GRÜNDE ALS DEFAULTS

JOHN F. HORTY
University of Maryland

Inhalt

[Bitte, wenn ohne Schaden möglich, die Schrift „Variable“ verwenden, die auf meinem Rechner hier nicht kommt. Sie wird aber vom File aufgerufen – zumeist die einzelnen Buchstaben „W“, „D“ und „S“. H.Rott]

[Seitenzahlen bitte nach Typesetting einfügen!]

1 EINLEITUNG

Ein Großteil der gegenwärtigen Literatur über Gründe konzentriert sich auf einen bekannten Bereich von Fragestellungen, der beispielsweise das Verhältnis zwischen Gründen und Motivation, Wünschen und Werten, die Frage nach Internalismus versus Externalismus in einer Theorie von Gründen, oder die Objektivität von Gründen umfasst. Diese Abhandlung beschäftigt sich mit einer anderen, orthogonalen Menge von Fragen: Was sind Gründe, und wie stützen sie Handlungen oder Konklusionen? Gegeben eine Sammlung von individuellen Gründen, die möglicherweise konfligierende Handlungen oder Konklusionen stützen: Wie können wir bestimmen, welche Handlungsweise oder welche Konklusion durch die Sammlung als Ganzes gestützt wird? Was ist der Mechanismus der Stützung? Die Antwort, die ich vorschlage, ist, dass Gründe durch Defaults[1] bestimmt werden und dass sie in Übereinstimmung mit der Logik des Default-Schließens Konklusionen stützen. Das Ziel dieses Textes ist es, diese Antwort verständlich zu machen und zu entwickeln.

Obwohl es keine einzelne Theorie gibt, von der wir jetzt sagen können, dass sie die korrekte Logik für Default-Schließen bereitstellt, beginne ich mit einer Beschreibung dessen, was mir eine besonders nützliche Weise der Entwicklung einer solchen Theorie zu sein scheint. Diese Logik wird hier nur so detailliert präsentiert, wie es nötig ist, um zu zeigen, dass hier wirklich eine konkrete Theorie am Werk ist, und um eine Vorstellung von der Form dieser Theorie sowie der Fragen, die damit in Verbindung stehen, zu geben. Nach der Darstellung dieser Default-Logik zeige ich, wie sie ausgearbeitet werden kann, um mit bestimmten Fragen umgehen zu können, die mit der Entwicklung einer stabileren Theorie von Gründen in Verbindung stehen. Ich konzentriere mich hier auf zwei solcher Probleme: erstens Situationen, in denen die Prioritätsrelationen zwischen Gründen oder Defaults selbst durch Default-Schließen gebildet zu werden scheinen; zweitens die Behandlung des „Schlagens“ von Gründen durch Unterminierung und die Behandlung von ausschließenden Gründen. Schließlich zeige ich in einer Anwendung, wie der resultierende Ansatz eine Thematik innerhalb der Theorie von Gründen erhellen kann, welcher in letzter Zeit viel Aufmerksamkeit geschenkt wurde: Jonathan Dancys interessantes und einflussreiches Argument, das von einem Gründe-Holismus zu einer Form des extremen Partikularismus in der Moraltheorie führt.

[1] Das englische Substantiv „default“, welches im Deutschen mit „Vorannahme“ oder „Standardregel“ nur unzureichend wiedergegeben wäre, bleibt in diesem Aufsatz unübersetzt. (Anm. des Übersetzers)

2 Eine Theorie des Default-Schließens

2.1 Default-Theorien und Szenarien

Wir nehmen als Hintergrund eine gewöhnliche aussagenlogische Sprache mit den üblichen Junktoren $\supset$, $\wedge$, $\vee$ und $\neg$ und mit **T** als einer speziellen Konstante, die das Verum repräsentiert. Das Zeichen $\vdash$ steht für gewöhnliche logische Folgerung.

Vor diesem Hintergrund wollen wir jetzt mit einem Standardbeispiel beginnen, das als „Tweety-Dreieck“ bekannt ist.[2] Wenn einem Akteur nur gesagt wird, dass Tweety ein Vogel ist, wäre es natürlich für den Akteur zu schließen, dass Tweety fliegen kann. Unsere alltäglichen Schlussfolgerungen scheinen durch einen allgemeinen Default geleitet zu werden, gemäß welchem Vögel in der Regel fliegen können; und nach dem Ansatz, der hier vorgeschlagen wird, ist es dieser auf Tweety angewandte Default, der einen Grund für die Folgerung bereitstellt, dass Tweety fliegen kann. Aber nehmen wir an, dass dem Akteur zusätzlich gesagt wird, dass Tweety ein Pinguin ist. Es gibt auch einen Default, gemäß welchem Pinguine in der Regel nicht fliegen können, der jetzt einen Grund für eine konfligierende Konklusion bietet. Weil der Default für Pinguine stärker ist als derjenige für Vögel, ist es natürlich anzunehmen, dass der erste dieser Gründe vom zweiten geschlagen wird, so dass der Akteur sein anfängliches Urteil überdenken und statt dessen schließen sollte, dass Tweety nicht fliegen kann.

Wenn A und B Formeln aus der Hintergrundsprache sind, repräsentiert $A \rightarrow B$ die *Default-Regel*, die uns erlaubt, per Default auf B zu schließen, wann immer festgestellt ist, dass A. Zur Illustration: Wenn V für die Aussage steht, dass Tweety ein Vogel ist, und F für die Aussage steht, dass Tweety fliegen kann, dann ist $V \rightarrow F$ die Regel, die uns erlaubt per Default zu schließen, dass Tweety fliegen kann, wenn festgestellt ist, dass Tweety ein Vogel ist. Dieser spezielle Default kann als Tweety-Instanz des allgemeinen Defaults

$$Vogel(x) \rightarrow Fliegt(x)$$

interpretiert werden, der uns sagt, dass Vögel in der Regel fliegen können (um von diesem allgemeinen Default zur speziellen Instantiierung $V \rightarrow F$ zu kommen, stellen Sie sich V als eine Abkürzung der Aussage *Vogel(Tweety)* und F als Abkürzung für *Fliegt(Tweety)* vor).

Wir gehen von zwei Funktionen – *Prämisse* und *Konklusion* – aus, die die Prämissen und Konklusionen von Default-Regeln herausgreifen: Wenn δ zum Beispiel der Default $A \rightarrow B$ ist, dann ist *Prämisse*(δ) die Aussage A, und *Konklusion*(δ) ist die Aussage B. Die zweite dieser

[2] Es wird aufgrund seiner dreieckigen Gestalt so genannt, wenn es in einem Vererbungs-Netzwerk, einer grafischen Repräsentation von Default-Relationen zwischen Klassen von Entitäten, dargestellt wird; Vgl. mein (1994) für einen Überblick über nichtmonotones Schließen durch Vererbungsrelationen.

Funktionen wird auf offensichtliche Weise von individuellen Defaults auf Mengen von Defaults erweitert, so dass wir, wenn D eine Menge von Defaults ist,

$$Konklusion(D) = \{Konklusion(\delta) : \delta \in D\}$$

als die Menge ihrer Konklusionen bekommen.

Wie wir gesehen haben, haben einige Defaults eine größere Stärke oder eine höhere Priorität als andere; einige Gründe sind besser als andere. Um diese Information zu repräsentieren, führen wir eine Ordnungsrelation < auf der Menge von Defaults ein, wobei $\delta < \delta'$ bedeutet, dass der Default δ' eine höhere Priorität besitzt als δ. Wir nehmen an, dass diese Ordnung der Prioritäten von Defaults transitiv ($\delta < \delta'$ und $\delta' < \delta''$ implizieren $\delta < \delta''$), und auch irreflexiv ($\delta < \delta$ ist immer falsch) ist; solch eine Ordnungsrelation nennt man eine *strikte partielle Ordnung*.

Die Prioritätsrelationen zwischen Defaults können verschiedene Quellen haben. Im Tweety-Dreieck hat die Priorität des Defaults für Pinguine vor dem Default für Vögel zum Beispiel mit Spezifität zu tun: ein Pinguin ist eine spezifische Art von Vogel, und deshalb haben Informationen über Pinguine im Besonderen Vorrang vor Informationen über Vögel im Allgemeinen. Aber es gibt auch Prioritätsrelationen, die nichts mit Spezifität zu tun haben. Verlässlichkeit ist eine andere Quelle. Sowohl der Wetterbericht als auch die Arthritis in meinem rechten Knie geben halbwegs verlässliche Vorhersagen für kommenden Niederschlag, aber der Wetterbericht ist verlässlicher. Und wenn wir uns von epistemischen zu praktischen Gründen bewegen, dann ist Autorität noch eine weitere Quelle für Prioritätsrelationen. Europäische Gesetze sind typischerweise Landesgesetzen übergeordnet, und neuere Gerichtsentscheidungen haben höhere Autorität als ältere. Unmittelbar erteilte Befehle haben Vorrang vor langfristig geltenden Befehlen, und Befehle des Obersts sind Befehlen des Majors übergeordnet.

Wir beginnen diesen Abschnitt, indem wir den Spezialfall betrachten, in dem alle Prioritätsrelationen zwischen Defaults im Voraus festgesetzt sind, so dass wir weder die Quelle dieser Prioritätsrelationen, noch die Art und Weise, wie sie festgestellt werden, beachten müssen, sondern nur ihre Wirkung auf die Konklusionen, die man durch Default-Schließen zieht. Formal definieren wir, wenn D eine Menge von Defaults mit der strikten partiellen Ordnung < ist und wenn W dazu eine Menge von gewöhnlichen Formeln ist, eine *Default-Theorie mit fester Priorität* als eine Struktur der Form $\langle W, D, < \rangle$. Eine solche Struktur – eine Menge von gewöhnlichen Informationen zusammen mit einer geordneten Menge von Defaults – repräsentiert das, was dem Akteur als Basis für seine Schlussfolgerungen zu Anfang gegeben ist.

Der größte Teil der Forschung in nichtmonotoner Logik ist durch die epistemische Interpretation von Defaults motiviert und konzentriert sich so auf das Problem der Charakterisierung der Mengen von Meinungen oder Überzeugungen,[3] die durch Default-Theorien gestützt werden. Die Defaults selbst werden als Regeln betrachtet, nach welchen die aus einer Menge von Formeln ableitbaren Überzeugungen über ihre klassischen Konsequenzen hinaus erweitert werden. Aus diesem Grund werden die Überzeugungsmengen, die durch Defaults gestützt werden, häufig als *Extensionen* bezeichnet. Wir werden uns hier allerdings nicht auf die Extensionen selbst konzentrieren, sondern auf *Szenarien*, wobei ein Szenario, das auf einer Default-Theorie $\langle W, D, < \rangle$ basiert, einfach als eine bestimmte Teilmenge S der Menge der Defaults D definiert ist, die in der Theorie enthalten ist. Intuitiv soll ein Szenario die Menge derjenigen Defaults repräsentieren, die vom Akteur an irgendeiner Stelle des Schlussfolgerungsprozesses akzeptiert werden und hinreichende Unterstützung für seine Konklusionen bereitstellen.

Das Konzept eines Szenarios besitzt sowohl für die epistemische als auch für die praktische Lesart von Default-Regeln eine natürliche Interpretation. In der epistemischen Lesart kann man den Akteur, der ein Szenario wählt, so verstehen, dass er diejenigen Defaults wählt, die er benutzen wird, um seine Anfangsinformationen zu einer vollen Menge von Glaubensinhalten auszuweiten: Wenn S ein auf $\langle W, D, < \rangle$ basierendes Szenario ist, können wir die von diesem Szenario generierte Überzeugungsmenge definieren als die Menge von Formeln, die aus seinen Konklusionen und den Anfangsinformationen des Akteurs ableitbar sind – das heißt, die aus $W \cup \mathit{Konklusion}(S)$ ableitbare Menge von Formeln. In der praktischen Lesart, in der Defaults eher als Handlungsempfehlungen denn als Stützungen von Propositionen verstanden werden, kann die Menge von Konklusionen der Defaults, die zu einem Szenario S gehören – das heißt, die Menge $\mathit{Konklusion}(S)$ –, als die Menge der Handlungen verstanden werden, für die sich der Akteur entschieden hat.

Unsere erste Aufgabe ist es, *geeignete Szenarien* zu definieren, wie wir sie nennen werden. Dies sind diejenigen Mengen von Defaults, die letztlich von einem idealen schlussfolgernden Akteur auf der Basis der Informationen akzeptiert werden könnten, die in einer geordneten Default-Theorie enthalten sind. Mit dieser Idee im Hinterkopf können die Extensionen von Default-Theorien – ideale Überzeugungsmengen – als diejenigen Mengen von Überzeugungen definiert werden, die von geeigneten Szenarien generiert werden. Von einer

[3] Das pluralisierbare englische Substantiv „belief“ wird in diesem Aufsatz wahlweise mit „Überzeugung“ oder „Meinung“ oder „Überzeugung oder Meinung“ wiedergegeben, obgleich „Überzeugung“ tendenziell zu stark und „Meinung“ tendenziell zu schwach ist. Ein genaues Äquivalent zu „belief“ gibt es im Deutschen nicht. (Anm. des Übersetzers)

praktischen Perspektive aus gesehen, kann ein geeignetes Szenario als eines verstanden werden, das eine Menge von Handlungen spezifiziert, für deren Ausführung sich ein rationales Individuum entscheiden könnte.

2.2 Bindende Defaults

Wir beginnen mit dem Begriff eines bindenden Defaults. Wenn Defaults Gründe liefern, dann sollen bindende Defaults diejenigen repräsentieren, die im Kontext eines bestimmten Szenarios *gute* Gründe liefern.
Der Begriff eines bindenden Defaults wird mittels dreier vorläufiger Ideen definiert, denen wir uns zunächst zuwenden – dem Auslösen, dem in-Konflikt-Stehen und dem Schlagen.
Natürlich ist nicht jeder Default in jedem Szenario anwendbar. Beispielsweise liefert der Default, dass Vögel fliegen, überhaupt keinen Grund für einen Akteur zu schließen, dass Tweety fliegt, wenn sich der Akteur nicht bereits auf die Proposition festgelegt hat, dass Tweety ein Vogel ist. Die ausgelösten Defaults, nämlich die, die in einem bestimmten Szenario anwendbar sind, sind einfach diejenigen, deren Prämissen durch dieses Szenario impliziert werden – das heißt diejenigen Defaults, deren Prämissen aus den Anfangsinformationen des Akteurs und den Konklusionen der Defaults, die der Akteur bereits akzeptiert hat, folgen.

Definition 1 (Ausgelöste Defaults) Wenn S ein Szenario ist, das auf der *Default-Theorie mit fester Priorität* $\langle W, D, < \rangle$ basiert, dann sind die Defaults aus D, die in S *ausgelöst* werden, diejenigen, die zu der Menge
$Ausgelöst_{W,D,<}(S) = \{\delta \in D : W \cup Konklusion(S) \vdash Prämisse(\delta)\}$
gehören.

Um dies zu illustrieren, sollen V, F, und W jeweils für die Sätze stehen, dass Tweety ein Vogel ist, dass Tweety fliegt und dass Tweety Flügel hat; und δ_1 und δ_2 sollen für die Defaults $V \rightarrow F$ und $F \rightarrow W$ stehen, also für Tweety-Instantiierungen der allgemeinen Defaults, dass Vögel fliegen und dass fliegende Tiere tendenzierte Flügel haben. Stellen Sie sich vor, dass ein Akteur die geordnete Default-Theorie $\langle W, D, < \rangle$ als Anfangsinformation besitzt, wobei $W = \{V\}$, $D = \{\delta_1, \delta_2\}$ und die Ordnung $<$ leer ist; und nehmen Sie an, dass der Akteur noch keinen der Defaults aus D akzeptiert hat, sodass sein Anfangsszenario einfach $S_0 = \varnothing$ ist. Wir bekommen dann $Ausgelöst_{W,D,<}(S_0) = \{\delta_1\}$, so dass in diesem Anfangsszenario nur δ_1

ausgelöst wird, was dem Akteur einen Grund für seine Konklusion bereitstellt, nämlich den Satz *F*.

Diese Diskussion über ausgelöste Defaults, die dem Akteur Gründe bereitstellen, führt zu einer terminologischen Frage: Sollten diese Gründe mit den Defaults selbst oder mit Sätzen identifiziert werden? Nehmen wir wie in unserem Beispiel an, dass die Hintergrundtheorie des Akteurs den Default $V \rightarrow F$ enthält, eine Tweety-Instantiierung des allgemeinen Defaults, dass Vögel fliegen, zusammen mit *V*, dem Satz, dass Tweety ein Vogel ist, so dass der Default ausgelöst wird. In diesem Fall scheint es klar zu sein, dass der Akteur einen Grund hat zu schlieißen, dass Tweety fliegt. Aber wie genau sollte dieser Grund reifiziert werden? Sollte er mit dem Default $V \rightarrow F$ selbst oder mit dem Satz *V* identifiziert werden?

Diese Frage ist, wie viele Fragen zur Reifikation, ein wenig künstlich. Offensichtlich sind sowohl der Default als auch der Satz daran beteiligt, den Akteur mit einem Grund für die Konklusion auszustatten, dass Tweety fliegt. Der Default hätte keine Auswirkung, wenn er nicht durch eine Tatsache, einen wahren Satz, ausgelöst werden würde; die Tatsache würde nichts als ein zufälliges Merkmal der Situation sein, wenn sie nicht einen Default auslösen würde. Wenn es um Reifikation geht, könnte die Grund-Relation also in jede Richtung projiziert werden, auf Defaults oder auf Sätze, und die Wahl ist weitgehend willkürlich. Dennoch scheint es unserem Alltagsgebrauch am ehesten zu entsprechen, wenn wir Gründe als Sätze reifizieren. Der vorliegende Artikel wird sich deshalb auf eine Analyse stützen, wonach *Gründe mit den Prämissen ausgelöster Defaults identifiziert werden*; und wir werden sagen, dass diese ausgelösten Defaults nicht selbst Gründe *sind*, sondern bestimmte Sätze – ihre Prämissen – als Gründe ihrer Konklusionen *anführen*. Zur Illustration: In unserem Beispiel werden wir sagen dass *V*, die Tatsache, dass Tweety ein Vogel ist, ein Grund ist für die Konklusion, dass Tweety fliegt, und dass dieser Grund durch den Default $V \rightarrow F$ bereitgestellt wird.

Auslösung ist eine notwendige Bedingung, die ein Default erfüllen muss, damit er als in einem Szenario bindend klassifiziert wird, aber sie ist nicht hinreichend. Sogar wenn ein Default ausgelöst wird, könnte er alles in allem nicht bindend sein; zwei weitere Aspekte des Szenarios könnten störend einwirken.

Der erste Aspekt ist einfach zu beschreiben. Ein Default wird, sogar wenn er ausgelöst wird, nicht als in einem Szenario bindend klassifiziert, wenn dieser Default einen Konflikt erzeugt – das heißt, wenn sich der Akteur bereits auf die Negation seiner Konklusion festgelegt hat.

Definition 2 (Widersprochene Defaults) Wenn S ein Szenario ist, das auf der *Default-Theorie mit fester Priorität* $\langle W, D, < \rangle$ basiert, dann sind die Defaults aus D, die in S *konfligieren*, diejenigen, die zu der Menge

$Widersprochen_{W,D,<}(S) = \{\delta \in D : W \cup Konklusion(S) \vdash \neg Konklusion(\delta)\}$

gehören.

Die intuitive Kraft dieser Restriktion kann mit einem weiteren Standardbeispiel verdeutlicht werden, das als „Nixon-Diamant" bekannt ist.[4] Q, R und P stehen jeweils für die Propositionen, dass Nixon ein Quäker ist, dass Nixon ein Republikaner ist und dass Nixon ein Pazifist ist. δ_1 und δ_2 repräsentieren die Defaults $Q \rightarrow P$ und $R \rightarrow \neg P$, Instantiierungen der Verallgemeinerungen, dass Quäker tendenziell Pazifisten und Republikaner tendenziell keine Pazifisten sind. Stellen Sie sich vor, dass die Anfangsinformationen des Akteurs von der Theorie $\langle W, D, < \rangle$ bereitgestellt werden, wobei $W = \{Q, R\}$, $D = \{\delta_1, \delta_2\}$ und die Ordnung $<$ wieder leer ist; und nehmen sie noch einmal an, dass der Akteur noch keinen dieser beiden Defaults akzeptiert hat, so dass sein Anfangsszenario $S_0 = \varnothing$ ist.
In dieser Situation haben wir $Ausgelöst_{W,D,<}(S_0) = \{\delta_1, \delta_2\}$; der Default δ_1 stellt einen Grund für die Konklusion P bereit, und der Default δ_2 stellt einen Grund für die Konklusion $\neg P$ bereit. Obwohl diese beiden Defaults konfligierende Konklusionen stützen, erzeugt keiner von beiden im Anfangsszenario einen Konflikt: $Widersprochen_{W,D,<}(S_0) = \varnothing$. Der Akteur muss deshalb einen Weg finden, mit den konfligierenden Gründen umzugehen, die sich in seinem epistemischen Zustand zeigen. Nehmen Sie nun an, dass sich der Akteur aus irgendwelchen Gründen für einen dieser beiden Defaults entscheidet – sagen wir δ_1 mit der Konklusion P – und sich so in das neue Szenario $S_1 = \{\delta_1\}$ begibt. In diesem neuen Szenario wird dem anderen Default nun widersprochen werden: $Widersprochen_{W,D,<}(S_1) = \{\delta_2\}$. Vom Standpunkt des neuen Szenarios aus kann der Grund, der von δ_2 bereitgestellt wird, nicht weiterhin als guter Grund klassifiziert werden, weil sich der Akteur bereits für einen Default entschieden hat, der einen Grund für eine konfligierende Konklusion bereitstellt.
Die zweite Restriktion, die für den Begriff eines bindenden Defaults maßgeblich ist, besagt, dass ein Default, sogar dann, wenn er ausgelöst wird, nicht als bindend klassifiziert werden kann, wenn er geschlagen wird. Auch wenn der Begriff eines geschlagenen Defaults um einiges schwieriger zu definieren ist als der eines widersprochenen Defaults, ist die Grundidee recht einfach: Ein Akteur sollte einen Default angesichts eines stärkeren Defaults, der eine konfligierende Konklusion stützt, nicht akzeptieren.

[4] Wiederum, weil seine Beschreibung als ein Vererbungsnetzwerk die Gestalt eines Diamanten besitzt.

Diese Idee kann verdeutlicht werden, indem wir zu unserem ursprünglichen „Tweety-Dreieck“ zurückkehren, wobei P, V und F die Propositionen repräsentieren, dass Tweety ein Pinguin ist, dass Tweety ein Vogel ist und dass Tweety fliegt. δ_1 und δ_2 sollen die Defaults $V \rightarrow F$ und $P \rightarrow \neg F$ sein – Instanzen der allgemeinen Regeln, dass Vögel fliegen und dass Pinguine nicht fliegen. Stellen Sie sich vor, dass der Akteur mit der Theorie $\langle W, D, < \rangle$ als Anfangsinformation ausgestattet ist, wobei $W = \{P, V\}$, $D = \{\delta_1, \delta_2\}$ und nun $\delta_1 < \delta_2$. Der Default für Pinguine hat eine höhere Priorität als der Default für Vögel. Und nehmen wir erneut an, dass der Akteur noch keinen dieser beiden Defaults akzeptiert hat, so dass sein Anfangsszenario $S_0 = \varnothing$ ist.

In dieser Situation haben wir wieder $Ausgelöst_{W,D,<}(S_0) = \{\delta_1, \delta_2\}$; der Default δ_1 bietet einen Grund für die Konklusion F, während der Default δ_2 einen Grund für die Konklusion $\neg F$ bereitstellt. Und wir haben wieder $Widersprochen_{W,D,<}(S_0) = \varnothing$; keinem dieser Defaults wird selbst widersprochen. Trotzdem scheint es nicht, dass der Akteur wie im vorhergehenden „Nixon-Diamant“ frei sein sollte, den Konflikt durch willkürliche Auswahl zu beheben. Hier scheint es intuitiv angemessen zu sagen, dass der Default δ_1, der die Konklusion F stützt, vom stärkeren Default δ_2 geschlagen wird, weil dieser Default ebenfalls ausgelöst wird und weil er die konfligierende Konklusion $\neg F$ stützt.

Durch dieses Beispiel motiviert ist es natürlich, eine Definition vorzuschlagen, nach der ein Default in einem Szenario geschlagen wird, wenn dieses Szenario einen stärkeren Default mit einer konfligierenden Konklusion auslöst.

Definition 3 (Geschlagene Defaults: vorläufige Definition) Wenn S ein Szenario ist, das auf der *Default-Theorie mit fester Priorität* $\langle W, D, < \rangle$ basiert, dann sind die Defaults aus D, die in *S geschlagen werden*, diejenigen, die zu der folgenden Menge gehören:

$Geschlagen_{W,D,<}(S) = \{\delta \in D$: es gibt einen Default $\delta' \in Ausgelöst_{W,D,<}(S)$, so dass

(1) $\delta < \delta'$,

(2) $Konklusion(\delta') \vdash \neg Konklusion(\delta)\}$.

Diese vorläufige Definition liefert im Fall des „Tweety-Dreiecks“ das richtige Ergebnis: Aus der Definition folgt $Geschlagen_{W,D,<}(S_0) = \{\delta_1\}$, weil $\delta_2 \in Ausgelöst_{W,D,<}(S_0)$ und weil sowohl (1) $\delta_1 < \delta_2$ als auch (2) $Konklusion(\delta_2) \vdash \neg Konklusion(\delta_1)$ gilt. Tatsächlich liefert diese vorläufige Definition in allen Beispielen, die hier besprochen werden sollen, die richtigen Ergebnisse, und wir können uns in diesem Text guten Gewissens auf diese Definition als unsere offizielle Definition verlassen. Allerdings ist die vorläufige Definition nicht

durchgängig präzise und führt in bestimmten komplizierteren Fällen zu falschen Resultaten; diejenigen Leser, die an den Feinheiten der Formulierung einer einwandfreien Definition interessiert sind, können eine vollständige Diskussion in Horty (2007) finden.
Nachdem wir den Begriff des Schlagens zur Verfügung haben, können wir die Menge der Defaults, die in einem bestimmten Szenario als bindend klassifiziert werden, recht einfach als diejenigen Defaults definieren, die in diesem Szenario ausgelöst werden, aber weder widersprochen noch geschlagen sind.

Definition 4 (Bindende Defaults) Wenn *S* ein Szenario ist, das auf der *Default-Theorie mit fester Priorität* $\langle W, D, < \rangle$ basiert, dann sind die Defaults aus *D*, die in *S bindend sind*, diejenigen, die zu der folgenden Menge gehören:

$$Bindend_{W,D,<}(S) = \{\delta \in D : \delta \in Ausgelöst_{W,D,<}(S),$$
$$\delta \notin Widersprochen_{W,D,<}(S),$$
$$\delta \notin Geschlagen_{W,D,<}(S)\}.$$

Weil die bindenden Defaults dazu gedacht sind, im Kontext eines bestimmten Szenarios die guten Gründe zu repräsentieren, ist es natürlich, den Begriff eines stabilen Szenarios so zu bestimmen, dass ein stabiles Szenario alle und nur diejenigen Defaults enthält, die in diesem Kontext bindend sind. Formal: Wenn *S* ein Szenario ist, das auf der Default-Theorie $\langle W, D, < \rangle$ basiert, können wir nur in folgendem Fall sagen, dass *S* ein *stabiles Szenario* ist:
$S = Bindend_{W,D,<}(S)$.
Ein Akteur, der eine Menge von Defaults akzeptiert hat, die ein stabiles Szenario bilden, ist in einer beneidenswerten Position. Solch ein Akteur hat bereits genau diejenigen Defaults akzeptiert, die er als gute Gründe liefernd anerkennt – und zwar im Kontext der Defaults, die er akzeptiert. Der Akteur hat deshalb keinen Anreiz, einen der Defaults, die er bereits akzeptiert hat, aufzugeben oder irgendwelche anderen Defaults zu akzeptieren.

2.3 Schlussfolgern mittels geeigneter Szenarien

Erinnern wir uns, dass es unser Ziel ist, geeignete Szenarien zu charakterisieren – diejenigen Mengen von Defaults, die ein ideal rationaler Akteur auf der Basis der Anfangsinformationen, die in einer Default-Theorie enthalten sind, akzeptieren könnte. Dürfen wir die geeigneten Szenarien einfach mit den stabilen Szenarien identifizieren? Ich biete wieder zwei Antworten auf diese Frage an, eine vorläufige und eine endgültige Antwort.

Die vorläufige Antwort ist, dass wir in der großen Mehrzahl normaler Fälle, einschließlich denen, die in diesem Text betrachtet werden sollen, die geeigneten Szenarien tatsächlich mit den stabilen Szenarien identifizieren können. Diese vorläufige Antwort kann zu einer vorläufigen Definition verdichtet werden.

Definition 5 (Geeignete Szenarien: vorläufige Definition) $\langle W, D, < \rangle$ sei eine geordnete Default-Theorie und S ein Szenario. Dann ist S ein *geeignetes Szenario*, das auf $\langle W, D, < \rangle$ basiert, genau dann, wenn $S = Bindend_{W,D,<}(S)$.

Unglücklicherweise ist aber die endgültige Antwort, dass es auch bestimmte sonderbare Theorien gibt, die stabile Szenarien erlauben, die nicht wirklich als geeignet klassifiziert werden können – das heißt als Szenarien, die ein ideal Schlussfolgernder akzeptieren würde. Weil uns diese sonderbaren Fälle hier nicht betreffen, werden wir uns in diesem Text wieder auf die vorläufige Definition als unsere offizielle Definition berufen. Diejenigen Leser, die an einer korrekten Definition interessiert sind, finden eine Diskussion darüber in Horty (2007).
Der Begriff eines geeigneten Szenarios kann veranschaulicht werden, indem wir zum „Tweety-Dreieck“ zurückkehren. Wie der Leser selbst nachprüfen kann, ist das geeignete Szenario, das auf dieser Default-Theorie basiert, $S_1 = \{\delta_2\}$, wobei δ_2 der Default $P \rightarrow \neg F$ ist, sodass $Konklusion(S_1) = \{\neg F\}$ gilt. In diesem Fall assoziiert der hier vorgeschlagene Ansatz mit der Default-Theorie ein einziges geeignetes Szenario, das die intuitiv korrekte Konklusion stützt, dass Tweety nicht fliegen kann. In anderen Fällen jedoch definiert dieser Ansatz – wie viele andere zu nichtmonotonem Schließen – eine Relation zwischen Default-Theorien und ihren geeigneten Szenarien, die aus einer konventionelleren logischen Perspektive als eine Anomalie erscheinen könnte: Bestimmte Default-Theorien können mit mehreren geeigneten Szenarien assoziiert werden.[5]
Das kanonische Beispiel einer Default-Theorie mit mehr als einem geeigneten Szenario ist der „Nixon-Diamant“, der zwei davon aufweist: $S_1 = \{\delta_1\}$ und $S_2 = \{\delta_2\}$, wobei δ_1 $Q \rightarrow P$ und δ_2 $R \rightarrow \neg P$ ist, sodass $Konklusion(S_1) = \{P\}$ und $Konklusion(S_2) = \{\neg P\}$ gelten. Worauf sollte der Akteur im Lichte dieser beiden Extensionen schließen, von denen die eine P und die andere $\neg P$ enthält: Ist Nixon Pazifist oder nicht? Allgemeiner: Wenn eine geordnete Default-Theorie mehr als ein geeignetes Szenario erlaubt – wie sollten wir dann die Konsequenzen der Theorie definieren?

[5] Und andere können mit gar keinen geeigneten Szenarien assoziiert werden – ein Faktum, das uns hier nicht weiter beschäftigen muss.

Diese Frage ist irritierend, und noch nicht einmal die Literatur über nichtmonotones Schließen hat sich angemessen damit befasst. Ich habe hier nicht den Raum, die Angelegenheit im Detail zu untersuchen, sondern werde einfach die am häufigsten gebrauchte Antwort übernehmen: Der Akteur sollte einer Konklusion nur in dem Fall zustimmen, dass sie durch jedes geeignete Szenario gestützt wird, das auf der ursprünglichen Default-Theorie basiert. Im Beispiel des „Nixon-Diamanten" würde der Akteur dann weder schließen, dass Nixon ein Pazifist ist, noch, dass der keiner ist, weil weder *P* noch ¬*P* durch beide geeigneten Szenarien gestützt werden. Diese Option wird heutzutage allgemein als *skeptisch* bezeichnet.

3 Ausarbeitung der Theorie

Die zentrale These dieses Artikels ist, dass es nützlich ist, Gründe als durch Defaults bereitgestellt zu denken, aber bis jetzt habe ich kaum mehr als einen Ansatz für das Default-Schließen gegeben. Jetzt möchte ich meine zentrale These stützen, indem ich zeige, wie dieser Ansatz ausgearbeitet werden kann, um zwei Problemen zu begegnen, die sich bei der Entwicklung einer robusteren Theorie von Gründen stellen. Erstens wurde bis jetzt davon ausgegangen, dass die Prioritäten zwischen den Defaults im Voraus feststehen, aber es gibt Fälle, in denen diese Prioritäten selbst erst durch anfechtbares Schließen ermittelt werden müssen. Und zweitens erfasst die hier definierte Auffassung nur eine Form des Schlagens, die üblicherweise „Schlagen durch Widerlegung" genannt wird und in der ein stärkerer Default einen schwächeren Default schlägt, indem er seiner Konklusion widerspricht. Es gibt mindestens eine andere Form – üblicherweise „Schlagen durch Unterminierung" genannt und verwandt mit der Diskussion ausschließender Gründe in der Literatur über praktisches Schließen –, in der ein Default einen anderen nicht dadurch schlägt, dass er seiner Konklusion widerspricht, sondern indem er seine Fähigkeit unterminiert, einen Grund bereitzustellen.

3.1 Default-Theorien mit variabler Priorität

Die Definition

Wir haben uns auf Default-Theorien mit fester Priorität konzentriert, in denen die Prioritätsrelationen zwischen den Default-Regeln im Voraus festgesetzt sind. Aber in Wirklichkeit sind einige der wichtigsten Dinge, über die wir nachdenken und über die wir auf

anfechtbare Weise nachdenken, die Prioritäten zwischen genau denjenigen Defaults, die unser anfechtbares Schließen leiten.

Unsere erste Aufgabe ist es daher zu zeigen, wie diese Art des Schließens in das hier vorgestellte allgemeine Rahmenwerk eingepasst werden kann. Was wir suchen, ist ein Ansatz, in dem unser Schließen wie zuvor von einer Menge von Defaults geleitet wird, die einer Prioritätsordnung unterliegen, aber in dem es jetzt möglich ist, dass die Prioritäten zwischen den Defaults in demselben Prozess des Schließens festgelegt werden, den sie leiten. Obwohl das kompliziert klingen mag – vielleicht bedrohlich kompliziert, vielleicht zirkulär –, es stellt sich heraus, dass die vorliegende Theorie durch die Adaption bekannter Techniken[6] in vier einfachen Schritten so erweitert werden kann, dass sie einen solchen Ansatz bereitstellt.

Der erste Schritt ist, unsere Objektsprache um Ressourcen zu erweitern, die formales Schlussfolgern mit Prioritäten zwischen Defaults ermöglichen: eine neue Menge von Individuenkonstanten, die als Namen von Defaults interpretiert werden, sowie ein Relationssymbol, das Priorität repräsentiert. Der Einfachheit halber werden wir annehmen, dass jede dieser neuen Konstanten die Form d_X, für irgendeinen Index X, hat, und dass jede dieser Konstanten sich auf den Default δ_X bezieht. Und wir werden auch annehmen, dass die Objektsprache jetzt das Relationssymbol $\prec$ aufweist, das Priorität zwischen Defaults repräsentiert.

Um diese Notation zu illustrieren, nehmen Sie an, dass δ_1 der Default $A \rightarrow B$ ist, dass δ_2 der Default $C \rightarrow \neg B$ ist, und dass δ_3 der Default $\mathbf{T} \rightarrow d_1 \prec d_2$ ist, wobei sich im Einklang mit unserer Konvention d_1 und d_2 auf die Defaults δ_1 und δ_2 beziehen. Was δ_3 dann aussagt ist, per Default, dass δ_2 eine höhere Priorität besitzt als δ_1. Als Resultat würden wir erwarten, dass, wenn beide dieser Defaults ausgelöst werden, – das heißt, wenn A und C der Fall sind –, der Default δ_1 von δ_2 geschlagen wird, weil die beiden Defaults konfligierende Konklusionen haben. Weil δ_3 selbst ein Default ist, ist natürlich auch die Information, die über die Priorität zwischen δ_1 und δ_2 bereitgestellt wird, anfechtbar und kann ebenfalls übertrumpft werden.

Der zweite Schritt ist, unsere Aufmerksamkeit weg von Strukturen der Form $\langle W, D, < \rangle$ – das heißt, von Default-Theorien mit fester Priorität – und hin auf Strukturen der Form $\langle W, D \rangle$ zu lenken, die eine Menge W von gewöhnlichen Formeln, sowie eine Menge D von Defaults, aber keine im Voraus festgelegte Prioritätsrelation zwischen den Defaults enthält. Stattdessen können W und D Anfangsinformationen über die Prioritätsrelationen zwischen den Defaults enthalten, und dann kommt man zu Konklusionen über diese Prioritäten wie zu allen anderen

[6] Die fundamentale Idee, die den hier zu beschreibenden Techniken zugrunde liegt, wurde zuerst von Gordon (1993) eingeführt. Die Idee wurde dann von einer Anzahl von Leuten weiterentwickelt und ausgebaut, vor allem von Brewka (1994, 1996), sowie von Prakken und Sartor (1995, 1996).

Konklusionen durch anfechtbares Schließen. Weil Konklusionen über die Prioritäten zwischen Defaults selbst, je nachdem welche Defaults der Akteur akzeptiert, variieren können, sind diese neuen Strukturen als *Default-Theorien mit variabler Priorität* bekannt. Wir fordern als Teil dieser Definition, dass die Menge *W* einer Default-Theorie mit variabler Priorität jede mögliche Instantiierung des Irreflexivitäts- und des Transitivitätsschemas

$\neg(d \prec d)$,
$(d \prec d' \wedge d' \prec d'') \supset d \prec d''$

enthalten muss, in dem die Variablen durch Namen der zu *D* gehörenden Defaults ersetzt werden.
Nehmen Sie jetzt an, dass der Akteur ein Szenario akzeptiert, das diese neuen Prioritätsaussagen enthält. Der dritte Schritt ist dann, die im Szenario des Akteurs implizit enthaltene Prioritätsordnung zu einer expliziten Ordnung zu erheben, die in unserem metasprachlichen Schließen benutzt werden kann. Dies wird auf die denkbar einfachste Art und Weise gemacht. Wenn S ein Szenario ist, das auf der Default-Theorie $\langle W, D \rangle$ basiert, verstehen wir jetzt die Aussage $\delta <_S \delta'$ so, dass der Default *δ' gemäß Szenario S eine höhere Priorität besitzt als δ*, wobei diese Relation folgendermaßen definiert ist:

$\delta <_S \delta'$ genau dann, wenn $W \cup \mathit{Konklusion}(S) \vdash d \prec d'$.

Das heißt: δ' hat gemäß Szenario S eine höhere Priorität als δ genau dann, wenn die Konklusionen der Defaults, die zu diesem Szenario gehören, zusammen mit der normalen Information der Hintergrundtheorie die Formel $d \prec d'$ implizieren, die besagt, dass δ' eine höhere Priorität besitzt als δ. Weil die normale Information aus W alle Instantiierungen von Transitivität und Irreflexivität enthält, ist die abgeleitete Prioritätsrelation $<_S$ mit Garantie eine strikte partielle Ordnung.
Der vierte und letzte Schritt ist die Definition eines geeigneten Szenarios für Default-Theorien mit variabler Priorität. Dies wird bewerkstelligt, indem unsere frühere Definition fruchtbar gemacht wird, die die Bedingungen festsetzt, unter welchen S als geeignetes Szenario für eine Default-Theorie mit fester Priorität $\langle W, D, < \rangle$ gilt, wobei < jede strikte partielle Ordnung der Defaults sein kann. Mit dieser früheren Definition können wir jetzt stipulieren, dass S nur für den Fall ein geeignetes Szenario für die Default-Theorie mit variabler Priorität $\langle W, D \rangle$ darstellt, in dem S ein geeignetes Szenario für die spezielle Default-Theorie mit

fester Priorität $\langle W, D, <_S \rangle$ ist, wobei W und D aus der Default-Theorie mit variabler Priorität übernommen werden, und $<_S$ die Prioritätsrelation ist, die vom Szenario S selbst abgeleitet wird.

Definition 6 (geeignete Szenarien: Default-Theorien mit variabler Priorität)
Sei $\langle W, D \rangle$ eine Default-Theorie mit variabler Priorität und S ein Szenario. S ist ein geeignetes Szenario, das auf $\langle W, D \rangle$ basiert, genau dann, wenn S ein geeignetes Szenario ist, das auf der Default-Theorie mit fester Priorität $\langle W, D, <_S \rangle$ basiert.

Das intuitive Bild sieht so aus: Auf der Suche nach einem geeigneten Szenario kommt der Akteur zu einem Szenario S, das dann Konklusionen über verschiedene Aspekte der Welt impliziert. Darunter sind auch Prioritätsrelationen zwischen den Defaults des Akteurs. Wenn diese abgeleiteten Prioritätsrelationen benutzt werden können, um zu rechtfertigen, dass der Akteur genau dieses ursprüngliche Szenario S akzeptiert, dann ist das Szenario ein geeignetes.

Ein Beispiel

Der hier beschriebene Ansatz kann durch eine Variante des „Nixon-Diamant“ veranschaulicht werden, in der es nützlich ist, nicht die Perspektive einer dritten Partei einzunehmen, die zu entscheiden versucht, ob Nixon ein Pazifist ist oder nicht, sondern stattdessen die praktische Perspektive des jungen Nixon selbst, der zu entscheiden versucht, ob er Pazifist werden soll oder nicht. Nehmen Sie also an, dass Nixon mit der Default-Theorie $\langle W, D \rangle$ konfrontiert ist, wobei W die Formeln Q und R enthält, die Nixon daran erinnern, dass er sowohl Quäker als auch Republikaner ist, und D nur δ_1 und δ_2 enthält, δ_1 ist der Default $Q \rightarrow P$ und δ_2 der Default $R \rightarrow \neg P$. Gegeben unsere neue Perspektive, sollten diese beiden Defaults jetzt als praktische Gründe interpretiert werden: δ_1 sagt Nixon, dass er als Quäker Pazifist werden sollte, während δ_2 aussagt, dass er als Republikaner kein Pazifist werden sollte. Nichts in seiner Anfangstheorie sagt Nixon, wie der Konflikt zwischen diesen beiden Defaults zu lösen ist, und so sieht er sich einem praktischen Dilemma ausgesetzt: Seine Anfangstheorie ergibt zwei geeignete Szenarien, nämlich die bekannten $S_1 = \{\delta_1\}$ und $S_2 = \{\delta_2\}$, die die konfligierenden Konklusionen P und $\neg P$ stützen.
Stellen Sie sich jetzt vor, dass Nixon sich entscheidet, sich mit einigen Autoritäten zu beraten, um sein Dilemma zu lösen. Nehmen wir an, dass er sein Problem zuerst mit einem Ältesten in seiner Quäker-Gemeinde bespricht, der ihm sagt, dass δ_1 Priorität über δ_2 haben sollte, dass er

aber auch mit einem Politiker der Republikaner spricht, der ihm genau das Gegenteil sagt. Der Rat dieser beiden Funktionäre kann repräsentiert werden, indem man die Menge D durch die neuen Defaults δ_3 und δ_4 ergänzt, wobei δ_3 der Default $\mathbf{T} \rightarrow d_2 \prec d_1$ und δ_4 der Default $\mathbf{T} \rightarrow d_1 \prec d_2$ ist. Gegeben unsere praktische Perspektive, sollten diese Defaults nicht als Informationen sondern als Ratschläge interpretiert werden; der Default δ_3 sollte zum Beispiel nicht so interpretiert werden, dass er Nixon einen Hinweis darauf gibt, dass δ_1 tatsächlich höhere Priorität *hat* als δ_2, sondern so, dass er vorschlägt, dass Nixon in seinen Abwägungen mehr Wert auf δ_1 *legen sollte*.[7]

Weil seine gewählten Autoritäten nicht übereinstimmen, hat Nixon sein praktisches Dilemma noch nicht gelöst, das jetzt durch die zwei geeigneten Szenarien $S_3 = \{\delta_1, \delta_3\}$ und $S_4 = \{\delta_2, \delta_4\}$ repräsentiert wird, die erneut konfligierende Handlungsoptionen befürworten. Gemäß Szenario S_1, das die Aussagen $d_2 \prec d_1$ und P stützt, sollte Nixon mehr Wert auf δ_1 als auf das konfligierende δ_2 legen und somit Pazifist werden. Gemäß Szenario S_2, das die Aussagen $d_1 \prec d_2$ und $\neg P$ stützt, sollte Nixon stattdessen mehr Wert auf δ_2 legen und kein Pazifist werden. Was an dieser Theorie besonders interessant ist, ist jedoch nicht, dass sie zwei geeignete Szenarien ergibt, die zwei Handlungsoptionen favorisieren, sondern dass sie *nur* zwei geeignete Szenarien ergibt, die nur zwei Handlungsoptionen favorisieren. Schließlich konfligiert der Default δ_1 mit δ_2, während der Default δ_3 mit δ_4 konfligiert. Wenn es zwei Konflikte gibt, von denen jeder auf zweierlei Art gelöst werden kann kann; warum gibt es dann nicht vier geeignete Szenarien, die vier Handlungsoptionen favorisieren?

Die Antwort ist, dass die zwei Konflikte nicht unabhängig sind. Jede Auflösung des Konflikts zwischen δ_3 und δ_4 legt Nixon auf eine bestimmte Prioritätsordnung zwischen δ_1 und δ_2 fest, die dann die Lösung dieses Konflikts bestimmt. Intuitiv wäre es für Nixon falsch, beispielsweise δ_3 zu akzeptieren, wonach mehr Wert auf δ_1 als auf δ_2 zu legen ist, aber dann trotzdem δ_2 zu akzeptieren und kein Pazifist zu werden. Diese Intuition wird formal erfasst, weil $S_5 = \{\delta_2, \delta_3\}$ – das Szenario, das die Kombination aus δ_3 und δ_2 enthält – zu einer abgeleiteten Prioritätsordnung führt, gemäß der $\delta_2 <_{S5} \delta_1$. Wenn wir unsere gegebene Theorie mit variabler Priorität durch diese abgeleitete Prioritätsordnung ergänzen und die Theorie mit fester Priorität $\langle W, D, <_{S5} \rangle$ erhalten, dann können wir sehen, dass der Default δ_2 im Kontext von S_5 durch den Default δ_1 geschlagen wird, einem stärkeren Default mit einer konfligierenden Konklusion. Das Szenario S_5 ist deshalb kein geeignetes Szenario, das auf der Theorie mit fester Priorität $\langle W, D, <_{S5} \rangle$ basiert, und deshalb kann es gemäß unserer

[7] Die Idee, dass unser Schlussfolgern selbst bestimmt, auf welche Gründe wir mehr Wert legen sollten, wird in Schroeder (2005) diskutiert und verteidigt.

Definition auch kein geeignetes Szenario sein, das auf der ursprünglichen Theorie mit variabler Priorität basiert.

Stellen wir uns schließlich vor, dass Nixon, der sich immer noch einem Konflikt ausgesetzt sieht, weitere Beratung sucht. Vielleicht geht er jetzt zu seinem Vater, der ihm sagt, dass der Rat des Kirchenältesten gegenüber dem des Politikers zu bevorzugen ist. Diese Information kann repräsentiert werden, indem die Regel δ_5 zu der Default-Menge D hinzugefügt wird, wobei δ_5 ist: $\mathbf{T} \rightarrow d_4 \prec d_3$. Mit diesem neuen Default hat Nixon endlich seinen Konflikt gelöst. Wie der Leser selbst bestätigen kann, führt die Theorie nun zu dem einzigen geeigneten Szenario $S_3 = \{\delta_1, \delta_3, \delta_5\}$, das die Konklusionen $d_4 \prec d_3$, $d_2 \prec d_1$ und P stützt und gemäß welchem Nixon δ_3 vor δ_4 und somit δ_1 vor δ_2 favorisieren und somit Pazifist werden sollte.

3.2 Default-Theorien mit Grenzwerten

Die Definition

Wir haben bisher nur eine Form des Schlagens betrachtet – im Allgemeinen Schlagen durch „Widerlegung“ genannt – gemäß der ein Default, der eine Konklusion stützt, von einem stärkeren Default geschlagen wird, der eine konfligierende Konklusion stützt. Es gibt auch eine zweite Form des Schlagens, gemäß der ein Default, der eine Konklusion stützt, von einem anderen Default nicht deshalb, weil dieser eine konfligierende Konklusion stützt, geschlagen wird, sondern weil dieser die Verbindung zwischen Prämisse und Konklusion des ursprünglichen Defaults in Frage stellt. In der Literatur über epistemische Gründe wird diese zweite Form des Schlagens gemeinhin als Schlagen durch „Unterminierung“ bezeichnet und wurde zuerst von John Pollock (1970) aufgezeigt.[8]

Die Unterscheidung zwischen diesen beiden Formen des Schlagens kann durch ein Standardbeispiel veranschaulicht werden. Nehmen Sie an, ein Objekt vor mir sieht rot aus. Dann ist es für mich vernünftig zu schließen, dass es rot ist, und zwar durch die Anwendung eines allgemeinen Defaults, gemäß dem Dinge, die rot aussehen, tendenziell auch rot sind. Aber stellen wir uns zwei Drogen vor. Die Wirkung von Droge Nr. 1 ist, dass rote Dinge blau und blaue Dinge rot aussehen. Im Gegensatz dazu besteht die Wirkung von Droge Nr. 2 darin, dass alles rot aussieht. Wenn das Objekt vor mir nun rot aussieht, ich aber Droge Nr. 1

[8] Einige der frühen Formalismen zur Wissensrepräsentation in künstlicher Intelligenz – wie etwa das von Fahlman (1979) beschriebe NETL System – erlaubten eine Form des Schlagens durch Unterminierung. Aber die Idee ist schnell wieder aus der Literatur über künstliche Intelligenz verschwunden und in diesem Untersuchungsfeld nicht wieder aufgetaucht, bis sie von Autoren, die explizit über Pollocks Arbeit nachgedacht haben, wieder eingeführt wurde.

genommen habe, ist es natürlich, sich auf einen anderen Default zu berufen, der stärker als der ursprüngliche ist und gemäß dem Dinge, die unter Einfluss von Droge Nr. 1 rot aussehen, tendenziell blau – und also nicht rot – sind. Dieser neue Default würde dann den ursprünglichen Default in dem Sinne schlagen, den wir bisher betrachtet haben, nämlich indem er einen stärkeren Grund für eine konfligierende Konklusion liefert. Wenn das Objekt andererseits unter Einfluss von Droge Nr. 2 rot aussieht, dann scheint es abermals so zu sein, dass ich nicht länger berechtigt bin, die Konklusion zu akzeptieren, dass das Objekt rot ist. Aber in diesem Fall wird der ursprüngliche Default nicht auf dieselbe Weise geschlagen. Es gibt keinen stärkeren Grund für den Schluss, dass das Objekt nicht rot ist; stattdessen ist es so, als ob der ursprüngliche Default selbst unterminiert wird und somit überhaupt keinen Grund mehr für seine Konklusion liefert.

Diese zweite Form des Schlagens – oder etwas sehr Ähnliches – wird auch in der Literatur über praktisches Schließen diskutiert, wo sie als Teil des allgemeinen Themas der „ausschließenden" Gründe behandelt wird, wie sie zuerst von Joseph Raz (1975) eingeführt wurden. Raz gibt eine Reihe von Beispielen, um das Konzept zu motivieren, aber wir betrachten hier nur den beispielhaften Fall von Colin, der entscheiden muss, ob er seinen Sohn auf eine Privatschule schickt. Wir müssen uns vorstellen, dass es verschiedene Gründe dafür und dagegen gibt. Auf der einen Seite wird die Schule für eine exzellente Ausbildung von Colins Sohn sorgen und ihm die Gelegenheit bieten, eine abwechslungsreichere Gruppe von Freunden kennen zu lernen. Auf der anderen Seite ist das Schulgeld hoch, und Colin ist besorgt, dass eine Entscheidung, seinen eigenen Sohn auf eine private Schule zu schicken, zu einer Schwächung der Unterstützung für öffentliche Ausbildung im Allgemeinen führen könnte.

Wie auch immer, Raz bittet uns, uns zusätzlich zu diesen gewöhnlichen Gründen dafür und dagegen auch vorzustellen, dass Colin seiner Frau versprochen hat, in allen Entscheidungen über die Ausbildung seines Sohnes nur diejenigen Gründe in Betracht zu ziehen, die direkt die Interessen seines Sohns betreffen. Und dieses Versprechen, so Raz, kann nicht richtigerweise einfach als ein weiterer gewöhnlicher Grund dafür betrachtet werden, seinen Sohn auf die Privatschule zu schicken, wie die Tatsache, dass die Schule für eine gute Ausbildung sorgt. Vielmehr muss das Versprechen als ein Grund von völlig anderer Art betrachtet werden – ein Grund „zweiter Ordnung" dafür, dass all die gewöhnlichen Gründe oder Gründe „erster Ordnung", die nicht die Interessen von Colins Sohn betreffen, aus der Überlegung ausgeschlossen werden. Ebenso wie ich, wenn ich Droge Nr. 2 genommen habe, den Default gemäß dem Dinge, die rot aussehen, tendenziell rot sind, nicht beachten sollte, so sollte

Colins Versprechen ihn dazu bringen, diejenigen Defaults nicht zu beachten, die nicht die Interessen seines Sohns betreffen. Ein ausschließender Grund ist nach dieser Interpretation nichts anderes als ein unterminierender *Defeater* in der praktischen Domäne.
Wie kann man nun diesem Phänomen des unterminierenden oder ausschließenden Schlagens Rechnung tragen? Die Standardpraxis ist es, das Schlagen durch Unterminierung als eine gesonderte und grundlegende Form des Schlagens zu postulieren, die neben dem Begriff des widerlegenden Schlagens zu analysieren ist. Diese Praxis wird vor allem von Pollock verfolgt.[9] Was ich jedoch vorschlagen möchte ist, dass wir, nachdem wir nun die Möglichkeit des Schlussfolgerns mit Prioritäten zwischen Defaults eingeführt haben, das Phänomen des Schlagens durch Unterminierung auf natürlichere Weise analysieren, und zwar als einen Spezialfall einer Prioritätenabstimmung. Die Grundidee ist einfach. Wir postulieren für unsere Prioritätsordnung einen speziellen Wert, sagen wir τ, als *Grenzwert*, der niedrig genug ist, dass wir uns sicher fühlen, wenn wir jeden Default unbeachtet lassen, dessen Priorität nicht über diesem Grenzwert liegt. Ein Default wird dann unterminiert, wenn unser Schließen zu der Konklusion führt, dass seine Priorität unter den Grenzwert fällt.[10]
Um diese Idee zu implementieren, ändern wir in einem ersten Schritt unsere frühere Definition der Auslösung, um die zusätzliche Anforderung zu berücksichtigen, dass ein Default nicht ausgelöst werden kann, wenn er nicht über dem Grenzwert liegt – dass wir also $\tau < \delta$ haben müssen, damit der Default δ ausgelöst wird.

Definition 7 (Ausgelöste Defaults; revidierte Definition) Wenn S ein Szenario ist, das auf der Default-Theorie mit fester Priorität $\langle W, D, < \rangle$ basiert, dann sind die Defaults aus D, die in S *ausgelöst* werden, diejenigen, die zu der Menge
$Ausgelöst_{W,D,<}(S) = \{\delta \in D : \tau < \delta \text{ und } W \cup Konklusion(S) \vdash Prämisse(\delta)\}$
gehören.

[9] Vgl. Pollock (1995) und die Artikel, die darin zitiert werden. Ein Überblick der Arbeit auf diesem Gebiet des nichtmonotonen Schließens, der weitgehend demselben Ansatz folgt, findet sich in Prakken und Vreeswijk (2002).
[10] Die allgemeine Idee, die hier vorgestellt wird, entspricht einer Anregung von Dancy (2004), der die Begriffe von „Verstärkern“ und „Abschwächern“ als Elemente einführt, die die Kraft von Gründen stärken oder schwächen, und der sich dann einen „Deaktivierer“ als ein Element vorstellt, das die Stärke eines Grundes mehr oder weniger vollständig abschwächt – in unserem gegenwärtigen Vokabular ist ein Deaktivierer also etwas, das einen Default, der einen Grund bereitstellt, so stark schwächt, dass er unter den Grenzwert fällt. Die Idee, dass Unterminierung eine graduelle Angelegenheit ist und dass das, was typischerweise als „Unterminierung“ bezeichnet wird, am besten als extremer Fall von Abschwächung der Stärke von Gründen zu analysieren ist, wird in ähnlicher Weise von Schroeder (2011) festgehalten, der diese Idee als „Unterminierungshypothese“ bezeichnet.

Aber natürlich ist diese eine Revision noch nicht genug. Wie die Dinge jetzt stehen, werden überhaupt keine Defaults ausgelöst, weil noch nicht gesichert ist, dass irgendwelche Defaults über dem Grenzwert liegen.

Um zu garantieren, dass die richtigen – und nur die richtigen – Defaults über dem Grenzwert liegen, müssen wir unsere Default-Theorie mit variabler Priorität um einige zusätzliche Informationen ergänzen. Wir führen deshalb den Begriff einer *Default-Theorie mit Grenzwert* als eine Default-Theorie mit variabler Priorität $\langle W, D \rangle$ ein, in der die Menge D von Defaults und die Menge W von Hintergrundtatsachen zwei weitere Bedingungen erfüllen müssen. Bei der Formulierung dieser Bedingungen nehmen wir Bezug auf die Art von Defaults, die wir bisher als *gewöhnliche Defaults* behandelt haben, und wir führen t als sprachliche Konstante ein, die τ entspricht: die Grenzgewichtung.

Die erste Bedingung besagt, dass es für jeden gewöhnlichen Default δ_X, der zu D gehört, auch einen speziellen *Grenzwert-Default* δ^*_X der Form $\mathbf{T} \rightarrow t \prec d_X$ gibt. Die Rolle des Grenzwert-Defaults δ^*_X ist es, uns per Default zu sagen, dass der dem gewöhnlichen Default δ_X zugeschriebene Prioritätswert über dem Grenzwert liegt. Und genau wie jeder gewöhnliche Default δ_X durch einen Term d_X der Objektsprache repräsentiert wird, nehmen wir an, dass jeder Grenzwert-Default δ^*_X von einem Term d^*_X repräsentiert wird.

Die zweite Bedingung besagt, dass die Menge W von gewöhnlichen Informationen zusätzlich zu den oben erwähnten Formeln, die eine strikte partielle Ordnung garantieren, jede Instantiierung der beiden Schemata enthalten muss:

$t \prec d^*$ und

$d^* \prec d$,

wobei d^* für beliebige Grenzwert-Defaults und d für gewöhnliche Defaults steht. Der Zweck dieser Formeln ist es sicher zu stellen, dass jeder Grenzwert-Default selbst über dem Grenzwert liegt, aber dass alle Grenzwert-Defaults eine niedrigere Priorität als gewöhnliche Defaults besitzen.

Ein Beispiel

Um diese Ideen zu veranschaulichen, beginnen wir mit der Beschreibung einer speziellen Default-Theorie mit Grenzwert $\langle W, D \rangle$, die das vorher skizzierte epistemische Beispiel mit den beiden Drogen repräsentiert. S, R, $D1$ und $D2$ sollen für die Sätze stehen, dass das Objekt vor mir rot aussieht, dass es rot ist und dass ich die Droge Nr. 1 beziehungsweise die Droge Nr. 2 genommen habe. δ_1, δ_2, δ_3 und δ_4 sind dann die gewöhnlichen Defaults, die zu D

gehören, wobei δ_1 für $S \rightarrow R$, δ_2 für $S \wedge D1 \rightarrow \neg R$, δ_3 für $\mathbf{T} \rightarrow d_1 \prec d_2$ und δ_4 für $D2 \rightarrow d_1 \prec t$ stehen. Gemäß dem ersten dieser Defaults ist es vernünftig zu schließen, dass das Objekt rot ist, wenn es rot aussieht; gemäß dem zweiten, dass das Objekt nicht rot ist, wenn es rot aussieht und ich Droge Nr. 1 genommen habe; gemäß dem dritten, dass der zweite Default stärker ist als der erste; und gemäß dem vierten, dass die Stärke des ersten Defaults unter den Grenzwert fällt, wenn ich Droge Nr. 2 genommen habe.

Weil $\langle W, D \rangle$ eine Default-Theorie mit Grenzwert ist, muss sie unsere beiden zusätzlichen Bedingungen erfüllen. Aufgrund der ersten Bedingung muss die Menge D auch vier Grenzwert-Defaults enthalten, wobei es zu jedem gewöhnlichen Default einen Grenzwert-Default gibt – das heißt, einen Default δ^*_i von der Form $\mathbf{T} \rightarrow t \prec d_i$ für jedes i von 1 bis 4. Und aufgrund der zweiten Bedingung muss die Menge W von gewöhnlichen Formeln vier Aussagen der Form $t \prec d^*_i$ für jedes i von 1 bis 4, sowie 16 Aussagen der Form $d^*_i \prec d_j$ enthalten, wobei i und j unabhängig von 1 bis 4 laufen.

Nehmen wir nun zunächst an, dass W zusätzlich die Formeln S und $D1$ enthält, was die Situation repräsentiert, in der das Objekt rot aussieht, ich aber Droge Nr. 1 genommen habe. In diesem Fall ergibt die Default-Theorie mit Grenzwert $S_1 = \{\delta^*_1, \delta^*_2, \delta^*_3, \delta^*_4, \delta_2, \delta_3\}$ als ihr einziges geeignetes Szenario, das die vier Aussagen der Form $t \prec d_i$ für i von 1 bis 4 sowie $\neg R$ und $d_1 \prec d_2$ stützt. Das Szenario S_1 erlaubt uns also zu schließen, dass jeder der gewöhnlichen Defaults über dem Grenzwert liegt, dass δ_2 eine höhere Priorität als δ_1 besitzt und dass das Objekt nicht rot ist. Der Default δ_1 wird in diesem Szenario geschlagen, weil seine Konklusion mit der des Defaults δ_2 in Konflikt steht, dessen höhere Stärke durch δ_3 begründet wird. Der Default δ_4 wird nicht ausgelöst, weil seine Prämisse im Kontext dieses Szenarios nicht impliziert wird.

Als nächstes nehmen wir stattdessen an, dass W die Formeln S und $D2$ enthält, was die Situation repräsentiert, in der das Objekt rot aussieht, ich aber Droge Nr. 2 genommen habe. Die Theorie ergibt jetzt $S_2 = \{\delta^*_2, \delta^*_3, \delta^*_4, \delta_3, \delta_4\}$ als ihr einziges geeignetes Szenario, das in diesem Fall nur drei Aussagen der Form $t \prec d_i$ für i von 2 bis 4 sowie $d_1 \prec d_2$ und jetzt auch $d_1 \prec t$ stützt. Das Szenario S_2 erlaubt uns zu schließen, dass die drei gewöhnlichen Defaults δ_2, δ_3 und δ_4 über dem Grenzwert liegen, dass δ_2 eine höhere Priorität besitzt als δ_1 und dass δ_1 tatsächlich unter dem Grenzwert liegt; über die wirkliche Farbe des Objekts können keine Konklusionen erzielt werden. Hier wird der Grenzwert-Default δ^*_1, der ansonsten δ_1 über dem Grenzwert platziert hätte, von δ_4 geschlagen, einem stärkeren Default, dessen größere Stärke sich aus der Formel $d^*_1 \prec d_4$ in W ergibt und der die konfligierende Konklusion stützt, dass δ_1 unter den Grenzwert fallen sollte. Weder δ_1 noch δ_2 werden ausgelöst; δ_2, weil seine

Prämisse im Kontext des Szenarios nicht impliziert wird, und δ_1, weil er unsere neue Anforderung nicht erfüllt, wonach ausgelöste Defaults über dem Grenzwert liegen müssen. Es lohnt sich, diese Situation vom Standpunkt unserer früheren Analyse eines Grundes als die Prämisse eines ausgelösten Defaults aus zu betrachten. Sobald ich Droge Nr. 2 genommen habe, sodass δ_1 unter den Grenzwert fällt, liefert dieser Default gemäß unserer Analyse keinen Grund mehr für den Schluss, dass das Objekt rot ist. Es ist aber nicht so, dass δ_1 im Sinne unseres üblichen Verständnisses von Schlagen durch Widerlegung geschlagen wird. Es gibt keinen stärkeren ausgelösten Default, der die konträre Konklusion stützt – tatsächlich gibt es überhaupt keinen Grund für den Schluss, dass das Objekt nicht rot ist. Stattdessen sind wir gezwungen zu schließen, dass der Default δ_1 unter dem Grenzwert liegen muss, sodass er selbst nicht ausgelöst werden kann und deshalb keinen eigenen Grund liefert.

4 Anwendung der Theorie

In diesem letzten Abschnitt möchte ich nun die in diesem Artikel entwickelte Theorie der Gründe auf ein Argument für den Partikularismus in der Moraltheorie anwenden, das kürzlich von Jonathan Dancy vorgebracht worden ist. Wir beginnen mit einigen allgemeinen Definitionen.

Es wird häufig angenommen, dass unsere Fähigkeit, sich für geeignete Handlungen zu entscheiden, oder zumindest unsere Fähigkeit, diese Handlungen als geeignet zu rechtfertigen, auf irgendeiner Ebene einen Rekurs auf allgemeine Prinzipien enthält. Lassen Sie uns jede derartige Ansicht als eine Form von *Generalismus* bezeichnen. Die hier vorgestellte Ansicht fällt sicherlich darunter, weil sie letztlich auf einem System von Prinzipien basiert, die dazu gedacht sind, anfechtbare Generalisierungen zu erfassen.

Im Gegensatz zum Generalismus steht die als *Partikularismus* bekannte Position, die die Wichtigkeit allgemeiner Prinzipien tendenziell herunterspielt und stattdessen eine Art von Sensibilität für die Eigenheiten bestimmter Situationen betont. Es ist allerdings nützlich, verschiedene Versionen dieser partikularistischen Perspektive zu unterscheiden. Wir können uns erstens einen *moderaten* Partikularismus vorstellen, der nur behauptet, dass zumindest ein signifikanter Teil unserer praktischen Bewertung nicht auf einem Rekurs auf allgemeine Prinzipien basiert.

Es ist schwer, sich mit dem moderaten Partikularismus auseinanderzusetzen, wenn er auf diese Weise formuliert wird. Es geschieht beispielsweise häufig, dass die Anwendung einer Menge von Prinzipien in einer bestimmten Situation ein von einem intuitiven Standpunkt aus

falsch erscheinendes Ergebnis ergibt, und wir haben deshalb das Gefühl, dass diese Prinzipien selbst geändert werden müssen. Aber wie können wir in einem Fall wie diesem darauf schließen, dass das ursprüngliche Ergebnis falsch war? Es kann nicht durch eine Anwendung unserer Prinzipien geschehen, weil es diese Prinzipien sind, die das Ergebnis generiert haben, mit dem wir jetzt nicht einverstanden sind. Und was leitet die Änderung unserer Prinzipien? Einige Autoren haben vorgeschlagen, dass wir uns auf eine fundamentalere Menge von Prinzipien berufen müssen, die irgendwie im Hintergrund liegen. Auch wenn dieser Vorschlag in vielen Fällen nützlich ist – wenn wir uns zum Beispiel auf moralische Prinzipien verlassen, um Fehler in einem Rechtssystem zu korrigieren –, ist seine Anwendbarkeit begrenzt. Solange wir eingestehen, dass jede endliche Menge von Prinzipien manchmal zu Fehlern in der praktischen Bewertung führen kann, muss die Berufung auf immer weitere Prinzipien zur Diagnose und Reparatur dieser Fehler letztlich entweder in einen Regress oder in eine Zirkularität münden.

Außer dem moderaten Partikularismus gibt es auch eine radikalere Position, die als *extremer* Partikularismus bezeichnet werden kann. Während die moderate Sichtweise einen Rekurs auf Prinzipien bei unserer alltäglichen praktischen Evaluation erlaubt und nur darauf besteht, dass es spezielle Umstände geben kann, in denen eine direkte Anwendung dieser Regeln falsche Ergebnisse liefert, behauptet der extreme Partikularismus, dass Prinzipien in der praktischen Evaluation überhaupt keine Rolle zu spielen haben.

Weil er die Legitimität eines jeden Rekurses auf Prinzipien verneint, ist der extreme Partikularismus rundweg inkonsistent mit dem Generalismus. Dennoch ist es genau diese radikale Position, die von Dancy vorgebracht wurde, der behauptet, dass der extreme Partikularismus aus einem allgemeineren *Holismus* der Gründe folgt – aus der Idee, dass die Kraft von Gründen variabel ist, so dass das, was in einer Situation als Grund für eine Handlung oder Konklusion gilt, in einer anderen Situation nicht dieselbe Handlung oder Konklusion stützten muss.[11]

4.1 Das Argument

Nach Dancys Ansicht ist Holismus ein allgemeines Phänomen, das sowohl bei praktischen als auch bei epistemischen Gründen Anwendung findet. Beide, so schreibt er, sind imstande ihre

[11] Das Argument wird mit kleineren Veränderungen in einer Reihe von Veröffentlichungen dargestellt, die mit Dancy (1983) beginnt. Ich konzentriere mich hier aber auf die Versionen, die in Dancy (1993; 2000; 2001) und besonders kanonisch in Dancy (2004) zu finden sind. Ähnliche Argumente wurden von anderen Autoren vorgelegt. Eine besonders klare Darstellung findet sich in Little (2000).

Polarität zu wechseln: Eine Überlegung, die in einem Kontext als Grund für eine Handlung oder Konklusion fungiert, muss in einem anderen Kontext nicht als Grund für die gleiche Handlung oder Konklusion dienen und könnte tatsächlich sogar ein Grund dagegen sein. Dancy präsentiert sowohl für den praktischen als auch für den theoretischen Bereich eine Auswahl von Fällen, die diese Möglichkeit aufzeigen sollen. Weil diese Beispiele einem gemeinsamen Muster folgen, betrachten wir nur zwei repräsentative Vertreter.

Um mit dem praktischen Bereich zu beginnen, stellen Sie sich vor, dass ich mir ein Buch von Ihnen geliehen habe. In den meisten Situationen würde mir die Tatsache, dass ich ein Buch von Ihnen geliehen habe, einen Grund geben, Ihnen das Buch zurückzugeben. Aber stellen Sie sich vor, dass ich entdecke, dass das geliehene Buch eines ist, das Sie vorher aus der Bücherei gestohlen haben. In diesem Kontext fungiert die Tatsache, dass ich das Buch von Ihnen geliehen habe, nach Dancy nicht länger als Grund dafür, Ihnen das Buch zurückzugeben; tatsächlich habe ich überhaupt keinen Grund mehr, es Ihnen zurückzugeben.[12] Um das gleiche Phänomen im epistemischen Bereich sichtbar zu machen, benutzt Dancy ein Standardbeispiel, das wir bereits betrachtet haben. In den meisten Situationen gibt mir die Tatsache, dass ein Objekt rot aussieht, einen Grund dafür zu denken, dass das Objekt rot ist. Aber stellen Sie sich vor, dass ich weiß, dass ich Droge Nr. 1 genommen habe, die, wie wir uns erinnern, zur Folge hat, dass rote Dinge blau und blaue Dinge rot aussehen. In diesem neuen Kontext fungiert die Tatsache, dass das Objekt rot aussieht, nach Dancy nicht länger als Grund dafür zu denken, dass das Objekt rot ist; sie ist stattdessen ein Grund dafür zu denken, dass das Objekt blau und also nicht rot ist.[13]

Räumen wir für den Moment ein, dass Beispiele wie diese hinreichend dafür sind, einen Holismus bezüglich von Gründen einzuführen. Wie soll diese holistische Sichtweise zu einem extremen Partikularismus führen, einer kompromisslosen Ablehnung jeglicher Rolle für allgemeine Prinzipien bei praktischer Bewertung? Um diese Frage zu beantworten, müssen wir die Natur der Verallgemeinerungen bedenken, die in diesen Prinzipien inbegriffen sind, und es ist nützlich, sich auf ein konkretes Beispiel zu konzentrieren. Betrachten wir also das Prinzip, dass es falsch ist, zu lügen. Was könnte das bedeuten?

Wir könnten dieses Prinzip zunächst als einen allquantifizierten materialen Konditionalsatz verstehen, gemäß welchem jede Handlung, die Lügen beinhaltet, dadurch falsch ist, und zwar unabhängig von den Umständen. Obwohl einige Autoren eine derartige Sichtweise befürwortet haben, erachten heute nur wenige Leute eine solch unnachgiebige Konzeption als

[12] Dieses Beispiel findet sich in Dancy (1993, S. 60). Danach folgen einige ähnliche Beispiele.
[13] Dancy (2004, S. 74); das Beispiel wird auch in Dancy (2000, S. 132) und in Abschnitt 3 von Dancy (2001) diskutiert.

haltbar. Es ist natürlich möglich, den Vorschlag abzuschwächen, indem man das einfache Prinzip, dass Lügen falsch ist, als eine Art von Abkürzung für eine viel kompliziertere Regel ansieht, immer noch ein materiales Konditional, aber eines, das mit Ausnahmeklauseln bestückt ist, die all die verschiedenen Umstände abdecken, in denen es akzeptabel sein könnte, zu lügen – die Rettung eines Lebens, Vermeidung sinnloser Beleidigungen und so weiter. Das Problem dieses Vorschlags ist jedoch, dass noch nie eine befriedigende Regel dieser Form aufgestellt worden ist, und es ist legitim, an unserer Fähigkeit zu zweifeln, solche Regeln mit vollständigen Ausnahmeklauseln auch nur zufrieden stellend zu formulieren – vom Lernen dieser Regeln und dem Schlussfolgern mit ihnen ganz zu schweigen.

Alternativ könnten wir das Prinzip, dass Lügen falsch ist, nicht als die Vorstellung verstehen, dass alle Handlungen falsch sind, die Lügen beinhalten, oder sogar, dass alle Handlungen falsch sind, die Lügen beinhalten und zusätzlich noch eine umfangreiche Liste von Bedingungen erfüllen, sondern einfach als die Vorstellung, dass Lüge immer ein Merkmal ist, das gegen eine Handlung spricht – ein „falsch-machendes“ Merkmal. In dieser Sichtweise würde die Tatsache, dass eine Handlung eine Lüge beinhaltet, immer als ein Grund gelten, die Handlung für falsch zu halten, auch wenn diese Handlung tatsächlich insgesamt als richtig erachtet werden könnte, wenn man verschiedene andere Gründe in Betracht zieht. Die Funktion von Prinzipien wäre es dann, allgemeine Gründe für oder gegen Handlungen oder Konklusionen zu artikulieren, die nicht entscheidend sein müssen, aber die zumindest eine gleich bleibende Rolle in unseren Abwägungen spielen, indem sie immer eine bestimmte Seite favorisieren.

Aber nehmen Sie jetzt an, dass der Holismus bezüglich von Gründen korrekt ist, so dass jede Überlegung, die in einer Situation ein Ergebnis favorisiert, in einer anderen Situation nicht das gleiche Ergebnis favorisieren könnte. In diesem Fall gäbe es keine allgemeinen Gründe, keine Überlegungen, die in unseren Abwägungen eine gleich bleibende Rolle spielen, indem sie unabhängig vom Kontext die gleiche Kraft ausüben. Wenn es dann die Funktion von Prinzipen ist, allgemeine Gründe zu spezifizieren, gibt es einfach nichts für sie zu spezifizieren; jedes Prinzip, das besagt, dass ein Grund eine bestimmte Rolle in unseren Abwägungen spielt, müsste falsch sein, weil es immer einen Kontext gäbe, in dem dieser Grund eine andere Rolle spielt. Das ist Dancys Konklusion – dass, wie er sagt, „ein auf Prinzipien basierender Ansatz zur Ethik inkonsistent ist mit dem Holismus von Gründen.“[14]

4.2 Bewertung des Arguments

[14] Dancy (2004, S. 77)

Beginnen wir mit dem epistemischen Beispiel, das Dancy anbietet, um den Holismus von Gründen zu etablieren. Dieses Beispiel wurde vorher als eine Default-Theorie mit Grenzwert dargestellt, die wir jetzt der Verständlichkeit halber etwas vereinfacht neu fassen.[15] Nehmen Sie wieder an, dass *S*, *R* und *D*1 jeweils für die Propositionen stehen, dass das Objekt vor mir rot aussieht, dass es rot ist und dass ich Droge Nr. 1 genommen habe; und nehmen Sie an, dass δ_1 für $S \rightarrow R$, δ_2 für $S \wedge D1 \rightarrow \neg R$ und δ_3 für $\mathbf{T} \rightarrow d_1 \prec d_2$ stehen. Sei $\langle W, D \rangle$ die Theorie, in der *D* δ_1, δ_2, und δ_3 enthält und in der *W S* und *D*1 enthält, gemäß der das Objekt rot aussieht, ich aber die Droge genommen habe. Weil das eine Default-Theorie mit Grenzwert ist, muss D zusätzlich zu diesen drei gewöhnlichen Defaults auch die korrespondierenden Grenzwert-Defaults δ^*_i der Form $\mathbf{T} \rightarrow t \prec d_i$ für jedes *i* von 1 bis 3 enthalten; und W muss Aussagen der Form $t \prec d^*_i$ für *i* von 1 bis 3 sowie Aussagen der Form $d^*_i \prec d_j$ für *i* und *j* von 1 bis 3 enthalten.

Es ist leicht zu sehen, dass diese Default-Theorie als einziges geeignetes Szenario $S_1 = \{\delta^*_1, \delta^*_2, \delta^*_3, \delta_2, \delta_3\}$ ergibt, das die drei Aussagen der Form $t \prec d_i$ für *i* von 1 bis 3 sowie $\neg R$ und $d_1 \prec d_2$ stützt. Die Theorie erlaubt uns also zu schließen, dass jeder der gewöhnlichen Defaults δ_1, δ_2 und δ_3 über dem Grenzwert liegt, dass δ_2 eine höhere Priorität besitzt als δ_1, und dass das Objekt nicht rot ist.

Dancys Argument hängt aber nicht so sehr von den Konklusionen ab, die in bestimmten Situationen gestützt werden, sondern vielmehr von Aussagen, die in diesen Situationen als Gründe oder als keine Gründe klassifiziert werden. Diese Angelegenheit kann auf nützliche Weise vom Standpunkt unserer Analyse aus untersucht werden, wonach ein Grund die Prämisse eines ausgelösten Defaults ist. Und wenn wir die Angelegenheit von diesem Standpunkt aus untersuchen und die Situation wie oben repräsentieren, sehen wir, dass die Proposition *S*, dass das Objekt rot aussieht, in der Tat als Grund klassifiziert wird. Warum? Weil ein Default im Kontext eines Szenarios genau dann ausgelöst wird, wenn seine Prämisse vom Szenario impliziert wird und der Default über dem Grenzwert liegt. Die Prämisse des Defaults δ_1, die Aussage *S*, ist in der gewöhnlichen Information, die von W bereitgestellt wird, bereits enthalten und wird also von jedem Szenario impliziert; und im Kontext von Szenario S_1 liegt der Default über dem Grenzwert. Dieser Default wird folglich ausgelöst; und deshalb wird seine Prämisse gemäß unserer Analyse als Grund für seine Konklusion klassifiziert – die Aussage *R*, dass das Objekt rot ist.

[15] Die Vereinfachung ist, dass wir jetzt Droge Nr. 2 ignorieren, die uns nicht weiter interessiert.

Wenn *S* ein Grund für *R* ist, warum stützt das Szenario dann aber diese Konklusion nicht? Nun, es ist natürlich zu sagen, dass ein Grund – eine Proposition, die eine Konklusion favorisiert – *geschlagen* wird, wann immer der Default, der diese Proposition als Grund bereitstellt, selbst geschlagen wird. Und im vorliegenden spezifischen Fall wird der Default δ_1, der *S* als Grund für die Konklusion *R* bereitstellt, vom stärkeren Default δ_2 geschlagen, der $S \wedge D1$ – die Proposition, dass das Objekt rot aussieht, ich aber Droge Nr. 1 genommen habe – als Grund für die konfligierende Konklusion $\neg R$ bereitstellt.

Unsere anfängliche Repräsentation dieser Situation verdeutlicht also die Möglichkeit einer alternativen Interpretation eines der Schlüsselbeispiele, auf die sich Dancy beruft, um einen Gründe-Holismus einzuführen. Nach Dancy bildet die Situation, in der sowohl *S* als auch *D*1 gelten – das Objekt sieht rot aus, aber ich habe die Droge genommen –, einen Kontext, in dem *S*, obwohl *S* normalerweise ein Grund für *R* ist, seinen Status als Grund gänzlich verliert. Was in normalen Fällen ein Grund ist, ist in diesem Fall kein Grund und wir werden zu einem Gründe-Holismus gedrängt. Im Gegensatz dazu wird *S* in unserer jetzigen Repräsentation immer noch als Grund für *R* klassifiziert – aber eben als geschlagener Grund.

Ich erwähne diese Möglichkeit hier nur, um zu zeigen, dass es verschiedene Arten der Interpretation von Situationen gibt, in denen eine bekannte Überlegung nicht ihre gewohnte Rolle als kraftvoller oder überzeugender Grund spielt. In jedem Einzelfall müssen wir uns fragen: Versagt die Überlegung vollständig darin, als Grund zu fungieren, oder fungiert sie zwar als Grund, aber einfach als einer, der geschlagen wird? Die Antwort hierauf bedarf häufig eines feinfühligen Urteils, und manchmal sind verschiedene Interpretationen derselben Situation möglich – wir werden auf diesen Punkt zurückkommen, wenn wir uns Dancys praktischem Beispiel zuwenden.

In dem hier betrachteten besonderen Fall trifft es sich, dass die Idee, dass rotes Aussehen immer noch ein Grund ist, aber eben einer, der geschlagen wird, von Dancy in Betracht gezogen, aber sofort zurückgewiesen wird:

> „Es ist nicht so, als ob es ein Grund für mich wäre zu glauben, dass etwas Rotes vor mir ist, dass dieser Grund allerdings von Gegengründen geschlagen wird. Es ist *überhaupt kein Grund* mehr zu glauben, dass etwas Rotes vor mir ist; stattdessen ist es ein Grund, das Gegenteil zu glauben.“[16]

[16] Dancy (2004, S. 74)

Und in diesem besonderen Fall müsste ich zustimmen. Wenn ich die Droge genommen habe, scheint es nicht so, als ob rotes Aussehen immer noch einen Grund für die Folgerung bereitstellt, dass das Objekt rot ist, der dann von einem stärkeren Grund für das Gegenteil geschlagen wird. In diesem Fall scheint es so zu sein, dass der Status von rotem Aussehen als Grund selbst unterminiert wird.

Wichtig ist allerdings zu sehen, dass diese – Dancys bevorzugte – Interpretation der Situation genauso innerhalb des hier vorgestellten Rahmens behandelt werden kann, und zwar durch eine Anwendung unseres Verfahrens für Schlagen durch Unterminierung. Sei δ_4 der neue Default $D1 \rightarrow d_1 \prec t$, gemäß welchem der vorherige Default δ_1 – der rotes Aussehen als Grund für Rotsein festlegte – nun in jeder Situation, in der ich die Droge nehme, unter den Grenzwert fällt. Die passenden Ergebnisse können dann erzielt werden, indem man die vorherige Theorie um diesen neuen Default ergänzt und die notwendigen Grenzwertbedingungen erfüllt: Formal sei $\langle W, D \rangle$ die Default-Theorie mit Grenzwert, in der D sowohl die gewöhnlichen Defaults δ_1, δ_2, δ_3 und nun δ_4 enthält als auch die korrespondierenden Grenzwert-Defaults δ^*_i der Form $\mathbf{T} \rightarrow t \prec d_i$ für i von 1 bis 4 und in der W S und $D1$ zusammen mit Aussagen der Form $t \prec d^*_i$ und $d^*_i \prec d_j$ für i und j von 1 bis 4 enthält.

Diese neue Default-Theorie ergibt als ihr einziges geeignetes Szenario $S_2 = \{\delta^*_2, \delta^*_3, \delta^*_4, \delta_2, \delta_3, \delta_4\}$, das drei Aussagen der Form $t \prec d_i$ für i von 2 bis 4 sowie $d_1 \prec d_2$, $\neg R$ und $d_1 \prec t$ stützt. Die Theorie erlaubt uns also zu schließen, dass die gewöhnlichen Defaults δ_2, δ_3 und δ_4 über dem Grenzwert liegen, dass δ_2 eine höhere Priorität besitzt als δ_1, dass δ_1 tatsächlich unter den Grenzwert fällt, und natürlich, dass das Objekt nicht rot ist.

Nun ist – ich wiederhole mich – ein Grund die Prämisse eines ausgelösten Defaults, und der Default δ_1, der zuvor S als Grund für R bereitgestellt hat, wird, gegeben diese neue Repräsentation der Situation, nicht mehr ausgelöst. Wieso nicht? Weil ein Default im Kontext eines Szenarios ausgelöst wird, wenn seine Prämisse in diesem Szenario impliziert wird und wenn er zusätzlich über dem Grenzwert liegt; und während die Prämisse von δ_1 tatsächlich impliziert wird, fällt der Default jetzt unter den Grenzwert. Weil S, das rote Aussehen, nicht mehr die Prämisse eines ausgelösten Defaults ist, wird es nicht als Grund klassifiziert. Was in normalen Fällen ein Grund ist, ist also in dieser Situation nicht nur ein geschlagener Grund, sondern überhaupt kein Grund mehr, genau wie Dancy behauptet.

Mit dieser Beobachtung können wir uns jetzt einer Bewertung von Dancys Argument zuwenden. Das Argument hängt an der Vorstellung, dass der extreme Partikularismus, also die Ablehnung allgemeiner Prinzipien, aus dem Gründe-Holismus folgt. Aber der hier

entwickelte Ansatz stützt den Gründe-Holismus, indem er die Möglichkeit eröffnet, dass eine Überlegung, die in einer Situation als Grund gilt, in einer anderen Situation nicht als Grund klassifiziert wird. Trotzdem basiert dieser Ansatz selbst auf einem System von Prinzipien, den Defaults, die als Instantiierungen von anfechtbaren Generalisierungen betrachtet werden können; was in einer bestimmten Situation ein Grund oder kein Grund ist, wird tatsächlich durch diese Defaults bestimmt. Der Ansatz stellt somit ein Gegenbeispiel zu der Vorstellung dar, dass der Gründe-Holismus einen extremen Partikularismus impliziert und dass der Holismus mit jeder Form von Generalismus inkonsistent ist. Der Gründe-Holismus ist zumindest mit dieser Form des Generalismus konsistent, und somit ist Dancys Argument genau genommen ungültig.

Offensichtlich gibt es hier eine Diskrepanz. Wie sollte die Diagnose aussehen? Wieso glaubt Dancy, dass der Holismus mit jeglichem Rekurs auf allgemeine Prinzipien inkonsistent ist, während es im hier entwickelten Ansatz so aussieht, als ob die Ideen des Holismus und des Generalismus in Einklang zu bringen sind?

Die Diskrepanz hat, so denke ich, ihre Wurzeln in unseren verschiedenen Ansichten über die Bedeutung von allgemeinen Prinzipien. Wir erkennen beide an, dass die Prinzipien, die unser praktisches Schließen leiten, im Großen und Ganzen nicht auf nützliche Weise als Ausdruck allquantifizierter materialer Konditionalsätze verstanden werden können. Das praktische Prinzip, dass Lügen falsch ist, kann nicht bedeuten, dass jede Handlung falsch ist, die Lügen beinhaltet. Was Dancy stattdessen vorschlägt, ist, wie wir gesehen haben, diese Prinzipen so zu verstehen, dass sie Überlegungen spezifizieren, die ausnahmslos eine Rolle als Gründe spielen. Das Prinzip, dass Lügen falsch ist, soll bedeuten, dass Lügen immer einen Grund gibt, eine Handlung als weniger positiv zu bewerten, sogar in den Fällen, in denen unsere allgemeine Bewertung der Handlung positiv ist. Und vermutlich muss das epistemische Prinzip, gemäß welchem Dinge, die rot aussehen, tendenziell rot sind, auch so verstanden werden, dass rotes Aussehen immer einen Grund für den Schluss gibt, dass ein Objekt rot ist, sogar in den Fällen, in denen es unsere Konklusion insgesamt ist, dass das Objekt nicht rot ist.

Gegeben dieses Verständnis von allgemeinen Prinzipien folgt nun sofort – ist es *offensichtlich* –, dass der Gründe-Holismus zur Ablehnung allgemeiner Prinzipien führen muss. Wenn der Holismus korrekt ist, so dass das, was in einer Situation ein Grund ist, in einer anderen Situation kein Grund sein muss, dann muss natürlich jedes Prinzip falsch sein, das einer Überlegung eine ausnahmslose Rolle als Grund zuweist. Wenn die Bewertung von „Lügen ist falsch“, ist, dass Lüge *immer* eine negative Bewertung einer Handlung begünstigt, und wenn

es gewisse Situationen gibt, in denen dies nicht der Fall ist, dann muss das praktische Prinzip selbst fehlerhaft sein, und es kann unsere Handlungen nicht auf geeignete Weise leiten. Wenn die Bedeutung von „rotes Aussehen zeigt Rotsein an“ ist, dass rotes Aussehen *immer* einen Grund für den Schluss gibt, dass ein Objekt rot ist, und wenn es gewisse Situationen gibt, in denen dies nicht der Fall ist, dann ist das epistemische Prinzip fehlerhaft.

Ich stimme also zu, dass der Gründe-Holismus die Ablehnung von allgemeinen Prinzipien implizieren muss, wenn man Dancys Auffassung dieser Prinzipien übernimmt. Mit der Entwicklung eines Ansatzes, in dem der Holismus konsistent mit allgemeinen Prinzipien ist, will ich mich daher auf eine andere Auffassung dieser Prinzipien stützen – sie spezifizieren keine ausnahmslosen Gründe, sondern modifizieren die Defaults, die unser Schließen leiten. Nach dieser Ansicht sollte das allgemeine Prinzip, dass Lügen falsch ist, einfach in dem Sinn verstanden werden, dass Lügen per Default falsch ist – das heißt in einer ersten Annäherung, dass wir, sobald wir erfahren, dass eine Handlung Lügen beinhaltet, diese Handlung als falsch beurteilen sollten, wenn nicht gewisse Komplikationen dazwischen kommen. Und das Prinzip, dass rotes Aussehen auf Rotsein hindeutet, sollte ebenso in dem Sinn verstanden werden, dass diese Relation per Default besteht – dass wir, sobald wir sehen, dass ein Objekt rot aussieht, schließen sollten, dass das Objekt rot ist, und zwar wieder wenn keine Komplikationen auftreten.

Diese Explikation von allgemeinen Prinzipien als Aussagen, die Defaults modifizieren, beinhaltet eine offene und explizite Berufung auf Ceteris paribus Einschränkungen. Die Prinzipien sagen uns, was wir bei Abwesenheit von externen Komplikationen schließen sollten. Ceteris-paribus-Aussagen wie diese werden manchmal als nichts sagend kritisiert (scherzhaft formuliert heißt es, dass eine Erklärung dieser Art die offenbar gehaltvolle Aussage, dass Lügen falsch ist, auf eine weniger gehaltvolle Aussage von der Art „Lügen ist falsch, außer wenn es nicht falsch ist“ reduziert). Es wird auch argumentiert, dass diese Aussagen, die die geeigneten Konklusionen bei Abwesenheit von Komplikationen spezifizieren, nichts über die üblicheren Fälle sagen, in denen Komplikationen auftreten. Beide Einwände haben, glaube ich, ihre Verdienste, wenn sie gegen viele der üblichen Berufungen auf Ceteris-paribus-Generalisierungen vorgebracht werden, weil diese Berufungen oft nicht über die Ebene einer ersten Annäherung hinausgehen. Für den vorliegenden Ansatz sind diese Einwände allerdings nicht zutreffend. Hier ist unsere erste Annäherung an die Bedeutung eines allgemeinen Prinzips nichts als eine Zusammenfassung der Arbeitsweise dieses Prinzips innerhalb der zugrunde liegenden Default-Theorie, die nicht nur im Detail spezifiziert, was Komplikationen sein können – wann ein Default

widersprochen, geschlagen oder unterminiert ist – sondern auch, wie die verschiedenen Probleme gelöst werden können, die durch diese Komplikationen hervorgerufen werden, und wie man in jedem Fall zur geeigneten Menge von Konklusionen kommt. Was der vorliegende Ansatz also beizutragen hat, ist eine konkrete Theorie des Default-Schließens zur Untermauerung der Explikation von allgemeinen Prinzipien als Defaults, eine Theorie, die präzise, durch unsere Intuitionen gestützt und konsistent mit dem Gründe-Holismus ist.[17]

LITERATUR

Gerhard Brewka: Reasoning about priorities in default logic, in: *Proceedings of the Twelveth National Conference on Artificial Intelligence* AAAI-94 (1994), MIT Press, 940-945.

Gerhard Brewka: Well-founded semantics for extended logic programs with dynamic preferences, in: *Journal of Artificial Intelligence* 4 (1996), 19-36.

Jonathan Dancy: Ethical particularism and morally relevant properties, in: *Mind* 92 (1983), 530-547.

Jonathan Dancy: *Moral Reasons*, Basil Blackwell Publisher 1993.

Jonathan Dancy: The particularist's progress, in: *Moral Particularism*, ed. by Brad Hooker and Margaret Little, Oxford University Press 2000, 130-156.

Jonathan Dancy: Moral particularism, in: *The Stanford Encyclopedia of Philosophy (Summer 2001 Edition)*, ed. by Edward Zalta, Stanford University 2001. Abrufbar auf http://plato.stanford.edu/archives/sum2001/entries/moral-particularism/.

Jonathan Dancy: *Ethics Without Principles*, Oxford University Press 2004.

Scott Fahlman: *NETL: A System for Representing and Using Real-world Knowledge*, The MIT Press 1979.

Thomas Gordon: *The Pleadings Game: An Artificial-Intelligence Model of Procedural Justice*, Dissertation, Technische Hochschule Darmstadt 1993.

John Horty: Some direct theories of nonmonotonic inheritance, in: *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, ed. by D. Gabby, C. Hogger and J. Robinson, Oxford University Press 1994, 111-187.

John Horty: Defaults with priorities, in: *Journal of Philosophical Logic* 36 (2007), 367-413.

[17] Die Übertragung des englischen Textes ins Deutsche wurde von Stefan Ruhland, Regensburg, besorgt.

Margaret Little: Moral generalities revised, in: *Moral Particularism*, ed. by Brad Hooker and Margaret Little, Oxford University Press 2000.
John Pollock: The structure of epistemic justification, in: *Studies in the Theory of Knowledge, American Philosophical Quarterly Monograph Series* 4 (1970), Basil Blackwell Publisher, 62-78.
John Pollock: *Cognitive Carpentry: A Blueprint for How to Build a Person*, The MIT Press 1995.
Henry Prakken and Giovanni Sartor: On the relation between legal language and legal argument: assumptions, applicability, and dynamic properties, in: *Proceedings of the Fifth International Conference in Artificial Intelligence and Law* ICAIL-95 (1995), The ACM Press.
Henry Prakken and Giovanni Sartor: A dialectical model of assessing conflicting arguments in legal reasoning, *Artificial Intelligence and Law* 4 (1996), 331-368.
Henry Prakken and Gerhard Vreeswijk: Logical systems for defeasible argumentation, in: *Handbook of Philosophical Logic (Second Edition)*, ed. by Dov Gabby and F. Guenther, Kluwer Academic Publishers 2002, 219-318.
Joseph Raz: *Practical Reasoning and Norms*, Hutchinson Company 1975. [Zweite Auflage Princeton University Press 1990, wieder abgedruckt von Oxford University Press 2002; Seitenangaben beziehen sich auf die Oxford-Ausgabe.]
Mark Schroeder: Weighting for a Humean theory of reasons, in: *Nous* (2005). Forthcoming.
Mark Schroeder: Holism, weight, and undercutting, in: *Nous* 45 (2011), 328-344.