

appeared in: Deutsche Zeitschrift für Philosophie 62(5), 2014, pp. 937-962.

## **Die Entwicklung der Autorität der Ersten Person**

**Johannes L. Brandl**

**Universität Salzburg**

### **Einleitung**

Dieser Aufsatz geht der Frage nach, welche Rolle die Entwicklung des reflexiven Selbstbewusstseins für die Autorität der Ersten Person spielt. Dabei verstehe ich unter einem reflexiven Selbstbewusstsein ein höherstufiges Bewusstsein in Form von kognitiven Akten, die sich auf das eigene Empfinden, Denken und Fühlen beziehen. Personen, die ein solches Meta-Bewusstsein entwickelt haben, verfügen damit über eine epistemische Qualität, die man die Autorität der Ersten Person nennt. Aber warum ist das so? Was verleiht dem reflexiven Selbstbewusstsein diese epistemische Relevanz?

Viele Abhandlungen über die Entwicklung des Selbstbewusstseins vermeiden es, auf diese erkenntnistheoretische Frage eine Antwort zu geben. Im Fokus dieser Arbeiten steht die Fähigkeit der Selbstbezugnahme, die zusammen mit anderen repräsentationalen Fähigkeiten grundlegend ist für die Entwicklung eines Selbstkonzepts (siehe z. B. Bermudez 1998). Auf diese Weise werden Fragen über die kognitive Entwicklung von Fragen bezüglich der Autorität der Ersten Person getrennt, indem man sich ganz auf die Fähigkeit der Selbstbezugnahme bzw. auf den Erwerb des Ich-Begriffs konzentriert. Kennzeichnend dafür ist die von Sydney Shoemaker angestossene Diskussion über die Immunität gegen einen Irrtum durch Fehlidentifikation (siehe Shoemaker, 1968). In dieser Debatte spielen epistemische Fragen zwar auch eine Rolle, aber sie betreffen nur den Erwerb eines Selbstkonzepts. Die Frage der Autorität der Ersten Person steht hier nicht zur Diskussion.

Dass man die Frage der Autorität der Ersten Person von der Frage der Entwicklung des Selbstbewusstseins abkoppeln kann, ist jedoch keineswegs selbstverständlich. Traditionsgemäß war die philosophische Beschäftigung mit dem Thema Selbstbewusstsein stets von einem erkenntnistheoretischen Interesse geleitet. Sowohl Rationalisten als auch Empiristen gehen davon aus, dass unser Selbstkonzept mehr ist als eine Summe von 'Ich'-Gedanken, deren Legitimität gesondert zu erörtern wäre. Selbstbewusstsein ist traditionell ein 'Bewusstsein vom Bewusstsein', dessen Besonderheit darin liegt, dass es die Quelle eines

unmittelbaren Selbstwissens ist. Unter dieser Voraussetzung stellt sich die Frage der epistemischen Relevanz des Selbstbewusstseins gar nicht, und es gibt auch keinen Raum für einen Zweifel an der Autorität der Ersten Person, denn an der Existenz eines reflektiven Selbstbewusstseins zu zweifeln, wäre geradezu absurd.

Die Autorität der Ersten Person ist jedoch nicht über jeden Zweifel erhaben. Wenn es daher, wie ich argumentieren werde, einen direkten Konnex zwischen der Entwicklung des Selbstbewusstseins und dieser erkenntnistheoretischen Problematik gibt, muss dieser Zusammenhang auf andere Weise erklärt werden. Was könnte jedoch an die Stelle der traditionellen Theorien des Selbstwissens treten, die die Autorität der Ersten Person durch einen epistemisch privilegierten Zugang zum eigenen Bewusstsein erklären? Seit Wittgenstein kennen wir eine mögliche Alternative: die Autorität der Ersten Person kann als ein *sprachliches* Phänomen begriffen werden. Sie rührt daher, dass wir lernen, unsere Empfindungen und Gedanken in Form von Selbstbekundungen (*avowals*) zum Ausdruck zu bringen (Wittgenstein 1953). Diese Erklärung hat meines Erachtens drei wesentliche Vorzüge: Zum einen kann diese Erklärung das Modell des privilegierten Zugangs zum eigenen Bewusstsein ersetzen, um unsere Intuitionen bezüglich der Autorität der Ersten Person verständlich zu machen. Zum zweiten läßt sich damit argumentieren, dass expressive Fähigkeiten sowohl für die Autorität der Ersten Person als auch für die Entwicklung des reflexiven Selbstbewusstseins grundlegend sind. Und drittens macht die beschränkte Glaubwürdigkeit expressiver Äußerungen deutlich, warum es durchaus berechtigte Zweifel an der Autorität der Ersten Person geben kann.

Die Auffassung, dass es eine Autorität der Ersten Person auch ohne privilegierten Zugang geben kann, läßt sich auf verschiedene Weise entwickeln.<sup>1</sup> Der expressivistische Ansatz von Wittgenstein scheint mir nach wie vor der vielversprechendste zu sein. Dies belegen eine Reihe von Arbeiten, die diesen Ansatz zuletzt aufgegriffen und systematisch weiter entwickelt haben, (siehe Falvey 2000, Bar On und Long 2001, Finkelstein 2003, Bar On 2004, 2010 und Green 2007). Meine Überlegungen stützen sich in vielen Punkten auf diese Arbeiten, weichen aber davon ab, wenn es darauf ankommt, gewissen Schwierigkeiten auszuweichen, die sich auftun, wenn man die Autorität der Ersten Person durch unsere expressiven Kompetenzen zu erklären versucht.

---

<sup>1</sup> Zu nennen wäre hier Rortys Versuch, das Cartesische Konzept der Infallibilität durch ein konventionalistisches Konzept der Inkorrigität zu ersetzen (Rorty 1970), sowie Andre Gallois' Versuch, die Autorität der Ersten Person durch die These von der Transparenz des Geistes zu erklären (Gallois 1996). Beide Ansätze übersehen meines Erachtens die wesentliche Rolle, die expressivistische Selbstbekundungen in diesem Zusammenhang spielen.

Abschnitt 1 erläutert nochmals die Ausgangslage und formuliert die zentralen Fragen, die eine Theorie der Autorität der Ersten Person zu beantworten hat. Abschnitt 2 stellt ein Konzept der Autorität vor, das bereits die Grundidee des von mir bevorzugten Deskriptiven Expressivismus in sich trägt. Diese Idee wird in Abschnitt 3 genauer expliziert und mit einem Vorbehalt versehen, solange die Frage falscher Selbstbekundungen noch ungelöst ist. Dieses Problem einer expressivistischen Analyse, auf das unter anderem David Rosenthal hingewiesen hat, ist Thema von Abschnitt 4. Als Antwort auf Rosenthal und als Lösung des Problems führe ich in Abschnitt 5 eine Unterscheidung zwischen aufrichtigen und echten Selbstbekundungen ein. In Abschnitt 6 kehre ich zur Hauptfrage zurück und verteidige die These, dass die Autorität der Ersten Person nicht auf einem privilegierten Zugang zum Bewusstsein beruht, sondern auf einer sprachlichen Kompetenz, die zugleich ein entscheidender Faktor in der Entwicklung des reflexiven Selbstbewusstseins ist.

## **1. Fragen zur Autorität der Ersten Person**

Dass die Entwicklung höherer kognitiver Fähigkeiten durch das Erlernen einer Sprache gefördert oder überhaupt erst ermöglicht wird, kann als selbstverständlich betrachtet werden. Strittig ist, welche der höheren kognitiven Fähigkeiten sich erst mit dem Spracherwerb ausbilden und welche bereits vorsprachlich erworben werden. Wie steht es diesbezüglich mit dem reflexiven Selbstbewusstsein und der Autorität der Ersten Person?

Aus erkenntnistheoretischer Sicht besteht zunächst kein Grund, diese Phänomene mit dem Spracherwerb in Verbindung zu bringen. Ausgangspunkt ist hier die These, dass sich ein Subjekt seiner eigenen Bewusstseinsinhalte unmittelbar bewusst ist – also ein Bewusstsein von seinem Bewusstsein besitzt. Dafür scheint es nicht nötig zu sein, dass man den Inhalt dieses Meta-Bewusstseins sprachlich artikulieren kann. Freilich stellt sich dann die Frage, wie sich ein nicht an sprachliche Fähigkeiten gebundenes unmittelbares Selbstwissen ausbilden könnte. Cartesianer bemühen dafür eine metaphysische Erklärung in Form einer ontologisch primitiven Beziehung, in der jedes Subjekt zu seinen eigenen *cogitationes* steht. Empiristen behaupten dagegen, dass diese unmittelbare Bekanntschaft das Ergebnis eines wahrnehmungsähnlichen Vorgangs ist, nämlich der Introspektion. Diese beiden klassischen Modelle des Selbstwissens sind zwar umstritten, sie haben aber nichts von ihrer prinzipiellen Attraktivität verloren, wie die anhaltende Debatte darüber zeigt (siehe Gertler 2011, Hatzimoysis 2011, Smithies and Stoljar 2012). In einer modernisierten Form bilden diese

Modelle auch die Grundlage für kognitivistische Erklärungen unseres Selbstwissens durch Repräsentationsprozesse, die wie ein interner Monitor andere mentale Repräsentationen registrieren und diese Information auf höherer Stufe weiter verarbeiten. Auch ein solcher interner Monitoring-Mechanismus ist nicht auf sprachlich erworbene Fähigkeiten angewiesen (siehe Newen 2005; Vosgerau 2009).

Es wäre jedoch falsch, würde man diese Beispiele als Indiz für einen sich abzeichnenden Konsens darüber betrachten, dass es ein nicht an Sprache gebundenes unmittelbares Selbstwissen gibt. Es ist nicht nur die schon erwähnte expressivistische These von Wittgenstein, die dem widerspricht. Als einen weiteren Vertreter einer ganz anderen Position möchte ich an dieser Stelle auf Donald Davidson verweisen, dessen Aufsatz "First Person Authority" sehr gut verdeutlicht, welche tiefgehende Auffassungsunterschiede hier zu berücksichtigen sind.

Davidson stellt die Frage nach der Autorität der Ersten Person von Anfang an in einen sprachphilosophischen Kontext. Seine Begründung dafür lautet folgendermaßen:

Das Problem [der Autorität der Ersten Person] kann entweder mit Blick auf die Sprache oder mit Blick auf die Epistemologie erörtert werden. [...] Wenn das eigene Wissen eine systematische Differenzierung aufweist, dann müssen Wissensäußerungen diese Differenzierung widerspiegeln. Ich nehme deshalb an, dass wir, wenn wir die Autorität der Ersten Person *in der Rede* erklären können, viel, wenn nicht alles, getan haben werden, was getan werden muss, um die epistemologischen Tatsachen zu beschreiben und zu erklären.“ (Davidson 1984/dt. 2005, 151; meine Hervorhebung)

Die zuvor genannten Vertreter unmittelbaren Selbstwissens würden dem natürlich widersprechen. Eine Asymmetrie zwischen Selbst- und Fremdzuschreibungen besteht aus ihrer Sicht deshalb, weil es ein unmittelbares Wissen von eigenen Bewusstseinsinhalten gibt. Was könnte daran verkehrt sein? Für Davidson geht es dabei um die methodische Frage, welche Erklärungsrichtung man wählt. Wenn es ein unmittelbares Selbstwissen gibt, dann macht es Sinn zu sagen, dass sich die besondere Qualität unseres Selbstwissens auf die Äußerungen überträgt, mit denen wir dieses Wissen zum Ausdruck bringen. Aber die "systematische Differenzierung" unseres Wissens, wie Davidson es nennt, kann auch auf andere Weise verständlich gemacht werden. Man kann die Autorität der Ersten Person auch

zuerst “in der Rede” analysieren und davon ausgehend dann eine Theorie des Selbstwissens entwickeln.<sup>2</sup>

Aber warum sollte man den von Davidson vorgeschlagenen Weg gehen? Davidsons Antworten auf diese Nachfrage zeigen, dass gute sachliche Argumente hier nicht leicht zu finden sind.<sup>3</sup> Für ihn ist der Verweis auf ein Bewusstsein vom eigenen Bewusstsein im besten Fall “empty”, d.h. ohne Erklärungskraft (Davidson 1993, 249), im schlechtesten Fall ein Teil jenes philosophischen Mythos, den er den ‘Mythos des Subjektiven’ nennt (Davidson 1988). Was diesen Mythos genau ausmacht, ist jedoch schwer zu sagen. Er hat etwas mit dem metaphysischen Dualismus zu tun und mit der Cartesischen Vorstellung von Infallibilität und Transparenz des Geistes. Darüberhinaus ist der Mythos aber nichts weiter als ein Ausdruck – und keine Begründung – der Überzeugung, dass die Autorität der Ersten Person auf einer besonderen Form des Selbstwissens beruht. Für die Frage nach der sachlichen Korrektheit dieser Überzeugung ist somit nichts gewonnen.

Da Davidsons Argumentation an diesem Punkt zu versagen droht, werde ich meine Kritik der Idee eines unmittelbaren Selbstwissens auf eine expressivistische Grundlage stellen. Ein entscheidender Unterschied dabei ist folgender: Davidson scheint es zweckmäßig, die Frage der Autorität der Ersten Person zunächst für Aussagen zu diskutieren, in denen ein Sprecher versichert, gewisse Überzeugungen, Wünsche, Hoffnungen oder Absichten zu haben. “Was für diese propositionalen Einstellungen gilt”, so Davidson, “das sollte, wie es scheint, auch für Empfindungen und alles übrige relevant sein.” (Davidson 1984/dt. 152). Aus expressivistischer Sicht ist es dagegen zweckmäßiger, die Frage der Autorität zuerst für Selbstbekundungen zu diskutieren, mit denen Sprecher ihre Empfindungen zum Ausdruck bringen. Was wir dabei über den Ausdruck von Empfindungen lernen, kann dann auch für andere mentale Zustände relevant sein, die Gegenstand unseres reflexiven Selbstbewusstseins sind.

Zur besseren Übersicht für die weitere Diskussion möchte ich an dieser Stelle die wesentlichen Fragen formulieren und die Reihenfolge erläutern, in der ich sie behandeln werde. Ich beziehe mich dabei auf folgende zwei Thesen, deren Verhältnis es zu klären gilt:

---

<sup>2</sup> Nach Davidson bliebe hier allerdings nicht mehr viel zu tun, falls mit dem ersten Schritt schon alles getan ist, um die epistemologischen Tatsachen zu beschreiben, wie er meint.

<sup>3</sup> Siehe Davidsons Antwort auf Thöle 1993.

(These A) Ein Sprecher, der sich selbst eine Empfindung zuschreibt, besitzt aufgrund seines privilegierten Zugangs zu seinen eigenen Empfindungen ein unmittelbares Selbstwissen.

(These B) Ein Sprecher, der sich selbst eine Empfindung zuschreibt, verfügt bei Selbstzuschreibungen dieser Art über die Autorität der Ersten Person.

Die These (A) entspricht den klassischen Modellen des Selbstwissens durch unmittelbare Bekanntschaft oder durch Introspektion. Verglichen damit ist die These (B) eine viel schwächere These, die nur behauptet, dass es eine Autorität der Ersten Person gibt, aber nichts über deren Ursprung aussagt. These (A) schließt daher These (B) ein, nicht aber umgekehrt.

Die erste Frage, die ich stelle, wird daher folgende sein:

(Frage 1) Macht es Sinn, These (B) ohne These (A) zu vertreten?

Bei dieser Frage geht es zunächst um den *Begriff* von Autorität. Können wir diesen Begriff überhaupt *verstehen*, ohne implizit vorauszusetzen, was in These (A) explizit gemacht wird. Ist die Antwort auf diese Frage positiv, wie ich argumentieren werde, dann stellt sich zweitens folgende Frage:

(Frage 2) Warum sollte These (B) wahr sein, wenn These (A) nicht gilt?

Anders als Davidson bin ich der Ansicht, dass es zur Beantwortung dieser Frage nicht genügt, Prinzipien für eine Interpretation sprachlicher Äußerungen zu benennen, aus denen folgt, dass Selbstzuschreibungen großteils wahr sind. Denn ein solches Prinzip der Nachsichtigkeit gilt ja ganz allgemein für das, was Sprecher behaupten. Ich setze hier auf eine expressivistische Analyse von Selbstzuschreibungen als Alternative zu den Modellen des unmittelbaren Selbstwissens. Meine dritte Frage lautet daher:

(Frage 3) Warum sind Selbstbekundungen, die mit der Autorität der Ersten Person geäußert werden, keine gewöhnlichen Behauptungen?

Wie sich zeigen wird, spielt auch bei dieser Frage die Unterscheidung zwischen den Thesen (A) und (B) eine tragende Rolle. Wäre nämlich die Autorität der Ersten Person, mit der wir

Selbstbekundungen äußern, im unmittelbaren Selbstwissen des jeweiligen Sprechers begründet, dann wären solche Äußerungen Ausdruck dieses Wissens und somit gewöhnliche Behauptungen. Sind dagegen Selbstbekundungen *keine* gewöhnlichen Behauptungen, dann spricht dies gegen die Annahme eines zugrundeliegenden Selbstwissens, das in solchen Äußerungen zum Ausdruck kommt. Diese Argumentation lässt erkennen, wie eng der Zusammenhang zwischen den ersten drei Fragen ist.

Es gibt aber noch eine vierte Frage, die ich in diesem Aufsatz behandeln werde. Sie betrifft die Rolle, die die Entwicklung des reflexiven Selbstbewusstseins in diesem Kontext spielt:

(Frage 4) Warum müssen Sprecher über ein reflektives Selbstbewusstsein verfügen, damit sie Selbstbekundungen mit der Autorität der Ersten Person äußern können?

Solange man an These (A) festhält, ist die Antwort auf diese Frage klar und einfach. Lässt man hingegen These (A) fallen und hält nur an These (B) fest, wird die Antwort schwierig. Ob es sich lohnt, diese Schwierigkeiten auf sich zu nehmen, gilt es nun zu ergründen.

## **2. Autorität ohne privilegierten Zugang**

Peter Carruthers beschreibt in seinem Buch *The Opacity of the Mind* (2011) das menschliche Bewusstsein als eine Art Schleier, hinter dem sich die für unser Verhalten bestimmenden kognitiven Vorgänge verbergen. Unser Selbstwissen von diesen Vorgängen, so Carruthers, beruhe zum größten Teil auf Mechanismen der Selbstinterpretation. Dass diese Mechanismen nicht besonders zuverlässig sind, zeigen die vielen Beispiele von Konfabulation und Selbsttäuschung. Wir können den Schleier des Bewusstseins ebensowenig lüften, wie wir einen direkten und unvermittelten Blick ins Bewusstsein anderer Personen werfen können. In beiden Fällen arbeiten wir mit mehr oder minder gut bestätigten Theorien über mögliche Handlungsgründe, die uns der Bewusstseinschleier verbirgt. Die Vorstellung, es gäbe eine klare Asymmetrie zwischen Selbstzuschreibungen und Fremdzuschreibungen mentaler Zustände, beruht daher nach Carruthers auf einem Irrtum, den es empirisch zu erklären gilt.

Auf ersten Blick handelt es sich hier um eine Position, die der Cartesischen Auffassung des menschlichen Geistes mindestens ebenso diametral entgegengesetzt ist wie

die Position von Davidson. Dem ist aber nicht wirklich so. Denn es gibt zwei Punkte, in denen Carruthers These der Opakheit des Geistes mit der Cartesischen These der Transparenz des Geistes übereinstimmt. Zum einen nimmt Carruthers an, dass es ein von Interpretation freies “sensory self-knowledge” gibt (Carruthers 2011, 72). Zum anderen hält er daran fest, dass in jenen Fällen, wo es eine Autorität der Ersten Person gibt, diese Autorität in einem kognitiven Mechanismus begründet liegt, der einem privilegierten Zugang zu den eigenen Bewusstseinsinhalten gleich kommt. Während für Descartes ein solches Privileg die Regel ist, ist es für Carruthers die Ausnahme. Dieses – zugegebenermaßen kleine – Zugeständnis gegenüber einem Vertreter der Idee des unmittelbaren Selbstwissens, wirft eine interessante Frage auf: Muss man dieses Zugeständnis machen, wenn man der Intuition gerecht werden will, dass es zumindest bei sinnlichen Empfindungen eine Autorität der Ersten Person gibt?

Anders als Carruthers bin ich der Auffassung, dass ein solches Festhalten an der Idee des privilegierten Zugangs auch in diesem Fall keinesfalls notwendig ist. Dabei räume ich ein, dass wir nicht wissen, wie opak unser Bewusstsein tatsächlich ist. Vielleicht sind viele unserer Selbstzuschreibungen tatsächlich von Selbstinterpretationen abhängig, die von Vorurteilen beeinflusst und daher unzuverlässig sind. Vielleicht gibt es aber auch noch andere Gründe, warum ein naives Konzept von Autorität vor empirischer Widerlegung nicht gefeit ist. Jedenfalls brauchen wir – in diesem Punkt stimme ich mit Carruthers überein – ein Konzept von erstpersonaler Autorität, das mit allen verfügbaren empirischen Daten in Einklang steht.

Begreift man die Autorität der Ersten Person als ein sprachliches Phänomen, gelangt man zu einem Konzept von erstpersonaler Autorität, das gute Aussichten hat, dieser Anforderung zu genügen. Ich stelle dieses Konzept zunächst in Form einer Definition vor, um es dann im Detail zu diskutieren:

DEF 1 Ein Sprecher S verfügt in Bezug auf eine Äußerung x über die Autorität der Ersten Person, genau dann wenn gilt:

- (i) die Äußerung x beschreibt ein mentales Ereignis oder einen Zustand, dessen sich S zum Zeitpunkt der Äußerung bewusst ist.<sup>4</sup>

---

<sup>4</sup> Streng genommen wäre es bei dieser Definition nicht nötig, die Autorität der Ersten Person auf mentale Phänomene *zum Zeitpunkt der Äußerung* zu beschränken. Wer von Erlebnissen berichtet, an die er sich lebhaft erinnern kann, kann für diese Berichte ebenfalls die Autorität der Ersten Person beanspruchen. Der Einfachheit halber halte ich jedoch an der üblichen Beschränkung auf das gegenwärtig Bewusste fest.



- (ii) S besitzt in Bezug auf x die sprachliche Kompetenz, die man von einem normalen Sprecher erwartet.
- (iii) die Tatsache, dass S x äußert, ist unter der Voraussetzung, dass die Bedingungen (i) und (ii) erfüllt sind, ein *prima facie* Grund zu glauben, dass x wahr ist.

Der Grundgedanke dieser Definition ist sehr einfach: Die Autorität der Ersten Person anzuerkennen, bedeutet, dass man bereit ist, einem Sprecher zu glauben, was er sagt, ohne dafür weitere Evidenzen heranzuziehen. Liegt eine mit Kompetenz geäußerte Selbstzuschreibung vor, ist dies Evidenz genug. Damit ist weder gesagt, dass Selbstzuschreibungen irrtumsfrei sind, noch dass sie den höchst möglichen Grad an Gewissheit besitzen. Selbstzuschreibungen können also auch falsch sein.<sup>5</sup> Doch handelt es sich dabei stets um Fehler, die auf besondere Umstände zurückgeführt werden können, also um Ausnahmefälle. *Prima facie* ist die Tatsache, dass eine Selbstbekundung geäußert wird, ein guter Grund anzunehmen, dass sie wahr ist.

Obwohl all dies intuitiv äußerst plausibel erscheint, lohnt es sich, die Bedingungen von DEF1 jeweils einzeln zu betrachten. Zur Veranschaulichung betrachte man den alltäglichen Fall eines Patienten, der mit Knieschmerzen zum Arzt kommt. Der üblichen Routine folgend ersucht der Arzt den Patienten zu demonstrieren, in welcher Position sein Knie schmerzt. Der Patient beginnt eine Kniebeuge, stoppt, und sagt dann auf sein linkes Knie zeigend:

(1) Jetzt tut es hier weh.

Es spielt keine Rolle, ob der Patient noch Näheres über seine Schmerzen sagen kann. Bereits seine Aussage (1) kann als eine minimale Beschreibung einer Empfindung gelten, sodass Bedingung (i) jedenfalls erfüllt ist.<sup>6</sup>

Sicher können wir uns dessen aber nur sein, wenn auch die Bedingung (ii) erfüllt ist. Stellen wir uns deshalb einen Fall vor, in dem die verbale Kompetenz eines Sprechers zweifelhaft ist. Angenommen es gäbe einen äußerst zuverlässigen Lügendetektor, mit dem sich feststellen lässt, ob jemand aufrichtig ist oder absichtlich die Unwahrheit sagt. Und nehmen wir weiters an, jemand, an dessen Aufrichtigkeit kein Zweifel besteht, behauptet, er habe Schmerzen, verhält sich aber ansonsten völlig normal und so fröhlich und gelassen, als hätte er gar keine Schmerzen. Dann wären wir ebenso verwirrt wie im Fall eines Kindes, das

---

<sup>5</sup> Einen solchen Ausnahmefall – den Fall der Schmerz-Angst – werde ich später behandeln.

<sup>6</sup> Ein Mißverständnis wäre es zu denken, Bedingung (i) könnte nur erfüllt sein, wenn ausser dem Schmerzgefühl noch ein weiterer kognitiver Zustand vorliegt, nämlich die Überzeugung ‘Ich habe jetzt Schmerzen’. Dass dies nicht nötig ist, werde ich in Abschnitt 3 zeigen.

noch nicht begriffen hat, was das Wort 'Schmerzen' bedeutet. Auch im Fall eines Erwachsenen, der scheinbar aus Lust und Laune sagt 'Ich habe Schmerzen', ohne Hilfe oder Mitleid zu erwarten, besteht daher der Verdacht, dass ein Missverständnis oder ein sprachlicher Missbrauch vorliegt. Dass solche Fälle nicht als Gegenbeispiele gegen die Autorität der Ersten Person herangezogen werden können, dafür soll Bedingung (ii) sorgen.

Ähnlich verhält es sich bei Bedingung (iii). Auch in diesem Fall sind es gewisse Grenzfälle, die zu dieser relativ komplizierten Bedingung Anlass geben. So kann sich ein Schwindler, der Schmerzen nur vortäuscht, die Tatsache zunutze machen, dass man seine Mitteilung gewöhnlich für glaubwürdig hält. Das bedeutet, dass wir auch dem erfolgreichen Schwindler die Autorität der Ersten Person zugestehen müssen, denn er nutzt sie ja aus, um Mitleid zu erregen oder jemanden zu täuschen. Wie steht es aber mit einer Person, von der allgemein bekannt ist, dass sie dazu neigt, sich Schmerzen einzubilden oder diese vorzutäuschen. Einem Hypochonder oder einem notorischen Schwindler können wir nicht absprechen, dass er manchmal die Wahrheit sagt. Man könnte sogar die Auffassung vertreten, dass es schmerzhaft empfindungen auch ohne klare physiologische Ursache gibt, weshalb auch ein eingebildeter Schmerz als etwas Schmerzhaftes gelten könnte. Was in diesem Fall wiederum in Zweifel gerät, ist jedoch die sprachliche Kompetenz solcher Personen. Sie können zwar immer noch sagen, sie hätten Schmerzen, aber können sie wirklich meinen, was sie sagen? Das hängt natürlich davon ab, was hier mit 'meinen' gemeint ist. Sicherlich können sie mit ihren Äußerungen nicht all das bewirken, was ein normaler Sprecher bewirkt, wenn er Schmerzen bekundet. Hypochonder und notorische Schwindler werden nicht ernst genommen. Aus Bedingung (iii) kann man ersehen, weshalb es sich bei dieser Bedingung um einen Ausnahmefall und kein Gegenbeispiel gegen die These der Autorität der Ersten Person handelt.

Wie steht es aber mit den vielen Konfabulationen und Selbsttäuschungen, auf die sich Carruthers beruft? Zeigen diese Beispiele, dass die meisten unserer Selbstzuschreibungen auf Selbstinterpretationen beruhen und daher mit keiner besonderen erstpersonalen Autorität versehen sind? Wenn dem so wäre, müssten wir der Definition als weitere Bedingung hinzufügen, dass die Äußerung des Sprechers "auf keiner Selbstinterpretation beruhen darf".

Der Normalfall des Patienten, der dem Arzt seine Schmerzen schildert, zeigt, warum eine solche Bedingung problematisch wäre. Denn wie sollte ein Arzt wissen, ob eine Selbstinterpretation im Spiel ist oder nicht? Würde man diese Bedingung hinzufügen, ließe sich nicht einmal in einem ganz gewöhnlichen Fall entscheiden, ob jemand über die Autorität der Ersten Person verfügt oder nicht. Die Praxis ist, dass der Arzt glaubt, was ein Patient sagt,

auch wenn die Möglichkeit einer Selbstinterpretation *nicht* ausgeschlossen ist. Für einen Skeptiker, wie Carruthers, mag dies eine naive Haltung sein. Für mich zeigt es nur, dass sich die Zuschreibung der Autorität der Ersten Person und das Vorliegen einer Selbstinterpretation *auf rein begrifflicher Ebene* nicht ausschließen müssen.

Schwieriger wird es, wenn wir zur Frage weiter gehen, *warum* Sprecher über diese Autorität verfügen. Dann kommt auch die Idee des privilegierten Zugangs ins Spiel und zwischen *dieser* These und der These, dass Selbstzuschreibungen auf Selbstinterpretation beruhen, gibt es zweifellos einen Konflikt. Aber was folgt daraus? Für Carruthers bedeutet es, dass in jenen Fällen – nämlich beim sinnlichen Selbstwissen – in denen die Autorität der Ersten Person nicht zu bestreiten ist, die Existenz eines privilegierten Zugangs in Form eines entsprechenden metakognitiven Mechanismus angenommen werden muss. Es gibt aber auch eine andere Schlussfolgerung, die man an dieser Stelle ziehen kann. Vielleicht zeigt dieser Konflikt ja, dass die Autorität der Ersten Person gar nicht auf der Annahme eines privilegierten Zugangs beruhen muss. Sie kann auch daher rühren, dass Selbstzuschreibungen eine *expressive* Funktion haben. Fehlt einer Selbstbekundung diese expressive Funktion, dann haben wir es mit einer Selbstzuschreibung zu tun, die auf einer Selbstinterpretation beruht und daher keine erstpersonale Autorität beanspruchen kann. Warum dies die richtige Schlussfolgerung ist, will ich nun erläutern.

### **3. Selbstbekundungen**

Als ein sprachliches Phänomene betrachtet, ist die Autorität der Ersten Person an Selbstzuschreibungen als eine besondere Form von verbaler Äußerung gebunden. In diesem Fall ist die Bezeichnung ‘Selbstzuschreibung’ jedoch zu allgemein. Wenn solche Äußerungen eine spezifische expressive Funktion haben, verdienen sie einen besonderen Namen. Vertreter dieser Auffassung sprechen deshalb von “Selbstbekundungen” (*avowals*), und dieser Tradition werde ich nun folgen (siehe Gasking 1966).

Mit der Wahl dieses Namens geht die Verpflichtung einher zu erklären, was die besondere Ausdrucksfunktion von Selbstbekundungen ist. Dies ist keine leichte Aufgabe, wie wir gleich sehen werden. Es tauchen dabei Fragen bezüglich der semantischen Eigenschaften und Verwendungsregeln für solche Äußerungen auf, darunter insbesondere die Frage, ob Selbstbekundungen sowohl wahr als auch falsch sein können, und wann eine

Selbstbekundung aufrichtig ist. Die grundlegende Frage, mit der wir hier beginnen müssen, ist jedoch, warum Selbstbekundungen keine gewöhnlichen Behauptungen sind.

Betrachten wir dazu jene Passage in Wittgensteins *Philosophischen Untersuchungen*, die als *locus classicus* für eine expressivistische Analyse des Selbstbekundens gilt (PU §244). Wittgenstein stellt sich hier die Frage: “Wie *beziehen* sich Wörter auf Empfindungen?” Die Schwierigkeit der Frage bestehe darin, so Wittgenstein, dass nicht klar ist, wie “ein Mensch die Bedeutung der Namen von Empfindungen lernt”. Als Antwort darauf beschreibt er folgende Möglichkeit:

“Es werden Worte mit dem ursprünglichen, natürlichen, Ausdruck der Empfindung verbunden und an dessen Stelle gesetzt. Ein Kind hat sich verletzt, es schreit; und nun sprechen ihm die Erwachsenen zu und bringen ihm Ausrufe und später Sätze bei. Sie lehren das Kind ein neues Schmerzverhalten.

>>So sagst Du also, dass das Wort >Schmerz< eigentlich das Schreien bedeute?< — Im Gegenteil; der Wortausdruck des Schmerzes ersetzt das Schreien und beschreibt es nicht.“ (PU §244)

Inwiefern lassen sich sprachliche Äußerungen mit einem natürlichen Ausdrucksverhalten vergleichen, wie z.B. Weinen, Schreien, Stöhnen, aber auch Lachen oder Jubeln? Wittgenstein meint, Worte wie “Schmerz” werden gelernt, indem Kinder diese Worte zu ihrem Repertoire des Schmerzverhaltens hinzufügen. Dies kann leicht als eine behaviouristische These missverstanden werden, und ich werde ganz am Schluss eine ähnliche These betrachten, die Wittgenstein vertreten haben könnte. Ausschließen können wir jedoch die Möglichkeit, dass das Wort “Schmerz” gar nicht den Schmerz bedeutet, sondern nur ein gewisses Schmerzverhalten. Diese Klarstellung, die Wittgenstein explizit trifft, beantwortet aber noch nicht die entscheidende Frage, was es bedeuten soll, dass der Wortausdruck den Schrei “ersetzt” und nicht “beschreibt”?

Einer verbreiteten Auffassung zufolge vertritt Wittgenstein hier folgende These:

(SEX) Selbstbekundungen haben keine beschreibende, sondern eine rein expressive Funktion, so wie ein natürliches Ausdrucksverhalten (z.B. Stöhnen, Weinen oder Lachen).

Das Akronym “SEX” steht hier für “simple expressivism”. So nennt Dorit Bar On die Position, die Wittgenstein häufig zugeschrieben wird (Bar On 2004, 228). Ob man Wittgenstein wirklich als “einfachen Expressivisten” interpretieren darf, ist allerdings strittig (siehe Finkelstein 2003).<sup>7</sup> Auf jeden Fall gibt es zwei gute Gründe, die gegen die expressivistische These in der vorliegenden Form sprechen.

Der eine Grund ist, dass Selbstbekundungen sprachliche Handlungen sind und als solche kein Ersatz für Verhaltensweisen wie Weinen oder Stöhnen sein können. Es ist zumindest voreilig zu behaupten, eine Sprachäußerung könne eine Empfindung in der gleichen Weise zum Ausdruck bringen wie eine Verhaltensweise. Bar On schlägt daher vor, die expressive Funktion von Selbstbekundungen nicht mit der natürlichen (kausalen) Form des Ausdrückens zu vergleichen, sondern als eine eigene genuine Form des Ausdrückens zu analysieren, die sie in Anlehnung an Sellars den “action sense of expressing” nennt (Bar On 2004, 216).

Der zweite Grund, der gegen (SEX) spricht, ist schwerwiegender. Selbstbekundungen verhalten sich in logischer und syntaktischer Hinsicht ganz wie gewöhnliche Aussagen. Man kann aus der Selbstbekundung ‘Ich habe Schmerzen’ folgern ‘Jemand hat Schmerzen’; man kann sie mit anderen Aussagen syntaktisch verknüpfen, indem man sie als Antezedens in einen Konditionalsatz integriert, und man kann Selbstbekundungen mit anderen Aussagen kontrahieren, indem man z.B. aus ‘Ich habe Schmerzen’ und ‘Er hat Schmerzen’ die komplexe Aussage bildet: ‘Wir beide haben Schmerzen’ oder ‘Er hat Schmerzen und ich auch’. All dies spricht dafür, dass Selbstbekundungen genauso wie andere Aussagen eine deskriptive Funktion haben. Sie bringen nicht nur Empfindungen, Gefühle oder Gedanken zum Ausdruck, sondern *beschreiben* diese auch. Ein blosser Schrei tut das nicht. In dieser Hinsicht sind Selbstbekundungen überhaupt nicht mit einem natürlichen Ausdrucksverhalten zu vergleichen (siehe Geach 1965, Wright 1998, Bar On 2004).

Trotzdem besteht vorerst noch kein Grund, von der These abzurücken, dass Selbstbekundungen aufgrund ihrer spezifischen Ausdrucksfunktion keine gewöhnlichen Behauptungen sind. Diese Position kann man zwar vertreten, wie wir im nächsten Abschnitt sehen werden. Zunächst fordern uns die Einwände gegen den *simple expressivism* nur dazu auf, genauer zu prüfen, was mit dieser spezifischen Ausdrucksfunktion gemeint sein könnte.

Grundsätzlich gibt es zwei Möglichkeiten, um eine Ausdrucksfunktion genauer zu bestimmen. Man kann dabei auf die *Genese* dieser Funktion Bezug nehmen oder man kann

---

<sup>7</sup> Man beachte, dass Wittgenstein an der zitierten Stelle nur sagt, das Wort ‘Schmerz’ beschreibe nicht *das Schreien*; er sagt nicht, es beschreibe nicht den Schmerz. Die These (SEX) läßt sich daher durch das vorliegende Zitat nicht belegen.

genauer charakterisieren, *was* dabei zum Ausdruck gebracht wird. Spricht man von einer “natürlichen Ausdrucksfunktion“, folgt man dem ersten Weg, indem man auf einen kausalen Ursprung der Funktion verweist. Auch der vorhin genannte “action sense of expressing”, den Bar On bevorzugt, gibt eine genetische Antwort auf unsere Frage. Ich möchte stattdessen vorschlagen, den anderen Weg zu gehen und die Besonderheit von Selbstbekundungen darin zu suchen, *was* sie zum Ausdruck bringen, nicht darin, *wie* sie es tun.

Vergleichen wir deshalb eine Selbstbekundung mit einer gewöhnlichen Behauptung unter diesem Gesichtspunkt. Was ist der Unterschied, wenn wir die Aussage ‘Ich habe Schmerzen’ einmal als Selbstbekundung (S) und im anderen Fall als gewöhnliche Behauptung (B) interpretieren? Im ersten Fall würde gelten

(S) S ist der verbale Ausdruck eines Bewusstseinszustandes M, in dem sich der Sprecher zum Zeitpunkt der Äußerung befindet.

Interpretieren wir dieselbe Aussage hingegen als gewöhnliche Behauptung, gilt folgendes:

(B) S ist der verbale Ausdruck der Überzeugung des Sprechers, dass er sich zum Zeitpunkt der Äußerung im Zustand M befindet.

Das Beispiel einer Schmerzäußerung, das wir bisher betrachtet haben, macht hier Schwierigkeiten, weil man die Überzeugung, Schmerzen zu haben, nicht so leicht von der Schmerzempfindung trennen kann. Klarer wird der Unterschied zwischen (S) und (B) deshalb, wenn wir das Beispiel wechseln und für M den Zustand des heimlichen Verliebtseins einsetzen. Solche Gefühle können auch dem betreffenden Subjekt verborgen bleiben, wenn z.B. jemand unsicher ist, ob er (oder sie) nun tatsächlich verliebt ist oder nicht. Es könnte daher sein, dass jemand von seinen Freunden davon überzeugt werden muss, dass er verliebt ist, weil er diese Gefühle selbst nicht wahr haben will. Solange dieser Zustand anhält, kann seine Behauptung ‘Ich bin verliebt’ keine Selbstbekundung sein. Sie kann höchstens als Behauptung interpretiert werden, mit der man eine Überzeugung kundtut, zu der einem andere verholfen haben. Genau genommen sollte jemand in dieser Situation daher sagen ‘Ich *glaube*, ich bin verliebt’.

Erklärt man den Unterschied zwischen Selbstbekundungen und gewöhnlichen Behauptungen auf diese Weise, liegen die Vorteile auf der Hand. Die heikle Frage, ob und in welchem Sinn es sich dabei um eine natürliche Ausdrucksfunktion handelt, brauchen wir jetzt

nicht zu lösen. Der größere Vorteil liegt aber darin, dass es jetzt keinen Grund mehr gibt, Selbstbekundungen einen deskriptiven Gehalt abzusprechen. Das besondere Merkmal einer Selbstbekundung ist, so können wir jetzt sagen, dass sie einen mentalen Zustand des Sprechers *sowohl* beschreibt *als auch* zum Ausdruck bringt. Genau darin sieht auch Kevin Falvey den Vorzug eines kognitiven Expressivismus gegenüber einem nicht-kognitiven Expressivismus:

"[We must] avoid the errors of non-cognitivist expressivism, and take seriously the idea that avowals have an expressive function, while retaining the natural and compelling idea that the sentences used in making them are capable of truth or falsity, and presented as true, so that such avowals also have the pragmatic status of assertions" (Falvey 2000, 73).

Will man den Punkt, auf den es hier ankommt, noch präziser machen, muss man genauer sagen, was den "pragmatischen Status" einer Behauptung ausmacht. Es könnte damit bloss gemeint sein, wie Falvey nahelegt, dass es sich dabei um wahrheitsfähige Aussagen handelt; es könnte aber auch eine stärkere These gemeint sein, gemäß der eine Selbstbekundung neben ihrer spezifischen Ausdrucksfunktion *alle* Funktionen einer gewöhnlichen Behauptung hat. Dann würde eine Selbstbekundung der Form 'Ich bin in M' mindestens *drei* Funktionen erfüllen. Sie würde (1) den Zustand M beschreiben, (2) den Zustand M ausdrücken, und zusätzlich (3) auch eine Überzeugung mit dem Inhalt [Ich bin in M] zum Ausdruck bringen.

Grundsätzlich spricht nichts dagegen, einer einzigen Äußerung alle drei Funktionen zuzuschreiben, es sollte dies aber keine Bedingung sein, die für *jede* Selbstbekundung gilt.<sup>8</sup> Denn man muss hier berücksichtigen, dass es elementare Selbstbekundungen von Kindern gibt, bei denen eine so komplexe Analyse sicher inadäquat wäre. Wenn ein Kind sagt 'Ich habe Durst', bringt es wohl nur den Zustand M zum Ausdruck, nicht aber eine Überzeugung mit dem Inhalt [Ich bin in M]. Ich ziehe daher eine Formulierung der expressivistischen Auffassung von Selbstbekundungen vor, die für solch elementare Selbstbekundungen Platz lässt. Dies nenne ich im Folgenden die These des *Deskriptiven Expressivismus*:

(DEX) Eine Selbstbekundung ist die Selbstzuschreibung eines mentalen Zustandes, der damit sowohl beschrieben als auch zum Ausdruck gebracht wird, ohne dass damit

---

<sup>8</sup> Hier stimme ich mit Bar On überein, wenn sie sich gegenüber einer "dual expression thesis" vorsichtig verhält (siehe Bar On 2004, 307ff.).

auch zum Ausdruck gebracht werden muss, dass der Sprecher überzeugt ist, zum Zeitpunkt der Äußerung in diesem Zustand zu sein.

Damit ist die Frage, worin sich Selbstbekundungen von gewöhnlichen Behauptungen unterscheiden, fürs Erste beantwortet. Fürs Erste deshalb, weil wir die These (DEX) auf "echte" Selbstbekundungen einschränken müssen. Dies hat mit einem Problem zu tun, das ich im nächsten Abschnitt behandeln werde.

#### **4. Das Problem falscher Selbstbekundungen**

Selbstbekundungen bringen genau jene Zustände zum Ausdruck, die sich ein Sprecher mittels dieser Selbstbekundung zuschreibt. Diese expressivistische These kann man vertreten, ohne sich festzulegen, ob es sich dabei um eine natürliche (kausale) oder eine andere Art von Ausdrucksfunktion handelt. Einer Frage kann sich ein Expressivist jedoch nicht entziehen: Gibt es gemäß seiner Auffassung auch falsche Selbstbekundungen? Wenn eine Selbstbekundung aufgrund ihrer expressiven Funktion genau den mentalen Zustand zum Ausdruck bringt, den sich ein Sprecher in dieser Selbstzuschreibung zuschreibt, wie kann es dann falsch sein, was der Sprecher sagt?

Zunächst scheint dies gar kein Problem zu sein; ist es doch möglich, dass jemand durch sein nicht-verbales Verhalten signalisiert, er habe Schmerzen, ohne dass dies zutrifft. Umso mehr sollte es möglich sein, dass jemand sagt 'Ich habe Schmerzen', obwohl es falsch ist. Weder in dem einen noch im anderen Fall ist ausgeschlossen, dass sich jemand so verhält, als wäre er in einem Zustand M, ohne in M zu sein.

Wie David Rosenthal deutlich gemacht hat, gibt es hier aber sehr wohl ein Problem, wenn man Selbstbekundungen sowohl eine deskriptive als auch eine expressive Funktion zuschreibt (siehe Rosenthal 2010). Betrachten wir dazu folgendes Beispiel einer falschen Selbstbekundung. Es gibt das Phänomen, dass Patienten aus Furcht vor einem drohenden Schmerz behaupten, Schmerzen zu empfinden, die gar nicht vorhanden oder noch gar nicht eingetreten sind. Die sogenannte *dental fear* ist ein Fall solcher "Schmerz-Angst". Dabei zeigen Patienten alle Anzeichen von Schmerzen, obwohl sie aus ärztlicher Sicht schmerzfrei sein müssten. Es wird z.B. ein Zahn behandelt, dessen Nerv abgestorben ist, oder es wurde eine lokale Anästhesie verabreicht, die jede Empfindung in der behandelten Region unterbindet. Ein Irrtum ist zwar in solchen Fällen nie ganz auszuschließen, aber die Wahrscheinlichkeit, dass der Patient sich die Schmerzen nur einbildet ist meist höher. Eine naheliegende Vermutung wäre, dass der Patient etwas spürt, z.B. das Vibrieren des Bohrers,



und dass diese an sich schmerzlose Empfindung zu einer Schmerzreaktion führt (siehe Rosenthal 2010, 109ff.).

Man kann sich bei der Analyse solcher Beispiele an folgendes Prinzip halten: Was es nicht gibt, kann auch nicht zum Ausdruck gebracht werden. Sind die Schmerzen nur eingebildet, kann sich der Betreffende nur so verhalten, als hätte er Schmerzen, *ausdrücken* kann er seinen (nicht vorhandenen) Schmerz nicht. Rosenthal geht hier aber noch einen Schritt weiter, wenn er aus diesem Beispiel folgende Schlussfolgerung zieht:

“The possibility of detectable factual error in such cases suggests that we should see remarks such as ‘I am in pain’ and ‘I think it’s raining’ as straightforward reports of individual’s mental states.” (Rosenthal, 2010, 109).

Das klingt zunächst so, als möchte Rosenthal den Unterschied zwischen der Perspektive der Ersten Person und der Perspektive der Dritten Person aufheben. Auch Selbstbekundungen – seien sie wahr oder falsch – sind nur Berichte, die jemand aus der Dritten Person Perspektive über sich selbst verfasst. Dies ist es jedoch nicht, was Rosenthal meint. Seine These läßt sich ohne den problematischen Begriff des “Berichtens” weniger mißverständlich auf folgende Weise formulieren:

(R) Selbstbekundungen bringen die erstpersionale Überzeugung eines Sprechers zum Ausdruck, dass er sich zum Zeitpunkt der Äußerung in dem darin beschriebenen mentalen Zustand befindet.

Damit ist klar, was Rosenthal bestreitet. Er bestreitet, dass es einen signifikanten Unterschied zwischen Selbstbekundungen und gewöhnlichen Behauptungen gibt. Weil Expressivisten wie Bar On nach einer solchen Differenz suchen, meinen sie, dass es im Fall von Selbstbekundungen keinen “brute error” – keinen simplen Irrtum – geben kann (siehe Bar On 2004, 200). Für Rosenthal zeigt das Beispiel der Schmerz-Angst jedoch, dass ein simpler Irrtum auch hier möglich ist, wenn Patienten zwei verschiedene Empfindungen miteinander verwechseln.

Rosenthal glaubt auch zeigen zu können, wodurch Expressivisten hier in die Irre geführt werden. Selbstbekundungen können mit einer gewöhnlichen Behauptung in einem besonderen Äquivalenzverhältnis stehen. Sie können “performance conditionally equivalent” sein, wie Rosenthal sagt (Rosenthal 2005, 274ff. und 2010, 118). Negiert man diese

Äquivalenz, indem man von zwei derartigen Aussagen nur eine verneint, kann man ein sogenanntes Moore-Paradoxon erzeugen, z.B. indem man gleichzeitig Folgendes behauptet:

- (1) Es regnet.
- (2) Ich glaube nicht, dass es regnet.

Eine ähnliche Inkonsistenz entsteht, wenn man im selben Atemzug 'Au weh!' sagt und dann hinzufügt 'Ich habe keine Schmerzen'. Auch damit verwickelt man sich in einen pragmatischen Widerspruch, weil die Selbstbekundung 'Ich habe keine Schmerzen' eine Überzeugung ausdrückt, die man nicht haben kann, wenn man Schmerzen hat, so wie man nicht die Überzeugung haben kann, nicht zu glauben, es regnet, wenn man genau dies glaubt. Es scheint also, dass sich Rosenthals Analyse verallgemeinern lässt, und dass die performativen Widersprüche, die hier entstehen, wesentlich mit der Ausdrucksfunktion von Selbstbekundungen zu tun haben.

Worin liegt nun das Problem für die expressivistische Auffassung von Selbstbekundungen? Da sich Selbstbekundungen in pragmatischer Hinsicht auf besondere Weise verhalten, zeigt Rosenthals Argument, dass jede weiter gehende Annahme bezüglich ihrer Ausdrucksfunktion für die Analyse pragmatischer Widersprüche verzichtbar ist (siehe Rosenthal 2010, 118ff.). Dies könnte man als einen methodischen Einwand gegen die expressivistische These werten. Darüberhinaus gerät jetzt aber eine weitere Annahme in Gefahr, die für den Expressivismus von grundlegender Bedeutung ist. Folgt man Rosenthals Analyse, ist es nicht mehr möglich, die Autorität der Ersten Person als ein sprachliches Phänomen unabhängig von der Idee des privilegierten Zugangs zum eigenen Bewusstsein zu erklären. Es gibt hier einen erklärungsbedürftigen Zusammenhang zwischen diesen pragmatischen Phänomenen und der Autorität der Ersten Person, der in folgende Frage mündet: Warum treten diese Widersprüche immer nur dann auf, wenn eine der beiden Aussagen mit der Autorität der Ersten Person geäußert wird? Rosenthal kommt hier nicht darum herum, die Idee des privilegierten Zugangs zu bemühen, wie man an seiner Antwort sieht: "When one describes oneself as believing something or as being disposed to do so, it's on the basis of one's subjectively unmediated access to the states." (Rosenthal 2010, 119).

Einen solchen Rekurs auf die Idee des privilegierten Zugangs möchte man jedoch gerade vermeiden, indem man die Autorität der Ersten Person als ein sprachliches Phänomen zu erklären versucht. Wenn wir an diesem Projekt weiter festhalten, müssen wir daher das Problem der falschen Selbstbekundungen anders lösen als Rosenthal vorschlägt.

## 5. Aufrichtige und echte Selbstbekundungen

Die Analyse pragmatischer Widersprüche ist ein eigenes Thema, das ich an dieser Stelle nicht weiter zu verfolgen brauche. Denn in einem Punkt hat Rosenthal sicher Recht: Aus der Tatsache, dass zwei Äußerungen die gleichen Performanzbedingungen haben, *folgt* nicht, dass sie dieselbe Ausdrucksfunktion besitzen. Es schließt dies allerdings auch nicht aus, und dies genügt, um die expressivistische These gegen Rosenthals Argument zu verteidigen.

Im Fall einer gewöhnlichen Behauptung wissen wir, wann eine solche Behauptung als aufrichtig gilt. Sie ist es dann, wenn derjenige, der die Behauptung macht, glaubt, was er behauptet. Unaufrichtig ist hingegen jemand, der das Gegenteil dessen glaubt, was er behauptet. Wären Selbstbekundungen gewöhnliche Behauptungen, wäre der Fall damit erledigt, weil sich ihre Aufrichtigkeit dann genau auf die gleiche Weise erklären liesse. Gemäß der These des Expressivismus sind Selbstbekundungen jedoch *keine* gewöhnlichen Behauptungen. Ein alternativer Erklärungsansatz ist also nötig.

Was wir an dieser Stelle brauchen, ist eine Unterscheidung zwischen *aufrichtigen* und *echten* Selbstbekundungen. *Aufrichtige* Selbstbekundungen können falsch sein, nicht aber *echte* Selbstbekundungen. Dass man diese Unterscheidung machen muß, lässt sich an der soeben beschriebenen performativen Äquivalenz von ‘Au weh!’ und ‘Ich habe Schmerzen’ verdeutlichen. Diese Äquivalenz sollte auch dann erhalten bleiben, wenn es sich um eine falsche Selbstzuschreibung handelt, weil der Sprecher beispielsweise unter Schmerz-Angst leidet. Was wir dem Patienten nicht absprechen können, ist seine Aufrichtigkeit, wenn er sagt, er habe Schmerzen. Dementsprechend können wir auch sein Jammern nicht als unaufrichtig bezeichnen. Wie kann dies jedoch sein, wenn er gemäß Voraussetzung gar keine Schmerzen hat?

Anstatt die Aufrichtigkeit einer Selbstbekundung durch das *Vorhandensein* einer entsprechenden Überzeugung zu erklären, kann ein Expressivist dafür das *Fehlen* der gegenteiligen Überzeugung heranziehen. Wir gelangen so zu folgender indirekten Definition von Aufrichtigkeit:

DEF2 Eine Selbstbekundung, mit der sich ein Sprecher S einen mentalen Zustand M zuschreibt, ist aufrichtig genau dann, wenn S zum Zeitpunkt der Äußerung *nicht* glaubt, *nicht* in M ist.

Die doppelte Negation mag hier zunächst irritieren. Was könnte dies anderes bedeuten, als dass S glauben muss, in M zu sein? Ich werde im Folgenden drei Gründe anführen, warum es sich bei dieser indirekten Definition um mehr als nur eine logische Spielerei handelt.

Der erste Grund, der dafür spricht, die Aufrichtigkeit bei Selbstbekundungen in dieser indirekten Weise zu definieren, ist ein empirischer. Wie ich schon erwähnte, müssen wir auch elementare Selbstbekundungen berücksichtigen, die Kinder in einer Phase ihrer kognitiven Entwicklung äußern, in der sie eine Vielfalt von Empfindungen haben, aber noch kein reflexives Selbstbewusstsein. Dass es diese Phase gibt, lässt sich auch mit einer ‘higher-order’ Theorie des Bewusstseins vereinbaren, wie Rosenthal sie vertritt. Denn gemäß dieser Theorie sind Empfindungen zwar erst bewußt, wenn es zu einer Selbstzuschreibung in Form eines höherstufigen Gedankens kommt, aber solche Gedanken sind von einem reflexiven Selbstbewusstsein – oder introspektivem Selbstbewusstsein, wie Rosenthal sagt – zu unterscheiden (siehe Rosenthal 2005, 47ff). Auch darauf brauche ich hier nicht weiter einzugehen. Entscheidend ist die Frage, ob Kinder, die noch kein reflexives Selbstbewusstsein haben, bei ihren Selbstbekundungen aufrichtig oder unaufrichtig sein können. Von dieser Frage hängt es ab, ob man ihnen auch schon eine Autorität der Ersten Person zugestehen kann oder nicht. Damit sich diese Frage überhaupt sinnvoll stellen läßt, muss man den Begriff der Aufrichtigkeit als einen indirekten Begriff definieren.<sup>9</sup>

Auch das vorhin beschriebene Phänomen der Schmerz-Angst kann hier als Argument dienen. Ich habe schon erwähnt, dass man einem Subjekt, das unter Schmerzangst leidet, kaum absprechen kann, aufrichtig zu sein, wenn es behauptet, Schmerzen zu empfinden. Unklar ist hingegen, ob man sagen soll, dass der Patient fälschlich *glaubt*, Schmerzen zu empfinden. Wie soll man entscheiden, ob eine solche Überzeugung vorliegt oder nicht? Die indirekte Definition der Aufrichtigkeit erspart uns diese Schwierigkeit. Man kann mit Sicherheit davon ausgehen, dass ein Patient in dieser Situation jedenfalls nicht davon überzeugt ist, *keine* Schmerzen zu haben. Also ist seine Selbstbekundung aufrichtig, im Unterschied zu einem Schwindler, der weiß (oder jedenfalls glaubt), keine Schmerzen zu haben, dies aber dennoch behauptet.<sup>10</sup>

---

<sup>9</sup> Auch der schon angesprochene Unterschied zwischen der Perspektive der Ersten Person und der Perspektive der Dritten Person läßt sich hier ins Treffen führen. Damit ein Bericht aus der Dritten Person Perspektive aufrichtig sein kann, muss derjenige, der diesen Bericht gibt, glauben, dass sein Bericht wahr ist. Wer dagegen aus der Perspektive der Ersten Person bekundet, dass er z.B. Schmerzen empfindet, muss nicht an die Wahrheit seiner Aussage glauben, um aufrichtig zu sein. Er darf bloss nicht das Gegenteil dessen glauben, was er bekundet.

<sup>10</sup> Interessant ist in diesem Zusammenhang das als-ob Spiel von Kindern. Dass Kinder dabei nicht unaufrichtig sind, könnte man zunächst so erklären: Im Rahmen eines als-ob-Spiels kann jemand sagen, er hätte Schmerzen, während er davon überzeugt ist, keine Schmerzen zu haben, weil der

Als dritten, und entscheidenden Grund zugunsten eines indirekten Verständnisses von Aufrichtigkeit möchte ich anführen, dass wir jetzt das Problem falscher Selbstbekundungen anders lösen können, als Rosenthal dies vorgeschlagen hat. Für ihn sind falsche Selbstbekundungen, wie wir gesehen haben, einfach Behauptungen, die eine falsche Überzeugung des Sprechers zum Ausdruck bringen. Es sind damit genauso echte Selbstbekundungen, wie falsche Behauptungen echte Behauptungen sind. Meine Intuition sagt hier etwas anderes. Wer Schmerzen auszudrücken versucht, die er gar nicht hat, dessen Aussage würde ich nicht als eine echte Selbstbekundung bezeichnen, und zwar auch dann nicht, wenn sie aufrichtig ist. Meines Erachtens sollten wir den Begriff der echten Selbstbekundung deshalb so zu definieren:

DEF3 Die Selbstbekundung eines Sprechers ist eine *echte* Selbstbekundung genau dann wenn sie *aufrichtig* und *wahr* ist.

Falsche Selbstbekundungen sind also gar keine echten Selbstbekundungen, sondern “defizitär”, wie man sagen könnte. Sie erfüllen nicht wirklich die Ausdrucksfunktion, die sie zu erfüllen vorgeben. Sie bringen nicht jene Zustände zum Ausdruck, die sich ein Sprecher dabei selbst zuschreibt. Diese Doppelfunktion haben nur echte Selbstbekundungen. Dies war der Vorbehalt, den ich am Ende des vorigen Abschnitts erwähnt habe: Die These (DEX) muss auf echte Selbstbekundungen eingeschränkt werden.

Für den Begriff der Autorität der Ersten Person gilt diese Einschränkung nicht. Dass eine Selbstbekundung mit der Autorität der Ersten Person geäußert wird, bedeutet ja nur, wie wir in Abschnitt 1 gesehen haben, dass ihre Äußerung durch einen kompetenten Sprecher ein *prima facie* Grund ist, dass es sich so verhält, wie der Sprecher sagt. Diese Bedingung ist auch dann erfüllt, wenn sich herausstellt, dass es eine falsche Selbstbekundung war.

## 6. Autorität und reflexives Selbstbewusstsein

---

Kontext des Spiels dafür sorgt, dass es zu keiner Täuschung anderer kommt. Versteht man Aufrichtigkeit hingegen in dem schwächeren Sinn von DEF2, bietet sich als Alternative folgende Erklärung an: Kinder glauben nicht, dass sie Schmerzen haben, wenn sie in einem als-ob-Spiel sagen, sie hätten Schmerzen. Da sie aber auch nicht glauben, keine Schmerzen zu haben, können sie so tun, als seien sie aufrichtig.

Ich komme damit zur Ausgangsfrage dieses Aufsatzes zurück: Welche Rolle spielt das reflexive Selbstbewusstsein für die Entwicklung der Autorität der Ersten Person? Dabei setze ich nun voraus, dass wir die Autorität der Ersten Person als ein sprachliches Phänomen im Sinne von DEF1 verstehen. Weiters setze ich voraus, dass die Erklärung, woher diese Autorität rührt, aus der These des Deskriptiven Expressivismus (DEX) folgt. Vor diesem Hintergrund stellt sich nun folgende vierte Frage:

(Frage 4) Warum müssen Sprecher über ein reflektives Selbstbewusstsein verfügen, damit sie Selbstbekundungen mit der Autorität der Ersten Person äußern können?

Gemäß der hier vertretenen Auffassung ergibt sich die Antwort auf diese Frage aus der spezifischen expressiven Kompetenz, die mit dem Gebrauch von Selbstbekundungen verknüpft ist. Sie ist gleichsam die gemeinsame Wurzel für beides: das reflexive Selbstbewusstsein und die Autorität der Ersten Person. Zu zeigen gilt es daher, dass die expressive Kompetenz, die hinreichend ist, um einer Person ein reflexives Selbstbewusstsein zuzuschreiben, auch hindreichend ist, um ihr die Autorität der Ersten Person zuzuschreiben, und umgekehrt. Erst wenn wir beides als wesentlich zusammengehörig begreifen – das reflexive Selbstbewusstsein und die Autorität der Ersten Person – wird klar, welche Rolle den expressiven Fähigkeiten hier zukommt: Sie erfüllen genau die Funktion, die in der traditionellen Erkenntnistheorie das unmittelbare Selbstwissen spielt.

Kehren wir an diesem Punkt nochmals zu den früher besprochenen Überlegungen von Davidson und Wittgenstein zurück, um zu sehen, wo wir im Vergleich dazu jetzt stehen. Davidson hat folgende Entscheidungsfrage in den Raum gestellt: Soll man die Autorität der Ersten Person “mit Blick auf die Sprache” oder “mit Blick auf die Erkenntnistheorie” erörtern? Die Schwierigkeit für Davidson bestand darin zu begründen, warum man nicht versuchen sollte, so wie es viele getan haben, die Cartesische Annahme einer unmittelbaren Bekanntschaft mit den eigenen Bewusstseinsinhalten oder das empiristische Modell von Introspektion weiter zu entwickeln. Nun sehen wir, woher diese Schwierigkeit rührt. Sie kommt daher, dass Davidson nicht erklären kann, weshalb es diese Autorität nicht ohne reflexives Selbstbewusstsein geben kann. Eine solche Erklärung lässt sich jedoch finden, wenn wir die expressive Funktion von Selbstbekundungen hinzunehmen.

Betrachten wir daher nochmals, was Wittgenstein über das Lernen von Empfindungswörtern sagte. Wittgenstein meint, Kinder lernen Ausdrücke wie ‘Schmerz’

einfach dadurch, dass sie das Äußern solcher Worte ihrem Verhaltensrepertoire hinzufügen. Sie sagen ‘Schmerz’ oder ‘Mir tut etwas weh’ anstatt zu Schreien. Wie wir gesehen haben, wollte Wittgenstein damit keine behaviouristische These in dem Sinn vertreten, dass das Wort ‘Schmerzen’ keine Empfindung, sondern nur das Schreien bezeichnet. So weit so gut. Aber es könnte eine andere behaviouristische These hinter dieser Bemerkung stecken. Vielleicht dachte Wittgenstein, es sei ein einfacher Konditionierungsprozess, der diesem Lernprozess zugrundeliegt: Ein Kind lernt, ‘Schmerz’ zu sagen, damit es getröstet wird; ‘Durst’ zu sagen, damit es zu trinken bekommt; ‘Hunger’ zu sagen, damit es zu essen bekommt, usw.<sup>11</sup>

Die exegetische Frage, ob es Wittgenstein so gemeint hat oder nicht, lasse ich wieder beiseite. Die relevante Frage ist: Warum ist der Sprachlernprozess kein solcher Konditionierungsvorgang? Das liegt daran, dass Kinder das mentale Vokabular *nicht nur* dadurch lernen, dass sie Selbstbekundungen von sich geben. Dieser Lernprozess ist Teil einer komplexen sozialen Interaktion, die reziprok verläuft: Kinder lernen gleichzeitig ihren Empfindungen Ausdruck zu verleihen und entsprechende Äußerungen anderer Personen zu verstehen. Auf diese Weise begreifen sie nicht nur, was jemand möchte, der ‘Au weh!’ oder ‘Durst’ sagt. Sie verstehen auch, wie es zu solchen Äußerungen kommt, und sie können schon mit zwei Jahren empathisch mit anderen Personen mitfühlen (siehe Bischof-Köhler 2011).

Als sprachlich kompetent betrachten wir ein Kind erst, wenn es die Doppelrolle als Sprecher *und* als Interpret anderer Personen hinreichend beherrscht. Ohne diese Kompetenz gibt es keine Autorität der Ersten Person, wie Definition DEF1 deutlich macht. Selbst einem noch so gut konditionierten Papagei würden wir nicht zubilligen, dass er seine Empfindungen mit der Autorität der Ersten Person äußern kann. Er sagt vielleicht im richtigen Augenblick ‘Ich habe Durst’, aber er versteht nicht, was er damit sagt, jedenfalls nicht so, wie wir es von einem normalen Sprecher erwarten. Der deskriptive Gehalt dieser Aussage erschließt sich erst in der sozialen Interaktion mit anderen Sprachbenützern.

Ein letztes Argument soll diese Schlussfolgerung noch untermauern: Wäre es denkbar, dass jemand den Sinn mentaler Ausdrücke zu verstehen lernt, ohne sich seiner eigenen Empfindungen, Gefühle oder Gedanken reflexiv bewusst zu sein? Dies können wir

---

<sup>11</sup> Eine Variante eines solchen Konditionierungsprozesses sind ‘Aufstiegsroutinen’, wie sie R. Gordon mit folgendem Beispiel veranschaulicht. “There is a [...] way of training children to use the form of a self-ascription of belief. It would be possible to take a child – say, a two-year old – who can make just a few simple assertions, like ‘It’s raining’, and in one easy step train her to sound *very* sophisticated. Just get her to preface her assertion, optionally, with ‘I believe’. She’ll then say, for example, ‘I believe it is raining’. (Gordon (1995), 61). Solche Routinen führen fürs erste nur zu ‘uncomprehending ascriptions’, so Gordon: “They will not have learned that they believe it is raining, or even that they have beliefs. As far as they are concerned, they are just parroting a formula before saying what they *really* mean to say, namely, that it’s raining.” (ibid.).

ausschließen. Ebenso können wir aber auch ausschließen, wie ich argumentiert habe, dass er diese Ausdrücke versteht, wenn sie von anderen Personen geäußert werden, ohne dass er sie selbst verwendet und dabei die Autorität der Ersten Person für sich in Anspruch nimmt. Dies widerspräche der interaktiven Art und Weise, wie mentale Ausdrücke gelernt werden. Es ist nicht vorstellbar, dass ein Sprecher im Zuge des Spracherwerbs nur begreift, dass andere gewisse Empfindungen, Gedanken und Gefühle haben, ohne auch zu begreifen, dass er selbst ähnliche Empfindungen, Gedanken und Gefühle hat bzw. haben kann. Damit wird diese sprachliche Kompetenz zum Bindeglied von zwei kognitiven Leistungen. Zum einen ist sie eine hinreichende Bedingung dafür, dass Sprecher ihre mentalen Zustände mit der Autorität der Ersten Person zum Ausdruck bringen können; zum anderen ist sie auch eine hinreichende Bedingung dafür, dass Subjekte ein reflexives Selbstbewusstsein erwerben.

An keiner Stelle müssen wir hier von einem unmittelbaren Selbstwissen reden. Daher möchte ich am Ende ein berühmtes Bild von Wittgenstein abwandeln, um die kritische Seite dieses Ergebnisses hervorzuheben. Der Käfer in der Schachtel, den nur ich sehen kann, gehört nicht zum Sprachspiel, sagt Wittgenstein. Das Ding in der Schachtel fällt als “irrelevant aus der Betrachtung heraus.” (PU §293). Was könnte dieses “Ding” sein, durch das “gekürzt” werden kann? Sicher nicht die Empfindung, und auch nicht das Wissen, dass man etwas empfindet. Was man “wegkürzen” kann, ist das *unmittelbare Selbstwissen*, also die traditionellen Vorstellungen von Introspektion und privilegierten Zugang zum eigenen Bewusstsein.<sup>12</sup>

## Literatur

Bar-On, Dorit (2004), *Speaking My Own Mind. Expression and Self-Knowledge* (Oxford: Clarendon Press).

——— (2010), ‘Avowals: Expression, Security, and Knowledge: Reply to Matthew Boyle, David Rosenthal, and Maura Tumulty’, *Acta Analytica*, 25, 47-63.

Bar-On, Dorit and Long, Peter (2001), ‘Avowals and first-person privilege’, *Philosophy and Phenomenological Research*, 62, 311-35.

Bermúdez, José Luis (1998), *The Paradox of Self-Consciousness* (Cambridge, MA: The MIT Press).

Bischof-Köhler, Doris (1991), ‘The development of empathy in infants.’, in M.E. Lamb and H.

Keller (eds.), *Infant Development: Perspectives from German-speaking countries* (Hillsdale:

---

<sup>12</sup> Für seine wertvollen Kommentare zu früheren Fassungen dieses Aufsatzes möchte ich Frank Esken herzlich danken.



- Lawrence Erlbaum), 245-73.
- Carruthers, Peter (2011), *The Opacity of Mind. An Integrative Theory of Self-Knowledge* (Oxford: Oxford University Press).
- Davdison, Donald (1984), 'First Person Authority', *Dialectica*, 38, 101-11 (dt. in: A. Newen, G. Vosgerau (Hrsg.): *Den eigenen Geist kennen. Selbstwissen, privilegierter Zugang und die Autorität der ersten Person*. Mentis: Paderborn 2005).
- (1988), 'The Myth of the Subjective', in M. Benedikt and R. Burger (eds.), *Bewusstsein, Sprache und die Kunst* (Wien: Edition S. Verlag), (Nachdruck in: D. Davidson: *Subjective, Intersubjective, Objective*, Oxford: Clarendon Press 2001, 39-52).
- (1993), 'Reply to Bernhard Thöle', in R. Stoecker (ed.), *Reflecting Davidson*. (Berlin: De Gruyter), 248-50.
- Falvey, Kevin (2000), 'The Basis of First Person Authority', *Philosophical Topics*, 28, 69-99.
- Finkelstein, David H. (2003), *Expression and the Inner* (Cambridge, Mass.: Harvard University Press).
- Gallois, André (1996), *The World Without, the Mind Within* (Cambridge: Cambridge University Press).
- Gasking, Douglas (1966), 'Avowals', in R. J. Butler (ed.), *Analytic Philosophy* (Oxford: Basil Blackwell), 154-69.
- Geach, Peter (1965), 'Assertion', *Philosophical Review*, 84, 449-65.
- Gertler, Brie (2011), *Self-knowledge* (London: Routledge).
- Gordon, Robert M. (1995), 'Simulation without introspection or inference from me to you', in Martin Davies and Tony Stone (eds.), *Mental Simulation. Evaluations and Applications* (Oxford: Blackwell), 53-67.
- Green, Mitchell S. (2007), *Self-Expression* (Oxford: Clarendon Press).
- Hatzimoysis, Anthony (ed.), (2011), *Self-Knowledge* (Oxford: Oxford University Press).
- Newen, Albert (2005), 'Warum genießen wir eine Autorität der ersten Person?', in A. Newen and G. Vosgerau (eds.), *Den eigenen Geist kennen. Selbstwissen, privilegierter Zugang und die Autorität der ersten Person* (Paderborn: Mentis), 165-83.
- Rorty, Richard (1970), 'Incorrigibility as the mark of the mental', *Journal of Philosophy*, 12, 399-424.
- Rosenthal, David M. (2005), *Consciousness and Mind* (Oxford: Clarendon Press).
- (2010), 'The mind and its expression', *Self, Language, and World* (Atascadero: Ridgeview), 107-26.
- Shoemaker, Sydney (1968), 'Self-reference and self-awareness', *Journal of Philosophy*, 65, 555-67.

- Smithies, Declan and Stoljar, Daniel (eds.) (2012), *Introspection and Consciousness* (Oxford: Oxford University Press).
- Thöle, Bernhard (1993), 'The Explanation of First Person Authority', in R. Stoecker (ed.), *Reflecting Davidson*. (Berlin: De Gruyter), 213-47.
- Vosgerau, Gottfried (2009), *Mental Representation and Self-Consciousness* (Paderborn: Mentis).
- Wittgenstein, Ludwig (1953), *Philosophische Untersuchungen* (Werkausgabe, Bd 1, Frankfurt: Suhrkamp (= PU)).
- Wright, Crispin (1998), 'Self-Knowledge: The Wittgensteinian Legacy', in Crispin Wright, Barry C. Smith, and Cynthia MacDonald (eds.), *Knowing Our Own Minds* (Oxford: Clarendon Press), 13-45.