

Martin Welsch\*

## Kant über den Selbstbetrug des Bösen

<https://doi.org/10.1515/kant-2019-0002>

**Abstract:** In Kantian philosophy, the evil heart is constituted as a system of self-degrading and self-deranging freedom by the coordination of two voluntary acts: the act of establishing radical evil and the act of a voluntary lie to oneself. The consequence is a kind of “madness of freedom”, which characterises the self-deception of evil. By discussing Kantian rhetoric as an elaborate art of writing, this structure will be explored via a new approach to reading ‘On Radical Evil in Human Nature’. The result is that it is the mere possibility of a lie to oneself originating in freedom that makes it impossible to cognise whether one’s heart is good or systematically evil.

**Keywords:** radical evil, evil heart, self-deception, rhetoric

Ist es möglich, sich selbst zu belügen? Folgt man Derridas „Geschichte der Lüge“<sup>1</sup>, so muss man in dieser Frage unweigerlich an die Grenzen des traditionellen Begriffs einer Lüge stoßen. Er besagt, dass ohne einen vorsätzlichen Willen zur Täuschung keine Lüge möglich sei.<sup>2</sup> Sich selbst zu belügen würde demnach heißen, sich mit Absicht etwas anderes zu sagen, als man weiß und dass man es denkt und tut, um sich selbst zu schaden. Das ist Wahnsinn („folie“), offenbar jedoch der Wahnsinn des Selbstbetrugs.<sup>3</sup> Folglich sind andere Strukturen zu denken, will man die Selbstlüge („mensonge à soi“) verstehen. Strukturen des Irrationalen sind gefragt, die über das Bewusstsein und intentionale Wissen hinausführen.<sup>4</sup>

Ähnliche Strukturen sind es, die der Kant-Forschung im Verständnis des Lehrstücks vom ‚radikalen Bösen‘ größte Schwierigkeiten bereiten. Immer wieder werden „Paradoxien und Aporien“ festgestellt, welche die Rationalität der Kantischen Moralphilosophie aufzuheben drohen.<sup>5</sup> Von „Gefahren“ ist die Rede, von

---

1 Derrida, Jacques: *Histoire du mensonge*. Paris 2012.

2 Derrida 2012, 22–24.

3 Derrida 2012, 18–20.

4 Derrida 2012, 95–97.

5 Horn, Christoph: „Die menschliche Gattungsnatur: Anlagen zum Guten und Hang zum Bösen“. In: *Klassiker Auslegen*. Band 41. Hrsg. von Otfried Höffe. Berlin 2011, 43–69, 64–68, im Anschluss

---

\*Kontakt: Martin Welsch, Universität Heidelberg, Philosophisches Seminar, Schulgasse 6, D-69117 Heidelberg; m.welsch@uni-heidelberg.de

denen man Kants Philosophie „gerne loslösen würde“ und die das Kantische Lehrstück als „philosophisch mißglückt“ erscheinen lassen.<sup>6</sup> Zwar gibt es gute Gründe, kritische Distanz zu dieser Forschungsposition einzunehmen.<sup>7</sup> Meines Erachtens fällt jedoch erst dann die Originalität der Kantischen Schrift in den Blick, wenn man dem Hinweis auf diese Problemstrukturen folgt.

Keineswegs zufällig ist es, dass der Kantische Hang zum Bösen als „Suchtmodell“<sup>8</sup> strukturiert ist: Moralische Delikte werden im Rahmen dieses Hanges nicht länger als vorsätzliche Pflichtverletzung aufgefasst, sondern als tendenziell entschuldbares Fehlverhalten einer psychisch kranken Person. Dasselbe gilt für das Unwissen über den Ursprung des Hanges zum Bösen in einem vorsätzlichen Willensakt. Um zurechenbar zu sein, müsste er zu gegenwärtigen Handlungen in einer „psychologisch-erinnerungsbasierten Kontinuität“<sup>9</sup> stehen, was aber offenbar nicht der Fall ist. Zufällig ist das deshalb nicht, da Kant den Akt der vorsätzlichen Entscheidung für das Böse mit einer „Unredlichkeit“ einhergehen lässt, „sich blauen Dunst vorzumachen“<sup>10</sup>. Später ist vom „Opium fürs Gewissen“ die Rede<sup>11</sup>, in der „Tugendlehre“ von „betäuben, oder in Schlaf bringen“<sup>12</sup>.

Hiermit liegen nicht nur die differenten Strukturen vor, auf welche Derrida mit Verweis auf den Marx'schen Ideologiebegriff und die Freud'sche Psychoanalyse hindeutet. Bei Kant ist auch das von Derrida aufgestellte Programm bereits erfüllt, diese Strukturen mit der Problematik des willentlichen Selbstbetruges zusammenzuführen.<sup>13</sup> Die Strukturen sind hier nämlich das systematische Produkt eines vorsätzlichen Selbstbetruges. Und der ist wiederum nach dem Muster des traditionellen Begriffes einer Lüge konzipiert: Die vorsätzliche Entscheidung für das Böse wird zugleich von einem ebenso vorsätzlichen Willen zum Selbstbetrug flankiert, der nach dem personalen Modell einer äußeren Lüge strukturiert ist.<sup>14</sup> Was durch beide bezweckt wird, ist die Aufhebung eben jener Rationalitätsstrukturen einer ‚gesunden‘ praktischen Vernunft, die er vor-

---

an Firestone, Chris L./Jacobs Nathan: *In Defense of Kant's Religion*. Bloomington/Indianapolis 2008, und Michalson, Gordon E.: *Fallen Freedom*. Cambridge 2009.

**6** Horn 2011, 64–68.

**7** Sirovátka, Jakob: *Das Sollen und das Böse in der Philosophie Immanuel Kants*. Hamburg 2015, 100–102, vgl. ferner Dörflinger, Bernd: „Kant über das Böse“. In: *Kant-Lektionen*. Hrsg. von Manfred Kugelstadt. Würzburg 2008, 81–107, 88–90, 103–105.

**8** Horn 2011, 48 f., 65.

**9** Horn 2011, 65 f.

**10** RGV, Erstes Stück, III,9, AA 06: 37.18–38.33.

**11** RGV, AA 06: 78.02, Anm.

**12** MS, TL, § 13,2, AA 06: 438.13–23.

**13** Vgl. Derrida 2012, 96 f.

**14** MS, TL, § 9,3, AA 06: 430.09–13.

aussetzt. Das ist der gewollte Schaden. Ihn hat Kant bereits früh als ‚selbstverschuldete Unmündigkeit‘ bezeichnet;<sup>15</sup> das radikale Böse sei weniger „Bosheit“ als vielmehr „Nichtswürdigkeit“<sup>16</sup>. Die „Paradoxien und Aporien“ des radikalen Bösen sind folglich ganz recht als Gefahren für die Kantische Moralphilosophie einzuschätzen, allerdings nicht im Sinn philosophischer Denkfehler. Vielmehr ist es das ideengeschichtliche Verdienst der Kantischen Bemühung, diese Degeneration als die immanente Logik eines Selbstbetruges sich selbst degradierender Freiheit erkannt zu haben. Damit erweist sich Kant ein weiteres Mal als großer Denker der Freiheit, dieses Mal jedoch, weil er ihren womöglich größten Abgrund aufgedeckt hat. Dies ist der „Wahnsinn der Freiheit“ (Kleist), der im vorsätzlich und frei gewählten Selbstbetrug des Bösen liegt.<sup>17</sup>

Bis jetzt ist die Tiefe des Abgrundes dieser Selbstlüge nicht ausgelotet, sie gilt es erst noch zu erforschen. Dazu bietet die einschlägige Kant-Literatur eine solide Basis. Vor allem zwei Punkte der (Standard-)Interpretation *Henry E. Allison* eignen sich hierzu vorzüglich: zum einen die sogenannte *Incorporation Thesis*, zum anderen die Annahme, das böse Herz sei mitsamt seiner drei Stufen ein durch und durch verlogenes Herz.<sup>18</sup> Die *Incorporation Thesis* ist dabei weniger als These Allison aufzufassen,<sup>19</sup> sondern vielmehr als griffige Bezeichnung jener Behauptung, die Kant zu Beginn der Schrift über das Böse selbst aufstellt. Sie besagt, der Mensch könne „durch keine Triebfeder zu einer Handlung bestimmt werden [...], als nur sofern [...] sie in seine Maxime aufgenommen hat (es sich zur allgemeinen Regel gemacht hat, nach der er sich verhalten will)“.<sup>20</sup> In der Literatur hat man diese These so verstanden: Die Neigung

15 WA, AA 08: 35.

16 RGV, AA 06: 38.23–27.

17 In ihrer großangelegten Studie, *Kant on Evil, Self-Deception and Moral Reform*. New York 2018, hat Laura Papish hingegen versucht, das Kantische Lehrstück von derartigen Problemen weitgehend zu befreien, vor allem indem sie den Selbstbetrug des Bösen nicht als Selbst-Lüge nach dem traditionellen Modell einer (äußeren) Lüge interpretiert – obwohl sie einräumt, 69 f., Kant spreche wiederholt von einer Selbstlüge („a lie to oneself“). Stattdessen begreift Papish jenen Selbstbetrug als Verstellung, die mit einer Selbsttäuschung einerseits einhergehe, 149 f., andererseits aber auch eine Verwandtschaft mit ihr aufweise, 118, 148 f. Doch selbst diese moralisch schwerwiegendere Selbsttäuschung sei lediglich eine Rationalisierungsstrategie („rationalization“), die unter anderen Voraussetzungen auch normativ erlaubt sei – und damit *grundsätzlich ungefährlich* ist, 69–79.

18 Allison, Henry E.: *Kant's Theory of Freedom*. Cambridge 1990, 147, 159.

19 So aber Rukgaber, Matthew S.: „Irrationality and Self-Deception within Kant's Grades of Evil“. In: *Kant-Studien* 2015 (2), 234–258, 248.

20 RGV, AA 06: 23.03–24.05, vgl. Sargentis, Konstantinos: „Das Böse als Selbstverhältnis des Guten. Zum Problem der Stufen des ‚bösen Herzens‘ bei Kant“. In: *Allgemeine Zeitschrift für Philosophie* 2011 (1), 49–69, 59.

gen des Menschen sind keine natürlichen Gegebenheiten, sondern Produkte seines Willens.<sup>21</sup> Daraus ergibt sich eine bemerkenswerte Freiheit über die eigene Psychologie,<sup>22</sup> und genau darum ist die *Incorporation Thesis* äußerst wichtig, will man jenen „Wahnsinn der Freiheit“ verstehen. Aus der These folgt aber auch, dass die erste Stufe des bösen Herzens notwendig eine (Selbst-)Lüge sein muss, was im Text freilich unausgesprochen bleibt: Wenn die „Gebrechlichkeit (*fragilitas*) der menschlichen Natur“ – „Wollen habe ich wohl, aber das Vollbringen fehlt“ – ganz vom menschlichen *Willen* abhängt, so kann es sich eben nicht um eine Gebrechlichkeit der menschlichen *Natur* handeln.<sup>23</sup> Daraus folgt: Willensschwäche ist Selbstbetrug. Davon ausgehend ist schließlich anzunehmen, dass der Selbstbetrug dritter Stufe sich nicht nur auf diese Stufe erstreckt, sondern auf das ganze menschliche Herz.<sup>24</sup>

Neuerdings sind diese zwei Punkte aber grundsätzlich in Zweifel gezogen worden, und zwar von *Matthew S. Rukgaber*: Einerseits wird die „absolut[e] Spontaneität der Willkür (der Freiheit)“ in Frage gestellt, über Maximen zu disponieren und so den Charakter *in toto* festzulegen; „[t]he idea that one’s ultimate character is under the empirical agent’s ‘voluntary control’ cannot be true“.<sup>25</sup> Andererseits sei aber auch das Phänomen der Willensschwäche (*akrasia*) in einer Form zu bewahren, die von der Selbstlüge des Bösen nicht betroffen ist.<sup>26</sup> So wird das böse Herz mit seinen drei Stufen nicht mehr als Einheit eines verlogenen Herzens begriffen, sondern als Typologie, die repräsentativ für die wichtigsten Formen immoralischer Handlungen sei, mit denen man es in der Erfahrung zu tun habe.<sup>27</sup> Von einem „Wahnsinn der Freiheit“ lässt sich so gesehen nur noch bedingt sprechen: Einerseits wird jene weitreichende Disposition über die Psyche verneint. Der Möglichkeitsraum der Freiheit ist infolge dessen so sehr eingeschränkt, dass die Zurechenbarkeit der durch die Selbstlüge des Bösen verursachten Defekte

21 Vgl. z. B. Guyer, Paul: *Kant on Freedom, Law and Happiness*. Cambridge 2000, 299.

22 Rukgaber 2015, 249.

23 RG, AA 06: 29.24–30. Was bei Kant 1793 unausgesprochen bleibt, kommt in Rousseaus ›*Émile*‹ zu Wort: „C’est la seule tiédeur de nôtre volonté qui fait toute nôtre foiblesse, et l’on est toujours fort pour faire ce qu’on veut fortement: *Volenti nihil difficile*“, „*Émile* ou De l’éducation“. In: *Œuvres complètes*, IV. Paris 1969, 651. Dem unmittelbarem Kontext dieser Textstelle entspringt auch das Galgenbeispiel der KpV, § 6, AA 05: 30.21–35.

24 Allison 1990, 159.

25 Rukgaber 2015, 244 f., 249 f., ferner Papish 2018, 128. Vgl. dagegen meine ausführliche Darlegung der Kantischen Position in: „Anthropologische Selbstbestimmung bei Kant, Rousseau und Nietzsche“. In: *Anthropologie in der Klassischen Deutschen Philosophie*. Würzburg. Im Erscheinen.

26 Rukgaber 2015, 239 f., 243.

27 Rukgaber 2015, 255, ferner Sargentis 2011, 69.

fragwürdig wird. Andererseits erfährt aber auch der Wirkungsbereich der Selbstlüge eine weitgehende Einschränkung wenn verneint wird, dass sie sich auf das ganze menschliche Herz erstreckt.

So sehr man diese Interpretation selbst wiederum in Frage stellen kann, so hat sie doch ein schlagendes Argument für sich: Zwar mag man von der *Incorporation Thesis* aus die menschliche Gebrechlichkeit als nichtige Selbstlüge entlarven; einen stichhaltigen Beleg dafür lässt sich im Text allerdings nicht finden.<sup>28</sup> Damit ist die Übertragung der Selbstlüge dritter Stufe auf alle Stufen des menschlichen Herzens eine interessante, letztlich aber haltlose Interpretation. Dasselbe wäre für meine weiterführende Deutung zu sagen, welche den Selbstbetrug des Bösen als Fundamentalproblem menschlicher Freiheit begreift. Im einschlägigen Text sind dem Problem nämlich nur zwei Absätze und eine Fußnote gewidmet.<sup>29</sup> Davor und danach scheint der fundamentale Selbstbetrug überhaupt keine Schwierigkeit mehr darzustellen: Zum einen wird das Böse ohne Bezug auf das Thema der Lüge eingeführt und bestimmt. Zum anderen scheint das Problem der Selbstlüge aber dann auch völlig unerheblich zu sein für die Überwindung des Bösen in einer moralischen Revolution und Reform.

Demgegenüber lassen sich im Text allerdings auch Anzeichen einer Rhetorik finden, die auf das Problem der Selbstlüge des Bösen zugeschnitten ist. So muss die Verlagerung des Problems vom Haupttext in eine Fußnote nicht notwendig als sachliche Marginalisierung ausgelegt werden. Vielmehr kann man sie als rhetorisches Verfahren lesen, das die Struktur der vorsätzlichen Selbstlüge imitiert: Das Problem, das zuvor als zentral identifiziert wurde, bleibt als solches bestehen, doch es wird auf eine externe und tieferliegende Ebene verlagert und letztlich ausgeblendet. An den Leser des Textes ist damit dieselbe Forderung gestellt, wie an das Subjekt der Selbstlüge: Beide müssen stets von der offensichtlichen Ebene auf eine vorgelagerte, nicht länger sichtbare Ebene reflektieren. Laut dem „Satz vom angeborenen Bösen“<sup>30</sup> ist der Leser jedoch mit diesem Subjekt identisch.

Der Gebrauch verhüllender und auf Enthüllung angelegter Rhetorik war zu Kants Zeit wohlgermerkt nicht unüblich. Darauf hat Leo Strauss in den 40-er und

---

<sup>28</sup> Rukgaber 2015, 258, ferner Sargentis 2011, 63, der „glaub[t], dass die stillschweigende Rolle des Eigendünkels der Schlüssel zu einem adäquaten Verständnis“ der Stufen des Bösen ist und davon ausgeht, man müsse „einiges zu Kants Gesagtem“ unterstellen, weil sonst nicht einsehbar wäre, „wie dem kantischen Bild des bösen Herzens Konsistenz verliehen werden kann“. Ferner sieht sich auch Papish, 2018, 69, 9 f., aufgrund der knapp bemessenen Textgrundlage dazu gezwungen, die Grundannahme ihrer Studie – der Selbstbetrug des Bösen sei keine Selbstlüge – auf Überlegungen zu gründen, die sie ausgehend von textexternen Referenzen anstellt.

<sup>29</sup> RGV, Erstes Stück, III, 8 und 9, AA 06: 37.18–38.33 sowie IV,4, Anm., AA 06: 42.17. Vgl. Papish 2018, 69.

<sup>30</sup> RGV, AA 06: 50.21 f.

50-er Jahren hingewiesen und Heinrich Meier ab den 80er Jahren:<sup>31</sup> Zum einen war die Verhüllung von Gedanken vonseiten der Autoren durch die Bedingung politischer Zensur und Verfolgung motiviert. Zum anderen sollte der Vorgang der Enthüllung aufseiten der Leser eine Übung im Selbstdenken sein; *faire penser les autres*, das sei die zu Grunde liegende Maxime gewesen. Im Rekurs auf solch eine *Art of Writing* würde sich schließlich erklären lassen, warum das Problem der Selbstlüge des Bösen im Kantischen Lehrstück nur angedeutet bleibt. Die „Höllenfahrt des Selbsterkenntnisses“, den Gang in die „Tiefen (Abgrund) des Herzens“,<sup>32</sup> hätte Kant den Lesern womöglich auch in philosophischer Hinsicht weitgehend selbst überlassen.

Darum soll im Folgenden der Zugang zum Kantischen Text auf dem Weg der Analyse seiner Rhetorik gefunden werden – Rhetorik, verstanden als Kunst der Rede und des Schreibens im weitesten Sinne.<sup>33</sup> Hierzu bespreche ich (1.) im Ausgang und in Rückbezug auf die Schrift über das Böse die rhetorische wie philosophische Schlüsselfunktion der Kantischen Charakterlehre. Erst vor ihrem Hintergrund kann erkennbar werden, inwiefern die (Selbst-)Lüge ein existenzielles Problem *par excellence* ist.<sup>34</sup> Auf dieser Grundlage erfolgt (2.) die Rekonstruktion des Zusammenhangs der beiden Entscheidungen für das radikale Böse einerseits und für die Selbstlüge andererseits. Beide Akte sind nicht identisch, vielmehr wohl zu unterscheiden, doch konstituieren zusammen das systematische Ganze eines ‚bösen Herzens‘.<sup>35</sup> Abschließend wird (3.) das Kantische Pro-

---

**31** Strauss, Leo: „Persecution and the Art of Writing“. In: *Social Research*. 1/1941, 488–504, ders.: „On a Forgotten Kind of Writing“. In: *Chicago Review* 1954 (1), 64–75, Meier, Heinrich: „Rousseaus Diskurs über die Ungleichheit – Ein einführender Essay über die Rhetorik und die Intention des Werkes“. In: *Diskurs über die Ungleichheit. Discours sur l'inégalité*. Paderborn 2008, XXI–LXXVII, XXIII–XXXI.

**32** MS, TL, § 14,2, AA 06: 441.12–19.

**33** Kant selbst war ein geschickter Rhetoriker. Darauf wurde in der Literatur bereits mehrfach hingewiesen, neuerdings von Maja Schepelmann: *Kants Gesamtwerk in neuer Perspektive*. Münster. 2017. Die Kantische Rhetorik wird von meiner Seite allerdings nicht von Vornherein (werkübergreifend) auf einen Begriff gebracht; vielmehr befrage ich vorrangig den Text auf die dort partikular ins Werk gesetzte Rhetorik.

**34** Dieser Arbeitsschritt ist auch als Gegenentwurf zu Papishs Ansatz zu lesen, demzufolge das Modell der (Selbst-)Lüge wie gesagt durch eine existenziell weitgehend ungefährliche Praxis der Verstellung und Rationalisierung ersetzt wird.

**35** Folgende Thesen sollen damit zurückgewiesen werden: (a) es gebe kein (Kantisches) System des Bösen, (b) die Selbstlüge sei die Ursache des Bösen sowie (c) die Ordnung des bösen Herzens bestehe weitgehend unabhängig von der Selbstlüge. Die erste These (a) wird von Paul Ricœur vertreten („Il n’y a pas de système du mal“), *La critique et la conviction*. Paris 1995, 168, vgl. in Bezug auf Kant Sirovátka 2015, 89; die zweite (b) von Nelson Potter: „Duties to Oneself, Motivational Internalism, and Self-Deception in Kant’s Ethics“. In: *Kant’s Metaphysics of Morals*:

gramm der Überwindung des Bösen in Rückbezug auf das Problem der Selbstlüge des Bösen gelesen und diskutiert. Auf diese Weise soll das Ausmaß des Problems eingeschätzt werden. Dabei wird sich jedoch zeigen: Die *bloße Möglichkeit der Selbstlüge aus Freiheit* reicht bereits aus, die moralische Urteilskraft dahingehend zu irritieren, dass es für sie unerkennbar ist, ob das Herz gut oder böse ist.

## 1 Die rhetorisch-philosophische Schlüssel- funktion der Charakterlehre

Das systematische Zentrum der Kantischen Schrift ist die Kopplung der Entscheidung für das radikale Böse mit dem Vorsatz zur Selbstlüge am Ende des dritten Kapitels. Die moralische Wertung der Lüge liegt damit auf der Hand, sie ist höchst verwerflich. Die Rhetorik des Textes scheint sich hierzu jedoch gegenläufig zu verhalten. „*Mutato nomine de te fabula narratur.*“/„Unter verwandeltem Namen wird in der Fabel von Dir erzählt.“<sup>36</sup> Das ist für Kant nicht nur ein bekennender Ausdruck seiner Würdigung der Heiligen Schrift als philosophisch brauchbare „Dichtkunst“,<sup>37</sup> sondern vielmehr ein Prinzip seiner eigenen philosophischen Rhetorik. Diese verharrt im Text über das Böse nicht nur oft im „Schwebezustand der List“<sup>38</sup>, sondern ist als Rhetorik der Täuschung zu identifizieren, wenn nicht gar der Lüge.

Einerseits gibt es eine klar angeordnete Begrifflichkeit praktischer Vernunft, die Kant gleich zu Beginn im (problematischen) dritten Absatz der Einleitung<sup>39</sup> vorlegt: Eine böse Handlung setzt eine böse Maxime voraus und die wiederum eine oberste böse Maxime. Solch eine oberste Maxime umfasst alle Handlungen des Subjekts und bestimmt zugleich, wie dieses von seiner „Freiheit überhaupt“ Gebrauch macht, entweder gut oder böse. Im Willkürakt der Entscheidung für diese Maxime legt das Subjekt dann seinen moralischen „Charakter“ fest, seine

---

*Interpretative Essays.* Hrsg. von Mark Timmons. Oxford 2002, 371–389, 386; die dritte (c) von Konstantinos Sargentis 2011, 63–68, vgl. aber auch Gressis, Robert: „Kant’s Theodicy and its Role in the Development of Radical Evil“. In: *Archiv für Geschichte der Philosophie* 2018 (1), 46–75, 66, der behauptet, durch Freiheit sei nur der Hang zum Bösen verursacht, nicht aber der Hang zur Falschheit.

<sup>36</sup> RGV, AA 06: 42.19 f., Horaz: *Sermones. Satiren.* Stuttgart 1972, I,1.

<sup>37</sup> RGV, AA 06: 10.06–08.

<sup>38</sup> Vgl. Brandt, Reinhard: *Universität zwischen Selbst- und Fremdbestimmung.* Berlin 2003, 39.

<sup>39</sup> Brandt, Reinhard: „Ein problematischer Absatz im ‚Ersten Stück‘ von Kants *Religion innerhalb den Grenzen der bloßen Vernunft*“. In: *Archivio di filosofia* (Pisa/Roma) 63, 1995, 337–349, 343–347.

„Gesinnung“, seinen „Hang“ zu guten oder bösen Handlungen sowie sein „Herz“.<sup>40</sup>

Andererseits wird diese klare Struktur der Freiheit gleich nach ihrer Exposition in die überkommene Sprache der Erbsündendoktrin übersetzt:<sup>41</sup> „der Mensch ist von Natur gut, oder er ist von Natur böse“; der gute oder böse „Charakter“ ist ihm „angeboren“; die „eine oder andere Gesinnung“ habe man „als angeborene Beschaffenheit von Natur“.<sup>42</sup> Wenn fortan also von Natur die Rede ist, dann soll darunter nicht wie gewöhnlich Determination und Unfreiheit verstanden werden, sondern Freiheit und Spontaneität.<sup>43</sup> Der Gegensatzbegriff zu dem der Freiheit habe für die Strukturen der Freiheit selbst zu stehen, *mutato nomine de te fabula narratur*. Doch so einfach ist das nicht, denn die traditionelle Semantik bleibt selbst für den aufmerksamsten Leser bestehen, Natur heißt nicht Freiheit. Auch lässt sich der Übersetzungsakt oft nicht reibungslos vollziehen, da Kant verschiedene Naturbegriffe zu verwenden scheint<sup>44</sup>. Was Kant daher mit dieser Einverleibung des traditionellen Sprachgebrauchs konstituiert, ist wenigstens eine Rhetorik der Täuschung.<sup>45</sup>

Das Prinzip der Verwandlung von Natur in Freiheit und Freiheit in Natur ist in der Kantischen Philosophie jedoch nicht nur ein rhetorisches Prinzip, sondern auch ein philosophisches. Nach der Einleitung der ›Kritik der Urteilskraft‹ ist es sogar das Prinzip der „Philosophie überhaupt“, in der theoretische (Natur-) und praktische (Freiheits-)Philosophie vereint sind.<sup>46</sup> Die Freiheit soll in der Natur wirklich und somit Natur in Freiheit verwandelt werden: „Die Wirkung nach dem Freiheitsbegriffe ist der Endzweck, der (oder dessen Erscheinung in der Sinnenwelt) existieren soll“.<sup>47</sup> Dafür steht der Mensch in seiner Einheit von physischem „Thiermensch[en]“ (*homo phaenomenon*) und geistigem Freiheits- und „Vernunftmensch“ (*homo noumenon*)<sup>48</sup>. Nach Kant ist hierzu eine zweistufige Übertragungsarbeit notwendig: Die sinnlich nicht darstellbare, da übersinnliche Freiheit muss einerseits doch im Bereich des Sinnlichen praktisch-kausal dargestellt werden. Andererseits muss diese Freiheit dort auch sinnlich wahrnehmbar

<sup>40</sup> RGV, AA 06: 21.24–28, 25.01–09, 29.04–15.

<sup>41</sup> Dörflinger 2008, 88–90.

<sup>42</sup> RGV, AA 06: 20.35–21.23, 21.24–22.09, 25.01–05.

<sup>43</sup> RGV, AA 06: 20.35–21.01.

<sup>44</sup> Sirovátka 2015, 102.

<sup>45</sup> Diese Rhetorik bereitet Interpreten erhebliche Schwierigkeiten, die Kants Hang zum Bösen konsequent aus der Freiheitsperspektive lesen möchten, so z. B. Gressis 2018, 68.

<sup>46</sup> Kap. I–III und IX.

<sup>47</sup> KU, AA 05: 195.30 f.

<sup>48</sup> Vgl. MS, TL, § 11,1–3, AA 06: 434.23–435.22.

sein, damit sich der Mensch als Mensch identifizieren kann. Dazu bedarf es der Vermittlungstätigkeit einer reflektierenden Urteilskraft.

Diesem zweistufigen Modell liegt die kantische Charakterlehre zugrunde mit ihrer Unterscheidung eines intelligiblen und empirischen Charakters des Menschen. Beide verhalten sich so zu einander, dass der übersinnlich-intelligible Charakter der Freiheit „völliger Spontaneität“ den empirisch-sinnfälligen kausal verursacht, der empirische aber wiederum das „sinnliche Zeichen“ des intelligiblen ist.<sup>49</sup> So heißt es nach Auskunft der *Anthropologie*, der Oberbegriff „Charakter“ gehöre zur „allgemeine[n], natürliche[n] (nicht bürgerliche[n]) Zeichenlehre (*semiotica universalis*)“; er heiße lediglich „Unterscheidungszeichen“.<sup>50</sup> In metaphorischer Weise kann man also sagen: Einerseits wird der menschliche Charakter in einem Akt der Zeichensetzung sozusagen *geschrieben*; vor allem dann, wenn in der Charakterbildung der intelligible Charakter den empirischen kausal verursacht. Andererseits muss der Charakter aber auch grundsätzlich in Akten der Interpretation in gewisser Weise *gelesen* werden; vor allem, wenn es gilt, ihn trotz seiner Verwandlung in Natur als Zeichen der Freiheit auszulegen.

Die Charakterlehre gibt aber auch den Rahmen vor, in dem sich die Lüge abspielt. Laut der „Tugendlehre“ ist die Lüge der Grunddefekt Nummer Eins auf der Seite der Bildung des Charakters, weil durch sie die Ordnung von Sprache, Denken und Freiheit willentlich gestört wird.<sup>51</sup> Dort heißt es, die Lüge sei eine Perversion der Sprache als Vermögen der „Mittheilung seiner Gedanken an jemanden durch Worte“. Das habe zur Folge, dass man als Lügner innerlich wie äußerlich nur noch „eine bloß täuschende Erscheinung vom Menschen“ sei, „nicht der Mensch selbst“. So wird im Grundsatz der Lüge einer gesehen, „sich es zum Grundsatz [zu machen], keinen Grundsatz, und so auch keinen Charakter, zu haben“.<sup>52</sup> Präziser, aber auch widersprüchlicher könnte man auch von einem *Charakter der Charakterlosigkeit* sprechen, der hier *gewollt* ist.<sup>53</sup> In gewisser Weise kann man das als *Schreibdefekt* verstehen, der für den Leser zum *Interpretationsproblem* wird. Man könnte vielleicht sogar sagen, *Unlesbarkeit* sei intendiert; das Stichwort wäre ‚gewollte Unlesbarkeit‘.

---

49 KrV, A 546 f./B 574 f., A 551 f./B 579 f., A 553 f./B 581 f., A 556 f./B 584 f. Vgl. Willaschek, Marcus: *Praktische Vernunft*. Stuttgart/Weimar 1992, 135, 131–142.

50 Anth, § 86, AA 07: 285 f.

51 MS, TL, § 9,1, AA 06: 429.

52 MS, TL, § 4,5, AA 06: 420.19–26.

53 Diesen (gewollten) existenziellen Defekt muss Laura Papish in ihrer Studie unbeachtet lassen, wenn sie davon ausgeht, 2018, 69 f., der Selbstbetrug des Bösen könne nicht als (Selbst-)Lüge gedeutet werden.

Zudem ist die Charakterlehre eine weitere Basisstruktur der Schrift über das Böse, ihrer Rhetorik, wie auch der dortigen Thematisierung der (Selbst-)Lüge. Wenn im Text beispielsweise von Natur die Rede ist, aber Freiheit gelesen werden soll, dann ist das eine große Zumutung. Ihr entspricht jedoch exakt die Forderung der Vernunft, den empirischen Charakter (Natur) als Zeichen des intelligiblen (Freiheit) zu interpretieren. Damit ist der kantische Text zugleich eine Art ‚Leseschule‘ der praktischen Philosophie. An sich ist es zwar bereits eine Herausforderung, die in Natur verwandelte Freiheit konsequent als Freiheitsprodukt zu deuten, also als Kultur. Zum Problem wird das jedoch eigentlich erst durch das Böse, weil sich hier die Freiheit nicht selbstbewusst bejaht, sondern verneint<sup>54</sup>. Sätze über die angeborene Natur des Menschen, sofern sie nicht freiheitlich umgedeutet werden, entsprechen nun genau diesem Vorsatz. Als Sprechakte reihen sie sich in die Summe rhetorischer Strategien ein, die Spuren der Freiheit zu verwischen, welche das Böse zurechenbar machen würde. Sie alle bezwecken, den Akt der kausalen Festschreibung des empirischen Charakters durch den intelligiblen zu leugnen.<sup>55</sup> Es handelt sich hier bereits um eine Rhetorik der Selbstlüge.

Kant möchte in der Schrift über das Böse jedoch nicht primär den empirischen Charakter unter Vorzeichen der Freiheit gedeutet wissen, sondern die naturalisierende Sprache der Erbsündenlehre als Bezeichnung für die intelligente Freiheit und ihre Charakterarbeit etablieren. Auf der rhetorisch-polemischen Ebene ist das als (Gegen-)Strategie zu identifizieren, die freiheitswidrige Sprache der Erbsündenlehre im Freiheitsdenken selbst aufzulösen. Allerdings eignet sich die konventionelle Rede auch dafür, das Determinationsmoment einer Charakterbildung aus Freiheit zu bestimmen – beispielsweise in der Rede vom Angeboren-Sein des Charakters. Die Determination der Freiheit setzt Kant nämlich als größtmöglich an. So bezieht sich die oberste Maxime, als subjektiver Grund davon, wie der Mensch von seiner Freiheit überhaupt Gebrauch macht,<sup>56</sup> in systematisch-einheitsstiftender Absicht auf alle Handlungen des Menschen.<sup>57</sup> Ihre Bestimmung soll nicht auf einen Lebensabschnitt begrenzt sein, sondern sich in ihrer Allgemeinheit auf das ganze Leben der Person erstrecken.<sup>58</sup> So gesehen ist die Rede vom *Angeboren-Sein* eine angemessene Figuration des Totalitätsanspruches der Freiheit. Problematisch wird es jedoch auch hier erst, wenn die böse, sich selbst degradierende Freiheit ins Spiel kommt. Dann sorgt die Totalitätsstruktur für einen entsprechend großen Schaden: Der ganze Freiheitsgebrauch

---

<sup>54</sup> Dörflinger 2008, 94.

<sup>55</sup> Vgl. Gressis 2018, 69, 71–73.

<sup>56</sup> RGV, AA 06: 21.01–04.

<sup>57</sup> RGV, AA 06: 36.34–37.07.

<sup>58</sup> Willaschek 1992, 144 f.

der Person wird zu einem *System des Bösen* arrangiert, besser aber derangiert. Denn auch die Erosion der Zurechnungs- und Verantwortlichkeitsstrukturen praktischer Vernunft ist als Systemeigenschaft zu berücksichtigen, welche durch die Selbstlüge des Bösen gewollt wird.<sup>59</sup> Tut man dies, so gelangt man zu jenem „Suchtmodell“<sup>60</sup> einer mental gestörten Vernunft, dem weder zu widerstehen, noch zu entrinnen ist.

Kant setzt der Determination aus Freiheit jedoch wiederum die Freiheit als Heilmittel entgegen. Jederzeit soll es möglich sein, die totalisierende Systemstruktur in einer „Revolution der Denkungsart“ aufzubrechen, so das vierte Kapitel. Das Modell ist hier das Sieyes'sche der *pouvoir constituant* und *constitué*:<sup>61</sup> Gleichgültig wie viel Knechtschaft vonseiten der konstituierten Ordnungsmacht ausgeübt wird, vonseiten der ursprünglich-souveränen Gewalt kann damit jederzeit gebrochen werden. Diese residiere als wahre Souveränität in einem herrenlosen Niemandsland und sei gegen jede Bestimmung immun. So sei es möglich, die oberste Maxime kraft erstursprünglicher Freiheit einfach abzuändern,<sup>62</sup> wie man auch jede böse Handlung kraft dieser Freiheit einfach unterlassen könne<sup>63</sup>. Auch die gewollten Defekte der Zurechnungs- und Verantwortungsstrukturen hätten letztlich keine Rolle zu spielen. Hiermit scheint das „Suchtmodell“ der menschlichen Bosheit aufgehoben und mit ihm auch der Wahnsinn einer sich selbst belügenden Freiheit. Doch ist das Problem so einfach aus der Welt zu bringen? Wird durch diese Lösung im vierten Kapitel nicht die massive Problemstruktur unterschätzt, die zuvor in der Schlüsselstelle des dritten Kapitels aufgeworfen wurde? Um das zu beantworten, bedarf es zunächst einer genauen Interpretation des dritten Kapitels.

---

59 Vgl. Allison, Henry E.: „Reflections on the Banality of (Radical) Evil: A Kantian Analysis“. In: *Rethinking Evil*. Hrsg. von María Pía Lara. Berkeley et al. 2001, 86–100, 96: „it seems reasonable to conclude that self-deception is already for Kant (as it was to become to Fichte and in our century Sartre) the major mechanism that human beings use to evade their responsibility“. Allison spricht dort, 2001, 93, auch von „a kind of systematic self-deception“.

60 Horn 2011, 48, 65.

61 Sieyes, Emmanuel Joseph: *Qu'est-ce que le Tiers état?* Genf 1970, Kap. 5.

62 RGV, AA 06: 44.15–24.

63 RGV, AA 06: 41.17–25.

## 2 Der Selbstbetrug des Bösen (Kap. III)

### 2.1 Der vorangehende Kontext (Kap. III, 2–6)

Der Zusammenhang der Entscheidungen für das Böse einerseits und die Selbstlüge andererseits wird im achten und neunten Absatz des dritten Kapitels präsentiert. Die vorhergehenden Absätze bereiten die Einführung dieses Nexus vor, müssen aber auch rückwirkend im Licht dieser Problematik gelesen werden. So wird im zweiten Absatz (III, 2) zuerst das Dasein des Bösen bewiesen. Dann wird im vierten angeführt, dass man damit noch nicht über seine „eigentliche Beschaffenheit“ belehrt sei (III, 4). Diese Belehrung (III, 5f.) untergräbt jedoch rückwirkend und unausgesprochen den Beweis, indem sie zwei Prämissen nicht nur als kurzsichtig und falsch erkennbar macht, sondern als symptomatisch für den fundamentalen Selbstbetrug des Bösen.

Die erste Prämisse ist die Behauptung der *Sichtbarkeit* und *Interpretierbarkeit des Bösen* – von äußerlich bösen Handlungen könne man auf die oberste Maxime des Bösen als „Natur des Menschen“ schließen<sup>64</sup> und weil jeder Mensch wohl mindestens eine begangen habe, ist ein Schluss auf die Gattung möglich. Die zweite ist die einer gewissen *Würde des Bösen* als satanischer Widerstandshandlung – das Böse sei äußerlich sichtbar, weil es in einem offensichtlichen „Widerstreit der menschlichen Willkür gegen das Gesetz“ bestehe<sup>65</sup>. Auch wenn man dem Beweisverfahren eine gewisse Plausibilität einräumen kann,<sup>66</sup> ändert das nichts daran, dass mit ihm der eigentümliche Schaden des Bösen als Selbstlüge kaschiert wird. Denn einerseits suggeriert der Schluss von einer bösen Handlung auf den bösen Charakter, das Modell der Charakterbildung und -interpretation werde nicht gestört. Die eigentümliche Logik der (Selbst-)Lüge liegt jedoch wie gesagt genau darin: Nach ihrem Grundsatz wird der Charakter mit Vorsatz so gebildet, dass er nur noch Täuschung ist; der Vorsatz ist einer, keinen Charakter haben zu wollen im Sinne eines verlässlichen Unterscheidungszeichens (s. o.). Darum heißt es in der „Tugendlehre“, die Lüge sei „Wegwerfung und gleichsam

<sup>64</sup> RGV, AA 06: 20:22–34.

<sup>65</sup> RGV, AA 06: 35.27–30.

<sup>66</sup> Vgl. z. B. Sirovátka 2015, 14, 90–102, Bojanowski, Jochen: *Kants Theorie der Freiheit*. Berlin et al. 2006, 271f. sowie Willaschek 1992, 152, dagegen aber Horn 2011, 66, Fleischer, Margot: *Mensch und Unbedingtes im Denken Kants*. Freiburg/München 2009, 295 und Allison 1990, 154 f. Die Untersuchung der genauen Beweis- und Argumentationsstruktur gehört immer noch zu den Forschungsfragen und -desideraten, so auch Wyrwich, Thomas: „Böses, radikales“. In: *Kant-Lexikon, Band 1*. Hrsg. von Marcus Willaschek et al. Berlin/Boston 2015, 305–310, 309.

Vernichtung der Menschenwürde<sup>67</sup>. Somit erscheint das Narrativ des Bösen als Würdefigur mit rebellischer Haltung selbst als beschönigende Lüge vonseiten des Bösen als „Lügne[r] von Anfang“<sup>68</sup>, der sich mit der weggeworfenen Würde lediglich bekleiden will. Dem setzt Kant eine weitere Umdeutung entgegen: Mit Blick auf das menschliche Böse sei selbst der Satan und Teufel noch zu gut; das Böse sei, wenn nicht „Bosheit“, dann „doch wenigstens Nichtswürdigkeit“<sup>69</sup>.

Die Struktur der „Nichtswürdigkeit“ des bösen Charakters wird im dritten Kapitel jedoch *vor* dem Akt des Selbstbetruges festgelegt und damit *unabhängig* von ihm. Im *fünften Absatz* soll das, was das Böse sei, in der Entwicklung des Begriffes dargelegt werden. Demnach besteht der Akt der Setzung der obersten Maxime des Bösen darin, dass der Mensch „die Triebfeder der Selbstliebe und ihre Neigungen zur Bedingung der Befolgung des moralischen Gesetzes macht“. Ein kategorisches Wollen kann es folglich nicht mehr geben, da durch die oberste Maxime *alles* Wollen strukturell von den Bedingungen der Selbstliebe abhängig gemacht wird, die nun als „Prinzip *aller* [...] Maximen“ fungiert<sup>70</sup>. Zwar mag es immer noch Handlungen autonomer Freiheit geben, diese haben jedoch fortan ihren *Preis*: „»Ein jeder Mensch hat seinen Preis, für den er sich weggibt«“<sup>71</sup>. Handlungen, die „über allen Preis erhaben“ sind, mithin „eine Würde“ haben,<sup>72</sup> sind unter der obersten Maxime des Bösen also strukturell nicht mehr möglich. Das liegt daran, dass der Mensch in der Setzung dieser Maxime bereits seine „Autonomie“ als „Grund“ der „Würde“ seiner „Natur“<sup>73</sup> verkauft, weggeworfen, oder vernichtet hat – wenn auch nicht unwiderruflich.

Im Übergang zur Exposition des Selbstbetruges im achten und neunten Absatz nimmt der *sechste* eine Übergangsposition ein, die ganz im „Schwebezustand der List“ verweilt. Einerseits weist er auf die fundamentale (Lese-)Störung im Charaktermodell hin, welche das Böse durch seine innere Logik zu bewirken vermag: Der „empirische Charakter“ könne „gut“ sein, auch wenn der „intelligible [...] böse ist“. Andererseits scheint in diesem Hinweis aber auch eine Würdigung der traditionellen Ethik zu liegen, welche das Prinzip der Glückseligkeit nicht als Prinzip des Bösen erkennt, sondern als Prinzip des Guten propagiert. Bereits die Funktion einer obersten Maxime, „Einheit der Maximen überhaupt“ herzustellen, also der Systemgedanke selbst, könne dazu führen, dass „die

---

67 MS, TL, § 9,1, AA 06: 429.

68 RGV, AA 06: 42.17, Anm.

69 RGV, AA 06: 38.23–27.

70 RGV, AA 06: 45.13–15; Kursivdruck v. Vf.

71 RGV, AA 06: 38.35.

72 GMS, AA 04: 434.31–34.

73 GMS, AA 04: 436.01–07.

Handlungen [...] so gesetzmäßig ausfallen, als ob sie aus echten Grundsätzen entsprungen wären“. Sogar ein *Ende der Lüge* stehe in Aussicht: Der Grundsatz der „Wahrhaftigkeit“ überhebe „uns der Ängstlichkeit [...], unseren Lügen die Übereinstimmung zu erhalten und uns nicht in den Schlangenwindungen derselben selbst zu verwickeln“. Die Prinzipien des Guten und Bösen würden demnach gleichermaßen das Prinzip der Wahrhaftigkeit fordern, welches das Gegenprinzip zu dem der Lüge ist<sup>74</sup>.

Vor dem Hintergrund der Kantischen Charakterlehre erscheint der Satz nahezu irrwitzig, der empirische Charakter sei gut oder gar wahrhaftig, auch wenn der intelligible böse ist. Einerseits besagt diese Lehre, der *empirische* Charakter könne überhaupt nur deshalb ein moralisch-geistiger Charakter sein, insofern er als ‚lesbares‘ Zeichen des intelligiblen fungiert.<sup>75</sup> Andererseits beruht diese ‚Lesbarkeit‘ als *natürliches Zeichen* nur darauf, dass der intelligible Charakter der *Freiheit* die *Ursache* des empirischen ist.<sup>76</sup> Ohne Rückbezug auf den intelligiblen Freiheitsursprung kann der empirische deshalb weder als *moralisch*-gut noch -böse im „eigentlich[en]“ Sinn ausgelegt werden.<sup>77</sup> Umgekehrt wird diese Charakterlosigkeit aber auch von intelligibler Seite bedingt und zwar durch das Prinzip des Bösen als struktureller „Nichtswürdigkeit“. Der vermeintlich gute und wahrhaftige empirische Charakter muss also notwendig ein *falscher* sein. Was allein wahrhaftig sein würde, wäre eine Darstellung der Charakterlosigkeit. Doch das ist widersinnig.

Was Kant mit dem sechsten Absatz einleitet, ist eine nicht direkt ausgesprochene, aber leicht zu identifizierende Kritik der traditionellen Ethik als Glückseligkeitslehre (nach Aristoteles). Wie in der *Allgemeinen Anmerkung* weiter ausgeführt wird, leiste sie zwar durch Habitualisierungsprozesse eine beachtliche „Reform“ des Verhaltens, die „Gründung eines Charakters“ bleibe aber aus.<sup>78</sup> Deshalb könne es zwischen dem offensichtlich bösen und dem lediglich empirisch guten Charakter keinen wahren Unterschied geben, und folglich auch kein verlässliches Unterscheidungszeichen: „»Es ist hier kein Unterschied, sie sind allzumal Sünder — es ist Keiner, der Gutes thue (nach dem Geiste des Gesetzes), auch nicht einer.«“<sup>79</sup>. Diesen lügenhaften Anspruch erhebt jedoch die Ethik der Glückseligkeit.

74 Vgl. MS, TL, § 9,1, AA 06: 429.

75 Vgl. KrV, A 546/B 574, A 553/B 581.

76 Anth, § 86, AA 07: 285 f., Willaschek 1992, 135, 141 f.

77 Vgl. RGV, AA 06: 29.04–11.

78 RGV, AA 06, 46 f., vgl. Aristoteles: *Die Nikomachische Ethik*. Düsseldorf 2007, II, 1 sowie Gressis 2018, 69.

79 RGV, AA 06: 39.05 f. [Kant zitiert den Apostel Paulus.]

## 2.2 Das böse Herz als System des Selbstbetrugs des Bösen (Kap. III, 8–9)

Die Eigentümlichkeit der abschließenden Absätze ist die Kopplung der Entscheidung für das Böse mit dem Akt des Selbstbetruges. Sie kommt überraschend, vor allem wenn man sich entsinnt, dass das Prinzip des Bösen angeblich die „Wahrhaftigkeit“ als Grundsatz anrate. Aber auch die Bestimmung der obersten Maxime des Bösen kündigte nicht an, dass ein Selbstbetrug sie begleiten würde. Die Struktur der Kopplung wird nun dargelegt, indem der Selbstlüge im dreistufigen System des bösen Herzens ihr Ort und ihre Funktion zugewiesen wird.

Dieser Stufenbau kennzeichnet sich nach dem zweiten Kapitel zuallererst durch eine bemerkenswerte Dopplung. Einerseits ist das böse Herz die *Wirkung* des intelligiblen Aktes der Entscheidung für die Maxime des Bösen.<sup>80</sup> Entsprechend ist die Hauptstufe der „Bösartigkeit“ die Stufe jener „Verkehrtheit“ der Ordnung von Freiheit und Selbstliebe, welche die oberste Maxime des Bösen anordnet.<sup>81</sup> Andererseits wird diese Ordnung jedoch von genau dieser Stufe aus erst verursacht: Sie könne auch „die Verkehrtheit (*perversitas*) des menschlichen Herzens heißen, weil sie die sittliche Ordnung in Ansehung der Triebfedern einer freien Willkür“ umkehre und „dadurch“ werde erst „die Denkungsart [...] in ihrer Wurzel (was die moralische Gesinnung betrifft) verderbt“ (ebd.; Kursivdruck v. Vf.). Einerseits ist das böse Herz also Wirkung und Ausdruck der obersten Maxime, andererseits werde die oberste Maxime erst auf Basis eines solchen Herzens festgelegt.

Die Widersprüchlichkeit mag irritierend sein, bringt jedoch lediglich die Logik der intelligiblen Freiheit zum Ausdruck: Einerseits legt sie eine höherstufige Ordnung der Kausalität fest,<sup>82</sup> andererseits kann sie sich aufgrund ihrer Spontaneität jedoch auch in Einzelhandlungen beliebig über diese Ordnung hinwegsetzen, sie aber auch ebenso beliebig affirmieren<sup>83</sup>. Darin liegt die „absolut[e] Spontaneität der Willkür (der Freiheit)“<sup>84</sup>. Daher ist es konsequent, dass die intelligible Tat der Entscheidung für die Ordnung des Bösen dem bösen Herz sowohl vorgelegt ist, als auch in seiner Hauptstufe eine strukturelle Entsprechung findet, als permanenter Akt der Bejahung dieser Ordnung. So will Kant im vierten Kapitel den „Ursprung“ und „Anfang“ des Bösen keineswegs auf einen singulären Akt frühester Jugend beschränkt wissen: „Daß wir es täglich eben so machen, mithin

<sup>80</sup> Stangneth, Bettina: *Kultur der Aufrichtigkeit*. Würzburg 2000, 52.

<sup>81</sup> RGV, AA 06: 30.09–18.

<sup>82</sup> Willaschek 1992, 144 f.

<sup>83</sup> Bojanowaki 2006, 261 f.

<sup>84</sup> RGV, AA 06: 24.04 f.

»in Adam alle gesündigt haben« und noch sündigen, ist aus dem obigen klar<sup>85</sup>. Mit Blick auf die Struktur des Herzens ist das so zu verstehen: Durch die Setzung der obersten Maxime muss die Maxime selbst zwar nicht immer wieder aufs Neue angenommen werden; nichtsdestoweniger ist sie jedoch auf eine fortwährende Affirmation, also *Befestigung* im Freiheitsgebrauch angewiesen.

Die Selbstlüge wird nun als „Denkungsart“ (III, 8) und „Charakter“ (III, 9) ebenfalls auf der Hauptstufe des bösen Herzens verortet. Im Gegensatz zur Exposition des Stufenbaus im zweiten Kapitel erfolgt die Verortung jedoch in organischer Rückbindung an die beiden ersten Stufen. Allerdings geschieht auch dies wieder im „Schwebezustand der List“, Strukturen der Lüge nicht als solche zu identifizieren. Es bleibt wie gesagt unausgesprochen, dass die Selbstlüge auf der Hauptstufe die strukturelle Verlogenheit der vorhergehenden Stufen bedingt<sup>86</sup> – was freilich nicht so zu verstehen ist, dass sie *das Böse* in diesen Stufen verursacht<sup>87</sup>.

Die strukturelle Verlogenheit der ersten beiden Stufen – der „Gebrechlichkeit“ und „Unlauterkeit“ – ließ sich bereits im zweiten Kapitel erkennen. Dort sollte das ganze böse Herz mitsamt seiner drei Stufen auf dem Akt der Verkehrung der Triebfederordnung gründen, der zugleich in der dritten Stufe strukturell verankert sei. Deshalb setzten die ersten beiden Stufen die Verkehrung dritter Stufe voraus. Die bewirkt jedoch mit der Degradierung der Triebfeder der Freiheit genau das Kraftdefizit, welches für die ersten beiden Stufen charakteristisch ist.<sup>88</sup> So ist, gemäß dem *Entschlüsselungscodex* der *Einleitung* (s. o.),<sup>89</sup> die „Gebrechlichkeit der menschlichen *Natur*, zu Befolgung seiner genommenen Grundsätze nicht stark genug zu sein“, eben die Gebrechlichkeit der menschlichen *Freiheit*, die Maxime des Böse zu wollen und damit die Gebrechlichkeit als zweite Natur (Kursivdruck v. Vf.). „Wollen habe ich wohl, aber das Vollbringen fehlt“ – diese „Klage eines Apostels“ (Paulus) ist darum die charakteristische (Selbst-)Lüge der ersten Stufe des menschlichen Herzens.<sup>90</sup> Denn das Böse „erhält hier den Namen einer Schwachheit“<sup>91</sup>. In der „Tugendlehre“ ist *das* bereits als „Unlauterkeit in Erklärungen, die man gegen sich selbst verübt“, zu sehen, folglich als (Selbst-)Lüge.<sup>92</sup> Was in 1793 unausgesprochen bleibt, kommt also 1797 explizit zur Sprache.<sup>93</sup>

85 RGV, AA 06: 42.20 f.

86 Allison 1990, 159.

87 Potter 2002, 386, vgl. dagegen Rukgaber 2015, 235, Fn. 4.

88 Ebd.

89 RGV, AA 06: 20.35–21.05.

90 RGV, AA 06: 29.24–30, vgl. Stangneth 2000, 53 f.

91 MS, TL, § 9,5, AA 06: 430.27–431.03.

92 Ebd., vgl. RGV, AA 06: 37.18 f.

93 Auch das relativiert den Einwand Rukgabers, 2015, 242, maßgeblich.

Diese Problemstruktur wird auf der zweiten Stufe des bösen Herzens nur fortgetragen.

Dass es sich auf den ersten beiden Stufen um eine unhintergehbare empirisch-anthropologische Schwäche handelt, wird jedoch in einer Rhetorik der Täuschung durchgehend suggeriert. Dies fängt mit der Verwendung des naturalistisch konnotierten Naturbegriffes an und gipfelt in der vermeintlichen Genealogie, die Bösartigkeit dritter Stufe entspringe aus der Kombination der ersten beiden und könne „mit einem, im allgemeinen guten Willen zusammen bestehen“<sup>94</sup>. Letzteres suggeriert, die oberste Maxime des Guten sei bereits gesetzt.<sup>95</sup> Das kann aber unter der Voraussetzung der „Gebrechlichkeit der menschlichen Natur“ nicht der Fall sein – wenn man das Wort ‚Natur‘ hier wie gefordert als ‚Freiheit‘ umdeutet.

Die Ursprungsgeschichte des Bösen dritter Stufen aus den beiden vorigen ist jedoch nicht nur ein rhetorisches Täuschungsmanöver. Sie bestimmt wie gesagt auch die primäre und zentrale Funktion der Selbstlüge auf der Hauptstufe des Bösen in organischem Bezug zu den beiden anderen Stufen. Der *Ansatzpunkt* ist dabei, dass die Kombination von Gebrechlichkeit und Unlauterkeit nicht notwendig das „Laster“ bewirkt, sondern eher eine weitgehende Legalität der Handlungen. So stand die „Unlauterkeit“ bereits in II,4 dafür, dass das Gute zwar getan wird, doch nicht allein um des Guten willen. Der Ansatzpunkt ist also die innere (Herzens-)Struktur, die dem scheinbar guten empirischen Charakter bei vorausgesetzter Bösartigkeit des intelligiblen zugrunde liegt, also das böse Herz des „sittlich guten Menschen“<sup>96</sup>. Wegen dieser vermeintlichen Gutartigkeit scheinen die ersten beiden Stufen keine Stufen der „Verkehrtheit des Herzens“ zu sein – zumindest solange man davon absieht, dass beide auf dem fundamentalen Akt der Verkehrung der Triebfederordnung gründen, dem Bösen dritter Stufe. Genau diese Bösartigkeit dritter Stufe wird jedoch mit der für die beiden ersten Stufen charakteristischen „Denkungsart“ ins Spiel gebracht. Die Denkungsart ist jedoch wiederum der intelligible und eigentliche Charakter des Menschen.<sup>97</sup> Damit liegt zugleich der Einheits Gesichtspunkt vor, der das menschliche Herz auf allen Stufen zu *einem* bösen, verkehrten Herzen macht.

Die einheitsstiftende „Denkungsart“ besteht nun in einem Akt der *Interpretation*, „die Abwesenheit desselben [des Lasters] schon für Angemessenheit der *Gesinnung* zum Gesetze der Pflicht (für *Tugend*) *auszulegen*“ (Kursivdruck v. Vf.). Der scheinbar gute empirische Charakter wird fälschlicherweise als guter

<sup>94</sup> RGV, AA 06: 37.23–31.

<sup>95</sup> Dies wird in der Literatur weitgehend angenommen; zur Übersicht und Kritik: Rukgaber 2015, 236f., 239.

<sup>96</sup> RGV, AA 06: 30.24–31.05, 36.34–37.07.

<sup>97</sup> KrV, A 551/B 579.

*intelligibler* Charakter gedeutet. Genau *dieser* Selbstbetrug dritter Stufe konstituiert nach III, 9 die „Tücke des menschlichen Herzens (*dolus malus*)“ als einer dreistufigen Ganzheit der Verlogenheit, „sich wegen seiner eigenen guten oder bösen Gesinnungen selbst zu betrügen“.

Weil der Selbstbetrug unmittelbar die *Gesinnung* betrifft, ist es nicht allzu schwer, den Zusammenhang zwischen dem Akt der Setzung der obersten Maxime des Bösen einerseits und dem Akt des Selbstbetrugs andererseits zu rekonstruieren – und mit ihm die Logik der Hauptstufe des bösen Herzens. Kant hatte nämlich in der Einleitung die Gesinnung und die oberste Maxime im Begriff des „erste[n] subjektive[n] Grund[es] der Annehmung der Maximen“ gleichgesetzt<sup>98</sup>. Der Akt der Setzung der obersten Maxime und der Akt des Selbstbetrug treffen sich damit auf der elementaren Ebene der Entscheidung, wie von der „Freiheit überhaupt“ Gebrauch zu machen sei<sup>99</sup>. Wie die Selbstlüge hier wirksam sein kann, darauf gibt die Rede von der „Unredlichkeit, sich selbst blauen Dunst vorzumachen“<sup>100</sup> einen Hinweis mit ihrer Anspielung auf die Technik der Zauberkünstler. Durch den Kunstgriff der Verwendung von blauem Dunst sollte der entscheidende Moment des Kunststückes den Augen des Zuschauers entzogen werden. Ebenso ist auch die Koordination der beiden Akte auf der Hauptstufe des bösen Herzens zu verstehen: Der entscheidende Akt der Setzung der obersten Maxime des Bösen gehe mit dem Akt der Leugnung jenes Aktes einher. Dabei ist der blaue Dunst der „Tugendschei[n]“<sup>101</sup> des vermeintlich guten empirischen Charakters. Er ist der Kunstgriff, mit dem der intelligible Charakter und mit ihm die wahre Denkungsart verdeckt werden soll.

Die Freiheitsakte der Maximensetzung einerseits und der Selbstlüge andererseits sind demnach zwei separate, beigeordnete Akte *vorsätzlicher Schuld*, die sich jedoch zu einer einzigen Denkungsart und einem einzigen intelligiblen Charakter des Bösen vereinigen. Der Wahnsinn solch einer Selbsttäuschung des Bösen („*dolus malus*“) besteht hierbei darin, dass er als Praxisform erstursächlicher Freiheit in der Hauptstufe des menschlichen Herzens quasi institutionalisiert ist. „Dass wir es täglich ebenso machen“ gilt also nicht nur hinsichtlich des Vorsatzes, die eigene Freiheit für einen Preis verkaufen zu wollen, sondern auch dafür, sich stets mit „Bewusstsein“ darüber belügen zu wollen. Das ist der *Wahnsinn* der *atemporalen Freiheitstat*: Einerseits soll die Verantwortungs- und Zurechnungsstruktur dem Bewusstsein entzogen werden. Die vermeintliche Unvorsätzlichkeit des Bösen auf den ersten beiden Stufen sind Ausdruck davon.

---

<sup>98</sup> RGV, AA 06: 25.05 f.

<sup>99</sup> RGV, AA 06: 21.01–04.

<sup>100</sup> RGV, AA 06: 38.23 f.

<sup>101</sup> RGV, AA 06: 33.22.

Andererseits soll diese Verdrängung aber auch zugleich und zeitgleich als vorsätzliche Verschuldung erfolgen, und das heißt mit „Bewusstsein“, Schaden anzurichten zu wollen<sup>102</sup> – eben als Selbst-*Lüge*. Um die atemporale Freiheit des Bösen steht es also noch schlimmer als Christoph Horn mit Blick auf das Erinnerungsproblem diagnostizierte<sup>103</sup>. Sie scheint diejenige einer geisteskranken Verunnt im buchstäblichen Sinne zu sein; vielleicht ist sie aber auch nur eine, die ihr wahres Selbst im blauen Dunst des Schlafmohns vergessen will – nach Kant vergeblich.<sup>104</sup>

Im Kantischen Spätwerk werden die wichtigsten Aussagen oft in Form von Einklammerungen präsentiert, die suggerieren, es handle sich um etwas Nebensächliches. Auf diese Weise wird in III, 9 auch der Grund angegeben, weshalb das „radical[e] Bö[s]e der menschlichen Natur [...] den faulen Fleck unserer Gattung ausmacht“: „(indem es die moralische Urtheilskraft in Ansehung dessen, wofür man einen Menschen halten solle, verstimmt und die Zurechnung innerlich und äußerlich ganz ungewiß macht)“.

Zwar hieß es zuvor, dass die Menschen die „Denkungsart [...] in ihrem Inneren wohl aufdecken könnten, wenn sie nur wollten“, wobei die sich selbst verdeckende böse Denkungsart *in toto* gemeint war. Um die „moralische Urtheilskraft in Ansehung dessen, wofür man einen Menschen halten solle“, also welchen moralisch-intelligiblen Charakter er habe, ging es auch hier. Nun soll dieses Vermögen jedoch durch das Böse grundsätzlich „verstimmt“ sein, wozu auch eine ebenso grundsätzliche Störung der Zurechnungsfunktion zu zählen sei. Die Zurechnung soll „innerlich und äußerlich *ganz ungewiß*“ sein (Kursivdruck v. Vf.). Im Kontext ist das so zu verstehen, dass man in keiner Weise gewiss sein könne, ob andere oder man selbst einen guten Charakter haben, gute Taten vollbringen, oder doch böse. Dabei ist die Verstimmung der moralischen Urteils-kraft auch zweifach durch die Selbstlüge bedingt: Weil die Selbstlüge systematisch am zentralen Punkt der Entscheidung über den ganzen Charakter situiert ist, bewirkt sie einerseits jene grundsätzliche ‚Unlesbarkeit‘ des ganzen Charakters einschließlich all seiner Einzelhandlungen. Das System der bösen obersten Maxime ist in Kombination mit der Selbstlüge ein allumfassendes *System der Lüge*. Andererseits muss die moralische Urteils-kraft aber auch selbst immer unter Verdacht stehen, Gutes wie Böses immer lügenhaft zurechnen zu wollen. Das Böse verstimme auch die moralische Urteilskraft selbst. Dieser doppelte und allumfassende Schaden der Selbstlüge des Bösen sei der „faul[e] Fleck unserer Gattung“, welche das „radical[e] Bö[s]e der menschlichen Natur“ ausmache.

<sup>102</sup> Vgl. MS, AA 06: 224.04–08.

<sup>103</sup> Horn 2011, 65 f.

<sup>104</sup> MS, TL, § 13, 2, AA 06: 438.13–23.

Der außerhalb der Klammer liegende Text suggeriert indes, man könne diesen „faulen Fleck“ einfach „herausbringen“. Nach der *Allgemeinen Anmerkung* soll das zunächst auch einfach so möglich sein, man müsse nur die oberste Maxime ändern.<sup>105</sup> Das habe eine „Revolution in der Gesinnung“ zur Folge, „eine Art von Wiedergeburt gleich als durch eine neue Schöpfung [...] und Änderung des Herzens“.<sup>106</sup> Das Problem der Selbstlüge scheint damit gelöst und der dadurch bedingte Schaden behoben.

Allerdings stellt sich am Ende heraus, dass (a) trotz vollzogener Herzensänderung die in der Klammer dargelegte Problemstruktur weiter fortbesteht.

Zur Überzeugung aber hievon kann nun zwar der Mensch natürlicherweise nicht gelangen, weder durch unmittelbares Bewußtsein, noch durch den Beweis seines bis dahin geführten Lebenswandels: weil die Tiefe des Herzens (der subjective erste Grund seiner Maximen) ihm selbst unerforschlich ist.<sup>107</sup>

Zugleich wird (b) der „Satz vom angeborenen Bösen“ aber auch auf die neue Ebene der „moralischen Asceetik“ verlagert. Sie soll Auskunft darüber geben, was man täglich machen müsse, um ein guter Mensch zu sein. Hier

[...] will dieser Satz mehr, aber doch nichts mehr sagen als: wir können [...] nicht von einer uns natürlichen Unschuld den Anfang machen, sondern müssen von der Voraussetzung einer Bösartigkeit der Willkür [...] anheben und, weil der Hang dazu unvertilgbar ist, mit der unablässigen Gegenwirkung gegen denselben.<sup>108</sup>

Allein durch die Änderung der obersten Maxime kann der „faule Fleck unserer Gattung“ also doch nicht so einfach herausgebracht werden. Man könnte also meinen, das Böse könne durch die Änderung der obersten Maxime letztlich doch nicht „vertilgt“ werden.<sup>109</sup> Führt man jedoch die Unerforschlichkeitsbehauptung (a) sowie die Funktion des Satzes vom Bösen in der Askeetik (b) auf das Problem der *Selbstlüge aus Freiheit* zurück, so lässt sich die Schwierigkeit klären. Zugleich lässt sich dadurch aber auch erst der Folgenreichtum des Problems umreißen.

**105** RGV, AA 06: 44.15–24.

**106** RGV, AA 06: 47.22–28.

**107** RGV, AA 06: 51.12–16.

**108** RGV, AA 06: 50.25–51.06.

**109** Vgl. Horn 2011, 65 sowie Klemme, Heiner F.: „Die Freiheit der Willkür und die Herrschaft des Bösen“. In: *Aufklärung und Interpretation*. Hrsg. von Heiner F. Klemme et al. Würzburg 1999, 125–151, 139. Rukgaber 2015, 244 f., 249, stellt hingegen die Möglichkeit der Disposition über die oberste Maxime (des Bösen) grundsätzlich in Abrede.

### 3 Konsequenzen des Selbstbetrugs des Bösen

Warum ist der Hang zum Bösen „unvertilgbar“? In der Literatur findet sich die plausible und letztlich Rousseau'sche Antwort, die empirische Natur (Anthropologie) habe sich unter der Maxime des Bösen dahingehend geändert, dass eine physische Abhängigkeitsstruktur entstanden sei, die den Menschen fortwährend zum Bösen dränge.<sup>110</sup> Dieser physische Hang zum Bösen bestehe auch dann noch, wenn der moralisch-intelligible längst überwunden sei – als „Versuchung“<sup>111</sup>. Dieser Interpretation soll nicht widersprochen werden. Sie gilt es lediglich um eine Tiefenebene zu ergänzen, die selbst dann bestehen würde, wenn sich kein physischer Hang zum Bösen herausgebildet hätte<sup>112</sup>.

Laut der Klammer am Ende von III, 9 ist das radikale Böse deshalb der „faul[e] Fleck unserer Gattung“, weil es „die moralische Urtheilskraft [...] verstimmt und die Zurechnung [...] ganz ungewiß macht“. Die *bloße Möglichkeit* der Selbstlüge des Bösen reicht jedoch bereits dafür aus, dass der Mensch (a) „natürlicherweise“ niemals zur „Überzeugung“ vom dauerhaften Gelingen seiner Revolution kommen kann. Denn im Gegensatz zur intelligiblen Einsicht in das Prinzip des Bösen<sup>113</sup> steht das Aufdecken des Prinzips des Guten immer im Schatten einer möglicherweise verlogenen Urteilskraft. Dadurch muss die Zurechnung „ganz ungewiss“ bleiben. Auch der „Beweis seines bis dahin geführten Lebenswandels“ muss ebenfalls immer unter Verdacht stehen, der Einsatz von „blau[e]m] Dunst“ zu sein, der die Entscheidung zum Bösen kaschieren soll. Allein die *Möglichkeit* all dessen reicht bereits aus, die „Tiefe des Herzens“ als „unerforschlich“ zu erklären: Ob der Mensch eine gute oberste Maxime hat, kann er niemals mit Gewissheit erkennen; obwohl das Gute „ursprünglich in unser Herz *geschrieben*“ ist,<sup>114</sup> ist letzteres sozusagen *unlesbar*. Epistemische Argumente müssen also nicht notwendig herangezogen werden, obwohl es selbstverständlich gute Gründe gibt, dies zu tun.<sup>115</sup> Doch angesichts der in III, 9 eingeräumten Einsicht in die böse Denkungsart ist auf jeden Fall zu fragen, inwieweit solche Argumente tatsächlich ausschlaggebend sind. Dort heißt es nämlich erstaunlicherweise: Die

**110** Vgl. Rousseau, Jean-Jacques: *Diskurs über die Ungleichheit. Discours sur l'inégalité*. Paderborn 2008, 228–233 sowie Sirovátka 2015, 97, Forschner, Maximilian: „Über die verschiedenen Bedeutungen des ‚Hangs zum Bösen‘“. In: *Klassiker Auslegen*, Band 41. Hrsg. von Otfried Höffe. Berlin 2011, 71–90, 88 f. und Klemme 1999, 139 f., 142.

**111** Forschner 2011, 88.

**112** Vgl. Bojanowski 2006, 273–277.

**113** RGV, AA 06: 38.17–20.

**114** Vgl. RGV, AA 06: 104.21–23, sowie Augustinus: *Confessiones. Bekenntnisse*. Stuttgart 2009, II,4; Kursivdruck v. Vf.

**115** Vgl. z. B. Forschner 2011, 82, Bojanowski 2006, 273–277, Sirovátka 2015, 100.

Menschen könnten die böse „Denkungsart [...] in ihrem Inneren wohl aufdecken [...], wenn sie nur wollten“;<sup>116</sup> die „intelligible Tat“ sei „bloß durch die Vernunft ohne alle Zeitbedingung erkennbar“<sup>117</sup>.

Aus dem Grund der bloßen Möglichkeit der Selbstlüge ist aber auch (b) der „Satz vom angeborenen Bösen“ einerseits für die moralische Urteilskraft unabdingbar und zwar als Prinzip ihrer Reflexion des Charakters und seiner Taten: Durch den Vorwurf, kraft falscher Zurechnungen die Leugnung der bösen Gesinnung auszuführen, darf sie nicht „von einer [...] natürlichen Unschuld den Anfang machen“, sondern muss „von der Voraussetzung einer Bösartigkeit der Willkür in Annehmung ihrer Maximen [...] anheben“. Andererseits müsse aber auch täglich „mit der unablässigen Gegenwirkung gegen denselben“ angesetzt werden. Das ist ebenfalls nicht erst aufgrund einer sinnlichen Versuchung zum Bösen als Erblast früher Zeiten notwendig, sondern bereits aufgrund jener Möglichkeit, sich über seine gute Gesinnung betrügen zu können. Sie macht bereits eine Praxis erforderlich, in der beständig den Gefahren der Selbstlüge entgegenwirkt wird – auch wenn man sie vielleicht noch niemals begangen hat. Zwar ist das aufgrund der augenscheinlichen Bösartigkeiten der Menschen<sup>118</sup> unwahrscheinlich, offenbar wird jedoch in Kauf genommen, dass man gegen Windmühlen kämpft. Interessant ist auch, dass in diesem Kontext der Begriff der (religiösen) Hoffnung etabliert wird<sup>119</sup>. Dass der Hang zum Bösen im Ersten Stück der Religionsschrift auch als systematisch wichtiges Grundelement eingeführt wird, darauf hat bereits Heiner F. Klemme hingewiesen<sup>120</sup>. Das Problem des Selbstbetruges aus Freiheit codiert dieses System jedoch auf eigenwillige Weise neu.

Die erforderte Praxis moralischer Asketik besteht in der Vereinigung von moralischer Revolution und Reform<sup>121</sup>. Auch sie gilt es angesichts des Problems der Selbstlüge zu interpretieren. So ist bezeichnenderweise von einer „allmähliche[n] Reform des Hanges zum Bösen, als verkehrter *Denkungsart*“ die Rede (Kursivdruck v. Vf.). Die Denkungsart ist es nämlich, von der die Selbstlüge ausgeht und auf die sie sich primär bezieht. Zwar verbindet Kant die „Revolution für die Denkungsart“ – die Änderung des intelligiblen Charakters durch die Umkehrung der obersten Maxime – mit der „allmähliche[n] Reform für die Sinnesart“. Unter letzterem ist das Projekt zu verstehen, einen guten *empirischen* Charakter zu erzeugen. Doch vor allem letzterer muss immer unter Verdacht

---

**116** RGV, AA 06: 38.17–20.

**117** RGV, AA 06: 31.32–33, Dörflinger 2008, 100 f.

**118** RGV, AA 06: 32.34–34.17.

**119** RGV, AA 06: 51.12–52.15.

**120** Klemme 1999, 127 f.

**121** RGV, AA 06: 47.29–48.16.

stehen, ein falsches Zeichen des Bösen zu sein. Einerseits ist deshalb sowohl der *empirische* als auch der *intelligible* Charakter fortwährend einer *kritischen Auslegung* zu unterziehen, um der Möglichkeit einer lügenhaft-boshaften Auslegung entgegenzuwirken. Andererseits muss der intelligible Charakter auf Basis dieser kritischen Aufmerksamkeit aber auch sich selbst sowie den empirischen Charakter im Geist jener Revolution *fortwährend neu konstituieren*. Denn wie der Akt der Setzung der obersten Maxime des Bösen sowohl dem bösen Herzen vorangeht, als auch in dessen Hauptstufe eine strukturelle Entsprechung findet, so muss ebenso die Revolution der Denkungsart in der Hauptstufe des guten Herzens institutionalisiert werden – als fortwährende Reform der „Denkungsart“. Auch hier gilt wieder: Zwar wird die Willkür mit dem Beschluss der guten obersten Maxime einmalig und für immer („unveränderlich“)<sup>122</sup> auf den Modus der Kür des guten Willens festgelegt. Doch die Festlegung ist durch die *freie Willkür* eben auch spontan im Geiste genau dieser Festlegung zu vollbringen.

Kants Schrift über das Böse geht jedoch nicht nur von der *Möglichkeit* des Bösen und der Selbstlüge aus, sondern auch von ihrer *Wirklichkeit*. In der moralischen Asketik müssen daher auch die Folgen dieser Wirklichkeit einkalkuliert werden, selbst dann, wenn man nur von der bloßen Möglichkeit ausgeht. Einerseits ist hier die Rousseau'sche Lesart vom *physischen Hang zum Bösen* mit dem Problem einer Selbstlüge aus Freiheit zusammen zu denken. So ist davon auszugehen, dass die empirische Natur auch durch den Willen zur Selbstlüge verändert wird. Hier ist zwar zuerst einmal an die Ebene der Bedürfnisse und Neigungen zu denken, auf die „Laster der Kultur“ gepropft werden können.<sup>123</sup> Primär wird sich jedoch der Selbstbetrug in psychischen Strukturen unbewusster Umdeutung, Verdrängung, aber auch Überredung verankern. Diese Strukturen werden sicher auch unter dem Willen zur moralischen Revolution eine Gefährdung bleiben.

Andererseits ist es nicht zufällig, dass das zentrale Problem der Selbstlüge des Bösen auch im Anfangsstück einer *Religionsschrift* exponiert wird. Schließlich bedingen die zwei Prinzipien des intelligiblen Charakters zugleich auch verschiedene Vorstellungen, Funktionalitäten und Wirklichkeiten von Staat, Religion und Kirche. Weil das Prinzip des Bösen mit dem der Selbstlüge einhergeht, so ist klar, dass sich die „Unredlichkeit, sich selbst blauen Dunst vorzumachen“ auch auf soziopolitischer Ebene als *Ideologie* manifestieren wird.<sup>124</sup> Das Problem des vermeintlich guten empirischen Charakters böser Gesinnung wird sich auch auf soziale Institutionen und Praktiken übertragen. So war nach Rousseau der

---

122 RGV, AA 06: 51.07–12.

123 RGV, AA 06: 26.12–27.26.

124 Vgl. Kants Diskussion von „Fetischglaube“ und „Pfaffentum“ im *Vierten Stück der Religionsschrift*.

Betrugsvertrag der Reichen dem Wortlaut nach mit dem Gesellschaftsvertrag identisch – und trotzdem hieß es von ersterem, er sei den Prinzipien des Naturrechts zuwider.<sup>125</sup> Doch wie lässt sich auf dieser Ebene der Unterschied zwischen ideologischen Strukturen und wahrhaftigen ausmachen?

Die Rousseau-Forschung hat darauf hingewiesen, dass Rousseau mit dem Projekt einer Analyse der sprachlichen Verfassung des Denkens unter Berücksichtigung der Konfiguration der Sprache durch Interessen die „stärkste Form der Ideologiekritik“ vorgelegt habe.<sup>126</sup> Der Kern dieses Projekts ist auch in der Kantischen Schrift über das Böse aufzufinden, und zwar in der dortigen Weiterführung der Charakterlehre: Wie die Deformation des Charakters durch die Lüge zeigt, konstituiert sich der Individualcharakter einer Person im Feld der Beziehung von Sprache, Denken und Wille. Hier wird er als Unterscheidungszeichen festgeschrieben. Durch einen Rückgang auf diese Ebene lasse sich zumindest (interpretativ) aufdecken, ob der angeblich gute äußere Charakter tatsächlich falsch, verlogen und böse sei. Überträgt man das Modell auf die soziopolitische Ebene, so gelangt man direkt zur Rousseau'schen Kritik von Sprechakten. Die Verlogenheit politischer Sprechakte, die äußerlich das Gute deklarieren, lasse sich aufdecken, wenn man die Artefakte der Wort- und Schriftebene auf die Akte ihrer Entstehung zurückführt; die Sprachanalyse ist Analyse der Rhetorik.<sup>127</sup> Die Rhetorik des Bösen sei jedoch die traditionelle Rhetorik als amoralische und letztlich lügenhafte Überredungskunst,<sup>128</sup> die des Guten müsse sich hiervon unterscheiden<sup>129</sup>.

Nach Rousseau ist die Sprache des wahrhaft Guten eine künftige, die man erst zu kultivieren habe.<sup>130</sup> Impliziert nun auch Kants moralische Revolution zu einem „neuen Menschen“<sup>131</sup> ebenfalls, wie bei Rousseau, die Kultivierung einer neuen Sprache? Insofern Kant den Anfang in der „Gründung eines Charakters“ sieht,<sup>132</sup> ist dies in gewisser Weise zu bejahen: Die moralische Revolution fordert auch ein neues Unterscheidungszeichen, das im Rahmen der *semiotica universalis* im weitesten, Rousseau'schen Sinn von Sprache zu verstehen ist. Diese neue

---

125 Rousseau 2008, 214–217.

126 Poppenberg, Gerhard: „*ins Freie*. Probleme figurativer Sprache nach Rousseau und de Man“. In: de Man, Paul: *Allegorien des Lesens II*. Berlin 2012, 271–346, 307.

127 Beziehe Rousseau 2008, 214–217 auf 117–131.

128 Rousseau 2008, 122f. Vgl. den Zusammenhang zwischen „Selbsttäuschung“, „Fetischglaube“ und „Überredung“ im *Vierten Stück* der *Religionsschrift*, so z. B. in AA 06: 193.18–25, 171.11–19 und 174.09–21.

129 Rousseau 2008, 120–123, 130f.

130 Ebd.

131 RGV, AA 06: 47.18–28, 47.36–48.01.

132 RGV, AA 06: 48.17–19.

Sprache bestimmt Kant in der *Schiller-Anmerkung*:<sup>133</sup> Sie ist die Sprache des Frohsinns, das Gute im Geiste der Freiheit und Autonomie zu tun, wahrhaftig, ohne „verborgenen Haß“ gegen die Freiheit und ihr Gesetz. Die fröhliche „Gemütsstimmung“ sei „das Temperament der Tugend“, die Sinnesart des Guten, der wahre gute empirische Charakter. Kant geht sogar so weit, ihn als „Zeichen der Echtheit tugendhafter Gesinnung“ zu deklarieren. Dieser Charakter ist dabei im Sinne einer *Körpersprache* zu verstehen, die Ausdruck einer Geisteshaltung ist, sich aber auch auf die *Wort- und Schriftsprache* erstreckt. Allerdings muss auch diese Sprache des Frohsinns angesichts der Möglichkeit der Selbstlüge notwendig in Verdacht stehen, ein Selbstbetrug des Bösen zu sein in seinem *pursuit of happiness*. Der „Wahnsinn der Freiheit“ erstreckt sich auch auf sie; die Selbstlüge des Bösen ist und bleibt ein Problem.

**Danksagung:** Den Gutachtern, Hans Friedrich Fulda sowie Brigitta Sophie von Wolff-Metternich danke ich für Anregung und Kritik.

---

133 RGV, AA 06: 23.01, Anm.