



# Die Mannigfaltigkeit möglicher Maximen als Problem für Kants Theorie des obersten Prinzips der Moral

Jasper Lohmar 

Angenommen: 15. August 2022 / Online publiziert: 21. September 2022  
© Der/die Autor(en) 2022

**Zusammenfassung** Diese Untersuchung widmet sich einem Problem, das sich für Kants Konzeption des kategorischen Imperativs aus der Möglichkeit einer unterschiedlich feinkörnigen Individuation von Handlungstypen bei der Bildung von Maximen ergibt. Das hier präsentierte Argument aus der Mannigfaltigkeit möglicher Maximen entwickelt das bekannte „Problem of Relevant Descriptions“ weiter und verschärft es. Ich argumentiere, dass es Fälle gibt, in denen dieselbe Handlung sowohl unter einer universalisierbaren als auch unter einer nicht universalisierbaren Maxime vollzogen werden kann, sodass Kants Moralthorie, unter der Annahme der Notwendigkeit moralischer Tatsachen, impliziert, dass es Handlungen gibt, die zugleich erlaubt und nicht erlaubt sind. Die Diskussion ergibt, dass eine hinreichende Begrenzung des Maximenbegriffs für das Ausschließen eines heterogenen Auswertungsergebnisses zweier Maximen, die derselben Handlung zugrunde liegen können, ohne ad hoc-Bestimmungen unmöglich ist. Folglich ist eine Modifikation von Kants Moralthorie erforderlich, um zumindest ihre Konsistenz sicherzustellen. Die Untersuchung endet mit der Diskussion zweier solcher Modifikationen: Zunächst untersuche ich den Lösungsvorschlag, den moralischen Status von Handlungen nur Maximen-relativ zu bestimmen und die Annahme aufzugeben, dass Handlungen *tout court* erlaubt oder verboten sind. Ein solcher Schritt stellt jedoch einen tiefen Eingriff in Kants Moralphilosophie dar und geht, wie ich zeigen werde, seinerseits mit gravierenden Problemen einher. Anschließend diskutiere ich die Möglichkeit, die Universalisierungsformel des kategorischen Imperativs dahingehend zu modifizieren, dass sie nicht bloß über aktuelle, sondern über alle möglichen Maximen quantifiziert. Obgleich auch dieser Lösungsvorschlag problembehaftet ist, erweist er sich im Rahmen dieser Untersuchung als der aussichtsreichste Ansatz.

---

Jasper Lohmar (✉)  
Fakultät für Geisteswissenschaften, Philosophisches Seminar, Universität Hamburg,  
Überseering 35, 4, 22297 Hamburg, Deutschland  
E-Mail: [jasper.lohmar@uni-hamburg.de](mailto:jasper.lohmar@uni-hamburg.de)



**Abstract** In this paper I highlight and discuss a problem for Kant's conception of the categorical imperative that arises from the possibility of a differently fine-grained individuation of act types in the formation of maxims. The "Problem from the Manifold of Possible Maxims", as it might be called, further develops and exacerbates the well-known "Problem of Relevant Descriptions." In particular, I argue that there are cases in which the same act can be performed both under a universalizable and under a non-universalizable maxim. But then Kant's moral theory implies, given the necessity of moral truths, that there are acts that are both permissible and impermissible. The discussion reveals that a limitation of the notion of maxims sufficient to rule out a heterogeneous evaluation of two maxims that can underlie the same action is impossible without ad hoc stipulations. Consequently, a modification of Kant's moral theory is required to at least ensure its consistency. The investigation ends with a discussion of two such modifications: First, I examine the suggestion of determining the moral status of actions only relative to some maxim and abandoning the assumption that actions are permissible or impermissible tout court. Such a move, however, is too deep an intervention in Kant's moral philosophy to count as a suitable modification and, as I will show, faces serious problems on its own. I then discuss the possibility of modifying Kant's universalization formula so that it quantifies not merely over actual maxims, but over all possible maxims. Although this proposed solution is also problematic, it turns out to be the most promising approach in the context of this investigation.

**Keywords** Kant · Kategorischer Imperativ · Maximen · Problem of Relevant Descriptions · Universalisierungsformel · Universality Test

## 1 Einleitung

Im Mittelpunkt der Moralphilosophie Kants steht, als oberstes Prinzip der Moral, der kategorische Imperativ (KI), der uns, als praktisches reines Vernunftgesetz zugleich Seins- und Erkenntnisgrund moralischer Pflichten, in seiner Grundform vorschreibt, nur nach derjenigen Maxime zu handeln, durch die man zugleich wollen kann, „daß sie ein allgemeines Gesetz werde“ (GMS, AA 04: 421.08). Es ist also die Universalisierbarkeit von Maximen, verstanden als subjektive praktische Grundsätze, die den moralischen Status einer Handlung, d. h., ob sie erlaubt oder verboten, pflichtgemäß oder pflichtwidrig ist, *bestimmt* und *erkennbar macht*.<sup>1</sup>

Insofern von ihnen vermittelt dieses obersten Prinzips der Moral epistemisch und ontologisch einzelne moralische Pflichten und Verbote (gar für konkrete Handlungssituationen) abgeleitet werden sollen, ist die Struktur der Maximen von zentraler Bedeutung für die kantische Ethik. Wie Christel Fricke (2008: 125) schreibt, ist es

<sup>1</sup> Von dem moralischen Status in dem hier eingeführten Sinne ist zu unterscheiden, was wir den *moralischen Wert* einer Handlung nennen können: Dass zwei Handlungen pflichtgemäß oder -widrig, erlaubt oder verboten sind garantiert nicht, dass sie denselben moralischen Wert haben. Auf diese Unterscheidung gehe ich in den Abschnitten 3 und 4.2 ausführlicher ein.

jedoch „notorisch unklar [...], was genau unter ‚Maximen‘ zu verstehen ist und wie ‚Maximen‘ in einer plausiblen Moralphilosophie untergebracht werden können“.

Ein zentrales Problem der Moralphilosophie Kants im Zusammenhang mit dem Maximenbegriff ergibt sich daraus, dass einer Handlung verschiedene Maximen mit unterschiedlichen Bestimmungen des Handlungstypus zugrunde liegen können.<sup>2</sup> Einem Beispiel von David Ross (1954) entsprechend könnte etwa das Äußern einer Lüge, so scheint es, auf Basis einer Maxime erfolgen, die den Handlungstypus als „einen mordwilligen Menschen anlügen“, „lügen“ oder „eine Aussage treffen“ charakterisiert. Dies ist insofern problematisch, als es *prima facie* möglich erscheint, dass der KI relativ zu den verschiedenen Maximen unterschiedliche Ergebnisse bezüglich des moralischen Status der Handlung liefert: Wenn von einer möglichen Maxime einer Handlung gedacht und gewollt werden kann, dass sie allgemeines Gesetz werde, von einer anderen, die derselben Handlung zugrunde liegen kann, hingegen nicht, ergibt sich insofern ein Problem für Kants Theorie, als ein und dieselbe Handlung folglich zugleich verboten und erlaubt wäre. Diese Konsequenz ist jedoch für eine Moralthorie inakzeptabel.<sup>3</sup>

Als „Problem of Relevant (Act) Descriptions“ (PRD) ist das verwandte und viel diskutierte Problem der kantischen Moralphilosophie bekannt, die Anwendbarkeit des Universalisierbarkeits-Tests trotz der unzähligen Möglichkeiten der Beschreibung einer Handlung bei der Bildung einer Maxime zu gewährleisten, also das Kriterium für die relevante Beschreibung der Handlung zu finden, die als Input für das Testverfahren dient. In der vorliegenden Untersuchung liegt der Fokus jedoch, anders als in der Diskussion des PRD (und entsprechend auch bei Ross), nicht vorrangig auf dem KI als *Erkenntnis*-, sondern als *Seinsgrund* des moralischen Status einer Handlung. Nicht die Mannigfaltigkeit der möglichen Handlungsbeschreibungen, aus denen bei der *Bildung* einer Maxime zu wählen ist, bzw., im Falle einer retrospektiven Anwendung des KI: die Identifikation der *aktualen* Maxime ist Ausgangspunkt des von mir diskutierten Problems. Vielmehr geht es in diesem Aufsatz um die Mannigfaltigkeit der Maximen, die einer Handlung zugrunde liegen *können*, mithin um die Menge *möglicher* Maximen. Dieses Problem bleibt bestehen, auch wenn sich die Frage, welche Handlungsbeschreibung für die Anwendung des Tests

<sup>2</sup> Das im Folgenden entwickelte Problem ist zwar auf Kants Moralphilosophie zugeschnitten, ist aber auch für andere Moralthorien relevant, die der Beschreibung einer Handlung eine entscheidende Bedeutung für die Bestimmung von deren moralischem Status zuschreiben, ohne Kriterien zu liefern, die die Mannigfaltigkeit der möglichen relevanten Handlungsbeschreibungen hinreichend einzuschränken. Die Tragweite des Problems für andere Moralthorien genauer zu suchen kann in diesem Aufsatz jedoch nicht geleistet werden.

<sup>3</sup> Verschiedene Varianten dieses Problems wurden in den vergangenen Jahrzehnten viel diskutiert. So stellt etwa Henry Allison fest: „As many have pointed out, the same course of action, say telling a lie or making a false promise, could fall under a variety of descriptions, some of which allow for and others appear to exclude universalization.“ (Allison 2011: 192). Insofern es mir um ein Problem bezüglich des moralischen Status von Handlungen oder Handlungstypen geht, steht meine Analyse und Problementwicklung jedoch im Widerspruch zu Allison's These, „that FLN [Die Naturgesetzformel] is conceived by Kant as a test of the permissibility of maxims and concerns the maxim on the basis of which an agent is actually contemplating a course of action, not, as Ross seems to have assumed, an action“ (ebd.). Nach meinen Ausführungen dient der Universalisierbarkeits-Test (zumindest auch) dem Überprüfen von *Handlungen* auf ihren moralischen Status, unter Rückgriff auf die ihnen zugrunde liegenden Maximen, nicht (bloß) der Überprüfung des moralischen Status der Maxime selbst. Für eine ausführliche Entwicklung des Problems siehe Abschnitt 3.

verwendet werden sollte, zufriedenstellend beantworten und das PRD entsprechend lösen lässt.<sup>4</sup>

Im Folgenden werde ich, im Sinne einer Untersuchung dieser Problemlage, zunächst die Begriffe der moralischen Pflicht, des kategorischen Imperativs und der *Maxime*, ihre wechselweisen Beziehungen und ihre Stellung im Rahmen der kantischen Moralphilosophie auf Grundlage der *Kritik der praktischen Vernunft* (KpV) und der *Grundlegung zur Metaphysik der Sitten* (GMS) in gebotener Kürze rekonstruieren. Im darauffolgenden dritten Abschnitt präsentiere und entwickle ich auf Basis dieser Rekonstruktion der Grundzüge von Kants Moralphilosophie das Argument aus der Mannigfaltigkeit möglicher Maximen und grenze es vom PRD ab.

Der vierte Abschnitt ist der Diskussion verschiedener Lösungsansätze gewidmet, die darauf abzielen, zumindest die Konsistenz der kantischen Ethik angesichts des entwickelten Problems sicherzustellen: Der naheliegende Versuch, die Kriterien zur Beschränkung der Mannigfaltigkeit möglicher Maximen abzuleiten oder zu entwickeln, um ein heterogenes Ergebnis der Bewertung von Handlungen anhand des KI auszuschließen, erweist sich als erfolglos. Ich werde zeigen, dass eine hinreichende Begrenzung des Maximenbegriffs bei Kant ohne *ad hoc*-Bestimmungen unmöglich ist und folglich eine Modifikation seiner Moraltheorie für die Konsistenz seiner Annahmen erforderlich ist. Zwei Möglichkeiten einer solchen Modifikation werde ich diskutieren: Erstens untersuche ich den Lösungsvorschlag, dass sich der moralische Status einer Handlung ausschließlich in Relation zur tatsächlich zugrunde liegenden *Maxime* bestimmen lässt. Dieser Ansatz ist jedoch revisionär und problembehaftet. Zweitens diskutiere ich die Möglichkeit, die Universalisierungsformel des KI dahingehend zu modifizieren, dass sie nicht bloß über aktuelle, sondern über alle möglichen Maximen quantifiziert. Ein solcher Ansatz hat einige gewichtige Vorteile, und die Probleme, die sich im Zuge der Diskussion ergeben, reichen nicht aus, um ihn final zurückzuweisen. Insofern liefert die Untersuchung zumindest einen produktiven Ausgangspunkt für die weitere Debatte.

## 2 Maximen und der kategorische Imperativ in Kants Moralphilosophie

Praktische Grundsätze sind in der kantischen Terminologie *allgemeine* Bestimmungen des *Willens*.<sup>5</sup> Anders als im theoretischen sind Grundsätze im praktischen Vernunftgebrauch nicht zwangsläufig Gesetze (die Notwendigkeit voraussetzen), da hier nicht bloß die „Beschaffenheit des Objects“ (KpV, AA 05: 20.01) bestimmend ist, sondern ein Subjekt mit einem Begehrungsvermögen.<sup>6</sup> Kant unterscheidet entsprechend zwischen *subjektiven* und *objektiven* praktischen Grundsätzen. Während letztere für *alle* Vernunftwesen gelten, sind erstere bloß *Maximen*, insofern ihre „Be-

<sup>4</sup> Für eine ausführliche Darstellung des PRD und einen Überblick über die dafür relevante Literatur siehe Abschnitt 3.2 dieses Aufsatzes, in dem ich eine Abgrenzung des von mir diskutierten Problems vom PRD vornehme.

<sup>5</sup> Vgl. KpV, AA 05: 19.

<sup>6</sup> Vgl. KpV, AA 05: 20.

dingung nur als für den Willen des Subjects gültig von ihm angesehen wird“ (KpV, AA 05: 19.09-10).

Maximen sind also *subjektive praktische Grundsätze* und als solche selbstgewählt. Nach dieser recht abstrakten Bestimmung ist für eine erste Annäherung das Betrachten zweier Beispiele für Maximen hilfreich, die Kant selbst anführt:

ich mache es mir aus Selbstliebe zum Princip, wenn das Leben bei seiner längern Frist mehr Übel droht, als es Annehmlichkeit verspricht, es mir abzukürzen. (GMS, AA 04: 422.04-06)

wenn ich mich in Geldnoth zu sein glaube [sic], so will ich Geld borgen und versprechen es zu bezahlen, ob ich gleich weiß, es werde niemals geschehen. (GMS, AA 04: 422.20-23)

Christian Illies formalisiert die Struktur von Maximen folgendermaßen: „Wenn es eine Situation X gibt, dann will ich Y tun, um Z zu erreichen“ (Illies 2007: 309. Vgl. auch Allison 2011: 97 und Timmons 1984: 162). Da diese Formulierung potenziell missverständlich ist, ist es wichtig, a) den (selbst auferlegten) *regelhaften* Charakter der Maximen im Blick zu behalten, der über bloße Absichtsbekundungen hinausgeht, und b) zu beachten, dass, etwa bei Maximen wie „ich will alle meine Versprechen halten“ die X- und Z-Variablen keinen bestimmten und explizierten Wert annehmen müssen.

Im Unterschied zu Maximen sind objektive praktische Grundsätze nicht selbstgewählt, sondern Nötigungen der Vernunft, deren normative Kraft bei Entitäten, die Vernunft- und Sinneswesen zugleich sind, „durch ein Sollen [...] bezeichnet [werden]“ (KpV, AA 05: 20.10-11). Aus diesem Grunde sind sie für uns *Imperative*.<sup>7</sup> Sofern die objektiven praktischen Grundsätze *bedingt* sind, handelt es sich um *hypothetische Imperative*, die bloß objektive praktische *Vorschriften* sind – sie haben die Form: „Wenn du X willst, sollst du Y tun!“. Da das Antezedens des Konditionals bei hypothetischen Imperativen durch seine Abhängigkeit vom Begehrungsvermögen bloß empirischen Charakter hat, und Notwendigkeit in Kants Verständnis grundsätzlich *a priori* ist, können hypothetische Imperative keine praktischen Gesetze sein.<sup>8</sup> Damit ergibt sich als weitere Bestimmung: Sofern objektive praktische Grundsätze *unbedingt* sind, handelt es sich um *kategorische Imperative*, die allein praktische *Gesetze* sind – sie haben die Form: „Du sollst y tun!“.<sup>9</sup>

<sup>7</sup> Ein reines Vernunftwesen, etwa Gott oder ein Engel, handelt *notwendigerweise* vernunftgemäß, wodurch praktische Gesetze für ein solches Wesen Naturgesetzen gleichen (wobei die Autonomie anders als im Falle der heteronomen Bestimmung durch Naturgesetze im Falle praktischer Gesetze durch diese Notwendigkeit nicht eingeschränkt wird) und keine normative Kraft haben. (Vgl. GMS, AA 04: 412f. und KpV, AA 05: 32.).

<sup>8</sup> Vgl. KpV, AA 05: 21.

<sup>9</sup> Nach diesen Erläuterungen ist klar, dass man sich objektive praktische Grundsätze zu subjektiven, also zu Maximen machen kann: So kann ich ausgehend vom kategorischen Imperativ „Du sollst deine Versprechen halten!“ die Maxime bilden „Ich will meine Versprechen halten.“.



## 2.1 Maximen als subjektive praktische Grundsätze

Bevor wir mit der Diskussion des Problems der Individuation von Handlungstypen bei der Bildung von Maximen beginnen können, sind noch einige Eigenschaften von Maximen als *subjektive praktische Grundsätze* herauszuarbeiten:

1.: Als *Grundsätze* sind Maximen nicht nur *allgemein*<sup>10</sup>, sondern auch besonders *wichtig* für das Subjekt und *dauerhaft*<sup>11</sup>. Als selbstgesetzte Grundregeln für unser Handeln sind sie nicht darauf ausgelegt, von Situation zu Situation variiert zu werden (auch wenn wir Maximen aufgeben und neue bilden können).

Fricke räumt einen möglichen Irrtum aus, indem sie feststellt: „Maximen sind, so Kant, nicht bestimmte, einzelne, raum-zeitlich lokalisierbare geistige Zustände einer Person“ (Fricke 2008: 125), sie sind also abzugrenzen von Absichten u. Ä. (vgl. ebd.: 126). Als selbstgesetzte Regeln<sup>12</sup> werden Maximen zwar zu einem bestimmten Zeitpunkt *gebildet* und bei Handlungsentscheidungen greifen wir zu einem bestimmten Zeitpunkt auf sie zurück, sie sind aber selbst keine zeitlichen Ereignisse oder Zustände, sondern allgemeine, regelhafte Sätze. Michael Albrecht weist entsprechend darauf hin, dass es sich bei Maximen nicht um Gewohnheiten oder Handlungsdispositionen handelt (auch wenn diese folgen mögen), sondern um autonom aus Freiheit gesetzte Regeln der eigenen Willensbestimmung (vgl. Albrecht 1994: 136 f.).

2.: Als *praktische Grundsätze* sind Maximen *willensbestimmend* und *handlungsanleitend*. In einer vielzitierten Fußnote in der GMS bezeichnet Kant Maximen daher auch als „das subjective Princip des Wollens“ (GMS, AA 04: 400) bzw. „das subjective Princip zu handeln“, die „die praktische Regel [enthalte], die die Vernunft den Bedingungen des Subjects gemäß [...] bestimmt, und ist also der Grundsatz, nach welchem das Subject *handelt*“<sup>13</sup> (GMS, AA 04: 420-421). Maximen sind also diejenigen Prinzipien, die unser Wollen und Handeln *tatsächlich* bestimmen.<sup>14</sup> Nach gängiger Auffassung liegen Handlungen nach Kant somit *immer* Maximen zugrunde. Nur so handelt es sich bei einem Tun um das Ergebnis vernünftiger Überlegung, was es erst zur Handlung qualifiziert.<sup>15</sup> Hier zeigt sich, dass Kant einen anspruchs-

<sup>10</sup> Vgl. auch: Beck 1966: 78 f. Dass dieses Attribut durchaus problematisch ist, diskutiert Albrecht 1994: Fn. 39.

<sup>11</sup> Vgl. Albrecht 1994: 131.

<sup>12</sup> Strittig ist, ob es sich bei Maximen (zumindest teilweise) um selbstgesetzte *Normen* handelt, oder ob sie eher deskriptiv, bzw., als prospektive Willensbekundung, *expressiv* zu verstehen sind. Der hier verwendete Regel-Begriff ist entsprechend vorerst als in dieser Hinsicht *neutral* zu verstehen und soll lediglich betonen, dass Maximen als Prinzipien *regelhaft* sind und über die einzelne Situation hinaus gehen.

<sup>13</sup> Passagen, die im Original gesperrt geschrieben sind, werden hier und im Folgenden stattdessen durch Kursivierung hervorgehoben.

<sup>14</sup> Um Missverständnisse zu vermeiden, sollte an dieser Stelle darauf hingewiesen werden, dass Kant zwischen *Begehrungsvermögen* und *Willen* unterscheidet, insofern letzterer *Vernunft* voraussetzt: „Nur ein vernünftiges Wesen hat das Vermögen, *nach der Vorstellung* der Gesetze, d. i. nach Principien, zu handeln, oder einen *Willen*.“ (GMS, AA 04: 412.26-28, vgl. auch ebd. 446) Maximen bestimmen also nicht alles Begehren, wohl aber den Willen.

<sup>15</sup> Illies schreibt in ähnlichem Sinne: „Maximen lassen sich für alle Handlungen angeben, sofern sie nur frei gewählt sind und bewußt ausgeführt werden.“ (Illies 2007: 309).



vollen Handlungs-Begriff voraussetzt: Unmittelbares Tun aus Gewohnheit, Instinkt oder Neigung ohne Rückgriff auf eine Maxime stellt für ihn keine Handlung dar.<sup>16</sup>

Da der Vorwurf einer unplausibel anspruchsvollen Handlungstheorie angesichts dieser These naheliegt, ist zu ergänzen, dass uns bei unseren Handlungen die zugrundeliegenden Maximen Kant zufolge nicht zwangsläufig jederzeit bewusst und transparent sind. Onora O'Neill schreibt: „they may be very elliptically formulated“ (O'Neill 2013: 99) und Fricke weist darauf hin, dass Maximen „nur in den seltensten Fällen von den Personen, die sie sich zu eigen gemacht haben, ausdrücklich formuliert“ werden (Fricke 2008: 130).

Aus der Annahme, dass jeder Handlung notwendigerweise eine Maxime zugrunde liegt, folgt jedoch nicht, dass individuelle Maximen konstitutiv für Handlungen sind oder Handlungen anhand der ihnen zugrunde liegenden Maximen individuiert werden.<sup>17</sup> Handlungen werden vielmehr vor dem Hintergrund einer Maxime (als allgemeine, handlungsleitende Regel) instrumentell genutzt. Entsprechend kann es zu einer Handlung verschiedene Maximen geben. Um zum Ausgangsbeispiel von Ross zurückzukehren: Bei mehreren möglichen Äußerungen der fraglichen Lüge auf Grundlage jeweils unterschiedlicher Maximen, kann es sich trotzdem um ein und dieselbe Handlung handeln (bzw. um ein und denselben Handlungstypus, der so fein individuiert ist, dass alle möglichen Instanziierungen desselbigen denselben moralischen Status teilen). Dies gilt auch, wenn sich in diesen Maximen, die schematisch den Zweck benennen, dem die Lüge dient, unterschiedlich grob- und feinkörnigen Beschreibungen der Handlung finden.<sup>18</sup>

3.: Als *subjektive* praktische Grundsätze sind Maximen *selbstgewählt* und an ein bestimmtes Subjekt gebunden. Deshalb sind sie nicht universell, sondern *bloß subjektiv gültig*.<sup>19</sup> Sowohl Ursprung als auch Wirkbereich der Maxime ist somit das Subjekt: Ich wähle meine Maximen und die Maximen bestimmen meinen Willen.

Die Abhängigkeit der Maximen von einem Subjekt mit Begehrungsvermögen bringt des Weiteren mit sich, dass sie, anders als praktische Gesetze, „ein *Object* (Materie) des Begehrungsvermögens als Bestimmungsgrund des Willens voraussetzen“ (KpV, AA 05: 21.14-15): „Alle Maximen haben nämlich [...] eine *Materie*, nämlich einen Zweck“ (GMS, AA 04: 436.13-19). Sie sind somit abhängig vom Begehrungsvermögen, das bestimmte Zwecke setzt.

<sup>16</sup> Fricke (2008) verteidigt diese Interpretation in ihrem Aufsatz: Maximen stünden als Auswahlkriterien im Hintergrund jeder praktischen Überlegung und damit jeder zurechenbaren Handlung. Rüdiger Bittner schreibt entsprechend: „Zentral für Kants Handlungsverständnis ist die These, daß, wer handelt, nach einer Maxime handelt.“ (Bittner 2005: 54). Eine ausführliche Verteidigung dieser weit verbreiteten „Maximenthese“ findet sich in der Dissertation von Tobias Kronenberg (2016). Kritik kommt u. a. von Michael Albrecht (1994: 136f.).

<sup>17</sup> Ein solcher Versuch, Handlungen über Maximen zu individuieren, würde Kants Unterscheidung zwischen pflichtgemäßen Handlungen aus Pflicht und solchen aus Neigung hinfällig machen oder zumindest ihre Begründung erheblich verkomplizieren, siehe 3 und 4.2.

<sup>18</sup> Ausführlicher verteidige ich diese These ab Abschnitt 3 in Form von P1 des dort vorgestellten Arguments.

<sup>19</sup> Vgl. Fricke 2008, 131f. Dies bedeutet selbstverständlich nicht, dass nicht mehrere Personen dieselbe Maxime als Handlungsprinzip wählen können. Selbst wenn *alle* Personen dieselbe Maxime wählen, ist diese in dem Sinne bloß subjektiv gültig, dass ihre Gültigkeit für ein Subjekt davon abhängt von ebendiesem gewählt worden zu sein.



## 2.2 Der kategorische Imperativ als oberstes Prinzip der Moral

Praktische Gesetze im Sinne Kants müssen aufgrund ihrer objektiven Notwendigkeit „durch Vernunft *a priori*, nicht durch Erfahrung [...] erkannt sein“ (KpV, AA 05: 26.22-23). Nun sind Objekte des Begehungsvermögens (deren „Wirklichkeit“ begehrt wird) jedoch immer empirisch (wie auch in der Konsequenz die praktischen Prinzipien, die ein solches Objekt zur Willensbestimmung voraussetzen – wie etwa Maximen, in Form der oben erwähnten Zweckvariable). Daraus folgt, dass praktische Gesetze, wenn es sie gibt, ohne Bezugnahme auf solche, „nicht der Materie, sondern bloß [sic!] der Form nach, den Bestimmungsgrund des Willens enthalten“ (KpV, AA 05: 27.05-06). Durch diesen Ausschluss jedes inhaltlichen Bestimmungsgrundes enthält das Gesetz „keine Bedingungen [...] auf die es eingeschränkt [ist]“, sodass nichts übrigbleibt „als die Allgemeinheit des Gesetzes überhaupt“ (GMS, AA 04: 421.01-03). Daraus ergibt sich Kant zufolge die Grundformel des KI: „*handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, daß sie ein allgemeines Gesetz werde*“ (GMS, AA 04: 421.07-08).

Das formale Prinzip der Willensbestimmung, das die erforderliche objektive Notwendigkeit durch seine Erkennbarkeit *a priori* garantiert, besteht also Kant zufolge in der *Universalisierbarkeit* der Maxime: Eine Maxime muss als allgemeines Gesetz erstens gedacht und zweitens gewollt werden können, um praktisches Gesetz (unter *dem* praktischen Gesetz des KI) sein zu können.<sup>20</sup> Wenn sie sich hingegen, als allgemeines Gesetz, „selbst vernichten würde“ (KpV, AA 05: 27.31), kann es sich um kein praktisches Gesetz handeln, da sie der Form nach keine objektive Notwendigkeit haben kann.<sup>21</sup>

Die Universalisierungsformel ist, als Grundformel des KI, Kant zufolge das oberste Prinzip der Moral. Als solches erfüllt sie eine Doppelfunktion<sup>22</sup>: Sie ist zum einen die *ontologische* Grundlage, d. h. der Seinsgrund des moralischen Status einer Handlung: Nur wegen der formalen Eigenschaft einer Maxime, universalisiert (nicht) gedacht und gewollt werden zu können, ist die entsprechende Handlung erlaubt, geboten oder verboten.<sup>23</sup> Zweitens jedoch hat sie auch eine epistemische Funktion: Ihre Anwendung ermöglicht uns, Maximen auf ihre Universalisierbarkeit hin zu überprüfen und in der Folge den moralischen Status der Handlungen, denen

<sup>20</sup> Vgl. GMS, AA 04: 424.1.

<sup>21</sup> Die Universalisierbarkeit einer Maxime stellt für sich genommen bloß die *Gesetzeskonformität* sicher, nicht den Status als praktisches Gesetz. Eine Handlung auf Basis einer solchen Maxime ist also immer erlaubt, aber nicht zwangsläufig auch *geboten*, siehe die weitere Diskussion.

<sup>22</sup> Vgl. Illies 2007: 306 und Wimmer 1982.

<sup>23</sup> Selbst wenn wir die Auffassung teilen, die Universalisierungsformel betreffe in erster Linie „moral permissibility, which as such, has only an indirect bearing on the determination of specific duties“ (Allison 2011: 180) ergibt sich das Problem, das ich in dieser Arbeit entwickle. Es genügt, dass sich die Unterscheidung zwischen erlaubten und nicht erlaubten Handlungen von der Universalisierungsformel ableiten lässt.

Diskutiert wird auch die Frage, ob der KI Kant zufolge zur moralischen „Selbstkontrolle“ grundsätzlich vor dem Vornehmen von Handlungen, oder lediglich vor dem Bilden von Maximen angewandt werden soll – vgl. Albrecht 1994: 129-146 und Fricke 2008: 125-136.



sie zugrunde liegen, zu *erkennen*, wodurch sie auch als handlungsleitendes Prinzip praktisch nutzbar wird.

Das Kriterium der *Denkbarkeit* der Universalisierung ist gut verständlich, aber eben auch sehr schwach: Den „Test“ der Universalisierungsformel besteht eine Maxime nach diesem Kriterium bereits, wenn ihre Universalisierung zu keinem logischen oder begrifflichen Widerspruch führt.<sup>24</sup> Damit nicht auch klarerweise unmoralische Maximen den „Test“ bestehen, muss die Theorie gemäß Kant durch das Wollbarkeits-Kriterium<sup>25</sup> angereichert werden, welches im Sinne eines *vernünftigen* Willens zu verstehen ist.<sup>26</sup> Mit Glückseligkeit und Vervollkommnung gibt es laut Kant universale Zwecke, die zumindest von allen *endlichen* Vernunftwesen notwendig gewollt werden.<sup>27</sup> Das Wollbarkeits-Kriterium zielt nun auf die Vereinbarkeit des vernünftigen Willens mit der Universalisierung der Maxime.<sup>28</sup>

So abstrakt diese Bestimmungen des obersten Prinzips der Moral sind, so beansprucht Kant doch, dass alle Vernunftwesen „das gedachte Princip jederzeit vor Augen“ haben (GMS, AA 04: 402.15, vgl. auch: KpV, AA 05: 35f.). Die explizite Ausformulierung und systematische Darlegung ist eine außerordentliche intellektuelle Leistung, doch jeder noch so unintelligente und ungebildete Mensch weiß

<sup>24</sup> Es wurden verschiedentlich, unter anderem von Korsgaard (1996), auch alternative Interpretationen der Funktionsweise des Universalisierbarkeits-Tests vorgeschlagen. Das in dieser Arbeit entwickelte Problem und die folgende Diskussion sind jedoch neutral bezüglich der Frage, wie der Test genau anzuwenden ist und welche Ergebnisse er im Einzelnen liefert, solange fixiert bleibt, dass er in Anwendung auf eine Handlung und die ihr zugrunde liegende Maxime jeweils genau ein bestimmtes Ergebnis liefert.

<sup>25</sup> GMS, AA 04: 424.3-14 erhebt dies zur Trennlinie zwischen vollkommenen und unvollkommenen Pflichten.

<sup>26</sup> Vgl.: Illies 2007: 311. Diese rein *systematische* Darstellung soll nicht nahelegen, Kant setze zunächst bei Denkbarkeit an und füge dann, weil diese nicht hinreichend ist, um alle moralischen Pflichten und Verbote herzuleiten, Wollbarkeit als weitere Bedingung hinzu. Hier geht es mir lediglich um die Konsequenzen für Kants Moraltheorie, die die beiden Kriterien jeweils haben, unabhängig von deren Genese im kantischen Text.

<sup>27</sup> In der KpV schreibt Kant etwa: „Glücklich zu sein, ist nothwendig das Verlangen jedes vernünftigen aber endlichen Wesens, und also ein unvermeidlicher Bestimmungsgrund seines Begehungsvermögens.“ (KpV, AA 05: 25.12-13) In der GMS heißt es: „Denn als ein vernünftiges Wesen will er nothwendig, daß alle Vermögen in ihm entwickelt werden.“ (GMS, AA 04: 423.13-15).

<sup>28</sup> Siehe etwa GMS, AA 04: 423.02-15 und 423.31-35. Vergleiche hierzu auch die Herleitung der eigenen Vollkommenheit und der fremden Glückseligkeit als „Zwecke, die zugleich Pflichten sind“ (MS, AA 06: 385.31) in der Metaphysik der Sitten.

Problematisch erscheint diese Interpretation allerdings angesichts einer Aussage Kants im zweiten Teil der GMS, assertorische Imperative, d. i.: hypothetische Imperative, die Glückseligkeit als Zweck beinhalten und somit notwendig für alle Vernunftwesen gelten, seien trotz allem „noch immer hypothetisch“ (GMS, AA 04: 416.05). In der KpV schreibt er denn auch: „Das Princip der Glückseligkeit kann zwar Maximen, aber niemals solche abgeben, die zu Gesetzen des Willens tauglich wären, selbst wenn man sich die *allgemeine* Glückseligkeit zum Objecte machte.“ (KpV, AA 05: 36.09-11) Er begründet dies damit, dass die *Anwendung* des Gesetzes von empirischen Fakten abhänge. Weshalb dieser Umstand problematisch sein soll, wo doch das *Prinzip* rein a priori wäre, bleibt unerklärt. Ich möchte die Diskussion dieser Passagen und des Wollbarkeitskriteriums an dieser Stelle offenlassen, da sie für den Zweck der Untersuchung nicht zwingend erforderlich ist und einer ausführlichen Behandlung bedürfte, die im Rahmen dieses Aufsatzes nicht geleistet werden kann.



Kant zufolge, sofern er ein Vernunftwesen ist, zumindest implizit um den KI als Grundgesetz der Moral und die sich daraus ergebenden moralischen Pflichten.<sup>29</sup>

Der Anspruch der kantischen Moralphilosophie ist, dass sich aus den Formeln des kategorischen Imperativs alle einzelnen kategorischen Imperative, also alle praktischen Gesetze, ableiten lassen.<sup>30</sup> So schreibt er, dass „aus diesem einigen Imperativ alle Imperativen der Pflicht [sic] als aus ihrem Princip abgeleitet werden können“ (GMS, AA 04: 421.09-10). *Erlaubt* sind solche Handlungen, deren zugrundeliegende Maximen universalisierbar sind, *verboten* solche, deren zugrundeliegende Maximen nicht universalisierbar sind, und *geboten*<sup>31</sup> solche, die zu unterlassen verboten ist (d. h.: Eine Handlung h ist geboten gdw. non-h verboten ist gdw. non-h nicht erlaubt ist, wie in Standardsystemen deontischer Logik üblich).

Die anderen Formeln des kategorischen Imperativs<sup>32</sup> sind laut Kant mit der Universalisierungsformel *äquivalent*. Insofern sollte alles, was aus ersteren ableitbar ist, auch aus letzterer ableitbar und die Universalisierungsformel folglich bereits ausreichend sein, die Gesamtheit der einzelnen moralischen Gebote und Verbote abzuleiten.<sup>33</sup> Kant schreibt in diesem Sinne: „Die angeführten drei Arten, das Princip der Sittlichkeit vorzustellen, sind aber im Grunde nur so viele Formeln eben desselben Gesetzes, deren die eine die anderen zwei von selbst in sich vereinigt“ (GMS, AA 04: 436.08-10, vgl. auch: 421, 437f., 447). Mehr noch: Kant empfiehlt sogar explizit, „in der sittlichen *Beurtheilung* immer nach der strengen Methode“ zu verfahren, „die allgemeine Formel des kategorischen Imperativs zum Grunde“ zu legen – die anderen Formeln seien lediglich dazu dienlich, die Anschaulichkeit zu erhöhen (GMS, AA 04: 436.29-32).

Illies weist darauf hin, dass diese Ableitung nicht immer *direkt* möglich oder nötig ist, sondern oftmals über die Ableitung allgemeinerer Prinzipien erfolgt:

<sup>29</sup> „Was ich also zu tun habe, damit mein Wollen sittlich gut sei, dazu brauche ich gar keine weit ausholende Scharfsinnigkeit.“ (GMS, AA 04: 403.18-19).

<sup>30</sup> Ido Geiger kritisiert diese gängige These, dass laut Kant das rein formale Kriterium des KI (ohne moralisches Vorwissen, das in das Ableitungs-Verfahren eingebracht wird) ausreicht, die einzelnen Moralgesetze – Pflichten und Verbote – abzuleiten: „It is a particularly unfair caricature that draws Kant attempting in vain to derive all duties ex nihilo from the formal-logical principle of non-contradiction.“ (Geiger Berlin 2008: 138). Doch Kant scheint genau das zu behaupten, wenn er bezüglich des Gesetzes, welches als einziges den freien Willen „nothwendig zu bestimmen tauglich ist“ (d. i., bezüglich des moralischen Gesetzes), schreibt, „die gesetzgebende Form, so fern sie in der Maxime enthalten ist, [sei] das einzige, was einen Bestimmungsgrund des Willens ausmachen kann“ (KpV, AA 05: 29.12-22).

<sup>31</sup> Wie Pflichten in Abgrenzung zu *bloß erlaubten* Handlungen hergeleitet werden, ist nicht unumstritten. Eine alternative Rekonstruktion findet sich etwa bei Beck (1966: 121f.). Hier ist die Idee die, dass neben der regulativen, negativen Funktion des KI, der zufolge eine Handlung universalisierbar sein muss, um nicht verboten zu sein, eine positiv bestimmende Funktion darin besteht, Handlungen zu gebieten, sofern ich eine Handlung vollziehen will, *weil* ich sie von anderen in hinreichend ähnlichen Umständen vollzogen sehen will.

<sup>32</sup> Abhängig von der Zählweise gibt es neben der Universalisierungsformel bis zu vier (oder, s. Schönecker und Wood 2004: 125f., sogar fünf) weitere Formeln des KI bei Kant: Die Naturgesetz-Formel, die Autonomie-Formel, die Selbstzweck-Formel und die Reich-der-Zwecke-Formel (vgl. etwa Wimmer 1982: 292).

<sup>33</sup> Kant demonstriert dies, indem er dieselben vier exemplarischen Pflichten bzw. Verbote, die er anhand der Universalisierungsformel abgeleitet hat, nach Einführung der Selbstzweck-Formel auch mittels dieser ableitet (vgl. GMS, AA 04: 429f.) und darauf hinweist, dies sei auch mit der Reich-der-Zwecke-Formel möglich (siehe GMS, AA 04: 432).

In der *Metaphysik der Sitten* spricht Kant zum Beispiel lediglich von zwei sehr grundsätzlichen Zwecken, die zugleich [...] Pflichten sind‘ (MS, AA 06: 385.13–14), nämlich der ‚eigene[n] Vollkommenheit‘ und der ‚fremde[n] Glückseligkeit‘ (385.32). Aus ihnen leitet er dann zahlreiche Einzelpflichten ab, etwa die Pflicht zur Selbsterhaltung und das Verbot des Selbstmords. (Illies 2007: 307.)

Dies macht das zunächst sehr einfache und transparente „Testverfahren“ ein wenig unklarer, sollte aber für die folgende Diskussion kein größeres Hindernis darstellen, da das Problem, das ich behandle, nicht von der Auswertung einer bestimmten Maxime abhängt, sondern allgemeineren Charakter hat.

### 3 Das Argument aus der Mannigfaltigkeit möglicher Maximen

Das Problem, das ich in dieser Untersuchung, auf der Basis der obigen Rekonstruktion der zentralen Begriffe der kantischen Moralphilosophie, entwickeln und diskutieren möchte, lässt sich durch das folgende Argument in Form einer *reductio ad absurdum* darstellen:

**P1:** Es gibt mindestens eine Handlung *h*, für die gilt: Es ist möglich, dass *h* eine universalisierbare Maxime zugrunde liegt, und es ist möglich, dass *h* eine nicht universalisierbare Maxime zugrunde liegt.

**P2:** Notwendigerweise gilt für jede Handlung *h*: *h* ist erlaubt, wenn *h* eine universalisierbare Maxime zugrunde liegt, und *h* ist nicht erlaubt, wenn *h* eine nicht universalisierbare Maxime zugrunde liegt.<sup>34</sup>

**P3:** Notwendigerweise gilt für jede Handlung *h*: Wenn *h* möglicherweise erlaubt ist, ist *h* erlaubt, und wenn *h* möglicherweise nicht erlaubt ist, ist *h* nicht erlaubt.

**K:** Es gibt mindestens eine Handlung *h*, für die gilt: *h* ist erlaubt und *h* ist nicht erlaubt.<sup>35</sup>

Die Konklusion ist logisch widersprüchlich und entsprechend für eine Moraltheorie inakzeptabel.<sup>36</sup> Nun scheint aber Kant auf alle drei Prämissen festgelegt zu sein,

<sup>34</sup> Da Kant, gemäß obiger Rekonstruktion, die Auffassung vertritt, dass jeder Handlung eine Maxime zugrunde liegt, ließe sich diese Prämisse auch mit einem Bikonditional formulieren.

<sup>35</sup> Das Argument lässt sich wie folgt formalisieren:

P1:  $\exists x(Hx \wedge (\Diamond \exists y((My \wedge Hx) \wedge (Zyx \wedge Uy))) \wedge \Diamond \exists y((My \wedge Hx) \wedge (Zyx \wedge \neg Uy))))$

P2:  $\Box \forall x((\exists y((My \wedge Hx) \wedge (Zyx \wedge Uy)) \rightarrow Ex) \wedge (\exists y((My \wedge Hx) \wedge (Zyx \wedge \neg Uy)) \rightarrow \neg Ex))$

P3:  $\Box \forall x(((Hx \wedge \Diamond Ex) \rightarrow Ex) \wedge ((Hx \wedge \Diamond \neg Ex) \rightarrow \neg Ex))$

K:  $\exists x(Hx \wedge (Ex \wedge \neg Ex))$

Dabei ist das Prädikat *H* als „...ist eine Handlung“ zu übersetzen, *M* als „... ist eine Maxime“, *E* als „... ist erlaubt“, *Z* als „... liegt ... zugrunde“ und *U* als „... ist universalisierbar“.

<sup>36</sup> Eine Theorie, der zufolge *K* zwar nicht wahr, aber zumindest weder wahr noch falsch ist, diskutiere und verwerfe ich unter 4.2.



von denen, da K falsch und das Argument gültig ist, mindestens eine falsch sein muss:

Die von P1 implizierte Annahme, dass es für manche Handlungen (oder Handlungsweisen bzw. Handlungstypen<sup>37</sup>) mehr als eine Maxime gibt, für die es möglich ist, der Handlung zugrunde zu liegen, geht m. E. unmittelbar aus der kantischen Maximentheorie hervor. So führt Kant selbst in der GMS einige exemplarische Fälle an, in denen einer Handlung verschiedene Maximen zugrunde liegen können: Die pflichtgemäße Handlung eines Händlers, einem unerfahrenen Kunden keine zu hohen Preise abzuverlangen, kann beispielsweise nicht nur aus „Grundsätzen der Ehrlichkeit“ erfolgen, sondern ebenso gut aus „eigennütziger Absicht“, um keine Kunden zu verschrecken und den eigenen Ruf zu wahren (GMS, AA 04: 397.27-32). Der Handlung könnte also beispielsweise sowohl die Maxime „Wenn ein unerfahrener Kunde bei mir einkauft, will ich ihn denselben Preis bezahlen lassen, wie alle anderen, um ihn nicht ungerecht zu behandeln“, als auch „...um meinen Ruf zu wahren und meine Kundschaft nicht zu vergraulen“ zugrunde liegen.

Fricke schreibt: „Von den beobachtbaren Eigenschaften einer einzigen Handlung lässt sich die ihr zugrunde liegende Maxime ohnehin nicht ablesen“ (Fricke 2008: 131) und fährt fort mit der Feststellung, dass die Maximen zwar zu der Handlung (und womöglich anderen Faktoren, wie der handelnden Person und der Entscheidungssituation) passen müssen, dies jedoch die Zahl in Frage kommender Maximen nicht grundsätzlich auf eins reduziert. Zu einer Handlung *h* passt also nicht *jede* Maxime, aber zumindest gelegentlich mehr als eine. Es ist nur ein kleiner und m. E. unproblematischer Schritt von der epistemologischen These Frickes, dass verschiedene Maximen zu einer Handlung *passen* können, zu meiner ontologischen Annahme, dass auch verschiedene Maximen einer Handlung *zugrunde liegen* können (wenn auch womöglich nicht gleichzeitig).

P2 wiederum gibt schlicht ein zentrales Element der kantischen Moraltheorie wieder, wie sie oben rekonstruiert wurde. Dadurch, dass P2 lediglich strikte Konditionale verwendet, bleibt sie zudem neutral bezüglich des genauen inhaltlichen Verhältnisses zwischen „erlaubt sein“ und „universalisierbar sein“ bzw. „verboten sein“ und „nicht universalisierbar sein“. Eine Festlegung auf eine bestimmte, kontroverse Kant-Interpretation ist somit nicht erforderlich. P3 schließlich ist Konsequenz der weitgehend anerkannten Annahme, dass moralische Tatsachen (in Bezug darauf, welche Handlungen erlaubt oder verboten sind) *notwendige* Tatsachen sind,

<sup>37</sup> Kant geht davon aus, dass der Universalisierbarkeits-Test nicht nur zeigt, ob eine Handlung erlaubt ist, sondern auch ganze *Handlungstypen* als erlaubt oder verboten erweisen kann. Unter dieser Annahme lässt sich das Argument sogar aus noch schwächeren Prämissen konstruieren, da nicht mehr von einer einzelnen Handlung als raumzeitlich singulärem Ereignis gesagt werden muss, ihr könnten verschiedene Maximen zugrunde liegen. Eine potenziell problematische modale These bezüglich der Identitätsbedingungen einzelner Handlungen über verschiedene mögliche Welten entfällt also.

Die Prämissen müssten dafür wie folgt modifiziert werden:

P1°: Es gibt ein *x*, ein *y* und ein *z*, sodass gilt: 1. *z* ist eine Handlung, *x* und *y* sind verschiedene Handlungstypen. 2. *z* fällt unter *x* und *y*. 3. Es gibt eine Maxime *m*<sub>1</sub>, die *x* vorgibt und eine Maxime *m*<sub>2</sub>, die *y* vorgibt. 4. *m*<sub>1</sub> ist universalisierbar und *m*<sub>2</sub> ist nicht universalisierbar.

P2°: Ein Handlungstyp ist notwendigerweise erlaubt gdw. die Maximen, die ihn vorgeben, universalisierbar sind, und eine Handlung ist notwendigerweise erlaubt, wenn ein Handlungstyp, unter den sie fällt, erlaubt ist, und nicht erlaubt, wenn ein Handlungstyp, unter den sie fällt, nicht erlaubt ist.

wie auch Kant explizit voraussetzt<sup>38</sup>. Wenn es also möglich ist, dass einer Handlung eine universalisierbare Maxime zugrunde liegt, ist es möglich, dass sie notwendig erlaubt ist, und folglich *ist* sie erlaubt (*mutatis mutandis* für verbotene und gebotene Handlungen).

Diese Annahme ist damit kompatibel, dass ein und dieselbe Handlung unterschiedlichen moralischen Wert haben (unterschiedlich gut oder schlecht sein) kann, abhängig etwa davon, auf Grundlage welcher Maxime sie tatsächlich vorgenommen wird und was die „Triebfeder“ ist. P3 ist also kompatibel damit, dass etwa, im Sinne Kants, nur pflichtgemäße Handlungen *aus Pflicht*, nicht aber etwa pflichtgemäße Handlungen *aus Neigung* moralisch wertvoll und folglich lobenswert sind.<sup>39</sup> Aber wenn Handlung h im ersteren Fall pflichtgemäß (und somit erlaubt) ist, muss sie es auch im zweiten Fall sein. Ob sie es ist, hängt nicht davon ab, weshalb die Handlung vorgenommen wird, bzw. welche Maxime ihr zugrunde liegt.<sup>40</sup>

Zwischen P1 und P3 besteht ein gewisses Spannungsverhältnis: Je grobkörniger das Evaluandum individuiert wird, desto plausibler ist P1, weil bei einer größeren Zahl möglicher Instanzierungen des Evaluandums (dessen moralischer Status festgestellt werden soll) plausibler ist, dass nicht alle davon die fragliche Eigenschaft, in diesem Falle *unter einer [nicht] universalisierbaren Maxime zu erfolgen*, teilen. Umgekehrt wird P3 umso plausibler, je feinkörniger das Evaluandum individuiert wird, da bei einer geringeren Zahl möglicher Instanzierungen plausibler ist, dass diese die fragliche Eigenschaft, in diesem Falle den *moralischen Status*, teilen. Obige Feststellung, dass P3 eine eher unkontroverse metaethische Annahme sein dürfte, gilt entsprechend nur, wenn wir über Einzelhandlungen sprechen, die so feinkörnig individuiert sind, dass sie kaum modale Varianz zulassen, wie sie für P1 erforderlich ist. Doch das muss uns nicht weiter bekümmern, da Kant darauf festgelegt ist, modale Invarianz hinsichtlich des moralischen Status auch eines sehr grobkörnig individuierten Evaluandums zu behaupten: Sogar Handlungstypen recht groben Zuschnitts schreibt er einen notwendigen, einheitlichen moralischen Status zu (s. etwa Fn 37). Hierdurch wird eine strukturelle Ursache des Problems deutlich, mit dem sich Kant konfrontiert sieht: Seine Moralthorie impliziert, dass der einheitliche moralische Status eines grobkörnig individuierten Evaluandums (Handlungen im Sinne

<sup>38</sup> „Jedermann muß eingestehen, daß ein Gesetz, wenn es moralisch, d. i. als Grund einer Verbindlichkeit, gelten soll, absolute Nothwendigkeit bei sich führen müsse; daß das Gebot: du sollst nicht lügen, nicht etwa bloß für Menschen gelte, andere vernünftige Wesen sich aber daran nicht zu kehren hätten, und so alle übrige eigentliche Sittengesetze;“ GMS, AA 04: 389.11-16; Vgl. auch ebd.: 408.12-19. Zu beachten ist hier selbstverständlich Kants besonderes Modalitäts-Verständnis. Für die Gültigkeit des Argumentes spielt dies jedoch, soweit ich sehen kann, keine Rolle.

<sup>39</sup> In GMS, AA 04: 397 ff. erklärt Kant anhand einiger Beispiele, dass lediglich (pflichtgemäße) Handlungen *aus Pflicht*, nicht aber solche aus Neigung einen moralischen Wert haben. In GMS, AA 04: 399.35-37 erläutert er diesbezüglich: „eine Handlung aus Pflicht hat ihren moralischen Werth *nicht in der Absicht*, welche dadurch erreicht werden soll, sondern in der Maxime, nach der sie beschlossen wird“. Dass es Kant zufolge Wertunterschiede zwischen zwei typidentischen Handlungen geben kann, so diese nach unterschiedlichen Maximen beschlossen werden, ist also unmittelbar ersichtlich. Das Problem, mit dem ich in dieser Arbeit befasst bin, betrifft aber nicht diesen moralischen Wert einer Handlung, sondern die Frage, ob die Handlung erlaubt, verboten oder geboten ist, sprich: ihren moralischen Status.

<sup>40</sup> Natürlich kann die Absicht entscheidend dafür sein, um welche Handlung es sich überhaupt handelt – so ist die Intention etwa im Falle des Lügens [Mordens] plausiblerweise Voraussetzung dafür, ein Akt des Lügens [Mordens] und nicht etwa eines anderen Handlungstypus zu sein.



von P1, oder Handlungstypen im Sinne von P1°, s. Fn 36) anhand eines wesentlich feinkörniger individuierten Evaluans (Maximen) bestimmt wird.

Da Kant P2 ohne eine gravierende Modifikation seiner Moraltheorie nicht aufgeben kann, P3 eine für ihn unverzichtbare metaethische Grundannahme und K kontradiktorisch ist, muss er offenbar P1 negieren. Im nächsten Abschnitt werde ich versuchen, diese Prämisse zu plausibilisieren. Denn das Gewicht des Problems, mit dem ich Kants Moralphilosophie konfrontiere, entspricht, sofern P2, P3 und non-K unangetastet bleiben sollen, dem Preis, den es seine Theorie kostet, P1 aufzugeben. Da ich bereits für die von P1 implizierte Annahme argumentiert habe, dass einer Handlung unterschiedliche Maximen zugrunde liegen können, ist dabei insbesondere zu zeigen, dass es auch möglich ist, dass diese Maximen bei Anwendung der Universalisierungsformel ein heterogenes Ergebnis hervorbringen. Unter 4.1 werde ich verschiedene Varianten diskutieren, das Problem im Sinne Kants dennoch durch ein Zurückweisen von P1 aufzulösen. Gelingt dies nicht, so müssten wesentliche Elemente der kantischen Ethik, etwa seine Maximentheorie, der Individuenbereich der Universalisierungsformel oder die Annahme eindeutiger Klassifizierungen von Handlungen als erlaubt und nicht erlaubt, aufgegeben werden.

### 3.1 Plausibilisierung von P1

Wie die obige Rekonstruktion ergab, sind Maximen als subjektive praktische Grundsätze selbstgesetzte Handlungsregeln oder -prinzipien, die verschiedentlich angewendet werden können: Sie bestimmen den Willen nicht über das Vorgeben einer einzelnen Handlung, sondern eines bestimmten Handlungstyps. Handlungstypen können offensichtlich verschieden feinkörnig definiert werden. Wenn wir den Typus einer Handlung extrem grobkörnig beschreiben, erhalten wir sehr allgemeine Definitionen, unter die sehr viele Handlungen fallen können. Der Handlungstypus „anderen Wesen Schaden zufügen“ kann beispielsweise auf unzählige Arten und Weisen instanziiert werden. Bei sehr *feinkörnigen* Beschreibungen sind weniger verschiedene Instanzierungen möglich: Ein Extrembeispiel für einen solchen Handlungstypus wäre beispielsweise „am 2.5.1994 in der Emilie-Mayer-Straße drei Stunden nach Apfelsaftkonsum einen Mord durch den Einsatz einer kleinkalibrigen Schusswaffe begehen“. Die Menge der möglichen Handlungen des letzteren Typus ist eine echte Teilmenge der Menge der möglichen Handlungen des ersteren Typus: Ein und dieselbe Handlung kann also beiden Handlungstypen entsprechen.

Da die Auswertung von Maximen per Universalisierungsformel die darin gesetzten Handlungstypen als erlaubt (bzw. geboten) oder verboten identifiziert und erst dadurch, dass der Typus einer Handlung (etwa „ein Versprechen brechen“) verboten ist, auch die Instanzierungen des Typus verboten sind, lässt sich die Konsequenz, dass ein und dieselbe Handlung verboten und erlaubt zugleich sein kann, nur vermeiden, wenn die möglichen Handlungstypen, unter die die Handlung fällt und die in einer Maxime benannt werden können, entweder alle erlaubt oder alle verboten sind.

Zumindest *prima facie* scheint es aber plausibel zu sein, anzunehmen, dass eine Maxime, die meiner Handlung *h* zugrunde liegen kann, je nachdem, wie sie den Handlungstypus individuiert, universalisierbar oder nicht universalisierbar sein

kann. Es scheint hinsichtlich der Universalisierbarkeit schlicht einen Unterschied zu machen, ob ich das Belügen des Meuchelmörders bezüglich des Aufenthaltsortes eines potenziellen Opfers unter den Typus „einen Menschen belügen“ fasse, oder unter den Typus „einen Menschen daran hindern, einen anderen umzubringen“. Dies lässt sich durch konkrete Beispiele demonstrieren.

Das Vorgehen bei der Entwicklung solcher Beispiele für das Problem der Mannigfaltigkeit möglicher Maximen, die zu obigem Widerspruch K führen, ist folgendes: Man suche eine Maxime (am besten eine allgemeiner und eine spezifischer formulierte als diejenigen Kants, um Beispiele auf unterschiedlichen Skalierungsstufen zu erhalten), unter die sowohl eine Handlung fallen kann, die nach Kant erlaubt ist (der also laut Kant eine universalisierbare Maxime zugrunde liegen kann), als auch eine, die nach Kant verboten ist (der also laut Kant eine *nicht* universalisierbare Maxime zugrunde liegen kann). Da die gewählte Maxime entweder universalisierbar oder nicht universalisierbar ist, gilt somit für eine der beiden Handlungen, dass ihr sowohl eine universalisierbare als auch eine nicht universalisierbare Maxime zugrunde liegen kann.

Je nachdem, ob die für das Beispiel gewählte Maxime universalisierbar ist oder nicht, müssten nach dem KI entweder alle oder keine der Handlungen, denen sie zugrunde liegen kann, erlaubt sein (worunter sowohl „erlaubt und geboten“, als auch „bloß erlaubt“ fallen). Da die Handlungen so gewählt wurden, dass eine laut Kant erlaubt und die andere verboten ist, ergibt sich zwangsläufig bei einer der beiden Beispielhandlungen ein Widerspruch im Sinne von K. Es ist nicht erforderlich, zu zeigen, *welche* der beiden Handlungen davon betroffen ist.

Dieses Schema lässt sich mit folgenden beiden Maximen veranschaulichen:

1.: Ich will immer so handeln, dass das Leid in der Welt minimiert, das Glück in der Welt hingegen maximiert wird. (Utilitarismus-Maxime)<sup>41</sup>

Wenn die Utilitarismus-Maxime universalisierbar ist, ergibt sich der Widerspruch, weil Lügen laut Kant verboten sind (weil eine Lügen-Maxime nicht universalisierbar wäre), in einzelnen Fällen aber zweifellos der Utilitarismus-Maxime entsprechen. Läge diese nun der Lüge zugrunde, so würde nach einer universalisierbaren Maxime gehandelt, weshalb die Handlung erlaubt wäre. Wenn die Utilitarismus-Maxime *nicht* universalisierbar ist, ergibt sich der Widerspruch, weil Menschen in Notlagen zu retten laut Kant erlaubt<sup>42</sup> (und *geboten*) ist (will eine entsprechende Maxime universalisierbar ist), aber in vielen Fällen auch der Utilitarismus-Maxime entsprechen. Läge diese nun der rettenden Handlung zugrunde, so würde nach einer nicht universalisierbaren Maxime gehandelt, weshalb die Handlung verboten wäre.

2.: Ich will immer die Wahrheit sagen, es sei denn, ich bin epistemisch gerechtfertigt, zu glauben, dass durch eine Falschaussage ein Menschenleben gerettet wird. (Bedingte-Wahrhaftigkeits-Maxime)

<sup>41</sup> Dass es sich hier um eine ungenaue Wiedergabe utilitaristischer Positionen handelt, spielt für das Argument keine Rolle. Die Maxime könnte auch einen beliebigen anderen Namen erhalten. Für einen Versuch, die Moraltheorie Kants mit dem Utilitarismus zu verbinden siehe etwa: Hare 2003.

<sup>42</sup> Gegeben, wie wir für den Beispielfall einfach stipulieren können, es besteht kein Konflikt zu einer engen Pflicht wie der, nicht zu lügen.



Wenn die Bedingte-Wahrhaftigkeits-Maxime universalisierbar ist, ergibt sich der Widerspruch, weil Lügen laut Kant verboten sind (weil eine Lügen-Maxime nicht universalisierbar wäre), in einzelnen Fällen aber der Bedingte-Wahrhaftigkeits-Maxime entsprechen. Wenn die Bedingte-Wahrhaftigkeits-Maxime *nicht* universalisierbar ist, ergibt sich der Widerspruch, weil die Wahrheit zu sagen laut Kant erlaubt (und *geboten*) ist (weil die entsprechende Maxime universalisierbar ist), in vielen Fällen aber auch der Bedingte-Wahrhaftigkeits-Maxime entspreche.

### 3.2 Das „Problem of Relevant Descriptions“

Bevor ich mich der Diskussion verschiedener Lösungsvorschläge widme, möchte ich in diesem Abschnitt erörtern, ob und inwiefern es sich bei dem nun entwickelten Problem um ein *neues* Problem für Kants Moralthorie handelt und wie es sich zum PRD verhält. Dabei zeigt sich, wie ich hoffe, dass es sich tatsächlich um ein basaleres Problem handelt, das bestehen bleibt, auch wenn sich eine Lösung für das PRD ergibt, und das – umgekehrt – auch dann lösenswert ist, wenn sich das PRD als unlösbar erweisen sollte.

In ihrem Essay „Acting on Principle“ (2013) setzt sich Onora O’Neill ausführlich mit dem PRD auseinander.<sup>43</sup> Ausgangspunkt ihrer Entwicklung des Problems ist die Feststellung, dass ein „universality test“ (wie Kants Universalisierungsformel, aber auch Marcus Singers Generalisierungsprinzip und Richard Hares universaler Präskriptivismus, vgl. O’Neill 2013: 63) als Testverfahren „action-guiding“ sein, uns also als praktische Entscheidungshilfe in moralischen Fragen dienen und unser Handeln anleiten können soll (ebd.: 60; 77). Das PRD entsteht nun O’Neill zufolge aufgrund des Umstandes, dass Prinzipien und Handlungen nicht „in one-one or one-many correspondence“ (ebd.: 60.) stehen, also nicht nur eine oder mehrere Handlungen unter genau ein Prinzip fallen, „but any given act exemplifies numerous principles“. Bisweilen besteht also eine „many-many“-Relation. Für praktische Entscheidungen in Bezug auf eine mögliche Handlung ist nun ein Kriterium erforderlich, um das relevante Prinzip zur Auswertung durch einen „universality test“ zu identifizieren: „So if conditions on principles are to be action-guiding [...] we must specify some way of deciding which of the principles covering an act it is relevant to assess in a given context“ (ebd.: 61).

Bis zu diesem Schritt kann O’Neill als repräsentativ für die sehr umfangreiche Diskussion des PRD in den vergangenen 100 Jahren gelten: Das PRD betrifft die Universalisierungsformel als handlungsleitendes Prinzip und Entscheidungs-Prozedur, insofern unklar ist, welche Maxime als Input für das Testverfahren gewählt werden muss (wobei in vielen Beiträgen unklar bleibt, ob es sich bloß um ein epistemisches oder auch um ein ontologisches Problem für die Anwendbarkeit des Tests handeln soll). Ewing schreibt etwa bereits im Jahr 1938 in diesem Sinne: “One difficulty about Kant’s method is to decide what exactly it is we are to universalize in a given case. [...] what degree of generality is required?” (Ewing 1938: 49).

<sup>43</sup> Zentrale Texte der Debatte, die O’Neill (2013) diskutiert, sind etwa: Anscombe 1957, D’Arcy 1963 und Lyons 1965.

Inzwischen klassisch ist die Formulierung des PRD bei Ross (1954), in der er das PRD als ein Problem der Anwendung des Testverfahrens wegen einer Unterbestimmtheit hinsichtlich der zu testenden Maxime charakterisiert:

Any individual act is an instance of a class of acts which is a species of a wider class; we can set no limit to the degrees of specification which may intervene between the *summum genus* “act” and the individual act. For example, if C tells a lie to the would-be murderer, this falls (i) under the subgenus “lies told to murderous persons,” (ii) under the species “lies,” (iii) under the genus “statements.” Kant pitches, arbitrarily, on the middle one of these three classes, and since acts of this class are generally wrong, and are indeed always *prima facie* wrong, he says that the particular lie is wrong. But the man who tells the lie may well retort to Kant “Why should the test of universalizability be applied to my act regarded in this very abstract way, simply as a lie? [...] The test of universalizability applied at one level of abstractness condemns the act; applied at another level of abstractness it justifies it. And since the principle itself does not indicate at what level of abstractness it is to be applied, it does not furnish us with a criterion of the correctness of maxims, and of the rightness of acts that conform to them. (Ross 1954: 32 f.)

Elizabeth Anscombe bezieht das Problem der Bestimmung des Inputs für das Testverfahren auf eine Unterbestimmtheit bei der *Bildung* von Maximen:

His own rigoristic convictions on the subject of lying were so intense that it never occurred to him that a lie could be relevantly described as anything but just a lie (e.g. as “a lie in such-and-such circumstances”). His rule about universalizable maxims is useless without stipulations as to what shall count as a relevant description of an action with a view to constructing a maxim about it. (Anscombe 1958: 2.)

Mark Timmons erklärt in diesem Sinne, das PRD betreffe die Frage der richtigen Formulierung der einzelnen Elemente einer Maxime: “The main interpretative question about maxims and thus about step 1 of the deliberative procedure is how one’s circumstances, action, and end are to be described for purposes of moral testing. This is the problem of ‘relevant descriptions’” (Timmons 2006: 163). Auch Henry Allison diskutiert das PRD als prozedurales Problem der Unterbestimmtheit des Testverfahrens, das nach einer Lösung im Sinne einer Präzisierung der Auswahl des Inputs verlangt. Es betreffe

what is to be tested for its universalizability. [...] it seems that the results of applying FLN depend crucially on which description is chosen and there is therefore need for some procedure, which Kant allegedly does not supply, to determine which description, is appropriate in each case. (Allison 2011: 192).<sup>44</sup>

O’Neill vertritt die Auffassung, dass der „universality test“ Kants die Lösung des Problems explizit nennt: Das relevante Prinzip der Bewertung einer Handlung

<sup>44</sup> Für weitere Beiträge zur Diskussion des PRD siehe auch: Schumski 2017, 1588-1613 und Timmons 1998, 389-417.



(der zu wählende Input für das Testverfahren) ist, so die so triviale wie richtige Antwort, diejenige Maxime, auf Basis derer die Person tatsächlich gehandelt hat oder zu handeln gedenkt.<sup>45</sup> Zumindest für die Bewertung aus der Innenperspektive sieht sie daher kein Problem bei der Theorie des KI, da hier das relevante Prinzip, also die jeweilige Maxime, klar identifizierbar ist.

O'Neill sieht also, dass mehrere Prinzipien einer Handlung zugrunde liegen können und die Mannigfaltigkeit korrekter Handlungsbeschreibungen ein Problem darstellen kann, löst jedoch nur einen Teil des Problems. Sie glaubt nämlich, dass Maximen und Handlungen standardmäßig in einer „one-many“-Relation, nicht in einer problematischen „many-many“-Relation stünden (O'Neill 2013: 63). In einer solchen „one-many“-Relation stehen aber allenfalls die *aktualen* Maximen zu Handlungen. Da aber auch *mögliche* Maximen relevant sind und in Bezug auf diese eine many-many-Relation zwischen Maximen und Handlungen nicht vermeidbar ist, ist das Problem, mit dem ich mich in dieser Untersuchung befasse, damit nicht gelöst.<sup>46</sup>

Dieser Unterschied zwischen dem PRD und dem in dieser Untersuchung verhandelten Problem basiert letztlich darauf, dass ersteres den KI vor allem als handlungsleitende Entscheidungshilfe untersucht und dementsprechend der Innenperspektive Vorrang gibt. Das PRD bedroht die *Anwendbarkeit* des KI als prospektiver „universality test“ und stellt somit seine epistemische Funktion als Erkenntnisgrund des moralischen Status einer Handlung in Frage. Ein Sicherstellen der Anwendbarkeit des Tests durch das Identifizieren der aktual zu testenden Maxime stellt dementsprechend bereits eine Lösung des PRD dar.

Das von mir behandelte Problem betrifft hingegen die Verbindung zwischen dem KI und dem *objektiven moralischen Status* von Handlungen, ihr Erlaubt- oder Verbotensein, und bedarf keiner Innenperspektive. Die Mannigfaltigkeit korrekter Handlungsbeschreibungen und folglich der möglichen Maximen bedroht nämlich nicht bloß die Anwendbarkeit des Testverfahrens, sondern das Vermögen des KI, einen eindeutigen, objektiven moralischen Status einer Handlung zu begründen. Dieser Schritt von dem Problem, aus einer Menge diverser Prinzipien das aktual zu testende zu identifizieren, zu der modalisierten Fassung des Problems, in der es auch um *bloß möglichen* Input für den Test geht, ist somit entscheidend. Er ergibt sich daraus, dass, entsprechend P3, eine Handlung, die *möglicherweise* erlaubt [verboten] ist, auch erlaubt [verboten] *simpliciter* ist, da moralische Tatsachen dieser Art (zumindest laut Kant) *notwendige* Tatsachen sind. Den richtigen Input für den Test

<sup>45</sup> Vgl. O'Neill 2013, 61.

<sup>46</sup> Timmons stellt genau dies fest: “Since actions and maxims stand in a many-many relation; i.e., for any action there are numerous maxims to which it corresponds and vice versa, one crucial question is how to select a maxim to be used in testing an action’s morality” (Timmons 1984: 295 f.). Allerdings bezieht auch er das Problem der “many-many“-Relation auf die Auswahl des Inputs für das Testverfahren und berücksichtigt nicht, dass diese Auswahl zwar in jedem einzelnen Fall eindeutig und unproblematisch ist, wie O'Neill zeigt, das eigentliche Problem sich jedoch aus der *Kombination* verschiedener einzelner (u. U. bloß möglicher) Fälle ergibt. Auf die gleiche Weise nimmt auch Williams dieselbe Beobachtung wie ich als Ausgangspunkt der Entwicklung des PRD, wenn er schreibt: „for any given act it is possible that there may be a large, and perhaps indefinite, number of maxims, of widely differing degrees of generality, which can relate to it“ (Williams 1968: 54), ohne es dann jedoch über das Problem der Input-Auswahl hinaus zu entwickeln.

in der aktualen Welt identifiziert zu haben löst das Problem also nur dann, wenn er *notwendig* der relevante Input für den Test ist.<sup>47</sup>

O'Neill ist bewusst, dass sie mit ihrer Lösung bloß die Universalisierungsformel als handlungsleitendes Prinzip verteidigt, nicht aber in ihrer Funktion, den objektiven moralischen Status einer Handlung zu bestimmen. So schreibt Timmons:

O'Neill admits that theories of right action typically aim to give us a story about the objective deontic status of actions, and so she admits that restricting Kant's universality tests to only giving us a decision procedure for determining the subjective deontic status of actions is quite a retreat from what the theory seems to promise. But after surveying a few proposals for solving the problem of objective descriptions and finding none of them acceptable, she concludes that a theory of subjectively right action is the best Kant can do. (Timmons 1998: 401)

Doch es hängt zu viel an dieser zweiten zentralen Funktion des KI, als dass Kant einfach darauf verzichten könnte: Der moralische Verpflichtung generierende Charakter der Grundform des KI ergibt sich schließlich unmittelbar aus deren Herleitung. Die Fokussierung der Debatte auf die Konsequenzen der Möglichkeit unterschiedlicher Handlungsbeschreibungen für den KI qua handlungsleitendes Prinzip trivialisiert daher das Problem.

Die genauere Betrachtung zeigt sogar, dass die Bezugnahme der Universalisierungsformel auf Maximen nicht nur keine Lösung des Problems der Mannigfaltigkeit möglicher Prinzipien liefert, sondern es im Gegenteil noch verschärft. Denn die Universalisierungsformel *beschränkt* die Menge der für die Auswertung relevanten Prinzipien zwar auf die echte Teilmenge der Maximen, *verhindert* jedoch zugleich eine weitere Beschränkung dadurch, dass sie impliziert, dass durch sie jede Maxime, nach der gehandelt werden kann, auch zur Bewertung einer Handlung herangezogen werden kann. Die problematische Mannigfaltigkeit möglicher Maximen muss also bereits durch den Maximenbegriff selbst hinreichend eingeschränkt werden, um das Problem gegensätzlicher Auswertungsergebnisse auszuschließen, da die Universalisierungsformel spätere Eingrenzungen nicht mehr zulässt. Man könnte also sagen, dass nicht eine Unterbestimmung des relevanten Inputs für den Test mittels der Universalisierungsformel problematisch ist, wie es in der Diskussion des PRD oftmals behauptet wird, sondern dass gerade die sehr explizite Bestimmung des relevanten

---

<sup>47</sup> Dass die beiden Probleme logisch unabhängig voneinander sind (auch wenn sie bei Kant verschränkt sind), zeigt sich daran, dass es strukturell möglich ist, jeweils eines der Probleme ohne das andere zu lösen. Oben habe ich erklärt, weshalb O'Neills Lösung des PRD nicht das von mir entwickelte Problem löst. Umgekehrt ist es aber auch denkbar, dass zwar einerseits eine bestimmte Menge relevanter Prinzipien feststeht, die durch ihre einheitliche Universalisierbarkeit oder Nicht-Universalisierbarkeit einen eindeutigen moralischen Status der fraglichen Handlung bestimmen, was das von mir entwickelte Problem lösen würde, andererseits jedoch diese Prinzipien für die Handelnden intransparent sind, sodass sich bei der Anwendung des Testverfahrens zur Handlungs-Anleitung durch die Möglichkeit verschieden feinkörniger Handlungstyp-Individuation ein praktisches Problem im Sinne des PRD ergibt.

Inputs (nämlich der jeweils der Handlung zugrundeliegenden Maxime) das Problem verursacht, indem sie die Hinzuziehung externer Relevanzkriterien unterbindet.<sup>48</sup>

## 4 Lösungsmöglichkeiten im Sinne Kants

In diesem Abschnitt werde ich verschiedene Lösungsansätze diskutieren, auf Basis derer der Widerspruch nach Möglichkeit ohne oder doch zumindest mit sehr geringen Änderungen der kantischen Moralphilosophie vermieden und somit das Problem der Mannigfaltigkeit möglicher Maximen gelöst werden kann. Ziel in diesem Sinne ist also beispielsweise, den KI zugleich möglichst unverändert als oberstes Prinzip der Moral aufrecht zu erhalten.

### 4.1 Restriktionen des Maximenbegriffs

Ein möglicher Lösungsansatz besteht darin, den Maximenbegriff weiter einzuschränken, sodass die Zahl der Maximen, die für eine Handlung in Frage kommen, so weit reduziert wird, dass non-P1 plausibel ist.<sup>49</sup> *Ceteris paribus* ist die Wahrscheinlichkeit bei einer geringeren Maximenzahl höher, dass bei der Auswertung wie gewünscht ein homogenes Ergebnis erzielt wird. Zusätzlich zu diesem quantitativen Aspekt würden restringierende Kriterien u. U. auch eine größere inhaltliche Ähnlichkeit der in Frage kommenden Maximen garantieren, was ein einheitliches Ergebnis noch wahrscheinlicher machen würde. Im Folgenden werde ich daher drei Kriterien zur Restriktion des Maximenbegriffs diskutieren, die von Kants Theorie nahegelegt werden.

#### 4.1.1 Streichung der Zweck-Variable

Eine erste Möglichkeit, die Zahl der Maximen, die einer Handlung *h* zugrunde liegen können, durch formale Restriktionen im Sinne Kants zu reduzieren, besteht

<sup>48</sup> Ein Relevanzkriterium, demzufolge nur solche Prinzipien für die Evaluation per Universalisierungsformel in Frage kommen, die keine zu fein- oder grobkörnige Handlungsbeschreibung beinhalten, kann beispielsweise nur dann herangezogen werden, wenn es zugleich notwendige Bedingung dafür ist, dass das fragliche Prinzip als Maxime klassifiziert werden kann. Ein anderes Beispiel wäre ein Kriterium, das Prinzipien ausschließt, die konditionale oder disjunktive Handlungsbeschreibungen beinhalten. Auch dieses kann aufgrund der Formulierung der Universalisierungsformel nur dann herangezogen werden, wenn zugleich der Status des Maxime-Seins davon abhängt. Externe, also vom Maximen-Status unabhängige Kriterien hingegen sind ausgeschlossen.

<sup>49</sup> Die eingrenzenden Kriterien müssen aus dem Maximenbegriff selbst hergeleitet sein und dürfen keinen externen Charakter haben, da es zum Kern der kantischen Moraltheorie gehört, dass dasjenige subjektive praktische Prinzip, das unser Handeln *tatsächlich* leitet auf seine Universalisierbarkeit getestet wird und nicht ein anhand externer Kriterien idealisiertes Prinzip. Aufgrund ihres implizierten Autonomieverlusts des Universalisierbarkeits-Tests scheitern Vorschläge wie derjenige von Timmons (wie er selbst erklärt): "So, presumably, a solution to the problem of relevant descriptions would specify which features of one's choice situation are morally relevant and thus ought to be included in a formulation of one's maxim and, by implication, which features lack moral relevance and ought not be included." (Timmons 2006: 163.) Auch Ross' Vorschlag „to envisage the act in its whole concrete particularity“ (Ross 1954: 34) ist aus diesem Grund nicht in Kants Moraltheorie zu integrieren.

darin, die dritte Variable aus der Formalisierung Illies' herauszustreichen: Maximen, bei denen sich entsprechend der Formel „Wenn es eine Situation X gibt, dann will ich Y tun, um Z zu erreichen“ nur der Z-Wert unterscheidet, sind demnach für das hier verhandelte Problem nicht als unterschiedliche Maximen zu werten. Das sieht man deutlich, wenn man sich die Auswertungsweise der Maximen bei Kant anschaut: Maximen, bei denen der vordere Teil „Wenn ein unerfahrener Kunde bei mir einkauft, will ich ihn übervorteilen“ lautet, werden immer gleich ausgewertet und entsprechend ist das Ergebnis immer identisch: Keine Maxime mit diesen X- und Y-Werten ist universalisierbar. Selbiges gilt für Maximen, die mit „Wenn ein Mörder mich nach dem Aufenthaltsort eines potenziellen Opfers fragt, will ich ihn belügen“ beginnen. Hier ist anzumerken, dass es Handlungen geben mag, für die eine bestimmte Intention definitorisch ist – so besteht der Unterschied zwischen Mord und bloßer Tötung, zwischen Lüge und bloßer Falschaussage gerade in den unterschiedlichen Intentionen. Doch in solchen Fällen würde die Intention eben bereits in der Handlungsbeschreibung, dem Y-Wert, enthalten sein. Der Zweck, der durch den Wert der Z-Variable ausgedrückt wird, ist einer, der den Handlungstyp nicht bestimmt.

Der Z-Wert gibt Aufschluss über die „Triebfeder“ und kann für den moralischen Wert der Handlung entscheidend sein. Sie hat aber keinen Einfluss darauf, ob eine Handlung pflichtgemäß oder -widrig, erlaubt oder verboten ist, da die Handlungstypenindividuation ausschließlich über die ersten beiden Variablen erfolgt. Somit vereinfacht dieses Restriktionskriterium zwar die Diskussion ein Stück weit, ist aber nicht geeignet, das Problem zu beheben, denn die beiden beispielhaften Problemfälle habe ich ohne die Angabe des Zwecks charakterisiert – das Problem besteht daher auch, wenn bloße Zwecksetzungs-Unterschiede ausgeklammert werden. Selbiges gilt für den Vorschlag, auch die Situations-Variable X bei der Auswertung auszuklammern. Das Problem, das sich ja aus der Möglichkeit unterschiedlich feinkörniger *Handlungstyp*-Individuation bei der Maximenbildung ergibt, betrifft primär die für die Auswertung unverzichtbare Handlungs-Variable Y.

Die folgenden Abschnitte werden also der Diskussion möglicher Kriterien zur Restriktion der zur Bildung von Maximen in Frage kommenden Handlungstypen gewidmet sein. Um den obigen Gegenbeispielen zu entgehen, müsste gezeigt werden, dass Handlungstypen für die Maximenbildung nicht auf die Art individuiert werden dürfen, wie es im Falle der Utilitarismus-Maxime und der Bedingte-Wahrhaftigkeits-Maxime geschieht.

#### 4.1.2 *Allgemeinheit*

Wie im zweiten Teil dieser Untersuchung herausgearbeitet, sind Maximen als Grundsätze *allgemein* formuliert, haben also im Vergleich zu den praktischen Regeln, die sie unter sich haben, einen breiteren Anwendungsbereich.<sup>50</sup> Für sich genommen ist so verstandene Allgemeinheit jedoch bloß eine *relative* und *gradueller* Eigenschaft: Ein Satz kann allgemeiner sein als ein anderer und zugleich weniger allgemein als ein dritter; dass es natürliche Schwellenwerte der Allgemeinheit gibt, ist äußerst

<sup>50</sup> Vgl. auch Fricke 2008: 127.



zweifelhaft. Der Begriff der Maxime soll jedoch, das geht aus Kants Argumentation klar hervor, keinesfalls vage sein.<sup>51</sup> Es braucht daher schärfere Kriterien für die Allgemeinheit einer Regel.

Betrachten wir etwa Kants Aussage, Grundsätze seien nicht abgeleitet, von ihnen seien aber andere Sätze ableitbar. Dass ein Satz *p* von einem Satz *q* abgeleitet ist und *q* von keinem anderen Satz abgeleitet ist, kann nicht so verstanden werden, dass die logische Struktur der Sätze bzw. ihre *Form* das Ableitungsverhältnis bestimmt: Denn jeder Satz *p* ist von einem anderen Satz ableitbar, etwa von „*p* und *q*“, und aus jedem Satz *p* sind andere Sätze ableitbar, etwa „*p* oder *q*“. Wenn Grundsätze solche Sätze sind, von denen andere Sätze abgeleitet werden können, ohne selbst ableitbar zu sein, kann es unter Voraussetzung des logisch-formalen Ableitbarkeits-Kriteriums also gar keine Grundsätze geben.

Ob ein Satz *p* von einem anderen Satz *q* im hier relevanten Sinne ableitbar ist, muss also von inhaltlichen Aspekten oder dem Kontext ihres Vorkommens abhängen. Die Grundidee ist, dass Grundsätze quasi axiomatischen Status haben, also nicht durch andere Sätze *begründet* werden, aber andere Sätze begründen. Der Grundsatz, nicht zu lügen, wird, so die Vorstellung Kants, nicht durch einen allgemeineren Satz begründet, aber der Satz „Ich will meine Freunde nicht belügen.“ lässt sich durch den Grundsatz, „Ich will nicht lügen.“ begründen. Ein Satz *p* ist also von einem Satz *q* abgeleitet, wenn die Gültigkeit von *p* im Kontext seines Vorkommens unter Berufung auf die Gültigkeit von *q* beansprucht wird: Ich will meine Freunde nicht belügen, *weil* ich nicht lügen will. Warum ich nicht lügen will, ist hingegen (in diesem Beispiel) nicht unter Berufung auf einen höheren subjektiven praktischen Grundsatz zu begründen. Daher handelt es sich hierbei um eine Maxime, bei ersterem hingegen nur um eine subjektive praktische Regel.

Das Ableitbarkeits-Kriterium lässt sich anhand verschiedener Vergleiche verdeutlichen: Die Zentrifugalkraft lässt sich beispielsweise unter Rückgriff auf andere, zugrundeliegende Naturgesetze begründen (die Massenträgheit), es handelt sich also nicht um einen „Grundsatz“ (bzw., als naturwissenschaftliches Äquivalent, um ein basales Naturgesetz). Die Naturgesetze, anhand derer sie erklärt werden kann, lassen sich hingegen nicht auf immer allgemeinere, basale Naturgesetze zurückführen, sondern haben grundsätzlichen Charakter. Ein anderes Beispiel finden wir etwa mit Artikel 1 des Grundgesetzes: „Die Würde des Menschen ist unantastbar“ hat in der Verfassung einen axiomatischen Charakter: Er ist gesetzt und nicht weiter begründet, andere Paragraphen lassen sich hingegen unter Rückgriff auf diesen Grundsatz begründen.

Dieses Beispiel verdeutlicht allerdings eine weitere Besonderheit des inhaltlichen Ableitungs- bzw. Begründungs-Kriteriums: es gilt immer nur relativ zu einem bestimmten Sätze-System. Wir können uns beispielsweise eine andere Verfassung vorstellen, die ebenfalls den Satz „Die Würde des Menschen ist unantastbar“ ent-

<sup>51</sup> Aus diesem Grunde etwa ist es auch nur bedingt hilfreich, „to exclude from the composite act descriptions in the subset reference to the act’s remoter circumstances and consequences“, um diesen Lösungsvorschlag für das PRD auf unser Problem zu übertragen (O’Neill 2013: 267). Ganz abgesehen davon, dass auch unter Annahme eines solchen Kriteriums ausreichend unterschiedliche Maximen übrigblieben, um P1 zu garantieren, ist die Bedingung schlicht zu vage.

hält, ihn allerdings von dem (seinerseits nicht abgeleiteten) Satz „Die Würde des *Vernunftwesens* ist unantastbar“ ableitet. In diesem Falle handelte es sich bei letzterem um einen Grundsatz, bei ersterem hingegen um einen abgeleiteten Satz. Ebenso kann exakt derselbe Satz, je nachdem, wie ein Subjekt ihn sich zu eigen macht und welche Stellung er im Regelsystem des Subjektes einnimmt, Maxime oder abgeleitete praktische Regel sein. Dies zeigt, dass es kein Charakteristikum (weder Inhalt noch logische Form) eines Satzes ist, das durch dieses Kriterium erfasst wird, sondern eine kontingente Relation, in der er zu anderen Sätzen und zum Subjekt steht. Damit eignet sich dieses Kriterium nicht, bestimmte Handlungstyp-Beschreibungen von der Maximenbildung auszuschließen.

Für eine Lösung des Problems kommt somit allenfalls das inhaltliche Kriterium der Allgemeinheit in Frage, das, wie zu Beginn dieses Abschnitts festgestellt, vage ist. Doch *selbst wenn* sich dieses Kriterium so weit präzisieren ließe, dass Handlungstypen, die zu fein individuiert werden, nicht mehr für die Maximenbildung in Frage kommen (so dass „Ich will am 2. Mai 1994 in der Emilie-Mayer-Straße drei Stunden nach Apfelsaftkonsum einen Mord durch den Einsatz einer kleinkalibrigen Schusswaffe begehen“ keine Maxime sein kann),<sup>52</sup> so blieben mindestens solche Maximen übrig, die genauso allgemein oder allgemeiner sind, als diejenigen, die Kant als Beispiele anführt. Das Allgemeinkriterium kann also allenfalls das Beispiel der Bedingte-Wahrheits-Maxime entkräften, nicht jedoch dasjenige der Utilitarismus-Maxime.

#### 4.1.3 Psychische Realität

Ein weiteres Kriterium, anhand dessen sich die Zahl von Maximen unterschiedlich feinkörniger Handlungstypenindividuation begrenzen lässt, findet sich womöglich darin, dass Maximen laut Kant *psychische Realität* haben. Klarerweise kommen einige Imperative nach diesem Kriterium nicht als Maximen in Frage: Wenn die Handlungstypen zu feinkörnig individuiert werden, ist es unplausibel, davon auszugehen, sie hätten psychische Realität und würden *faktisch* unser Handeln anleiten, wie Kant es verlangt: I. d. R. halten wir ein Versprechen nicht aufgrund einer Regel wie „Wenn ich am 17.12. eine Erkältung habe und einem einbeinigen Norweger etwas verspreche, so will ich mich an mein Versprechen halten“, sondern aufgrund einer Regel wie etwa: „Ich will meine Versprechen halten“.

Dieses Kriterium ist also offenbar geeignet, einige Sätze, die eine maximenähnliche Form haben, aus der Menge der Maximen auszuschließen. Aber ist es geeignet, die Menge der in Frage kommenden Maximen je Handlung so weit zu reduzieren, wie es für die Problemlösung nötig ist? Ich denke, die Antwort muss auch hier *nein* lauten: Es ist m. E. recht offensichtlich, dass Menschen unterschiedlich allgemeine oder spezifische „Regeln“ haben, die ihren Handlungen zugrunde liegen. Die beiden Regeln „Töte kein empfindungsfähiges Lebewesen!“ und „Töte keine Menschen!“ scheinen mir plausible Kandidaten für Regeln zu sein, die faktisch das Handeln von Menschen anleiten. Aber auch im Falle der Utilitarismus-Maxime ist nicht er-

<sup>52</sup> Für eine solche Strategie siehe etwa: Allison 2011: 200f.



sichtlich, weshalb sie nicht zumindest in einigen Fällen psychische Realität haben soll.

Das noch gravierendere Problem ist hier jedoch, dass es a posteriori und kontingent ist, welche Sätze mit Maximenform dieses Kriterium erfüllen. Selbst wenn alle Menschen einander psychisch so ähnlich sein sollten, dass die Maximen, nach denen sie handeln, einander so sehr ähneln, dass das Problem unterschiedlicher Auswertungsergebnisse der Universalisierungsformel nicht auftritt, so scheinen doch Vernunftwesen möglich zu sein, bei denen allgemeinere oder weniger allgemein gefasste Regeln psychische Realität haben. Da *notwendig* gelten soll, dass keine Handlung erlaubt und verboten zugleich ist, ist auch dieses Kriterium also keine Hilfe.

## 4.2 Maximenrelativismus hinsichtlich des moralischen Status einer Handlung

Auch wenn ich nicht beanspruchen kann, alle möglichen Kriterien untersucht zu haben, die die Zahl möglicher Handlungstypen, die einer Handlung zugeordnet werden können, bei der Maximenbildung so weit reduzieren, dass das Problem nicht mehr auftritt, so ist m. E. zumindest deutlich geworden, dass es überaus schwierig ist, ein solches Kriterium zu finden. Insbesondere dadurch, dass sich nach obigem Schema beliebig viele weitere Gegenbeispiele mit unterschiedlich feinkörniger Handlungstypenindividuation konstruieren lassen, ist eine Lösung des Problems auf den oben versuchten Wegen nicht möglich. Aus der Maximen-Definition Kants geht kein taugliches Kriterium hervor, sodass zumindest Ergänzungen seiner Theorie zur Lösung des Problems erforderlich sind. Dies zeigt jedoch auch, dass eine solche Ergänzung der Bestimmung des Maximenbegriffs (so sie überhaupt möglich ist) *ad hoc* wäre, da sie sich nicht aus der systematischen Erschließung dieser Kategorie von Grundsätzen ergäbe, sondern *zum Zwecke* der Problembhebung vorgenommen würde.

Angesichts dieser Problemlage stellt sich nun die Frage, ob ein gangbarer Weg darin besteht, eine der anderen Annahmen aufzugeben. Da Kant, wie wir gesehen haben, auf P2 und P3 aufgrund zentraler Elemente seiner Moraltheorie festgelegt ist, ginge die Verneinung einer dieser Thesen zwangsläufig mit gravierenden Modifikationen der kantischen Philosophie einher. Zunächst möchte den Vorschlag diskutieren, die Annahme aufzugeben, dass eine Handlung für sich genommen bereits verboten oder erlaubt ist, und stattdessen Verboten-Sein und Erlaubt-Sein einer Handlung nur *als unter einer bestimmten Maxime vorgenommen* zuzuschreiben. Dies ergäbe eine Modifikation von P2 und P3:

**P2'**: Notwendigerweise gilt für jede Handlung *h* und jede Maxime *m*: *h* ist *unter m* erlaubt, wenn *h* *unter m* fällt und *m* universalisierbar ist, und *h* ist *unter m* nicht erlaubt, wenn *h* *unter m* fällt und *m* nicht universalisierbar ist.

**P3'**: Notwendigerweise gilt für jede Handlung *h* und jede Maxime *m*: Wenn *h* *unter m* möglicherweise erlaubt ist, ist *h* *unter m* erlaubt, und wenn *h* *unter m* möglicherweise nicht erlaubt ist, ist *h* *unter m* nicht erlaubt.

Die Konklusion des Argumentes lautet nun lediglich, (K') dass es eine Handlung *h* gibt, für die es zwei Maximen *m*<sub>1</sub> und *m*<sub>2</sub> gibt, sodass *h* *unter m*<sub>1</sub> erlaubt und

unter  $m_2$  nicht erlaubt ist. Diese Konklusion ist logisch konsistent, also anders als  $K$  keine Reductio des Prämissen-Terzetts.<sup>53</sup>

Es gibt, aufgrund einer Skopus-Ambiguität in der Formulierung von  $P_2'$  und  $P_3'$  hinsichtlich des „unter Maxime  $m$ “-Modifikators, zwei verschiedene Lesarten dieser veränderten Prämissen: Entweder ist das Evaluandum diesem Vorschlag gemäß nicht eine Handlung, sondern ein Paar bestehend aus einer Handlung und der ihr zugrundeliegenden Maxime.  $K'$  zufolge ist also nach dieser Lesart „ $h$  unter  $m_1$ “ erlaubt. In diesem Falle sind  $P_2$  und  $P_3$  falsch (oder weder wahr noch falsch), weil sie fälschlicherweise Handlungen als Evaluanda identifizieren. Oder wir verstehen  $P_2'$  und  $P_3'$  so, dass zwar das Evaluandum nach wie vor die Handlung selbst ist, dass aber deren moralischer Status von der jeweils zugrundeliegenden Maxime abhängt.  $K'$  zufolge ist also nach dieser Lesart  $h$  „unter  $m_1$  erlaubt“, was konditional zu verstehen ist als: (Notwendig) erlaubt, *wenn*  $h$  unter  $m$  vorgenommen wird. In diesem Falle sind  $P_2$  und  $P_2'$  kompatibel (und wohl sogar äquivalent) und lediglich  $P_3$  ist falsch: Es gibt Handlungen, für die es sowohl möglich ist, erlaubt zu sein (weil es für sie möglich ist, unter einer universalisierbaren Maxime vorgenommen zu werden), als auch möglich ist, verboten zu sein (weil es für sie möglich ist, unter einer nicht universalisierbaren Maxime vorgenommen zu werden). Beide Varianten würden das Problem der Mannigfaltigkeit möglicher Maximen vermeiden<sup>54</sup>, aber beide Varianten scheitern aus denselben Gründen, wie ich im Folgenden darlegen möchte.

Es mag zunächst scheinen, als würde ein solcher Maximenrelativismus hinsichtlich des moralischen Status einer Handlung mit Kants Konzentration auf den *guten Willen* als alleinigen Träger moralischen Wertes harmonisieren und diese Position nur dadurch radikalisieren, dass Handlungen für sich genommen dieser Modifikation zufolge keinerlei moralische Evaluation zulassen. Doch wie wir im Folgenden sehen werden, ist das Gegenteil der Fall: Dass Handlungen einen eindeutigen, nicht Maximen-relativen moralischen Status haben, ist unverzichtbare Voraussetzung für Kants Theorie des guten Willens.

Der Maximenrelativismus beraubt uns klarerweise vieler unserer alltäglichen moralischen Urteile: Wir gehen für gewöhnlich davon aus, dass bestimmte Handlungen (etwa einen unschuldigen Menschen umzubringen) verboten bzw. pflichtwidrig sind, unabhängig davon, aufgrund welcher Maxime sie vorgenommen werden. Die Maxime mag für die Zuschreibung moralischen Wertes wichtig sein, nicht jedoch für den moralischen Status einer Handlung, verboten, erlaubt oder geboten zu sein. Ebenso selbstverständlich gehen wir davon aus, dass bestimmte Handlungen (etwa einen Menschen vor dem Ertrinken zu retten) erlaubt bzw. pflichtgemäß sind,

<sup>53</sup> Wie nach diesem Vorschlag mit  $K$  selbst zu bewerten ist, ist nicht ganz klar: Entweder handelt es sich dabei nach wie vor um eine Kontradiktion, oder Aussagen der Form „Handlung  $h$  ist erlaubt“ beinhalten einen Kategorienfehler (da Handlungen nur *unter einer Maxime* einen moralischen Status haben) und trage keinen (binären) Wahrheitswert.

<sup>54</sup> Dies gilt zumindest, wenn wir annehmen, dass es nie mehr als eine *aktuale* Maxime pro Handlung gibt. Diese Annahme ist nicht trivial: Es müsste gezeigt werden, dass es unmöglich ist, eine Handlung auf Basis mehrerer Maximen gleichzeitig vorzunehmen. Allerdings werde ich diesen Punkt nicht weiter verfolgen und, *for the sake of the argument*, im Sinne des Maximen-Relativismus annehmen, dass dies in der Tat unmöglich ist.



unabhängig davon, aufgrund welcher Maxime sie vorgenommen werden. Selbst wenn wir aufgrund niederer Motive, etwa wegen zugrundeliegenden Eigennutzes, der Handlung keinen moralischen Wert zuschreiben wollen, so wollen wir doch ausdrücken können, dass sie nichtsdestotrotz *erlaubt* war. Diese Annahmen teilt Kant (s. Abschnitt 3). So bezeichnet er, ohne weitere Bedingungen anzuführen, bestimmte Handlungen (und ganze Handlungstypen) als pflichtgemäß bzw. pflichtwidrig: Zu lügen ist pflichtwidrig (verboten), unabhängig von der dahinterstehenden Maxime. Selbstmord zu begehen ist pflichtwidrig, unabhängig von der zugrundeliegenden Maxime. Seine Versprechen zu halten ist pflichtgemäß (erlaubt), unabhängig von der zugrundeliegenden Maxime.

Die Unterscheidung zwischen der Eigenschaft einer Handlung *pflichtgemäß* zu sein und *moralisch wertvoll* zu sein ist sogar von großer Bedeutung für seine Moralphilosophie. So hat eine pflichtgemäße Handlung, die aus Pflicht vorgenommen wird, nur deshalb moralischen Wert, weil sie wegen des Wissens um ihr Pflichtgemäß-Sein vorgenommen wird. Ließe sich von einer Handlung nicht mehr sagen, sie sei, unabhängig von der zugrundeliegenden Maxime, erlaubt bzw. pflichtgemäß, so fiel auch die Grundlage des moralischen Wertes weg. Maximenrelativismus hinsichtlich des moralischen Status einer Handlung ist also keine Lösung für das Problem der Mannigfaltigkeit möglicher Maximen.

### 4.3 Integrieren eines Möglichkeitsoperators in die Universalisierungsformel

Die letzte Lösungsmöglichkeit, die ich diskutieren möchte, besteht ebenfalls in einer Modifikation von P2, die jedoch nicht in obigem Sinne maximenrelativistisch ist: Da das Problem dadurch entsteht, dass einer Handlung verschiedene Grundsätze zugrunde liegen können, der moralische Status einer Handlung jedoch anhand der Auswertung *eines einzelnen* dieser Grundsätze erfolgt, ist es möglich, das Problem durch eine Ausweitung des Individuenbereichs der Universalisierungsformel zu vermeiden, sodass statt der *aktualen* Maxime die *Gesamtheit* der Maximen, die einer Handlung zugrunde liegen *könnten*, für den Test zu verwenden ist. So ergäbe sich beispielsweise:

**P2\*:** Für jede Handlung  $h$  gilt notwendigerweise:  $h$  ist erlaubt gdw. es für  $h$  *mindestens eine* mögliche Maxime gibt, die universalisierbar ist.

(Für jede Handlung  $h$  gilt notwendigerweise:  $h$  ist verboten gdw. es für  $h$  *keine* mögliche Maxime gibt, die universalisierbar ist.)

Die beiden exemplarischen Dilemmata ließen sich durch dieses modifizierte P2-Prinzip lösen: Wenn die Utilitarismus-Maxime *nicht* universalisierbar ist, so ist Lügen, ganz im Sinne Kants, verboten, da *keine* mögliche Maxime, die einer Lüge zugrunde liegen kann, universalisierbar ist.<sup>55</sup> Die Wahrheit zu sagen bliebe hingegen

<sup>55</sup> Ich habe hier selbstverständlich nicht alle möglichen Maximen betrachtet, sondern lediglich die Utilitarismus-Maxime und die Maxime, die Kant selbst in seinem Beispiel anführt, um die Pflichtwidrigkeit des Lügens zu erweisen. Es sollte jedoch klar sein, wie die Lösung auf die übrigen möglichen Maximen (von denen es vermutlich unendlich viele gibt) zu erweitern wäre. Ob *tatsächlich* keine Maxime, die einer Lüge zugrunde liegen kann, universalisierbar ist, scheint dabei zweifelhaft zu sein.

erlaubt, da einer solchen Handlung zwar mit der Utilitarismus-Maxime *auch* eine nicht universalisierbare Maxime zugrunde liegen kann, aber die Bedingung erfüllt ist, dass *mindestens eine* mögliche Maxime universalisierbar sein muss. Wenn die Utilitarismus-Maxime hingegen universalisierbar ist, so wäre Lügen (und entsprechend natürlich auch die Wahrheit sagen) nach obigem Schema erlaubt, solange die Lüge Glück-maximierend und Leid-minimierend ist (bzw. wenn sie diesen Zweck verfolgt). Jeder Handlung (oder, *mutatis mutandis*, jedem Handlungstyp) käme nach obigem Schema also ein eindeutiger moralischer Status zu.

Das auf diese Art modifizierte P2-Prinzip (das eine entsprechende Anpassung der Universalisierungsformel mit sich brächte) ist plausibler als folgende Variante:

**P2\*\*:** Für jede Handlung *h* gilt notwendigerweise: *h* ist erlaubt gdw. es für *h* keine möglichen Maxime gibt, die *nicht* universalisierbar ist.

(Für jede Handlung *h* gilt notwendigerweise: *h* ist verboten gdw. es für *h* *mindestens eine* mögliche Maxime gibt, die *nicht* universalisierbar ist.)<sup>56</sup>

Ein Problem von P2\*\* besteht darin, dass sich offenbar für *jede* Handlung eine Maxime konstruieren lässt, die *nicht* universalisierbar ist. Ich kann beispielsweise für jede Handlung *h* im Falle einer kontingenterweise nicht erfüllten Bedingung *b* die Maxime bilden „Ich will, gegeben dass *b*, Selbstmord begehen [was laut Kant verboten ist], anderenfalls aber *h* tun.“. Dies sollte daher nicht ausreichen, um eine Handlung verboten sein zu lassen. Eine Maxime, die problemlos universalisierbar ist, lässt sich hingegen weniger leicht für jede Handlung konstruieren. Es ist beispielsweise nicht ersichtlich, welche universalisierbare Maxime der grundlosen Tötung eines anderen Vernunftwesens zugrunde liegen könnte.<sup>57</sup>

Zudem ergäbe sich bei P2\*\* ein Problem in Bezug auf *gebotene* Handlungen. Während nach meinem Vorschlag eine Handlung *h* geboten ist gdw. es für non-*h* keine mögliche Maxime gibt, die universalisierbar ist, wäre *h* nach diesem Ansatz geboten gdw. es für non-*h* mindestens eine mögliche Maxime gibt, die nicht universalisierbar ist. Da es aber gut sein kann, dass es sowohl für *h* als auch für non-*h* jeweils mindestens eine mögliche Maxime gibt, die nicht universalisierbar ist, droht sich das Problem zu wiederholen, dass der moralische Status einiger Handlungen uneindeutig ist, da *h* in diesem Falle verboten und geboten zugleich wäre. (Analog ergäbe sich für Verbote das Problem, dass es eine Handlung *h* gäbe, für die gilt, dass sowohl *h* als auch non-*h* verboten sind.) Dieses Problem betrifft nicht P2\*, sofern es keine Fälle gibt, in denen weder für *h* noch für non-*h* eine universalisierbare Maxime möglich ist, was *prima facie* plausibel ist.

<sup>56</sup> D. h. mein erster Vorschlag, P2\*, lautet, in formaler Schreibweise:  $\Box \forall x(Hx \rightarrow (Ex \leftrightarrow \Diamond \exists y((My \wedge Hx) \wedge (Zyx \wedge Uy))))$ . Der hier aufgeführte Vorschlag, P2\*\*, den ich im Folgenden zurückweise, lautet in formaler Schreibweise:  $\Box \forall x(Hx \rightarrow (Ex \leftrightarrow \neg \Diamond \exists y((My \wedge Hx) \wedge (Zyx \wedge \neg Uy))))$ .

<sup>57</sup> Der Einwand, dass das Nach-vorne-Schieben eines Messers erlaubt ist, das Erstechen eines Menschen hingegen nicht, beides aber eine Beschreibung derselben Handlung ist, also das Vorhandensein einer universalisierbaren Maxime nicht ausreichen kann, eine Handlung erlaubt sein zu lassen, halte ich für nicht überzeugend, denn das nach vorne Schieben des Messers scheint nicht der problematische Aspekt am Erstechen eines Menschen zu sein. Diese Beschreibung wählt lediglich ein Element der Gesamthandlung aus, sodass auch nicht die Gesamthandlung bewertet wird, wenn wir dieses Element auswerten.

Kant könnte mit P2\* somit K vermeiden und auch seine einzelnen moralischen Urteile, etwa, dass zu lügen pflichtwidrig ist, verteidigen, sofern er zeigen kann, dass nicht nur die aktuelle, sondern alle möglichen Maximen, unter die diese Handlung fallen kann, nicht universalisierbar sind. Zudem ergibt sich eine interessante Differenzierungsmöglichkeit für die moralische Bewertung von Handlungen, analog zu Kants Unterscheidung pflichtgemäßer Handlungen *aus Pflicht* und solcher, die aus anderen Motiven erfolgen: Eine Handlung, der eine universalisierbare Maxime bloß zugrunde liegen *kann*, ist plausibler Weise anders zu bewerten, als eine, der eine solche Maxime auch *tatsächlich* zugrunde liegt, auch wenn die Handlung in beiden Fällen erlaubt ist.<sup>58</sup>

Doch diese Modifikation hat ihren Preis: Sie verkompliziert das Testverfahren erheblich, von dem Kant doch beansprucht, es sei für jedermann, selbst das ungebildetste Vernunftwesen, problemlos nutzbar. Während die eigene, tatsächlich zugrundeliegende Maxime womöglich noch ohne größere Probleme getestet werden kann, erfordert die Auswertung *aller möglichen* Maximen geistige Kapazitäten, über die wohl die wenigsten Vernunftwesen verfügen. Wie gravierend dieses epistemische Problem ist, ist unterdessen nicht klar: Schließlich ist es bereits anhand *einer* möglichen Maxime, die universalisierbar ist, möglich, festzustellen, dass eine Handlung nach obigem Vorschlag erlaubt ist – es geht also sicher nicht *alles* moralische Wissen verloren.

Problematischer ist unter Annahme von P2\* Wissen darüber, dass etwas verboten ist: Dafür muss ich nach obigem Vorschlag schließlich wissen, dass *keine* mögliche Maxime universalisierbar ist. Da ein induktives Schließen auf die Nichtexistenz einer universalisierbaren Maxime ausgehend von einer für die endliche Denkerin händelbaren Stichprobe möglicher Maximen dem Geist der kantischen Philosophie widerspräche, müssen wir eine Alternative suchen. So könnten wir beispielsweise als *Default* annehmen, dass eine Handlung verboten ist – und erst, wenn eine universalisierbare mögliche Maxime gefunden ist, vom Erlaubt-Sein ausgehen. Dies würde einem allgemeinen Vorsichtsprinzip in Fragen der Moral entsprechen, das mir zumindest nicht völlig unplausibel zu sein scheint. Ein naheliegender Einwand ist allerdings, dass nach diesem Vorschlag Fälle eintreten könnten, in denen wir *weder* für eine Handlung *h* *noch* für non-*h* eine universalisierbare Maxime finden und so, per *Default*, so handeln sollen, als seien *h* *und* non-*h* verboten. Solche moralischen Dilemmata könnten ein zu hoher Preis für meinen Lösungsvorschlag sein.

Außerdem wäre erst zu testen, ob die Ergebnisse des neuen Testverfahrens mit unseren moralischen Intuitionen verträglich sind – auch wenn Kant selbst in seiner Moralphilosophie überwiegend Top-Down argumentiert, sollten wir stark kontraintuitive Ergebnisse nicht einfach so hinnehmen. Sollte sich beispielsweise herausstellen, dass es kaum Fälle gibt, in denen *keine* mögliche Maxime universalisierbar ist, sodass also kaum eine Handlung verboten bzw. pflichtwidrig wäre (ein Verdacht, der naheliegt, da die Menge verbotener Handlungen nach P2\* eine Teilmenge der verbotenen Handlungen nach P2 ist), wäre dies ein starker Grund, P2\* abzulehnen.

<sup>58</sup> Über pflichtgemäße Handlungen *aus Pflicht* ließe sich, im Rahmen einer ersten Annäherung, etwa sagen: Eine Handlung *h* erfolgt aus Pflicht gdw. nicht bloß keine universalisierbare Maxime non-*h* zugrunde liegen kann, sondern dieser Umstand Teil der tatsächlichen Maxime, die *h* zugrunde liegt, ist.

## 5 Fazit

Die obige Diskussion hat gezeigt, dass das Problem, das sich aus der Funktion des KI, der Formulierung der Universalisierungsformel und dem kantischen Maximenbegriff ergibt, nicht ohne größere Eingriffe in die Moralphilosophie Kants zu lösen ist. Die Universalisierungsformel, aus der der moralische Status einzelner Handlungen hervorgehen soll, liefert angesichts der Mannigfaltigkeit möglicher Maximen, die sich aus unterschiedlich feinkörniger Individuation von Handlungstypen ergibt, teilweise widersprüchliche Ergebnisse. Eine Restriktion des Maximenbegriffs führte in der Diskussion zu keinem befriedigenden Ergebnis, insofern sie sich als entweder unzureichend oder *ad hoc* erwies. Auch einer Handlung ausschließlich *insofern ihr eine bestimmte Maxime zugrunde liegt* einen eindeutigen moralischen Status zuzuweisen, stellt keinen gangbaren Ausweg dar.

Daher bleiben m. E. zwei Möglichkeiten, die Universalisierungsformel des KI zu verteidigen: Entweder muss der Maximenbegriff Kants aufgegeben und durch einen strengeren ersetzt werden (wobei unklar ist, ob ein solches Vorhaben gelingen kann), oder aber die Universalisierbarkeitsbedingung darf nicht auf die *aktuelle* Maxime allein angewendet werden, sondern auf die Menge aller möglichen Maximen. Letztgenannte Variante vermeidet das Problem der Mannigfaltigkeit möglicher Maximen und fügt sich hervorragend in Kants Theorie des guten Willens und seine Unterscheidung zwischen pflichtgemäßen Handlungen aus Pflicht und solchen aus Neigung. Wie der erstgenannte Lösungsvorschlag hätte jedoch auch dieser an anderer Stelle gravierende Folgen für Kants Moralphilosophie und ist seinerseits problembehaftet. Es ist gut möglich, dass dieser Preis für die Sicherstellung der Konsistenz einer Moralphilosophie kantischer Bauart zu hoch ist.

Aufgrund der Äquivalenzthese Kants sollten die entwickelten Probleme eigentlich auch alle anderen Formeln des KI betreffen. Doch nach obigem Befund erscheint es mir angebracht, die Äquivalenzthese selbst in Zweifel zu ziehen und zu untersuchen, ob womöglich die anderen Formeln zu leisten vermögen, was mit der Universalisierungsformel nicht zu erreichen ist. Insbesondere in Bezug auf die *Selbstzweckformel* ist dieses Unternehmen indiziert, da diese nicht auf Maximen Bezug nimmt<sup>59</sup>, wodurch das Problem der Individuation von Handlungstypen bei der Bildung von Maximen hier keine Rolle spielt. Eine solche Lösung ist jedoch stark revisionär und allenfalls als letzter Ausweg für die kantische Ethik in Betracht zu ziehen.

**Danksagung** Für zahlreiche Anregungen und hilfreiche Kommentare danke ich Daniel Eggers, Thomas Grundmann, Achim Lohmar, Jakob Lohmar sowie zwei anonymen Gutachter\*innen.

<sup>59</sup> Siehe GMS, AA 04: 429. Tobias Kronenberg (2016), Jens Timmermann (2003) und Onora O'Neill (1989) schlagen vor, die Selbstzweckformel unter Rückgriff auf den Maximenbegriff zu reformulieren. Vor dem Hintergrund der vorliegenden Untersuchung halte ich es im Gegenteil für fruchtbar, gerade das *Fehlen* des Maximenbegriffs als Ausgangspunkt für die Bewertung ihrer Tragweite zu nutzen. Timmons hält dieses Hinzuziehen der Selbstzweckformel für die „pretty obvious solution“ des PRD. (Timmons 1998: 399). Anders als Timmons denke ich, dass dies eine der kantischen Moralphilosophie letztlich fremde Lösung und bestenfalls das geringste Übel ist, die erfordert, das Scheitern des ursprünglichen Ansatzes der kantischen Moralphilosophie zu akzeptieren. Während Timmons schreibt: „I am a happy-go-lucky optimist since I don't think the problem is very much of a problem in Kant's ethics“ (ebd.), ziehe ich, nach obigen Überlegungen, somit eine weit negativere Bilanz.



**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Albrecht, Michael. 1994. Kants Maximenethik und ihre Begründung. *Kant-Studien* (85/2), 129-146. <https://doi.org/10.1515/kant.1994.85.2.129>.
- Allison, Henry. 2011. *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press.
- Anscombe, G. E. M. 1957. *Intention*. Oxford: Blackwell.
- Anscombe, G. E. M. 1958. Modern Moral Philosophy. *Philosophy* (33/124), 1-19. <https://doi.org/10.1017/S0031819100037943>.
- Beck, Lewis White. <sup>4</sup>1966. *A Commentary on Kant's Critique of Practical Reason*. Chicago/London: University of Chicago Press.
- Bittner, Rüdiger. 2005. *Aus Gründen handeln*. Berlin/New York: De Gruyter.
- D'Arcy, Eric. 1963. *Human Acts. An Essay in their Moral Evaluation*. Oxford: Clarendon Press.
- Ewing, A. C. 1938. The Paradoxes of Kant's Ethics. *Philosophy* (13/49), 40–56.
- Fricke, Christel. 2008. Maximen. In: *Recht und Frieden in der Philosophie Kants. Band 3*, Hg. Valerio Rohden, 125-136. Berlin: De Gruyter.
- Geiger, Ido. 2008. How Do We Derive Moral Laws?. In: *Recht und Frieden in der Philosophie Kants. Band 3*, Hg. Valerio Rohden, 137-148. Berlin: De Gruyter.
- Hare, Richard. 2003. Could Kant Have Been a Utilitarian?. In: Ders.: *Sorting Out Ethics*, 147-166. Oxford: Clarendon Press.
- Illies, Christian. 2007. Orientierung durch Universalisierung: Der Kategorische Imperativ als Test für die Moralität von Maximen. *Kant-Studien* (98), 306-328. <https://doi.org/10.1515/kant.2007.018>.
- Korsgaard, Christine. 1996. Kant's Formula of Humanity. In: Dies.: *Creating the Kingdom of Ends*, 106-132. Cambridge: Cambridge University Press.
- Kronenberg, Tobias. 2016. *Maximen in Kants praktischer Philosophie – Über das verknüpfende Element der Kantischen Theorie des Handelns, der Freiheit des Willens und der Moralität*. Karlsruhe: Karlsruher Institut für Technologie (KIT).
- Lyons, David. 1965. *Forms and Limits of Utilitarianism*. Oxford: Clarendon Press.
- O'Neill, Onora. 1989. *Constructions of Reason. Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.
- O'Neill, Onora. 2013. *Acting on Principle – An Essay on Kantian Ethics*, Second Edition. Cambridge: Cambridge University Press.
- Ross, David. 1954. *Kant's Ethical Theory*. Oxford: Clarendon Press.
- Schönecker, Dieter und Allen W. Wood. 2004. *Kants „Grundlegung zur Metaphysik der Sitten“*. Ein einführender Kommentar. Paderborn: Schöningh (UTB).
- Schumski, Irina. 2017. The Problem of Relevant Descriptions and the Scope of Moral Principles. *European Journal of Philosophy* (25/4), <https://doi.org/1588-1613>. <https://doi.org/10.1111/ejop.12246>.
- Timmermann, Jens. 2003. *Sittengesetz und Freiheit. Untersuchungen zu Immanuel Kants Theorie des freien Willens*. Berlin/New York: De Gruyter.

- Timmons, Mark. 1984. Contradictions and the Categorical Imperative. *Archiv für Geschichte der Philosophie* (66), 294–312. <https://doi.org/10.1515/agph.1984.66.3.294>.
- Timmons, Mark: 1998. Decision Procedures, Moral Criteria, and Relevant Descriptions in Kant's Ethics. In: *Jahrbuch für Recht und Ethik* (5), 389–417.
- Timmons, Mark. 2006. The Categorical Imperative and Universalizability. In: *Groundwork for the Metaphysics of Morals*, Hgg. Christoph Horn und Dieter Schönecker, 158-199. Berlin/New York: De Gruyter. <https://doi.org/10.1515/9783110204551.3.158>.
- Williams, T. C. 1968. *The Concept of the Categorical Imperative. A Study of the Place of the Categorical Imperative in Kant's Ethical Theory*. Oxford: Clarendon Press.
- Wimmer, Rainer. 1982. Die Doppelfunktion des kategorischen Imperativs in Kants Ethik. *Kant-Studien* (73/3), 291-320. <https://doi.org/10.1515/kant.1982.73.1-4.291>.