

Moralische Gefühle aus der Sicht der kognitiv-motivationalen Theorie der Emotion

Rainer Reisenzein

Endfassung: 1.3. 2010

1. Einleitung

Soziale und moralische Normen werden sowohl in der Philosophie als auch der Psychologie traditionell mit bestimmten Emotionen in Verbindung gebracht, die oft als „moralische Gefühle“ bezeichnet werden. Beispiele für moralische Gefühle sind Schuld, Empörung und moralische Befriedigung; aber auch das Mitleid wird sehr oft zu diesen Emotionen gerechnet. In diesem Aufsatz gehe ich den Fragen nach, wie die moralischen Gefühle entstehen und was sie ihrer Natur nach sind. Die Klärung dieser Fragen erfolgt auf dem Hintergrund eines komputationalen Modells der Emotion (Reisenzein, 2009a; 2009b; vgl. auch Reisenzein, 2001). Deshalb werde ich im Abschnitt 2 des Aufsatzes zunächst dieses Emotionsmodell vorstellen. Im Abschnitt 3 werde ich dann exemplarisch am Beispiel von Mitleid und Schuld untersuchen, wie die vorgestellte Theorie die moralischen Gefühle erklärt. Abschließend benenne ich einige Fragen für die zukünftige Forschung.

2. Die kognitiv-motivationale Theorie der Emotion und ihre Naturalisierung¹

2. 1 Die kognitiv-motivationale Theorie der Emotion

Ausgangspunkt der in Reisenzein (2009a; 2009b) skizzierten komputationalen Emotionstheorie ist die *kognitiv-motivationale* oder *Glauben-Wunsch-Theorie der Emotion* (englisch: *belief-desire theory of emotion*, im folgenden: BDTE). Die BDTE gehört zur Klasse der kognitiven Emotionstheorien, welche die psychologische und philosophische Diskussion über Emotionen der letzten 30 Jahre dominiert haben. Kognitive Emotionstheorien innerhalb der Psychologie vertreten z. B. Arnold (1960), Frijda (1986), Lazarus (1991), Ortony, Clore und Collins (1988) und Scherer (2001); in der Philosophie z. B. Kenny (1963), Lyons (1980), Nussbaum (2001) und Solomon (1976). Als eigenständige Variante des kognitiven Ansatzes in der Emotionsforschung ist die kognitiv-motivationale Theorie der Emotion primär von Philosophen propagiert worden (siehe insbes. Davis, 1981; Green, 1992; Marks, 1982; Searle, 1983). Eine frühe

1 Die Ausführungen in diesem Abschnitt beruhen zum Großteil auf Reisenzein (2009a; b).

Version der BDTE findet sich bei Alexius Meinong (1894; zusammenfassend Reisenzein, Meyer & Schützwohl, 2003; Reisenzein, 2006a).

Die Mehrheit der existierenden kognitiven Emotionstheorien sind *kognitiv-evaluative* Theorien: Sie postulieren, dass für Emotionen eine bestimmte Art von Kognitionen notwendige Voraussetzung ist, nämlich evaluative Kognitionen oder Bewertungen (englisch *appraisals*). In ihrer paradigmatischen Form sind diese evaluativen Kognitionen Wertüberzeugungen, wie z. B. die Überzeugung, dass ein Ereignis gut oder schlecht, gefährlich oder frustrierend ist. Im Gegensatz dazu ist die BDTE eine *kognitiv-motivationale* Theorie: Sie postuliert, dass Emotionen nicht nur auf Überzeugungen bzw. Glaubensannahmen (also kognitiven oder informational Zuständen) beruhen, sondern auch auf Wünschen (also motivationalen Zuständen).

Um den Unterschied zwischen den beiden Versionen der kognitiven Emotionstheorie zu verdeutlichen, betrachten wir den Fall von Maria, die sich darüber freut, *dass Schroiber zum Kanzler gewählt wurde*. Ich bezeichne diesen Sachverhalt abkürzend als *p*. Nach der kognitiv-evaluativen Theorie der Emotion freut sich Maria über *p*, wenn sie (1) den (festen) Glauben erwirbt, dass *p* besteht und (2) *p* als gut für sich bewertet, d. h., glaubt, dass *p* gut für sie ist. Nach der kognitiv-motivationalen Theorie der Emotion freut sich Maria dagegen dann über den Sachverhalt *p*, wenn sie (1) den Glauben erwirbt, dass *p* besteht und wenn sie (2) sich *p* wünscht (Abbildung 1). Beide Versionen der kognitiven Emotionstheorie nehmen also an, dass eine bestimmte Tatsachenüberzeugung – die Überzeugung, dass *p* der Fall ist – eine notwendige Voraussetzung der Freude Marias über *p* ist. Die Theorien unterscheiden sich jedoch in Bezug auf die Frage, was zu dieser Tatsachenüberzeugung hinzukommen muss, damit Maria Freude über *p* erlebt. Nach der evaluativen Theorie der Emotion ist der zusätzlich notwendige Faktor eine weitere Überzeugung Marias, jedoch eine mit wertendem Inhalt, also eine Wertüberzeugung; nämlich Marias Überzeugung, dass *p* gut für sie ist. Nach der motivationalen Theorie ist der zusätzliche Faktor dagegen ein motivationaler Zustand, nämlich Marias Wunsch nach *p*.

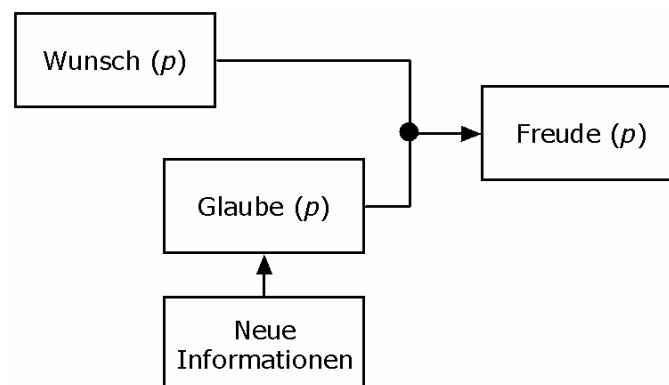


Abb.1.: Entstehung von Freude nach der kognitiv-motivationalen Theorie der Emotion

Die Unterscheidung zwischen der evaluativen und der motivationalen Version der kognitiven Emotionstheorie ist allerdings nicht ganz scharf. Denn viele Vertreter der kognitiven Emotionstheorie, insbesondere in der Psychologie – wo sie *Bewertungs- oder Einschätzungstheoretiker (appraisal theorists)* heißen – nehmen an, dass die Bewertung von Ereignissen durch eine Person die Relevanz dieser Ereignisse für ihre Wünsche oder Motive repräsentiert. „*p* ist gut für mich“ bedeutet nach dieser Auffassung also so viel wie „*p* ist geeignet, meine Wünsche zu erfüllen“ oder „*p* stimmt mit dem überein, was ich mir wünsche.“ Entsprechend bedeutet „*p* ist schlecht für mich“ so viel wie „*p* frustriert meine Wünsche“, „*p* widerspricht dem, was ich mir wünsche.“ Aber auch dann, wenn man dies berücksichtigt, bleiben wichtige Unterschiede zwischen der Bewertungstheorie und der BDTE bestehen. Der wichtigste Unterschied ist folgender: Nach der Bewertungstheorie wird die Beziehung zwischen den Wünschen der Person und ihren Emotionen durch Wertüberzeugungen vermittelt²; nach der BDTE beruhen die Emotionen dagegen *direkt* auf Wünschen und Überzeugungen (Green, 1992; Reisenzein, 2009b; vgl. auch Castelfranchi und Miceli, 2009).

Durch diese Annahme entgeht die BDTE einem zentralen empirischen Einwand gegen die Bewertungstheorie: Nämlich dem Einwand, dass faktische und evaluative Kognitionen für Emotionen *nicht hinreichen*. In der Tat rufen die Gedanken, dass *p* besteht, und dass *p* gut für einen ist, durchaus nicht immer Freude hervor. Beispielsweise ist einem die eigene Gesundheit zumeist viel wert; der Gedanke daran, dass man gesund ist (Tatsachenurteil), und dass dies gut sei (Werturteil), löst aber nur unter besonderen Bedingungen Freude aus – etwa wenn man erst kürzlich von einer Krankheit genesen ist. Die kognitiv-motivationale Emotionstheorie kann dies erklären: Entscheidend für Freude über *p* ist nicht die Überzeugung, dass *p* gut für einen ist, sondern der Wunsch nach *p*. Was man für gut hält, muss man sich jedoch nicht unbedingt wünschen. Insbesondere wünscht man sich keine Sachverhalte, die man bereits für bestehend hält; denn wie Meinong (1917) sagt, werden Wünsche durch ihre Erfüllung ausgelöscht. Deshalb löst der Gedanke daran, dass man gesund ist und dass dies gut für einen ist, in der Regel kein Freudegefühl aus: Der Wunsch, gesund zu sein, ist zum Zeitpunkt dieses Gedankens meist längst erfüllt.

2 Es gibt auch Bewertungstheoretiker, die die von ihnen postulierten emotionsrelevanten Bewertungen *nicht* als Wertüberzeugungen verstanden wissen möchten, zumindest nicht für alle Emotionen. Beispielsweise postuliert Arnold (1960), dass auch sinnliche Lust- und Unlustgefühle (z. B. das angenehme Gefühl, das durch den Duft einer Rose hervorgerufen wird) durch Bewertungen verursacht werden; dabei kann es sich aber kaum um Wertüberzeugungen im üblichen Sinn handeln (Meinong, 1894). In der Tat betont Arnold (1960) selbst, dass sie den Begriff der Bewertung in diesem Fall in einer technisch erweiterten Bedeutung verwendet, die hypothetisch angenommene, unbewusste (und wahrscheinlich nichtpropositionale) bewertungsanaloge Prozesse mit umfassen soll. Eine ähnliche Position scheint Scherer (2001) zu vertreten. Zu den Problemen dieser Position siehe Reisenzein (2006a).

Die Annahme, dass Personen zum Erleben von Freude Wertkognitionen der beschriebenen Art bilden müssen, ist aber auch aus einem anderen Grund problematisch. Denn diese Wertüberzeugungen, wie z. B. „der Sachverhalt p stimmt mit meinen Wünschen überein“ sind *Metakognitionen*: Es sind Kognitionen über (unter anderem) die eigenen repräsentationalen Zustände, nämlich die eigenen Wünsche. Die kognitiv-evaluative Theorie der Emotionen scheint also zu implizieren, dass das Erleben von Freude oder Leid über einen Sachverhalt p eine Überzeugung über die eigenen Wünsche bezüglich p voraussetzt. Für die Entstehung von Gefühlen wie Überraschung und Enttäuschung über p scheint zudem eine Überzeugung über die eigenen Überzeugungen bezüglich p nötig zu sein („ p widerspricht dem, was ich erwartet habe“). Diese Annahme widerspricht der Introspektion und ist aus komputationaler, evolutionspsychologischer und entwicklungspsychologischer Perspektive unplausibel (Reisenzein, 2009a). So gibt es empirische Belege dafür, dass bereits kleine Kinder Gefühle wie Freude und Überraschung erleben; zu Metakognitionen der genannten Art scheinen sie jedoch erst ab etwa dem dritten Lebensjahr fähig zu sein (vgl. dazu Reisenzein, 2009a).

Die BDTE-Theoretiker vermeiden all diese Probleme durch die Annahme, dass Emotionen direkte „Produkte“ von faktischen Überzeugungen und Wünschen sind. Sie unterscheiden sich allerdings untereinander in Bezug auf die Frage, was hier unter „Produkt“ genau zu verstehen ist. So identifiziert Marks (1982) die auf p bezogene Emotion einfach mit dem Glauben-plus-Wunsch nach p . Andere Autoren postulieren dagegen, dass die Emotion aus einer gestaltartigen Integration von Glauben und Wunsch (Green, 1992) oder von Glauben und Wunsch zusammen mit dadurch verursachten Lust- oder Unlustgefühlen (Castelfranchi & Miceli, 2009) hervorgeht. Im Unterschied zu diesen Theorien über die Natur der Emotionen folge ich Meinong (1894; 1906) in der Annahme, dass Überzeugungen und Wünsche die *Ursachen* von Emotionen sind, welche davon verschiedene, eigenständige mentale Zustände sind – nämlich Gefühle der Lust und Unlust, der Überraschung und der „Erwartungsbestätigung“ (mehr dazu im Abschnitt 2.2). Damit vertrete ich eine *kausalistische Version* der BDTE.

Ungeachtet dieses Unterschieds nehmen alle BDTE-Theoretiker explizit oder implizit an, dass die am Beispiel der Freude erläuterte Glauben-Wunsch-Analyse von Emotionen (Abbildung 1) in angemessen modifizierter Form für die meisten in der Umgangssprache unterschiedenen Emotionen zutrifft. Für die hier vertretene kausalistische Version der BDTE kann diese Annahme folgendermaßen präzisiert werden: Alle Emotionen, die auf propositionale Objekte (d. h. Sachverhalte) gerichtet sind, sind letztlich verstehbar als Reaktionen auf die kognizierte (1) tatsächliche oder potentielle Erfüllung (z. B. Freude, Hoffnung) oder Frustration (z. B. Leid, Furcht) von Wünschen sowie (2) in einigen Fällen (z. B. Überraschung, Enttäuschung) die Bestätigung oder Widerlegung

von Überzeugungen (Reisenzein, 2009a; 2009b). Die folgenden Beispiele stützen diese These:

- Maria freut sich darüber, dass p (z. B. dass Schroiber zum Kanzler gewählt wurde), wenn sie sich wünscht, dass p besteht, und nun zu der festen Überzeugung kommt (d. h., sich sicher ist), dass p besteht.
- Maria leidet darunter, dass p , wenn sie nicht möchte, dass p besteht bzw. aversiv gegen p ist (ich analysiere dies im Folgenden vereinfachend als: Maria wünscht sich, dass $\neg p$) und nun zu der festen Überzeugung kommt, dass p besteht.
- Maria hofft, dass p , wenn sie sich wünscht, dass p besteht, sich aber unsicher ist, ob p besteht oder nicht (d. h., mit Unsicherheit glaubt, dass p).
- Maria hat Angst davor, dass p , wenn sie sich wünscht, dass $\neg p$, und unsicher ist, ob p besteht oder nicht.
- Maria ist überrascht darüber, dass p , wenn sie bisher glaubte, dass $\neg p$, und jetzt zu der festen Überzeugung kommt, dass p der Fall ist.
- Maria ist enttäuscht, dass p , wenn sie sich wünscht, dass p besteht, und bislang glaubte, dass p der Fall sei, nun aber zu der festen Überzeugung kommt, dass $\neg p$ der Fall ist.
- Maria ist erleichtert, dass p , wenn sie sich wünscht, dass p , und bislang glaubte, dass $\neg p$, nun aber zur festen Überzeugung kommt, dass p besteht.

Diese Gefühle machen natürlich nur eine kleine Teilmenge der Emotionen aus, die in der Alltagssprache unterschieden werden. Aus der Perspektive der BDTE sind sie allerdings grundlegend in folgendem Sinn: Die meisten anderen Emotionen sind Varianten bzw. spezielle Formen dieser Emotionen, die ihre Existenz der Tatsache verdanken, dass Menschen komplexe Überzeugungen und Wünsche – d. h. Überzeugungen und Wünsche mit komplexen Inhalten – haben. Für die in Abschnitt 3 behandelten moralischen Gefühle ist vor allem bedeutsam, dass Menschen als „ultrasoziale“ Wesen (Richerson & Boyd, 1998) fast zwangsläufig ein Interesse am Schicksal anderer Personen, ihren Handlungen und Einstellungen sowie den Auswirkungen der eigenen Handlungen auf andere Personen haben (vgl. auch den Beitrag von Niedenzu in diesem Band). Diese „sozialen Sachverhalte“ sind deshalb ein bevorzugter Gegenstand der menschlichen Überzeugungen, Wünsche und Emotionen (vgl. auch Heider, 1958).

2.2 Ein computationales Modell der Emotion

Die BDTE ist eine Explikation von alltagspsychologischen Annahmen. Sie ist damit – wie auch die meisten anderen traditionellen Theorien der Psychologie (vgl. Reisenzein & Rudolph, 2008) – auf der „intentionalen Ebene“ der Systemanalyse sensu Dennett (1971) angesiedelt. Das Ziel der modernen kognitiven Psychologie, die sich in der zweiten Hälfte des vorigen Jahrhunderts entwickelt

hat, ist es, die intentionale Analyse durch eine Analyse der psychischen Systeme auf der „Designebene“ (Dennett, 1971) zu ergänzen; d. h. durch eine Analyse auf der Ebene der Kausalmechanismen, die den Phänomenen der intentionalen Ebene zugrunde liegen. Die zentrale Annahme der kognitiven Psychologie bei diesem Unternehmen ist, dass psychische Prozesse analog zu Informationsverarbeitungsprozessen in Computern als Berechnungsprozesse (Komputationen) über internen Repräsentationen verstanden werden können. Das Ziel der kognitiven Psychologie ist dementsprechend, die mentalen Repräsentationen und darauf ablaufende Berechnungsprozesse zu rekonstruieren, die den intentionalen Phänomenen zugrunde liegen. Als Idealziel wird von vielen Kognitionswissenschaftlern die Entwicklung von Theorien psychischer Prozesse in Form eines *komputationalen Modells* angesehen, d. h. einer Nachbildung dieser Prozesse in Form eines Computerprogramms (vgl. Sun, 2008). Nach Ansicht der philosophischen Vertreter des komputationalen Paradigmas belegt die Entwicklung eines empirisch bewährten komputationalen Modells die Naturalisierbarkeit (naturwissenschaftliche Erklärbarkeit) des modellierten psychischen Phänomenbereichs (Sterelny, 1991). In jedem Fall aber ermöglicht ein solches Modell – indem es die Mechanismen und Prozesse spezifiziert, die den Phänomenen der intentionalen Ebene zugrunde liegen – ein besseres Verständnis dieser Phänomene (vgl. auch Cummins, 1983): Wir verstehen jetzt im Detail, wie sie zustande kommen, nämlich als Resultate oder Manifestationen von bestimmten Berechnungen auf internen Repräsentationen.

Es ist meine Überzeugung, dass auch das Verständnis der Emotionen durch eine komputationale Analyse wesentlich vertieft werden kann. Noch mehr: Meiner Meinung nach lassen sich einige zentrale Fragen der Emotionspsychologie überhaupt *nur* auf der Ebene der kognitiven Architektur – der Ebene der Repräsentationen und Algorithmen – beantworten. Dazu gehören z. B. die Fragen, welche Prozesse zwischen Wünschen und Überzeugungen einerseits und Emotionen andererseits vermitteln; was Emotionen selbst eigentlich sind; und wozu Emotionen im psychischen System gut sind. Um diese Fragen zu beantworten und allgemeiner, um besser zu verstehen, welche Rolle Emotionen im kognitiven System spielen und was sie ihrer Natur und Funktion nach sind, muss man die Ebene des Erlebens und des Verhaltens verlassen und auf die diesen Oberflächenphänomenen zugrunde liegende Ebene der Systemarchitektur „hinabsteigen.“ Diese Überlegung war es, die mich zu dem im Folgenden beschriebenen Vorschlag für eine komputationale „Unterfütterung“ der kognitiv-motivationalen Emotionstheorie (*computational belief-desire theory of emotion*, CBDTE) motiviert hat (ausführlicher dazu Reisenzein, 2009a; 2009b).

Ausgangspunkt der CBDTE war die Frage: Welche kognitive Architektur wäre geeignet, die in der BDTE beschriebenen Zustände und Prozesse zu realisieren? Da die BDTE Emotionen als Produkte von Überzeugungen und Wünschen auffasst, lässt sich diese Frage am besten beantworten, indem man sich

zunächst überlegt, welche kognitive Architektur für Überzeugungen und Wünsche geeignet ist. Glücklicherweise gibt es bereits eine plausible und transparente komputationale Analyse von Überzeugungen und Wünschen (Fodor, 1987). Nach Fodor enthalten die Repräsentationssysteme des Menschen (neben anderen möglichen Repräsentationsformen) im Kern ein propositionales Repräsentationssystem. Damit ist ein sprachähnliches Repräsentationssystem gemeint, das allerdings nicht mit einer natürlichen Sprache identisch ist. Es handelt sich vielmehr um eine „Sprache des Denkens“ (*language of thought*; Fodor, 1987). Ein propositionales Repräsentationssystem bietet sich für die komputationale Modellierung von Überzeugungen und Wünschen unmittelbar an; denn die intentionalen Objekte von Überzeugungen und Wünschen sind nach allgemeinem Verständnis Propositionen (Sachverhalte), die durch Sätze beschrieben werden.³ Ausgehend von der Annahme, dass psychische Prozesse symbolverarbeitende Prozesse sind, schlägt Fodor vor: Die mentalen Zustände des Glaubens und des Wünschens sind spezielle Verarbeitungszustände von propositionalen Repräsentationen, d. h. von Sätzen in der Sprache des Denkens (vgl. auch Anderson & Lebiere, 1998; Aydede, 2004; Gratch & Marsella, 2004). Die unterschiedlichen Verarbeitungszustände der Repräsentationen in einem kognitiven System sind dabei durch ihre jeweilige kausale Rolle im System definiert. Zur Veranschaulichung dieser kausalen Rollen verwendet Fodor die anschauliche Metapher von unterschiedlichen Speichern, in denen sich Repräsentationen befinden können und deren Inhalte vom System unterschiedlich verarbeitet werden. Den Sachverhalt p zu glauben, heißt dann auf der komputationalen Ebene, dass sich ein Exemplar eines Satzes s , der p repräsentiert, momentan in einem speziellen Speicher befindet (der dementsprechend „Glaubensspeicher“ heißt). Zu sagen, dass sich ein Satz im Glaubensspeicher befindet, steht abkürzend dafür, dass dieser Satz im kognitiven System eine bestimmte kausale Rolle spielt – nämlich genau die Rolle, die Sätze spielen, die wir als wahr behandeln: Der Satz wird als gültige Prämisse in Ableitungen verwendet; wir setzen ihn als wahr voraus, wenn wir Handlungen planen; er erzeugt Überraschung, wenn er sich als falsch erweist, usw. z. B. glaubte Maria vor der Wahl, dass Schroiber nicht gewinnen würde. Auf der komputationalen Ebene entspricht dem der Umstand, dass Marias Glaubensspeicher unter anderem den Satz „Schroiber gewinnt die Wahl nicht“ enthält.

Analog bedeutet, sich p zu wünschen, auf der komputationalen Ebene: ein Exemplar eines Satzes, der p repräsentiert, befindet sich in einem anderen speziellen Speicher (dem „Wunschspeicher“). Beispielsweise wünschte sich Maria, dass Schroiber die Wahl gewinnt. Auf der komputationalen Ebene entspricht

3 In der psychologischen Literatur werden meist die mentalen Repräsentationen von Sachverhalten (die Sätze der Sprache des Denkens) und nicht die dadurch repräsentierten Sachverhalte als „Propositionen“ bezeichnet. Ich folge hier dem philosophischen Sprachgebrauch.

dem: Marias Wunschspeicher enthält unter anderem den Satz „Schroiber gewinnt die Wahl.“

Um Emotionen modellieren zu können, müssen jedoch zusätzliche Annahmen gemacht werden. Nach der BDTE erlebt Maria Freude über die Wahl Schroibers zum Kanzler dann, wenn sie sich diesen Sachverhalt p gewünscht hat und nun zum Glauben gelangt, dass p tatsächlich besteht. Um den Prozess des Erwerbs neuer Überzeugungen zu modellieren, nehme ich an, dass neu erworbene Überzeugungen in einem weiteren speziellen Speicher abgelegt werden. Komputational gesprochen, erlebt Maria somit Freude über p wenn, oder kurz nachdem, ein Satz s , der p repräsentiert, in ihrem Speicher für neue Überzeugungen abgelegt wurde.

Aus der komputationalen Perspektive betrachtet wird jedoch sofort klar, dass die *gleichzeitige Anwesenheit* der Überzeugung, dass p , und des Wunsches nach p im kognitiven System nicht ausreicht, damit man Freude über p erlebt. Zusätzlich muss das kognitive System diese beiden Fakten *miteinander in Beziehung setzen*: Es muss erkennen oder feststellen, dass der Inhalt des neu erworbenen Glaubens mit dem Inhalt eines Wunsches identisch ist. Dazu wird ein Mechanismus bzw. eine Informationsverarbeitungsprozedur benötigt, der die Inhalte der neu erworbenen Überzeugungen mit denen der vorhandenen Wünsche vergleicht – also ein „Überzeugung-Wunsch-Vergleicher“ (*belief-desire comparator*, BDC). Paralleles gilt für Marias Überraschung über die Kanzlerwahl Schroibers: Überraschung tritt auf, wenn das kognitive System feststellt, dass der Inhalt von neu erworbenen Überzeugungen dem Inhalt einer bereits vorhandenen Überzeugung widerspricht. Dazu wird ein Mechanismus benötigt, der die Inhalte von neu erworbenen Überzeugungen mit denen der bereits vorhandenen Überzeugungen vergleicht, also ein „Überzeugung-Überzeugung-Vergleicher“ (*belief-belief comparator*, BBC).

Diese zwei Mechanismen vergleichen unermüdlich die Sätze im Speicher für neu erworbene Überzeugungen mit denen im Wunsch- und Glaubensspeicher. Wird dabei eine Übereinstimmung (der Satz s_{neu} ist identisch mit einem der vorhandenen Sätze) oder eine Nichtübereinstimmung (s_{neu} ist identisch mit der Negation eines vorhandenen Satzes) festgestellt, dann erzeugen die Vergleichsmechanismen bzw. Kongruenzdetektoren eine Ausgangsinformation, die dem kognitiven System die Entdeckung einer Übereinstimmung bzw. eines Widerspruchs mitteilt. In unserem Beispiel stellt Marias BBC fest, dass s_{neu} – der mentale Satz, der den Sachverhalt „Schroiber gewinnt die Wahl“ repräsentiert – dem Inhalt einer vorhandenen Überzeugung widerspricht. Dagegen stellt Marias BDC fest, dass s_{neu} mit dem Inhalt eines bestehenden Wunsches übereinstimmt. Marias BBC gibt deshalb die Information aus, dass ein Widerspruch entdeckt wurde, was bedeutet, dass eine der Überzeugungen Marias soeben durch eine neue Information entkräftet wurde. Und Marias BDC gibt die Information aus,

dass eine Übereinstimmung entdeckt wurde, was bedeutet, dass einer der Wünsche Marias soeben erfüllt worden ist.

Nach der CBDTE hat die Ausgangsinformation, die vom BBC und BDC produziert wird, mehrere wichtige funktionale Konsequenzen im kognitiven System. Erstens wird die Aufmerksamkeit automatisch auf die Inhalte der neu erworbenen Überzeugung gelenkt, die zur Übereinstimmung oder zum Konflikt geführt hat (z. B. im Fall Marias den unerwarteten, aber erwünschten Wahlsieg Schroibers). Zweitens findet, ebenfalls automatisch, eine minimale Aktualisierung des Überzeugung-Wunsch-Systems statt: Sätze, die widerlegte Überzeugungen repräsentieren, werden aus dem Glaubensspeicher und Sätze, die die Inhalte erfüllter Wünsche repräsentieren, werden aus dem Wunschspeicher gelöscht. Drittens: BBC und BDC-Outputs, die eine bestimmte Intensitätsschwelle (dazu unten) überschreiten, führen (direkt oder indirekt) zum Erleben von einzigartigen, d. h. für Emotionen spezifischen Gefühlsqualitäten: den Gefühlen der Überraschung oder der „Erwartungsbestätigung“ (BBC) bzw. den Gefühlen von Lust oder Unlust (BDC). Die allgemeine Funktion dieser Gefühle, so nehme ich an, ist die gleiche wie die Funktion aller bewussten Erlebnisse: Nämlich die Funktion, Informationen systemweit zur Verfügung zu stellen und sie damit in die Lage zu versetzen, globale Kontrolle auszuüben (z. B. Baars, 1988; Block, 1995).

Nun zur Frage, wie die postulierten Vergleichsmechanismen implementiert bzw. realisiert sind. Der erste Gedanke, der einem in den Sinn kommen könnte, ist folgender: Es handelt sich um Schlussfolgerungsprozeduren, die prinzipiell von derselben Art sind wie andere Inferenzprozeduren, die auf propositionalen Repräsentationen operieren (z. B. Anderson & Lebiere, 1998). Eine gewisse Besonderheit dieser Prozesse wäre nur, dass sie metakognitive Prozesse sind (insofern sie sich auf Überzeugungen und Wünsche des Erlebenden beziehen). Um diese Idee zu illustrieren: Maria überlegt beim Erwerb der Überzeugung, dass p , wie folgt: „Ich habe mir gewünscht, dass p ; jetzt weiß ich, dass p besteht; damit ist nun einer meiner Wünsche erfüllt.“ Diese Metakognition – der Gedanke „einer meiner Wünsche wurde erfüllt“ – ist dann die unmittelbare Ursache von Marias Freudegefühl über p . Wie ich früher anlässlich der Kritik der Bewertungstheorie der Emotion bereits angedeutet und an anderer Stelle genauer ausgeführt habe, erscheint mir diese Theorie des BBC und BDC höchst unplausibel: Sie stimmt weder mit der Introspektion überein, noch mit entwicklungspsychologischen Daten zum Erwerb von Überzeugungen und Wünschen, noch mit komputationalen und evolutionären Überlegungen (Reisenzein, 2009a). Damit soll selbstverständlich weder geleugnet werden, dass erwachsene Menschen zu Metakognitionen der genannten Art in der Lage sind, noch dass sie solche Kognitionen tatsächlich mehr oder weniger häufig bilden und dass diese ihr Denken und Handeln und ihre Emotionen beeinflussen können. In Frage gestellt wird ausschließlich, dass diese Metakognitionen die proximatn Ursa-

chen (oder sogar Bestandteile; Lazarus, 1991) von Emotionen sind, wie die Bewertungstheorie impliziert.

Ich habe deshalb folgende Alternative vorgeschlagen: Der BBC und der BDC sind keine gewöhnlichen Inferenzprozeduren, sondern spezielle Mechanismen, die im Gehirn „fest verdrahtet“ sind. Genauer gesagt: Der BBC und der BDC sind Bestandteile eines angeborenen Mechanismus, der das Glauben-Wunsch-System überwacht und wenn nötig aktualisiert. Teilweise als Folge dieser besonderen Herkunft weisen die zwei Kongruenzdetektoren – so nehme ich an – einige besondere Merkmale auf. Insbesondere vergleichen sie alle neu erworbenen Überzeugungen *automatisch* (ohne Absicht und vorbewusst) und *parallel* mit den Überzeugungen und Wünschen, die sich momentan im Arbeitsspeicher befinden. Darüber hinaus – das ist für das Verständnis des Emotionserlebens von besonderer Bedeutung – wird die Ausgangsinformation der Vergleichsmechanismen in einem *nichtpropositionalen Format* repräsentiert. D. h., das Ergebnis der Informationsverarbeitung durch den BBC und BDC ist *nicht* ein (weiterer) Satz in der Sprache der Gedanken. Vielmehr geben die Vergleichsmechanismen nichtpropositionale Signale aus, d. h. Signale, die in Art und Intensität variieren, jedoch keine interne Struktur aufweisen (vgl. Oatley & Johnson-Laird, 1987). Diese Signale sind damit einfachen Empfindungen vergleichbar, wie Ton- oder Temperaturempfindungen (Wundt, 1896). Die Signale enthalten Informationen über den Grad der (Un-)Erwartetheit und (Un-)Erwünschtheit des Inhalts von neu erworbenen Überzeugungen; aber sie repräsentieren nicht deren Inhalt selbst.

Zusammengefasst postuliert die CBDTE, dass das Glauben-Wunsch-System über angeborene, festverdrahtete Überwachungs- und Aktualisierungsmechanismen verfügt, den BBC und den BDC. Diese Mechanismen sind sensorischen Wandlern vergleichbar (d. h. den Sinnesorganen für z. B. Farbe, Ton, Berührung oder körperliche Veränderungen); insbesondere sind ihre unmittelbaren Produkte nichtpropositionale Signale. Anstatt jedoch die Welt oder den eigenen Körper zu „erspüren“ (zumindest unmittelbar), erfassen die „internen Sensoren“ den aktuellen Zustand und Zustandsänderungen des zentralen Repräsentationssystems des Menschen – des Glauben-Wunsch-Systems – während dieses sich mit den neuen Informationen auseinandersetzt, die andauernd durch Wahrnehmungs- und Schlussfolgerungsprozesse erzeugt werden. Eine zentrale architektonische Annahme der CBDTE lautet somit, dass es zusätzlich zu den Sinnesorganen, die uns über den Zustand der Welt informieren, und denen, die uns über den Zustand unseres Körpers informieren, auch Sensoren gibt, die unser zentrales Repräsentationssystem überwachen (vgl. auch Clore, 1994). Emotionen entstehen immer dann, wenn diese Sensoren – die Kongruenzdetektoren – eine Übereinstimmung oder einen Widerspruch von neu erworbenen Überzeugungen und existierenden Überzeugungen oder Wünschen feststellen.

Daraus ergibt sich, dass Emotionen eng mit der Aktualisierung des Glaubens-Wunsch-Systems zusammenhängen. Der Zusammenhang könnte in der Tat kaum enger sein: Laut CBDTE sind die Vergleichsmechanismen, die den Kern der Aktualisierungsmechanismen ausmachen – der BBC und der BDC – *zugleich die zwei grundlegenden, emotionserzeugenden Mechanismen*. Dementsprechend postuliert die CBDTE, dass die *evolutionäre Funktion* der Emotionsmechanismen *nicht* in der Lösung von bereichsspezifischen Problemen besteht (wie z. B. Ekman, 1992, oder McDougall, 1908/1960 postulieren). Vielmehr besteht ihre Funktion darin, die bereichsunspezifische Aufgabe zu lösen, Inkongruenzen zwischen neu erworbenen Überzeugungen und vorhandenen Überzeugungen und Wünschen (1) festzustellen und (2) sobald sie erkannt worden sind, das kognitive bzw. kognitiv-motivationale System auf den Umgang mit ihnen allgemein vorzubereiten.

Wie in Reisenzein (2009a; 2009b) genauer ausgeführt wird, ermöglicht das skizzierte komputationale Modell der BDTE die Beantwortung einer ganzen Reihe von klassischen Fragen der Emotionspsychologie. Zunächst einmal erlaubt die CBDTE eine präzise theoretische (d. h. theorie-basierte; Reisenzein, 2007) Definition von Emotionen: Emotionen sind die nichtpropositionalen Signale, die vom Überzeugungs- und Wunsch-Kongruenzdetektor erzeugt und subjektiv als Gefühle erlebt werden. Auf der Grundlage dieser Definition kann der Bereich der Emotionen im Prinzip exakt abgegrenzt werden. Ferner erlaubt die CBDTE im Gegensatz zu anderen evolutionären Emotionstheorien (z. B. Ekman, 1992; McDougall, 1908/1960) eine an einem klaren Kriterium orientierte Abgrenzung von Basisemotionen: Die Menge der Basisemotionen umfasst exakt die unterschiedlichen Outputs der Kongruenzdetektoren. Gleichzeitig liefert die CBDTE jedoch keine Rechtfertigung für eine Unterscheidung zwischen „Basisemotionen“ und „Nichtbasisemotionen.“ Im Gegenteil: *Alle* Emotionen, die von der CBDTE beschrieben werden – wie komplex oder „kulturell bestimmt“ sie in anderen Hinsichten auch sein mögen – sind nach der CBDTE gleichermaßen grundlegend in dem Sinn, dass sie allesamt Produkte des BBC und des BDC sind.

Die CBDTE liefert ferner eine plausible Erklärung der besonderen Erlebensqualität von Emotionen – des „wie-es-sich-anfühlt, eine Emotion zu haben“ (Reisenzein, 2009b; vgl. auch Reisenzein & Döring, 2009). (1) Dass Emotionen überhaupt eine Erlebensqualität haben, kann dadurch erklärt werden, dass sie sinnesempfindungsähnliche, nichtpropositionale Zustände sind oder zumindest einen „sensorischen Kern“ haben. (2) Emotionale Erlebnisse fühlen sich anders an als nichtemotionale Erlebnisse, weil ihr sensorischer Kern durch Mechanismen erzeugt wird, die für Emotionen spezifisch bzw. für sie „reserviert“ sind. (3) Die Intensität einer Emotion (z. B. die Intensität von Marias Freude über die Wahl Schroibers) ist einfach die Intensität des jeweiligen Gefühls, das der BBC oder BDC erzeugt. (4) Verschiedene Emotionen (z. B. Freude und Enttäu-

schung) fühlen sich – zumindest zum Großteil – deshalb im Erleben unterschiedlich an, weil sie verschiedene emotionale Gefühle oder Kombinationen von Gefühlen sind bzw. enthalten.

Was die komplexeren Emotionen, inklusive der im Folgenden besprochenen moralischen Gefühle betrifft, nehme ich an, dass diese sich in ihrer Erlebensqualität nicht grundlegend von anderen Lust- und Unlustgefühlen unterscheiden. Was sie jedoch von diesen anderen Gefühlen unterscheidet, ist ihr besonderer kognitiv-motivationaler Hintergrund und als Folge davon ihr besonderes „formales Objekt“ (Kenny, 1963).

3. Moralische Gefühle in der CBDTE

Die CBDTE nimmt mit der BDTE an, dass die meisten Emotionen Varianten von wenigen Grundformen sind (vgl. Abschnitt 2.1). Die Reichhaltigkeit des emotionalen Erlebens ist daher nicht dem Umstand zu verdanken, dass es zahlreiche separate Emotionsmechanismen gibt, sondern dem Umstand, dass Menschen relativ komplexe Überzeugungen und Wünsche haben. Um diese Annahme im Detail zu belegen, ist es notwendig, zusätzlich zu den im Abschnitt 2 untersuchten Emotionen – Freude, Leid, Angst, Hoffnung, Überraschung, Enttäuschung und Erleichterung – Analysen weiterer Emotionen oder Emotionsgruppen aus der Perspektive der kognitiv-motivationalen Emotionstheorie und spezieller der CBDTE durchzuführen (vgl. auch Castelfranchi & Miceli, 2009). Als Beitrag zu diesem Projekt untersuche ich im Folgenden in exemplarischer Form die moralischen Gefühle.

Wie bereits einleitend erwähnt wird der Ausdruck „moralische Gefühle“ traditionell verwendet, um eine Gruppe von Emotionen zu bezeichnen, die – so wird jedenfalls angenommen – mit sozialen und moralischen Normen eng zusammenhängen. Über diese Bestimmung hinausgehend ist die Begriffsverwendung allerdings uneinheitlich. Einige Autoren klassifizieren die moralischen Emotionen nach ihren Ursachen als diejenigen Gefühle, die prima facie als Reaktionen auf wahrgenommene Verletzungen und Erfüllungen von Normen auftreten, wie z. B. Schuld, Scham Empörung und moralische Befriedigung (z. B. Strawson, 1962; Wildt, 1993). Diese Emotionen kann man als *moralische Gefühle im engeren Sinn* bezeichnen. Andere Autoren bestimmen die moralischen Gefühle dagegen nach ihren Wirkungen als Gefühle, die für das moralische Handeln und Urteilen besonders bedeutsam sind (z. B. de Sousa, 2001; Haidt, 2003; Weiner, 2006; vgl. dazu auch den Beitrag von v. Scheve in diesem Band). Zu diesen *moralischen Gefühlen im weiteren Sinn* gehören nach Auffassung vieler Autoren neben den normbasierte Emotionen (wie Schuld oder Empörung) auch eine Reihe anderer Emotionen, die Normen nicht essentiell voraussetzen scheinen. Insbesondere gehören dazu auch die von Meinong (1894)

so genannten „Sympathie- und Antipathiegefühle“: die emotionalen Reaktionen auf das positive oder negative Schicksal anderer Personen. Die prototypischen Mitglieder dieser „Emotionsfamilie“ (Castelfranchi & Miceli, 2009) sind Mitfreude, Mitleid, Schadenfreude und Neid (Meinong, 1894; Ortony et al., 1988).

Mein Anliegen im Folgenden ist es, anhand einer exemplarischen Analyse je einer Emotion (Mitleid bzw. Schuld) aus diesen zwei Gefühlsgruppen zu untersuchen, ob und wie die moralischen Gefühle mit der CBDTE in Einklang zu bringen sind bzw. wie sie sich im Rahmen dieser Theorie rekonstruieren lassen. Grundlage der folgenden Analyse ist also einerseits die CBDTE; andererseits sind es meine Intuitionen über die untersuchten Emotionen.⁴

3.1 Analyse der Sympathie- und Antipathiegefühle am Beispiel des Mitleids

Wenn die CBDTE zutrifft, dann entstehen die Sympathie- und Antipathiegefühle im Prinzip auf dieselbe Art wie „einfache“ selbstbezogene Gefühle von Freude und Leid. Speziell für Mitleid heißt das: Mitleid ist eine Form von Leid und tritt wie alle Leidgefühle dann auf, wenn der BDC entdeckt, dass der Inhalt einer neu erworbenen Überzeugung dem Inhalt eines Wunsches widerspricht. Nehmen wir zur Illustration an, Marias Arbeitskollege Karl sei entlassen worden (F), und Maria erlebt nun Mitleid mit Karl (o) über diesen Sachverhalt Fo . Der Inhalt p der neu erworbenen Überzeugung Marias ist hier der Sachverhalt Fo , dass Karl entlassen wurde. Der Inhalt des Wunsches, der durch Fo frustriert wird, ist dementsprechend $\neg Fo$, der Sachverhalt, dass Karl nicht entlassen wird bzw. dass er weiterbeschäftigt wird. Die unmittelbaren Ursachen von Marias Mitleid mit Karl sind somit ihr Glaube, dass Karl entlassen wurde, und ihr Wunsch, dass Karl nicht entlassen wird. Marias BDC stellt fest, dass der Inhalt der neu erworbenen Überzeugung (Fo) dem Inhalt eines ihrer Wünsche ($\neg Fo$) widerspricht; daraufhin erzeugt der BDC ein Signal, das diese Wunschinkongruenz signalisiert und von Maria als Unlustgefühl erlebt wird. Dieses Gefühl ist nach der CBDTE, wie alle Gefühle, für sich genommen nicht objektgerichtet bzw. hat keinen propositionalen Gehalt. Ich nehme jedoch an, dass es durch nachfolgende kognitive Prozesse an das propositionale Objekt Fo „angebunden“ werden kann und dadurch zum *subjektiven Eindruck* führt, das Gefühl würde sich auf dieses Objekt beziehen (genaueres dazu in Reisenzein, 2009a, 2009b). Es *erscheint* Maria dann subjektiv so, als ob sie Leid *über* $\neg Fo$ empfindet. Dieses Verständnis vorausgesetzt, kann auch im Rahmen der CBDTE problem-

4 Ich fasse diese Intuitionen als Resultat der mentalen Simulation von hypothetischen Situationen auf und schreibe ihnen empirischen Stellenwert zu. Es handelt sich deshalb im Folgenden nicht um eine Sprachanalyse, auch wenn ich Emotionswörter der Umgangssprache (wie „Mitleid“ und „Schuld“) als Einstiegspunkte in die Analyse verwende (vgl. dazu auch Castelfranchi & Miceli, 2009; Ortony et al., 1988).

los vom intentionalen Objekt des Mitleids und anderer Emotionen gesprochen werden.

Die bisherige Analyse reicht aber nicht aus, um Mitleid als eigenständige emotionale Qualität – als eigenständige Form eines Leidgefühls – zu bestimmen. Sie gestattet es nämlich nicht einmal, Mitleid von selbstbezogenem Leid zu unterscheiden. Stellen wir uns z. B. vor, Maria leidet ausschließlich deshalb unter Karls Entlassung, weil sie glaubt, dass sich dadurch ihre eigene berufliche Situation verschlechtert (sie glaubt, dass sie einen Teil von Karls Arbeit übernehmen muss, was erhebliche unbezahlte Mehrarbeit bedeutet). Abgesehen davon ist ihr Karls Schicksal vollständig gleichgültig. In diesem Fall – so sagt mir meine Intuition – leidet Maria zwar darunter, dass Karl entlassen wurde; aber sie leidet nicht *mit ihm* über seine Entlassung. Mit anderen Worten, Maria erlebt zwar Leid *über* Karls Entlassung, aber kein Mitleid *mit* Karl über seine Entlassung (für empirische Belege siehe Reisenzein, 2002). Der Umstand, dass ein unerwünschtes Ereignis (unmittelbar) eine andere Person betrifft anstatt einen selbst, ist also nicht hinreichend dafür, dass man Mitleid mit der anderen Person erlebt. Damit Maria Mitleid mit Karl erlebt, muss mehr vorhanden sein als Marias Wunsch, dass Karl nicht entlassen wird und ihr Glaube, dass er entlassen worden ist.

Nach der CBDTE kommt jedoch ein weiterer mentaler Zustand als *unmittelbare Ursache* des Mitleids nicht in Frage. Denn laut CBDTE hat Leid über p nur zwei unmittelbare Ursachen, den Glauben, dass p , und den Wunsch, dass $\neg p$: Nur diese beiden mentalen Repräsentationen sind unmittelbare Inputs des BDC, des Mechanismus, der Leidgefühle erzeugt. Deshalb lässt sich das Problem der Bestimmung des Mitleids als spezielle Form des Leides nicht dadurch lösen, dass zum Wunsch nach $\neg Fo$ und dem Glauben, dass Fo , einfach eine weitere Teilursache des Mitleids hinzu genommen wird.

Es gibt aber noch eine andere Lösungsmöglichkeit: Auch wenn sich Mitleid und selbstbezogenes Leid nicht eindeutig durch die *unmittelbar* gefühlverursachenden Überzeugungen und Wünsche unterscheiden lassen, könnten sie sich doch in ihrem kognitiv-motivationalen *Hintergrund* unterscheiden; d. h. in den Überzeugungen und Wünschen, die den proximativen Gefühlsursachen zugrunde liegen. In unserem Beispiel: Der Unterschied zwischen Marias Mitleid mit Karl wegen dessen Entlassung und ihrem selbstbezogenen Leid darüber könnte in den *Gründen* zu suchen sein, aus denen Maria Karls Entlassung unerwünscht ist.

3.1.1 Mitleid hat einen besonderen kognitiv-motivationalen Hintergrund

Meist wünscht man sich Sachverhalte nicht um ihrer selbst willen, sondern weil man glaubt, dass sie zu anderen, gewünschten Sachverhalten führen oder zumindest die Chance ihres Eintretens erhöhen. Mit anderen Worten, die meisten Wünsche sind aus anderen Wünschen abgeleitet. Diese Ableitung wird gewöhnlich mit Hilfe von Mittel-Ziel-Überzeugungen bewerkstelligt (Reisenzein,

2006b; vgl. auch Lahnos Beitrag in diesem Band): Man wünscht sich p_1 , weil man sich p_2 wünscht und glaubt, dass p_1 zu p_2 führt; man wünscht sich p_2 , weil man sich p_3 wünscht und glaubt, dass p_2 zu p_3 führt usw.⁵, bis man schließlich bei Zuständen anlangt, die man sich nicht mehr als Mittel für andere Ziele wünscht. Die Wünsche nach solchen Zuständen sind die *Grundmotive*. Plausible Kandidaten für Grundmotive sind biologische Motive (der Wunsch nach Nahrung, Sex, körperlicher Unversehrtheit usw.), aber auch das hedonistische Motiv, das Machtmotiv, das Motiv nach Geselligkeit, nach Erkenntnisgewinn und andere „höhere“ Motive werden oft als Grundmotive betrachtet (Reisenzein, 2006b). Nach Ansicht einiger Motivationstheoretiker, denen ich mich hier anschließen möchte, gibt es neben egoistischen auch altruistische Grundmotive, also uneigennützig-wünsche nach dem Wohlergehen (geeigneter) anderer Personen. Empirische Belege für diese Annahme wurden insbesondere von C. D. Batson und seinen Mitarbeitern (z. B. Batson & Shaw, 1991) beigebracht.

Marias Wunsch, dass Karl nicht entlassen wird, und vergleichbar konkrete, auf andere Personen bezogene Wünsche sind sicherlich keine Grundmotive, sondern aus anderen Wünschen abgeleitet.⁶ Deshalb ist es *prinzipiell* möglich, dass die unmittelbar gefühlsverursachenden Wünsche im Fall von Mitleid aus anderen Wünschen hergeleitet sind als im Fall von selbstbezogenem Leid. Dass dem *tatsächlich* so ist, wird durch unser Illustrationsbeispiel – Maria fühlt Leid über Karls Entlassung, aber kein Mitleid mit Karl – ganz direkt nahe gelegt: Im Fall des selbstbezogenen Leides wünscht sich Maria nur deshalb, dass Karl nicht entlassen wird, weil sie glaubt, dass Karls Entlassung ihrem eigenen Wohlergehen schadet oder kurz: schlecht für *sie* ist. Im Fall des Mitleids wünscht Maria Karls Nichtentlassung dagegen deshalb, weil sie glaubt, dass die Entlassung dem Wohlergehen Karls schadet oder kurz: schlecht für *ihn* ist.

Marias Glaube, dass Karls Entlassung schlecht (bzw. seine Nichtentlassung gut) für ihn ist, reicht für die Entstehung ihres Wunsches nach Karls Weiterbeschäftigung alleine aber nicht aus. Abgeleitete Wünsche entstehen ja, wie gesagt, nicht ausschließlich aus Überzeugungen, sondern setzen andere Wünsche voraus, aus denen sie abgeleitet werden können. Und diese Wünsche müssen so beschaffen sein (d. h. einen solchen Inhalt haben), dass eine Ableitung auch

5 Diese Ableitung kann durch bewusste Denkprozesse geschehen, in denen man sich die eigenen Wünsche vergegenwärtigt; grundlegend sind m. E. jedoch Prozeduren, die beim Vorliegen des Wunsches nach p_2 und der dazu „passenden“ Mittel-Ziel-Überzeugung $p_1 \Rightarrow p_2$ automatisch einen abgeleiteten Wunsch nach p_1 generieren (s. auch Conte & Castelfranchi, 1995).

6 Diese Wunschableitung findet oft erst zum Zeitpunkt der Entstehung des Mitleids statt bzw. wird erst durch das auslösende Ereignis veranlasst. Sie kann natürlich aber schon früher stattgefunden haben, analog zum Fall der Freude Marias über Schroibers Wahlerfolg (Abb. 1). Beispielsweise könnte Maria den Wunsch, dass Karl nicht entlassen werden möge, erworben haben, als sie von bevorstehenden Personalentlassungen hörte.

tatsächlich möglich ist. In unserem Beispiel müssen wir deshalb annehmen, dass Maria noch einen weiteren Wunsch hat, aus dem mit Hilfe der Überzeugung „Die Entlassung Karls ist schlecht für Karl“ der Wunsch „Karl soll nicht entlassen werden“ abgeleitet werden kann. Der offensichtliche Kandidat dafür ist Marias Wunsch, dass Karl Gutes widerfährt und dass er von negativen Ereignissen verschont bleibt – zumindest in angemessenen Grenzen (z. B.: Karl bekommt dasjenige, was er verdient; vgl. Ortony et al. 1988). Im Fall von Marias selbstbezogenem Leid über Karls Entlassung ist Marias Wunsch nach Karls Nichtentlassung dagegen aus dem Wunsch abgeleitet, Zusatzbelastungen durch Mehrarbeit zu vermeiden, zusammen mit der Überzeugung, dass dieses Ziel durch Karls Weiterbeschäftigung erreicht werden kann.

Damit ist die Analyse des Mitleids aber noch nicht abgeschlossen. Denn man kann weiter fragen, woher Marias Wunsch nach dem Wohlergehen Karls seinerseits stammt. Meine These ist: Damit Maria Mitleid mit Karl erlebt und nicht bloß selbstbezogenes Leid, darf ihr Wunsch nach Karls Wohlergehen nicht aus egoistischen Wünschen abgeleitet sein. Vielmehr muss er (wenigstens teilweise) altruistisch sein. Für diese Annahme gibt es sowohl theoretische als auch empirische Gründe.

Zuerst die theoretischen Gründe: Wenn Maria Karls Wohlergehen nur aus egoistischen Gründen wünscht, dann ist auch ihr aus dem Wohlergehenswunsch abgeleiteter, konkreter Wunsch nach Karls Nichtentlassung egoistisch. Dann befindet sich Maria aber in derselben motivationalen Situation, wie wenn sie von Karls Weiterbeschäftigung direkt zu profitieren hofft. Die Entlassung Karls sollte deshalb bei Maria erneut nur selbstbezogenes Leid auslösen, kein Mitleid. Ferner: Der Wunsch, dass es einem anderen wohl ergehen möge, ist wie gesagt der Wunsch, dass dem anderen (in angemessenen Grenzen) positive Ereignisse zustoßen und dass er von negativen Ereignissen verschont bleibt. Wie ausgeführt scheint jedoch die Frustration eines *einzelnen* dieser Wünsche, sofern er rein egoistisch motiviert ist, kein Mitleid hervorzurufen. Dann ist aber nicht zu sehen, wie der Wunsch nach *mehreren* derartigen Ereignissen die Grundlage für Mitleid abgeben kann.

Der empirische Grund sind die Ergebnisse der mentalen Simulation von geeigneten hypothetischen Situationen. Wie würde man emotional reagieren, wenn einem anderen, dessen Wohlergehen einem ausschließlich aus egoistischen Gründen wichtig ist, etwas Negatives zustoßt? Man könnte sich dazu beispielsweise in die Rolle eines Sklaven versetzen, der seinen grausamen Herrn zutiefst verabscheut, aber dennoch um sein Wohlergehen besorgt ist, weil sein eigenes Schicksal vollständig von dem des Herrn abhängt: Jede Verschlechterung seines Wohlergehens lässt er den Sklaven sofort spüren. Welche Gefühle würde dieser Sklave erleben, wenn er erfährt, dass sein Herr beispielsweise einen herben geschäftlichen Verlust erlitten hat? Selbstbezogenes Leid und wahrscheinlich Furcht; aber nach meiner Intuition kein Mitleid.

Wenn Marias Wunsch nach dem Wohlergehen Karls aber nicht aus anderen, egoistischen Motiven abgeleitet ist, dann ist dieser Wunsch entweder ein Grundmotiv – das wäre dann ein personenspezifisches, d. h. auf die Person Karls bezogenes, altruistisches Grundmotiv – oder er ist aus grundlegenden nichtegoistischen Wünschen abgeleitet. Er könnte z. B. aus dem Grundmotiv abgeleitet sein, geeigneten nahe stehenden Personen möge in angemessenen Grenzen Gutes geschehen und sie mögen von Schlechtem verschont bleiben.

Die vorgeschlagene Analyse des Mitleids kann wie folgt zusammen gefasst werden.

CBDTE-Analyse des Mitleids

Mitleid mit o über $p = Fo$ ist eine Form von Leid, die man erlebt wenn:

- 1a. man glaubt, dass Fo ; mit Fo : ein Sachverhalt des Typs F , der (unmittelbar) eine andere Person (oder eine Gruppe) o betrifft; sowie
- 2b. man wünscht, dass $\neg Fo$.
3. der Wunsch nach $\neg Fo$ beruht auf:
 - 3a. dem Glauben: (Fo ist schlecht für o) bzw. ($\neg Fo$ ist gut für o); sowie
 - 3b. dem Wunsch: (o widerfährt – in angemessenen Grenzen – Gutes und nichts Schlechtes).
4. Der Wunsch 3b (o widerfährt Gutes und nichts Schlechtes) ist nicht aus egoistischen Gründen abgeleitet, sondern altruistisch.

Weil der Wunsch nach dem Wohlergehen des anderen (3b) laut Annahme altruistisch ist, ist auch der daraus abgeleitete, unmittelbar gefühlsverursachende Wunsch nach $\neg Fo$ (2b) altruistisch. Deshalb kann man auch verkürzt sagen: Mitleid mit o über Fo ist eine Form von Leid, die durch die wahrgenommene – genauer, vom BDC entdeckte – Frustration eines altruistischen Wunsches hervorgerufen wird. Ist der frustrierte Wunsch dagegen egoistisch, dann entsteht selbstbezogenes Leid. Natürlich ist es auch möglich, dass der durch das Schicksal des anderen frustrierte, proximate Wunsch teilweise altruistisch und teilweise egoistisch fundiert ist. Beispielsweise kann Maria die Nichtentlassung Karls sowohl deshalb wünschen, weil ihr Karls Wohlergehen am Herzen liegt, als auch deshalb, weil sie von Karls Weiterbeschäftigung zu profitieren hofft. In diesem Fall erlebt Maria eine Mischung von Mitleid und selbstbezogenem Leid (Reisenzein, 2002).

3.1.2. Mitleid hat ein besonderes intentionales Objekt

Es gibt es noch eine zweite Möglichkeit, das Mitleid im Rahmen der CBDTE zu analysieren. Die Grundidee dieses zweiten Analyseansatzes ist, dass die *intentionalen Objekte* des Mitleids von besonderer, für Mitleid eigentümlicher Art sind. Dahinter steht die allgemeinere Idee, dass die unterschiedlichen Vorkommnisse eines Emotionstyps (Freude, Leid, Mitleid usw.) zwar unterschiedli-

che konkrete Objekte haben, dass jedoch allen diesen konkreten Objekten etwas gemeinsam ist, das deshalb zur Bestimmung des Emotionstyps verwendet werden kann. Philosophen nennen diese gemeinsame Eigenschaft aller konkreten Objekte einer Emotion das „formale Objekt“ dieser Emotion (Kenny, 1963; de Sousa, 1987).

Die Möglichkeit der Analyse des Mitleids (und anderer Emotionen) durch Angabe eines formalen Objekts besteht auch im Rahmen der CBDTE; sie fällt nämlich einfach als Nebenprodukt der oben (Abschnitt 3.1.1) vorgeschlagenen Analyse an. Dort wurde Mitleid analysiert als eine Form von Leid, das auftritt, wenn man glaubt, dass ein Sachverhalt p vorliegt, der für eine andere Person negativ ist, und wenn man sich diesen Sachverhalt aus altruistischen Gründen nicht wünscht. Daraus lässt sich sofort ein formales Objekt für Mitleid konstruieren. Denn gemeinsam ist allen konkreten Objekten p von Mitleid offenbar: Sie sind – aus der kognitiv-motivationalen Perspektive des Erlebenden – vorliegende, für eine andere Person negative, altruistisch unerwünschte Sachverhalte. Das formale Objekt des Mitleids ist demnach einfach eine Beschreibung des intentionalen Objekts von Mitleid, die von der konkreten Natur dieses Objekts (z. B. Entlassung, Krankheit) absieht und es stattdessen relational charakterisiert, nämlich durch Bezugnahme auf die dem Mitleid zugrunde liegenden Überzeugungen und Wünsche. Auf diese Weise kann die kognitiv-motivationale Grundlage von Emotionen in die Beschreibung des Emotionsobjekts „verpackt“ werden, was die Darstellung vereinfachen kann (z. B. Lazarus, 1991; Ortony et al., 1988; Reisenzein et al., 2003): Mitleid ist Leid über einen vorliegenden, für einen anderen negativen, altruistisch unerwünschten Sachverhalt.

Die Charakterisierung des Mitleids durch Angabe eines formalen Objekts setzt also die im Abschnitt 3.1.1 vorgeschlagene Analyse voraus und fügt dieser Analyse zunächst nichts Substantielles hinzu. Ein substantieller Unterschied würde sich erst dann ergeben, wenn man annimmt, dass die vom Emotionsforscher konstruierten, formalen Objekte von Emotionen auch (in Form konkreter Realisierungen) die intentionalen Objekte der Emotionen der erlebenden Person sind; also konkret wenn postuliert würde, dass Maria, um Mitleid mit Karl zu erleben, dessen Entlassung *selbst* als einen „vorliegenden, für eine andere Person negativen, altruistisch unerwünschten Sachverhalt“ kognitiv repräsentieren muss. Nach der CBDTE ist eine derart komplexe Kognition für die Entstehung von Mitleid *nicht* erforderlich. Man kann zwar sagen, dass Maria die Entlassung Karls zum Zeitpunkt der Emotionsentstehung implizit als ein vorliegendes, für Karl negatives Ereignis repräsentiert; denn sie glaubt ja, dass Karl entlassen wurde, und sie glaubt auch, dass Karls Entlassung schlecht für ihn ist. Ein Teil der beschriebenen „Formalobjektkognition“ ist also zumindest implizit vorhanden. Der andere Teil dieser Kognition – „Karls Entlassung ist mir aus altruistischen Gründen unerwünscht“ – muss jedoch laut CBDTE nicht (auch nicht implizit) vorliegen, damit Mitleid erlebt wird. Im Gegensatz dazu scheint die

Bewertungstheorie der Emotion genau das zu implizieren (vgl. Abschnitt 2.1). Sofern die hier vorgeschlagene Analyse des Mitleids zutrifft, hieße das: Zur Entstehung von Mitleid ebenso wie von selbstbezogenem Leid über p ist nicht nur die Wertüberzeugung „ p widerspricht meinen Wünschen“ erforderlich, sondern auch die Überzeugung, dass diese Wünsche im einen Fall altruistisch und im anderen egoistisch sind. Diese Implikation der Bewertungstheorie ist nicht plausibel.

Da die CBDTE die Form der Beschreibung von Sachverhalten in der „Sprache des Denkens“ nicht einschränkt, schließt diese Theorie jedoch nicht aus, dass die Objekte von Emotionen auch von der erlebenden Person selbst als (Teil-)Realisierungen formaler Objekte *explizit* repräsentiert werden und dann unter dieser Beschreibung Gefühle auslösen können. So ist es nach der CBDTE zumindest theoretisch möglich, dass Maria Karl nicht (nur) dafür bemitleidet, dass er entlassen wurde, sondern (auch) dafür, dass ihm etwas Schlechtes zugestoßen ist. Damit letzteres geschieht, muss Maria den *expliziten* Glauben bilden, dass Karl etwas Schlechtes zugestoßen ist, ebenso wie den expliziten Wunsch, dass ihm dieses Schlechte nicht zustoßen möge. Denn nur wenn diese expliziten Repräsentationen vorliegen, kann der BDC darauf zugreifen und ein Leidgefühl erzeugen. Ein mögliches Beispiel wäre der Fall, in dem Maria erfährt, dass Karl etwas Schlechtes zugestoßen ist, aber noch nicht weiß, worin dieses Schlechte besteht. Laut CBDTE sollte Maria Karl dann zunächst dafür bemitleiden, dass ihm etwas Schlechtes zugestoßen ist, und anschließend – wenn sie erfährt, dass Karl entlassen wurde – noch einmal und spezieller dafür, dass er entlassen worden ist. Normalerweise verhält es sich allerdings zeitlich genau umgekehrt: Maria erfährt zuerst, dass ein konkretes Ereignis eingetreten ist (z. B. dass Karl entlassen wurde) und bildet erst danach, wenn überhaupt, eine abstraktere Überzeugung über dieses Ereignis wie „Karl ist etwas Schlechtes passiert.“ *Falls* Maria diese Überzeugung bildet, sollte sie nach der CBDTE jedoch erneut unterschiedliche Mitleidsgefühle erleben, auch wenn diese Gefühle in diesem Fall im Erleben kaum unterscheidbar sein dürften.

3.2 Analyse der normbasierten Gefühle am Beispiel der Schuld

Nach der vorausgehenden, relativ ausführlichen Untersuchung des Mitleids kann die Analyse der moralischen Gefühle im engeren Sinn (wie Schuld, Empörung und moralische Befriedigung), als deren Stellvertreter ich hier die Schuld analysiere, kürzer ausfallen. Grundsätzlich gehe ich davon aus, dass die traditionelle Sichtweise der Entstehung dieser Emotionen – es sind emotionale Reaktionen auf wahrgenommene Normverletzungen – korrekt ist. Damit widerspreche ich Autoren, die behaupten, dass beispielsweise Schuld auch auftreten kann, ohne dass die Person glaubt, etwas Normwidriges getan zu haben (vgl. dazu Wildt, 1993). Ich bestreite nicht, dass es Fälle gibt, die prima facie eine solche Analyse

nahe legen (z. B. das Phänomen der „survivor guilt“, Schuldgefühle von Überlebenden von Katastrophen); ich glaube aber, dass diese Fälle einer genauen Analyse nicht standhalten (vgl. auch Jäger & Bartsch, 2006). Auch wenn sich diese Einschätzung als falsch herausstellen sollte, ist doch weitgehend unbestritten, dass Schuld *normalerweise* das Resultat von wahrgenommenen Normverletzungen ist. Diese Normalfälle müssen von jeder Emotionstheorie erklärt werden. Ferner wäre dann erst noch zu erklären, wie Schuldgefühle (wenn es wirklich *Schuldgefühle* sind) in jenen Fällen entstehen, in denen sie nicht durch eine wahrgenommene Normverletzung hervorgerufen werden. Auf diese Frage ist meiner Meinung nach bislang keine überzeugende Antwort gegeben worden.

Wenn die CBDTE zutrifft, dann entstehen auch die moralischen Gefühle im engeren Sinn – die normbasierten Emotionen – im Prinzip auf dieselbe Art wie Freude und Leid. D. h., negative moralische Gefühle wie Schuld oder Empörung sind spezielle Formen des Leidens, die wie alle Leidengefühle von der Entdeckung einer Wunschfrustration durch den BDC hervorgerufen werden. Positive moralische Gefühle wie moralische Befriedigung treten dagegen auf, wenn der BDC eine Wunscherfüllung entdeckt.

Zur Illustration nehmen wir an: Maria (Person *a*) hat ihre Freundin Berta in einer Angelegenheit ohne Not belogen (Handlung *H*) und Maria fühlt sich nun schuldig über ihre Handlung *Ha*. Die Grundannahme der kognitiv-motivationalen Analyse von Schuld ist wie gesagt, dass Schuld eine Form von Leid ist. Schuld wird laut CBDTE von Maria deshalb dann erlebt, wenn Marias BDC entdeckt, dass der Inhalt einer neu erworbenen Überzeugung mit dem Inhalt eines ihrer Wünsche nicht übereinstimmt bzw. einen Wunsch frustriert. Der Inhalt der neu erworbenen Überzeugung *p* ist hier der Sachverhalt *Ha*, dass Maria Berta belogen hat. Die unmittelbaren Ursachen von Marias Schuldgefühl sind somit ihr Glaube, dass sie Berta belogen hat, und ihr Wunsch, dies nicht getan zu haben; Marias BDC stellt fest, dass der Inhalt der neu erworbenen Überzeugung (*Ha*) dem Inhalt eines ihrer Wünsche ($\neg Ha$) widerspricht; daraufhin erzeugt der BDC ein Signal, das diese Wunschinkongruenz signalisiert und von Maria als Unlustgefühl erlebt wird.

Analog zum Fall des Mitleids reicht diese Analyse aber nicht aus, um Schuld als eigenständige Form des Leidens von selbstbezogenem Leid abzugrenzen. Stellen wir uns z. B. vor, dass Maria ausschließlich deshalb Leid über ihre Lüge erlebt, weil diese sich als sehr nachteilig für Maria erwiesen hat. Moralische Bedenken wegen ihrer Lüge hat Maria dagegen keine. In diesem Fall – so sagt meine Intuition – wird Maria es zwar bedauern, Berta belogen zu haben; Schuldgefühle erleben wird sie aber nicht.

Was ist dann aber das Besondere des Schuldgefühls; worin liegt das Spezielle des Leidens, das es darstellt? Meine These lautet analog zum Mitleid: Das Besondere des Schuldgefühls ist in den Wünschen und damit verbunden den Überzeugungen zu suchen, die den kognitiv-motivationalen Hintergrund der Schuld

bilden. Als hochsoziale Wesen (Richerson & Boyd, 1998) haben Menschen auch Wünsche und Überzeugungen betreffend die Einhaltung von sozialen und moralischen Normen. Schuld wird erlebt, wenn man zur Überzeugung kommt, etwas getan zu haben, das einer (in einem speziellen motivationalen Sinn; siehe dazu unten) internalisierten Norm widerspricht.

Anstatt die CBDTE-Analyse von Schuld analog zum Fall des Mitleids schrittweise zu entwickeln, präsentiere ich sie sofort und kommentiere sie anschließend.

CBDTE-Analyse der Schuld

Schuld über $p = Ha$ ist eine Form von Leid, die man erlebt, wenn:

- 1a. man glaubt, dass Ha ; mit Ha : eine Handlung der Typs H , die man selbst (a) ausgeführt hat; sowie
- 1b. man wünscht, dass $\neg Ha$ (dass man die Handlung nicht ausgeführt hätte).
2. der Wunsch nach $\neg Ha$ beruht auf:
 - 2a. dem Glauben: (a ist ein Akteur des Typs A und a befindet sich in einer Situation des Typs S); sowie
 - 2b. dem Wunsch: (in S führen Akteure des Typs A Handlungen des Typs H nicht aus).
3. Der Wunsch 2b (nach Regeleinhaltung) beruht auf:
 - 3a. dem Glauben: (in S sind für Akteure des Typs A Handlungen des Typs H durch den Normautor P verboten);
 - 3b. dem Wunsch: (die Gebote und Verbote von P werden befolgt).
4. der Wunsch 3b (die Gebote von P werden befolgt) ist nicht egoistisch motiviert.

In unserem Beispiel ist demnach Marias Wunsch, Berta nicht zu belügen (1b) abgeleitet aus ihrem Wunsch, eine bestimmte Verhaltensregel einzuhalten (2b). Verhaltensregeln können allgemein formuliert werden als Regeln der Art „in S führen Akteure des Typs A Handlungen des Typs H (nicht) aus“ (vgl. hierzu die Beiträge von Iorio und Siegwart in diesem Band). In unserem Beispielfall lautet die relevante Verhaltensregel etwa „Einen Freund belügt man nicht ohne Not“ oder etwas ausführlicher: „Wenn man mit einem Freund kommuniziert und kein wichtiger Grund besteht, zu lügen, dann belügt man den anderen nicht.“ Aus Marias Wunsch, dass diese Regel eingehalten wird (2b), lässt sich mit Hilfe ihrer Überzeugung, sich in der Situation S (eine Kommunikationssituation, in der es keinen wichtigen Grund zum Lügen gibt) zu befinden und ein Akteur des Typs A (ein Freund des anderen, der diesem etwas mitteilt) zu sein (2a), ihr konkreter Wunsch ableiten, Berta nicht zu belügen (1b).

Ich nehme ferner an: Marias Wunsch nach Einhaltung des genannten Lügenverbots ist seinerseits abgeleitet aus (3a) Marias Überzeugung, dass die Verhaltensregel durch einen Normautor P geboten wurde, und (3b) Marias Wunsch,

dass die Gebote dieses Normautors befolgt werden. Der Normautor *P* kann eine einzelne Person, die Gruppe oder eine überindividuelle oder abstrakte Instanz sein. Die Ableitung des Wunsches (2b) aus (3a) und (3b) ist eine Explikation des Prozesses der Internalisierung von Normen. Sie entspricht weitgehend der Sichtweise, die Baurmann (in diesem Band) im Anschluss an Hart (1994, Original 1961) entwickelt. Eine ganz ähnliche Sichtweise der Norminternalisierung findet sich auch in Teilen der Psychologie (z. B. Ajzen, 1991) und der Kognitionswissenschaft (Conte & Castelfranchi, 1995). Eine Norm zu *internalisieren* oder zu *akzeptieren* beinhaltet demnach nicht nur, die Norm zu kognizieren, d. h. zu erkennen, dass ein Gebot besteht (3a), sondern auch die „Importierung“ der gebotenen Handlungsregel in das eigene motivationale System; d. h. die Entstehung des Wunsches, dass die Regel befolgt wird (2b). Nach der vorgeschlagenen Explikation entsteht dieser Wunsch erneut durch seine Ableitung aus einem grundlegenden Wunsch; nämlich dem Wunsch, dass die Gebote des Normautors *P* befolgt werden (3b).

Beispielsweise geht die Internalisierung des Gebots „Man soll einen Freund nicht ohne Not belügen“ durch Maria folgendermaßen vor sich: Zuerst kommt Maria zur Überzeugung, dass dieses Gebot besteht, d. h., dass die Einhaltung der Handlungsregel „Einen Freund belügt man nicht“ von einem Normautor *P* (z. B. ihren Eltern) gewünscht wird (3a). Weil Maria wünscht, dass die Gebote von *P* befolgt werden (3b), entsteht daraufhin in ihr der abgeleitete Wunsch, dass die Handlungsregel befolgt wird. Die Überzeugung 3a repräsentiert den von Baurmann (in diesem Band) unterschiedenen hermeneutischen Aspekt der Normakzeptanz; der Wunsch 3b den motivationalen Aspekt der Normakzeptanz. Nach der CBDTE muss eine Normakzeptanz im motivationalen Sinn vorliegen, damit eine wahrgenommene Erfüllung oder Verletzung der Norm eine emotionale Reaktion hervorrufen kann. Baurmann differenziert den motivationalen Aspekt der Normübernahme in den Wunsch, dass andere die geforderte Regel einhalten („Normakzeptanz im Forderungssinn“), und den Wunsch, dass man selbst die Regel einhält („Normakzeptanz im Befolgungssinn“). Diese Differenzierung ergibt sich in der vorliegenden Darstellung in Abhängigkeit davon, ob sich die Person selbst zur Gruppe der Akteure zählt, für die die Handlungsregel gilt, oder nicht. Ist dies nicht der Fall, dann liegt eine Normakzeptanz im Forderungssinn vor; ist es der Fall, dann liegt auch eine Normakzeptanz im Befolgungssinn vor. Für Schuld ist offenbar letzteres erforderlich (2a). Denn wenn sich die Person nicht zu den Akteuren zählt, für die die Handlungsregel gilt, dann kann im Fall einer normwidrigen Handlung *Ha* der Wunsch nach $\neg Ha$ nicht aus dem Wunsch nach Regeleinhaltung abgeleitet werden. Jedoch sollte die Person emotional (z. B. empört) auf die normverletzenden Handlungen anderer Personen reagieren, für welche die Norm ihrer Meinung nach gilt. Denn für diese Personen *kann* ein entsprechender Wunsch abgeleitet werden.

Schließlich nehme ich – analog zur Analyse des Mitleids – an, dass der Wunsch 3b (die Gebote von *P* werden befolgt) nicht egoistisch motiviert ist. Befolgt man eine Norm nämlich ausschließlich aus egoistischen Interessen, z. B. um Reputation zu gewinnen oder Sanktionen zu vermeiden, dann – so sagt meine Intuition – erlebt man im Fall einer eigenen normabweichenden Handlung zwar möglicherweise Bedauern, aber keine Schuld. Der Wunsch 3b könnte altruistisch motiviert sein oder durch die Einnahme einer Gruppenperspektive zustande kommen (vgl. auch Lahnos Beitrag in diesem Band). Möglicherweise gibt es auch eine angeborene Bereitschaft, Normen zu internalisieren, die man unabhängig von ihrem Autor als gültig betrachtet (vgl. dazu auch Conte & Castelfranchi, 1995).

Ebenso wie im Fall des Mitleids ist auch im Fall der Schuld eine alternative Analyse möglich, die auf das formale Objekt der Schuld abhebt anstatt auf ihren kognitiv-motivationalen Hintergrund. Schuld ist demnach Leid über eine eigene Verletzung einer Norm, die man nicht (nur) aus egoistischen Gründen akzeptiert.

3.3 Zu den Funktionen der moralischen Gefühle aus Sicht der CBDTE

Nach der CBDTE besteht die Funktion von negativen Emotionen darin, Inkongruenzen zwischen neu erworbenen Überzeugungen und Wünschen zu signalisieren und das kognitive System auf den Umgang mit ihnen global vorzubereiten. Die moralischen Gefühle passen gut in dieses Bild. Wieder am Beispiel von Mitleid und Schuld illustriert: Mitleid signalisiert nach den Ergebnissen des Abschnitts 3.2 die Frustration eines altruistischen Wunsches nach dem Wohlergehen eines anderen; Schuld die Frustration eines nichtegoistischen Wunsches nach eigener Normeinhaltung. Gleichzeitig mit dem Gefühlserleben wird die Aufmerksamkeit automatisch auf den gefühlsverursachenden Sachverhalt gelenkt – das negative Los des anderen; die eigene normabweichende Handlung – was die Voraussetzung für die bewusste Analyse der Ursachen und Folgen dieser Ereignisse ist (Reisenzein et al., 2003; 2009a). Die begleitenden Unlustgefühle können die Motivation verstärken, die Auslösesituation zu ändern (Reisenzein, 2009a; 2009b), z. B. indem man dem anderen hilft oder einen zugefügten Schaden wieder gut zu machen versucht. Zudem können antizipierte Schuldgefühle – Schuldgefühle, die auftreten, wenn man sich einen möglichen Normverstoß lebhaft vorstellt und die Emotionsmechanismen dabei im „Simulationsmodus“ betreibt (Reisenzein, 2009a; Reisenzein et al., 2003) – eine Normverletzung im Vorhinein verhindern (vgl. auch von Scheves Beitrag in diesem Band).

Wegen der unwillkürlichen, automatischen Auslösung von Emotionen durch die Kongruenzdetektoren ist das Auftreten von Mitleid und Schuld bei anderen Personen ein gutes Motivdiagnostikum: Das Mitleid des anderen zeigt uns an,

dass ihm unser Los „wirklich“ (d. h. nicht bloß aus egoistischen Gründen) am Herzen liegt; die Schuldgefühle des anderen zeigen an, dass ihm die Einhaltung der Norm, die er durch seine Handlung verletzt hat, ein „echtes“ Anliegen ist (d. h. dass er die Norm nicht bloß aus egoistischen Gründen, etwa aus Furcht vor Sanktionen befolgt). Mangelnde Schuld angesichts einer Normverletzung bedeutet allerdings nicht unbedingt, dass der andere die Norm nicht akzeptiert; sie kann auch bedeuten, dass der andere sich (vielleicht irrtümlich) nicht zu den Akteuren rechnet, für die die Norm gilt. In diesem Fall sollte sich die Normakzeptanz der Person jedoch zumindest in ihrer Empörung über die Normverletzung anderer Personen äußern. Die Tatsache, dass die moralischen Gefühle sowohl uns selbst als auch anderen Personen unsere sozialen Motive verlässlich anzeigen, ist m. E. von zentraler Bedeutung für die Stabilisierung sozialer Ordnung; zumindest aber für das Verständnis der Rolle, die Emotionen dabei spielen (vgl. von Scheves Beitrag in diesem Band).

4. Ausblick

Die vorgeschlagene Analyse von Mitleid lässt sich mit geringen Änderungen auf die übrigen Sympathie- und Antipathieemotionen (Mitfreude, Schadenfreude und Neid) übertragen. Analog kann die Analyse von Schuld mit geringen Abänderungen auf die Analyse anderer normbasierter Gefühle wie Empörung und moralische Befriedigung übertragen werden.

Einige für die moralischen Emotionen bedeutsame Einflussfaktoren wurden in der vorgeschlagenen Analyse nicht explizit berücksichtigt. Ausgeklammert wurde insbesondere der Faktor „wahrgenommene Verantwortlichkeit“ (vgl. Ortony et al., 1988; Reisenzein et al., 2003; Weiner, 2006). Eine Möglichkeit, diesen Faktor im Rahmen der vorgeschlagenen Analyse zu berücksichtigen, ist folgender: Die Verantwortlichkeit modifiziert den Grad, zu dem die Person sich bzw. andere als Akteure sieht, für die eine Norm gilt. Dieser Vorschlag weist bereits auf die Notwendigkeit einer weiteren wichtigen Erweiterung der Analyse hin, nämlich die Entwicklung eines quantitativen Modells, durch das nicht nur die Qualität, sondern auch die Intensität der moralischen Emotionen vorhergesagt werden kann. Für die im Abschnitt 2.1 analysierten einfacheren Emotionen liegen quantitative Modelle bereits vor (vgl. Reisenzein, 2009a; 2009b).

Schließlich muss die vorgeschlagene Glauben-Wunsch-Analyse der moralischen Gefühle empirisch überprüft werden. Die am meisten stützungsbedürftige Annahme ist sicherlich, dass die moralischen Gefühle (zumindest Mitleid und Schuld) nichtegoistische Motive voraussetzen. Denn die Existenz solcher Motive ist trotz stützender Befunde (Batson & Shaw, 1991) nach wie vor umstritten. Gerade in diesem Punkt eröffnet jedoch die Untersuchung der moralischen Gefühle neue empirische Möglichkeiten (Reisenzein, 2002).

Literatur

- Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah: Erlbaum.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Arnold, M. B. (1960). *Emotion and personality (Vol. 1 & 2)*. New York: Columbia University Press.
- Aydede, M. (2004). The language of thought hypothesis. *Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/entries/language-thought/>.
- Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Batson, C. D. & Shaw, L. L. (1991). Evidence for altruism: Toward a pluralism of prosocial motives. *Psychological Inquiry*, 2, 107-122.
- Block, B. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227-287.
- Castelfranchi, C. & Miceli, M. (2009). The cognitive-motivational compound of emotional experience. *Emotion Review*, 1, 223-231.
- Clore, G. L. (1994). Why emotions are felt. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion* (pp. 103-111). Oxford: Oxford University Press.
- Conte, R. & Castelfranchi, C. (1995). *Cognitive and social action*. London: UCL Press.
- Cummins, E. E. (1983). *The nature of psychological explanation*. Cambridge: MIT Press.
- Davis, W. (1981). A theory of happiness. *Philosophical Studies*, 39, 305-317.
- De Sousa, R. (1987). *The rationality of emotions*. Cambridge: MIT Press.
- De Sousa, R. (2001). Moral emotions. *Ethical Theory and Moral Practice*, 4, 109-126.
- Dennett, D. C. (1971). Intentional systems. *Journal of Philosophy*, 68, 87-106.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6, 169-200.
- Ellsworth, P. C. & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572-595). Oxford: Oxford University Press.
- Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge: MIT Press.
- Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852-870). Oxford: Oxford University Press.
- Hart, H. L. A. (1994). *The Concept of Law*. Oxford: Clarendon Press.

- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Gratch, J. & Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5, 269-306.
- Green, O. H. (1992). *The emotions: A philosophical theory*. Dordrecht: Kluwer.
- Jäger, C. & Bartsch, A. (2006). Meta-emotions. *Grazer Philosophische Studien*, 73, 179–204.
- Kenny, A. (1963). *Action, emotion, and will*. London: Routledge and Kegan Paul.
- Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- Lyons, W. (1980). *Emotion*. Cambridge: Cambridge University Press.
- Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- Marks, J. (1982). A theory of emotion. *Philosophical Studies*, 42, 227-242.
- McDougall, W. (1908/1960). *An introduction to social psychology*. London: Methuen.
- Meinong, A. (1894). *Psychologisch-ethische Untersuchungen zur Werth-Theorie*. Graz, Austria: Leuschner & Lubensky. Wiederabdruck in R. Haller & R. Kindinger (Hg.) (1968). *Alexius Meinong Gesamtausgabe* (Band III, S. 3-244). Graz: Akademische Druck- und Verlagsanstalt.
- Meinong, A. (1906). Über Urteilsgefühle: was sie sind und was sie nicht sind. *Archiv für die gesamte Psychologie*, 6, 22-58.
- Meinong, A. (1917). Über emotionale Präsentation. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, Philosophisch-Historische Klasse* (Band 183, S. 1-181). Wien: Hölder. Wiederabdruck in: R. Haller & R. Kindinger (Hg.) (1968), *Alexius Meinong Gesamtausgabe* (Band III, S. 285-465). Graz: Akademische Druck- und Verlagsanstalt.
- Nussbaum, M. C. (2001). *Upheavals of thought: The intelligence of emotions*. Cambridge: Cambridge University Press.
- Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, 1, 29-50.
- Ortony, A., Clore, G. L. & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.
- Reisenzein, R. (2001). Appraisal processes conceptualized from a schematheoretic perspective: Contributions to a process analysis of emotions. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 187-201). Oxford: Oxford University Press.
- Reisenzein, R. (2002). Die kognitiven und motivationalen Grundlagen der Empathie-Emotionen. *Referat auf dem 43. Kongress der DGPs in Berlin 2002*.

- Reisenzein, R. (2006a). Arnold's theory of emotion in historical perspective. *Cognition & Emotion*, 20, 920-951.
- Reisenzein, R. (2006b). Motivation. In K. Pawlik (Hg.), *Handbuch Psychologie* (S. 239-247). Berlin: Springer.
- Reisenzein, R. (2007). What is a definition of emotion? And are emotions mental-behavioral processes? *Social Science Information*, 46, 424-428.
- Reisenzein, R. (2009a). Emotions as metarepresentational states of mind: Naturalizing the belief-desire theory of emotion. *Cognitive Systems Research*, 10, 6-20.
- Reisenzein, R. (2009b). Emotional experience in the computational belief-desire theory of emotion. *Emotion Review*, 1, 214-222.
- Reisenzein, R. & Döring, S. (2009). Ten perspectives on emotional experience: Introduction to the special issue. *Emotion Review*, 1, 195-205.
- Reisenzein, R., Meyer, W.-U. & Schützwohl, A. (2003). *Einführung in die Emotionspsychologie, Band III: Kognitive Emotionstheorien*. Bern: Huber.
- Reisenzein, R. & Rudolph, U. (2008). The discovery of common-sense psychology. *Social Psychology*, 39, 125-133.
- Richerson, P. & Boyd, R. (1998). The evolution of ultrasociality. In I. Eibl-Eibesfeldt, & F. K. Salter (Eds.), *Indoctrinability, ideology and warfare* (pp. 71-96). New York: Berghahn Books.
- Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-129). Oxford: Oxford University Press.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Solomon, R. C. (1976). *The passions*. Garden City: Anchor Press/Doubleday.
- Sterelny, K. (1991). The representational theory of mind: an introduction. Oxford: Blackwell.
- Strawson, F. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 187-211.
- Sun, R. (2008). *The Cambridge handbook of computational psychology*. Cambridge: Cambridge University Press.
- Weiner, B. (2006). *Social motivation, justice, and the moral emotions: An attributional approach*. Mahwah: Erlbaum.
- Wildt, A. (1993). Die Moralspezifität von Affekten und der Moralbegriff. In H. Fink-Eitel & G. Lohmann (Hg.), *Zur Philosophie der Gefühle* (S. 188-217). Frankfurt/Main: Suhrkamp.
- Wundt, W. (1896). *Grundriss der Psychologie*. Leipzig: Engelmann.