

DAS UNIVERSALISIERBARKEITSPROBLEM IN DER MORALPHILOSOPHIE*

Georg Meggle (Saarbrücken)

EINLEITUNG

Der Titel meines Vortrags ist irreführend. Er unterstellt, daß es *das* Universalisierbarkeits-Problem in der Moralphilosophie gibt. Das ist falsch. Zum einen laufen dort unter dem Stichwort *Universalisierbarkeit* mehrere Dinge. Und bei nahezu jedem davon gibt es jeweils eine ganze Menge Probleme. Natürlich packe ich heute nur ein paar davon an.

Daß schon der Term *Universalisierbarkeit* mehrdeutig ist, dieses Manko ist leicht zu beheben. Ich erkläre hiermit, daß es mir im folgenden nur um eine ganz bestimmte Art von Universalisierbarkeit geht. Nämlich um die, wonach in einer Situation etwas nur dann geboten ist, wenn dasselbe in jeder Situation, die der erstgenannten hinsichtlich jeder universellen Eigenschaft gleicht, ebenfalls geboten ist.

– Dieses Universalisierungsprinzip ist nicht zu verwechseln mit dem Verallgemeinerungsprinzip. Letzteres besagt, daß eine Handlung nur dann erlaubt ist, wenn es auch erlaubt ist, daß alle Leute sie vollziehen.

– Das Universalisierungsprinzip ist auch nicht zu verwechseln, aber wird verwechselt, mit dem Generalisierungsprinzip. Letzteres besagt, daß moralische Urteile für möglichst viele Situationen einschlägig sein sollten. In diesem Sinne ist etwa „Du sollst nicht lügen.“ allgemeiner als „In allen Situationen, in denen durch eine Lüge kein Menschenleben gerettet wird, sollst Du nicht lügen.“ In einer Lesart von „Individualisierung“ ist Generalisierung, nicht Universalisierung, der Gegenbegriff zu Individualisierung. Unter der Überschrift „Generalisierung vs. Individualisierung“ wären etwa die Fragen zu diskutieren, ob Verhaltenskodizes viele mögliche Fälle einzeln durchspielen oder eher allgemein verfahren sollten und ob überhaupt eine Kodifizierung denkbar ist, die Einzelfällen gerecht wird.

– Das Universalisierungsprinzip sollte auch nicht verwechselt werden mit der Goldenen Regel. Die besagt, in Luthers Worten, „Und wie ir wolt das euch die Leute thun sollen, also thut inen gleich auch ir.“ (Lucas VI) Die Goldene Regel nimmt im Unterschied zum Universalisierungsprinzip keinen Bezug auf eine Gleichheit der Rahmenbedingungen. Ein Masochist, der seinen nicht-masochistischen Nachbarn quält, könnte sich auf die Goldene Regel berufen. Nicht aber auf das Universalisierungsprinzip; selbst dann nicht, wenn wir einmal voraussetzen, daß es, da sie ja Spaß daran haben, geboten ist, Masochisten zu quälen. Denn daß der nicht-masochistische Nachbar nicht gequält werden will, ist eine universelle Eigenschaft, hinsichtlich derer sich die eine Situation, also seine Tortur, von der anderen, also der Tortur des Masochisten unterscheidet. Also sagt das Universalisierungsprinzip nicht, daß, da das eine geboten ist, auch das andere geboten ist.

* Den großen Spaß an und einen Gutteil meines eigenen Erkenntnisgewinns aus der Arbeit an diesem Vortrag verdanke ich Diskussionen mit Christoph Fehige. Sein Input steckt vor allem in IV.

I UNIVERSALISIERBARKEIT, INFERENTIALITÄT UND SUPERVENIENZ

Universalisierbar zu sein, sei, so ist generell zu hören, nichts besonderes – auch nicht für moralische Urteile. Im Gegenteil: Universalisierbarkeit sei gerade die Eigenschaft, die moralische Urteile – und im übrigen auch alle sonstigen Werturteile – mit deskriptiven Urteilen gemeinsam haben. Ja, es sei sogar so, daß sich die Universalisierbarkeit von Moralurteilen gerade daraus ergebe, daß auch diese eine deskriptive Bedeutungskomponente besitzen. Was liegt also näher, als das Universalisierbarkeits-Merkmal zunächst einmal anhand der generell als weniger problematisch erachteten deskriptiven Urteile erläutern zu wollen.

Richard M. Hare, der vehementeste Universalisierer in der zeitgenössischen Ethik, tut das in *Freiheit und Vernunft (FV)* (S. 25 f.) für das einfache partikuläre Urteil „a ist rot“ in zwei Schritten so:

1. Schritt: „Wer von einem Gegenstand behauptet, daß er rot ist, der hat sich damit darauf festgelegt, auch alle diejenigen Dinge rot zu nennen, die diesem Gegenstand in den relevanten Gesichtspunkten ähnlich sind.“

Gegen diese Formulierung läßt sich aber einwenden, daß sie schon aus formalen Gründen nicht streng universell sei, da in ihr bei der Bestimmung der relevanten Ähnlichkeiten auf den betreffenden Gegenstand selbst noch Bezug genommen wird. Diesem Einwand trägt Hare in seinem 2. Schritt Rechnung.

2. Schritt: „Man kann ... jeden singulären Ausdruck, der von dem Wort „ähnlich“ oder einem Synonym davon regiert wird, zu einem universellen machen. Man braucht nur für ‚diesem Ding ähnlich‘ einen Ausdruck einzusetzen, der die Gesichtspunkte beschreibt, unter denen man von dem in Frage stehenden Gegenstand behauptet, daß er jenem Ding ähnlich ist. ... Und damit gilt: Wenn sich jemand durch die Äußerung

(1) a ist rot

auf die Behauptung

(2) Alle Dinge, die a in den relevanten Punkten ähnlich sind, sind ebenfalls rot

festgelegt hat, dann ist er auch auf die folgende Behauptung festgelegt:

(3) Es gibt eine Eigenschaft, von welcher gilt: Sie kommt a zu, und alle Dinge, die diese Eigenschaft besitzen, sind rot.“

Nun war aber, daß wir mit der Behauptung (1) auf die Behauptung (2) festgelegt sind, schon im 1. Schritt gesagt worden. Beide Schritte zusammengenommen besagen so, daß wir mit Behauptung (1) auch auf Behauptung (3) festgelegt sind.

Die Behauptung (3) macht, wie man auch sagen könnte, die in Behauptung (1) lediglich implizit enthaltene Universalität explizit. Wir definieren nun genau in diesem Sinne, wobei $\&$, $>$ und \equiv für die Konjunktion, Implikation bzw. Äquivalenz stehen:

Ein (partikuläres) Urteil von der Form Ea ist *implizit universell* gdw. gilt:

$\forall G(Ga \ \& \ \Lambda x(Gx \ > \ Ex))$

Es gibt eine Eigenschaft G, von welcher gilt: Sie kommt a zu und alle Dinge, die die Eigenschaft G haben, haben die Eigenschaft E

Man kann nun leicht sehen, daß im eben definierten Sinne jedes partikuläre Urteil implizit universell ist. Denn eine Eigenschaft der geforderten Art, die die implizite Universalität

von Ea garantiert, gibt es immer. Nämlich die Eigenschaft E selbst. Denn selbstverständlich gilt: $\Lambda x(Ex > Ex)$ – und aus dieser Tautologie allein folgt bereits die implizite Universalität von Ea. Die Behauptung „a ist rot“ ist demnach schon deshalb implizit universell, weil es 1. eine Eigenschaft gibt, die a zukommt, nämlich die Eigenschaft der Röte: und 2. für alle Dinge mit dieser Eigenschaft trivialerweise gilt, daß sie rot sind. Entsprechendes gilt für jedes (partikuläre) Moralurteil. Gilt z.B., daß a die Handlung F tun soll, d.h., ist es dem a geboten, F zu tun (wofür wir im folgenden – mit O wie „obligatorisch“ – auch kurz schreiben: OFa), so gilt trivialerweise wiederum: Es gibt 1. eine Eigenschaft, die a zukommt, nämlich die des F-Tun Sollens, und 2. ist es auch jederman, dem es geboten ist, F zu tun, geboten, F zu tun: $\Lambda x(OFx > OFx)$. Geboten ist, was geboten ist. Auch jedes (partikuläre) Moralurteil ist also in diesem Sinne implizit universell.

Sollte mit der Universalisierbarkeit wirklich nicht mehr als diese implizite Universalität gemeint sein? Nun, dann wäre die These, daß alle partikulären Urteile (insbesondere auch alle partikulären Moralurteile) universalisierbar sind, zwar in der Tat richtig – aber doch wohl zu trivial, um auch nur von irgendeinem Interesse zu sein (vgl. jedoch Hare *FV*, S. 26). Diese Deutung wäre zudem ganz unverträglich mit der Idee, wonach die Universalisierbarkeit eines Urteils etwas mit dessen Begründbarkeit zu tun haben soll. Und so erst recht mit der stärkeren Idee, wonach sogar gelten soll: Universalisieren heißt begründen (so Hare in *FV*, S. 19). Wie immer man, was echte Begründungen von Scheinbegründungen unterscheidet, des näheren charakterisieren mag: Zirkuläre Begründungen wird man nicht zulassen. Von dieser Sorte waren aber unsere eben vorgeführten trivialen Explizitmachungen impliziter Universalitäten ganz sicher: Wir rekurrierten zur Begründung eines partikulären Urteils letztlich nur auf dieses selbst. Die universelle Komponente – d.h. $\Lambda x(Ex > Ex)$ bzw. $\Lambda x(OFx > OFx)$ – spielte infolge ihres rein tautologischen Charakters keinerlei begründungsrelevante Rolle.

Will man über die implizite Universalität zu einem gehaltvolleren Begriff von Universalisierbarkeit kommen, sind derartige Trivialisierungsmöglichkeiten auszuschließen. Die in den partikulären Urteilen implizit enthaltenen universellen Urteile dürfen keine derartigen Tautologien sein. Das sei i. f. vorausgesetzt.

Ersparen wir uns weitere Zwischenüberlegungen – und fragen uns gleich, wie man das Ziel, Universalisierbarkeit in klar begründungsrelevanter Weise zu erklären, auf dem mit der Implizitheitstheese nun schon einmal begonnenen Wege wohl am schnellsten erreichen dürfte. Nun, eben so: Die Implizitheitstheese ist im Grunde eine Inferentialitätstheese. Denn: Daß ein partikuläres Urteil von der Form Ea implizit universell ist, heißt ja – und jetzt stellen wir das obige Definiens nur unwesentlich um – nichts anderes als:

Es gibt eine (von E verschiedene Eigenschaft) G, so daß gilt:

1. $\Lambda x(Gx > Ex)$
2. Ga

Und aus diesen beiden Prämissen folgt das fragliche Urteil Ea bereits prädikatenlogisch. Implizit universell ist ein Urteil also gerade insofern, als es die Schlußfolgerung eines Schlusses darstellt, in dem (wie man früher gesagt hätte: als Obersatz) ein universelles Urteil wesentlich vorkommt – was heißt, daß die betreffende Schlußfolgerung (auch unter Voraussetzung der 2. Prämisse, dem Untersatz) ohne das betreffende universelle Urteil nicht gezogen werden könnte. Kurz: Die These, daß jedes partikuläre deskriptive wie auch moralische Urteil implizit universell ist, besagt nichts anderes, als daß es zu jedem derartige Urteil einen solchen Schluß gibt – mit jeweils wahren Prämissen, versteht sich.

Die Schlußform, auf die bei dieser Inferentialitätstheese abgehoben wird, ist ganz klar eine deduktive. Und so liegt es nahe, sich an dieser Stelle an die sogenannten deduktiv-nomologischen Erklärungen als speziellen Fall einer deduktiven Begründung zu erinnern. Deduktiv-nomologische Erklärungen entsprechen, zumindest der von Hempel-Oppenheim

vertretenen Standardauffassung zufolge, in ihrer Form genau dem obigen Schlußschema – mit dem einen kleinen Unterschied: Es reicht für solche Erklärungen nicht, daß der Obersatz die Form einer universellen Aussage hat: Die betreffende Aussage muß darüber hinaus auch noch gesetzesartigen Charakter haben. D.h. insbesondere: Sie muß auch eine Begründung für irrationale Konditionalsätze liefern, also auch Aussagen stützen können, die bloß hypothetische Fälle betreffen. Kurz und für manchen Wissenschaftstheoretiker sicher erschreckend grob gesagt: Der Obersatz muß nicht nur wahr sein, er muß mit (kausaler) Notwendigkeit gelten. Soll die Implizitheits- bzw. Inferentialitäts-These in diesem Fall also tatsächlich eine Universalisierbarkeit in dem speziellen Sinne einer deduktiv-nomologischen Begründbarkeit liefern, so ist die 1. Prämisse also wie folgt um den Notwendigkeitsoperator N zu verstärken:

$$1^*: N \wedge x(Gx \supset Fx)$$

Je nachdem, ob diese Verstärkung vorgenommen wird oder nicht, spricht man dann auch von starker versus schwacher Inferentialität. Universalisierbarkeit im Sinne einer deduktiv-nomologischen-Begründbarkeit setzt also, wie gezeigt, starke Inferentialität voraus.

Was das mit der Moral zu tun hat? Nun, die Moral der eben eingeblendeten Kausal-Geschichte könnte die sein, daß eine genau parallele Verstärkung vielleicht ja auch für den universellen Obersatz einer von den Universalisierbarkeits-Theoretikern avisierten Begründung eines partikulären Moralurteils einschlägig ist. Auch eine solche Begründung soll ja schließlich eine Warum-Frage beantworten helfen. Nämlich die, warum jemandem etwas bestimmtes geboten ist. Sei das fragliche Moralurteil wiederum OFa, so besagt die Inferentialitätsthese, daß es eine Eigenschaft G gibt derart, daß sich OFa als Schlußfolgerung eines Schlusses mit den beiden folgenden wahren Prämissen ergibt:

1. $\wedge x(Gx \supset OFx)$
2. Ga

Zu fragen ist nun: Reicht diese Inferentialität als Begründung dafür, daß OFa, hin oder ist auch hier der Obersatz zu einer entsprechenden Notwendigkeit zu verstärken? Die Antwort ist: Aber freilich. Denn andernfalls ließe sich ja wiederum nicht ableiten, daß jemand, falls er die Eigenschaft G hätte, ebenfalls verpflichtet wäre, seinerseits F zu tun. Kurz: Andernfalls könnte man auch hier nicht sagen, daß a deshalb F tun soll, weil er die Eigenschaft G hat. G könnte, mit anderen Worten, nicht als GRUND dafür gelten, daß es a geboten ist, F zu tun.

Doch von welcher Art von Notwendigkeit ist in der verstärkten Prämisse

$$1^{**}: N \wedge x(Gx \supset OFx)$$

die Rede? Wie man 1^{**} zu lesen hätte, weiß ich wohl: Nämlich so, daß in jeder möglichen Welt gelte, daß es jedem mit der G-Eigenschaft geboten ist, F zu tun. Aber das sagt mir nicht, um welche Art von Notwendigkeit es dabei gehen soll. Eine analytische kann es, da das betreffende universelle Gebot ja ein inhaltliches sein soll, nicht sein. Eine kausale Notwendigkeit auch nicht. Eine epistemische – d.h. hier: eine auf die Sicht des Universalisierers bezogene? Oder eine spezifisch moralische? Nun: Ich weiß nicht, ich weiß nicht.

Lassen wir, wie so vieles andere, auch diese Frage jetzt einfach offen. Geschafft haben wir immerhin eines: Ein erstes Verständnis dessen, wie Universalisierbarkeit mit Begründbarkeit zusammenhängt. Nämlich schlicht so: WENN ein Urteil im Sinne strikter Inferentialität universalisierbar ist, DANN ist es auch begründbar. Aber diese Erkenntnis ist nun wirklich nichts besonderes. Denn schließlich haben wir ja den Universalisierbarkeits-Begriff genau so hingetrimmt, daß genau dieser Zusammenhang garantiert ist. Aber eines ist auch damit noch

nicht garantiert: Daß jedes Urteil- insbesondere also auch jedes Moralurteil tatsächlich in diesem Sinne universalisierbar ist.

Bei Wert und Moralurteilen wird die Universalisierbarkeit des weiteren auch mithilfe der sogenannten *Supervenienz* von Werteigenschaften zu erklären versucht. Hares (*Sprache und Moral*, S. 182) Beispiel:

Angenommen, wir sagen: „Der Heilige Franz war ein guter Mensch“. Es ist logisch unmöglich, das zu sagen und gleichzeitig zu behaupten, daß es einen anderen Menschen – in genau denselben Umständen wie der Heilige Franz gegeben haben könnte, der sich auch genau so wie der Heilige Franz verhielt, der sich jedoch von diesem in der einen Hinsicht unterschied, daß er kein guter Mensch war. Ich nehme natürlich an, daß das Urteil in beiden Fällen ein Urteil über das vollständige Leben des Betreffenden ist, das ‚innere‘ wie das ‚äußere‘ Leben.

Allgemeiner gesagt: Es ist nicht möglich, daß sich zwei Dinge in allen (relevanten) nicht-normativen Eigenschaften gleich sind, sich aber in ihren normativen Eigenschaften unterscheiden. Oder nochmals anders: Notwendigerweise gilt: Haben a und b die gleichen nicht-normativen Eigenschaften, dann auch die gleichen normativen. Die normativen Eigenschaften sind, kurz gesagt, bezüglich der nicht-normativen Eigenschaften supervenient.

Bezeichnen wir den Bereich der nicht-normativen Eigenschaften kurz mit D (wie deskriptiv), so können wir die Behauptung, daß OFa bezüglich D *supervenient* ist, präziser (für beliebige a und b) auch so ausdrücken:

$$(SV) \quad N(\Lambda G(G \in D > (Ga \equiv Gb)) > (OFa \equiv OFb))$$

Wie bereits Kim (1984) gezeigt hat, ist diese Forderung nun aber äquivalent mit

$$(SV') \quad N(OFa > \forall G(G \in D \& Ga \& \Lambda x(Gx > OFx)))$$

und das entspricht, wie man leicht sieht, genau der uns von oben schon bekannten schwachen Inferentialität.

Die uns bislang bekannte Supervenienz ist somit genauso schwach. Wie schwach diese ist, zeigt sich insbesondere wie folgt.

Angenommen, es gelte, die normative O-Eigenschaft sei supervenient bezüglich der- und-der deskriptiven D-Eigenschaften. Dann verträgt sich das voll und ganz etwa damit, daß z.B. (i) in dieser Welt zwar jeder mit den-und-den D-Eigenschaften auch die O-Eigenschaft hat, in einer anderen möglichen Welt aber keiner mit den entsprechenden D-Eigenschaften auch Träger der O-Eigenschaft ist; oder damit daß, (ii) In einer anderen möglichen Welt, die der unsrigen in Hinsicht auf die Verteilung der D-Eigenschaften genau gleicht, jeder die O-Eigenschaft hat; usw.

Das aber heißt mit anderen Worten: Wiederum kann man nicht sagen, daß die betreffenden D-Eigenschaften der *Grund* für das Vorliegen der O-Eigenschaft sind. Daß der Heilige Franz von Assisi ein guter Mensch war, das ist zwar supervenient bezüglich der Tatsache, daß er so und so gefühlt, gelebt und gehandelt hat. Aber diese Art von (wie gesagt: schwacher) Supervenienz berechtigt uns alleine noch nicht zu der Behauptung, daß er deshalb ein guter Mensch war, *weil* er so und so gefühlt, gelebt und gehandelt hat.

Nach den obigen Betrachtungen der Bedingungen für eine starke Inferentialität dürfte aber inzwischen jedem klar sein, wie wir dieses Manko der offensichtlich zu schwachen Supervenienz beseitigen können. Nämlich genauso wie das Manko der (für Begründungszwecke ebenfalls noch zu schwachen) schwachen Inferentialität. Und so verstärken wir denn auch die Supervenienzbestimmung entsprechend wie folgt:

(SV*) $N(\text{OFa} > \text{VG}(G \in D \ \& \ Ga \ \& \ N \ \wedge \ x(\text{Gx} > \text{OFx}))$

Und das entspricht evidentermaßen genau unserer starken Inferentialitätsbedingung. Aus den starken Inferentialitäts- wie Supervenienz-Forderungen folgen die schwachen, aber natürlich, (sonst wären schließlich die schwachen nicht schwächer) nicht umgekehrt. Und obwohl sich, wie gesagt, beide Fassungen der Supervenienz wie der Inferentialität jeweils genau entsprechen, sind Supervenienz und Inferentialität doch nicht dasselbe. Die Inferentialität ist für Urteile ganz generell definiert bzw. – wie wir auch sagen könnten, für alle Arten von Eigenschaften. Die Supervenienz hingegen ist nur für Paare unterschiedlicher Arten von Eigenschaften definiert – und zwar an dieser Stelle nur für normative Eigenschaften in Hinsicht auf deskriptive. In den Supervenienz-Definitionen ist daher der Bezug auf D enthalten, in den Inferentialitäts-Definitionen nicht. Insofern ist Inferentialität gegenüber Supervenienz allgemeiner.

Die übliche Universalisierbarkeits-Debatte geht freilich nicht von diesen Inferentialitäts- wie Supervenienz-Bestimmungen aus, sondern lediglich von der schon eingangs formulierten Universalisierbarkeitsthese U – wobei U natürlich nicht nur für ein Geboten-Sein, sondern auch für alle anderen normativen (Wert-) Eigenschaften sowie für beliebige a und b gelten soll:

U: Stimmen a und b hinsichtlich all ihrer (universellen) D Eigenschaften überein, so gilt: Ist es einem von beiden (unter den und den Umständen) geboten, F zu tun, so (unter den gleichen Umständen) auch jeweils dem anderen

U nenne ich i. f. auch einfach so, wie sie heißt: U – und nichts weiter. Und die in U enthaltene Bedingung, wonach a und b hinsichtlich all ihrer (universellen) D-Eigenschaften übereinstimmen sollen, nenne ich i. f. naheliegenderweise einfach die *Zwillingsbedingung*.

II LOGISCH – ODER WAS ?

Ein Großteil der U-Debatte kreist und kreist um die Frage, welchen Status U hat. Ist's eine logische These? Oder ist's ein inhaltliche moralische Norm? Nun, wie steht's damit? Kurz so: Auch diese Sache ist schon etwas komplizierter als man oft denkt.

Trotzdem ist die erste Frage zumindest heute ganz einfach zu beantworten. Nämlich mit einem klaren Nein. U ist heute keine logische Wahrheit. Aber das allein heißt nicht, daß U nicht eines Tages doch noch zu einer solchen Wahrheit werden könnte. Es heißt nur: Wir verwenden derzeit keine Logik, in der U bereits heute beweisbar wäre. Aber auch das könnte sich ändern.

Was wäre ein Grund für eine solche Änderung? Nun, Logiken haben den Zweck, analytische Wahrheiten einzufangen.

Nächste Frage also: Gibt es einen solchen Grund? Ist U analytisch wahr – so, wie „Junggesellen sind unverheiratet.“?

Ob U in einer Sprechergemeinschaft analytisch wahr ist, das finden erstens am besten nicht Philosophen, sondern Linguisten heraus. Zweitens hängt daran philosophisch nichts. Sprachkonventionen können auch unvernünftig sein, und wie das Wort „sollen“ von Kreti und Pleti verwendet *wird*, ist kaum ein Indiz dafür, wie es verwendet werden *sollte*. Insbesondere dürfen wir nicht vergessen, daß sich uns moralische Fragen in Entscheidungssituationen stellen und jedwede Ethik höchstens dann adäquat ist, wenn sie uns entscheidungsrelevante Erwägungen präsentiert, wofür sie letztlich auf unsere Präferenzen und Überzeugungen rekurrieren muß. Gehe ich mit einem linguistischen Schmetterlingsnetz durch die Welt und erhasche,

was mir an semantischen Brocken bezüglich des Wortes „sollen“ zu Ohren kommt, so ist es eine offene Frage, ob meine Beute das genannte Adäquatheitskriterium erfüllt und auch nur im entferntesten etwas mit entscheidungsrelevanten Aspekten zu tun hat. Dieses Adäquatheitskriterium ernstzunehmen und dadurch die meisten gordischen Knoten der normalsprachlich ausgerichteten Metaethik durchschlagen zu haben, ist gerade das zentrale Verdienst von Fehiges entscheidungstheoretischer Metaethik.

Allemaal hätte die Frage also zu lauten, ob wir U analytischen Status einräumen *wollen* oder nicht. Diese Frage stellt sich frühestens wieder am *Ende*, nicht im Verlaufe, von Gedanken über Us Witz, Us genaue Formulierung und Us Plausibilität, und angesichts der vielen Rätsel in diesem Bereich scheint es geraten, diese Anschlußfrage offen zu halten. Bedeutungswahrheiten haben nicht den Job, Plebiszite, sondern als fruchtbar erkannte Begriffe zu fixieren.

III UNIVERSALISIERBARKEIT – i.e. GENAU

Ein weiterer, und vielleicht der für jeden an Universalisierbarkeit Interessierten bedrückendste Problemkreis läßt sich in folgende Frage fassen: Gibt es überhaupt eine Chance, U in einer Weise zu formulieren, die U nicht gradewegs und offenkundig tautologisch, hanebüchen oder witzlos werden läßt?

Erwähnen wir zunächst Kutscheras Paradoxon. Angenommen, einer Person a sei es geboten, eine Handlung F zu tun. Für a's Vollzug der Handlung F schreiben wir „Fa“. Definieren wir nun das Prädikat H wie folgt:

$$Hx := (x = a \ \& \ x \text{ tut } F) \text{ oder } (x \neq a \ \& \ x \text{ tut } F \text{ nicht})$$

Angenommen, es gelte in einer bestimmten Situation das Gebot OFa. Betrachten wir dann eine zweite Person, b. Angenommen, diese Person und die Situation, in der sie steckt, sind Zwillinge von Person a und von der Situation, in der es a geboten ist, F zu tun. Us Zwillingensbedingung ist erfüllt. Und so gilt mit U wegen OFa auch (i) OFb. Da Fa nach der H-Definition äquivalent ist mit Ha, gilt also, da auch logische Folgen des Gebotenen geboten sein sollen, wegen OFa auch OHa. Nach U also auch OHb. Hb ist nun aber nach H-Definition äquivalent mit non-Fb. Also auch O(non-Fb), insbesondere also auch (ii) non-OFb – in klarem Widerspruch zu (i) OFb.

Das ist ein starkes Stück: Daraus, daß es in einer Situation Person a geboten war, F zu tun, konnten wir mit U herleiten, daß es a's Zwilling b in der Zwillingssituation geboten ist, gerade F nicht zu tun. O jemine! Wir könnten froh sein, wenn dies schlicht und einfach unplausibel wäre. Es ist aber nicht nur ein Schlag ins Gesicht unserer moralischen Überzeugungen, es ist ein Schlag ins Gesicht von U selbst.

„Aber ist H“, so könnte eingewendet werden, „in dessen Definition ja immerhin eine Individuenkonstante, also ein Eigename, auftritt, überhaupt eine universelle Eigenschaft?“ Dieser Einwand geht aus fundamentalen Gründen, die nichts mit Logeleien zu tun haben, am Ziel vorbei. Angenommen nämlich, H ist keine universelle Eigenschaft – was dann? Das Paradox bleibt. Damit U einschlägig ist und uns von der Gebotenheit von Ha auf die von Hb zu schließen erlaubt, müssen sich a und b resp. die Situationen, in denen sie sich befinden, in allen universellen Eigenschaften gleichen, sonst in nichts. Wenn sie sich auch in einer nicht-universellen Eigenschaft, wie, ex hypothesi, H eine ist, gleichen, nun, dann wird U dadurch nicht weniger einschlägig. (Genau wie ein weißer Kühlschrank nicht dagegen spricht, daß auch Schwäne weiß sind.) Im übrigen sagt niemand, daß sie sich bezüglich H gleichen, H taucht in diesem Paradoxon ja nur *im Gebot selbst* auf.

Hier könnte der Einwander nachsetzen und sagen: „Was soll es denn heißen, daß zwei Gegenstände eine nicht-universelle Eigenschaft gemein haben? Handelt es sich dann nicht nur um einen Gegenstand?“ Aber auch dieser Einwand geht daneben. Der Einwander selbst wollte ja H als nicht-universelle Eigenschaft bezeichnen, und daß zwei verschiedene Dinge beide die Eigenschaft H haben können, ist, wie ein kurzer Blick auf Hs Definition zeigt, eine logische Trivialität.

Hier ist noch ein Einwand: „Versteckt, nämlich in der Definition von H, tritt in H eine Individuenkonstante auf. U besagt, daß was einer Person a in einer Situation geboten ist, auch deren Zwillingperson in der Zwillingssituation geboten ist. Buchstabieren wir das gebotene Ha aber aus, so erhalten wir

$$H_a \equiv (a = a \ \& \ a \text{ tut } F) \text{ oder } (a \neq a \ \& \ a \text{ tut } F \text{ nicht}).$$

Um herauszufinden, was gemäß U nun auch b geboten ist, müssen wir *alle* Vorkommnisse von a in dieser ausbuchstabierte Version durch b ersetzen. Geboten ist b also die Handlung

$$(b = b \ \& \ b \text{ tut } F) \text{ oder } (b \neq b \ \& \ b \text{ tut } F \text{ nicht})$$

Also ist es b, genau wie a, geboten, F zu tun.“ An diesem Einwand ist etwas dran; nur will es nicht gelingen, präzise zu fassen, was. Es gibt keine offensichtliche Möglichkeit, in der Deontischen Logik das Auftreten von pathologischen Prädikaten wie H zu vermeiden oder die Forderung, immer mit den sozusagen ausbuchstabierte Prädikaten zu arbeiten, adäquat zu präzisieren.

Der Eindruck bleibt, daß in Kutscheras Paradoxon am Geist der Universalisierbarkeit vorbeigelogelt wird. Philosophie ist nun aber gerade die Kunst, sich von solchen Geisterbeschwörungen zu lösen, und somit ist dieser Eindruck argumentativ nicht viel wert. Er müßte umgesetzt werden in die Suche nach einer paradoxie-immunen Formulierung von U, und diese Suche ist sehr schwierig. Im Falle des Großen Bruders von Kutscheras Paradoxie, Goodmans Paradoxie der Induktion, ist sie schon seit einigen Jahrzehnten erfolglos.

Nun zu einem weiteren Problem. Das beste, was wir zum Thema Identität sagen können, geht wesentlich in Richtung des Leibnizschen Prinzips identitatis indiscernibilium: a ist identisch mit b genau dann, wenn a genau die universellen Eigenschaften hat, die auch b hat. Das ist sehr plausibel. Doch dann wird U nie und nimmer einschlägig. Es kann dann einfach, allein aus begrifflichen Gründen, keine zwei Situationen geben, die sich in Hinsicht auf alle universellen Eigenschaften gleichen: Gleich Situation a Situation b auf diese Weise, dann *ist sie auch schon dieselbe* (konkrete) Situation. U besagt dann nur, daß eine Situation so zu bewerten ist, wie sie selbst zu bewerten ist. Das ist tautologisch und nicht hilfreich.

„Aber können denn nicht zwei verschiedene Personen in derselben Situation sein?“ Nun, unterscheiden sich diese Personen hinsichtlich einer universellen Eigenschaft, dann ipso facto auch die Situationen, in denen der eine resp. der andere eine bestimmte Rolle spielt. Unterscheiden sich die Personen hinsichtlich keiner universellen Eigenschaft, so handelt es sich, nach Leibniz, um dieselbe Person.

Die Moral von Leibniz scheint zu sein, daß *alle* universellen Eigenschaften zu viele sind; und des Universalisierers Flucht vor Leibniz scheint die Form eines Us annehmen zu müssen, das auf weniger als alle universellen Eigenschaften rekurriert. Es ist vorgeschlagen worden, auf alle *moralisch relevanten* Eigenschaften zu rekurrieren; in der Tat scheint ja die Hoffnung begründet, daß in einer vernünftigen Ethik nicht alle universellen Eigenschaften moralisch relevant sind (z.B. plädiere ich nachdrücklich für die moralische Irrelevanz der Farbe der Socken, die jemand trägt); und gibt es auch nur eine universelle moralisch nicht-relevante Eigenschaft, so würde das Haben oder Nicht-Haben dieser Eigenschaft ja eine Per-

son von einer anderen oder eine Situation von einer anderen *unterscheiden*, und wir wären der Leibnizschen Gefahr entronnen. Ein dergestalt revidiertes Relevanz-U würde also besagen, daß, falls Person a in einer Situation etwas geboten ist, jeder Person, die in einer Situation, die der von a in jeder moralisch relevanten Hinsicht ähnelt, a's Rolle innehat, das gleiche geboten ist wie a in dessen Situation.

Aber das Relevanz-U ist wieder tautologisch: Es besagt, daß einen Unterschied für die moralische Bewertung nur das macht, was moralisch relevant ist. Aber „moralisch relevant“ *bedeutet* nichts anderes als „einen Unterschied für die moralische Bewertung machend“.

Die nächste Station auf der Flucht vor Leibniz wird also sein, U in nicht-tautologischer Weise von weniger als von allen universellen Eigenschaften reden zu lassen. Nichts einfacher als das: Benennen wir einfach die Eigenschaften, die unserer Ansicht nach moralisch relevant sind. Ein theonomer Moralist würde sagen: „Zwei Situationen, die sich hinsichtlich der Befolgung von Gottes Willen in nichts unterscheiden, müssen gleich bewertet werden.“ Ein rassistischer Spinner würde sagen: „Zwei Situationen, die sich nicht hinsichtlich der Anzahl getöteter Juden unterscheiden, müssen gleich bewertet werden.“ Ein Utilitarist würde sagen: „Zwei Situationen, die sich nicht hinsichtlich der Frage unterscheiden, wieviele Präferenzen welcher Stärke befriedigt werden, müssen gleich bewertet werden.“ Etc. So bekommen wir ein Gottes-U, ein Nazi-U, ein Utilitaristen-U und wer-weiß-wie-viele andere Us. Das Problem ist, daß nun die ganze Sache witzlos geworden ist. Sie hat keine interessante Struktur mehr. Jede auch noch so perverse Bewertung kann in diesem Sinne ihr eigenes U formulieren: Sie listet dazu ihre perversen Kriterien auf und behauptet, daß, wo hinsichtlich dieser Kriterien kein Unterschied besteht, auch kein moralischer Unterschied ist. Wieder scheint uns der Geist, mindestens der Witz, von U entglitten zu sein, ohne daß irgendjemand zu sagen wüßte, was dieser Geist oder Witz ist.

Kommen wir nun zu einem anderen Bereich, dem dessen, was aus U folgt, und wenden uns einer seiner heißdiskutierten Fragen zu. Angenommen, U ließe sich, allem, was wir gerade gehört haben, zum Trotz, befriedigend formulieren. Wäre dann, wer U akzeptiert, auch schon darauf festgelegt, Utilitarist zu sein? Ziemlich genau diese überraschende Behauptung verfiicht Richard Hare.

Die Zeit reicht hier nur für eine grobe Skizze von Hares Weg, oder vermeintlichem Weg, von U zum Utilitarismus. Hares These ist, daß, würde ein allwissendes und voll rationales Wesen universalisierbar wollen wollen, er, aus begrifflichen Gründen, auch utilitaristisch würde wollen wollen. In Hares Argumentation, die ich gleich an Hares eigenem Beispiel entwickeln möchte, spielt folgendes Prinzip P1 eine wichtige Rolle:

P1: Glaubt a, daß b eine Präferenz einer bestimmten Stärke dafür hat, daß p, so hat a selbst eine Präferenz gleicher Stärke dafür, daß, würde er in b's Haut stecken, p der Fall wäre

Formal: $G(a, P(b, p)) > P(a, a = b > . p)$

(Wobei $G(a, A)$ für: a glaubt, daß A; $P(a, A)$ für: a will (hat eine starke Präferenz dafür), daß A; und $>$ für die strikte Implikation stehen.)

Schauen wir uns nun einen beliebigen Interessenkonflikt an. Sagen wir, b hat eine starke Präferenz dafür, daß a's Fahrrad um einen Meter weiter nach links gestellt wird, damit er, b, am jetzigen Fahrradstandort sein Auto parken kann; a hingegen hat eine schwache Präferenz dafür, das Fahrrad stehen zu lassen, wo es ist. (Die Präferenzstärken könnten wir uns beispielsweise so erklären, daß bei b nur Faulheit im Spiel ist, weil das Fahrrad da, wo es steht, 1 Meter näher an seiner Haustür steht, bei a aber dringliche Parkplatznot mit Zeitdruck herrscht.) Nehmen wir nun an, daß a voll informiert ist, daß er universalisierbar wollen will und daß das Wollensprinzip P1 auf ihn zutrifft. Was dann? Nun, aus a's Vollinformiertheit und aus den Tatsachen folgt, daß a weiß, daß b eine starke Präferenz dafür hat, daß a's Fahr-

rad verschoben wird. Daraus, daß er das weiß, folgt allemal, daß er es glaubt. Daraus folgt mit P1, daß er selbst eine ebenso starke Präferenz dafür hat, daß, würde er in b's Haut stecken, das Fahrrad verschoben würde. Nun will er aber universalisierbar wollen: Er will, daß er für die Zwillingssituation, in der er in b's Haut stecken würde, das gleiche will wie für die tatsächliche Situation, in der er a ist. Zunächst ist, was er für die beiden Situationen will, noch nicht das gleiche: Für die tatsächliche Situation hat er eine schwache Fahrrad-Nicht-Verschiebe-Präferenz, für die Zwillingssituation eine starke Verschiebe-Präferenz. Eine von beiden Präferenzen muß er, will er als Universalisierer für beide Situationen das gleiche wollen, aufgeben. Welche? Nun, mittlerweile haben wir es mit einer prudentiellen, nicht mehr mit einer moralischen Frage zu tun: Gegeben zwei Präferenzen, von denen ich eine aufgeben will – welche aufzugeben ist das Vernünftigste? Ganz klar: Die schwächere. Unser vernünftiger a wird also die schwache Nicht-Verschiebe-Präferenz aufgeben und wollen, daß sein Fahrrad umgestellt wird. Analog werden, so Hare, in jedem Interessenkonflikt für den vollinformierten Universalisierer die Präferenzen der anderen zu eigenen Präferenzen; und im prudentiellen Bereich ist ja, anders als im moralischen, die Maximierung der Präferenzbefriedigung das unbestritten Vernünftigste.

Was ist von dieser, hier nur skizzierten, Theorie zu halten? Sicher gehört sie zu dem Pffiffigsten, was die Ethik-Geschichte aufzuweisen hat. Aber ist sie auch korrekt?

Ein (recht allgemeines) Problem besteht darin, daß weder Hare noch seine Anhänger die entscheidende Behauptung in der Präferenzlogik oder Entscheidungstheorie präzisiert oder bewiesen haben. Wir haben es mit normalsprachlichen Fahrrad- und Autogesichten mit zwei beteiligten Parteien zu tun und mit der Versicherung, der entscheidende Punkt sei ebenso klar wie die Übertragung auf beliebige andere Fälle mit beliebig vielen Beteiligten. Ein großes Problem bei solch einer Präzisierung (und ein echtes Problem der Theorie) ist bereits absehbar: Rationale Entscheidungstheorien gehen, sonst wären sie nicht rational, von Rationalitätsannahmen über die Präferenzen des Entscheiders aus. Den input, den Hare rationalentscheidend verarbeiten will, kann keine der handelsüblichen Entscheidungstheorien verarbeiten: Den nämlich, daß jemand eine starke Präferenz für p und eine weniger starke Präferenz für non-p hat. Hare hat angedeutet, was für Theorien ihm vorschweben. Er hat sich auf die Seite des Moralischen Equiprobabilismus geschlagen. Er behauptet, der allwissende universalisierende Woller würde rationalerweise die Welt wollen, deren Welt-Lotterie den höchsten Gewinnerwartungswert hat, wobei die zu einer Welt gehörige Weltlotterie diejenige ist, die den Spieler für jeden der, sagen wir n, Einwohner der Welt mit der Wahrscheinlichkeit $1/n$ das Los gewinnen läßt, ebendieser Einwohner in ebendieser Welt zu sein. Aber erstens ist der Moralische Equiprobabilismus selbst unterdeterminiert, denn es ist selbst bei voller Informiertheit nicht klar, was eine rationale Bewertung der zur Debatte stehenden Lebens-Lose ist. Zweitens ist auch schleierhaft, wieso Hare glaubt, daß ihn *seine Prämissen* zum Moralischen Equiprobabilismus *führen*.

Hare hat sich in der gleichen Angelegenheit und im gleichen Atemzug auch auf die Seite einer anderen Theorie geschlagen; wiederum, ohne zu sagen, wie er darauf kommt, daß der allwissende universalisierende Woller bei ihr landet, und auch ohne zu sagen, wie er zu der zweifelhaften Annahme kommt, daß diese Theorie auf das gleiche hinausläuft wie der Moralische Equiprobabilismus, also stets die gleiche Wahl induzieren wird wie jener. Diese zweite, auf C.I. Lewis zurückgehende Theorie sagt: Der Entscheider stellt sich am besten vor, in jeder der zur Debatte stehenden möglichen Welten die Rollen aller Beteiligten hintereinander durchleben zu müssen, und wählt dann die Welt, die dergestalt durchzuleben am angenehmsten ist. Außer an den erwähnten Problemen und daran, daß durch dieses Kriterium die Wahl wiederum unterdeterminiert ist, krankt die Theorie daran, daß mögliche Welten zeitlich *unbegrenzt* sind; die Vorstellung des Nacheinanderdurchlebens *aller* Leben in so einer Welt ist daher möglicherweise schlecht definiert; des weiteren kann so eine Methode Präferenzen ewigen Inhalts nicht gerecht werden, m. a. W.: Es kann die Frage, ob jemandes Präferenzen

ewigen Inhalts erfüllt sind oder nicht, nicht berücksichtigt werden. Das ist ein unerquicklicher Allgemeinheitsverlust und umso unerquicklicher, als die meisten von uns solche Präferenzen haben, einige davon recht stark: Ich wette zum Beispiel, daß Sie *nie* leiden und *immer* glücklich sein wollen, daß Sie wollen, daß, wenn möglich, das gleiche für jedes bewußte Wesen gilt, das je das Licht der Welt erblicken wird, und daß sie zu keinem Zeitpunkt als Verbrecher in Erinnerung sein möchten, etc.

Soviel zu Hares nicht-trivialen Präzisierung- und Beweislasten.

Grundsätzlicher ist die Frage, was es mit P1 auf sich hat. P1 ist das Kernstück der Argumentation und schafft die Präferenzen für die nicht-tatsächlichen Zwillingssituationen herbei. Beachten wir zunächst, daß P1 etwas ganz anderes ist als

P2 Glaubt a, daß b eine Präferenz einer bestimmten Stärke dafür hat, daß p, so hätte a, wenn er in b's Haut stecken würde, eine Präferenz gleicher Stärke dafür, daß p der Fall wäre

Formal: $G(a, P(b, p)) > (a = b > . P(b, p))$

P2 ist, zumindest in einer naheliegenden Lesart des kontrafaktischen Konditionalsatzes, wahr: klar, wenn a in b's Haut stecken würde, dann hätte er eben alle universellen Eigenschaften von b. (Von den Leibnizschen Problemen, die das aufwirft, wollten wir ja zeitweise einmal absehen.) Und da b die besagte Präferenz hat, hätte dann auch a sie. Das plausible P2 kann die von Hare benötigte Brücke natürlich nicht abgeben. Es besagt nicht, daß a irgendwelche Präferenzen bezüglich der Zwillingssituation *hat*, sondern daß er sie unter bestimmten Bedingungen *hätte* (von denen wir sogar schon wissen, daß sie nicht vorliegen). Wir brauchen aber eine tatsächliche Präferenz, die beim Universalisierer a einen tatsächlichen Sieg über seine tatsächliche schwache Fahrrad-Nicht-Verschiebe-Präferenz davontragen könnte.

In der Hinsicht ist P1 hilfreicher, aber, leider, auch weniger plausibel als P2. P1 ist, so Hare, eine Bedeutungswahrheit, zumindest für den Begriff des rationalen Präferierens. Das Konsequens von P1 redet von einer Präferenz bezüglich eines kontrafaktischen Konditionalsatzes, um dessen Kontrafaktizität der Präferierende weiß (denn er weiß ja, daß er er und nicht jemand anderes ist). Schauen wir uns die beste handelsübliche Semantik für kontrafaktische Konditionalsätze an. Sie besagt, daß ein Satz „Wenn p der Fall wäre, dann auch q“ wahr ist genau dann, wenn in allen unserer Welt ähnlichen Welten die materiale Implikation „Wenn p, dann q“ wahr ist. Nun, ich habe überhaupt keine Präferenz hinsichtlich der Wahrheit von Implikationen in möglichen Welten, die nicht die tatsächliche Welt sind. Ich habe sogar überhaupt keine Präferenz hinsichtlich von Sachverhalten in anderen möglichen Welten. Genausowenig, wie ich Mitleid mit einem möglichen Kind habe, das verhungert – nicht, solange es nicht auch ein tatsächliches Kind ist. Hungersnöte in bloß möglichen Welten lassen mich kalt, Lottogewinne auch. Mich juckt nur *eine* mögliche Welt, nämlich die tatsächliche, und ich komme mir dabei nicht sonderlich irrational vor. Um P1, Hares entscheidende Brücke, ist es schlecht bestellt.

Ich fasse zusammen. Aus der bunten Vielzahl von Fragen, die sich in der Ethik um das Universalisierbarkeitsprinzip ranken, habe ich einige, aus verschiedenen Bereichen, herausgegriffen. Ich habe, erstens, gefragt, was Universalisierbarkeit mit Begründbarkeit zu tun hat. Die Antwort war: kompliziert. Ich habe, zweitens, gefragt, ob U eine logische These ist. Die Antwort war: Derzeit nicht. Ich habe, drittens, gefragt, ob sie eine Bedeutungswahrheit ist. Die Antwort war, in der einen, empirischen Lesart der Frage: Mag sein – doch wen juckt's?; in der anderen, kritischen Lesart: Mal seh'n. Ich habe, viertens, gefragt, ob für U überhaupt eine Formulierung zu finden ist, die uns nicht geradewegs in himmelschreiende Inadäquatheiten stürzt, und die Antwort war: Vermutlich nicht. Ich habe, fünftens, gefragt, ob aus

U der Utilitarismus folgt, und die Antwort war: Nach dem prominentesten Herleitungsvorschlag zu urteilen: nein.

Fazit vom Fazit: Universalisierbarkeit ist und bleibt eine harte Nuß.

Literatur:

Fehige, C., *Soll ich?*, Berlin-New York, 1992.

Hare, R. M., *The Language of Morals*, Oxford 1952; dtsh. *Die Sprache der Moral*, Frankfurt 1983.

Hare, R. M., *Freedom and Reason*, Oxford, 1962; dtsh. *Freiheit und Vernunft*, Frankfurt, 1983.

Kim, J., *Concepts of Supervenience*, in: *Philosophy and Phenomenological Research*, 45, 1984, S. 153-176.

Kutschera, F. von, *Grundlagen der Ethik*, Berlin-New York, 1982.