. Es gibt allerdings einen zweiten, wenngleich aufwendigeren Weg, zum selben Ergebnis zu gelangen:

Die Hypothese h , „ $\forall x(\text{Rabe}(x) \rightarrow \text{Schwarz}(x))$ “ ist nicht nur äquivalent mit h' , sondern auch mit „ $\neg \exists x(\text{Rabe}(x) \wedge \neg \text{Schwarz}(x))$ “ (Hypothese h''), was leicht einzusehen ist. Nun ist h'' wahr genau dann, wenn die Gegenbehauptung „ $\exists x(\text{Rabe}(x) \wedge \neg \text{Schwarz}(x))$ “ falsch ist. Finden sich also Daten, welche die Gegenbehauptung falsifizieren, dann liegt die Nachher-Wahrscheinlichkeit von h'' , angesichts dieser Daten, bei 1. Das ist im vorliegenden Beispiel dann der Fall, wenn alle 80 nicht-schwarzen Objekte untersucht und dabei keine Raben gefunden wurden. Entdeckt man dagegen einen bunten Raben, hat man ein Datum, das die Gegenbehauptung enthält, womit diese 1 und h'' folgerichtig 0 als Nachher-Wahrscheinlichkeit erhielte. Wenn aber 80 nicht-schwarze Nicht-Raben geeignet sind, die Hypothese h'' maximal zu stützen, weil sie die Möglichkeit der Falsifikation der Hypothese ausräumen, dann wird es schwer, zu erklären, weshalb dasselbe nicht auch für die maximale Stützung von h gelten soll.

Ähnlich dürfte es sich auch dann verhalten, wenn es nicht um die Frage *maximaler* Stützung geht. Sammeln wir nicht-schwarze Objekte, von denen wir jeweils feststellen, dass sie keine Raben sind, dann ist jedes dieser Objekte ein Schritt in die Richtung, die Gegenbehauptung von h'' zu falsifizieren, und verringert umgekehrt die Wahrscheinlichkeit, ein für h'' , und damit für h , falsifizierendes Objekt zu finden, weil die Menge potentiell falsifizierender Objekte kleiner wird. Entsprechend senkt jeder nicht-schwarze Nicht-Rabe die Wahrscheinlichkeit der Gegenbehauptung und hebt damit jene von h'' , beziehungsweise h .

Zu wissen, dass keine bunten Raben existieren, sagt uns natürlich noch nichts darüber, ob es schwarze Ra-

ben gibt. Wer aber behauptet, alle Raben seien schwarz, setzt damit nicht die Existenz von Raben voraus. Stattdessen legt einen h bloss auf die Position fest, dass falls es einen Raben gibt, dieser sicher schwarz ist. Und das wäre auch dann der Fall, wenn gar keine Raben existieren würden. Daher braucht man streng genommen keinen Raben, um h zu stützen. Es reicht ein rabenfreier Datensatz, solange der gross genug ist, um die Wahrscheinlichkeit zu senken, dass es nicht-schwarze Raben gibt.

Vor diesem Hintergrund ist die Annahme, die Beobachtung vieler weisser Schuhe könne die Raben-Hypothese beträchtlich stützen, nicht mehr gleichermassen abwegig.

7. Schlussfolgerungen

Siebel kritisiert an Howson und Urbachs Lösungsversuch für das Raben-Paradox, es werde nicht klar, weshalb ein weisser Schuh die Hypothese überhaupt stützen solle, weil die Raben-Hypothese sich ausschliesslich mit Eigenschaften von Raben beschäftige. Nun beschreibt aber die Raben-Hypothese nicht nur Eigenschaften von Raben, sondern auch von allen nicht-schwarzen Objekten. Sie enthält damit logische Konsequenzen für Nicht-Raben, anhand derer sie geprüft werden kann, weshalb es sinnvoll ist, anzunehmen, dass auch Nicht-Raben die Raben-Hypothese in geringem Ausmass stützen können.

In seinem zweiten Einwand argumentiert Siebel, eine möglicherweise vernachlässigbare Stützung durch einen einzelnen weissen Schuh laufe auf eine nicht mehr vernachlässigbare Stützung durch viele weisse Schuhe hinaus. Tatsächlich nimmt die Stützung der Raben-Hypothese mit dem Umfang des Datensatzes zu. Allerdings wären deutlich mehr nicht-schwarze Nicht-Raben für eine substantielle Stützung notwendig als Siebel in seinem Beispiel annimmt – so viele, dass die Wahrscheinlichkeit, ein die Hypothese falsifizierendes Objekt zu finden wesentlich gesenkt würde. Die bayesianische Differenzierung der Konklusion des Raben-Paradoxes ist daher akzeptierbar.

Zu klären bleibt indes, ob die hier am Beispiel einer Welt bestehend aus 100 Objekten unternommenen Überlegungen auch auf eine grössere Welt angewandt werden könnten. Für eine Welt mit endlich vielen Objekten dürfte das der Fall sein, unklar ist, ob das ebenso für eine Welt mit unendlich vielen Objekten gilt. Ansonsten liefert uns der Bayesianismus gute Gründe, die Konklusion des Raben-Paradoxes anzunehmen, ohne aufgeben zu müssen, dass Raben für die Raben-Hypothese um ein Vielfaches relevanter sind als weisse Schuhe.

Literatur

- Siebel, Mark. 2004. „Der Rabe und der Bayesianist“. In: Zeitschrift für allgemeine Wissenschaftstheorie 35, 313-329.
- Hempel, Carl Gustav. 1965. „Studies in the Logic of Confirmation“. In: Hempel, Carl Gustav (Hg.). Aspects of Scientific Explanation. And Other Essays in the Philosophy of Science. New York: The Free Press. 3-46.
- Howson, Colin/Urbach, Peter. 1989. Scientific Reasoning. The Bayesian Approach. Chicago/Lasalle Ill.: Open Court.

Simon Richmond (25) studiert im 8. Semester des Bachelors Philosophie und Geschichte. Er interessiert sich für politische Philosophie, Wissenschaftsphilosophie und Metaethik.

Thought Experiments in Ethics

This contribution first appeared in: Stuart, Michael T.; Yiftach Fehige; James Robert Brown (eds). 2018 *The Routledge Companion to Thought Experiments*. Abingdon/New York: Routledge. 195–210. © 2018, Georg Brun. Thanks to Routledge for the permission to reprint this paper in meta(φ).

1. Introduction

In normative and applied ethics, thought experiments are frequently used in debates on questions such as: Are we morally obliged to give large amounts to poverty relief? Can the number of lives saved be ethically decisive? Is torture ever morally permissible? At the heart of every thought experiment is a scenario with a question. Here are sketches of some of the best-known ethical thought experiments, which have also become prominent in political discourse:

Trolley: in the so-called *Driver's Two Options* (there are many other versions), a runaway trolley is heading towards five workers on the track who have no escape; is it morally permissible for the driver to steer the trolley onto another track where it will kill one worker? (Foot 2002; Thomson 2008)

Pond: a child has fallen into a pond and is about to drown; ought you hurry to rescue him even if doing so ruins your shoes and thwarts your plans for the day? (Singer 1972)

Violinist: while asleep, your body has been hooked up to the body of a violinist who has an illness that makes it necessary for her to be connected to your metabolism for nine months; is it morally permissible for you to ask to be severed from her even if this will kill her? (Thomson 1972)

Ticking Bomb: a terrorist has hidden a bomb that will kill thousands of people if set off; are we morally permitted to torture her given that we know that only in this way we will learn how to defuse the bomb? (Shue 1978)

Original Position: behind a "veil of ignorance," in a situation in which no one knows his place in society, his

status or his natural strengths and weaknesses, which fundamental principles of justice would free and rational persons accept as being in their interest? (Rawls 1999)

In contrast to the first four scenarios, *Original Position* is not a stock example. One explanation for this is that *Original Position* has a specific function in Rawls's theory of justice, whereas the other scenarios are used in more familiar ways as parts of arguments for or against an ethically relevant claim about, for example, abortion, torture or our obligation to help others. However, the first four scenarios are sometimes used differently as well, for example, to illustrate the consequences of some moral principle. This calls for a typology of ethical thought experiments according to their functions, which will be introduced in Section 3. Before this issue can be tackled, we need a more detailed characterization of the structure of thought experiments which explains how scenarios, questions and arguments figure in thought experiments (Section 2). Section 4 then investigates the two most widely discussed functions of thought experiments, namely supporting and refuting ethical claims. Since such epistemic functions cannot be analysed in a metaethically neutral way, I introduce some elements of the method of reflective equilibrium as a more specific background in moral epistemology, which enjoys relatively broad acceptance. In Section 5, I finally discuss a range of challenges and touch upon specifically moral objections to ethical thought experiments.

2. Reconstructing ethical thought experiments

The philosophical literature usually focuses on analysing thought experiments of a certain kind, most often, thought experiments that aim at refuting a general claim. For such thought experiments, specific schemes of reconstruction have been developed, typically tied to some philosophical approach (e.g. Sorensen 1992, ch. 6; Häggqvist 1996, ch. 5, 2009; and see the entries

in Part III of Stuart, Fehige and Brown 2018.). In this paper, I rely on a simpler reconstruction, which enables us to address the variety of ethical thought experiments and to distinguish two common uses of the term “thought experiment”:

- (1) A scenario and a question are introduced.
- (2) The experimenter goes through (imagines, thinks about, etc.) the scenario and arrives at some result.
- (3) A conclusion is drawn with respect to some target (e.g., an ethically relevant claim or distinction).

A “core” thought experiment just comprises (1) and (2). An “extended” thought experiment includes some additional reasoning (3), which can be reconstructed as an argument; the result of (2) is used as a premise from which, typically together with a range of often implicit assumptions, a conclusion is drawn with respect to some target.

In *Trolley*, for example, the scenario describes a situation with a runaway trolley, two tracks, a driver, several workers and a switch. The question is whether the driver is morally permitted to throw the switch. The experimenter imagines or thinks about the situation and arrives at, say, the conviction that the driver is permitted to change the switch. This is the result of the core thought experiment. The extended thought experiment then uses this result as a premise in an argument for, say, the target thesis that (body) numbers count morally. An additional assumption involved in this argument may be that the difference in numbers of workers on the tracks is the reason why switching is permissible.

In more complex cases, one extended thought experiment draws on the results of several core thought experiments. Smart (1972, 27–8), for instance, uses two scenarios, one with one and another with two million equally happy people, to illustrate the difference between average (both scenarios yield the same result) and total utilitarianism (the situation with more people fares better).

Conversely, the same core thought experiment can be used in different arguments. This provides a more natural and flexible analysis of reasoning by thought experiment than an analysis which uses “thought experiment” only for extended thought experiments, associating every thought experiment with a specific

structure (1)–(3) (see, e.g., Häggqvist 1996, 2009).

The elements (1)–(3) require several explanations. The scenario describes a situation which is declared or assumed to be possible in some sense, for example, (meta)physically, logically or conceptually.¹ If the thought experiment is to be successful, the scenario must be a consistent, meaningful description (see Rescher 2005, ch. 9) and the person running the thought experiment must be able to make appropriate use of the scenario by, for example, conceiving of the described situation or applying a given principle to it. Often, the scenario is qualified as fictional (e.g. Elgin 2014), hypothetical or counterfactual (e.g. Gendler 2000, 17). This is not meant to imply that the described situation must not be real or realizable, but only that it need not be so, and that, if the described situation happens to be realized, this does not affect the thought experiment. Passersby come across drowning children, and *Trolley*-style accidents do in fact happen, but as thought experiments, *Pond* and *Trolley* are independent of those tragedies. This is so because even if situations as described in the scenario are real, the scenario, not these situations, sets the stage for the thought experiment. As all descriptions of non-abstract states of affairs, scenarios are invariably less rich than the actual and possible situations that fit the description (does the violinist like eggplant casserole?). This selectivity serves to establish at least a presumption about which aspects of a situation are relevant for answering the given question, without implying the unrealistic assumption that the scenario explicitly mentions all and only the relevant information. Common ground and implicatures play an important role as well.

The element (2) is deliberately framed in unspecific terms to leave room for a broad range of activities and results. More specific descriptions can be used to characterize specific kinds of thought experiments, for example, thought experiments in which an intuition is prompted by imagining a situation or thought experiments which use modal reasoning. We should, however, resist the temptation to postulate any such specific description as a general characterization. It is, for example, sometimes said that ethical thought experiments are used to elicit intuitions (see Stich and Tobia 2018). What this boils down to depends on what one means by “intuition.” A necessary but not sufficient requirement for a belief (or other propositional attitude) to be an intuition is that it is not held *just* because it

1 There is also reasoning per impossibile with counterfactuals with logically or mathematically false antecedents (see Rescher 2005, ch. 8.8), but as far as I can see, it plays no role in ethics.

has been inferred consciously (see Brun 2014). But as a general characterization of the result in (2), this minimal condition is still too narrow.² There are many ethical thought experiments in which the result of (2) is a judgement gained by means of an explicit argument, typically involving additional assumptions not explicitly mentioned in the scenario.

Similar problems result if thought experiments are distinguished from counter-factual or hypothetical reasoning in general by the requirement of having an experimental element in the sense that something is imagined not merely by contemplating propositional content, but visually, tactually or in another perception-related way (Brown and Fehige 2014). If this idea is to be effective at all, the element of experience must be more than the visual, auditory or other associations routinely evoked by any description of a non-abstract situation. Understood in this way, however, an implausibly narrow notion of thought experiment results, which excludes many philosophical (e.g. Putnam's H₂O vs. XYZ on Twin Earth) and ethical thought experiments such as *Ticking Bomb* and *Trolley* if they are described as simply as at the beginning of this section. The argumentation reconstructed in (3) may take any form of argument ranging from modal and deontic deductive reasoning to arguments from analogies and inference to the best explanation. The target mentioned in (3) allows for many different kinds of conclusions: they can be about a concept, a proposition, a theory, or about a relation between two or more such objects (e.g. a difference between two theories). As Gendler (2000, 25) argues, there is a tendency (but no strict rule) that thought experiments in different disciplines aim at different results in (2). Scientific thought experiments primarily ask what would happen in the situation described in the scenario; thought experiments in metaphysics, epistemology and philosophy of language more often ask how we should describe what would happen; and thought experiments in ethics and aesthetics very commonly ask how what would happen should be evaluated. Accordingly, scientific thought experiments typically target empirical claims; in metaphysics, epistemology and philosophy of language, the target is typically a proposal for a conceptual analysis or, more generally, a characterization of a concept (e.g. Gettier-cases against knowledge as justified true belief); in ethics, the target is usually a claim about a normative concept or some normative principle or

theory, for example, that numbers count morally, that abortion is permissible, or that the right to life includes a right to be sustained in living.

The explanations given so far obviously do not amount to a definition of "ethical thought experiment," but together with the examples above they should suffice to give a fairly clear idea of what ethical thought experiments are. On this basis, we can now explore the functions of ethical thought experiments in more detail.

3. Functions of ethical thought experiments

Typologies of extended thought experiments according to their goal are of special importance since the evaluation of a thought experiment must frequently be relative to what it is supposed to achieve. In what follows, I suggest a typology that can accommodate the broad range of functions of thought experiments in ethics (for alternatives, see, e.g., Brown 1991; Walsh 2011; Davis 2012). It is important to keep in mind that core thought experiments are normally put to more than one use, and that their use can change over time, for instance, because they are employed in the context of different background assumptions and with different goals (Elgin 2014). The proposed classification therefore primarily deals with extended thought experiments, and only derivatively with core thought experiments.

To simplify, we can speak of, for example, an "epistemic thought experiment" where we strictly speaking would have to say "extended thought experiment with an epistemic function" or "core thought experiment used in an extended thought experiment with an epistemic function."

3.1. Epistemic thought experiments

The goal of epistemic thought experiments is to provide a reason which speaks in favour of or against a claim with respect to the target.³ Accordingly, we can draw a basic distinction between constructive and destructive epistemic functions. Constructive thought experiments come in two varieties. Some are meant to show that something is possible. Shue (1978), for example, uses *Ticking Bomb* to show that there are possible exceptional circumstances in which torture is permissible, and according to one analysis, *Violinist* aims at showing "that abortion could be morally permissible even when the fetus has a right to life" (Brown and Fehige 2014). Other constructive thought experiments are intended

² Cappelen (2012) even argues that there are no philosophical thought experiments that rely on eliciting intuitions as opposed to reasoning; he specifically discusses *Violinist* and *Trolley*.

³ This is a rather narrow sense of "epistemic." In a wider sense, heuristic functions may also be classified as epistemic.

to provide only some (non-conclusive) support for a claim. *Pond* is regularly interpreted in this way, either as intended to support a general principle that obliges us to help if we can do so without significant moral costs, or as an argument from analogy which supports the claim that we should give significant amounts of money to help poor people. That thought experiments can provide support for an ethical principle at all, is a view that needs metaethical backing. On Singer's own view (2009, 15–17), for example, intuitions are epistemically dubious, including those prompted by *Pond*; and Dancy (e.g., 1985, 1993, 64–6) has argued that analogies from core thought experiments to real cases are inherently problematic. Section 4 discusses epistemic functions in more detail against a specific background in moral epistemology.

Destructive thought experiments provide a counterexample to a target claim. Often, they show that something can be the case and use this in an argument against an incompatible claim. *Violinist*, for example, is standardly reconstructed as providing such a refuting argument against the claim that the right to life of one person under all circumstances outweighs the right of another person to decide what happens with his or her body. Obviously, this type of refuting use of a scenario is most closely related to the possibility-showing constructive use. *Ticking Bomb* shows that torture is permissible in some circumstances just in case it refutes the claim that torture is impermissible under all conditions whatsoever. Another way of using the result of going through the scenario is to turn it against some background assumption. *Violinist*, for example, can also be reconstructed as challenging the tacit assumption that the right to life entails a right to be sustained in living (Brown and Fehige 2014).

It is natural to think that epistemic functions of thought experiments are normally refuting rather than constructive because general claims can be decisively undermined but not conclusively supported by individual cases. And in fact, there is a culture of “counterexample philosophy,” which widely employs refuting thought experiments which target a conceptual claim that is part of some account of, for example, moral permissibility, autonomy or consent. The thought experiment then comes to the result that an act described in the scenario does (not) fall under that concept in contrast to what follows from the account in question. Whether that actually amounts to a refutation of this account depends on methodological assumptions, in particular on the conditions an adequate “analysis” or “explication” of a concept is supposed to meet. If

the goal is an analysis which preserves all clear-cut cases, counterexamples to such cases invariably refute the proposed analysis. If the goal is an explication that may include some conceptual revision, counterexamples are not automatically decisive.

In the philosophical discussion, epistemic or, more specifically, destructive epistemic functions are generally taken to be the basic or the most important functions of thought experiments. There is, however, not one function all ethical thought experiments essentially have. And even if one considers non-epistemic functions as entirely secondary, it remains important to recognize them because ethical thought experiments usually have an epistemic function non-exclusively, and therefore other, non-epistemic functions can influence the epistemic function or the two functions can get conflated.

3.2. Illustrative and rhetorical thought experiments

Illustrative thought experiments render a concept, a proposition, a theory or some other target clearer and more understandable, but insofar as we focus exclusively on its illustrative function, we do not see a thought experiment as providing reason for or against its target. Singer's original use of *Pond* is clearly intended as an illustration of a principle he independently defends. He writes (1972, 231): “If it is in our power to prevent something very bad from happening, without thereby sacrificing anything morally significant, we ought, morally, to do it. An application of this principle would be as follows.” The case of the drowning child makes then vivid what the quoted principle may ask of us in a concrete situation. Another example is the difference between total and average utilitarianism. One may know in abstract terms that it will become manifest only in a comparison of two groups with a different number of people, but going through specific scenarios makes the difference more readily graspable. Furthermore, textbook-examples which are thought experiments are usually illustrative (Cohnitz 2006, 77). Closely related are rhetorical thought experiments, which are intended to convince people to accept or reject some ethical idea. *Pond* is certainly apt for rhetorical use and actually functions this way in some of Singer's later writings (e.g. 2009). Rhetorical functions very often piggyback on an epistemic function, but certainly not always. Ticking-bomb scenarios, for example, are notorious for having been used in media and political discourse as a means for campaigning in favour of torture and other illegal or contested practices.

Of course, illustrative thought experiments can be misleading, and rhetorical thought experiments can be used in propagandistic or demagogic ways, but this is not necessarily so. We expect successful epistemic thought experiments to have the rhetorical power of convincing us, and there is nothing inherently problematic about that, in contrast to an illustrative or rhetorical function of a thought experiment which is interpreted or sold as epistemic (for a paradigmatic example, see Dennett's 1980 critique of Searle's Chinese room argument as an "intuition pump").

3.3. Heuristic thought experiments

In the simplest case, heuristic thought experiments are just devices for generating ideas. A core thought experiment is run, but its result is then investigated independently. More ambitious heuristic thought experiments systematically explore morally relevant factors and differences.

One possibility is to try to isolate factors which are potentially morally relevant by developing variations of a scenario and seeing whether the resulting moral intuitions or judgements co-vary. Thomson's work on trolley cases is a well-known example. She extensively varies the scenario in order to find whether it makes a difference if, for example, the driver or a bystander can change the switch, or if the choice is described as one between killing five and killing one, or between killing one and letting five die. Actually, this procedure can also run into problems. Many more factors can be tested for moral relevance, including the age of the people on the track, whether they are employees of the train-system, or whether they are members of an ethnic minority. In fact, there is empirical research showing that skin colour has an influence on the judgement of people who are strongly committed to its moral irrelevance (see Uhlmann et al. 2009). This raises the question of whether such an effect undermines the value of the elicited judgements, or whether we should rather conclude that a trolley scenario which alludes to skin colour is badly designed, or that philosophers need to be more careful in running thought experiments (see Ludwig 2018; Stich and Tobia 2018).

Another heuristic strategy systematically explores the consequences of various theories for a range of scenarios with the goal of, for example, finding out in which situations two rival theories actually lead to different results. Again, comparisons of various versions of utilitarianism provide examples.

Applications of the two strategies just described can actually be analysed as heuristic or as another type of

thought experiment, depending on what we take to be their target. Consider the example of the two versions of utilitarianism. If the target is the claim that there is a difference between total and average utilitarianism, then Smart's thought experiment is most naturally interpreted as illustrating this difference or as having the epistemic function of showing that there is indeed such a difference. If the target is, say, total utilitarianism, then Smart's thought experiment may be interpreted as having the heuristic function of tracking down a situation in which total utilitarianism differs from average utilitarianism, but in itself this neither supports nor undermines total utilitarianism. The last interpretation shows that the heuristic potential of thought experiments can be exploited even by those who are deeply sceptical about epistemic functions of ethical thought experiments (e.g. Dancy 1993, 65).

In the literature there is a tendency to see illustrative and heuristic functions of thought experiments as of merely limited interest, especially in comparison with epistemic functions. However, another view emerges, for example, in Hills's recent discussion of moral understanding. According to her account, to understand that q is why p is morally right or wrong, requires that one has the ability to "draw the conclusion that p' (or that probably p') from the information that q' (where p' and q' are similar to but not identical to p and q)" (Hills 2009, 102). In other words, one needs the ability to run heuristic and illustrative thought experiments. Furthermore, Hills argues that moral understanding is not only instrumental in reliably doing the right thing and justifying yourself to others, but also essential for having a good character and for morally worthy action. On this account, the ability to run thought experiments is of central interest not only to ethical theorists, but to all moral agents.

3.4. Thought experiment with a theory-internal function

Borsboom, Mellenbergh and Van Heerden (2002) argue that there are thought experiments which perform some specific function within a theory, different from the functions discussed so far. Their prime example is the scenario of a "counterfactual long run" in frequentist statistics. A prominent example of an ethical thought experiment with a theory-internal function is Rawls's *Original Position*. The scenario models fair conditions for social cooperation as well as restrictions on arguments for principles of justice (Rawls 1999, 16–9; 2001, 17, 83, 85; cf. Gendler 2007 for a different analysis). Also Kant's categorical imperative involves a

thought experiment in which we have to test whether we can will the generalization of the maxim guiding some way of acting (see Parfit 2011, 285, 328–9). A closer analysis and evaluation of such thought experiments inevitably requires an in-depth discussion of the respective ethical theories and cannot be undertaken here.

4. Ethical thought experiments and reflective equilibrium

Epistemic functions of thought experiments are philosophically the most prominent and hence deserve a closer analysis. But as pointed out in Section 3.1, such functions depend on metaethical assumptions. I therefore introduce in this section a background in moral epistemology, which is relatively widely accepted, namely reflective equilibrium (Section 4.1). On this basis, the central functions of thought experiments can be characterized more precisely (Section 4.2), and some consequences can be drawn regarding the results of core thought experiments and with respect to what we can expect to accomplish by thought experiments (Section 4.3).

4.1. Reflective equilibrium

The method of (so-called “wide”) reflective equilibrium can be characterized most succinctly by two key ideas. Firstly, judgements and principles are justified if judgements, principles and background theories are in equilibrium. Secondly, this state is reached through a process that starts from judgements and background theories, proposes systematic principles and then mutually adjusts judgements and principles (and possibly also background theories).⁴

The contrast between principles and judgements is not to be understood as a matter of their content, but as the contrast between propositions that are part of some given moral theory or system of principles and propositions to which somebody is actually committed. Commitment is to be understood as an epistemically relevant status which implies at least a minimal degree of credibility. Commitments can be expressed explicitly or merely revealed in action and they can have any degree of firmness from feeble to unwavering. In the tradition of Rawls and Daniels, commitments are required to be considered judgements; that is, judgements not made under circumstances prone to error such as inattention, excessive self-interest, etc.

Both, commitments and elements of the theory, may be changed in the process of mutual adjustment. Every stage of the process of developing a reflective equilibrium is characterized by a position consisting of the current commitments and the current theory.

4.2. Functions of thought experiments

Within the framework of reflective equilibrium, thought experiments can play several roles. Constructive epistemic thought experiments can generate commitments at any stage in the process of developing a reflective equilibrium. In simple cases, the commitment is just the result of the core thought experiment. The experimenter considers *Trolley*, for example, and resolves that throwing the switch is morally permissible. In more complex cases, a conclusion of an extended thought experiment is supported by the result of a core thought experiment and additional reasoning. In this way, *Trolley* may support the claim that it is morally permissible to kill a person if this is the only way to save several others. While such supportive thought experiments can be independent of the current theory, this need not be the case. Elements of the current theory can enter the reasoning, and candidates for commitments can also be generated just by applying the current theory to the scenario. In *Trolley*, a philosopher with deontologist leanings could argue that the prohibition on killing implies that it is morally impermissible to change the switch, and then conclude that this result should be accepted because it does not conflict with any of her other commitments.

Of course, conflicts will often arise, and then destructive thought experiments become important. In such cases, the result of a core thought experiment is a premise in an argument which shows that there is an incoherence within or between the current commitments and the theory, as in the *Trolley*-examples of the preceding paragraph. On this basis, an extended thought experiment argues for giving up a commitment or some element of the current theory.

If the process of mutual adjustments is to be carried out in a systematic way, heuristic thought experiments play a crucial role as well. They are needed for exploring the results of applying the current theory to systematic variations of scenarios. Again, the many variants of *Trolley* provide an example.

4.3. Consequences

We can now turn to some debated issues in the literature on ethical thought experiments. Let us begin with the results of core thought experiments.

⁴ These two ideas can be found in all standard accounts of reflective equilibrium, specifically in Rawls (1999) and Daniels (2011). In what follows, I rely on an understanding which also draws on ideas of Elgin (1996); see Brun (2014) for a more extensive outline.

Firstly, the result of a thought experiment is always a commitment that is explicitly expressed in public or in one's mind, not one merely revealed in action – for the thought experiment must not rely on the scenario being real. It has been argued that this is problematic and we should rather rely on how we act in real situations (Davis 2012). As a general strategy to determine what is morally right, this is implausible because, regrettably, we cannot count on our actions being more reliably right than our explicit commitments. The method of reflective equilibrium, however, calls for initially recognizing and also for possibly revising all kinds of commitments, including those resulting from thought experiments and those merely revealed in action.

A second issue concerns the role of intuitions. We have seen in Section 2 that an implausibly narrow conception of thought experiments results if one insists that all core thought experiments elicit intuitions. Many further points about the role of intuitions depend on a theory of intuitions, but some basic results are available independently (for discussion, see Copp 2012; Brun 2014; Ludwig 2018; Stich and Tobia 2018). On the relatively uncontroversial assumption that intuitions cannot be based on explicit inference, the method of reflective equilibrium leaves room for intuitions as results of core thought experiments, as long as they are not gained by explicit reasoning; this excludes appealing to the theory we are currently developing when going through the scenario. And the conclusion of an extended thought experiment can be an intuition only if it is not reached by explicit argumentation. Furthermore, reflective equilibrium does not exempt intuitions from critical evaluation. They can be revised just as any other commitment. Sometimes, we will not accept an intuition elicited by a scenario as its proper result because we have good reason to think that the intuition is inadequate. Someone may hold that changing the switch is impermissible even if he or she is not able to get rid of his or her recalcitrant intuition that doing so is permissible in this particular case. In other cases, epistemic background theories may give us reason to discredit intuitions as based upon, say, cultural stereotypes.

Closely related considerations also make clear that requiring commitments to be considered judgements poses no principled obstacles to the use of thought experiments. It just means that the conclusion of the extended thought experiment depends on the (possibly implicit) assumption that the result of the core thought experiment meets certain standards.

Turning to extended thought experiments, we can first note that although destructive thought experiments can play an important role, this does not mean that the result of a core thought experiment always “wins” against incompatible commitments or elements of the current theory. It is rather one of the characteristics of reflective equilibrium that revisions may well take the other direction and lead to overriding the result of a thought experiment (as in the example of the recalcitrant intuition above). A refuting thought experiment can show the need for a revision, but it alone cannot settle what is to be revised.

A second point concerns analogies. It is sometimes assumed that supportive thought experiments paradigmatically take the form of analogies (see Section 5.2). Take *Violinist* as an example: is its main point not that the result of the core thought experiment can be transferred by analogy to cases of abortion? In the context of the method of reflective equilibrium, however, the standard way of transferring a moral judgement from one case to another is not by direct analogy but by explicitly introducing principles that cover both cases. The most important consequence is more general. If we rely on the method of reflective equilibrium, then the fact that a proposition is the result of a core thought experiment is not sufficient for this proposition to be epistemically justified to the degree required for knowledge (and the same is true a fortiori for conclusions of analogies based on such a result). If reflective equilibrium is necessary for epistemic justification, the primary object of justification is not a single proposition but an entire position consisting of the commitments and the theory in equilibrium. Hence, thought experiments may play important epistemic roles according to the method of reflective equilibrium, but they cannot by themselves justify a commitment or a theory. Defenders of reflective equilibrium should therefore object to the view that a destructive thought experiment alone may suffice to refute a claim and to the idea that being supported by an intuition elicited by a thought experiment or by an analogy based on a core thought experiment may be sufficient for justification.

5. Challenges to ethical thought experiments

Although ethical thought experiments are frequently challenged, wholesale criticism of ethical thought experiments in all their functions is implausible. There are, however, challenges which target a specific function of thought experiments. Well-known examples are Dancy's attacks on supportive thought experiments,

which are briefly discussed below. But most challenges to ethical thought experiments raise a problem that may affect specific thought experiments, and I focus on such challenges in what follows. I first give an overview in Section 5.1 and then discuss three issues which have been prominent in debates about ethical thought experiments (Section 5.2). Finally, I address the question of how thought experiments in ethics can raise specifically moral problems (Section 5.3).

5.1. A general overview of challenges

The structure (1)–(3) from Section 2 can be used as a basis for an overview of challenges which address specific thought experiments in any of the functions distinguished in Sections 3 and 4 (see also Gendler 2000, 22–4). Challenges to core thought experiments frequently point out a problem with the scenario or the question. Sometimes we have difficulties going through the described situation because the scenario is incoherent. In other cases, the situation is difficult to imagine because it is underdescribed; that is, described too schematically and relevant details remain undetermined. Or the scenario in effect undermines the conditions for a meaningful application of the concepts used in the question (this kind of objection is best known from the debate about personal identity, targeting, e.g., Parfit’s 1987 teletransportation scenarios; see Rescher 2005, ch. 9). Two points which will be discussed further below are the question of how realistic scenarios should be, and the charge that a scenario or a question is misleading.

Other challenges to core thought experiments point out problems with “going through” the scenario and coming up with a result. Maybe no intuition or judgement results because the situation is underdescribed. In such cases, the problem is not that we cannot imagine a situation as described, but that the scenario does not provide enough information to answer the question in a non-arbitrary way (Wilkes 1988, e.g., claims that this problem arises for personal-identity thought experiments because “person” is not a natural-kind term). In other cases, an intuition or a judgement results, but in a problematic way. Whether thought experiments are unreliable if they rely on intuitions is currently a much debated issue (see Stich and Tobia 2018; specifically for moral intuitions: Burkard 2012; Copp 2012). Many writers hold that intuitions are not reliable truth-trackers, but frequently the product of prejudice, tradition or stereotypes. One strand of experimental philosophy is devoted to investigating empirically whether such accusations can be substantiated by showing that peo-

ple’s reactions to a scenario co-vary with, for example, cultural factors (see Ludwig 2018; Stich and Tobia 2018). If the result of a core thought experiment is arrived at not intuitively but discursively, the reasoning may be criticized as fallacious if it involves an invalid or weak argument, an unwarranted assumption (invited, e.g., by an underdescribed scenario), or a conclusion that is irrelevant in the context at hand.

Challenges to an extended thought experiment attack the reasoning which is supposed to establish the conclusion about the target. One type of criticism is that the reasoning is fallacious in one of the ways just mentioned. Since thought experiments using arguments from analogy are quite common in ethics, such objections often home in on weak analogies (see the discussion below). Other challenges point out that it is not clear enough what the conclusion of the thought experiment is supposed to be or that it misses what is at issue in a given debate. For example, does a refuting thought experiment address utilitarianism in general or only some specific version?

5.2. Common challenges to ethical thought experiments

Let us now turn to three types of challenges which are frequently raised in debates about ethical thought experiments. A first question is whether thought experiments, especially epistemic and heuristic ones, should be realistic. The issue is not whether the scenario gives a vivid and detailed description, nor whether the described situation is improbable or bizarre, but rather whether the described situation is modally remote, involving, for example, biologically or technologically impossible elements such as the fission of humans.

On the one hand, there are reasons for appealing to unrealistic situations. They are relevant to universally valid statements and definitions (Sorensen 1992, 279; Walsh 2011). They can produce results that we have a lot of confidence in (Parfit 1987, 200). And if we want to use heuristic thought experiments to isolate morally relevant factors, we need to compare similar situations which differ just in the factor in question (e.g. would cruel treatment of animals be permissible if animals were insentient?); appeal to unrealistic scenarios is then needed if rival theories agree on realistic situations or if realistic situations differ in additional respects (Sorensen 1992, 91; Kamm 1993, 7).

On the other hand, unrealistic epistemic thought experiments have been challenged on various grounds (see Elster 2011). Firstly, some approaches to ethics seek to develop practical, action guiding, moral prin-

ciples which are meant to lead to the right results in our world, but not necessarily in unrealistic situations. Unrealistic thought experiments are then irrelevant to the justification of these principles (e.g. Hare 1981, 47–9). However, insofar as an approach to ethics must also deal with “deeper” ethical principles that capture moral truths or definitions of moral terms, unrealistic cases remain relevant.

Secondly, there is the worry that unrealistic scenarios lead to core thought experiments with unreliable results. This challenge needs qualification. Reliability is independent of whether the scenario is realistic overall. Unrealistic problems can have easy answers (If binary fission is iterated three times, how many persons result?), realistic cases can pose hard questions and real situations can include too many factors which distract from important features, trigger biases or interfere with our judgement in other ways. More plausibly, unreliability looms if the unrealistic aspects undermine the resources needed to answer the question at stake (Rescher 2005, ch. 9). Even if the scenario is coherent, we may not be able to determine the morally relevant factors and their exact import, or to work out the necessary background assumptions for the unrealistic situation, especially if the unrealistic aspects concern central elements of our web of belief (Sorensen 1992, 43–4). With respect to unrealistic thought experiments that ask for an intuition, many argue that our capacity for intuitive judgement is unreliable in situations for which it has not been developed by evolution or social practice (e.g. Kitcher 2012). This line of argument sometimes leads to more general scepticism about epistemic thought experiments involving intuitions (see, e.g. Singer 2005; Machery 2011).

As a second type of challenge, many epistemic thought experiments are accused of being misleading (Wood 2011), independently of whether they are realistic. At least three charges can be distinguished. Firstly, the really important moral questions are replaced by dubious problems that do not occur in reality. Instead of, say, “Who is responsible for the trolley disaster?”, the experimenter asks “Should you throw the switch?” Secondly, thought experiments use forced-choice questions which artificially limit the range of admissible courses of action (“Change the switch, yes or no?”) and thereby rule out giving the right answers (e.g. “Try to save all six.”). Thirdly, problematic assumptions are introduced, especially by way of implicature or presupposition. *Trolley*, for example, suggests that it asks an important question, that at least one of the available answers is morally legitimate, that moral decisions are

a matter of ranking states of affairs resulting from actions, and that further information about the situation is irrelevant (e.g. “Do the workers have permission to be on the track?”).

However, those who advance such objections must deal with the rejoinder that they do not take seriously the thought experiments they reject, either by refusing to deal with them, by filling in details that favour their own views or by insisting on changing the subject. Moreover, unrealistic epistemic thought experiments cannot simply be dismissed as irrelevant if they target a universal moral statement or definition which covers such situations. But it is true that they are often not worked out in sufficient detail and need reconstruction. In *Trolley*, for example, the intuition or judgement that we should change the switch may contribute to refuting the view that numbers are never decisive, though it certainly does not show that consequentialism is right, that numbers always count or that rights and entitlements are unimportant. Suggesting that such conclusions follow is bad rhetoric – just as much as insinuating without further evidence that users of trolley cases draw or implicate such conclusions. Finally, a lot of the scepticism may stem from the worry that scenarios such as *Trolley* or *Ticking Bomb* are used to argue directly for a conclusion about some other situation. Typically such a move involves an analogy, and this leads to a third type of challenge to epistemic thought experiments.

Violinist and *Pond* are routinely interpreted as analogies in which the result of the core thought experiment – “It is permissible to unplug yourself from the violinist.” or “You ought to rescue the child.” – supports a conclusion about abortion or poverty relief respectively. In Section 4.3, I explained that in the context of method of reflective equilibrium such supportive functions are usually better reconstructed as operating more indirectly. But if extended thought experiments are interpreted as analogies, two sorts of objections need to be considered.

Firstly, Dancy (1985, 1993, 64–6) has mounted an attack on supportive thought experiments, which he assumes to employ an analogy from the result of a core thought experiment to a real case. According to Dancy, such analogies are not reliable because, in contrast to the real case, the information on the situation described in the scenario is inevitably limited. We must therefore either qualify our conclusion about the real case with the condition that it does not relevantly differ from the situation described in the scenario, or admit that any new information about the real case can spoil the

analogy. One answer operates on the metaethical level and argues against Dancy's specific form of particularism and holism about reasons, which underwrites his claim that we cannot conclusively evaluate situations which are neither real nor variations of real situations (see Häggqvist 1996, ch. 2.7; Cohnitz 2006, ch. 4.2.2 for discussion). Another answer challenges Dancy's understanding of the analogy involved. He asks for an argument that, given the result of the core thought experiment, conclusively shows a claim about the real case. Arguments from analogy, however, are a form of plausible reasoning resting on a premise which states that the real case is relevantly similar to the situation described in the scenario. Analogies are therefore not deductively valid, and strong analogies can be overthrown by further arguments. Nonetheless, for supporting, in contrast to proving, a conclusion a strong analogy suffices (Rescher 2005, 59).

Objections of the second sort target weak analogies in specific extended thought experiments. Examples abound in the debate about thought experiments which use a ticking-bomb scenario to argue that torture may be morally permissible in real life. Critiques of such arguments very often point out that even if torture should be permissible in *Ticking Bomb*, real-life situations differ from this scenario in ways that effectively undermine the analogy. For example, the scenario (explicitly or implicitly) assumes that we know that only torture will get us the right information in time, but in reality this is uncertain, or that torture is strictly limited to some very specific and rare exceptions, whereas in reality it is always an institutionalized practice with a strong tendency to spread (Luban 2005).

5.3. Moral objections to ethical thought experiments

If an ethical thought experiment in fact suffers from one of the shortcomings discussed so far, this can obviously be morally relevant. Bad ethical thought experiments can lead to bad ethics and consequently to morally wrong action. Accordingly, Parfit's ethical project, for example, has been criticized for employing misleading thought experiments (Wood 2011). And attempts to legitimize post-9/11 torture practice with the help of ticking-bomb scenarios have been criticized as irresponsible because they involve an analogy which is invalidated by discrepancies between the scenario's assumptions and easily knowable empirical facts (Luban 2005).

There is, however, also the worry that some ethical thought experiments raise moral problems in other

ways. Specifically, it has been argued that discussing certain scenarios and questions is a symptom of moral corruption or leads to moral corruption of the author or the audience of the thought experiment (see, e.g., Anscombe 2005; Williams 1972, 92). Such worries can hardly be defended as a general critique of (unrealistic) ethical thought experiments. After all, one may also argue that the debate about ticking-bomb scenarios, and especially the analysis of the disanalogies involved, has considerably contributed to a better understanding of how torture might not be justified. What rather seems really problematic is the rhetorical use of thought experiments such as *Ticking Bomb* in political discourse and propaganda (see Luban 2005 for an analysis and examples). Glossing over crucial assumptions and selling gerrymandered questions as well-founded challenges can have morally bad consequences and it is incompatible with an adequate self-understanding of a philosopher.

Acknowledgments

For helpful comments, I thank Christoph Baumberger, Claus Beisbart and Anna Goppel. My research has been supported by the Swiss National Science Foundation (project 150251).

References

- Anscombe, Gertrude Elizabeth Margaret. 2005 [1957]. "Does Oxford Moral Philosophy Corrupt Youth?" In *Human Life, Action and Ethics*. Exeter: Imprint Academic. 161–167.
- Borsboom, Denny; Gideon J. Mellenbergh; Jaap Van Heerden. 2002. "Functional Thought Experiments". *Synthese* 130, 379–387.
- Brown, James Robert. 1991. *The Laboratory of the Mind. Thought Experiments in the Natural Sciences*. London: Routledge.
- Brown, James Robert; Yiftach Fehige. 2014. "Thought Experiments". In *Stanford Encyclopedia of Philosophy*. (<http://plato.stanford.edu/archives/fall2014/entries/thought-experiment/>)
- Brun, Georg. 2014. "Reflective Equilibrium without Intuitions?" *Ethical Theory and Moral Practice* 17, 237–252.
- Burkard, Anne. 2012. *Intuitionen in der Ethik*. Münster: Mentis.
- Cappelen, Herman. 2012. *Philosophy without Intuitions*. Oxford: Oxford University Press.
- Cohnitz, Daniel. 2006. *Gedankenexperimente in der Philosophie*. Paderborn: Mentis.
- Copp, David. 2012. "Experiments, Intuitions and Methodology in Moral and Political Philosophy". In Shafer-Landau, Russ (ed.). *Oxford Studies in Metaethics*. Vol. 7. Oxford: Oxford University Press. 1–36.
- Dancy, Jonathan. 1985. "The Role of Imaginary Cases in Ethics". *Pacific Philosophical Quarterly* 66, 141–153.
- Dancy, Jonathan. 1993. *Moral Reasons*. Oxford: Blackwell.
- Daniels, Norman. 2011. "Reflective Equilibrium". In *Stanford Encyclopedia of Philosophy*. (<http://plato.stanford.edu/archives/spr2011/entries/reflective-equilibrium/>)
- Davis, Michael. 2012. "Imaginary Cases in Ethics. A Critique". *International Journal of Applied Philosophy* 26, 1–17.

- Dennett, Daniel C. 1980. "The Milk of Human Intentionality". *The Behavioral and Brain Sciences* 3, 428–30.
- Elgin, Catherine Z. 1996. *Considered Judgment*. Princeton: Princeton University Press.
- Elgin, Catherine Z. 2014. "Fiction as Thought Experiment". *Perspectives on Science* 22, 221–241.
- Elster, Jakob. 2011. "How Outlandish Can Imaginary Cases Be?" *Journal of Applied Philosophy* 28, 241–258.
- Foot, Philippa. 2002 [1967]. "The Problem of Abortion and the Doctrine of the Double Effect". In *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Clarendon Press. 19–32.
- Gendler, Tamar Szabó. 2000. *Thought Experiment. On the Powers and Limits of Imaginary Cases*. New York: Garland.
- Gendler, Tamar Szabó. 2007. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium". *Midwest Studies in Philosophy* 31, 68–89.
- Häggqvist, Sören. 1996. *Thought Experiments in Philosophy*. Stockholm: Almqvist and Wiksell.
- Häggqvist, Sören. 2009. "A Model for Thought Experiments". *Canadian Journal of Philosophy* 39, 55–76.
- Hare, Richard Mervyn. 1981. *Moral Thinking. It's Levels, Method, and Point*. Oxford: Clarendon Press.
- Hills, Alison. 2009. "Moral Testimony and Moral Epistemology". *Ethics* 120, 94–127.
- Kamm, Frances Myrna. 1993. *Morality, Mortality. Vol. 1. Death and Whom to Save from It*. New York: Oxford University Press.
- Kitcher, Philip. 2012. "The Lure of the Peak". *New Republic* 243/1, 30–35.
- Luban, David. 2005. "Liberalism, Torture, and the Ticking Bomb". *Virginia Law Review* 91, 1425–1461. (<http://scholarship.law.georgetown.edu/facpub/148>)
- Ludwig, Kirk. 2018. "Thought Experiments and Experimental Philosophy". In Stuart, Michael T.; Yiftach Fehige; James Robert Brown (eds). *The Routledge Companion to Thought Experiments*. Abingdon/New York: Routledge. 385–405.
- Machery, Edouard. 2011. "Thought Experiments and Philosophical Knowledge". *Metaphilosophy* 42, 191–214.
- Parfit, Derek. 1987. *Reasons and Persons*. Oxford: Clarendon Press.
- Parfit, Derek. 2011. *On What Matters. Vol. 1*. Oxford: Oxford University Press.
- Rawls, John. 1999. *A Theory of Justice. Revised ed.* Cambridge, MA: Belknap Press.
- Rawls, John. 2001. *Justice as Fairness. A Restatement*. Cambridge, MA, Harvard University Press.
- Rescher, Nicholas. 2005. *What If? Thought Experimentation in Philosophy*. New Brunswick: Transaction.
- Shue, Henry. 1978. "Torture". *Philosophy and Public Affairs* 7, 124–143.
- Singer, Peter. 1972. "Famine, Affluence, and Morality". *Philosophy and Public Affairs* 1, 229–243.
- Singer, Peter. 2005. "Ethics and Intuitions". *The Journal of Ethics* 9, 331–352.
- Singer, Peter. 2009. *The Life You Can Save. Acting Now to End World Poverty*. London: Picador.
- Smart, J.J.C. 1972. "An Outline of a System of Utilitarian Ethics". In Smart, J.J.C.; Bernard Williams. *Utilitarianism For and Against*. Cambridge: Cambridge University Press. 1–74.
- Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.
- Stich, Stephen; Kevin Tobia. 2018. "Intuition and its Critics". In Stuart, Michael T.; Yiftach Fehige; James Robert Brown (eds). *The Routledge Companion to Thought Experiments*. Abingdon/New York: Routledge. 369–384.
- Stuart, Michael T.; Yiftach Fehige; James Robert Brown (eds). 2018. *The Routledge Companion to Thought Experiments*. Abingdon/New York: Routledge.
- Thomson, Judith Jarvis. 1972. "A Defense of Abortion". *Philosophy and Public Affairs* 1, 47–66.
- Thomson, Judith Jarvis. 2008. "Turning the Trolley". *Philosophy and Public Affairs* 36, 359–374.
- Uhlmann, Eric Luis; David A. Pizarro; David Tannenbaum; Peter H. Ditto. 2009. "The Motivated Use of Moral Principles". *Judgment and Decision Making* 4, 476–91.
- Walsh, Adrian. 2011. "A Moderate Defence of the Use of Thought Experiments in Applied Ethics". *Ethical Theory and Moral Practice* 14, 467–481.
- Wilkes, Kathleen V. 1988. *Real People. Personal Identity without Thought Experiments*. Oxford: Clarendon Press.
- Williams, Bernard. 1972. "A Critique of Utilitarianism". In Smart, J.J.C.; Bernard Williams. *Utilitarianism For and Against*. Cambridge: Cambridge University Press. 75–150.
- Wood, Allen. 2011. "Humanity as an End in Itself". In Parfit, Derek. *On What Matters. Vol. 2*. Oxford: Oxford University Press. 58–82.

Georg Brun ist Dozent am Institut für Philosophie der Universität Bern. Er interessiert sich besonders für philosophische Methoden und hat über Fragen der Philosophie der Logik, der epistemischen Rechtfertigung, des Verstehens, der Metaethik, der Symboltheorie und der Umweltphilosophie gearbeitet.

Moral Philosophers in Wonderland

How Outlandish May Ethical Thought Experiments Be?

1. Introduction

There can be no doubt that the method of thought experiment occupies a prominent place in the philosopher's toolbox. Particularly in the last century or so, this device has become ubiquitous in the discipline as a whole, yet it is arguably employed most extensively and liberally in ethics (see Brown and Fehige 2017; Archard 2017, 25). Indeed, devising and assessing thought experiments has become an integral part of the work of moral philosophers.

There is however something deeply curious about the kinds of thought experiments used in ethics: Although most do not extend beyond innocuous everyday scenarios, there abound thought experiments which put many a science fiction story to shame, featuring bizarre and alien worlds, situations, or beings: *outlandish* thought experiments. I am thinking of thought experiments such as Rawls' *original position*, Nozick's *experience machine*, or Thomson's *people-seeds*, the last of which I will discuss in detail in section 3. It is (prima facie) remarkable that practitioners of ethics, which is arguably first and foremost a practical i.e. action-guiding discipline, concern themselves so extensively with scenarios so completely removed from our world (Walsh 2011, 468; Archard 2017, 19). This tension has, perhaps surprisingly, rarely been scrutinized by moral philosophers. However, given the method's prevalence in contemporary ethics it would be deeply troublesome if not to say disastrous, were there to emerge serious problems with outlandish (ethical) thought experiments. Do we have reason to think that there are? Two promising avenues of criticism can be found in the literature, both holding the employment of outlandish thought experiments in ethics to be deeply flawed: Firstly, Jonathan Dancy (1985) objects to outlandish thought experiments (and indeed thought experiments in general) on the grounds that they are typically under-described, hence morally *indeterminate*. Secondly, both Russell Hardin (1997) and R. M. Hare (1981) argue that outlandish cases are *irrelevant* to the evaluation of our moral theories, since the scope of these theories is limited to our world. This essay will examine these two objections and argue that although neither are ultima-

tely convincing, they nonetheless provide insights into the problems inherent in outlandish (ethical) thought experiments which are invaluable for constructing as well as assessing such thought experiments.

The body of this essay divides into three sections: Section 2 covers some brief preliminaries regarding terminology and scope. Sections 3 and 4 discuss the objection of indeterminacy and the objection of irrelevance respectively.

2. Preliminaries

Firstly, what makes a scenario outlandish? I take the term to refer to a notion partially grounded in intuition: Some thought experiments just seem far-out to us in that some elements of them feel completely removed from reality (Elster 2011, 244). However, one might plausibly specify the term to cover scenarios which are sufficiently *modally distant* from our world, in the sense that a world in which the present essay does not exist is less modally distant from ours than one in which human beings do not communicate through language (Walsh 2011, 468–73). Moreover, this essay will be concerned exclusively with *genuinely* outlandish thought experiments, by which I mean outlandish thought experiments which may *not* be reformulated in a non-outlandish version while (virtually) fully preserving their argumentative function (Elster 2011, 244–45; Sorensen 1998, 280).

Secondly, this essay will focus on a specific class of thought experiments used in ethics that is marked by its function: Typically, ethical thought experiments are employed *against* opposing moral theories or principles by presenting a counter example to the theory or principle, i.e. a case in which it conflicts with our moral intuitions. I will refer to this class of ethical thought experiments as *destructive* thought experiments. Moreover, I will assume our moral intuitions to be epistemically relevant. Indeed, I will take it for granted that the intuitions involved in the thought experiments examined here are strictly non-negotiable, i.e. the option to abandon them (as there exists in the framework of the reflective equilibrium for example) is excluded.

3. The Objection of Indeterminacy

The most prominent criticism of the thought experiment method in ethics argues that thought experiments are typically under-described and hence *indeterminate*, i.e. they lack adequate background or context, in absence of which the case's moral status is indeterminate. Note that Dancy's objection is *not* limited to outlandish cases in ethics but directed at ethical thought experiments in general (Dancy 1985, 142). However, the fact that lack of adequate background or context seems to be a flaw that affects most if not all outlandish cases, combined with the fact that under-description presents a particularly grave problem in such cases (more on this shortly), makes this one of the most powerful objections against outlandish thought experiments in ethics (see Rescher 2005, 150–51).

In this section, I will first sketch the argument as formulated by Dancy, then discuss its merit. Throughout, I will illustrate Dancy's objection by reference to an example thought experiment. Consider the following slightly less famous sister thought experiment of Judith Jarvis Thomson's *Violinist*:

People-Seeds

[S]uppose it were like this: people-seeds drift about in the air like pollen, and if you open your windows, one may drift in and take root in your carpets or upholstery. You don't want children, so you fix up your windows with fine mesh screens, the very best you can buy. As can happen, however, and on very, very rare occasions does happen, one of the screens is defective; and a seed drifts in and takes root. Does the person-plant who now develops have a right to the use of your house? (Thomson 1971, 59)¹

3.1. The Challenge

Dancy's point of departure is what he regards as an obvious yet crucial point of difference between actual cases and imaginary ones, namely the amount of context they provide. Thus, what seems to be the greatest strength of imaginary cases, i.e. their abstraction, is according to Dancy at the same time their greatest flaw: In contrast to a fully determined actual case, nothing beyond the explicit description of a thought ex-

periment (save for the logical implications of this description) is determined (Dancy 1985, 144). While we may never know every detail of an actual case, there is nothing *to* know about an imaginary case beyond the description of the thought experiment. Consequently, there are no answers to questions about the background of the case (What color is the carpet in *People-Seeds*?), barring new stipulations on the part of the thought experimenter (Rescher 2005, 114–15). Imaginary cases are unavoidably indeterminate.

Why worry about this difference? Dancy's argument is straightforward: If a thought experiment is under-described, i.e. its background is largely indeterminate, so is the moral status of the case as it is described to us (henceforth "foreground") indeterminate and consequently the status of the case as a whole (Dancy 1985, 144–45). The moral status of the foreground will differ depending on how the background is filled out. This problem persists even if we suppose the isolated foreground to possess a definitive moral status by itself, as it may still be impossible to coherently fill in the background in a way that preserves the supposed moral status of the foreground: The burden then lies on the thought experimenter to show that this is indeed possible (Davis 2012, 6).

Turning to Thomson's thought experiment, Dancy might object that we are not given sufficient information about the background of the case needed for us to be able to judge the case. A world in which people-seeds are floating about is surely a world in which basic anthropological and social facts will differ in morally significant respects from our world. (Do people-seeds need to be watered and fed? Are there people-seed orphanages?) Yet it is utterly unclear *how* these facts will differ.

One obvious answer to this puzzle available to the thought experimenter is to stipulate that the case's background is free of anything which would significantly change the moral relevance of the properties presented in the foreground of the case (Dancy 1985, 144). According to this answer, we are to fill in the details ourselves in whatever way we please, if it does not significantly alter the moral nature of the foreground. But this is, in Dancy's view, for the thought experimenter to beg the question: It is the aim of a destructive thought experiment to show there to exist an exception to the principle it attacks, yet the thought experimenter now seems to be *presupposing* a major part of what was to be shown, namely that a possible background exists which preserves the supposed moral status of the foreground (Dancy 1985, 147–49). The

¹ I take *People-Seeds* to be a destructive thought experiment in that it presents an attack on the principle that the right to life of human creatures is inviolable.

problem persists: In under-described thought experiments in general, and in outlandish thought experiments such as *People-Seeds* in particular, it is unclear in what context we are to place the case, i.e. how this world would differ from ours given the stipulations of the thought experiment. Yet without this information we cannot judge these cases. Dancy concludes that the present preoccupation of moral philosophers with thought experiments is largely unwarranted.

3.2. The Indeterminacy of Fiction

I hold Dancy's objection to be seriously flawed. To see where his argument goes wrong, one must examine how exactly a real-world counterexample differs from a fictional counterexample. Dancy is correct about the fact that actual and imaginary cases differ in their *metaphysical* determinacy, i.e. there is as much to find out about actual cases as we please, while the same cannot be said for imaginary ones. Metaphysical indeterminacy is however irrelevant for the matter at hand: The fact that a case is metaphysically indeterminate does in no way hinder our ability to make an intuitive judgment on the basis of the information we *do* have.

However, *both* actual and imaginary cases are *epistemically* indeterminate to us (Sorensen 1998, 246–48; see also Häggqvist 1996, 41–43): We obviously never know all there is to know about an actual case, indeed we usually only ever possess an extremely narrow knowledge of an actual case's background. Yet I doubt that anybody would dispute that we are nonetheless able to reliably pass an intuitive judgment on many actual cases. We need not know the entirety of a case's background to reach an adequate intuitive judgment of it, since we may (or indeed must) make reasonable *assumptions* about the background of the case.

Therefore, because we must make assumptions about the background of both actual cases as well as imaginary ones, our intuitive judgment of both necessarily comes with a qualification, namely that our assumptions about the case are indeed reasonable. Yet there remains a crucial difference between actual and imaginary cases, particularly outlandish ones: We are familiar with the background of the former, since, of course, the world of the actual case is *our* world. In the case of scenarios such as *People-Seeds* however, we simply do not know in what context to place the situation; we are unable to make reasonable assumptions about its background, and thus unable to pass an intuitive judgment on it.

3.3. Stipulation and Its Limits

How may imaginary cases be salvaged in the face of this problem? An obvious answer available to the defender of thought experiments is that under-description may be avoided by more extensive stipulation on the part of the thought experimenter. Specifically, by stipulating that, aside from what is explicitly stated in the thought experiment, *everything else remains the same* as it is in our world. Indeed, this is arguably tacitly assumed in all thought experiments (and works of fiction in general): If the author remains silent about some aspect of his work, we must ourselves fill in the information in a way that preserves the integrity of the information the author *does* give us (Sorensen 1998, 247). The thought experimenter may thereby anchor the scenario in our world and thus in our realm of intuitional competence, or so the idea.

However, stipulation comes at a price: The stipulator commits herself to everything entailed by her stipulations (Häggqvist 1996, 146–47). Some rules must be followed: Needless to say, the overall thought experiment (foreground and background) ought to remain coherent. It must not be contradictory, since no one could plausibly demand that our moral theories must cover cases that are logically impossible as well (Davis 2012, 6; see also Rescher 2005, 147).

A second more demanding criterion is that the foreground must not entail radical changes in the basic anthropological or social facts (henceforth "human nature") of our world. This is problematic *not* because moral principles ought not cover worlds of this kind as well (they do), but because knowledge of these background facts is crucial to reaching an adequate intuitive moral judgment of a case. *People-Seeds* fails to meet this criterion since the existence of floating people seeds is plainly incompatible with human nature as we know it (Häggqvist 1996, 157–58). Human nature in the world of *People-Seeds* must differ significantly from what it is in our world, yet it is utterly mysterious in what way; we are unable to make reasonable assumptions about the background of such cases. (It seems conceivable, for example, that the existence of people-seeds would lead to a two-tiered society in which people born in this way were cognitively impaired and socially isolated, thus living on average a very brief and miserable life compared to ordinary people. This fact would no doubt be of moral relevance to *People-Seeds*.) Therefore, knowing *how* human nature differs is crucial for the moral evaluation of the case hence we may not reach any intuitive moral judgments regarding such cases (Hardin 1997, 23).

Thomson might object that, despite its shortcomings, people seem to have intuitions about *People-Seeds* when presented with the thought experiment. She might agree that there are problems with filling in the background of the case but insist that we “just answer” anyway (see Davis 2012, 6). However, in outlandish cases like *People-Seeds*, where it is unclear how one may fill in the background consistently, intuitive judgments expressed by people are likely the result of them unwittingly filling out the gaps in the description of the thought experiment with knowledge about *our* world (Elster 2011, 251). In this way an intuition is elicited, yet just as the case has been hopelessly muddled (remember that the foreground is incompatible with our world) so will the resulting intuitive judgment be confused and hence worthless for the purposes of a destructive thought experiment (Davis 2012, 6).²

3.4. Dancy’s Argument Reconsidered

What remains of Dancy’s argument in view of these considerations? Plainly, the conclusions warranted by the argument from indeterminacy are not as far-reaching as Dancy takes them to be; the argument does *not* spell doom for outlandish thought experiments in ethics in general. Nevertheless, the argument does provide a reason to avoid outlandishness in ethical thought experiments, since the more outlandish a case, the more difficult it is to make reasonable assumptions about its background and the higher the risk that we will encounter a problem when attempting to fill in the background in a way that is consistent with the foreground (Häggqvist 1996, Ch. 8).

Thought experimenters must thus exercise caution when devising ethical thought experiments: Less outlandish versions of thought experiments should always be preferred and far-reaching consequences concerning basic anthropological and social facts, knowledge of which is crucial to the moral evaluation of a case, must be avoided. *People-Seeds* arguably fails to meet this standard. Unless one may somehow accommodate the scenario in a world which does not suffer from fatal under-description, it must be rejected as a destructive thought experiment. It is however unlikely that this scenario may be properly accommodated as it features an element (floating people-seeds) which undoubtedly entails radical changes of what I called human nature at its core.

4. The Objection of Irrelevance

Outlandish thought experiments are open to a second line of criticism which holds our moral intuitions regarding outlandish cases to be *irrelevant*, since our moral principles need not or cannot ever be strictly universal. I will present a brief sketch of this objection as articulated by Russell Hardin, then discuss the objection in detail. Parallel to the previous section, I will refer to a thought experiment as an example to illustrate the objection. Consider the following version of Nozick’s utility monster (1974, 41):

Utility Monster

Suppose there exists a being which experiences both pleasure and suffering at a magnitude immensely greater than human beings do. Indeed, the pleasure experienced by this creature by inflicting harm on others is so great that it will always outweigh the suffering thereby caused to its victims. Consequently, to maximize utility, each one of us would have to offer themselves up to the monster (presumably to be tortured and killed). Ought we do so?

4.1. The Challenge

To do justice to the evident frustration that accompanies this objection in both Hare and Hardin it is, I think, necessary to let them voice it themselves. Hardin does not mince words:

No moral theory need to prevail in the face of [outlandish] examples. [...] That such test cases [...] have any respectability in the conclusions of moral theorists is an oddity, an aberration of intellectual fecundity. No moral theory can withstand the intuitionist implications of such tests absolutely, because a test case can be made strategically dependent on one’s moral theory or principle. [...] The relevant tests of our moral [...] principles is whether they would be good principles to follow in actual life, not in fanciful circumstances. Too many of the examples of moral theorists are so contrived and unrealistic with their deus ex machina devices that it is hard to draw any—pardon the word—morals from them. (Hardin 1997, 24–26)

Hardin’s point is this: We should in general reject outlandish thought experiments as ponderings on

² Or as Davis puts it more bluntly, “Garbage in, garbage out.”

“desert-island morality” with few if any applications to practical morality. He believes that the tension between ethics as a *practical* discipline and outlandish thought experiments should be taken seriously: Since ethics is essentially practical, our principles need only have a limited scope, i.e. they must apply to *our* world and *our* circumstances, nothing more (see Elster 2011, 246). A moral theory need not accommodate a case that is frankly impossible to ever occur in our world: Utility monsters do not and (most likely) never will exist (Hardin 1997, 27). Therefore, to object to a moral principle on the basis of our intuition towards an outlandish thought experiment, Hardin contends, is absurd in the same way as it is to object to a proposed bridge design with the words “But what if Satan says he will smash the bridge and all the people on it if we build it that way?” (1997, 24). By doing so, the objector points out that the principle in question is not applicable to a contrived outlandish world constructed specifically against it; *so what?* No moral theory is exempt from such counter-examples (Hardin 1997, 24–26).³ Thus, according to Hardin, to show that a moral principle fails to account for our intuitions in an outlandish thought experiment is to show little if anything at all (1997, 24–25). To make outlandish counter examples a central part of one’s critique of a moral principle or theory is to gravely misunderstand the proper essence of ethics, namely its practical nature.

4.2. Two Levels of Moral Principles

To see why Hardin’s rejection of outlandish counter examples is unsatisfactory, it is helpful to consider the following distinction, as formulated by G. A. Cohen (2008), between two classes of moral principles which I will refer to as *conditional* principles and *ultimate* principles respectively. Conditional principles are *fact-sensitive*: they are conditional upon some contingent anthropological or social fact. These are the principles we utilize to guide and justify our behaviour in everyday life and which only apply to our particular circumstances as human agents or as members of a certain society. The principle that it is wrong to steal, for example, does not apply in a society in which the concept of personal property does not exist. Ultimate principles, on the other hand, are *fact-insensitive*: they are strictly universal, i.e. they apply to all possible worlds (Cohen 2008, Ch. 6). For example, most utilitarians presumably hold the principle that utility is the

highest good to be an ultimate principle.

Plainly, Hardin’s objection does indeed apply to *conditional* principles: One may always ad-hoc think up a situation in which the most undisputed conditional moral principle fails, by stipulating away the anthropological or social facts on which that principle is contingent on. In contrast, ultimate principles are by definition susceptible to outlandish counter examples, since they claim to hold true in *every* imaginable world (see Rescher 2005, 150–51). However, Hardin holds ethics to be concerned *exclusively* with finding appropriate conditional principles. Why should moral philosophers worry about outlandish destructive thought experiments which target their (supposed) conditional principles, e.g. why should Hardin, as a utilitarian, be concerned about *Utility Monster*?

4.3. Why Ultimate Principles Matter

Again, utility monsters arguably cannot ever exist in our world. Yet the point of *Utility Monster* is obviously not the possibility of us encountering a utility monster, but that utilitarianism seems to get something fundamental *wrong*. The thought experiment shows that there is *something* about the principle in question (i.e. the utilitarian principle without qualification) that is deeply repugnant to us. (Here roughly: Unqualified utilitarianism treats agents as mere vessels of value.) Hardin might object that if this supposed flaw does not affect how well the principle in question guides our actions in our day to day lives, it is irrelevant to our evaluation of this (conditional) principle (Elster 2011, 248). There are however several reasons to consider this answer lacking:

Firstly and trivially, since definitions claim to have strictly universal validity, they are susceptible to outlandish counter examples. By denying that outlandish destructive thought experiments have any place in ethics, Hardin denies (unconditional) definitions to be a part of ethics, which is deeply problematic (Sorensen 1998, 280).

Secondly, by embracing Hardin’s position one bars oneself from using ultimate principles to justify one’s conditional principles. This is however an indispensable means of justification in ethics: Were one to examine the justificatory chain of any conditional principle, i.e. *why* it holds true in our world, one would inevitably arrive at some ultimate principle (Cohen 2008, 233–34; see Walsh 2011, 473–74). It is up to Hardin to show that the conditional principles he endorses may be justified *without* appealing to an ultimate principle. Lastly and most importantly, we rely on outlandish

³ That is no theory which somehow accommodates *consequences*, which Hardin takes to be a trivial criterion for an adequate moral theory.

thought experiments to arrive at appropriate conditional principles. They help us identify which properties are morally relevant: It might be a contingent peculiarity of our world that properties F and G always coincide, hence a principle which holds F to be morally relevant and one which holds G to be morally relevant are practically equivalent, i.e. they both prescribe identical actions in our world. An outlandish counter example might however show that G is in fact the morally relevant property, while F is morally irrelevant (Sorensen 1998, 91; see also Elster 2011, 244–45). This conjunction, and thus the practical equivalence of these two principles, might however cease to hold true in the future. The principle which withstands the (destructive) outlandish thought experiment is then “future-proof”, if you will, in a way that the second principle is not.

Moreover, outlandish thought experiments may help us distinguish between practically equivalent principles which involve different *justifications*. Consider Kant’s infamous argument against animal cruelty: Kant argued that one ought to avoid animal cruelty only to prevent one’s moral corruption (Ak. 27: 459). Applying this principle, one might plausibly arrive at the same conclusions regarding how we ought to treat animals as through a more plausible principle. Yet it is obviously mistaken: The person acting on this principle would plainly be acting on the *wrong* reasons, and the thought experiment of taking an anti-corruption pill before inflicting animal cruelty shows this to be the case (Sorensen 1998, 91).⁴ Not all practically equivalent conditional principles are created equal and outlandish thought experiments may help us identify the appropriate ones.

In conclusion, Hardin is correct about the fact that outlandish counter examples do not threaten conditional principles in the same way non-outlandish counter examples do. However, since we cannot dispense with ultimate principles, one may not evade responsibility for the result of one’s theory in outlandish cases by arguing that such cases are irrelevant to practical morality. If a conditional principle is shown to be inadequate in an outlandish thought experiment, we gain ethically valuable knowledge about this principle, which the proponent of this principle must address.

5. Conclusion

This essay examined two objections about the usage of outlandish thought experiments in ethics and argued that both are unsuccessful. The argument of indeterminacy fails because we can (in many cases) fill out the background by making reasonable assumptions about it, whereas the argument of irrelevance fails because even the philosopher concerned with conditional principles must not ignore fundamental principles and thus must deal with outlandish thought experiments formulated against her proposed principles.

Nevertheless, the two objections make points relevant to thought experimentation in ethics: Outlandish thought experiments ought to be devised mindful of what the scenario entails about the case’s background, particularly regarding human nature, as pointed out by Dancy’s objection. If a scenario entails major changes in human nature without elaborating on the nature of these changes, as does *People-Seeds*, we are unable to adequately fill in the gaps ourselves and the thought experiment must consequently be rejected. Outlandish destructive thought experiments are also less argumentatively powerful than non-outlandish ones, since they cannot be legitimately employed against conditional principles, as pointed out by Hardin’s objection. These are then two reasons to avoid outlandishness in ethical thought experiments.

Note further that even if the two objections were successful, an important function of (outlandish) thought experiments would still be left untouched: Thought experiments may enable us to *rethink* a problem, particularly in cases where discussion has reached an impasse (Walsh 2011, 473–78).

⁴ Note that the anti-corruption pill thought experiment is arguably not an outlandish scenario. However, it seems highly likely that there are practically equivalent principles with differing justifications for which one must employ outlandish thought experiments to differentiate them.

References

- Archard, David. 2017. "The Methodology of Applied Philosophy." In *A Companion to Applied Philosophy*. Vol. 25, edited by Kasper Lippert-Rasmussen, Kimberley Brownlee, and David Coady, 18–33. Blackwell Companions to Philosophy. Chichester, West Sussex: Wiley Blackwell.
- Brown, James Robert, and Yiftach Fehige. 2017. "Thought Experiments." Accessed December 23, 2018. <https://plato.stanford.edu/archives/sum2017/entries/thought-experiment/>.
- Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.
- Dancy, Jonathan. 1985. "The Role of Imaginary Cases in Ethics." *Pacific Philosophical Quarterly* 66 (1-2): 141–53.
- Elster, Jakob. 2011. "How Outlandish Can Imaginary Cases Be?" *Journal of Applied Philosophy* 28 (3): 241–58.
- Hardin, Russell. 1997. *Morality Within the Limits of Reason*. Chicago: University of Chicago Press.
- Hare, R. M. 1981. *Moral Thinking: Its Levels, Method, and Point*. Oxford: Clarendon Press.
- Kant, Immanuel. 1900–. *Kants gesammelte Schriften*, edited by Königlich-Preußische Akademie der Wissenschaften. Berlin: Walter de Gruyter.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Rescher, Nicholas. 2005. *What If? Thought Experimentation in Philosophy*. New Brunswick, NJ: Transaction Publishers.
- Sorensen, Roy A. 1998. *Thought Experiments*. New York: Oxford University Press.
- Walsh, Adrian. 2011. "A Moderate Defence of the Use of Thought Experiments in Applied Ethics." *Ethical Theory and Moral Practice* 14 (4): 467–81.

Maurice Grüter (22) studiert im 7. Semester des Bachelors Philosophie mit Nebenfach Geschichte. Seine Interessen gelten der Ethik, insbesondere der Kantianischen, Metaethik sowie Religionsphilosophie der Aufklärung.

Zwischen Eigentlichkeit und Uneigentlichkeit

Nachweis einer graduellen Unterscheidung anhand von Heideggers Befindlichkeitstheorie

1. Einleitung

Einführung

Heideggers Frühwerk (1926-1933) beschäftigt sich primär mit der Frage nach dem *Sinn des Seins* (vgl. Kapitel 2). Hierbei geht es Heidegger zunächst nicht um die inhaltliche Bestimmung, was ein erfülltes und sinnvolles Leben ausmacht, sondern darum, die Frage überhaupt erst richtig zu stellen (Heidegger, 2006: 4). Hierzu bedarf es laut Heidegger einer speziellen Philosophie, welche er *Existenzialontologie* nennt. In ihr geht es darum, die menschliche Existenz in ihren verschiedenen Seinsverfassungen zu untersuchen. Ausgehend von dieser sollte der Schluss zum Sinn von Sein möglich werden. Zwei dieser erwähnten Seinsverfassungen nennt er *Eigentlichkeit* und *Uneigentlichkeit* (Heidegger, 2006: 42f). Bei der Lektüre seiner Philosophie mutet es der Leserschaft zuerst an, diese beiden Seinsmodi als exklusiv zu verstehen. Exklusiv bedeutet, dass die eine Existenzform die andere notwendigerweise ausschliesst; entweder eigentlich oder uneigentlich, jedoch nicht beides. Innerhalb von Heideggers Werk finden sich jedoch Passagen, welche darauf hinweisen, dass eine exklusive Trennung nicht notwendig ist und dass es zwischen Eigentlichkeit und Uneigentlichkeit eine graduelle, stufenweise Trennung gibt (vgl. bspw. Heidegger, 2006: 164 und 384). Ebenfalls kann angenommen werden, dass eine graduelle Unterteilung der Komplexität der menschlichen Existenz eher entsprechen würde. Wie diese zwei Existenzformen sich nun zueinander verhalten, ist in der wissenschaftlichen Debatte bisher jedoch wenig beachtet worden (vgl. aber Poci, 1996 und Rentsch, 2015). Eine Untersuchung hierzu erscheint deshalb lohnenswert.

Arbeitshypothese

Ziel der vorliegenden Arbeit ist es, die Hypothese stark zu machen, dass die zwei Existenzformen Eigentlichkeit und Uneigentlichkeit graduell und daher stufenweise ineinanderfliessen. Insofern schliessen sie sich nicht gegenseitig aus und ein dynamischer Über-

gang sollte möglich sein. Für dieses Unterfangen wird Heideggers Befindlichkeitstheorie herbeigezogen. Ein solcher Versuch wurde schon über die *Sprache* gestartet (Hüsch, 2014: 189). Die Befindlichkeitstheorie ist jedoch vom Umfang her weitaus differenzierter dargestellt und bietet deshalb mehr Informationen an. Ebenfalls kann man der Befindlichkeitstheorie eine grössere Gewichtung in Heideggers Frühwerk beimessen. Heidegger verweist in seiner Befindlichkeitstheorie darauf, dass es uneigentliche (die Furcht) und eigentliche (die Angst) Stimmungen gibt. In seinem späteren Werk beschreibt er auch die verschiedenen Stufen der Langeweile. Die vorliegende Arbeit beschäftigt sich nun mit dem Nachweis, dass diese Stimmungen (Furcht, Angst und Langeweile) morphologisch ähnlich sind und in ihren wesentlichen Strukturmerkmalen, die in Zusammenhang mit der Eigentlichkeit und Uneigentlichkeit stehen, übereinstimmen. Da die Langeweile jedoch fließende Übergänge zwischen ihren Stufen aufweist (Heidegger, 2010: 234f), kann man daraus schliessen, dass Eigentlichkeit und Uneigentlichkeit auch hier ineinanderfliessen.

Überblick

Die Arbeit gliedert sich in folgende Teile: Zu Beginn soll eine knappe Einführung in Heideggers Denken und in seine Begriffswelt erfolgen (Kapitel 2). Dies dient zum besseren Verständnis der gesamten Arbeit. Anschliessend werden die zwei zentralen Begriffe ‚*Eigentlichkeit*‘ und ‚*Uneigentlichkeit*‘ definiert (Kapitel 3). Von hier aus ist es nun möglich, sich Heideggers Befindlichkeitstheorie zu widmen. Zuerst sollen jene Stimmungen dargestellt werden, die laut Heidegger insofern ausgezeichnete Befindlichkeiten sind, als sie eine Unterscheidung zwischen uneigentlichen und eigentlichen Stimmungen ermöglichen. Dies wären die Furcht und die Angst (Kapitel 4). Daraufhin werden die drei Stufen der Langeweile skizziert und mit den Stimmungen Furcht und Angst verglichen (Kapitel 5). Es

wird sich zeigen, dass die Stimmungen morphologisch ähnlich sind und dies die gestellte Arbeitshypothese bekräftigt (Kapitel 6). Die Arbeit wird durch einige Schlussbemerkungen abgerundet.

Kapitel 2: Begriffliche Grundlagen

Bevor die Analyse von Eigentlichkeit und Uneigentlichkeit im Hinblick auf die Stimmungen beginnt, sollen zuerst wesentliche Begriffe aus Heideggers Denken kurz erläutert werden. Dies soll dazu dienen, das Verständnis der heideggerschen Befindlichkeitstheorie zu vertiefen und diese in einen konkreten Begriffsrahmen seiner Philosophie zu stellen. Selbstverständlich können hier nicht alle Begriffe dargestellt werden, da dies den Rahmen der vorliegenden Arbeit sprengen würde. Dementsprechend stehen also die Begriffe *Dasein*, *Welt* und *Stimmung* im Zentrum, mit der Bemerkung, dass auch besagte Begriffe nur zusammenfassend und vereinfacht erklärt werden. Auf eine kritische, exegetische Auslegung wird hier also grösstenteils verzichtet. Als Hauptquelle für die Begriffsklärung dient Heideggers Werk *Sein und Zeit*.

Dasein

„*Dasein*“ als zentraler Begriff in Heideggers Denken bedeutet eine Form des präsenten¹ Seins: Dasein bedeutet hier nicht primär ein räumliches ‚da‘-sein, sondern ein anwesend, aufmerksames, erschliessendes Seiendes (Heidegger, 2006: 12f). Was unter diesem Begriff genauer verstanden wird, soll nun geklärt werden. Zuerst wird erörtert, weshalb das Dasein eine vorrangige Stellung für Heideggers Philosophie besitzt und auch für die vorliegende Untersuchung relevant ist. Es wird sich zeigen, dass das Dasein in jener Seinsweise existiert, welche es ihm ermöglicht, die Frage nach dem Sinn von Sein überhaupt erst stellen zu können. Anschliessend gilt es, den Begriff ‚*Dasein*‘ von anderen Begriffen abzugrenzen. Zum Schluss sollte jene spezifische *Seinsweise* des Daseins - die *Existenz* - noch betrachtet werden.

Dasjenige Seiende, welches sich die Frage nach dem *Sein* und dessen Sinn überhaupt stellen kann, nennt Heidegger das ‚*Dasein*‘ (Heidegger, 2006: 7). Das Dasein sind wir - also die Menschen², welche durch ihre

Existenz in einem direkten Verhältnis zum Sein stehen und folglich auch ein *Seinsverständnis* besitzen. Erst dadurch, dass das Dasein ein Seinsverständnis besitzt, kann es sich also die Frage nach dem Sinn von Sein stellen. Das besagte Seinsverständnis, so Heidegger, ist jedoch dem Dasein vorerst nur *ontisch* nahe, was ungefähr so viel bedeutet, dass es sich des Seins zwar im Alltagsgebrauch bewusst ist, dies aber ontologisch/thematisch noch nicht erschlossen hat.³ Dennoch ist das Dasein ontisch dadurch ausgezeichnet, dass es eben *ontologisch* ist, also dazu disponiert, sich überhaupt erst diese Frage nach dem Sinn von Sein stellen zu können (Luckner, 2007: 19). Diese Zirkularität betrachtet Heidegger als notwendig, was in seinen weiteren Ausführungen noch deutlicher wird (vgl. Heidegger, 2006: 153). Was bedeutet es aber nun, wenn Heidegger meint, das Seinsverständnis sei dem Dasein in seiner Alltäglichkeit ontisch nahe? Dies hat damit zu tun, dass das Dasein *existiert*. Hierunter versteht sich: Es ist in die Zeit eingebunden und deshalb über sein eigenes Sein *besorgt*, was auch der Anlass dazu ist, sich die Frage nach dem Sinn von Sein zu stellen (Heidegger, 2006: 17). Was genau unter ‚*Existenz*‘ zu verstehen ist, wird unten noch genauer erläutert. Somit sollte nun jedoch die vorrangige Stellung des Daseins geklärt sein und es ist jetzt daran, diesen Begriff von anderen Begriffen abzugrenzen.

Da, wie schon erwähnt, sich das Dasein um sein eigenes Sein *sorgt*, steht es in einem speziellen Zusammenhang zum Sein, nämlich zu seinem je eigenen. Heidegger nennt dies die *Jemeinigkeit* (Heidegger, 2006: 42). Insofern ist es dem Dasein nicht gleichgültig, was mit seiner Zeit geschieht oder was es aus seiner Existenz macht. Dieser grundlegende, *eigentliche* Bezug (vgl. Kapitel 3) zum Sein individuiert somit ein einzelnes Dasein unter den jeweils anderen. Trotz dieser Individuation ist es jedoch laut Heidegger nicht angebracht, von einem Subjekt, einer Seele oder Substanz zu sprechen (vgl. Grondin in Rentsch, 2015: 15). Da Begriffe wie beispielsweise ‚*Substanz*‘ nach Heidegger in der philosophischen Tradition sowohl für Menschen als auch für Dinge verwendet wurden, hat er für den Menschen den Begriff ‚*Dasein*‘ eingeführt. Diesen Schritt vollzieht Heidegger dementsprechend, weil er versucht, mit der traditionellen Metaphysik

formuliert wird (vgl. Heidegger, 2007 und 2010).

1 Hierbei ist nicht eine Verfallenheit an die Gegenwart gedacht (was ‚präsentisch‘ wäre, vgl. Gumbrecht, 2016), sondern die Bedeutung erklärt sich aus dem folgendem Nebensatz.

2 Heidegger verzichtet zwar darauf, den Begriff ‚Mensch‘ zu verwenden (vgl. Luckner, 2007: 16), da dieser Begriff schon ontologisch vorbelastet ist. Dennoch ist es anzunehmen, dass der Begriff ‚Dasein‘ der Spezies Mensch als Seiendem zugeschrieben werden kann, was so auch in seinen späteren Werken explizit

3 Dieses Erschliessen wäre dann eine ontologische/thematische Auseinandersetzung und nicht eine ontische (Heidegger, 2006: 12). Diese zwei Begriffe müssen hier streng voneinander getrennt werden. Heidegger selbst versucht in seinem Werk das Seinsverständnis von der ontischen auf die ontologische Ebene zu heben.

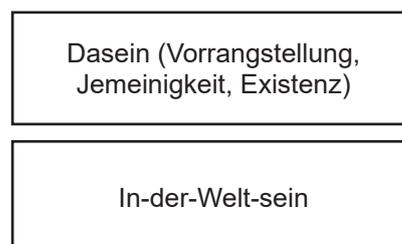
zu brechen, welche die Menschen - also das Dasein - bisher als dinghaftes Seiendes beschrieben haben. Das Dasein ist deshalb ontologisch nicht mit Steinen, Wolken oder anderen Dingen gleichzusetzen, sondern nimmt eine ontologische Sonderstellung ein. Dasein ist insofern auch nicht synonym mit *Leben* oder *Bewusstsein*, da diese Begriffe aus der Biologie und der Psychologie stammen und sich, so Heidegger, an den philosophischen Begriffen wie ‚*Subjekt*‘ und ‚*Substanz*‘ orientieren oder auf diesen aufbauen (Heidegger, 2006: 45ff). Weshalb die metaphysische Tradition bis anhin den Menschen als Ding verstanden und definiert hat, begründet Heidegger durch ein Fluchtverhalten, das eine Eigentümlichkeit des Daseins ausmacht (Heidegger, 2006: 44). Das Dasein tendiert meistens dazu, in seiner Existenz verdinglicht zu werden, oder ist dem eigenen ontologischen Status gegenüber zumindest indifferent.⁴ Nun habe ich zwei Aspekte des *Daseins* ausgearbeitet: Erstens nimmt das Dasein eine Vorrangstellung ein, wenn es um die Frage nach dem Sein geht. Zweitens ist das Dasein ein spezielles Seiendes, welches in der Seinsweise seiner je eigenen *Existenz* ist. Es zeichnet sich also nicht dadurch aus, *was* es ist, sondern *wie* es ist.

Das Dasein gleicht also keinem Ding mit irgendeiner Essenz und zufälligen Eigenschaften, es *hat zu sein* - muss also sein Leben führen - und dies nennt Heidegger *Existenz* (Heidegger, 2006: 12). Wie bereits erwähnt, ist Existenz daher die *Seinsweise* des Daseins. Im Unterschied zu Dingen und deren Seinsweisen ist die Existenz ein Vollzug, eine Bewegung - im räumlichen wie im zeitlichen Sinne (Safranski, 2015: 145). Diese Seinsweise besteht daher darin, sich mit dem *wie*, also den erschliess- und umsetzbaren Möglichkeiten, auseinanderzusetzen. Indem es sich für Möglichkeiten entscheidet, bzw. sich auf seine Seinsweisen hin *entwirft*, übernimmt das Dasein die Verantwortung über sein je eigenes Sein (Heidegger, 2006: 42f). Diese Auseinandersetzung die Verantwortung zu übernehmen und seine eigene Existenz zu entwerfen, nennt Heidegger auch *Sorge* (vgl. Kapitel 3 und Heidegger, 2006: 182). Die Sorge entsteht dadurch, dass das Dasein zeitlich begrenzt ist, was das ausschlaggebende Element der Existenz ausmacht.⁵

4 Diese Seinsvergessenheit ist auch der ausschlaggebende Grund, weshalb sich Heidegger mit der Frage nach dem Sinn von Sein beschäftigt und weshalb er es zum Anlass nimmt, diese Untersuchung zu legitimieren (Heidegger, 2006: 1).

5 Obwohl das Werk unvollendet ist, will Heidegger auf den Schluss hinaus, dass das Sein direkt mit der Zeit zusammenhängt (Heidegger, 2006: 1).

Nun schliesst sich also der Kreis, den ich hier nochmals kurz zusammenfassen möchte: Das Dasein ist ein spezielles Seiendes, das kein Ding ist, eben deshalb, weil es in der Seinsweise der Existenz *zu sein hat*. Dieser Bezug ist grundlegend dem einzelnen Dasein zugeschrieben und insofern *jemeinig*. Dadurch, dass die Existenz zeitlich begrenzt ist, kommt das Dasein auch zur Vorrangstellung, überhaupt erst nach dem Sinn von Sein zu fragen. Durch die Beschäftigung mit dieser Frage entwirft sich so das Dasein auf seine eigentlichen und uneigentlichen Möglichkeiten hin (vgl. Kapitel 3). Die Eigentümlichkeiten des Daseins werden also aus dessen Vorrangstellung, dessen Jemeinigkeit und dessen Existenz gebildet (vgl. Luckner, 2007: 31). Um nun auf die folgenden Begriffe überleiten zu können, muss noch erwähnt werden, dass das Dasein in seiner Existenz keine Essenz oder Eigenschaften wie ein Ding besitzt, sich aber dennoch durch ein *Strukturmoment* auszeichnet. Dieses Strukturmoment wird *In-der-Welt-sein* genannt (Heidegger, 2006: 53). *In-der-Welt-sein* ist daher eine notwendige und zugleich hinreichende Bedingung für das Dasein. Das *In-der-Welt-sein* wird in dieser Form geschrieben, um anzudeuten, dass alle Aspekte dieser Seinsverfassung *gleichursprünglich* sind, sich also nicht voneinander ableiten lassen, und in einem einheitlichen, zusammenhängenden Bezug zueinanderstehen (Heidegger, 2006: 53). Die folgenden Begriffe (*Welt*, *In-Sein*, und *Selbst*) sind also Aspekte des *In-der-Welt-seins*.



Grafik 1: *Dasein und seine Seinsverfassung*

Welt

Wie schon erwähnt, kann das Strukturmoment *In-der-Welt-sein* in drei gleichursprüngliche Aspekte unterteilt werden, welche die Seinsverfassung des Daseins ausmachen. Die drei Aspekte behandeln das *Wo*, das *Wer* und das *Wie*. Im Folgenden soll nun auf alle drei Aspekte eingegangen werden, begonnen mit dem *Wo*: die *Welt*.⁶ Hierbei soll zuerst wieder eine Abgrenzung

6 ‚Welt‘ darf hier nicht mit dem Weltbegriff aus Heideggers Spätwerk gleichgesetzt werden, welchen Heidegger dem Begriff der ‚Erde‘ gegenüberstellt (vgl. Heidegger, 2012).

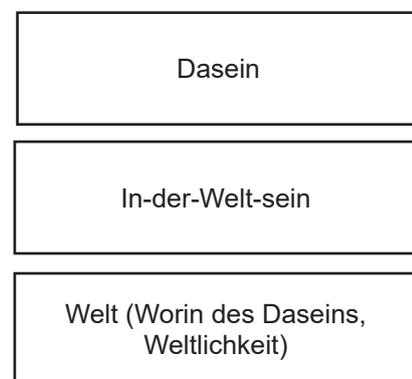
zu anderen Begriffsverwendungen vollzogen und anschliessend auf die heideggersche Bedeutung der *Welt* eingegangen werden.

Die *Welt*, so wie sie Heidegger phänomenologisch und fundamentalontologisch versteht, unterscheidet sich von den traditionellen Weltbegriffen der westlichen Philosophie. Insofern ist die Welt nicht die Summe der vorhandenen Dingen oder ein ontologischer Terminus für die Regionen besagter Dinge (Heidegger, 2006: 64). Die Welt ist zudem auch kein leerer Raumcontainer, der mit Gegenständen, Menschen etc. gefüllt ist (Luckner, 2007: 38). Somit grenzt sich der Begriff der ‚Welt‘ auch explizit vom cartesianischen Paradigma ab (vgl. Heidegger, 2006: §19ff). Die *Welt* kann hingegen auf zwei andere, ergänzende Weisen verstanden werden. Erstens bildet die Welt existenziell das *Worin*, in welchem das Dasein lebt (Heidegger, 2006: 65). Darunter versteht Heidegger einen Horizont - eine sich meldende Umwelt, in welcher das Dasein anderem Seienden begegnet (Pocai in Rentsch, 2015: 58). Die Welt ist folglich kein schlichter Raumcontainer, sondern eine dynamische, sich zeigende Sphäre, in der das Dasein in seiner Alltäglichkeit *da* ist. Nun stellt sich die Frage, wie denn das Dasein überhaupt *da* - also *anwesend* ist in dieser Welt. Heidegger geht hierfür in seiner Untersuchung von einer pragmatischen Seinsweise aus, indem er behauptet, dass das Dasein primär und zunächst handelnd, *besorgend* in der Welt zugange sei (Heidegger, 2006: 67). Damit postuliert Heidegger zugleich, dass der Umgang in der Welt nicht ein primär theoretisch-kontemplativer Zugang ist, sondern ein praktischer. Das bedeutet, wir sind meistens mit irgendetwas beschäftigt, betätigen uns mit Gegenständen, Projekten oder Menschen.⁷ Innerhalb dieser grundlegenden Beschäftigung ist das Dasein mehr oder minder *vertraut* mit dem, was es gerade unternimmt, und so erscheint ihm die Welt in verschiedenen *Bedeutungszusammenhängen*. So *flackert* die Welt in unterschiedlichen Weisen auf, wenn etwas auffällig wird, schief läuft, kaputt geht oder fehl am Platz ist, und somit den erwähnten Zusammenhang nicht erfüllen kann (Heidegger, 2006: 73).

Ausgehend von dieser ersten Bestimmung der Weltbedeutung hat es nun Sinn, auf die zweite Bedeutung der Welt zu wechseln. Die Welt ist eine existenziale Ermöglichungsinstanz für das Dasein, was Heidegger

Weltlichkeit nennt (Heidegger, 2006: 65). Indem das Dasein die Welt in den oben genannten Bedeutungszusammenhängen je schon zu einem gewissen Grad erschlossen hat, verfügt es über eine praktische Vertrautheit mit den gegebenen Gegenständen (Heidegger, 2010: 8). Hieraus leitet er später auch die räumliche Dimension der Welt ab, die bis dahin vernachlässigt wird (vgl. Heidegger, 2006: 102ff). Hierbei wird noch einmal der Bruch mit der traditionellen Philosophie explizit hervorgehoben. Das Dasein entdeckt in der Welt seine Möglichkeiten zu handeln, *Dasein ist weltlich*, und findet sich so mehr oder weniger in dieser Sphäre zurecht. Deutlich wird dies anhand der Äusserung Heideggers, der Mensch sei *weltbildend*, das Tier *weltarm* und der Stein *weltlos* (Heidegger, 2010: 263). Der Mensch als Dasein besitzt also, da er weltbildend ist, nicht nur die Fähigkeit, dem Seienden innerhalb der Welt zu begegnen, sondern auch diejenige, eigene Pläne zu *entwerfen* und zu verfolgen (Pocai in Rentsch, 2015: 59). Die Welt ist dem Dasein jedoch nie vollständig erschlossen, was zugleich zu einer Unvertrautheit oder gar einem Ohnmachtsgefühl führen kann (Pocai, 1996: 21). Hierauf wird in den späteren Teilen (vgl. Kapitel 4 und 5) noch genauer eingegangen.

Zusammenfassend lässt sich also sagen, dass die Welt jene Sphäre ist, in der sich das Dasein aufhält und in welcher es anderem Seienden begegnet. Diese Begegnungen von Gegenständen und Menschen sind primär dadurch gekennzeichnet, dass das Dasein praktisch mit den Gegenständen und den Menschen umgeht. Indem es mit verschiedenen Zusammenhängen vertraut ist, kann es sich dementsprechend verhalten und seine eigenen Pläne verfolgen. Welt als Strukturmoment des Daseins besagt, dass Welt da ist, wenn Dasein existiert.



Grafik 2: Welt als Aspekt des In-der-Welt-seins

7 Auf den Umgang mit Zeug (Gegenständen) und mit anderen Menschen (Mitsein) kann hier nicht ausführlich eingegangen werden. Für ersteres vgl. Heidegger (2006: §15ff) und letzteres (Heidegger, 2006: §25ff), sowie Kapitel 3 dieser Arbeit. Ausführliche Kommentare zu dieser Thematik finden sich bspw. bei Brandon (1983), Christensen (2007), Haugeland (2005) und Rentsch (2015).

Stimmung

Das *Wo* des *In-der-Welt-seins* wurde jetzt bestimmt. Nun können wir uns dem zweiten Aspekt, dem *Wie* - dem *In-Sein* als solchem - zuwenden. Die *Stimmung* (ontologisch-existential; die *Befindlichkeit*) ist ein Aspekt davon.⁸ Dazu muss erwähnt werden, dass die Stimmung kein *Vorhandenes* ist, kein Ding einer subjektiven Psyche, sondern das vorher erwähnte *Wie-es-ist* (Mugerauer, 2015: 83). Insofern gibt es mehrere Stimmungen, die das *In-Sein* des Daseins beeinflussen und qualitativ⁹ unterschiedlich prägen (vgl. Kapitel 4 und 5, sowie Heidegger, 2007). Was eine Stimmung im heideggerischen Sinne ausmacht, soll nun geklärt werden.

Da die Stimmung ein Aspekt des Strukturmomentes des Daseins ist, folgt daraus, dass das Dasein je schon gestimmt ist (Heidegger, 2006: 134). Selbst wenn wir uns ungestimmt fühlen, ist dies eine spezifische Form der Stimmung, zwar jene der *Gleichgültigkeit* (Heidegger, 2006: 134).¹⁰ Wir vermögen es zwar, uns in Stimmungen hineinzusteigern, aber es bleibt der Stimmung eigentümlich, dass sie nicht völlig kontrollierbar ist. Das Dasein *befindet sich* folglich in einer Stimmung. *Befinden* wird hier sowohl im räumlichen als auch im emotiven Sinn verstanden (Luckner, 2007: 62). In der Stimmung erschliesst sich dem Dasein sein Selbst, sowie auch der Weltbezug in unterschiedlichen Qualitäten und Massen. Ein fröhlicher Mensch wird deshalb die Welt anders wahrnehmen als ein zorniger. Es soll hier der Definition halber kurz auf drei wesentliche Aspekte der Stimmung eingegangen werden. Anschliessend wird noch ein Übergang zum *Verstehen* formuliert.

Erstens zeigt sich in der Stimmung, wie schon erwähnt, *wie* es dem Dasein ist. Durch die Stimmung erschliesst das Dasein sich selbst und auch die Welt und ihre Bedeutungszusammenhänge (Mugerauer, 2015: 85). Diese Erkennungsform ist noch präreflexiv und

vortheoretisch (Grosse, 2008: 44). Deshalb muss hier der Begriff von ‚*Erschlossenheit*‘ vom theoretischen *Erkennen* strikt getrennt werden. Nur weil das Dasein sich selbst und die Welt erschliesst, folgt daraus noch nicht, dass es nun schon alles erkannt und verstanden hat (Heidegger, 2006: 133).

Zweitens offenbart sich in der Stimmung der *Lastcharakter* der Existenz (Heidegger, 2006: 345). In der Stimmung zeigt sich dem Dasein, dass es *zu sein hat* (vgl. oben). Dass es seine Existenz vollziehen muss und sich je schon in einer Situation - der Welt - befindet, nennt Heidegger *Geworfenheit* (Heidegger, 2006: 135).

Drittens folgt daraus eine Spannung, die durch diesen Lastcharakter aufkommt. Da Stimmung im heideggerischen Sinne als passiver Aspekt verstanden werden kann, ist das Dasein diesbezüglich in seiner Handlungsfähigkeit eingeschränkt (Pocai, 1996: 45). Es wird vor die Aufgabe gestellt, sich sein eigenes Sein und dessen Last auf die Schultern zu werfen. Wenn sich das Dasein nun aber je schon in einer Stimmung befindet, dadurch diese Last erfährt und zugleich aber seine Existenz vollziehen muss, eröffnen sich zwei Auswege. Ein Ausweg besteht darin, zu versuchen, sich der Stimmung zu entziehen und den Lastcharakter zu verdecken (vgl. Kapitel 3 und Heidegger, 2006: 135ff). Der andere Ausweg besteht in der *Angänglichkeit* (Heidegger, 2006: 137). Das bedeutet, die Stimmung *betrifft* das Dasein in solchem Masse, dass es die Möglichkeit wahrnimmt, den Lastcharakter seiner Existenz zu übernehmen.

Hierzu bedarf es jedoch eines produktiven Parts des *In-Seins*, welchen Heidegger *Verstehen* nennt (Heidegger, 2006: 142). Als Strukturmoment ist das Dasein auch je schon verstehend und dementsprechend auch dafür verantwortlich, wie die Welt und das Selbst erschlossen werden. Durch das Verstehen wird dem Dasein die Möglichkeit eröffnet, seine Existenz zu vollziehen, Entscheidungen zu treffen und Pläne zu *entwerfen* (Heidegger, 2006: 146). Verstehen wird hier also in einem weiten Sinne verstanden und nicht als ein rein sprachliches Phänomen (Jung, 2012: 105).

Folgende Punkte können zusammenfassend für die Stimmung aufgeführt werden: Das Dasein ist je schon gestimmt und erschliesst damit mehr oder weniger sowie qualitativ das eigene Selbst und die Welt. Dadurch meldet sich beim Dasein der Lastcharakter der eigenen Existenz - dass es zu sein hat. Hierauf versucht das Dasein jene Stimmung zu verdecken, oder mittels Verstehen die gegebenen Möglichkeiten wahrzunehmen und umzusetzen.

8 Sowohl Befindlichkeit als auch Verstehen und Rede und ihre Modi des Verfallens sind Aspekte des In-Seins. Auf eine Betrachtung der anderen Aspekte muss in dieser Arbeit aus Platzgründen verzichtet werden.

9 Die Bezeichnung ‚qualitativ‘ ist hier und weiter unten als ontische und nicht als fundamentalontologische im Sinne Heideggers zu begreifen.

10 Diese Bemerkung von Heidegger ist in zweierlei Hinsicht interessant. Einerseits weist er später darauf hin, dass solche Stimmungen nur schlafen und erst geweckt werden müssen (vgl. Heidegger, 2010: 91). Welchen Einfluss dies auf seine Befindlichkeitstheorie hat, muss hier vernachlässigt werden. Andererseits finden wir in der fahlen Ungestimmtheit morphologische Ähnlichkeiten mit der Langeweilekonzeption, welche Heidegger erst später entwickelt (vgl. Kapitel 5, sowie Safranski, 2015: 138, oder Pocai, 1996: 48).



Grafik 3: In-Sein als Aspekt des In-der-Welt-seins

Kapitel 3: Das Man und das Selbst: zwischen Uneigentlichkeit und Eigentlichkeit

Zwei der insgesamt drei Aspekte des In-der-Welt-seins wurden bisher beleuchtet. Nun kann der letzte Aspekt betrachtet, d.h. die Frage nach dem *Wer* gestellt werden. Diese Arbeit folgt der Idee von Heideggers Alltäglichkeitskonzeption und wird somit zuerst den Begriff des ‚Mans‘ - und den damit verbundenen Begriff der *Uneigentlichkeit* definieren. Anschliessend wird der Fokus auf das *eigentliche* Selbst verlagert. Hierzu muss noch eine Klärung vorab folgen, die für diese Begriffserläuterungen relevant ist. Die Begriffe ‚*Eigentlichkeit*‘ und ‚*Uneigentlichkeit*‘, so scheint es, sind bei Heidegger in *Sein und Zeit* zweideutig, auch wenn die zwei Verwendungsarten zusammenhängen. Einerseits kann man diese zwei Begriffe in einem handlungstheoretischen Sinne verstehen, wobei sich das Dasein in seinen Entwürfen entscheiden kann.¹¹ Ein Dasein, das sich selbst wählt, wird eigentlich (vgl. unten). Andererseits kann man diese Begriffe auch als existenzial-ontologisch, also als *Seinsweisen* auffassen. Das bedeutet, das Dasein existiert während unterschiedlichen Zeitabschnitten als eigentlich oder uneigentlich. Die zweite Bedeutungsform, welche sich innerhalb der Stimmungen zeigt, steht für diese Untersuchung und ihre Arbeitshypothese im Zentrum (vgl. Kapitel 1, sowie 4 und 5).

Uneigentlichkeit und das Man

Auch beim *Wer* führt Heidegger seine Kritik an der traditionellen Metaphysik weiter. Das Dasein ist kein einzelnes Ding oder ein isoliertes Subjekt, sondern

¹¹ Für diesen Hinweis bin ich Marcello Ruta dankbar, der mich auf diese Zweideutigkeit aufmerksam gemacht hat. Dieses Verständnis der Begriffe, welches anscheinend der Philosophie Kierkegaards folgt (vgl. Kierkegaard, 2014), kann hier jedoch nicht weiter erläutert werden. Vgl. hierzu auch Pocai (1996: 84f).

es steht je schon mit anderen Menschen in einem Verhältnis. Dass also andere daseinsmässige Seiende schon zur Struktur des Daseins gehören, wird dadurch begründet, dass die Bedeutungszusammenhänge der Welt schon auf andere Menschen verweisen (Demmerling in Rentsch, 2015: 86). Ein hergestellter Gegenstand, bspw. die gesamte begegnende Umwelt, verweist auf andere

Akteure oder wird von anderen Personen mitverwendet etc.¹²

Heidegger geht nun in seiner Untersuchung davon aus, dass das Dasein zumeist und in seiner *Alltäglichkeit* jenes Seiende ist, welches in eine solche Gemeinschaft eingebettet ist - ihr *verfallen* ist. Was und wer wir sind, wird demgemäss primär durch die praktischen Lebensvollzüge der Gemeinschaft geregelt. Diese Gemeinschaft nennt er das *Man* (Heidegger, 2006: 126). Im *Mitsein* im Man erschliesst sich dem Dasein das Selbst und die Welt, indem es teilnimmt an der Art und Weise, wie das Man sich selbst und die Welt erschliesst. Demzufolge hat das Man je schon gewisse Deutungsmuster entworfen, welche sich in jedem Dasein reproduzieren und aus welchen auch dessen *Vertrautheit* mit sich selbst und der Welt resultiert (vgl. Kapitel 2). Das Man legt dementsprechend die Normen fest, welche für das Welt- und Selbstverständnis grundlegend sind. So betrachtet, übt das Man eine unscheinbare *Diktatur* auf das Dasein aus (Heidegger, 2006: 126). Unscheinbar ist die Macht deshalb, weil das Man niemand Bestimmtes und doch jeder zugleich ist und sich so nicht auf einzelne Personen oder Institutionen konzentriert.¹³

Nun gilt es zu klären, weshalb sich das Dasein dieser Diktatur des Mans unterwirft. Wie weiter oben schon erwähnt wurde, muss das Dasein den Lastcharakter seiner Existenz übernehmen. Indem es diese Last dem Man überantwortet, versucht es, die eigene Existenz zu erleichtern (Heidegger, 2006: 268). Das Man übernimmt die Entscheidungen des Daseins und entwirft

¹² Diese Verhältnisse und Begegnungen zu anderen Menschen unterscheiden sich von den Bezügen des Daseins zu nichtdaseinsmässigen Seienden (Luckner, 2007: 57). Erstere bezeichnet Heidegger mit dem Begriff Fürsorge und letztere mit dem Begriff Besorgen (Heidegger, 2006: 121), welche beide jedoch in einem Verhältnis zur existenzialen Sorge stehen.

¹³ Vgl. hierzu besonders Foucault, der diese Idee als Diskurs weiterentwickelt (Demmerling in Rentsch, 2015: 88 und bes. Foucault, 2014).

so für jedes Dasein ein normkonformes Leben. Dies erklärt auch die früher erwähnte Fluchttendenz von den Stimmungen des Daseins (Heidegger, 2006: 184). Im Man wird das Dasein verdinglicht. Es will damit seinem eigenen Selbst-sein-können (und -müssen) entkommen. Dass diese These der Fluchttendenz kein zeitgenössisches Phänomen, oder primär für westliche Kulturen intelligibel ist, versucht Heidegger dadurch zu vermeiden, dass er darauf hinweist, dass dies fundamentalontologisch zur Struktur des Daseins gehört (vgl. Heidegger, 2006: 176 und §10). Dieses *Verfallen* an die Diktatur des Mans ist infolgedessen unhintergebar, da das Man ein Aspekt des Strukturmoments des Daseins bildet. Dasein tendiert demgemäss je schon zur Uneigentlichkeit.

Das Dasein, das sein Selbst dem Man überantwortet, ist somit in seiner Seinsweise *verfallen* (Heidegger, 2006: 167). In diesem Sinne ist es nicht *sich selbst*, sondern eben *uneigentlich*. Wie sich dieses Verfallen auf die Existenz des Daseins auswirkt (*Gerede, Neugier, Zweideutigkeit*), darauf kann hier nicht weiter eingegangen werden. Es zeichnet sich jedoch durch eine schnellebige, unbestimmte Oberflächlichkeit aus, in der das Dasein sich Selbst und die Welt als und durch Innerweltliches versteht (vgl. Kapitel 4 und 5). Diese Verdinglichung nennt Heidegger das *Sein-bei* (Heidegger, 2006: 192). *Bei* den Dingen und den anderen Menschen *sein*, ist daher die ausdrückliche Form des Verfallens. Das Dasein wird dadurch zerstreut und muss sich, wenn es sich selbst - also *eigentlich* -sein will, überhaupt erst finden (Heidegger, 2006: 129). Obwohl Heidegger darauf hinweist, dass das Verfallen nicht negativ konnotiert ist, da es ein Strukturmoment des Daseins ist (vgl. Heidegger, 2006: 128 und 167), so ist die Debatte darüber doch nicht eindeutig. Denn es ist sehr wahrscheinlich anzunehmen, dass er mit der Konzeption des Daseins in seiner Existenz auf das *eigentliche Selbst-sein* hinauswill.¹⁴

Das Dasein ist also in seiner Alltäglichkeit nicht sich selbst, sondern uneigentlich und dem Man verfallen. Es übernimmt die vorherrschenden Bedeutungszusammenhänge und versteht sich selbst und die Welt durch innerweltliche Entwürfe des Seinsverständnisses. Dadurch entledigt sich das Dasein des Lastcharakters seiner eigenen Existenz. Obwohl das Dasein meistens uneigentlich ist, ist die *Eigentlichkeit* als Seinsweise ontologisch gleichursprünglich. Dieser wollen wir uns nun zuwenden.

Eigentlichkeit und das Selbst

Dieser Abschnitt gliedert sich in zwei Teile. Im ersten wird es darum gehen, die *Eigentlichkeit* im Hinblick auf die Abgrenzung zur Uneigentlichkeit zu definieren. Der zweite Teil beschäftigt sich anschliessend mit der Frage, wie das Dasein sich selbst - also eigentlich - werden kann. In diesem Sinne ergänzen sich die beiden Teile, überlappen sich aber auch thematisch in einigen Punkten.

Teil 1, Begriffsklärung:

Wie schon erwähnt, ist die *eigentliche* Seinsverfassung ontologisch gleichursprünglich wie das Leben im Man, auch wenn das Dasein zumeist ans Man verfallen ist. Dennoch ist die *eigentliche* Seinsweise nur deshalb möglich, weil es das Man gibt (Luckner, 2007: 59). Eigentlichkeit und Uneigentlichkeit sind folglich Seinsmöglichkeiten. Dies soll aber nicht in einem kategorialen Sinne verstanden werden (kein *was es ist*), sondern in einem existenzialen, so gesehen in dem Sinne, *wie* die Existenz vollzogen wird.

Ein eigentliches Dasein (das *Selbst*) ist damit als Abgrenzung zum uneigentlichen Man eine selbständige, selbstbestimmte Existenz, welche sich die möglichen Lebensentwürfe durch Entschlüsse zu *eigen* macht (Heidegger, 2006: 42). Daraus folgt, dass es sich nicht auf die vorher erwähnten, schon bestehenden Bedeutungsmuster des Mans verlässt, sondern sich selbst auf seine eigenen Möglichkeiten hin entscheidet, indem es einige Bedeutungsmuster übernimmt und andere abändert, oder neue entwickelt (Hüsch, 2014: 139). Das Dasein erschafft dadurch Sinn. Ihm erschliesst sich die Welt und sein Selbst auf seinen eigenen Sinn hin.

Heidegger verzichtet hier absichtlich darauf, inhaltliche Punkte anzugeben, welche Aktivitäten und Deutungsmuster Sinn ergeben, weil dies seiner Existenzialontologie widersprechen würde. Wäre der Sinn von aussen her herantragbar, so wäre Sinn nur eine kategoriale Bestimmung des Daseins. Er möchte jedoch existenzial bestimmen, wie das Dasein *eigentlich* sein kann, nämlich genau dadurch, dass es Sinn erschafft (Heidegger, 2006: 303). Was der Inhalt dieses Sinnes ist, bleibt somit für seine Untersuchung irrelevant.

Eigentlich sein bedeutet also, sich für einen selbstbestimmten Sinn zu entschliessen. Darin liegt offenbar ein Aspekt der *Vereinzelung* (Safranski, 2015: 322). Das Dasein wird deshalb vom Man gelöst und wird zu seinem je eigenen Selbst. Jedoch muss sich dieses Selbst nicht notwendig vom Man oder von der Öffentlichkeit entfernen und sich aus der gemeinsam erschlossenen Welt zurückziehen, sondern das Dasein *eignet* sich die

¹⁴ Vgl. Beiträge zur dieser Debatte bspw. Adorno (2015), Demmerling und Luckner in Rentsch (2015), Poci (1996) sowie Safranski (2015).

öffentliche Welt überhaupt erst an. Die Welt wird dadurch seine Welt (Luckner in Rentsch, 2015: 148). Dies ist auch unter dem oben erwähnten Punkt zu verstehen, dass das Dasein nicht alle Deutungsmuster und Lebensentwürfe zu verwerfen hat, sondern es wählt sich unter den ergreifbaren Möglichkeiten seine je eigenen aus. Zusammengefasst erzeugt das eigentliche Dasein seinen eigenen Seinssinn für seine Pläne anhand aller möglichen Entwürfe, die ihm zur Verfügung stehen.

Da das Dasein zumeist aber schon an das Man verfallen ist, stellt sich die Frage, wie es überhaupt zur Eigentlichkeit, zu seinem Selbst finden kann. Dies soll nun erörtert werden.

Teil 2, Selbstwerdung:

Das Dasein muss, um eigentlich sein zu können, sich selbst wählen (Heidegger, 2006: 268). Diese Wahl besteht nicht zwischen dem Man und dem Selbst, als könnte es auch die Option in Betracht ziehen, uneigentlich zu sein. Zur Uneigentlichkeit zieht es das Dasein nur aufgrund der Flucht tendenz, den Lastcharakter seiner Existenz abzugeben. Eine Entscheidung für die Uneigentlichkeit ist demzufolge aber nicht möglich (Luckner in Rentsch, 2015: 147). Auch das Man wählt sich nicht, es reproduziert schlicht und bleibt unentschieden. Insofern ist die Wahl zur Eigentlichkeit alternativlos. ‚Wähle, oder unterlasse es zu wählen/bleibe unentschieden‘, ist somit Heideggers These.

Die Wahl zur Eigentlichkeit wird dadurch ermöglicht, dass sich beim Dasein das *Gewissen* meldet (Heidegger, 2006: 271f). Der Begriff des ‚*Gewissens*‘ darf hier nicht im herkömmlichen Sinne verstanden werden, als wäre es etwas wie eine innere Stimme, Gott, gesellschaftliche Normen, das Über-Ich o.ä., welches das Dasein auf ein *Sollen* hindrängt. Der Ruf des Gewissens kommt vom Dasein selbst, das sich über seine Existenz *sorgt* (Heidegger, 2006: 286). Da das Dasein meist schon verfallen ist, kann der Gewissensruf irgendwann und plötzlich auftreten.¹⁵ So unterscheidet sich der Ruf von moralischen Gewissenskonzeptionen, welche sich erst bei ethischen Erwägungen Gehör verschaffen. Ein weiterer relevanter Unterschied zu den üblichen Gewissensvorstellungen besteht darin, dass der *Rufcharakter* des heideggerschen Gewissens im *Schweigen* besteht und sich nicht artikuliert. Nur der schweigenden Unheimlichkeit gelingt es, das *Gerede* des öffentlichen Mans zu übertönen (Heidegger, 2006: 273). Erst durch diese schweigende Leere kann das

Gewissen mit seinem Ruf Aufmerksamkeit erlangen und dem Dasein über sein eigenes *Selbst-sein-können* Klarheit verschaffen. Das *Selbst-sein-können* besteht nun darin, so Heidegger, dass sich das Dasein seiner eigenen Endlichkeit bewusst wird (Luckner, 2007: 114). Dies ist allein im Man nicht möglich. Der Tod, der zwangsläufig eintreten und der eigenen Existenz ein zeitliches Ende setzen wird, erzeugt erst die Last der Existenzialität und ermöglicht dadurch, dass Sinn überhaupt erst erschaffen werden kann.

Dadurch, dass der Tod notwendigerweise eintreffen wird, muss sich das Dasein in seiner verbleibenden Zeit entwerfen und damit andere Lebensentwürfe ausschliessen. Insofern wird das Dasein *schuldig*. Ebenso darf hier dieser Begriff nicht im umgangssprachlichen Sinne verstanden werden, sondern existenzial. Das Dasein ist je schon schuldig (Heidegger, 2006: 283). Die Schuld, sich selbst zu entwerfen, sein eigenes Leben zu führen und Sinn zu schaffen, wird also durch das Gewissen offenbar. Dies erklärt auch die Flucht tendenz des Daseins ins Man (vgl. oben und Kapitel 2). Um dem Selbst und dieser offenbarten Schuld zu entgehen, übergibt das Dasein den Lastcharakter dem Man, das die vorgefertigten Lebensentwürfe dem Dasein aufdrängt (Hüsch, 2014: 141). So wird das Dasein leicht, aber dafür uneigentlich.

Wählt das Dasein jedoch den Weg, auf den Ruf seines Gewissens zu hören, seine je eigenen Möglichkeiten im Hinblick auf den eigenen Tod zu ergreifen und sich jene anzueignen, so wird es dadurch eigentlich. Das Dasein *entschliesst* sich zur Eigentlichkeit und damit zu seinem Selbst (Heidegger, 2006: 298). Der Begriff ‚*Entschlossenheit*‘ ist hier aus zwei Gründen relevant. Erstens übernimmt das Dasein die *Verantwortung* für seine *Wahl* zur Eigentlichkeit und die damit einhergehenden Lebensentwürfe, die es sich angeeignet hat. Es bekennt sich zur existenzialen Schuld. Hierdurch gestaltet es seine Existenz - seine Zeit - selbst (Hüsch, 2014: 117). Zweitens ist das Dasein durch seine *Entschlossenheit* *offen* für das Gewissen, es kann es erschliessen und auf den stillen Ruf hören. Offen bedeutet hier sowohl räumlich (*Situation*) als auch zeitlich (*Augenblick*), worin sich die Thematik der Stimmung in ihrer Erschlossenheit wieder aufrollen lässt (vgl. Kapitel 4 und 5, sowie Luckner, 2007: 122).

Zum Schluss sollen die relevanten Punkte nochmals vorgestellt werden: Das meist schon verfallene Dasein im Man erhält die Möglichkeit eigentlich zu werden dadurch, dass sich sein eigenes Gewissen meldet. Das Gewissen ruft im Schweigen und weist auf die Endlichkeit der Existenz hin. Dafür aufgeschlossen, über-

¹⁵ Vgl. hierzu auch das ‚Absurde‘ bei Camus, welches in dieser Hinsicht grosse Ähnlichkeiten mit dem heideggerschen Gewissen aufweist (Camus, 2013).



Grafik 4: Das Selbst als Aspekt des In-der-Welt-seins

nimmt das Dasein die Verantwortung über sein eigenes Selbst und entwirft selbstbestimmt seine Existenz anhand der ergreifbaren Möglichkeiten und wird dadurch eigentlich.

Anhand der zwei nächsten Kapitel (4 und 5) soll nun gezeigt werden, wie sich der Ruf zur Eigentlichkeit innerhalb von Stimmungen meldet. Dafür werden zwei geeignete Stimmungen - die Angst und die tiefe Langeweile - ausgewählt und im Gegenzug noch defizitäre Stimmungen vorgestellt (Furcht und oberflächliche Langeweile), die sich nach Heidegger als uneigentlich erweisen.

Zum Abschluss und der Vollständigkeit halber noch eine Grafik der Inhalte von Kapitel 2 und 3 (Grafik 4).

Kapitel 4: Furcht und Angst als ausgezeichnete Stimmungen

In diesem sowie im nächsten Kapitel sollen nun Stimmungen untersucht werden, um die Hypothese der Arbeit zu überprüfen (vgl. Kapitel 1). Da Eigentlichkeit und Uneigentlichkeit nicht nur im handlungstheoretischen Sinne, sondern auch als Seinsverfassungen betrachtet werden können, soll nun gezeigt werden, wie das Dasein eigentlich und uneigentlich ist, während es sich gestimmt in der Welt befindet. Der Aufbau dieses Kapitels soll dem Schema Heideggers in *Sein und Zeit* entsprechen, wobei zuerst die uneigentliche Stimmung, die Furcht, erläutert wird und anschliessend die Angst, welche hingegen eine eigentliche Stimmung ist.

Furcht

Die Furcht als Stimmung ist nach Heidegger ein Modus der Befindlichkeit (Mugerauer, 2015: 89). In der Furcht erschliesst das Dasein die Welt und das Selbst auf eine bestimmte Weise (vgl. Kapitel 2).

Interessant ist hier, dass Heidegger eine Unterteilung einführt, die zwischen alltäglichen und nicht-alltäglichen Stimmungen unterscheidet (Heidegger, 2007:

33). Furcht wäre in diesem Sinne eine alltägliche Stimmung. Auf diese Aussage wird am Schluss dieses Abschnitts noch weiter eingegangen. Um die phänomenologische Struktur dieser Befindlichkeit nun zu erläutern, orientiere ich mich an Heideggers Idee, die Furcht auf das *Wovor*, das *Fürchten* selbst und das *Worum* des Fürchtens hin zu beleuchten. Diese erweisen sich als die Strukturmomente der einzelnen Stimmungen. Anschliessend folgen einige Bemerkungen darüber, weshalb die Furcht nun eine uneigentliche Stimmung ist.

Das Dasein fürchtet sich vor dem *Furchtbaren*. Das Furchtbare ist ein Innerweltliches, sei es etwas Dinghaftes oder andere Menschen (Heidegger, 2006: 140). Die Furcht ist somit eine *intentionale* Stimmung, das bedeutet, ihr Charakter ist dadurch gekennzeichnet, dass sie sich auf einen Gegenstand oder auf andere Menschen (evtl. auch Ereignisse) bezieht. Wir fürchten uns beispielsweise vor dem Vater, Käfern, juristischen Urteilen o.ä.¹⁶ Heidegger betont hierbei besonders das Furchtbare, welches das Dasein in seinen Entwürfen bedroht (Pocai, 1996: 68). Das Furchtbare zeichnet sich hierbei durch verschiedene Punkte aus (vgl. Heidegger, 2006: 140f). Erstens dadurch, dass es dem Dasein in einem gewissen Sinne abträglich ist. Das Dasein erschliesst also das innerweltlich Furchtbare als etwas Negatives. Zweitens muss das Furchtbare in der Nähe des Daseins sein oder es muss sich an das Dasein herannahen. Durch diesen Umstand ist das Furchtbare dem Dasein nicht geheuer, da es zwar in einer Form anwesend, aber zugleich auch verhüllt und somit abwesend ist. Drittens ist das Furchtbare auch ein innerweltlich Seiendes, das vom Dasein in seinen

16 Um die Plausibilität und Anwendbarkeit von Heideggers Stimmungskonzeptionen zu stärken, verweise ich bei jeder Stimmung auf illustrierende Beispiele aus literarischen Werken, die meiner Meinung nach diese Befindlichkeit wiedergeben. Für die Furcht sind die Werke von Kafka - die Verwandlung (2007) und der Prozess (2012) - überaus passend.

Entwürfen nicht beherrscht werden kann. So kann es das Bedrohliche nicht oder nur schwer in seine eigenen Pläne integrieren. Die letzte Eigentümlichkeit des Furchtbaren besteht darin, dass es nur die Möglichkeit besitzt einzutreffen und dies das Dasein noch weiter verunsichert.

Das *Fürchten* selbst besteht darin, dass das Dasein von diesem besagten Furchtbaren betroffen ist (Heidegger, 2006: 141). Im Unterschied zu anderen Stimmungskonzeptionen behauptet Heidegger, dass die Befindlichkeit Furcht vorrangig sei, indem das Dasein zuerst gestimmt ist und erst anschliessend das Furchtbare erschlossen wird. Hierunter ist speziell der entwerfende Aspekt des In-der-Welt-seins zu verstehen, da das innerweltlich Furchtbare die Entwurfspläne des Daseins durchkreuzen kann (Pocai, 1996: 70). Das Dasein, mit seinen Entwürfen, beginnt sich um jene zu fürchten und entdeckt dadurch das Furchtbare. Dies ist jedoch nur bei uneigentlichen Stimmungen der Fall, was sich bei der Angst und bei der tiefen Langeweile noch deutlicher zeigen wird.

Nun fürchtet sich das Dasein *um* sich selbst. Jedoch fürchtet es sich nicht um sein eigentliches Selbst, sondern um sich, wie es sich aus der Welt und durch das Man versteht (Heidegger, 2006: 141). Da die Furcht eine alltägliche Stimmung ist und das Man die alltägliche Seinsweise des Daseins, ist das Dasein je schon an Innerweltliches verfallen - das *Sein-bei* (vgl. Kapitel 3) - und versteht sich und die Welt in dieser Weise. Es fürchtet sich um die fremdbestimmten Entwürfe (von Welt und Selbst) des Mans und ist demgemäss in einer existenziell-ontischen Stimmung. Heidegger geht zum Schluss noch auf den Punkt ein, dass wir uns auch um Gegenstände und andere Menschen fürchten können. Um diese, so argumentiert er, wird jedoch aus dem Selbstverständnis des Daseins selbst gefürchtet, indem es sie zu sich selbst rechnet (Heidegger, 2006, 142). Auch der Umgang mit der Furcht ist primär uneigentlich, da wir zumeist versuchen, dem Furchtbaren durch geschäftiges Treiben, durch Vermeidung oder Ignoranz zu entgehen (Mugerauer, 2015: 114). Das Dasein flieht grundsätzlich vor dem Furchtbaren (Heidegger, 2006: 185).

Die Furcht ist dementsprechend eine intentionale Stimmung, die sich auf Innerweltliches richtet. Insofern ist die Seinsweise innerhalb dieser Stimmung primär eine verfallene und daher auch *uneigentlich*, da sie sich allein mit Innerweltlichem beschäftigt (Heidegger, 2006: 189). Das Selbst- und das Weltverständnis entspricht somit den Entwürfen des Mans. Die eigentliche Stimmung als Äquivalent, die Angst, soll nun betrach-

tet werden. Es wird sich zeigen, dass man die Furcht als uneigentlicher Modus der Angst betrachten kann (Heidegger, 2006: 189).

Angst

Die Angst ist hingegen keine alltägliche Stimmung. Denn sie stört das uneigentliche Verfallen ans Man (Merker in Rentsch, 2015: 113). So ist, laut Heidegger, die Angst eine Grundbefindlichkeit, eine latente Stimmung, die jederzeit zum Vorschein kommen kann. Eine Grundbefindlichkeit ist sie deshalb, weil es in ihr um das ganze Dasein geht.¹⁷ In ihr offenbart sich der Lastcharakter des Daseins, sie fragt nach dem Sinn des eigenen Seins (Safranski, 2015: 176). Insofern ist die Angst eine existenzial-ontologische Stimmung und man erkennt aus vorigen Überlegungen (vgl. Kapitel 3), dass das Dasein versucht ist, diese Stimmung zu verdrängen, vor sich selbst zu fliehen und uneigentlich zu bleiben (Heidegger, 2006: 189). Wir richten uns nun für die Erläuterung auch nach der Struktur dieser Befindlichkeit in Hinsicht auf die Abgrenzung zur uneigentlichen Furcht. Zum Schluss sollen noch drei Punkte aufgeführt werden, die sich im späteren Werk Heideggers zur Angst herauskristallisieren.

In der Angst ängstigt sich das Dasein nicht vor etwas Innerweltlichem, sondern von der *Welt als solcher* (Heidegger, 2006: 187). Kein Teil der Welt ist der Ursprung oder der Grund der Angst, sondern das In-der-Welt-sein als solches. Das bedeutet, die vom Man fremdbestimmten Bedeutungszusammenhänge, welche die Weltlichkeit ausmachen, werden irrelevant. Das Dasein verliert seine Vertrautheit. Es gibt insofern *nichts* was beängstigen könnte und daher bleibt das *Wovor* der Angst unbestimmt (Heidegger, 2006: 186). Das In-der-Welt-sein, die Eigentümlichkeit des Daseins, kommt somit dem Dasein selbst abhanden. Es wird von einer Fremdheit überwältigt, indem die vertrauten Zusammenhänge entgleiten, es wird ihm *unheimlich*, weil es seine Lebenspläne nicht mehr wie gehabt entwerfen kann. Der Unterschied, um es zu verdeutlichen, liegt also darin, dass sich das Dasein in der Furcht um die uneigentlichen Pläne und deren Misslingen fürchtet. In der Angst hingegen ängstigt es sich davor, diese Pläne überhaupt nicht mehr übernehmen und entwerfen zu können.

Dadurch ängstigt sich das Dasein. Es verliert seine gewohnte Weltlichkeit und wird von einer aufkom-

¹⁷ Ebenfalls ist die tiefe Langeweile eine Grundbefindlichkeit und es wird sich auch zeigen, dass diese zwei Stimmungen morphologisch viele Parallelen aufweisen. Dennoch ist die Angst von der tiefen Langeweile zu unterscheiden (vgl. hierzu Safranski, 2015: 183 und Pocai, 1996: 107f).

menden, undefinierbaren Unheimlichkeit entmachtet (Luckner, 2007: 83). Kein innerweltliches Ding und keine anderen Menschen vermögen es, dem Dasein diese verlorene Beruhigung zurückzubringen. So wird dem Dasein sein Entwerfen-Können nach dem Diktat des Mans verhindert. Dem Dasein offenbart sich das nackte *Dass es zu sein hat* (vgl. Kapitel 2). Hierdurch wird das Dasein vereinzelt und auf sich selbst zurückgeworfen (Luckner, 2007: 82). Das Dasein erlebt in der Angst eine Selbsterfahrung. Deshalb ermöglicht die Angst auch den Weg zur Eigentlichkeit (Heidegger, 2006: 188). Insofern bedarf es auch des Mutes, sich dieser Angst zu stellen, um schliesslich eigentlich werden zu können (vgl. Safranski, 2015: 171).

So erklärt sich nun auch, *worum* sich das Dasein ängstigt. Wovor und Worum werden das Selbe. Das Dasein ängstigt sich um das In-der-Welt-sein, seine eigenen Möglichkeiten, weil diese unbestimmt sind, oder ihre Bedeutung verlieren. Angst erschliesst somit überhaupt erst das eigentliche *Möglichsein* (Heidegger, 2006: 188).

Es soll nun noch genauer darauf eingegangen werden, weshalb die Angst eine eigentliche Stimmung ist. Aufgrund des Strukturmerkmals ist das Dasein zumeist schon an das Man verfallen (vgl. Kapitel 3). Es versteht sich durch das Man, verhält sich dementsprechend so zu den Dingen und anderen Menschen (Sein-bei). Dieses Verfallen kann als eine Fluchtbewegung gegenüber der Angst verstanden werden. Fragt man sich nun, ob das uneigentliche Verfallen oder die Stimmung Angst grundlegender ist, so antwortet Heidegger, dass die Flucht vor der Angst nur möglich ist, wenn diese Stimmung grundlegender ist.¹⁸ Eine Flucht vor sich selbst ist nur möglich, wenn sich das Dasein schon als Ganzes erschlossen hat (Heidegger, 2006: 184). Erschlossen ist hier nicht mit vollständig erkannt gleichzusetzen, sondern das Dasein hat nur eine schlummernde Vorahnung von sich, welche sich in der Angst und schliesslich durch das Gewissen melden kann (vgl. Kapitel 3). Die Flucht besteht nun darin, dass das Dasein nicht vor Innerweltlichem flüchtet, wie es dies bei der Furcht tut, stattdessen will es gerade zum Innerweltlichen hin fliehen, um sich vor dem Ausbrechen der Angst zu schützen (Luckner, 2007: 82). Bricht die Angst nun doch unerwartet aus, so wird das Dasein zu seinem eigentlichen Seinkönnen hingetrieben. Es ist der Angst und dem Gewissensruf ausgeliefert. Es muss hier selbst entwerfen. Die Angst ist somit der Weg von

der Uneigentlichkeit zur Eigentlichkeit hin. Das Dasein lernt sich selbst und die Welt durch seine eigenen Entwürfe zu verstehen und übernimmt daher die Verantwortung zur existenzialen Schuld. Diesen Prozess kann das Dasein in der Furcht nicht einholen, weshalb Heidegger folgert, dass die Furcht die uneigentliche, verfallene Form der Angst ist (Heidegger, 2006: 189).

Bevor wir uns nun der Langeweile zuwenden, will ich hier noch auf drei Änderungen eingehen, die Heidegger selbst bei seinem Angstbegriff vorgenommen hat. Insofern hat er in *Was ist Metaphysik?* einige unterbestimmte Punkte bei der Angst von *Sein und Zeit* verändert und verbessert.

Erstens zeigt sich in der Angst das totale *Nichts*, das sich dem Seiendem gegenüberstellt, nicht als Verneinung, sondern als das Andere (Heidegger, 2007: 34). Das Nichts wird der tätige Grund, weshalb das Seiende im Ganzen dem Dasein entgleitet und es daraufhin mit dieser Unvertrautheit in Berührung kommt. Das Nichts *nichtet* im Moment der Angst (Heidegger, 2007: 37). In dieser *Nichtung* lösen sich die Bedeutungszusammenhänge der Welt auf und sie zeigt sich dem Dasein als das Nicht-Entworfene. Ein umfangliches, vertrautes Verstehen (vgl. Kapitel 2) wird dadurch verunmöglicht. Dies löst das vorher erwähnte Gefühl der *Befremdlichkeit* aus (Pocai, 1996: 109). So zwingt es das Dasein zum selbständigen Entwerfen.

Zweitens folgert er auch daraus, dass das totale Nichts erst die Endlichkeit - der Tod als das Ende der Existenz - des Daseins offenbart. Somit wird ein Grund identifiziert, weshalb sich das Gewissen meldet. Dies ist also der Punkt, der das Dasein zur Selbstwerdung führt (Heidegger, 2007: 38). Wir nehmen in der Angst wahr, dass wir sind, und erschliessen damit auch, wie wir sein können.

Drittens grenzt Heidegger die Befindlichkeit noch genauer von der verwirrenden, aufreibenden Furcht ab, indem er das Ängstigen als bedrückende und doch negativ empfundene *Ruhe* definiert (Heidegger, 2007: 34). Erst in dieser Stille kann sich auch das Gewissen melden.

Zusammengefasst ist die Angst eine latente Grundbefindlichkeit, die bei ihrem Aufkommen dazu führt, dass das Dasein seinen gewohnten Zugang zur Welt verliert und zu seinem eigentlichen Selbstseinkönnen hingetrieben wird. Dies ist jedoch ein unangenehmer Prozess für das Dasein und es versucht dieser Stimmung auszuweichen, indem es zu innerweltlichen Dingen flieht. Die Furcht wird dadurch ein uneigentlicher Modus der ursprünglicheren Angst.¹⁹

¹⁸ Dass sich hier ein Widerspruch aufspannt, wird sofort auffällig, da Heidegger vorher versucht hatte, Eigentlichkeit und Uneigentlichkeit ontologisch gleichzusetzen (vgl. hierzu Merker und Kiesel in Rentsch, 2015, sowie Pocai, 1996).

¹⁹ Faust (Goethe, 2000) und Hamlet (Shakespeare, 2013) dienen als geeignete Beispiele für Angst.

	Furcht	Angst
Wovor	Innerweltliches	In-der-Welt-sein
Betroffenheit	Fürchten im Sein-bei	Existenziales ängstigen
Worum	Verfallene Dasein	In-der-Welt-sein
Seinsweise	uneigentlich	eigentlich

Grafik 5: Furcht und Angst als ausgezeichnete Stimmungen

Kapitel 5: Die drei Stufen der Langeweile

Auch die *Langeweile* ist ein Modus der Befindlichkeit. Heidegger unterteilt sie jedoch in drei verschiedene Stufen, die alle unterschiedliche Stimmungen repräsentieren (Heidegger, 2010: 122). Wie beim letzten Kapitel beginnen wir auch hier zuerst mit der alltäglichen Stimmung, der oberflächlichen Langeweile. Folgend wird die zweite Form dargestellt und am Schluss soll sich der Fokus auf die tiefe Langeweile richten, welche, ähnlich wie die Angst, eine latente *Grundstimmung* ist (Heidegger, 2010: 89). Ebenso wie zur Angst, verfügt das Dasein die Disposition zur tiefen Langeweile. Die tiefe Langeweile ist auch durch ihre *Angänglichkeit* (vgl. Kapitel 2) eine ausgezeichnete Stimmung, die das Dasein veranlasst eigentlich werden zu können. Auch in ihr kann sich das Nichts zeigen, kann sich das Gewissen melden (Safranski, 2015: 60). Obwohl Heidegger hier auf die Begriffe ‚*Eigentlichkeit*‘ und ‚*Uneigentlichkeit*‘ weitgehend verzichtet, soll in diesem Kapitel aufgezeigt werden, dass die jeweiligen Formen der Langeweile morphologisch den Stimmungen Furcht und Angst ähnlich sind. Dadurch soll der Nachweis gezogen werden, dass auch diese Stimmungen in einem gewissen Grade eigentlich und uneigentlich sind. Während die Langeweile in Heideggers Werk zuerst nur wenig Beachtung erfährt, wird sie in den *Grundbegriffen der Metaphysik* zum zentralen Teil seiner Forschung. Das hat zweierlei Gründe: Erstens ist die Langeweile eine Stimmung, welche ein engeres Verhältnis mit der Zeit aufweist, als dies die Furcht oder die Angst tut, und zweitens liegt es auch, so Heidegger, an der Aktualität dieser Stimmung (Pocai, 1996: 199). Die tiefe Langeweile ist in der Moderne deutlich ausgeprägt. Sie wird zum *Hintergrundrauschen* des Daseins. Dieses Werk ist folglich auch die Grundlage für das vorliegende Kapitel.

Einfachheitshalber übernehme ich die Befindlichkeitsstruktur, wie sie im vorigen Kapitel angewendet wurde. Das bedeutet, ich untersuche die verschiedenen Langeweileformen auf ihr *Wovor*, das *Sich-Langweilen* und das *Worum*. Hierzu lassen sich folgende

Punkte bei Heidegger finden. Das, wovor sich das Dasein langweilt, ist das *Leerlassende* (Heidegger, 2010: 130). In jeder Stufe ist dies etwas anderes. Heidegger verzichtet hier absichtlich darauf, die kausale Ursache anzugeben, ob das

Attribut ‚*langweilig*‘ vom Subjekt oder vom Objekt her gedacht werden muss (Hüsch, 2014: 66). Eine strikte Trennung von Subjekt und Objekt scheint ihm hier sowieso nicht sinnvoll, da dies wiederum dem Paradigma der traditionellen Philosophie entsprechen würde. Das *Sich-Langweilen*, als zweites Strukturmoment, besteht in der *Gleichgültigkeit* (Heidegger, 2010: 115). Das Dasein verliert den Bezug zu dem, wovon es gelangweilt wird, indem besagtes an existenziellem Wert verliert - es wird dem Dasein belanglos. Das letzte Strukturmoment, das *Worum*, ist die Zeit selbst. Sie wird dem Dasein schlicht *hingehalten* und es muss diese Unannehmlichkeit ertragen (Heidegger, 2010: 120). Das *Leerlassende*, die *Gleichgültigkeit* und die *Hingehaltenheit* sind notwendige und gemeinsam hinreichende Bedingungen für die Langeweile und hängen direkt miteinander zusammen (Heidegger, 2010: 161f). Ebenfalls, so lässt sich zeigen, gibt es auch ein Fluchtverhalten vor dieser Stimmung, den *Zeitvertreib* (Heidegger, 2010: 136). Durch den Vertreib der Zeit will das Dasein die inhaltende Zeit beschleunigen und dadurch der Langeweile entgegen.

Wenden wir uns nun zur Veranschaulichung dieser Struktur der ersten Stufe der Langeweile zu. Die weitere Untersuchung folgt der heideggerschen Methode, indem er die zunehmende Tiefe der Langeweile dadurch verdeutlichen will.

Erste Stufe der Langeweile

In der oberflächlichen, alltäglichen Langeweile ist das Dasein von etwas Innerweltlichem gelangweilt. Folglich ist die erste Stufe der Langeweile wiederum eine intentionale Stimmung, die sich auf ein Ding, ein Ereignis etc. als Inhalt richtet (Grosse, 2008: 32). Das Dasein wird von diesem Innerweltlichem *leergelassen*, weil es das Dasein in Ruhe lässt, es wird ein nacktes *Vorhandenes*, etwas Unbrauchbares, Belangloses (Heidegger, 2010: 153). Das Innerweltliche zeichnet sich dadurch aus, dass es nicht spannend ist, dass es dem Dasein nichts bieten kann.²⁰ Das Dasein wird demge-

²⁰ Dies mag beispielsweise ein langweiliger Bahnhof sein, bei welchem man auf einen verspäteten Zug warten muss. Geeignete literarische Beispiele findet man besonders bei Beckett (2015).

mäss sich selbst überlassen, solange es von der Langeweile festgehalten ist (Heidegger, 2010: 138ff).

Das Sich-Langweilen besteht nun darin, dass das Dasein in die Langeweile *herauszappelt* (Grosse, 2008: 32). Es bemerkt, dass die Situation, welche erschlossen wurde, nicht die Fähigkeit besitzt, das Dasein anzuregen und auszufüllen. Die oberflächliche Langeweile ist deshalb nicht mit einem schlichten Warten zu wechseln, da ein Warten oft von einer inneren Spannung begleitet wird (Heidegger, 2010: 141). Diese zwei Phänomene hängen jedoch oft zusammen. Trotzdem ist das Dasein in einer bestimmten Weise *betroffen* und zwar in seinem Bezug zur Zeit.

Sprich, das Dasein langweilt sich *um* seine Zeit. Diese hält sich dem Dasein schlicht hin, sie vergeht zu langsam - sie *zögert* (Heidegger, 2010: 147). Wie ist dies möglich? Laut Heidegger liegt das daran, dass jedes Ding seine Zeit hat (Heidegger, 2010: 159). Dies ist nicht fatalistisch aufzufassen, sondern in dem Sinne, dass das Dasein je schon Pläne entworfen hat. Es hat Zusammenhänge und Prozesse entworfen, die es für sich anwenden will. Hierfür benötigt es gewisse Dinge zu einer bestimmten Zeit. Werden diese Entwürfe also unterbunden oder hinausgezögert, nimmt das Dasein die umliegende Welt als belanglos und unbrauchbar wahr.

Insofern nimmt das Dasein die Langeweile als negative Stimmung auf. Dies erklärt auch die Fluchtbewegung des Zeitvertreibs (Hüsch, 2014: 64). Durch den Zeitvertreib, das *Sich-Ablenken*, das *Beschäftigt-Sein*, versucht es, die zögernde Zeit anzutreiben und damit die Langeweile wegzutreiben (Heidegger, 2010: 140). Die dabei auftauchende Ungeduld ist hingegen nur ein Zeichen eines scheiternden Versuchs, die Zeit anzutreiben (Heidegger, 2010: 146). Da nun das Dasein zumeist schon an das uneigentliche Man verfallen ist, versucht es auch ständig gegen die Stimmung der Langeweile anzukämpfen, indem es sich von einer Tätigkeit zur nächsten sputet (Heidegger, 2010: 143). Die permanente Unterhaltung wird dadurch zum Mittel gegen die Langeweile.²¹ Der Kampf gegen die Langeweile ist also eine Fluchtbewegung, indem versucht wird, ihr vorab auszuweichen, oder beim Aufkommen für sich eine Ablenkung zu finden.

Für die Zusammenfassung der ersten Stufe der Langeweile werde ich sie sogleich auch mit der Furcht vergleichen. Ähnlich wie die Furcht ist die oberflächliche Langeweile eine alltägliche Stimmung. In beiden versteht sich das Dasein durch Innerweltliches

(Sein-bei) und entwirft sich auch entsprechend den fremdbestimmten Entwürfen des Mans. Es versteht sich selbst und die Welt durch jenes, was gerade langweilt oder bedroht, indem die Ursachen der Stimmungen die Pläne des Daseins durchkreuzen können. Auch die oberflächliche Langeweile erweist sich demgemäss als existenziell-ontische Stimmung. Zudem entspricht der geschäftige Zeitvertreib dem Fluchtverhalten der Furcht, bei welchem sich das Dasein von dieser Stimmung ablenken will, indem es dem Langweiligen oder dem Furchtbaren auszuweichen versucht. Hieraus lässt sich nun der Schluss folgern, dass die erste Stufe der Langeweile morphologisch der Furcht in den relevanten Strukturen gleicht und somit auch als uneigentliche Stimmung definiert werden kann.

Zweite Stufe der Langeweile

Da die erste Stufe der Langeweile sowohl in Dauer wie auch in der Intensität variieren kann, zeichnet sich die zweite Stufe durch ein anderes Merkmal aus. Sie ist *tiefer* (Heidegger, 2010: 162). Das bedeutet, sie ist weniger konkret auf äusseres gerichtet und von diesem beeinflusst, sondern sie ist tiefer im Wesen des Daseins selbst verortet (Hüsch, 2014: 71). Betrachten wir nun die Strukturmomente dieser Form der Langeweile.

Das Wovon dieser Langeweile ist nicht mehr intentional bestimmbar, sondern es umfasst die gesamte Situation (Heidegger, 2010: 172).²² Es ist hier nicht mehr möglich, einen einzelnen Gegenstand für die Langeweile verantwortlich zu machen. Daher nennt Heidegger diese Form auch das *Langweilen bei* (Heidegger, 2010: 164). Es ist folglich ein *Sich-langweilen bei* etwas. Das Dasein ist nicht mehr vom äusseren Langweiligen gefangen genommen, sondern die Langeweile entsteht aus dem Dasein selbst. Das Dasein langweilt sich ab sich selber. Insofern ist diese Stufe tiefer als die vorangehende, weil diese Stimmung in einem engeren Verhältnis zum Wesen des Daseins steht (Hüsch, 2014: 72). Dies verändert natürlich auch die Art und Weise, wie das Dasein betroffen ist. Das Sich-langweilen bei etwas ist eine Gleichgültigkeit, die, wie schon erwähnt, aus der eigenen Tiefe des Daseins aufkommt (Grosse, 2008: 32). Im Gegensatz zur oberflächlichen Form der Langeweile sind die einzelnen Dinge nicht *leerlassend*, da sie es nicht vermögen, das Dasein auszufüllen, sondern die Leere bildet sich erst (Heidegger, 2010: 196). Und diese Leere macht sich im Dasein breit. Wie ist diese Leere zu verstehen? Es ist eine Sinnleere, eine

²¹ Vgl. hierzu besonders Breuninger & Schiemann (2015), sowie Han (2018) und Svendsen (2016).

²² Ein Beispiel hierfür wäre ein zwar unterhaltsamer Abend unter Bekannten, welcher jedoch im Nachhinein als langweilig und verschwendet betrachtet wird. Das Dasein lebt an sich selbst vorbei, wie dies bei Camus' Meursault (2014) der Fall ist.

Gleichgültigkeit, die durch den uneigentlichen Gegenwartsbezug evoziert wird. Das Dasein lässt sein Leben dahinplätschern, es tut, was das Man eben so unternimmt, ohne wirklich davon betroffen zu sein (Heidegger, 2010: 180). Es ist eine Form des Sein-bei, die sich eben in jener fahlen Ungestimmtheit (vgl. Kapitel 2) bemerkbar macht. Das Dasein lebt so an seinem eigenem Selbst vorbei und empfindet dadurch eine Leere. Dieses scheinbare Entgleiten des eigenen Selbst, was erst in der dritten Stufe vollständig abhandenkommt, wird aber hier in dieser Stimmung schon angedeutet (Heidegger, 2010: 177).

Dies erklärt nun auch, *worum* sich das eigene Dasein bei diesen Situationen langweilt. Das Dasein wird nicht entlassen aus einer *stehenden Zeit*. Sie lässt uns in der besagten Leere zurück (Heidegger, 2010: 183). Die Zeit steht, weil sich das Man an die Gegenwart richtet. Die Zeit wird so zu einem *gedehnten Jetzt*, das die gesamte Situation umschliesst (Hüsch, 2014: 75). Indem nun das Dasein die Möglichkeit verpasst, eigentlich zu werden und im Man dahinplätschert, erkennt es, wie es die eigene Zeit *verschwendet* (Heidegger, 2010: 194).

Wenn das Dasein seine eigene Zeit verschwendet und sich dabei selbst langweilig wird, so zeigt es sich, dass der Zeitvertreib und das Wovor der zweiten Langeweile das Selbe sind (Heidegger, 2010: 170). Sprich, das Dasein wird selbst zum eigenen Zeitvertreib. Der Zeitvertreib ist in dieser Form kühl und ruhig und unterscheidet sich zum hastigen, zappelndem Unterhalten der oberflächlichen Langeweile. Dies liegt an dem Umstand, dass der Zeitvertreib in dieser Form unauffälliger ist. Denn das Dasein, das sein eigener Zeitvertreib wird, bemerkt dies nur in einer Vorahnung oder rückblickend, da es in dieser Form der Langeweile weiterhin dem Man verfallen ist.

Zusammenfassend lässt sich also sagen, dass die zweite Stufe der Langeweile aus dem Dasein selbst heraus entsteht. Es langweilt sich im Man, weil es bemerkt, wie es seine eigene Zeit verschwendet. Der Zeitvertreib wird zugleich zum Wovor der Langeweile: das Dasein selbst. Für die vorliegende Untersuchung ist diese Zwischenstufe der Langeweile relevant, da sie den Übergang zur dritten, tiefen Langeweile darstellt. In dieser zweiten Stufe meldet sich zum ersten Mal die *Schwere* (Lastcharakter) des Daseins (Heidegger, 2010: 193). Es bemerkt in dieser Stimmung sein eigenes Verfallen an das Man und die damit einhergehende Zeit, die dem Dasein abhandenkommt. In dieser Zwischenstufe steht das Dasein am Scheideweg, ob es sich in die tiefe Langeweile stürzt, oder ob es weiter Ablenkung am Innerweltlichen sucht (Pocai, 1996: 194).

Dritte Stufe der Langeweile

Es wird hier nicht möglich sein, auf alle Aspekte der tiefen Langeweile einzugehen. Ich beschränke mich deshalb auf die Strukturmomente, die auch bei den anderen Stimmungen angeführt wurden und beziehe mich anschliessend noch auf den Bezug zur Eigentlichkeit und weise auf die Parallelen zur Angst hin.

Wie zu Beginn dieses Kapitels schon erwähnt wurde, ist die tiefe Langeweile eine Grundstimmung. Sie bildet das Hintergrundrauschen der modernen menschlichen Existenz (Safranski, 2015: 220). Dennoch ist sie keine alltägliche Stimmung. Anders als die anderen Formen der Langeweile ist diese Stufe weitaus stiller und privater. Sie ist in diesem Sinne auch tiefer als die anderen Stufen, weil sie das Dasein direkt in seiner eigenen *Zeitlichkeit* betrifft (Heidegger, 2010: 201). Daraus lässt sich nun eine Parabel folgern, in welcher das Dasein von der oberflächlichen Langeweile zur tiefen Langeweile hinuntergezogen werden kann (Kisiel in Rentsch, 2015: 261). Auch hier, so zeigt sich, geht Heidegger vom ontisch alltäglichen und nächsten zum ontologisch grundlegendsten. Folgen wir nun den Strukturmomenten.

Das Wovor der tiefen Langeweile ist vollkommen unbestimmt. Nicht einmal mehr das Dasein langweilt sich bei etwas, sondern es kommt zum Zustand des *es ist einem langweilig*²³ (Heidegger, 2010: 202f).²³ Das völlig Unbestimmte und Unbekannte erweist sich hier als das gesamte Seiende im Ganzen, und zudem auch das Nichts (Heidegger, 2010: 207). Also wird nicht nur innerweltliches Seiendes oder das Dasein selbst langweilig, sondern wiederum das In-der-Welt sein. Nicht mehr einzelne Gegenstände oder eine Situation haben dem Dasein nichts zu bieten, sondern die Möglichkeit selbst, Entwürfe zu produzieren, entfällt dem Dasein. Selbst das Dasein verliert in dieser Stimmung den Bezug zu sich selbst und wird zu einem *Es* - einem indifferenten Niemand (Heidegger, 2010: 203). Dieses indifferente Niemand ist jedoch nicht einfach nichts oder mit dem unpersönlichen *Man* zu verwechseln, sondern das Dasein ist in der tiefen Langeweile vereinzelt, hat jedoch sein eigenes Selbst verloren. Es ist dem Dasein im Vergleich zur zweiten Stufe gänzlich abhanden gekommen und hat dies nun auch erschlossen.

Betroffen ist das Dasein nun, indem sein Selbst stehen gelassen wurde. Das *Es* wird allem gegenüber gleichgültig, alles erstarrt und bedeutet nichts mehr (Heidegger, 2010: 207). Nun erschöpft sich auch der deutsche Begriff ‚Lange-Weile‘. Das eigene Weilen, die Existenz,

²³ Wahrscheinlich wurde Heidegger hierbei von Lichtenbergs ‚es denkt‘ (2005) inspiriert.

wird dem Dasein zu lange. Sie dauert zu lange, weil das Dasein nichts mit seiner Existenz anzufangen weiss (Heidegger, 2010: 228f). Pocai schliesst hieraus, dass das Dasein sich dabei in einer Vorstufe zur pathologischen Melancholie befindet. In der pathologischen Melancholie wiederholt sich diese Betroffenheit in einer stets wiederkehrenden Zukunft und lässt dadurch das Dasein verzweifeln (Pocai, 1996: 232). Auch wenn die Verzweiflung mit der tiefen Langeweile zusammenhängen kann, sind sie nicht miteinander zu verwechseln (Heidegger, 2010: 211). Man muss die tiefe Langeweile eher als Überdruß verstehen (Safranski, 2015: 183). Heidegger identifiziert hier jedoch einen Prozess, der das Dasein aus dieser Langeweile wieder entheben kann. Währendem in der ersten Stufe der Langeweile das Dasein seine eigene Stimmung versucht zu überschreien und es in der zweiten Stufe nicht anhören will, so wird es in dieser tiefen Langeweile zum anhören gezwungen. Der *Gewissensruf* (vgl. Kapitel 3) macht sich zwangsläufig bemerkbar (Hüsch, 2014: 78). Indem sich die fremdbestimmten Entwürfe der Welt und des Selbst dem Dasein *entsagen*, so weisen sie zugleich auf die brachliegenden Möglichkeiten des Daseins hin (Heidegger, 2010: 212). Das Dasein befindet sich hier vor dem Tor, um eigentlich zu werden. Bevor darauf weiter eingegangen wird, soll noch kurz das Worum in der Relation zur Zeit dargestellt werden. Während die Entwürfe zum Selbst und zur Welt dem Dasein entgleiten, *bannt* die Zeit das Dasein (Heidegger, 2010: 221). Die Zeit darf hier wiederum nicht im herkömmlichen Sinne verstanden werden, denn sie wird bei Heidegger existenzial-phänomenologisch verstanden. Dementsprechend stehen nicht die Uhren still, oder die physikalische Zeit verläuft grundlos langsamer, sondern diese *Zeitigung* des Daseins befindet sich ausserhalb dieser Verständnisse. Das Dasein (mit den Strukturmomenten) ist in dieser Zeit gebannt - die Zeit wird *steinern* und erlebt es so (Hüsch, 2014: 83). Das Dasein hat demzufolge keinen Bezug mehr zur Zeit. Durch den Umstand, dass die Zeit nun bannt und die Möglichkeiten brachliegen lässt, weist sie zugleich auf die Möglichkeiten zum Selbstsein hin, sie verweist dadurch auf eine *eigentliche* Zeit. In dem Moment, in welchem diese eigentliche Zeit, in welcher das Dasein seine Existenz selbst verwaltet, angezeigt wird, dass es nun sein Leben selbst zu führen hat, diesen Moment nennt Heidegger den *Augenblick* (Heidegger, 2010: 224). Der Augenblick ist somit der Blick des Daseins auf seine eigene *Entschlossenheit*. Durch die Wahl zur Eigentlichkeit vermag es nun das Dasein sich selbst und die Welt auf seine eigentliche Art und Weise zu

entwerfen. Diese Entschlossenheit bricht den vorangegangenen Zeitbann und erlöst das Dasein auch aus der Stimmung der tiefen Langeweile (Heidegger, 2010: 226).

Aus dem bisher gesagten wird nun verständlich, dass das Dasein im Moment der bannenden Zeit über keine Möglichkeit verfügt, sich die Zeit zu vertreiben. Ein Zeitvertreib ist daher versperrt (Heidegger, 2010: 202): Da die tiefe Langeweile trotzdem eine negative Empfindung ist, versucht das Dasein dieser Stimmung präventiv zu entgehen, indem sie ans Innerweltliche und ins Man flüchtet (Hüsch, 2014: 64). Um also die tiefe Langeweile schlummern zu lassen, sie überhaupt nicht wach werden zu lassen, wird die geschäftige Unterhaltung als Gegenmittel verwendet. Dadurch kann auch die oberflächliche Langeweile als Gegenpart, als Verfallsform der tiefen Langeweile verstanden werden (Pocai, 1996: 197).

Um die tiefe Langeweile zusammenzufassen, werde ich nun nochmals auf den beschriebenen Prozess der Selbstwerdung (vgl. Kapitel 3) eingehen. In der ersten Phase der tiefen Langeweile verliert das Dasein die Bezüge zu sich Selbst und zur Welt (Heidegger, 2010: 221). Das gesamte Seiende wird in Frage gestellt und verschwindet in einem Nebel der Gleichgültigkeit. Es wird einem langweilig. Hierin findet man eine dreifache Negativität; das Dasein verliert den Bezug zu sich selbst, die Welt wird ihm nichtig und die Bezüge verlieren an Bedeutung (Safranski, 2015: 226). Dies sind die drei Strukturmomente, welche das Dasein überhaupt ausmachen (Selbst, Welt, In-Sein), wie sie im Kapitel 2 vorgestellt wurden. In diesem Moment, in welchem Selbst, Welt und In-Sein belanglos werden, bannt die Zeit das Dasein. In diesem Bann, liegen sämtliche Möglichkeiten brach und aus dieser Ruhe und Leere meldet sich das Gewissen. Das Dasein muss in diesem Moment auf das Gewissen hören. Das Gewissen ruft nun das Dasein zur Eigentlichkeit auf. Durch den Entschluss bricht das Dasein die bannende Zeit und beschreitet den eigentlichen Weg der Selbstwerdung (Heidegger, 2010: 215). Es beginnt, eigene Pläne zu entwerfen, und übernimmt die Verantwortung der eigenen Existenz. In diesem *Augenblick* wird folglich das Selbst erschaffen (Safranski, 2015: 223). Man kann deshalb die tiefe Langeweile als *Umschlag* in die Eigentlichkeit verstehen (Grosse, 2008: 39).²⁴

Es sollte nun offensichtlich werden, dass die Angst

²⁴ In der Diskussion zur tiefen Langeweile hat sich die Frage aufgedrängt, ob es solch eine Stimmung überhaupt gibt (vgl. Hüsch, 2014: 79). Auch Heidegger verzichtet hier auf ein ausführliches Beispiel. Ich denke jedoch, dass diese Stimmung bei Roquentin (Sartre, 2013) auftaucht.

und die tiefe Langeweile morphologisch ähnlich sind. Erstens sind beide Befindlichkeiten unangenehm und beunruhigen das Dasein in existenzialer Weise (Hüsch, 2014: 81). Zweitens verweisen beide auf eine Leere, in welcher der Gewissensruf hörbar wird (Safranski, 2015: 203). Dementsprechend erfolgt auch drittens ein Umschlag in die Eigentlichkeit, der in dem oben beschriebenen Prozess stattfindet (Safranski, 2015: 226). Beide Stimmungen sind daher nicht alltägliche Stimmungen, sondern die Möglichkeit dazu, das Dasein eigentlich werden zu lassen. Zum Schluss lässt sich auch viertens ableiten, dass die Verfallsformen, die uneigentlichen Stimmungen (Furcht und oberflächliche Langeweile), aus diesen Grundstimmungen heraus erfolgen (Heidegger, 2010: 235). Das Dasein versucht die Grundstimmungen nicht wachwerden zu lassen und flieht daher an Innerweltliches und ins Man. In der alltäglichen Seinsweise kommen daher zumeist nur die uneigentlichen Formen auf. Hier wiederum eine Tabelle zur Veranschaulichung der relevanten Strukturmomenten. Die zweite Stufe wurde für den Vergleich mit Furcht und Angst ausgelassen.

	1. Stufe Langeweile	3. Stufe Langeweile
Wovor	Innerweltliches	In-der-Welt-sein
Betroffenheit	Leergelassenheit im Sein-bei	Existenziale Gleichgültigkeit
Worum	Zögernde Zeit	Zeitlichkeit des Daseins
Seinsweise	uneigentlich	eigentlich

Grafik 6: Morphologische Ähnlichkeit der Langeweile (vgl. Kapitel 4)

6. Schlussteil

Konklusion

Was folgt aus dem Bisherigen für die zwei Seinsweisen *eigentlich/uneigentlich*? Es wurde gezeigt, dass es alltägliche und nicht alltägliche Stimmungen gibt. Zu letzteren gehören die Angst und die tiefe Langeweile. Diese zwei Befindlichkeiten sind Grundstimmungen, die tief im Wesen des Daseins schlummern. Es lohnt sich, so Heidegger, diese Stimmungen wach werden zu lassen und auf ihren Ruf zu hören. Das Wahrnehmen der Leere - oder das erscheinende Nichts - ist kein Indikator für ein misslingendes Leben, sondern ermöglicht erst die Selbstwerdung (Pocai, 1996: 230). Denn durch diesen Prozess öffnen diese ausgezeichneten Stimmungen das Tor zur Eigentlichkeit. Wird das Dasein eigentlich, so kann es das verfallene und zerstreute Selbst finden und verwirklichen. Es ist von da an in der Lage, seine Zeit selbst zu schaffen und auszufüllen. Dementsprechend gilt es auch, sich nicht

ständig mit Institutionen zu beschäftigen, die jene Grundstimmungen unterdrücken und dem Dasein den Prozess der Selbstwerdung versperren (Heidegger, 2010: 268). Dass diese Selbstwerdung nicht nur einer philosophischen Elite zugeschrieben ist, sondern fundamentalontologisch zum Sein des Daseins gehört, wird dadurch ersichtlich, dass Heidegger ‚*können*‘ nicht in einem aussagelogischen Verständnis auffasst, sondern im Sinne von ‚*sein können*‘ (Hüsch, 2014: 140). Alle, die ihre Existenz selbst in die Hand nehmen wollen, können dies. Daher ist es keine schier unlösbare Aufgabe, vor welche das Dasein gestellt wird. Es geht in der Entschlossenheit nicht darum, dieses oder jenes zu sein, sondern nur sich selbst (Kierkegaard, 2014: 728).

Wenden wir uns nun der Arbeitshypothese aus Kapitel 1 nochmals zu. Vorerst wurde gezeigt, dass die oberflächliche Langeweile und die Furcht als uneigentliche Stimmungen, währendem die tiefe Langeweile und die Angst als eigentliche Stimmungen zu verstehen sind. Dies wird durch die morphologische Ähnlichkeit begründet. Daraus lässt sich jedoch die Hypothese noch

nicht bestätigen. Hierzu sind einige Bemerkungen von Heidegger hilfreich. Er weist darauf hin, dass bei der Langeweile die Grenzen der einzelnen Formen ineinander fließen (Heidegger, 2010: 165). Das bedeutet, dass sich das Dasein nicht zu einem

bestimmten Zeitpunkt in der ersten Stufe der Langeweile befindet und es zu einem anderen Zeitpunkt zur dritten hinüber schwappt. Die drei Formen bilden eine dynamische Parabel (vgl. Kapitel 5 und Heidegger, 2010: 235). Das Dasein kann beispielsweise von der ersten Stufe über die zweite hin zur dritten Stufe hinuntergezogen werden. Ebenfalls kann es vorkommen, dass das Dasein die erschlossene Stimmung mit einer anderen verwechselt (Heidegger, 2010: 205). Dies ist insofern interessant, als dass es bedeutet, dass epistemische Fehler passieren können. Das Dasein befindet sich möglicherweise kurz vor der tiefen Langeweile und glaubt sich aber in der ersten Stufe zu befinden. Von hier aus kann es wahrscheinlich wieder ins Man fliehen und verpasst dadurch den Gewissensruf. Es würden sich daher auch epistemische Probleme für eine strikte Trennung von Eigentlichkeit und Uneigentlichkeit entwickeln (Hüsch, 2014: 170).

Dies sind selbstverständlich noch keine zwingenden Gründe, anzunehmen, die Grenzen von Eigentlichkeit

und Uneigentlichkeit seien auch fließend. Man könnte hier einwenden, dass das Dasein eigentlich ist, sobald es fehlerlos erschlossen hat, sich in der dritten Stufe der Langeweile zu befinden. Hierauf lässt sich jedoch antworten, dass aber sogar die dritte Stufe selbst problematisch ist. Da die dritte Stufe der Langeweile einen *Umschlag* vollführt, kann beim gelangweilten *Es* (vgl. Kapitel 5) noch nicht von einem eigentlichen Dasein gesprochen werden (Pocai, 1996: 203). Dieses ‚es ist einem langweilig‘ ist aber auch kein rein verfallenes Dasein mehr, hat sich jedoch noch nicht im Augenblick für sein Selbst entschlossen. Pocai spricht hier von einem *voreigentlichen Dasein* (Pocai, 1996: 211). Folglich würde es sich auch bei der morphologisch ähnlichen Angst so verhalten. Verbinden wir nun diesen Prozess, der schrittweise in der dritten Stufe der Langeweile abläuft, mit der vorher erwähnten Dynamik und den epistemischen Problemen, wird es sehr wahrscheinlich, dass es Zwischenstufen von Eigentlichkeit und Uneigentlichkeit und einem dynamischen Fluss dazwischen geben muss. Die Hypothese lässt sich daher nicht zwingend, jedoch mit grosser Wahrscheinlichkeit bestätigen.

Kritik

An der vorliegenden Arbeit können einige Punkte bemängelt werden. Ich möchte hier auf jene Punkte eingehen, die für die Konsistenz der Arbeit wichtig sind und daher für weitere Forschungen zu beachten wären. Für dies strukturiere ich die Kritikpunkte in Heidegger-Diskurs interne und externe Probleme. Es sollen zuerst die Heidegger-Diskurs internen Punkte vorgestellt werden.

Intern kann erstens bemängelt werden, dass nicht auf alle Begriffe, die in Heideggers Befindlichkeitstheorie vorkommen, eingegangen wurde. Termini wie ‚*Zeitlichkeit*‘, ‚*Mächtigkeit*‘, ‚*Durchsichtigkeit*‘ und ‚*Wahrheit*‘ wurden hier absichtlich ausgelassen, da der Rahmen dieser Arbeit dies nicht zulässt. Eine Begrenzung auf das Wesentliche war hier nötig. Dennoch würden diese Begriffe wohl neues Licht auf die Abhandlung werfen. Zweitens wurden ebenso diverse Quellen vernachlässigt, da die jeweiligen Interpretationen wiederum mehr Platz erfordern würden. Interessant wären, hier aber absichtlich nicht in die Arbeit eingewoben, die Ansichten von Adorno (2015), Arendt (2016), Geier (2017), Lévinas (2008) und Rentsch (2003). Weiter existieren noch unzählige andere Quellen, auf die hier nicht eingegangen werden konnte. Das Verständnis der heideggerschen Philosophie könnte durch eine weitere Auseinandersetzung mit diesen Quellen nur

profitieren. Drittens stellt sich die Frage, ob die Werke von Heidegger, die für diese Arbeit verwendet wurden, untereinander konsistent sind. Es kann aber angenommen werden, dass Heideggers Denken zwischen 1926 und 1933 zur selben Phase gehören. Eine ausführliche Diskussion dazu findet sich in Pocai (1996) und Rentsch (Thomä in Rentsch, 2015).

Formale und nicht Heidegger-Diskurs interne Kritik, betreffen mehr seine Philosophie als die vorliegende Arbeit. Starke Mängel am Grundstoff wirken sich jedoch auch auf die vorliegende Arbeit und zukünftige Forschungen aus. Erstens wird bemängelt, dass Heideggers Befindlichkeitstheorie in sich nicht konsistent ist und schwere inhaltliche Fehler aufweist (vgl. Pocai, 1996). Ein Beispiel hierfür findet sich bei der Fussnote Nummer 18 der vorliegenden Arbeit. Zweitens stellt sich die Frage, ob die Begriffe ‚*Eigentlichkeit*‘ und ‚*Uneigentlichkeit*‘ nur in dieser Form verwendet werden können und müssen (vgl. hierzu Pocai, 1996: 211 und Hüsch, 2014: 133). Eine weitere Untersuchung hierzu wäre durchaus sinnvoll. Die nächsten vier Punkte betreffen primär die Stimmungen, so wie sie von Heidegger beschrieben wurden. Drittens also, so kann vorgebracht werden, ist Heidegger bei der Unterscheidung von intentionalen (uneigentlichen) Stimmungen und nichtintentionalen (eigentlichen) Stimmungen unvorsichtig. Denn es scheint eher sinnvoll zu sein, intentionale Stimmungen als *Emotionen* zu definieren (vgl. Breuninger & Schiemann, 2015 und bes. Ekman, 2011). Emotionen verhalten sich anders als Stimmungen und stehen wahrscheinlich auch in einem anderen Verhältnis zu Stimmungen, als dass sie schlicht deren Verfallsformen darstellen. Diese Ansichten haben sich erst nach Heidegger in der zweiten Hälfte des vergangenen Jahrhunderts entwickelt und dadurch an Bedeutung gewonnen. Viertens stellt sich die Frage nach der adäquaten Beschreibung der Langeweile. Wahrscheinlich ist die Langeweile eine historisch kontingente und schwer greifbare Stimmung, welche soziologische Einflüsse und die Thematik der *Willensschwäche* auch berücksichtigen müsste (Grosse, 2008: 17 und 140, sowie Bieri, 2016). Fünftens drängt sich auch die Frage auf, ob die tiefe *Langeweile* überhaupt das philosophische und existenziale Potenzial besitzt, um das Dasein in die Eigentlichkeit zu führen (vgl. hierzu Grosse, 2008: 34f und Cioran, 2013). Der letzte Punkt, der hier noch erwähnt werden muss, behandelt die Frage, inwiefern Heideggers Befindlichkeitstheorie mit seiner politischen Agenda zusammenfällt. Die kämpferische, vereinzelte Entschlossenheit beschwört jenen Kriegerpathos, der auch für den Nationalsozialismus von

Bedeutung war (vgl. hierzu Han, 2018 und ausführlich Pocaï, 1996). Insofern könnte bemängelt werden, dass Heidegger hier die Stimmungen phänomenologisch nicht adäquat beschrieben hat. Da die Phänomenologie von einer unvoreingenommenen Perspektive ausgeht - das sich an ihm selbst zeigende (Heidegger, 2006: 28) -, müsste hier eine strikte Trennung der eigenen politischen Agenda und dem Phänomen an sich gezogen werden. All die hier erwähnten Punkte müssten für ein besseres Verständnis und für den Erfolg dieser Arbeit weiter abgeklärt werden.

Zusammenfassung

Trotz der vorgebrachten Kritikpunkte lassen sich aus dieser Arbeit fruchtbare Einsichten gewinnen. Die vorliegende Arbeit hat sich mit Heideggers Existenzialontologie beschäftigt und ist der Frage nachgegangen, ob Eigentlichkeit und Uneigentlichkeit exklusiv voneinander getrennt sind, oder ob sie dynamische Grenzen aufweisen. Hierzu wurde Heideggers Befindlichkeitstheorie untersucht. Es wurde aufgezeigt, dass Furcht (uneigentlich) und Angst (eigentlich) der ersten und dritten Stufe der Langeweile morphologisch ähnlich sind. Daraus lässt sich schliessen, dass die oberflächliche Langeweile eine uneigentliche Stimmung und die tiefe Langeweile eine eigentliche Stimmung ist. Da nun die Stufen der Langeweile selbst dynamisch ineinanderfließen und in der letzten Stufe noch ein Umschlag stattfindet, kann daraus die Hypothese bekräftigt werden. Dass es graduelle Abstufungen zwischen Eigentlichkeit und Uneigentlichkeit gibt, ist somit sehr wahrscheinlich. Auch wenn die oben erwähnten Kritikpunkte die heideggersche Philosophie in Frage stellen, so lassen sich aus der hier gewonnenen Einsicht dennoch fruchtbare Elemente sowohl für die Heideggerforschung, wie auch für die praktische und Existenzial-Philosophie allgemein gewinnen. Das hier dargestellte Ergebnis wirft ein neues Licht auf die bisherigen Forschungen und entwickelt so neue Wege zu weiteren Arbeiten und verbessert das Verständnis des heideggerschen Denkens. Demgemäss wird es möglich, die Philosophie Heideggers besser zu verstehen, als er seine Philosophie selbst in seinen Werken vermitteln konnte. Dies war auch das Ziel seiner eigenen Tätigkeit (Heidegger, 2010: 232). Und wie kann man jemanden besser verstehen, als allein durch die zuhandenen Quellen? In dem man auf mögliche neue Einsichten und Fehler hinweist (Heidegger, 2010: 232),²⁵ kurz; indem man seine Philosophie überwindet.

²⁵ Vgl. hierzu auch Hampe (2014).

Grafikverzeichnis

Sämtliche Grafiken dieser Arbeit wurden selbst erstellt und anhand der angegebenen Quellen interpretiert.

Literatur

- Adorno, Theodor W.: Jargon der Eigentlichkeit, Zur deutschen Ideologie, 2015, 18. Auflage, Suhrkamp Taschenbuchverlag, Frankfurt a. M.
- Arendt, Hannah: Vita activa, oder vom tätigen Leben, 2016, 18. Auflage, Piper Verlag GmbH, München.
- Beckett, Samuel: Drei Stücke, Warten auf Godot, Endspiel, Glückliche Tage, 2015, 6. Auflage, Suhrkamp Taschenbuchverlag, Frankfurt a. M.
- Bieri, Peter: Das Handwerk der Freiheit, Über die Entdeckung des eigenen Willens, 2016, 8. Auflage, Carl Hanser Verlag, München.
- Brandom, Robert: Heidegger's Categories in Being and Time, 1983, The Monist, S. 387-409, Oxford University Press, Oxford.
- Breuninger, Renate und Schiemann, Gregor (Hrsg.): Langeweile, Ein philosophisches Lesebuch, 2015, 1. Auflage, Campus Verlag GmbH, Frankfurt a. M.
- Camus, Albert: Der Fremde, 2014, 69. Auflage, Rowohlt Taschenbuch Verlag, Reinbek.
- Camus, Albert: Der Mythos des Sisyphos, 2013, 15. Auflage, Rowohlt Taschenbuch Verlag, Reinbek.
- Christensen, Carleton B.: What are the Categories in *Sein und Zeit*? Brandom on Heidegger on *Zuhandenheit*, 2007, European Journal of Philosophy, S. 159-185, Blackwell Publishing Ltd., Oxford.
- Cioran, Emile M.: Der Absturz in die Zeit, 2013, 6. Auflage, J. G. Cotta'sche Buchhandlung Nachfolger GmbH, Berlin.
- Ekman, Paul: Gefühle lesen, Wie Sie Emotionen erkennen und richtig interpretieren, 2011, 2. Auflage, Spektrum Akademischer Verlag, Heidelberg.
- Foucault, Michel: Die Ordnung des Diskurses, Mit einem Essay von Ralf Konersmann, 2014, 13. Auflage, Fischer Taschenbuchverlag, Frankfurt a. M.
- Geier, Manfred: Wittgenstein und Heidegger, Die letzten Philosophen, 2017, 1. Auflage, Rowohlt Verlag GmbH, Reinbek.
- Goethe, Johann Wolfgang: Faust, Der Tragödie erster Teil, 2000, Philipp Reclam jun. GmbH + Co. KG, Stuttgart.
- Grosse, Jürgen, Philosophie der Langeweile, 2008, 1. Auflage, J.B. Metzler Verlag, Stuttgart.
- Gumbrecht, Hans U.: Präsenz, 2016, 2. Auflage, Suhrkamp Taschenbuch Verlag, Berlin.
- Hampe, Michael: Die Lehren der Philosophie, Eine Kritik, 2014, 2. Auflage, Suhrkamp Verlag, Berlin.
- Han, Byung-Chul: Gute Unterhaltung, 2018, 1. Auflage, MSB Mathes & Seitz Berlin Verlagsgesellschaft mbH, Berlin.
- Haugeland, John: Reading Brandom reading Heidegger, 2005, European Journal of Philosophy, S. 421-428, Blackwell Publishing Ltd., Oxford.
- Heidegger, Martin: Der Ursprung des Kunstwerkes, 2012, 1. Auflage, Vittorio Klostermann, Frankfurt a. M.
- Heidegger, Martin: Die Grundbegriffe der Metaphysik, Welt - Endlichkeit - Einsamkeit, 2010, 2. Auflage, Vittorio Klostermann, Frankfurt a. M.
- Heidegger, Martin: Sein und Zeit, 2006, 19. Auflage, Max Niemeyer Verlag, Tübingen.
- Heidegger, Martin: Was ist Metaphysik?, 2007, 16. Auflage, Vittorio Klostermann, Frankfurt a. M.
- Hüsch, Sebastian: Langeweile bei Heidegger und Kierkegaard, Zum Verhältnis philosophischer und literarischer Darstellung, 2014, 1. Auflage, Narr Francke Attempto Verlag GmbH + Co. KG, Tübingen.
- Jung, Matthias, Hermeneutik, Zur Einführung, 2012, 4. Auflage, Junius Verlag GmbH, Hamburg.
- Kafka, Franz: Brief an den Vater, 2008, 1. Auflage, Suhrkamp Basisbibliothek, Frankfurt a. M.
- Kafka, Franz: Der Prozess, 2012, 6. Auflage, Suhrkamp Taschenbuchverlag, Frankfurt a. M.

- Kafka, Franz: Sämtliche Erzählungen, 2007, 1. Auflage, Anaconda Verlag GmbH, Köln.
- Kierkegaard, Sören: Entweder - Oder, Teil I und II, 2014, 12. Auflage, Deutscher Taschenbuchverlag GmbH & Co. KG, München.
- Kierkegaard, Sören: Die Krankheit zum Tode, Furcht und Zittern, Die Wiederholung, Der Begriff der Angst, 2017, 7. Auflage, Deutscher Taschenbuchverlag GmbH & Co. KG, München.
- Lévinas, Emmanuel: Ethik und Unendliches, 2008, 4. Auflage, Passagen Verlag GmbH, Wien.
- Lichtenberg, Georg Christoph: Sudelbücher, 2005, 1. Auflage, 3 bändige Ausgabe, Deutscher Taschenbuchverlag GmbH & Co. KG, München.
- Luckner, Andreas: Martin Heidegger >>Sein und Zeit<<, 2007, 2. Auflage, Verlag Ferdinand Schöningh, Paderborn.
- Mugerauer, Roland: Klartext >>Sein und Zeit<<, Heideggers Hauptwerk dechiffriert, 2015, 1. Auflage, Tectum Verlag, Marburg.
- Poci, Romano: Heideggers Theorie der Befindlichkeit, Sein Denken zwischen 1927 und 1933, 1996, Alber Verlag, München.
- Rentsch, Thomas: Heidegger und Wittgenstein, 2003, 1. Auflage, J. G. Cotta'sche Buchhandlung Nachfolger GmbH, Stuttgart.
- Rentsch, Thomas (Hrsg.): Martin Heidegger: Sein und Zeit, 2015, 3. Auflage, Walter de Gruyter GmbH, Berlin.
- Safranski, Rüdiger: Ein Meister aus Deutschland, 2015, 9. Auflage, Fischer Taschenbuchverlag, Frankfurt a. M.
- Safranski, Rüdiger: Zeit, Was sie mit uns macht und was wir aus ihr machen, 2017, 1. Auflage, Fischer Taschenbuchverlag, Frankfurt a. M.
- Sartre, Jean-Paul: Der Ekel, 2013, 56. Auflage, Rowohlt Taschenbuchverlag, Reinbek.
- Shakespeare, William: Hamlet, 2013, 3. Auflage, Fischer Taschenbuchverlag, Frankfurt a. M.
- Svendsen, Lars: Philosophie der Einsamkeit, 2016, 1. Auflage, Berlin University Press in der Verlagshaus Römerweg GmbH, Wiesbaden.

Omar Ibrahim (26) studiert Philosophie und Sozialanthropologie im Master. Er interessiert sich für Phänomenologie, Diskursanalyse, Existenzialismus und insbesondere für dialogische Praxis.