

Sie sollten den Titel des Vortrags und die Tatsache, dass es der letzte Vortrag ist, nicht so verstehen, dass ich den Versuch unternehmen möchte, die verschiedenen Diskussionsstränge, die wir hatten, und die ganze Vielfalt von Beiträgen hier etwa zusammen zu führen, eine kohärente Beurteilung abzugeben oder gar die unterschiedlichen Ansätze zu dieser Thematik - bzw. zu den beiden Teilen unseres Interdisziplinären Projektes „Naturgeschichte der Freiheit“ und „Funktionen des Bewusstseins“ - zu bewerten. Das, was ich hier vortrage, ist vielmehr eine Fortsetzung eigener Arbeiten, wobei ich einen Teil der Überlegungen, die ich schon publiziert habe, hier noch einmal soweit präsentiere, dass auch diejenigen, die sie nicht kennen, der Argumentation folgen können. Die Thesen werden zum Teil kontrovers sein. Es ist also auch von daher kein Abschlussvortrag (Abschlussvorträge sollten befrieden). Ich werde schon aus Zeitgründen manche eigentlich notwendige Differenzierung unterschlagen, aber wir haben ja dann noch Zeit für eine Diskussion in der das nachgeholt werden kann.

Der Vortrag zerfällt in vier Teile. Ich sage zunächst etwas zu den *Formen von Bewusstsein*, allerdings nur sehr kurz, weil dazu in unserer Runde schon intensiv diskutiert worden ist. Der eigentlich systematische Teil beginnt mit dem zweiten Teil, der sich mit dem Verhältnis von *Bewusstsein und Intentionalität* befasst, es folgt ein dritter Teil zum Verhältnis von *Intentionalität und Freiheit* und schließlich ein vierter zu *Selbstbewusstsein und Freiheit*.

I. Formen von Bewusstsein

Es gibt eine weit verbreitete Darstellung, die etwa folgendermaßen beginnt: Wir verstehen sehr Vieles und sehr Unterschiedliches unter diesem Begriff „Bewusstsein“, manchmal wird er so gebraucht, dann wieder ganz anders und diese unterschiedlichen Bedeutungen sollte man sorgfältig auseinander halten. Ein Aufsatz von Peter Bieri beginnt etwa so: Das Wort „Bewusstsein“ sei vieldeutig und es bedeute in unterschiedlichen Kontexten ganz verschiedenes. „Bewusstsein“ könne etwa die Fähigkeit zu integriertem Verhalten, aber auch bestimmte kognitive Fähigkeiten bezeichnen. Ich glaube, es lohnte sich – aber das wäre ein eigenes Thema – einmal genau zu prüfen, ob das eigentlich die richtige Herangehensweise ist. Ist nicht das meiste, was hier in guter Absicht und in guter analytischer Tradition

¹ Dieser Text ist die nur leicht redigierte Abschrift des Vortrages am 14. März in der Berlin-Brandenburgischen Akademie der Wissenschaften. Der mündliche Stil wurde belassen.

auseinander genommen wird, nicht doch besser als lediglich unterschiedliche Aspekte des Selben zu interpretieren? Wenn wir z.B. sagen „Der ist offenbar nicht bei vollem Bewusstsein“, weil er sich so merkwürdig verhält — er kann offenbar seine konativen und epistemischen Einstellungen nicht hinreichend integrieren, d.h nicht hinreichend kohärent machen, uns ist daher nicht klar, welche konativen und epistemischen Einstellungen wir ihm zuschreiben sollen — dann ist sollten wir das als einen Hinweis darauf nehmen, dass Bewusstsein ein holistisches und kein punktuell Phänomen ist. Und deswegen reden wir dann zu Recht so: Der ist nicht recht bei Bewusstsein, z.B., weil er hochgradig betrunken ist. Bewusstsein von etwas ist immer Teil eines komplexen Netzwerkes von Intentionalität. Wir werden gleich noch differenzieren, wie Intentionalität und Bewusstsein zusammen hängen. Aber jedenfalls müssen die sich im Verhalten und Urteilen äußernden Einstellungen in sich halbwegs stimmig sein, damit wir bestimmte Bewusstseinsformen überhaupt zuschreiben können. Bewusstsein jeder Form äußert sich in Ich-Kontrolle, in kohärenten Überzeugungen und Handlungen, wobei das immer *gradualistisch* zu verstehen ist. Die einzelnen Elemente eines epistemischen und konativen Systems fügen sich mehr oder weniger kohärent zu den anderen und ihre Individuierung ist isoliert gar nicht möglich, sondern lediglich im größeren Kontext anderer Elemente. (Zuvor ist eine sorgfältige Trennung vorgenommen worden zwischen dem Qualia-Problem und anderen Problemen des Bewusstseins. Auch hier bin ich skeptisch. Mir scheint eher, dass es sich dabei um verschiedene Aspekte von Bewusstsein handelt. Wenn wir Bewusstsein zuschreiben, schreiben wir ipso facto immer auch Qualia zu.)

Bewusstsein beinhaltet epistemische Zustände erster und zweiter Ordnung, also *Wissen dass* (und *Wissen wie*) und *Wissen, dass ich weiß, dass* (bzw. *wie*). Gibt es auch nicht-propositionales Wissen zweiter Ordnung? – ich denke ja. Das geht zunächst in die Richtung derjenigen, die sagen, Bewusstsein ist als eine Form von Wissen zu charakterisieren, allerdings mit einer wichtigen Abweichung von dieser „Schule“ der *epistemischen* Bestimmung von Bewusstsein: Die Zuschreibung von epistemischen Zuständen setzt nicht das Verfügen von Sprache im üblichen Sinne voraus. Das wäre ein Streitpunkt mit Donald Davidson, der genau die Gegenposition sehr scharf vertreten hat. Es spricht nichts dagegen auch Tieren epistemische Zustände zuzuschreiben - diese Zuschreibung ist die plausibelste Interpretation und Erklärung ihres Verhaltens. Welche Rolle dann die Sprache spielt und welche zusätzlichen epistemischen Differenzierungen erst möglich sind bei Lebewesen, die über Sprache verfügen, das ist wiederum ein eigenes schwieriges Thema. Hinsichtlich der zweiten Ordnung wird auch die empirische Basis, was Tiere angeht, etwas unsicher. Immerhin haben wir hier mindestens drei Mal, wenn ich mich recht entsinne, das Phänomen

des reflexiven Wissens bei Tieren diskutiert.² Von den empirischen Befunden, die dabei herangezogen werden, sind manche schon allgemein bekannt und vielleicht schon etwas abgenutzt, wie der berühmte Kreidefleck. Interessanter sind Phänomene des Täuschungsverhaltens und die Frage, ob Täuschungsabsichten überhaupt zugeschrieben werden können, ohne dass die Akteure den anderen Akteuren bestimmte Erwartungen und mentale Zustände zuschreiben. Wenn das der Fall ist, haben sie wohl auch reflexives Wissen, was ihre eigenen mentalen Zustände angeht. Dass sie nur fremde mentale Zustände hypothetisch annehmen und nicht selbstbezüglich eigene, ist hochgradig unplausibel.

Meine These lautet also: Es ist - jedenfalls als heuristische Maxime - interessanter, nicht mit einer Aufzählung von verschiedenen Verwendungsweisen des Ausdrucks Bewusstsein zu beginnen, sondern erst einmal zu prüfen, wie weit man mit der Gegenthese kommt: Dass es nämlich im Wesentlichen um ein- und dasselbe Phänomen geht, um unterschiedliche Indikatoren oder Aspekte dieses Phänomens, die in unterschiedlichen Kontexten eine Rolle spielen. Dies nur als Vorbemerkung.

II. Bewusstsein und Intentionalität.

Die Intentionalitäts-Debatte ist gerade im deutschsprachigen Raum vorbelastet durch einen Strang philosophischen Denkens, der vor allem auf Brentano zurückgeht, und ein ganz spezifisches Verständnis von Intentionalität begründet. Ich möchte hier von den Reminiszenzen an die verschiedenen klassischen Textkorpora absehen und systematisch argumentieren und zunächst feststellen, dass es zweifellos *nicht-intentionales Bewusstsein* gibt. Das scheint mir auf der Hand zu liegen, obwohl das einer weit verbreiteten Auffassung entgegen steht, die erstens besagt, Intentionalität hänge mit Intensionalität eng zusammen, die Berichte über intentionale Zustände seien *intensional*, d.h. es lassen sich die vorkommenden Prädikate und Individuennamen nicht extensionsgleich ersetzen; zweitens habe Bewusstsein immer diesen Charakter. Ich glaube, das ist offenkundig falsch. Es gibt eine Vielzahl von Bewusstseinsformen, die nicht intentional sind, wobei diejenigen, die intentional sind, dann wieder unterschieden werden müssten in solche mit propositionalem und solche ohne propositionalen Gehalt. Jedenfalls für das nicht-intentionale Bewusstsein besteht dieser Zusammenhang mit Intensionalität auf keinen Fall.

Ich plädiere für eine realistische Interpretation von Intentionalität, d. h. dort, wo es sich um intentionale Bewusstseinszustände handelt, ist der Inhalt eine Entität, die sich von den Wahrnehmungseindrücken unterscheidet. Wenn ich z.B. ein Auto wahrnehme, nehme ich ein

² Vgl. z.B. Julia Fischer: „Metakognition bei Tieren“ in: „Naturgeschichte der Freiheit“. Humanprojekt Band 1.

Auto wahr und habe nicht lediglich das Wahrnehmungsbild eines Autos. Jetzt gibt es natürlich das Problem der Halluzinationen, der Täuschungen und dergleichen, aber das widerlegt diese Sichtweise nicht. Eine realistische Interpretation unserer Erfahrungen und eine holistische Konzeption von Bewusstsein stützen sich wechselseitig. Die plausibelste Interpretation des Gesamts unserer intentionalen Zustände ist, dass bestimmte Gegenstände (oder bestimmte Tatsachen) objektiv — als deren Erfüllungsbedingungen existieren. Im Falle von propositionalen intentionalen Zuständen eben solche Tatsachen, die man in dass-Sätzen (dass etwas der Fall ist) formulieren kann. Wenn man dem lingualistischen Programm folgt, was ich nicht tue³, dann würde man in Satztypen, die bestimmte intentionale Bewusstseinszustände beschreiben, differenzieren und diesen Satztypen entsprechende Sachverhaltstypen zuzuordnen versuchen. Ich bin mir nicht sicher, ob sich das wirklich durchhalten lässt. Es ist sicher heuristisch sinnvoll, möglichst weit in diese Richtung zu gehen, aber ich glaube, es wird keine klaren grammatischen Regeln geben, die die Typeneinteilung charakterisieren. Wie steht es z.B. mit so etwas wie Hass? Ist Hass ein intentionaler Bewusstseinszustand oder nicht? — Ich würde sagen, unter Normalbedingungen ist Hass ein intentionaler Bewusstseinszustand, der gerichtet ist auf jemanden („Ich hasse jemanden“) und damit ist er intentional. Er kann sogar einen propositionalen Gehalt haben, „Ich hasse es, dass (oder wenn) XY passiert“.

In dem sehr schönen Aufsatz von Max Scheler zum Ressentiment schildert er in abstrakten Formulierungen eine Veränderung des Bewusstseins. Es beginnt – das sind natürlich nicht Schelers Begriffe – mit propositionaler Intentionalität, dann intentionaler aber nicht-propositionaler Intentionalität und mündet am Ende, so würde ich es interpretieren, in einem nicht-intentionalen Zustand, nämlich Ressentiment. Das Ressentiment ist dann am Ende nur noch eine Art Krankheit der Seele. Dem muss man natürlich nicht folgen, es lässt sich auch sagen, Ressentiment ist immer intentional, es richtet sich bspw. gegen eine bestimmte Minderheit. Aber man kann unter Ressentiment auch eine Art Vergiftungszustand der Seele verstehen — und das scheint mir der Befund dieses Aufsatzes von Scheler zu sein — der sich gegen niemanden mehr richtet. So ergibt sich eine graduelle Veränderung von ursprünglich propositionalem intentionalem Gehalt zu einem intentionalen nicht-propositionalen und schließlich zu einem nicht-intentionalen Bewusstseinszustand, einer Emotion ohne intentionalen Gehalt.

Jetzt möchte ich noch ausbuchstabieren, was ich zuvor „pragmatischen Holismus“ genannt habe. Ich taste mich gewissermaßen vor und beginne dort, wo der Konsens wahrscheinlich noch am größten ist, nämlich der Zuschreibungspraxis propositionaler Einstellungen. Es gibt

³ Vgl. JNR: „Die Grenzen der Sprache“ in *Von der Logik zur Sprache*, hgg von R. Bubner und G. Hindrichs, Stuttgart 2006.

ein sehr erfolgreiches reduktionistisches Programm der praktischen Philosophie, das sich folgendermaßen charakterisieren lässt: Man kann das Gesamt der praktisch relevanten Bewusstseinszustände oder mentalen Zustände auf zwei spezifische Typen propositionaler Einstellungen reduzieren oder zurückführen, nämlich *belief* auf der einen und *desire* auf der anderen Seite: *belief* — ich meine, glaube, vermute, erwarte usw., dass etwas der Fall ist — und *desire* — ich wünsche, präferiere, ich hoffe usw. Wobei „hoffen, dass etwas der Fall ist“ eine merkwürdige Ambivalenz hat, denn „hoffen“ hat einen epistemisch-konativen Doppelcharakter. „Hoffen“ heißt epistemisch „Ich habe eine Erwartung höher als..., dass es eintritt“, aber zugleich „Ich habe einen Wunsch, dass es eintritt“. Wenn diese Typologie durchginge, dann könnten wir die Praxis von Individuen und die Interaktionspraxis von Gruppen, Gemeinschaften, Sprachgemeinschaften, Kulturgemeinschaften usw. rekonstruieren über diese Zuschreibungspraxis zweier Typen propositionaler Einstellungen — *belief* und *desire*. Attraktiv an diesem Modell ist, dass es die hochentwickelte rationale Entscheidungstheorie und eventuell auch die Spieltheorie mit ihren Ablegern, *collective choice*-Theorie, in die Analyse mit einbeziehen kann. Der Holismus ist interessanterweise zumindest in den interessanteren Modellen der Entscheidungstheorie eingebaut. Das bekannteste ist das Modell von Richard Jeffrey in „Logic of Decision“ aus der Mitte der 60er Jahre, in der sehr schön schrittweise gezeigt wird, dass die Zuschreibung von *belief* und *desire* (wobei er interessanterweise nicht von *desire* sondern von *desireability* spricht, was eine normative Komponente hineinbringt, und er begründet es auch so, dass man es nur normativ interpretieren kann) nur als simultane Zuschreibung gelingt.

Die ursprüngliche Idee, die aus dem Positivismus hervorgegangen ist und die ganz wesentlich wurde für die Entstehung der Ökonomie als Wissenschaft um die Wende vom 19. zum 20. Jahrhundert, war ja, auf alle metaphysischen Begriffe zu verzichten und zu fragen: Was hat man als Datum? — Als Datum hat man eine bestimmte Entscheidung oder genauer gesagt ein bestimmtes Verhalten angesichts gegebener Alternativen. Das ist das Datum und die empirische Basis. Nun schreitet man weiter, bleibt aber im empiristischen Rahmen: Alles, was jetzt an Interpretation kommt, darf sich nicht entfernen von diesem Datum. Das ist das *revealed preference*-Konzept der Entscheidungstheorie: Wir entfernen uns nicht von dem Datum. Doch dann wird es kompliziert, denn es gibt natürlich eine Vielzahl von *belief/desire*-Paaren, die dieses Einzelne, den *token*, dass ich mich nämlich so und nicht anders entschieden habe, rational machen würden. D.h. es ist unterbestimmt, ich weiß nicht, welche *beliefs* oder welchen Komplex von *beliefs* und *desires* ich zuordnen soll. Aufgrund dieses Datums ist das unterbestimmt. Hier kommt der Holismus ins Spiel: Es ist nicht die einzelne Entscheidung, sondern eine Folge von Entscheidungen. Da diese Folgen von Entscheidungen nicht gegeben sind, sind sie kontrafaktisch („Wie würde die Person wohl, wenn sie...“). Wir haben so eine ganze Folge von Daten in diesem Sinne (man könnte noch

viel Genaueres hierzu sagen), so dass ich jetzt zwei reellwertige, bis auf positiv-lineare Transformationen eindeutige Funktionen dieser Person zuschreibe, nämlich eine *belief* oder eine *desire* oder *utility function*, für die gilt, dass die Person sich genau dann so verhält, wie sie sich als empirisches Datum verhält, wenn sie diese beiden Funktionen, genauer gesagt: den Erwartungswert der *utility function*, maximiert. Das ist die These, die sich über einige Kohärenzbedingungen axiomatisieren lässt. Die Kohärenzbedingungen der Präferenzen einer Person sind im Wesentlichen vier:

1. Reflexivität. Diese Bedingung ist trivial, dass ich nämlich dieselbe Alternative mindestens so gut finde wie diese Alternative.
2. Transitivität. Wenn ich A gegenüber B und B gegenüber C vorziehe, muss ich auch A C gegenüber vorziehen.
3. Monotonie. Wenn ich wahrscheinliche Alternativen - eine Lotterie: eine Wahrscheinlichkeitsverteilung für zwei mögliche Ergebnisse und eine zweite Wahrscheinlichkeitsverteilung über diese beiden gleichen möglichen Ergebnisse - habe, dann ziehe ich diejenige Lotterie oder diejenige Wahrscheinlichkeitsverteilung vor, die das von mir Gewünschtere wahrscheinlicher macht.
4. Die Stetigkeitsbedingung besagt, wenn ich drei Alternativen und eine (strikte) Präferenz für A gegenüber B und B gegenüber C habe, dann gibt es genau eine Wahrscheinlichkeitsverteilung über A und C, bei der ich gegenüber dieser Wahrscheinlichkeitsverteilung und B, dem Mittleren, indifferent bin.

Das sind die vier Bedingungen. Es gibt verschiedene, z.T. kompliziertere Axiomatisierungen, aber das hier Vorgestellte ist der Kern. Merkwürdigerweise ist in einem Gutteil der ökonomischen Literatur dieser kohärentistische Charakter der modernen Nutzentheorie vergessen. Es wird immer konsequentialistisch interpretiert – die Leute wollen irgendetwas, streben nach irgendetwas und wählen ihre Handlungen als geeignete Mittel, das Erstrebte als kausale oder probabilistische Folge eigener Handlungen zu erreichen –, der theoretische Kern der modernen Ökonomie jedoch ist kohärentistisch, und zwar in einem sehr weit gehenden Sinne. Erstens deswegen, weil diese Postulate alle Kohärenzpostulate sind. Sie sagen gar nichts darüber aus, was mir inhaltlich wünschenswert erscheint, etwa ob ich meinen Eigennutzen maximiere (es wird immer noch die Gespensterdebatte in der Ökonomie geführt, ob man Eigennutzen maximierende Individuen zu Grunde legen kann oder nicht, was aber nicht die Annahme der modernen, sondern der alten Ökonomie des 19. Jahrhunderts ist). Spätestens seit von Neuman / Morgenstern ist der theoretische Kern der modernen Ökonomie strikt kohärentistisch. Zweitens ist er in dem Sinne kohärentistisch, dass er eine offene, beliebig viele Alternativen umfassende Alternativenmenge annimmt,

gegenüber der sich die Person kohärent verhält. Also die rationale Person verhält sich kohärent gegenüber einer beliebig großen Menge von Alternativen, wobei unter diesen Alternativen auch solche sind, die Wahrscheinlichkeitsmischungen zwischen festen Alternativen darstellen. Entgegen dem manchmal in modernen ökonomische Büchern Vertretenen ist der rationale Akteur im Sinne der modernen Ökonomie durch eine Zuschreibungspraxis zweier propositionaler Einstellungen charakterisiert – *desire* und *belief*. Diese Zuschreibungspraxis beruht auf Bedingungen der Kohärenz, der Kohärenz der Praxis.

Das ist, würde ich sagen, im analytischen Feld der praktischen Philosophie die Standardherangehensweise. Es muss möglich sein, Rationalität in dieser Weise zu modellieren, was allerdings zu der Merkwürdigkeit führt, dass der ganze Kosmos unterschiedlicher mentaler Zustände, zumal Bewusstseinszustände intentionaler und nicht-intentionaler Art, sich reduzieren lassen müsste auf diese zwei Grundtypen propositionaler Einstellungen. Ich will eine sicherlich brutale Antwort geben: In einem bestimmten Sinne ist das mit Sicherheit falsch. Die Vielfalt von mentalen Zuständen, die relevant sind für unsere Praxis, für unsere Interaktionen, lässt sich nicht reduzieren auf diese zwei propositionalen Einstellungen. Dennoch kann es sinnvoll sein, dass im theoretischen Modell eine solche radikale Reduktion vorgenommen wird und man kann dieses theoretische Modell trotz der Nichtreduzierbarkeit dadurch retten, dass diese beiden Funktionen eben nicht konkrete Bewusstseinszustände der Person repräsentieren, sondern lediglich die Kohärenz, die diese Person in der Vielfalt von intentionalen und nicht-intentionalen Zuständen herstellt. Anders formuliert: Wenn diese ganze Vielfalt kohärent ist, lässt sie sich so repräsentieren, als ob die Person diese zwei propositionalen Einstellungen hätte. D.h. es sind Kohärenzforderungen, die falls erfüllt, diese Darstellungsweise erlauben.

Zur Verteidigung dieses theoretischen Kerns der modernen Ökonomie kann man in der Tat anführen, dass die vier genannten Postulate so harmlos sind, dass sie eigentlich erfüllbar sein sollten für beliebige Spektren von intentionalen Zuständen, die relevant sind für unsere Praxis. In dieser Sichtweise ist dies keine Interpretation intentionalen Bewusstseins insgesamt, sondern das sind Kohärenzbedingungen, also Postulate der Kohärenz, die dann dazu führen, dass diese Darstellungsform möglich ist. Aber man darf diese Darstellungsform nicht reduktionistisch missverstehen.

Welche Rolle spielt dabei die Sprache? – Ich glaube, es gibt zwei große Stränge in der Diskussion um diese Frage. Der eine Strang ist *lingualistisch*, der andere ist nicht-lingualistisch. Der eine behauptet, dass alles, was an mentalen Zuständen relevant ist, um menschliches Verhalten zu verstehen, sich auch sprachlich darstellen lassen muss. Man muss es beschreiben können und eine bestimmte Teilmenge von Bewusstseinszuständen lässt sich dann mit intensionalen Sätzen beschreiben. Alles was nicht darstellbar ist, ist auch für die Interpretation menschlichen Verhaltens nicht verwendbar. Man könnte sagen: Nichts

trivialer als das. Wie denn sonst? Bleibt uns denn noch etwas außer Sprache? Die Sache ist allerdings vertrackter, denn Sprache hängt ja davon ab, dass es geteilte Bedeutungen gibt, d.h. dass die Äußerungen eine Bedeutung haben in einer Sprachgemeinschaft. Die behavioristische Form des Lingualismus sagt, wir verstehen Bedeutung als nichts anderes als die faktische Konformität der Sprachteilnehmer oder der Sprecher dieser Gemeinschaft mit bestimmten Verhaltensregeln (In bestimmten Situation äußert er dies und jenes. Wenn der eine das äußert, äußert der andere das usw.). Viele Behavioristen meinen, das hätte Wittgenstein behauptet. Ich bin mir ziemlich sicher, dass das nicht der Fall ist, obwohl man sicherlich bestimmte Passagen so lesen kann. Die mir viel plausibler erscheinende Analysemethode ist die, dass sich im Sprachverhalten ein spezifischer Typ von Bewusstseinszuständen, nämlich derjenige intentionaler, propositionaler Bewusstseinszustände, die im Deutschen „Absichten“ heißen (ich habe den Intentionalitätsbegriff viel weiter als den Absichtenbegriff verwendet), verstehen lässt. Oder anders formuliert: Dass bedeutungsvolles Äußerungsverhalten, d.h. erfolgreiche Verständigung, die angesprochene Kohärenz der Beteiligten bezüglich ihrer intentionalen Zustände voraussetzt. Die bekannte Form, in der dieses Programm seit Ende der 50er Jahre diskutiert wird, ist die der Grice'schen⁴ Semantik (neben David Lewis⁵, war Jonathan Bennett einer der drei großen Vertreter der intentionalistischen Semantik). Ich glaube, dass das im Kern der richtige Ansatz ist, natürlich mit der zu erwartenden Modifikation, dass das, was zu Grunde gelegt wird, nicht wie bei Grice das Einwirken auf den Anderen ist, sondern das Bereitstellen von Gründen, etwas zu glauben oder etwas zu tun⁶. Wesentlich ist, dass hier das Verhältnis gegenüber dem behavioristischen Programm gewissermaßen umgekehrt wird. Es ist die Zuschreibung von bestimmten Bewusstseinszuständen, und zwar ganz speziellen (nämlich Absichten oder Gründe geben), die bedeutungsvolles Äußerungsverhalten und sprachliche Verständigung erst möglich machen. Um plausibel zu sein, muss dieser Ansatz deswegen bei bedeutungsvollem Äußerungsverhalten beginnen in Fällen, in denen es keine konventionelle Bedeutung gibt. Wenn ich sage „Draußen scheint die Sonne“, dann hat der Satz eine konventionelle Bedeutung (Unter normalen Bedingungen kann niemand sagen „Draußen scheint die Sonne“ und meinen „Draußen regnet es“, aber es gibt Ausnahmen, wie etwa der ironische Sprachgebrauch, der erlaubt, dass man mit einer Äußerung das Gegenteil des konventionell Gesagten meint und die Äußerung dann auch so verstanden wird). Es ist in den meisten Fällen konventionell weitgehend festgelegt, welche

⁴ Vgl. H.P.Grice: „Meaning“, *Philosophical Review* 66 (1957), S. 377-88 und ders. „Utterer's Meaning and Intentions“, *Philosophical Review* 78 (1969), S. 147-177 und zur Diskussion dieses Ansatzes: R. E. Grandy und R. Warner (Hrsg.), *Philosophical Grounds of Rationality. Intentions, Categories, Ends*. Oxford 1989.

⁵ David Lewis, *Convention*, Cambridge/Mass. 1969

⁶ JNR: Strukturelle Rationalität. Stuttgart 2001. Kap V und „Grice und Gründe“, erscheint in der Festschrift für Georg Meggle 2009.

Bedeutung, unsere Äußerungen haben, und Grice und seine Schule müssen daher ungewöhnliche Situationen erfinden, in denen es keine konventionelle Bedeutung bestimmter Zeichen gibt, die gebrauchten Zeichen aber dann dadurch eine Bedeutung bekommen, dass zwischen Sprecher und Adressat aufgrund des Zeichengebrauchs eine gemeinsame wechselseitige Zuschreibung von Intentionalität erfolgt.

Hier ist eine Bemerkung zum Problem Linguismus erforderlich: Es ist sehr plausibel anzunehmen, dass wir große Teile unserer intentionalen Zustände nicht beschreiben können, weil das uns zur Verfügung stehende Vokabular zu dürftig ist. Da könnten Neurowissenschaftler und Psychologen viele Beispiele anführen. Z.B. ist die diskriminatorische Fähigkeit bezüglich Farbnuancen um ein Vielfaches höher als das, was sprachlich an Farbschattierungen angeboten wird. Erinnern Sie sich noch an die beliebte Legende, die Inuit hätten fünfzig unterschiedliche sprachliche Ausdrücke für „Weiß“ und die Hopi-Indianer hätten aufgrund Ihrer Sprache eine andere Physik entwickelt, als die die uns vertraut ist?⁷ Das ist deswegen Unfug, weil die Alltagspraxis – z.B. im Umgang mit einer meist schnee- und eisbedeckten Umwelt - von den sprachlichen Möglichkeiten nur in engen Grenzen festgelegt ist. Wenn Menschen einer Gruppe etwas unterscheiden können, äußert sich das unter geeigneten Bedingungen auch in Ihrer lebensweltlichen Praxis, unabhängig davon, ob ihre Sprache über Termini dieser Diskrimination verfügt. Deswegen ist es – entgegen der linguistischen Doktrin – möglich, dass Äußerungen Bedeutungen haben, dass Verständigungen stattfinden, dass Kommunikation erfolgreich ist, dass wechselseitige Zuschreibung von Intentionen stattfindet, ohne dass sich die Beteiligten einer Sprache im üblichen Sinne bedienen.

Sprachlose Kommunikation setzt ebenso wie sprachliche gemeinsames Wissen voraus. Ja, genau besehen, sogar in einem deutlich höheren Maße. Der Begriff „Kommunikation“ impliziert das Gelingen, sonst handelte es sich lediglich um einen Kommunikationsversuch. Die sprachlichen Regeln entlasten die Kunst der angemessenen und für Kommunikation konstitutiven Zuschreibung von Intentionen. Wer ein Versprechen unter Verwendung der üblichen sprachlichen Mittel gibt, offenbart damit dem Adressaten seine Intentionen und Erwartungen. Natürlich ist es möglich, dass jemand andere Intentionen hat, als die, die für den betreffenden Sprechakt konstitutiv sind, dann handelt es sich um einen problematischen Fall, wenn dies dem Adressaten verborgen bleibt (Austin spricht da von infelicity⁸), der zu einem Fehlschlag wird (Austin nennt das fallacy), den Vollzug des Sprechaktes selbst also

⁷ Vgl. Benjamin Lee Whorf: Language, Thought and Reality, London 1956

⁸ John L. Austin: How to do things with words, Oxford 1962

unmöglich macht, wenn sich das dem Adressaten offenbart. Die, die jeweiligen Sprechakte begleitenden Intentionen sind konventionell in dem Sinne, dass wir sie als Mitglieder der Sprachgemeinschaft haben (sollten), wenn wir die Sprechakte vollziehen, diese uns wechselseitig zuschreiben und bezüglich dieser über ein gemeinsames Wissen verfügen, das heißt, der Sprecher weiß, dass der Adressat annimmt, dass er diese Intentionen hat und der Adressat weiß, dass das der Sprecher weiß etc. – die Interaktionen sind lediglich pragmatisch begrenzt. Auch für sprachlose Kommunikationen spielt gemeinsames Wissen in diesem Sinne eine konstitutive Rolle, nur, dass die grammatischen Regeln seinen Inhalt nicht festlegen, dieser also nicht-konventionell ist. Sprachlose Kommunikation kann es daher nur vor dem Hintergrund geteilter Lebensformen und der für diese ausschlaggebenden Intentionen geben, unabhängig davon ob sich diese einer mehr oder weniger universalistisch verstandenen *conditio humana* verdanken oder den kulturellen, aber von den Kommunikationsbeteiligten geteilten Besonderheiten.

III Intentionalität und praktische Freiheit

Ich führe nun die Fäden im Verhältnis von Intentionalität und praktischer Freiheit zusammen. Handlungen haben eine äußere, öffentliche, man kann auch sagen physikalische Dimension und eine innere in Gestalt einer je spezifischen, den Handlungstyp individuierenden Intentionalität. Obwohl ziemlich künstlich scheint es mir sinnvoll, dabei drei Formen dieser spezifischen Intentionalität zu unterscheiden. Die erste Form sind motivierende Absichten. Das ist eine Form propositionaler Intentionalität oder propositionaler intentionaler Bewusstseinszustände. Wenn ich also gefragt werde: „Warum hast du denn vorhin deinen Nachtisch nicht aufgegessen?“ und ich sage „Weil ich rechtzeitig mit dem Vortrag beginnen wollte“, dann ist das eine Erläuterung im Sinne motivierender Absicht. Es war mein Motiv, dies zu tun. Ich glaube, es gibt keine Handlung, die nicht wenigstens rudimentär von motivierenden Absichten geprägt ist. Dies wird umstritten sein.

Zweitens gibt es keine Handlung, die nicht von vorausgehenden Absichten geleitet ist. Vorausgehende Absichten unterscheiden sich von motivierenden Absichten dadurch, dass ihre Erfüllungsbedingungen die Handlungen selbst sind. Wenn ich gehandelt habe, habe ich die vorausgehende Absicht erfüllt. An dieser Stelle gibt es das besondere Phänomen der Entscheidung. Was ist die Entscheidung in diesem Zusammenhang? Mir scheint, die Entscheidung ist eine besondere Form vorausgehender Absichten, die man etwa

folgendermaßen charakterisieren kann: Erstens: Sie wird erfüllt durch die Handlung. Wenn ich mich entschieden habe um 4 Uhr den Raum zu verlassen, dann erfülle ich diese Absicht dadurch, dass ich diesen Raum um 4 Uhr verlasse, ganz egal, welche Motive ich dafür habe. Das Motiv ist vielleicht, die S-Bahn zu erreichen, und es erfüllt sich dadurch, dass ich die S-Bahn erreiche. Erfüllt wird die vorausgehende Absicht durch die Handlung des diesen Raum-Verlassens. Aber das besondere einer Entscheidung ist zweitens, dass sie einen Deliberationsprozess stoppt – das wäre jedenfalls mein Vorschlag –, d.h. auf Dispens setzt, etwa in dem Sinne: Ich (bzw. bei kollektiven Entscheidungen „wir“) werde(n) jetzt keine weiteren Abwägungen mehr vornehmen. Wenn nichts dazwischen kommt, wenn ich nicht noch mal Grund habe, neu anzusetzen, dann ist das jetzt meine endgültige vorausgehende Absicht, die durch meine Handlung erfüllt werden wird. Wenn das stimmt, handelt es sich um ein sehr interessantes Phänomen, denn dann wäre dieses kognitive Element der Abwägung von Gründen in den Handlungs- und Entscheidungsbegriff selbst eingebaut, die stoizistische Auffassung *prohairesis krisis estin*, wäre keine exotische philosophische These, sondern unaufgebarbarer Teil unserer Praxis. Zu dieser Praxis gehörte das für Entscheidungen charakteristische Phänomen, Stellung zu nehmen und das heißt die Absicht zu haben, weitere Abwägungen zu unterlassen. „Schluss“ zu sagen, wertend Stellung zu nehmen, sich zu entscheiden wären gewissermaßen ein einziger Akt und die dann folgende Handlung nur noch das Resultat. Jede Handlung ist eine Stellungnahme, jede Präferenz ist eine Stellungnahme, d.h. das Ergebnis der Abwägung von Gründen und sei sie noch so rudimentär. Das ist nicht so abwegig, wie es zunächst zu sein scheint.

Nun kommt die dritte Form von Intentionalität, die vor allem in den 60er Jahren ganze Bibliotheken mit Beispielen und Gegenbeispielen gefüllt hat, ins Spiel, nämlich dass das, was meine vorausgehenden Absichten erfüllt, selbst wieder von Intentionalität begleitet sein muss. Ein Beispiel (vom Typ sind diese Beispiele immer gleich): Jemand will seinen Onkel ermorden, deswegen ist er aufgeregt und fährt zu schnell. Er überfährt deswegen einen Passanten, der sein Onkel ist. Jetzt kann man sagen: Er hat eine vorausgehende Absicht. Diese war kausal relevant für das, was passiert, nämlich, dass der Onkel zu Tode kommt. Doch irgendetwas stimmt hier nicht, deswegen muss noch das dritte Element hinzu, nämlich diese, sagen wir, *verhaltensbegleitende* Intentionalität. Nur wo diese drei Formen von Intentionalität eine Rolle spielen, können wir von Handlungen eines Typs sprechen, der durch die jeweiligen charakteristischen Bewusstseinszustände bestimmt ist.

Damit ergibt sich eine Kritik der *desire/belief*-Theorie, die man in der radikalsten Form folgendermaßen präsentieren kann: In letzter Instanz ist für die Frage, ob eine Handlung

rational ist oder nicht, ausschlaggebend, ob meine normative Überzeugung, dass diese Handlung die richtige ist, wohlgegründet ist oder nicht. Die letzte Instanz sind also nicht Wünsche, die sich in Absichten umsetzen und Absichten, die erfüllt werden durch etwas, sondern wir interpretieren jetzt diese Absichten als Ausdruck einer normativen Stellungnahme. Das erklärt, warum es keine Handlung, ohne eine wenigstens rudimentäre Abwägung von Gründen geben kann. Das ist eine sehr radikale Kritik der *desire/belief*-Theorie, die man auch vorsichtiger formulieren kann: Es gibt Handlungen, für die es erforderlich ist, eine solche Interpretation vorzunehmen. Die universelle Anwendung der *desire/belief*-Theorie scheitert also an diesem Phänomen bei bestimmten Handlungen. Das lässt sich tatsächlich beweisen.⁹ Die weitergehende These als Allsatz lässt sich schwer beweisen, aber ich habe noch kein überzeugendes Gegenargument gehört. Der normative, unaufgeregte, nämlich nicht ontologisch angereicherte Realismus, für den ich plädiere, besagt, dass es uns hier bei der Abwägung von Gründen letztlich darum geht festzustellen, welche normativen Sachverhalte bestehen, welche Handlungen die besseren Argumente für sich haben. Deswegen reichere ich die natürliche Welt nicht notwendigerweise mit moralischen Eigenschaften an, vielmehr ist das in meinen Augen eine adäquate Beschreibung dessen, was wir alle in unserer alltäglichen Verständigungspraxis tun.

Ich möchte den zentralen Punkt folgendermaßen zuspitzen: Angenommen es stimmt, dass ich bei jedem Verhalten, das Handlungscharakter hat, auf Fragen des Typs „Warum hast du das getan?“ eine Antwort geben kann, die keine Ursachen nennt, sondern Gründe angibt. Ich kann eine Antwort auf eine Warum-Frage spezifischen Typs geben und der Frager wäre äußerst enttäuscht, wenn ich bei meiner frühkindlichen traumatischen Erfahrung beginnen würde, um zu erklären, warum ich mich so und nicht anders verhalten habe, als eine Kausalgeschichte erzähle. Er will wissen, was für einen Grund ich hatte, so und nicht anders zu handeln. Wir haben hier etwas Merkwürdiges: Wenn ich Gründe angeben kann, dann nehme ich (um den Terminus zu verwenden) reflexiv Stellung. Ich sage, was mich bewogen hat, es so und nicht anders zu tun, warum ich überzeugt bin, es war das Richtige – normativ. Was nun mit der These, ein Gutteil unserer Handlungen habe nicht diesen Charakter, sondern geschehe irgendwie unbewusst oder – wie zuvor die These war – es gebe viele Emotionen, die keinen Bewusstseinszuständen entsprechen und diese könnten sich in einem konkreten Abwehrverhalten der betreffenden Person äußern? Nehmen wir einmal an, ein solches Forschungsprogramm ginge auf, man könnte also bestimmte Verhaltensweisen so erklären. Dann würden wir der Person Emotionen zuschreiben, die der Person nicht bewusst sind, die dieses Verhalten aber erklären. Jetzt fragen wir die Person, warum hast du das so

⁹ Vgl JNR: Strukturelle Rationalität. Stuttgart 2001, Kap 1

und nicht anders gemacht, und die Person gibt Gründe, die natürlich nicht auf diese Emotionen Bezug nehmen könne, da diese der Person ja nicht bewusst sind, wie angenommen wurde. Nun können wir versuchen die Sache mit der Behauptung zu retten, die Person irre sich. Sie wisse eben nicht, was sie dazu gebracht hat, sich so und nicht anders zu verhalten. Sie gibt zwar Gründe an, aber eben nur *ex post*, diese hatten keine handlungsleitende Rolle. Dies wäre eine Variante der *error-theory*, wir sind alle in einer Irrtumswelt befangen, wir geben Gründe an, ohne dass diese Gründe eine Rolle spielen für das, was wir tun. Wer das glauben will, den kann man nicht widerlegen, aber sehr plausibel ist das nicht. Die Tatsache, dass wir – schärfer formuliert – für diejenigen Verhaltensbestandteile, die Handlungscharakter haben, immer Gründe angeben können und diese Gründe sich dann in jener komplexen Intentionalität strukturieren, die ich vorher skizziert habe, mit motivierenden, vorausgehenden und begleitenden Absichten – diese Tatsache spricht gegen die These, dass das Handeln in hohem Maße bestimmt sei durch Vorgänge, die nicht ins Bewusstsein treten.

Es ist mir völlig klar, dass die These der umfassenden Steuerung unseres Handelns durch Gründe zunächst einmal unplausibel erscheint. Deswegen versuche ich es etwas plausibler zu machen mit einem Beispiel, das ich bereits vor vielen Jahren in der „Kritik des Konsequentialismus“ (München/Wien 1993) verwendet habe, das mir aber nach wie vor illustrativ erscheint: Wenn Sie einen schweren Unfall mit einem Bein-, oder Armbruch hatten und Sie dann wieder anfangen, sich zu bewegen, dann vollziehen Sie eine Vielzahl von Handlungen, die der normale Akteur, der keinen solchen Unfall hatte, allesamt nicht vollzieht. Sie sagen etwa zu sich: Damit ich zu diesem Glas komme, sollte ich erst einmal meine Schulter zurücknehmen, weil ich weiß, so direkt komme ich nicht hin, dann sollte ich mich mit meinem Oberkörper zur Seite drehen, damit das Gelenk eine Drehung mit kleinerem Winkel machen kann. Schließlich müssen Sie eventuell alle Ihre Willenskraft zusammennehmen, damit die Muskeln die für diese Bewegung notwendige Arbeit verrichten. Um ein anderes Beispiel zu nehmen: Für jemanden der gerade Autofahren gelernt hat, ist Rechts-Abbiegen eine dichte Folge von einzelnen Handlungen. Er ist schweißgebadet und überlegt sich, wann muss ich jetzt den Blinker setzen, von der Bremse gehen, die Kupplung drücken und das alles kurz hintereinander oder gar gleichzeitig. Derjenige, der gut Auto fährt, reduziert – mit einem Begriff von Danto – die Anzahl nicht mehr zerlegbaren Basishandlungen dramatisch. Der nach 25 Jahren Fahrpraxis wirkliche Geübte will nur noch von A nach B, seine Handlung setzt diese Absicht um. Es ist nicht erforderlich, ja psychologisch kaum noch möglich, diese Handlung in eine Vielzahl von Teilhandlungen zu zerlegen und diese einzeln zu vollziehen. Würde er das tatsächlich tun, erschiene sein Fahrstil wieder wie der eines Anfängers. Es gibt da viele, teilweise weit komplexere Beispiele: Jemand fährt 30 Kilometer nachts, er ist während dessen nicht eingeschlafen und merkt dennoch plötzlich, ich bin ja schon 30

Kilometer weiter gefahren, als mir bewusst war. Ist er bewusstlos gefahren, ist er während dessen bewusstlos gewesen? Nein, die naheliegendere Interpretation ist, dass die intentionale Steuerung, die handlungsleitende Rolle von Gründen hier auf ein Minimum reduziert ist und daher das Fahrverhalten weitgehend automatisiert ist, auch ohne bewusste Kontrolle auskommt (ganz ungefährlich ist das nicht, da erfahrungsgemäß in solchen Phasen die Reaktionszeiten wesentlich herabgesetzt sind).

Das Bild der Handlung ändert sich also. Es ist nicht etwa so, dass Verhalten neben Handlung steht, sondern, solange ich bei Bewusstsein bin, ist alles handlungsstrukturiert, über Handlungen strukturiert. Aber die strukturierenden Elemente sind unterschiedlich groß und damit sind auch die Modi des Begründens andere. Der eine fragt z.B.: „Warum hast das gemacht?“ – Dann sagt der erfahrene Autofahrer: „Weil ich nach B wollte“. Das genügt. Der Ungeübte sagt: „Weil ich erstens nach B wollte und weil mir zweitens der Fahrlehrer gesagt hat, wenn ich das will, muss ich dies und jenes machen usw.“ Man könnte sagen, der *phronimos* handelt immer weniger, was nicht heißt, dass Teile seines Verhaltens keinen Handlungscharakter hätten. Sondern die kleinsten, nicht mehr in Einzelhandlungen zerlegbaren Teile seines Verhaltens, die Handlungscharakter haben, werden größer. Das ist die Idee. Damit sind wir schon nahe an der wesentlichen These. *Gründe heben überzeugungs- und handlungsleitende Intentionalität ins Bewusstsein (da gibt es in der Tat ex post-Phänomene) und strukturieren diese.* Gründe strukturieren die Praxis, einschließlich der Urteilspraxis. Ich überspringe hier die Thematik der *naturalistischen Unterbestimmtheit von Gründen*, zu der ich in unserer Runde schon vorgetragen habe¹⁰ und komme zur Frage, wie das, was wir gerne als „Autonomie“ oder „Selbstbestimmung“ bezeichnen, mit Gründen zusammenhängt.

Die These lautet: Autonomie konstituiert sich über Gründe, nicht über Wünsche zweiter Ordnung – das ist gegen die Diskussionen im Anschluss an Harry Frankfurt gerichtet. Ich bin ich (ich bin nicht ich lediglich auf der Metaebene der Volitionen zweiter Stufe), ich bin verantwortlich, für das was ich tue, sofern das, was ich tue (und übrigens auch das, was ich glaube, und größtenteils auch das, was ich empfinde, zumindest, wenn es um bestimmte Typen von *moral sentiments* geht), sich über Gründe konstituiert. D.h. ich kann Auskunft geben, warum ich jemanden hasse, warum ich von etwas überzeugt bin oder warum ich etwas getan habe (wenn es denn Handlungscharakter hat): Autonomie als praktische Freiheit, aber diese Freiheit ist wohlgermerkt konstituiert über die besondere Fähigkeit, die eigene Lebensform, seine Überzeugungen, seine Interaktionspraxis, bestimmte handlungsleitende Gefühle, von Gründen leiten zu lassen. In der gegenwärtigen Literatur ist es weitverbreitet, dass die Anhänger einer Autonomieinterpretation von Freiheit und die

¹⁰ Julian Nida-Rümelin: Freiheit als naturalistische Unterbestimmtheit von Gründen. In: Naturgeschichte der Freiheit, Humanprojekt Band 1.

Anhänger einer Interpretation von Freiheit, die die Existenz von alternativen Möglichkeiten voraussetzt, als zwei Positionen angesehen werden, die eine unvereinbare Interpretation von Freiheit zu Grunde legen. Mir scheint das unplausibel zu sein, denn die Konfusionen in der zeitgenössischen *free will*-Debatte hängen genau damit zusammen, dass nicht geklärt wird, was es eigentlich heißt alternative Möglichkeiten zu haben und bezüglich welcher Bedingungen diese bestehen sollen. Da gibt es etwa das Beispiel, das z.B. Christine Korsgaard sehr beschäftigt hat: Wenn jemand ein moralisch so anständiger Mensch ist, dass das Abwägen von Gründen nur dazu führen kann, dass er dieses tut, dann ist er zwar verantwortlich für das, was er tut, dann ist er autonom, weil es ja seine Wünsche, Einstellungen und moralischen Überzeugungen sind, die dieses moralische Verhalten prägen, aber er war nicht frei, etwas anderes zu tun, weil er als Person ja so ist, dass er nichts anderes tun konnte, als das, was er tat. Nun, das ist völlig richtig, nur wird nicht genau gesagt, unter welchen Bedingungen dies geschieht. Denn diese Bedingungen, die eingeführt werden und die die alternativen Möglichkeiten ausschließen, beziehen sich ja schon auf den Inhalt der Abwägung der Gründe, die das betreffende Individuum als eine anständige moralische Person vornimmt. Meine These der Existenz alternativen Möglichkeit ist hingegen, dass die Deliberation eine Rolle spielt für das, was ich tue, und nicht schon vor aller Deliberation immer schon festliegt, was ich tue. Das ist eine ganz andere These und wenn man will, kann man das rekonstruieren über die Spielräume, die bleiben, wenn man eine vollständige Beschreibung der Verhaltensbedingungen vornimmt in der Sprache und mit den Gesetzesannahmen der Naturwissenschaft, also in einem *naturalistischen Modus*. Das, das alles *science fiction* ist, muss ich hier nicht betonen.

IV Selbstbewusstsein und Freiheit“.

Während ich beim ersten Schritt — Gibt es eine Einheit des Bewusstseins oder müssen wir wirklich verschiedene Typen unterscheiden? — eher einen heuristischen Vorschlag gemacht habe und mir nicht sicher bin, ob er sich wirklich durchhalten lässt, und ich im Zwischenteil, der sich mit dem Verhältnis von Bewusstsein und Intentionalität (II) und dem von Intentionalität und praktischer Freiheit (III) auseinandersetzt, ziemlich zuversichtlich bin, dass er sich gut verteidigen lässt, bin ich hier im vierten Teil wieder unsicherer.

Zunächst einmal ist deutlich zu unterscheiden zwischen der Frage „Was ist Selbstbewusstsein?“ und der Frage „Was sind die Indikatoren von Selbstbewusstsein?“. Ich hatte ein Beispiel genannt: Der Spiegeltest oder das Täuschungsverhalten bei Tieren ist am einfachsten interpretiert als Ausdruck von Selbstbewusstsein. Aber was ist Selbstbewusstsein selbst? Nehmen wir „Selbstbewusstsein“ zunächst wörtlich: Es ist ein Bewusstsein meiner selbst. Ich habe ein Bewusstsein für meine Rolle in der sozialen,

kulturellen oder physikalischen Welt. Selbstbewusstsein setzt voraus, dass ich eine Art Selbstbeobachtung habe: Ich stelle fest, was ich tue, ich kann rückblickend sagen, was ich getan habe usw. In diesem Sinne gibt es keine Rationalität über Gründe ohne Selbstbewusstsein. Wesen können Wünsche haben ohne Selbstbewusstsein, aber keine Gründe. Denn Gründe setzen immer ein bestimmtes Maß an Reflexion, eine Stellungnahme voraus („Ich sollte das tun“). Die These wäre – wieder gegen Harry Frankfurt –, die Person ist identifiziert über Gründe, die sie hat für ihre Überzeugungen, ihr Handeln und ihre moralischen Gefühle. Das macht den Persönlichkeitskern aus. Die Person nimmt Stellung zu ihrer Rolle in der Welt, indem sie Gründe gibt und Gründe nimmt, Anderen begründet, warum sie es so und so getan hat, Andere kritisiert dafür, dass sie es falsch gemacht haben usw. Ein wechselseitiger Prozess also, bei dem die jeweiligen Rollen in der Welt betrachtet werden und man sich auseinandersetzt mit den Kriterien der richtigen oder der falschen Rolle.

Wie hängen nun Selbstbewusstsein und personale Identität miteinander zusammen? Das ist ein nächstes Großproblem. Man könnte sagen, in einem synchronen Sinn ist Selbstbewusstsein mit personaler Identität verknüpft, so wie ich es formuliert habe: Ich nehme Stellung zu meiner Rolle, ich sehe mich, ich nehme Stellung zu mir als Person, als Akteur, als Urteilender usw. Diachron betrachtet wird es schwierig, was einer der wenigen Streitpunkte ist, die ich mit meiner Schwester seit unserer Kindheit habe. Ich sagte damals zu ihr, das, was Du meinst, wäre nur verständlich, wenn die „Engelhypothese“ stimmt: Nur wenn wir vorher und nachher Engel (derselbe Engel davor und danach) wären, die ihrerseits keine Entwicklung mehr durchmachen, also zeitlos sind, kann man verstehen, was (numerische) Identität über die Zeit, über alle Veränderungen im diesseitigen Leben, heißen soll. Ansonsten verstehe ich es nicht. Alle möglichen scharfsinnigen Argumente lösen nämlich die Grundproblematik nicht, dass diese These unterbestimmt ist. Ohne die ganze Diskussion hier führen zu können, will ich wenigstens zeigen, dass sich kohärent vertreten lässt, dass Selbstbewusstsein und personale Identität synchron miteinander verknüpft sind, sich diachron aber in einem gewissem Umfang entkoppeln. Jetzt bin ich wieder Gradualist! Denn wenn jemand von heute auf morgen völlig andere Begründungen für sein Handeln und Tun hervorbringt – und das passiert öfters –, dann haben wir ein Problem, ihn als Person überhaupt ernst zu nehmen. Wir verstehen nicht mehr, mit wem wir es eigentlich zu tun haben. Ist er noch der Gleiche? Wenn das allzu oft vorkommt, dann gibt es ein Problem der Zuschreibung von propositionalen Einstellungen, die über Gründe strukturiert sind. Denn dass sie über Gründe strukturiert sind, macht ja – wenn diese Analyse stimmt – den Persönlichkeitskern aus.

Das ist ein gradualistisches Konzept und es passt wunderbar in die Rechtspraxis: Warum nimmt die strafrechtliche Verantwortung mit zunehmendem Alter zu? Da zieht man willkürlich

bestimmte Altersstufen ein, aber das geschieht ja nur, um sozusagen auf der sicheren Seite zu sein. Warum verbinden wir strafrechtliche Verantwortlichkeit nicht mit dem IQ, sondern mit dem Alter? – Mir scheint, diese Fähigkeit, sein Verhalten über Gründe zu strukturieren, nimmt in der Tat mit dem Alter zu. Da gibt es auch interessante psychologische Forschungen, die sogar die erfreuliche Botschaft haben, dass diese Fähigkeit bis ins hohe Alter im Laufe des Lebens sogar weiter zunimmt, es sei denn man wird dement. Auch wenn die Fähigkeit, parallel Informationen schnell zu verarbeiten, nachlässt, nimmt dennoch offenbar diese Fähigkeit, kohärent zu urteilen und zu handeln, zu. Dies alles steht dann in wunderbarer Übereinstimmung damit, dass man davon spricht, dass sich die Persönlichkeit entwickelt bzw. weiterentwickelt hat und dass wir sie dann in höherem Maße für das, was sie lebt, verantwortlich machen als zuvor. Das ist also ein gradualistischer Begriff von personaler Identität, der in seinem Kern normativ ist: Wir *verlangen* von Menschen, dass sie halbwegs kohärente Gründe haben für das, was sie tun und glauben. Wir *verlangen*, dass sie ein gewisses Maß an diachroner Identität in diesem Sinne aufweisen.

Nun zur Rolle des Selbstbewusstseins für Interaktionen. Diese ließe sich noch mal sehr schön im Übergang von Entscheidungs- zu Spieltheorie erläutern. Das Merkwürdige ist ja, dass die Entscheidungstheorie, wie gerade geschildert, abgeschlossen ist: vier Postulate und ein Kriterium, das der Erwartungswertmaximierung. Warum wird dieses Kriterium in der Spieltheorie nicht angewandt? Warum sind da auf einmal Gleichgewichtspunkte ausschlaggebend? – Die grobe Antwort ist folgende: Dass impliziter in der Spieltheorie die Fähigkeit vorausgesetzt wird, dem anderen Motive zu unterstellen, in der gleichen Weise, wie man selbst Motive hat und Gründe zuzuschreiben, die diese Motive (vorausgehenden Absichten usw.) leiten, und zu wissen, dass nicht nur man selbst, sondern auch die anderen Beteiligten diese Fähigkeiten haben und alle wissen, dass sie wissen, dass die anderen diese Fähigkeiten haben etc., so dass nur solche Empfehlungen gegeben werden können, die mit diesem gemeinsamen, iterierten Wissen kollektiv kompatibel sind. Deswegen dieser scharfe Bruch zwischen Entscheidungs- und Spieltheorie. Anders ausgedrückt: In der Zuschreibung von Selbstbewusstsein ist ein gewisses Element von Universalisierung eingebaut. Ich bestimme meine Rolle in der Welt und gegenüber anderen, setzt voraus, dass ich mich in den anderen soweit hinein versetzen kann, dass ich überhaupt rational interagieren kann. Dieses Sich-hinein-versetzen-können in die Lage des anderen und dessen Interessen setzt eine gewisse Distanzierung (eine gewisse – gradualistisch!) vom eigenen Standpunkt voraus.

Nun zur letzten und wohl besonders umstrittenen meiner Thesen: *Selbstbewusstsein impliziert Freiheit*, und zwar in dem starken Sinne der naturalistischen Unterbestimmtheit von Gründen und der Existenz von alternativen Möglichkeiten. Warum das so ist, kann die Philosophie im Grunde nicht für sich beantworten. Das ist keine begriffliche Wahrheit, die ich

hier behaupte, sondern eine letztlich empirische, wenn auch durch Experimente nur schwer belegbare. Ich kann mir nicht vorstellen, dass Selbstbewusstsein entstehen kann, wenn nicht zugleich die Erfahrung da ist, dass das, was ich will, relevant ist für das, was ich tue, oder anders: Dass meine Intentionen kausal auf die Welt einwirken, dass die Welt ohne diese Intentionen und die sie realisierenden Handlungen anders wäre, als sie ist. Ohne diese Erfahrung der Freiheit, kann ich mir nicht vorstellen, dass ein Wesen als Spezies und als Individuum betrachtet Selbstbewusstsein entwickeln kann.