

Warum ich weiß, dass ich kein Zombie bin

Thomas Grundmann, Universität zu Köln

Wir alle oder – etwas vorsichtiger formuliert – zumindest die meisten von uns sind in einem bestimmten Sinne Cartesianer. Wir sind ohne einen Rest von Zweifel davon überzeugt, dass wir wissen, *was* wir denken, wissen *wie* (d.h. mit welcher Einstellung) wir es denken und wissen *dass* wir überhaupt Gedanken haben. Und wir sind außerdem davon überzeugt, dass wir dies nicht aufgrund irgendwelcher empirisch-psychologischen Untersuchungen oder Beobachtungen unseres Verhaltens wissen, sondern auf eine direkte, privilegierte Weise, eine Weise die keinem äußeren Beobachter zur Verfügung steht. Kurz: Wir haben introspektives Selbstwissen. Das nenne ich die „cartesianische Grundintuition“. Für sich genommen impliziert diese Intuition weder, dass unser Selbstwissen infallibel ist, noch, dass es auf innerer Beobachtung beruht. Die cartesianische Grundintuition besagt nur, dass wir Selbstwissen haben, das nicht auf äußerer Beobachtung beruht.

Fred Dretske und Sven Bernecker haben neuerdings dafür argumentiert, dass diese cartesianische Grundintuition nur in begrenztem Umfang mit einer gegenwärtig in der Philosophie des Geistes sehr beliebten Position verträglich ist; ich meine den repräsentationalistischen Externalismus (Bernecker 2000; Dretske 2003a und Dretske 2003b). Dieser Position zur Folge sind alle geistigen Zustände repräsentationaler Natur, d.h. sie repräsentieren Gegenstände als so-und-so beschaffen, und der repräsentationale Gehalt der geistigen Zustände hängt von den Beziehungen des Systems zu seiner Umwelt ab. Wenn diese Position richtig ist, so sagen Dretske und Bernecker, dann folgt daraus zwar nicht, dass wir überhaupt kein introspektives Wissen haben können, aber es folgt daraus, dass wir etwas ganz Bestimmtes *nicht* introspektiv wissen können, nämlich *dass wir Wesen mit einem Geist sind und keine geistlosen Zombies*. Wir können demnach nicht auf eine privilegierte Weise wissen, dass wir überhaupt Gedanken und Bewusstsein haben. Das ist ein Faustschlag ins Gesicht des Cartesianers. Bernecker bezeichnet diese Konsequenz als „Zombie-Skeptizismus“ (Bernecker 2000, 19). Als Zombie wird in diesem Zusammenhang jemand verstanden, der in seinem Aussehen und Verhalten menschenähnlich ist, aber über keinerlei mentale Zustände verfügt. Was ihn von uns unterscheidet ist also nicht nur die Abwesenheit von phänomenalem Bewusstsein (Chalmers 1996, 95¹), sondern die Abwesenheit von jeder Art von geistigem Innenleben (Bernecker 2000, 17). Die Dretske-Bernecker These lautet genauer:

¹ Chalmers spricht vom „phänomenalen Zombie“.

(DB) Wenn der Geist repräsentationaler Natur ist und wenn die Existenz repräsentationaler Eigenschaften von der Beziehung eines Systems zu seiner Umwelt abhängt, dann haben wir *introspektives* Wissen nur von den spezifischen Inhalten unserer Repräsentationen, aber kein *introspektives* Wissen davon, *dass* wir einen Geist haben.²

Nun sind Dretske und Bernecker beide repräsentationalistische Externalisten. Und sie erkennen an, dass diese Position mit der cartesianischen Grundintuition nur begrenzt in Einklang zu bringen ist. Da die Position jedoch introspektives Wissen nicht in Bausch und Bogen verwirft, sondern in gewissem Umfang zulässt, halten sie die Position für vertretbar. Meiner Meinung nach wäre aber auch eine andere Diagnose möglich: Wenn Dretske und Bernecker Recht haben, dann widerspricht der Gehaltsexternalismus unserer cartesianischen Grundintuition, und sollte deshalb aufgegeben werden.³ Weil die skeptischen Konsequenzen, die Dretske und Bernecker aus dem Externalismus ziehen, unter Umständen also auch als *reductio ad absurdum* des Externalismus verstanden werden können, möchte ich in diesem Artikel noch einmal genauer untersuchen, ob der Gehaltsexternalismus tatsächlich die behaupteten skeptischen Konsequenzen hat.

Ich werde erstens zu zeigen versuchen, dass sich das Problem des Zombie-Skeptizismus nicht allein aus dem Gehaltsexternalismus ergibt, sondern auch von Dretskes spezifischen informationstheoretischen Wissensbegriff abhängt. Ich werde diesen Wissensbegriff nicht in Frage stellen, sondern zweitens dafür argumentieren, dass dieser Wissensbegriff nicht die skeptischen Konsequenzen hat, die Dretske und Bernecker für unausweichlich halten. Schließlich werde ich zeigen, dass auch unter der Annahme eines anderen externalistischen Wissensbegriffs, nämlich dem von Robert Nozick, die skeptischen Konsequenzen vermieden werden können. Sollte ich Recht haben, dann wäre der repräsentationalistische Externalismus mit unserer cartesianischen Grundintuition vollkommen verträglich und das würde diese Position natürlich deutlich stärken.

Beginnen möchte ich mit einigen Erläuterungen zur Dretske-Bernecker These.

² Dretske 2003b, 137: „What introspection gives us is the content of our cognitive states (...), not the fact that it is content.“ Bernecker 2000, 2: „Self-knowledge provides us with knowledge of what is in our minds, but not that we have minds.“

³ Auch Dretske gibt durchaus zu, dass die Konsequenzen seiner Auffassung zumindest kontraintuitiv sind. In diesem Sinne verstehe ich den Ausdruck „paradox“ im folgenden Zitat: „I know this conclusion sounds paradoxical. It flies in the face of a cherished and deeply entrenched Cartesian doctrine, the doctrine that the first and indubitable fact is that we think.“ (Dretske 2003b, 137)

(i) Dort ist von „introspektivem Wissen“ die Rede. Gemeint ist dabei nicht etwa irgend ein positives Modell, wie das der inneren Wahrnehmung, das ja bekanntlich schwer lösbare Probleme aufwirft,⁴ sondern nur eine Wissensquelle, die von der äußeren Wahrnehmung (durch Sehsinn, Tastsinn, Gehör, Geruch oder Geschmack) unabhängig ist. Wir haben es mit einer rein negativen Charakterisierung, einem minimalen Begriff der Introspektion zu tun. Introspektives Wissen ist Wissen, das unabhängig von äußerer Wahrnehmung ist (Bernecker 2000, 24f).

(ii) Als nächstes müssen wir uns Dretskes Wissensbegriff genauer ansehen, der – wie noch zu zeigen sein wird – eine wichtige Rolle in der Argumentation für die Dretske-Bernecker These spielt. Nach Dretske weiß ein Subjekt S, dass p, genau dann, wenn S die Überzeugung hat, dass p, und diese Überzeugung kausal auf die *Information*, dass p, gestützt ist (Dretske 1981, 86). Was *Information* ist, hat Dretske allerdings nicht immer ganz einheitlich definiert. Klar ist jedoch, dass die Information, dass p, impliziert, dass p wahr ist. Und deshalb macht die Informationsbedingung des Wissens die übliche Wahrheitsbedingung des Wissens redundant. In *Knowledge and the Flow of Information* hat Dretske einen *wahrscheinlichkeitstheoretischen* Informationsbegriff vertreten. Danach trägt das Signal r die Information, dass p, gdw. p gegeben r eine bedingte Wahrscheinlichkeit von 1 hat (Dretske 1981, 65). Christoph Jäger hat kürzlich darauf aufmerksam gemacht, dass bedingte Wahrscheinlichkeiten von 1 unter logischer Implikation geschlossen sind (Wenn die bedingte Wahrscheinlichkeit von p, gegeben r, gleich 1 ist und p logisch q impliziert, dann ist auch die bedingte Wahrscheinlichkeit von q, gegeben r, gleich 1. (Jäger 2004)) Dretske lehnt jedoch die Geschlossenheit von Information und Wissen ab.⁵ Und er muss dies auch für seine Argumentation in unserem Zusammenhang tun, wie wir gleich noch sehen werden. Deshalb empfiehlt es sich, eine andere Definition der Information aufzugreifen, die Dretske in seinem frühen Aufsatz *Conclusive Reasons* unter dem Etikett des zwingenden Grundes präsentiert hat (Dretske 1971). Nach dieser *modalen Definition* trägt das Signal r die Information, dass p, gdw. r nicht aufgetreten wäre, wenn p nicht der Fall gewesen wäre (wobei zu beachten ist, dass die Bewertung des kontrafaktischen Satzes in den nächsten und nicht beliebigen nicht-p Welten stattfindet). Die modale Definition der Information hat das gewünschte Ergebnis – Information ist nicht geschlossen unter logischer Implikation. Das kann ein Beispiel zeigen:

⁴ Vgl. zur Kritik dieses Modells etwa Shoemaker 1996.

⁵ Vgl. Dretske 2004, im Manuskript S. 6: „I think there are some things we know to be implied by what we know that we do not – perhaps cannot – know to be true. My reason for rejecting closure, however, is not my externalism. It is that I think I have some reasonable clear idea of what kind of evidential relation is required for knowledge (a ‘conclusive reason’ or ‘information’) and this relation is itself not closed under known implication.”

Meine Erfahrung eines Tisches trägt die Information, dass vor mir ein Tisch ist, weil ich in der nächsten Welt, in der sich kein Tisch vor mir befinden würde, keine Erfahrung eines Tisches hätte. *Dass sich vor mir ein Tisch befindet* impliziert nun logisch, *dass ich kein von einem bösen Neurowissenschaftler getäushtes Gehirn im Tank in einer tischlosen Welt bin*. Aber meine Tischerfahrung trägt nicht die Information, dass ich kein solches Gehirn im Tank bin, denn wenn ich ein solches Gehirn wäre, hätte ich dieselbe Tischerfahrung. Die modale Definition der Information führt also zur Aufgabe des Geschlossenheitsprinzips.

(iii) Die Dretske-Bernecker These kann nur dann konsistent sein, wenn eine spezifische Version des Geschlossenheitsprinzips – die Transitivität der Gründe – falsch ist. Das Transitivitätsprinzip lautet:

(T) Wenn *S aufgrund von r* weiß, dass *p*, und aus *p* logisch *q* folgt, dann weiß *S aufgrund von r*, dass *q*.⁶

Wenn dieses Prinzip korrekt wäre, dann würde *S*, wenn er *introspektiv* wüsste, dass er *p* (und nicht *q*) repräsentiert (was von DB ja zugegeben wird), automatisch auch *introspektiv* wissen, dass er repräsentiert (und nicht ohne Geist ist), weil *S repräsentiert, dass p* logisch impliziert *S repräsentiert*. Doch diese Konsequenz wird ja gerade von der DB-These bestritten. Also muss das Transitivitätsprinzip aufgegeben werden. Folglich kann die These nur vertreten werden, wenn der modale Informationsbegriff zugrunde gelegt wird.

Soweit meine Erläuterungen der DB-These. Sehen wir uns jetzt die Argumente für diese These an. Meines Erachtens lassen sich zwei Argumente unterscheiden. Das erste Argument möchte ich als „Argument von den fehlenden rationalisierenden Gründen“ bezeichnen.⁷

ARGUMENT I

- (1) Wissen erfordert Gründe, die die Überzeugung rationalisieren. (Prämisse)
- (2) Ein Grund rationalisiert eine Überzeugung, wenn es eine inferentielle Beziehung zwischen dem repräsentationalen Gehalt des Grundes und dem repräsentationalen Gehalt der Überzeugung gibt. (analytisch)

⁶ Dretske lehnt dieses Prinzip explizit ab (vgl. Dretske 2003b, 139). Geschlossenheit im Allgemeinen besagt nur, dass man alle Dinge weiß, die (bekannter Weise) von etwas logisch impliziert werden, das man weiß. Aber dieses Prinzip sagt nichts darüber aus, *wie* man diese Dinge weiß. Das Transitivitätsprinzip behauptet dagegen, wenn man von etwas aufgrund bestimmter Gründe weiß, dass man dann aufgrund derselben Gründe auch alles das weiß, was (bekannter Weise) durch es logisch impliziert wird. Vgl. zur Unterscheidung von Geschlossenheit und Transitivität der Gründe Klein 1981 sowie Barke 2002, 75ff.

- (3) Introspektiv zugänglich sind nur die Gehalte der Repräsentationen und die internen Eigenschaften der Zustände des Organismus. (Prämisse)
- (4) Dass es sich um einen gehaltvollen Zustand handelt ist nicht Teil des Gehalts einer Repräsentation. (Prämisse)
- (5) Da der Gehaltsexternalismus wahr ist, lässt sich die Tatsache, dass repräsentationaler Gehalt vorliegt, auch nicht aus den internen Eigenschaften der Zustände des Organismus ableiten. (Prämisse)
- (6) Die Gehalte der Introspektion rationalisieren deshalb nicht die Überzeugung, dass repräsentationaler Gehalt vorliegt. (2), (3), (4), (5)
- (7) Also gibt es kein introspektives Wissen davon, dass repräsentationaler Gehalt vorliegt. (1), (6)

Dretske argumentiert in seinem Aufsatz *How do you know you are not a zombie?* vor allem für Prämisse (4). Die Quintessenz seiner Überlegungen lässt sich durch das folgende Beispiel veranschaulichen: Es gibt, sagt Dretske, selbstverständlich einen Unterschied zwischen Dingen, die man sieht, und solchen Dingen, die man nicht sieht. Und dieser Unterschied macht unsere Perspektive auf die Welt aus. Aber wenn man die Welt um einen herum sieht, dann sieht man seine eigene Perspektive auf die Welt nicht einfach mit (Dretske 2003a, 3). Diese Perspektive ist kein Teil des Inhalts unserer Wahrnehmung. Bernecker argumentiert komplementär dazu in *Knowing the World by Knowing One's Mind* für die Prämisse (5). Bernecker beklagt dort das Fehlen eines intern zugänglichen rationalisierenden Grundes, wenn er sagt: „Just by looking inside, I cannot decide whether or not my so-called thoughts have content or are contentless states.“ (Bernecker 2000, 12) Ich halte Dretskes und Berneckers Argumente für die Prämissen (4) und (5) für überzeugend. Und auch die Prämissen (2) und (3) haben eine überwältigende Plausibilität.

Angreifbar ist jedoch die Prämisse (1). Wenn man *erkenntnistheoretischer Externalist* ist, wozu sich Dretske und Bernecker ja offiziell bekennen, dann erfordert Wissen keine rationalisierenden Gründe. Wenn wir beispielsweise Dretskes Wissensdefinition zugrunde legen, dann erfordert das Wissen, dass p, die Information, dass p. Diese Information, dass p, ist jedoch nicht notwendigerweise ein rationalisierender Grund, weil die für einen solchen rationalisierenden Grund nötige inferentielle Beziehung nur zwischen Zuständen mit

⁷ Dieses Argument findet sich explizit weder bei Bernecker noch bei Dretske. Es handelt sich vielmehr um meinen Versuch einer rationalen Rekonstruktion. Am ehesten entspricht ihr vielleicht Berneckers „zentrales Argument“ in Bernecker 2000, 12.

repräsentationalem Gehalt bestehen kann. Ein Zustand kann jedoch Information über eine bestimmte Tatsache tragen, ohne diese Tatsache zu repräsentieren.⁸ Es ist ebenso möglich, dass ein Zustand eine bestimmte Tatsache repräsentiert, ohne Information über sie zu tragen (besonders deutlich im Fall der Fehlrepräsentation). Information und Repräsentation sind voneinander wechselseitig unabhängig. Allerdings kann die für Wissen erforderliche Information auch von einer Repräsentation getragen werden. Dretskes Wissensdefinition ist also mit der Existenz rationalisierender Gründe verträglich, aber sie verlangt solche Gründe nicht unbedingt.

Das erste Argument kann also nicht überzeugen, aber es gibt daneben noch ein weiteres, das Hauptargument für die DB-These. Ich möchte es als „informationstheoretisches Argument“ bezeichnen. Am besten ausbuchstabiert ist es in Dretskes Aufsatz *Externalism and Self-Knowledge*. Das Argument lautet in meiner Rekonstruktion wie folgt:

ARGUMENT II

- (8) Wissen erfordert Überzeugungen, die auf ihnen entsprechende Information gestützt sind.
- (9) In einem repräsentationalen System gibt es keine für introspektives Wissen relevante Information, *dass* dieses System repräsentiert. (Es gibt nur – introspektiv relevante - Informationen darüber, *was* dieses System repräsentiert.)

- (10) Ein repräsentationales System kann kein introspektives Wissen davon haben, dass es repräsentiert.

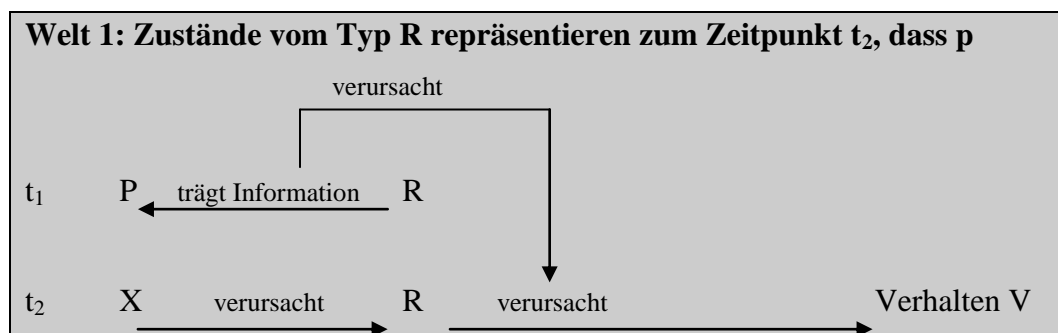
Die Prämisse (8) drückt natürlich Dretskes informationstheoretischen Wissensbegriff aus, den ich hier nicht weiter diskutieren möchte. Doch warum ist Prämisse (9) wahr? Nun, es gibt interne Informationen darüber, *was* das repräsentationale System repräsentiert. Wenn das System nämlich etwas anderes repräsentieren würde, als es tatsächlich repräsentiert, dann würde in ihm ein anderer intern individuierter Realisierer von repräsentationalem Gehalt auftreten. Der repräsentationalen Gehalt tragende Zustand selbst trägt also Information darüber, *was* repräsentiert wird. Es gibt jedoch nach Dretske keine intern verfügbare Information darüber, *dass* das repräsentationale System repräsentiert. Denn wenn der

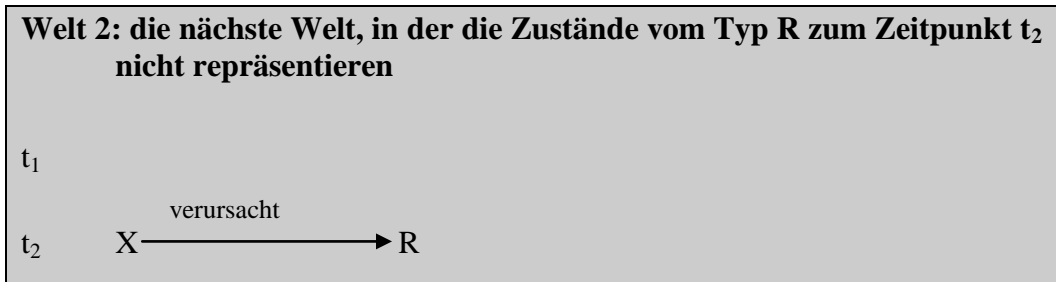
⁸ Informationeller Gehalt ist nämlich nicht hinreichend für repräsentationalen Gehalt, weil ein Zustand nach Dretske einen Sachverhalt nur dann repräsentiert, wenn er die Funktion hat, Information über ihn zu tragen. Ein Zustand kann jedoch Information tragen, ohne die Funktion zu haben, diese Information zu tragen. Ein Beispiel: Baumringe tragen Information über das Alter des jeweiligen Baumes, aber sie haben normalerweise keinerlei Funktion, diese Information zu tragen.

Gehaltsexternalismus richtig ist, dann würden dieselben internen Zustände, die aktual repräsentationalen Gehalt haben, auch dann auftreten, wenn sie keinen repräsentationalen Gehalt hätten, weil keinerlei externe gehaltskonstitutive Faktoren bestehen. Der repräsentationale Zustand trägt also keine Information darüber, *dass* er repräsentiert (Dretske 2003b, 136f).

Das wird etwas deutlicher, wenn man sich Dretskes Erklärung repräsentationalen Gehalts ansieht. Danach bekommt ein System repräsentationale Eigenschaften, wenn seine Zustände die *Funktion* haben, Informationen über bestimmte Eigenschaften zu tragen. Diese Funktion können Vorkommnisse des Typs R entweder – im Falle von Instrumenten – durch die Intention der Designer bzw. Benutzer oder – im Falle von kognitiven Systemen – auf natürlichem Wege erwerben. Letzteres geschieht, indem die informational Eigenschaften der Zustände vom Typ R historisch verursacht haben, dass Nachfolger vom selben Typ R im kognitiven System eine bestimmte funktionale Rolle übernehmen (Vgl. dazu und zum Folgenden Dretske 1988, 52). Ein Beispiel: Wenn neuronale Zustände vom Typ R im Gehirn eines Frosches in der Vergangenheit die Information getragen haben, dass gerade eine Fliege vorbeifliegt, dann war ein durch diese Zustände ausgelöstes Fangverhalten des Frosches erfolgreich. Das hat (als so genannte *strukturierende* Ursache) evolutionär dazu geführt, dass Nachfolgerzustände vom Typ R immer dieses Fangverhalten auslösen. Wenn das so ist, dann erwerben Zustände vom Typ R den Gehalt, dass eine Fliege gerade vorbeifliegt, und haben ihn (im Falle einer Fehlrepräsentation) auch dann, wenn die Zustände dieses Typs tatsächlich nicht durch eine vorbeifliegende Fliege ausgelöst werden.

Wenn das die richtige Erklärung des repräsentationalen Gehalts ist, dann ist leicht einsehbar, warum die repräsentierenden Zustände nicht die Information tragen, dass sie repräsentieren. Zustände vom Typ R würden nämlich auch dann auftreten, wenn sie keinen repräsentationalen Gehalt hätten, weil die gehaltserklärende Vorgeschichte nicht eingetreten wäre.





Nun zu meinem Einwand. Ich glaube, Dretske hat Recht damit, dass die repräsentationalen Zustände selbst keine Information darüber tragen, dass sie (oder das System, um dessen Zustände es sich handelt) repräsentieren. *Aber daraus folgt nicht*, dass es keinerlei Information im System darüber gibt, *dass* das repräsentationale System repräsentiert. Es gibt nämlich andere Zustände im System, die genau diese Information tragen, und zwar die Tatsache, dass die durch ihre syntaktische Form individuierten Träger des repräsentationalen Gehalts Realisierer einer ganz bestimmten funktionalen Rolle sind. Das sollte aus dem Schaubild unmittelbar hervorgehen. Wenn die Zustände, die aktuell repräsentationalen Gehalt haben, keinen solchen Gehalt hätten, dann würden auch die gehaltskonstitutiven historischen Faktoren nicht bestehen. Doch wenn diese Faktoren nicht bestünden, dann würde auch die strukturierende Ursache für die funktionale Rolle dieser Zustände wegfallen (wie in Welt 2). Es gilt folglich: Wenn die syntaktisch individuierten Zustände, die aktuell repräsentieren, keinen repräsentationalen Gehalt hätten, dann hätten sie nicht die funktionale Rolle, die sie aktuell haben. Also trägt die Tatsache, dass sie diese funktionale Rolle haben, die Information, dass die Zustände repräsentieren. Contra Dretske und Bernecker gibt es also interne Zustände des Systems, die die Information tragen, dass das System repräsentiert.

Eine ergänzende Bemerkung am Rande. In einem früheren Aufsatz mit dem Titel *Die Grenzen des Selbstwissens* hat Bernecker auch Zweifel daran angemeldet, dass wir introspektiv wissen können, mit welcher Einstellung wir repräsentieren – ob wir also eine Überzeugung haben, eine Wahrnehmung, einen Wunsch oder einen repräsentationalen Gehalt bloß erwägen (Bernecker 1997). Doch meines Erachtens gibt es auch interne Tatsachen, die Information darüber tragen, *mit welcher Einstellung* ein bestimmter Sachverhalt repräsentiert wird. Die Tatsache, dass ein syntaktisch individuiertes Zustand eines bestimmten Typs in einem bestimmten Modul, sagen wir dem Überzeugungsmodul, auftritt (und nicht in einem anderen, beispielsweise dem Wahrnehmungsmodul) trägt die Information, dass der

entsprechende repräsentationale Gehalt *geglaubt* und nicht *wahrgenommen* wird. Hätte nämlich keine Überzeugung sondern eine Wahrnehmung vorgelegen, dann wäre der (syntaktisch individuierte) repräsentationale Zustand nicht im Überzeugungs-, sondern im Wahrnehmungsmodul aufgetreten.⁹

Nun scheint es jedoch gleich auf den ersten Blick eine vernichtende Replik auf meinen Vorschlag zu geben, dass die funktionale Rolle eines repräsentationalen Zustandes die Information tragen soll, dass dieser Zustand repräsentiert. Könnten nicht Systeme auftreten, deren Zustände dieselbe funktionale Rolle hätten wie unser repräsentationales System, ohne dass deren funktionale Architektur durch gehaltskonstitutive Faktoren verursacht wäre? Kurz: Könnte es nicht funktionale Duplikate von uns geben, die keinen Geist hätten? Gerade der Gehaltsexternalismus scheint doch eine solche Möglichkeit zuzulassen. Davidsons Sumpfmännchen dramatisiert diese Möglichkeit auf anschauliche Weise.¹⁰ Warum sollte es nicht prinzipiell möglich sein, dass ein System mit der gleichen internen Beschaffenheit und funktionalen Verschaltung wie wir, durch einen Blitzschlag in den Morast eines Sumpfes plötzlich entsteht? Dieses System hätte aus der Perspektive des Externalismus nicht die für repräsentationalen Gehalt erforderliche Vorgeschichte. Ein funktionales Duplikat ohne Geist scheint also möglich. Auf diese Möglichkeit berufen sich Dretske und Bernecker auch in ihrer Argumentation (Bernecker 2000, 17f; Dretske 2003b, 135).

Meine Antwort darauf lautet: Ja, solche Szenarien sind selbstverständlich möglich. Aber sie sind nur in Welten realisiert, die für die Bewertung informationaler Eigenschaften *irrelevant* sind. Ein Signal *r* trägt die Information, dass *p*, wenn in der *nächsten* Welt, in der *p* und alle logisch und kausal damit zusammenhängenden Tatsachen *nicht* der Fall gewesen wären, dieses Signal nicht aufgetreten wäre. Genau diese Bedingung wird jedoch von der funktionalen Rolle erfüllt. In der nächsten Welt, in der die gehaltskonstitutiven Ursachen für die funktionale Rolle des Realisierers entfallen, gibt es keine alternative Ursache für die

⁹ Hier liegt nun vielleicht der folgende Einwand nahe: Das Auftreten des (syntaktisch individuierten) Trägers eines bestimmten repräsentationalen Gehalts in einem bestimmten (durch seine funktionale Architektur individuierten) kognitiven Modul kann keine Information über die Einstellung zu einem bestimmten Gehalt tragen, weil diese Tatsache die Einstellung konstituiert. Nach Dretske kann jedoch kein Zustand Information über sich selbst tragen. *Antwort*: Streng genommen wird die Tatsache, dass eine bestimmte Einstellung zu einem bestimmten repräsentationalen Gehalt vorliegt, nicht durch das Auftreten eines syntaktisch individuierten Tokens in einem Modul konstituiert. Der repräsentationale Gehalt ist extern und eine Einstellung (wie Glauben, Wünschen, Wahrnehmen, Hoffen etc.) zu ihm liegt auch erst vor, wenn ein bestimmter repräsentationaler Gehalt existiert, auf den die Einstellung bezogen ist. Auch das Bestehen der Einstellung selbst ist also eine extern individuierte Tatsache, über die eine interne Tatsache (das Auftreten des syntaktisch individuierten Tokens in einem Modul) sehr wohl Information tragen kann.

¹⁰ Vgl. dazu Davidson 1986.

Entstehung dieser Rolle des Realisierers.¹¹ Anders gesagt: Sumpfmänn-Welten sind zwar mögliche Welten, das möchte ich nicht bestreiten, aber es handelt sich um sehr weit entfernte, exotische Welten. Was müsste nicht alles anders sein, als in der aktuellen Welt, damit tatsächlich ein funktionales Duplikat von uns durch einen Blitzschlag in den Morast eines Sumpfes entstehen würde. Nach Dretskes Informationstheorie gefährdet es den Informationsgehalt eines Signals jedoch nicht, wenn in sehr weit entfernten Welten das Signal auch dann auftreten würde, wenn die Tatsache, über die das Signal Information trägt, nicht bestehen würde. Sonst könnten auch perzeptuelle Zustände keine Information über Tatsachen in der materiellen Außenwelt tragen, denn in sehr weit entfernten Welten könnten sie auch (ohne materielle Außenwelt) durch einen bösen Dämon hervorgerufen werden. Gerade diese Konsequenz lehnt Dretske jedoch ab. Relevant sind also nur die nächsten nicht-p Welten, und in denen würde die funktionale Rolle nicht ohne die gehaltskonstitutiven Faktoren bestehen. Soweit meine Erwiderung auf den Einwand.

Es gibt jedoch noch einen weiteren Einwand gegen meinen Vorschlag: Es mag schon sein, dass die funktionale Rolle der repräsentationalen Zustände eines Systems u.a. die Information trägt, dass diese Zustände repräsentieren. Doch warum sollten Überzeugungen, die auf dieser Information beruhen, Instanzen *introspektiven* Wissens sein? Dagegen scheint u.a. Folgendes zu sprechen: Wenn der Gehaltsexternalismus richtig ist, dann beinhaltet die Information, dass die Zustände repräsentieren, eine Information über externe gehaltskonstitutive Tatsachen. Kann ein solches Wissen über die Außenwelt tatsächlich introspektiv sein?

Um diese Frage zu beantworten, muss man sich zunächst einmal klar machen, dass Quellen des Wissens – wie Wahrnehmung, Introspektion oder Erinnerung – nicht durch die Gegenstandsbereiche des Wissens individuiert werden. Man kann also nicht einfach sagen, dass Wahrnehmungswissen dadurch definiert ist, dass es von der Außenwelt handelt, introspektives Wissen dadurch definiert ist, dass es von der Innenwelt handelt, und Erinnerungswissen dadurch definiert ist, dass es von der Vergangenheit handelt. Es ist offensichtlich möglich, dass dieselben Gegenstände mit Hilfe verschiedener Quellen des Wissens zugänglich sind. So kann man beispielsweise von seiner Aufregung vor einer Prüfung introspektiv wissen, man kann sie sich aber auch durch Beobachtung seines eigenen Verhaltens erschließen oder auf sie durch Dritte aufmerksam gemacht werden. Der

¹¹ Um Missverständnisse zu vermeiden: Genau genommen trägt nicht die funktionale Rolle als solche die besagte Information, sondern die *Tatsache*, dass dieser (syntaktisch individuierte) Realisierer diese Rolle hat. Das meine ich streng genommen immer dann, wenn ich sage, dass die funktionale Rolle diese Information trägt.

introspektive Charakter meines Wissens davon, dass ich Gedanken und Bewusstsein habe, wird also nicht allein dadurch unterminiert, dass die gewussten Tatsachen Implikationen hinsichtlich der Außenwelt haben.

Doch um die introspektive Natur meines Wissens, dass ich repräsentiere, zu verteidigen, bedarf es mehr. Ich muss erklären, was den eigentümlichen Charakter der Introspektion ausmacht. Meines Erachtens zeichnet sich introspektives Wissen im Unterschied zu perzeptuellem Wissen dadurch aus, dass es nicht auf rationalisierenden Gründen beruht. Im Fall von perzeptuellem Wissen, beispielsweise meinem visuellen Wissen, dass sich vor mir ein Rechteck befindet, habe ich für meine Überzeugung, dass sich vor mir ein solches Rechteck befindet, einen Grund, nämlich mein visuelles Erlebnis, dessen repräsentationaler Gehalt meine Überzeugung inferentiell rationalisiert. Ich sehe, dass vor mir etwas Rechteckiges ist, und dieser Grund rationalisiert meine Überzeugung, dass vor mir etwas Rechteckiges ist (Vgl. dazu auch Grundmann 2004). Ganz anders verhält es sich im Fall der Introspektion. Damit Wissen besteht, muss zwar eine Information darüber vorliegen, was, wie oder dass repräsentiert wird. Aber diese Information ist uns nicht in Form einer Repräsentation gegeben, sondern durch eine nicht-repräsentationale (und unbewusste) Tatsache wie die funktionale Rolle. Introspektives Wissen stützt sich nicht auf rationalisierende Gründe. Es ist nicht-inferentiell. Das macht die besondere Eigenart dieses Wissens aus.

Wenn also die für mein Wissen, *dass* ich repräsentiere, erforderliche Information nicht durch eine perzeptuelle Repräsentation gegeben ist (sondern durch die funktionale Rolle), dann lässt sich die These verteidigen, dass es sich um introspektives Wissen handelt. Selbstverständlich wird dadurch nicht die Möglichkeit ausgeschlossen, auch auf perzeptuell-empirischem Wege von dieser Tatsache zu wissen. Ein solches empirisches Wissen von meinem Geist würde auf einer Theorie beruhen, die auf die Beobachtung meiner Interaktionen mit der Umwelt gestützt wäre.¹²

Wenn meine Charakterisierung introspektiven Wissens richtig ist, lässt sich auch nicht einwenden, dass uns die spezifischen funktionalen Rollen unserer repräsentationalen Zustände zumindest introspektiv gar nicht zugänglich sind. Denn die für unser introspektives Wissen relevanten Informationen müssen nur vorliegen. Sie müssen nicht ihrerseits introspektiv

¹² Aber es ist vollkommen unplausibel, dass wir nur durch Beobachtung unserer Umwelt oder durch Dritte davon wissen können, wie Dretske meint (vgl. Dretske 2003b, 140f.)

bewusst sein. Wenigstens kann man eine solche Forderung nicht aus Dretskes Wissensdefinition herleiten.

So wie es aussieht, kann auch der Externalist bezüglich des mentalen Gehalts das introspektive Selbstwissen beanspruchen, dass man einen Geist hat und kein geistloser Zombie ist. Daran ändert Dretskes informationstheoretischer Wissensbegriff nichts. Wie verhält sich die Sache nun aus der Sicht Nozicks? Dieser hat in den achtziger Jahren die viel beachtete „truth-tracking“ Definition des Wissens präsentiert (Vgl. Nozick 1981, Kap. 3). Nach dieser Definition weiß ein Subjekt S, dass p, genau dann, wenn das Subjekt die Überzeugung hat, dass p, wenn diese Überzeugung wahr ist und wenn zwei weitere Bedingungen erfüllt sind: die Sensitivitätsbedingung (sensitivity) und die Stabilitätsbedingung (stability). Die Sensitivitätsbedingung (das eigentliche „truth-tracking“) besagt, dass S nicht glauben würde, dass p, wenn p nicht der Fall wäre. Es muss also – anders als bei Dretske – keine sensitiven Gründe für eine Überzeugung geben, es genügt, wenn die Überzeugung selbst sensitiv ist. Die Stabilitätsbedingung besagt darüber hinaus, dass S auch dann glauben würde, dass p, wenn p auch unter geringfügig veränderten Hintergrundsbedingungen der Fall wäre.

Nehmen wir an, Nozicks Wissensdefinition sei korrekt und der Gehaltsexternalismus würde gelten. Kann die Überzeugung (b) von S, *dass er selbst einen Geist hat bzw. repräsentiert*, die Wissensbedingung erfüllen? Nun, wenn diese Überzeugung vorliegt, dann ist sie automatisch wahr, weil S einen Geist hat, wenn er irgendeine Überzeugung hat. Die Überzeugung ist also selbstverifizierend. Die ersten beiden Bedingungen sind damit erfüllt. Selbstverifizierende Überzeugungen sind jedoch automatisch auch sensitiv. Wenn das, was die Überzeugung (b) besagt, falsch wäre (weil die externen für den geistigen Bedeutungsgehalt konstitutiven Faktoren generell nicht vorliegen), dann gäbe es auch die Überzeugung (b) nicht. In einem Zombie gibt es nämlich keinerlei repräsentierende Zustände, nicht einmal die Überzeugung, dass es solche Repräsentationen gibt. Die Sensitivitätsbedingung ist also ebenfalls erfüllt. Doch auch die Stabilitätsbedingung kann erfüllt werden. Das ist zumindest dann der Fall, wenn unsere geistigen Zustände selbstpräsentierend sind und automatisch die Überzeugung hervorrufen, dass repräsentiert wird. Nach Nozicks Wissensdefinition ist es also leicht zu wissen, dass man einen Geist hat.

Doch wie ist es mit der Überzeugung (c) zweiter Stufe, dass ich gerade irgendwelche Gedanken erster Stufe habe? Diese Überzeugung ist offensichtlich nicht selbstverifizierend,

weil sie nicht selbstbezüglich ist. Wenn (c) in einem kognitiven System mit repräsentationalen Zuständen erster Ordnung auftritt, dann sind Nozicks erste beiden Wissensbedingungen erfüllt. Was ist jedoch mit den anderen Bedingungen? Nehmen wir an, im kognitiven System gibt es einen kausalen Mechanismus, so dass die Realisierer von Gedanken erster Stufe immer Realisierer der Überzeugung (c) hervorbringen. Nehmen wir ferner an, Dretskes Erklärung des repräsentationalen Gehalts sei korrekt. Dann ist die Überzeugung (c), dass ich gerade irgendwelche Gedanken erster Stufe habe, nur möglich, wenn die typidentischen Vorgänger des Realisierers dieser Überzeugung Information über das gegenwärtige Vorliegen von Gedanken erster Stufe getragen haben. Die Realisierer der Gedanken erster Stufe müssen also zumindest in der Vergangenheit repräsentationalen Gehalt gehabt haben. Doch wenn der historische Externalismus korrekt ist, haben sie diesen Gehalt permanent. Das ist eine Konsequenz der für die Überzeugung (c) gehaltskonstitutiven Faktoren. Hätten die Realisierer meiner aktuellen Gedanken erster Ordnung also gegenwärtig keinen repräsentationalen Gehalt (weil die externen Bedingungen nicht erfüllt sind), dann hätte der Realisierer meiner Überzeugung (c) auch nicht die Bedeutung, dass ich gerade irgendwelche Gedanken erster Stufe habe. Damit scheint die Sensitivitätsbedingung erfüllt zu sein. Doch wenn, wie angenommen, auch ein Mechanismus besteht, so dass Realisierer von Gedanken erster Stufe immer Realisierer der Überzeugung (c) hervorbringen, dann ist auch die Stabilitätsbedingung erfüllt.

Aus meinen Überlegungen ergibt sich das folgende Gesamtbild. Solange wir annehmen, dass Wissen internalistische Bedingungen (wie rationalisierende Gründe) enthält, hat der Externalismus hinsichtlich des mentalen Gehalts weitreichende skeptische Konsequenzen hinsichtlich der Möglichkeit von introspektivem Selbstwissen. Erkenntnistheoretische Externalisten können hingegen daran festhalten, dass wir introspektiv wissen, *was* wir denken, *wie* wir es denken und *dass* wir überhaupt Gedanken haben, und zwar auch dann, wenn der Externalismus hinsichtlich des mentalen Gehalts korrekt ist. Aus der Perspektive des erkenntnistheoretischen Externalismus ist der Gehaltsexternalismus also vollkommen verträglich mit der cartesianischen Grundintuition. Ich hoffe, dass meine Diskussion der Dretske-Bernecker These vor dem Hintergrund verschiedener externalistischer Wissenskonzeptionen dies gezeigt hat.

Ein durchgängiger Externalismus des Gehalts und der Erkenntnis scheint also mit der cartesianischen Grundintuition vollkommen verträglich zu sein. Wenn es für eine Position

spricht, dass sie mit unseren grundlegenden Intuitionen verträglich ist, dann stärkt das Ergebnis meiner Überlegungen die Position des Gehaltsexternalismus in nicht unerheblichem Maße. Ganz besonders natürlich dann, wenn man annimmt, dass der erkenntnistheoretische Externalismus aus unabhängigen Gründen plausibel ist. Ich bin dieser Auffassung. Aber das wäre eine andere Geschichte.^{13, 14}

¹³ Vgl. dazu meine ausführliche Verteidigung des erkenntnistheoretischen Externalismus in Grundmann 2003.

¹⁴ Ich möchte insbesondere Sven Bernecker und Fred Dretske für ihre kritischen Bemerkungen danken. Mein Dank gilt auch Frank Hofmann, Joachim Horvath, Gottfried Vosgerau und Woldai Wagner für ihre Anmerkungen zu früheren Fassungen dieses Aufsatzes.

Literatur

- Barke, A. 2002, *The Closure of Knowledge in Context*, mentis, Paderborn.
- Bernecker, S. 1997, „Die Grenzen des Selbstwissens“, *Zeitschrift für philosophische Forschung* **Band 51**, S. 216-30.
- Bernecker, S. 2000, „Knowing the world by knowing one’s mind“, *Synthese* **Band 123**, S. 1-34.
- Chalmers, D. 1996, *The Conscious Mind. In Search of a Fundamental Theory*, University Press, Oxford.
- Davidson, D. 1986, „Knowing One’s Own Mind“, *Proceedings and Addresses of the American Philosophical Association* **Band 60**, S. 441-458.
- Dretske, F. 1971, „Conclusive Reasons“, *Australasian Journal of Philosophy* **Band 49**, S. 1-21.
- Dretske, F. 1981, *Knowledge and the Flow of Information*, MIT Press, Cambridge, MA.
- Dretske, F. 1988, *Explaining Behavior*, MIT Press, Cambridge, MA.
- Dretske, F. 2003a, “How do you know you are not a zombie?”, in B. Gertler (ed.), *Privileged Access*, Ashgate, Madison, S. 1-13.
- Dretske, F. 2003b, “Externalism and Self-Knowledge”, in S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge*, MIT Press, Cambridge, MA, S. 131-42.
- Dretske, F. 2004, „Externalism and Modest Contextualism“, *Erkenntnis* **Band 61**.
- Grundmann, T. 2003, *Der Wahrheit auf der Spur. Eine Verteidigung des erkenntnistheoretischen Externalismus*, mentis, Paderborn.
- Grundmann, T. 2004, „Perceptual Representations as Basic Reasons“, in R. Schumacher (ed.), *Perception and Reality. From Descartes to the Present*, mentis, Paderborn, S. 287-304.
- Jaeger, C. 2004, „Skepticism, Information, and Closure: Dretske’s Theory of Knowledge“, *Erkenntnis* **Band 61**.
- Klein, P. 1981, *Certainty: A Refutation of Scepticism*, University of Minnesota Press, Minneapolis.
- Nozick, R. 1981, *Philosophical Explanations*, University Press, Oxford.
- Shoemaker, S. 1996, “Self-Knowledge and ‘Inner Sense’”, in ders., *The First-Person Perspective and Other Essays*, University Press, Cambridge, S. 201-68.