

- Shadlen, Michael/Newsome, William (2001): Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. In: *Journal of Neurophysiology* 86, 1916–1936.
- Shadmehr, Reza/Mussa-Ivaldi, Fernando (1994): Adaptive representation of dynamics during learning of a motor task. In: *Journal of Neuroscience* 14, 3208–3224.
- Shenoy, Krishna/Sahani, Manesh/Churchland, Mark (2013): Cortical control of arm movements. In: *Annual Review of Neuroscience* 36, 337–359.
- Skinner, Burrhus (1953): *Science and Human Behavior*. New York. [dt.: *Wissenschaft und menschliches Verhalten*. Stuttgart 1973]
- Stanford, Terrence/Shankar, Swetha/Massoglia, Dino/Costello, Gabriela/Salinas, Emilio (2010): Perceptual decision making in less than 30 milliseconds. In: *Nature Neuroscience* 13, 379–385.
- Todorov, Emanuel/Jordan, Michael (2002): Optimal feedback control as a theory of motor coordination. In: *Nature Neuroscience* 5, 1226–1235.
- Tolman, Edward (1948): Cognitive maps in rats and men. In: *Psychological Review* 55, 189–208.
- Uno, Yoji/Kawato, Mitsuo/Suzuki, Ryoji (1989): Formation and control of optimal trajectories in human multijoint arm movements. In: *Biological Cybernetics* 61, 89–101.
- Wohlschläger, Andreas (2000): Visual motion priming by invisible actions. In: *Vision Research* 40, 925–930.
- Wolpert, Daniel/Miall, Chris/Kawato, Mitsuo (1998): Internal models in the cerebellum. In: *Trends in Cognitive Sciences* 2, 338–347.

Constantin A. Rothkopf

16. Repräsentation

Einleitung

Wir gehen täglich mit Texten, gesprochener Sprache, Stadtplänen und Landkarten, Diagrammen, Bildern, Filmen, Piktogrammen und Ähnlichem um. Bei jedem davon handelt es sich um eine Repräsentation, die für etwas anderes steht – ein Aussagesatz etwa steht für einen Sachverhalt, ein Stadtplan für die räumliche Struktur einer Stadt usw. Repräsentationen sind Träger semantischer Eigenschaften, d. h. sie haben Bedeutung (s. Kap. IV.20). Es gibt verschiedene Arten von Repräsentationen: Aussagen sowie generell sprachliche Ausdrücke sind sprachliche Repräsentationen, Verkehrszeichen, Piktogramme usw. sind konventionelle Repräsentationen und Fotografien sowie die Anzeige der Tankuhr im Auto sind ebenfalls Repräsentationen, wenn auch keine sprachlichen oder konventionellen. Repräsentationen wie diese sind *externe* Repräsentationen, die prinzipiell vielen Individuen zugänglich sind. In der Kognitionswissenschaft hingegen spielen v. a. *interne* oder »mentale« Repräsentationen eine zentrale Rolle, d. h. Repräsentationen, die wir (in erster Annäherung und ganz grob gesagt) »im Kopf« haben (s. Kap. IV.9), die sich aber dennoch ebenso wie externe Repräsentationen durch ihre semantischen Eigenschaften auszeichnen, also z. B. etwas bedeuten. Beispiele interner Repräsentationen wären etwa Gedanken, Vorstellungen, Wahrnehmungen oder Wünsche, aber auch unbewusst vorliegende Repräsentationen von Bewegungsmustern oder vom Aufbau des eigenen Körpers. Verstehen wir den Begriff der Repräsentation so, dann ist die Kognitionswissenschaft in gewisser Weise die Wissenschaft interner Repräsentationen. Während manche es vorziehen, aufgrund der Vielschichtigkeit des Phänomens verschiedene Repräsentationsbegriffe zu unterscheiden, werden wir im Folgenden versuchen, die grundlegende Gemeinsamkeit aller Verwendungsweisen herauszuarbeiten, um auf diese Weise ein Grundverständnis des Begriffs der Repräsentation zu präsentieren, das den verschiedenen Disziplinen der Kognitionswissenschaft zugrunde gelegt werden kann, wobei es naturgemäß im Wesentlichen um interne Repräsentationen gehen wird.

Zunächst soll geklärt werden, warum wir überhaupt von internen Repräsentationen sprechen: Warum schreiben wir kognitiven Systemen mentale Repräsentationen zu? Die Beantwortung dieser Frage führt zu einem Grundverständnis der zentra-

len Idee von Repräsentation, das es ermöglicht, verschiedene Eigenschaften von Repräsentationen zu diskutieren. Diese genauere Charakterisierung von Repräsentationen bildet die Grundlage für die Diskussion der Frage, ob Repräsentationen für eine Theorie des Geistes oder der Kognition überhaupt notwendig sind. Anschließend wird das Verhältnis von Repräsentationen zum Körper und der Umwelt eines kognitiven Systems erörtert und auf zwei zentrale Fragenkomplexe eingegangen: (1) Wie werden mentale Repräsentationen individuiert, d. h. wie wird festgelegt, ob es sich bei zwei verschiedenen Vorkommnissen um Vorkommnisse ein und derselben Repräsentation oder um Vorkommnisse zweier verschiedener Repräsentationen handelt? (2) Wie können verschiedene Arten von Repräsentationen unterschieden werden und welche Arten von Repräsentationen sollten sinnvollerweise unterschieden werden? Abschließend wird auf den gegenwärtigen und zukünftigen Stellenwert von Repräsentationen innerhalb der Kognitionswissenschaft eingegangen.

Warum sprechen wir von mentalen Repräsentationen?

Warum wir in der Kognitionswissenschaft von Repräsentationen sprechen und welche zentrale Rolle diesem Begriff zukommt, lässt sich am besten an einem der ersten Experimente verdeutlichen, die die ›Geburt‹ der Kognitionswissenschaft einläuteten. In der ersten Hälfte des 20. Jh.s war in der Psychologie v. a. in den USA der sog. Behaviorismus die dominierende Strömung, demzufolge alle Verhaltensweisen idealerweise durch Konditionierung, also durch den Aufbau starrer Reiz-Reaktions-Verknüpfungen, zu erklären seien. Als Edward Tolman mit seinen Mitarbeitern jedoch eine ganze Reihe von Experimenten durchführte, in denen Ratten darauf trainiert wurden, in einem Labyrinth Futter zu finden, zeigten diese Verhaltensweisen, die nicht im Sinne des Behaviorismus zu erklären waren, sondern die Postulierung interner Repräsentationen erforderten. In einem der klassischen Experimente (Tolman 1948) lernten Ratten, in einem Labyrinth den kürzesten von drei unterschiedlich langen Wegen zum Ziel (dem Futter) zu benutzen. Nach der Trainingsphase wurde dieser Weg kurz vor dem Ziel so blockiert, dass gleichzeitig auch der mittellange Weg blockiert war. Die Ratten rannten bis zur Blockade, kehrten dann zum Start zurück und entschieden sich sofort für die (einzig mögliche) lange Variante, ohne den mittellangen Weg auch nur auszuprobieren. Dieses Verhalten

lässt sich durch die einfachen und starren Reiz-Reaktions-Schemata des Behaviorismus nicht erklären. Die Ratten schienen vielmehr eine interne Repräsentation des Labyrinths aufgebaut zu haben, die es ihnen erlaubte, sofort zu bemerken, dass aufgrund der Anordnung der beiden Wege auch der mittellange Weg blockiert war. Sie hatten, wie Tolman es nannte, eine ›mentale Karte‹ (*mental map*), also eine interne Repräsentation des Labyrinths, erworben.

Die Grundidee hinter der Einführung von Repräsentationen lässt sich durch ein anderes Beispiel weiter verdeutlichen: Bestimmte Wüstenameisen sind in der Lage, nach einer unsystematischen Suche nach Futter auf einer geraden Linie zu ihrem Nest zurückzukehren (Wehner 1999), auch wenn sie zu diesem Zeitpunkt keinerlei Sinnesinformationen über seinen Ort erhalten, es also weder riechen noch sehen können. Offenbar muss es etwas in ihnen geben, das es ihnen erlaubt, einen Ort zu finden, den sie momentan nicht direkt wahrnehmen können: Sie müssen über eine interne Repräsentation des Ortes des Nestes verfügen. Die Ameisen sind also in der Lage, von jedem beliebigen Punkt ihres Suchpfades direkt zu ihrem Nest zurückzukehren, weil sie während ihrer Suche permanent über eine Repräsentation des Ortes ihres Nestes verfügen, so dass sie zu jedem Zeitpunkt ihrer Reise wissen, in welche Richtung und wie viele Schritte sie laufen müssen, um ihr Nest zu erreichen. Diese Repräsentation kann falsch sein: Wenn z. B. ein Wissenschaftler eine Ameise nach ihrer Nahrungssuche einen Meter nach Osten versetzt, dann wird sie ›glauben‹, ihr Nest sei einen Meter weiter östlich und dorthin laufen.

Während die Repräsentation der Ameise auch dann als interner Stellvertreter für den Ort ihres Nestes fungieren kann, wenn er ihr aktuell nicht sinnlich zugänglich ist, umkreisen Motten eine Lichtquelle nur solange sie angeschaltet ist; schaltet man sie ab, beenden die Motten ihren Kreisflug. Anders als bei den Ameisen muss zur Erklärung des Verhaltens der Motten keine interne Repräsentation der Lichtquelle postuliert werden, und anders als die Ameisen können sich die Motten aus diesem Grund auch niemals über den Ort der Lichtquelle täuschen: Sie verfügen überhaupt nicht über eine interne Repräsentation, die stellvertretend für die Lichtquelle steht und es ihnen erlaubte, sie auch dann weiter anzusteuern, wenn sie nicht mehr sichtbar ist – genau das wäre aber die Voraussetzung dafür, dass sie den Ort der Lichtquelle überhaupt falsch repräsentieren, d. h. sich irren, könnten.

In Beispielen wie diesen zeigt sich ein ›liberales‹ Bild von Repräsentationen, wonach Repräsentatio-

nen jene (internen) Stellvertreter von Gegebenheiten in der Welt sind, die zur Erklärung flexiblen, intelligenten Verhaltens herangezogen werden müssen, das nicht (wie beim Kreisflug der Motte) in starren Verhaltensmustern besteht, sondern (wie bei der räumlichen Orientierung der Ameise) flexible Reaktionen auf einen Stimulus erlaubt bzw. sich auf ein nicht wahrgenommenes Objekt richtet (natürlich sagt die bloße Postulierung interner Repräsentationen noch nichts darüber aus, wie genau diese aussehen und funktionieren – um das herauszufinden, müssen v. a. empirische Fragen gestellt werden, mit denen sich die Kognitionswissenschaft beschäftigt und die mit zunehmender Komplexität des jeweiligen Verhaltens zunehmend schwieriger zu beantworten sind). Legt man ein solches liberales Verständnis von Repräsentation zugrunde, dann besagt die These des *Repräsentationalismus* lediglich, dass flexibles Verhalten u. a. auf dem Vorliegen interner Zustände beruht, die eine Stellvertreterfunktion für andere Dinge übernehmen.

Aufbauend auf Experimenten wie denen von Tolman entwickelte sich der Begriff der Repräsentation zu einem der zentralen Begriffe der Kognitionswissenschaft, weil er letztendlich das entscheidende Element in der Erklärung von flexiblem Verhalten darstellte. Das ist bis heute so geblieben, auch wenn der Repräsentationsbegriff in unterschiedlichen Disziplinen inzwischen unterschiedlich gebraucht wird. So spricht man etwa in den Verhaltenswissenschaften immer dann von mentalen Repräsentationen, wenn eine Information intern ›gespeichert‹ wird, um Verhalten auch unabhängig von der Wahrnehmung der betreffenden Eigenschaft zu steuern. In den Neurowissenschaften (s. Kap. II.D) steht der Repräsentationsbegriff für bestimmte Aktivitätsmuster im Gehirn, die sich bei Ausübung ganz bestimmter kognitiver Leistungen zeigen. In der Robotik (s. Kap. II.B.2) bezeichnet er vorwiegend Darstellungen für unterschiedlichste, aus Sicht eines Roboters externe und interne Entitäten (Objekte, Zustände, Wissen) und Relationen zwischen ihnen. Und in der Philosophie schließlich (s. Kap. II.F) wird der Repräsentationsbegriff häufig gebraucht, um auf komplexe Gedanken zu verweisen, die nicht für ein einzelnes Objekt, sondern einen ganzen Sachverhalt stehen.

In der Geschichte der Kognitionswissenschaft wurden und werden bis heute jedoch auch viel stärkere Repräsentationsbegriffe eingeführt, die von einer Repräsentation sehr viel mehr verlangen als die soeben skizzierte Grundidee (z. B. eine propositionale Struktur; s. u.). Allerdings werden diese stärkeren Begriffe in der Kognitionswissenschaft meist

von sog. Anti-Repräsentationalisten verwendet, die dann zu zeigen versuchen, dass der Repräsentationalismus unhaltbar ist, weil es nichts gibt, was diese stärkeren Anforderungen erfüllt. Dabei geht jedoch die zentrale Idee der Kognitionswissenschaft verloren, dass flexibles Verhalten durch Rekurs auf interne Zustände des jeweiligen Akteurs zu erklären ist, so dass ein zu starker, und insofern inhaltsleerer, Repräsentationsbegriff letztlich das Kinde mit dem Bade ausschüttet. Der liberale Begriff, der auf die Wurzeln der Kognitionswissenschaft zurückgeht und im Folgenden verteidigt werden soll, kommt der wissenschaftlichen Praxis in der Kognitionswissenschaft hingegen viel näher und bewahrt ihre fruchtbaren Grundsätze, mit denen z. B. eine Abgrenzung zum Behaviorismus gelingt.

Wie können mentale Repräsentationen charakterisiert werden?

Eine der gängigsten Charakterisierungen mentaler Repräsentationen besteht darin, sie als ›intentionale‹ Zustände zu beschreiben. Die Verwendung des Begriffs ›Intentionalität‹ in der modernen Debatte geht auf Franz Brentano (1874) zurück, der unter Intentionalität die Eigenschaft von (bewussten) mentalen Zuständen verstand, sich auf etwas in der Welt beziehen zu können, auf etwas ›gerichtet‹ zu sein. Die verschiedenen Beschreibungen Brentanos wurden auf unterschiedliche Weise interpretiert; hier sollen nur zwei gängige Interpretationen skizziert werden:

- Mentale Repräsentationen haben eine Stellvertreterfunktion, d. h. sie können stellvertretend für Gegenstände (oder Eigenschaften oder Sachverhalte) benutzt werden. Statt also den kürzesten Weg zur nächsten Bushaltestelle zu suchen, kann ich auch meine ›mentale Karte‹ benutzen, um mir vorher den günstigsten Weg zu überlegen.
- Mentale Repräsentationen stehen für Sachverhalte, und diese sind entweder der Fall oder nicht. Mentale Repräsentationen sind also genau dann wahr, wenn der durch sie repräsentierte Sachverhalt tatsächlich besteht, in allen anderen Fällen sind sie falsch. Diese Lesart ist allerdings beschränkt auf Überzeugungen und Wahrnehmungen und daher v. a. in der philosophischen Debatte relevant. (Oft werden die Ausdrücke ›wahr‹ und ›falsch‹ vermieden und durch Ausdrücke wie ›veridisch‹ und ›Fehlrepräsentation‹ ersetzt.)

Viele Autoren gehen davon aus, dass diese beiden Interpretationen eng miteinander verknüpft sind; dass

das aber nicht so sein muss, wird später noch deutlich werden, wenn es um funktionale Kriterien der Individuierung von Repräsentationen geht. Wie genau Intentionalität zu verstehen ist und ob dieser Begriff geeignet ist, mentale Repräsentationen zu charakterisieren, ist weiterhin Gegenstand philosophischer Debatten.

Um genauer über die verschiedenen Eigenschaften von Repräsentationen sprechen zu können, müssen einige grundlegende Unterscheidungen getroffen werden.

Zunächst einmal haben Repräsentationen immer eine materielle Realisierung – im Fall von Wörtern z. B. sind dies Schallwellen oder Druckerschwärze, im Fall von mentalen Repräsentationen Gehirnzustände oder -prozesse. Diese materielle Seite von Repräsentationen bezeichnet man als ihre ›Vehikel‹. Die Vehikel sind die Träger des *Inhaltes*. Mit ›Inhalt‹ (oder auch ›Gehalt‹) ist jene Eigenschaft von Repräsentationen gemeint, die festlegt, welcher Gegenstand vertreten wird. Im Hinblick auf das folgende Vorkommnis des Wortes ›Tisch‹ z. B. ist die zwischen den beiden Anführungszeichen befindliche Druckerschwärze in ihrer spezifischen Verteilung das Vehikel, während der Inhalt dasjenige ist, das bestimmt, welche Art von Gegenstand dieses Wort bezeichnet. Eine Repräsentation ist demnach etwas, das physische Eigenschaften (ein Vehikel) und semantische Eigenschaften (einen Inhalt) hat.

Verdeutlichen lässt sich das Verhältnis von Repräsentation, Vehikel und Inhalt anhand eines Geldstücks: Ein Geldstück ist eine Münze (physische Eigenschaften), die einen gewissen monetären Wert (vergleichbar mit dem Inhalt) hat. Als physischer Gegenstand ist die Münze zwar Träger des monetären Wertes, aber für sich genommen noch kein Geldstück – zu einem Geldstück wird sie erst durch die Zuweisung eines monetären Wertes. Ganz analog ist bei mentalen Repräsentationen das Vehikel (etwa ein Gehirnzustand) zwar der Träger des Inhalts (der bestimmt, was repräsentiert wird), selbst aber noch keine Repräsentation – zu einer Repräsentation wird das Vehikel erst durch die Zuweisung eines Inhalts.

Diese beiden Aspekte von Repräsentationen spielen in den unterschiedlichen Disziplinen eine mehr oder weniger prominente Rolle: Während z. B. die Neurowissenschaften (s. Kap. II.D) stark auf die Vehikel fokussiert sind und in der Kognitionspsychologie (s. Kap. II.E.1) die Inhalte von Repräsentationen die zentrale Rolle spielen, wird in der Metaphysik des Geistes (s. Kap. II.F.1) gerade das Verhältnis zwischen Vehikeln und Inhalten analysiert. Diese Art

der Arbeitsteilung ist in einem echten interdisziplinären Projekt wie der Kognitionswissenschaft allerdings nicht unbedingt wünschenswert, da beide Aspekte von Repräsentationen, Vehikel und Inhalt, notwendigerweise gemeinsam auftreten und somit auch gemeinsam untersucht werden sollten.

Eine weitere Unterscheidung, die in diesem Zusammenhang von großer Bedeutung ist, ist die Unterscheidung zwischen *Inhalt* (Intension) und *Repräsentiertem* (Extension). Der Begriff ›Extension‹ bezieht sich auf die von einer Repräsentation herausgegriffenen Gegenstände, also das Repräsentierte: Die Extension des Ausdrucks ›Tisch‹ z. B. besteht schlicht aus der Menge aller Tische. Wird ein Einzelding repräsentiert, spricht man statt von ›Extension‹ häufig auch vom ›Referenten‹. Die Extension ist zu trennen von der Intension einer Repräsentation, d. h. von ihrem Inhalt, der festlegt, welche Gegenstände bezeichnet werden: Da tatsächlich alle Lebewesen, die ein Herz haben, auch eine Niere haben und umgekehrt, haben z. B. die Begriffe ›Lebewesen mit Niere‹ und ›Lebewesen mit Herz‹ dieselbe Extension; ihre Intension ist allerdings verschieden, denn wir beziehen uns auf diese Lebewesen in unterschiedlicher Weise: Um etwa herauszufinden, ob etwas unter den jeweiligen Begriff fällt oder nicht, wenden wir unterschiedliche Methoden an, und jemand, der den tatsächlichen Zusammenhang zwischen Nieren und Herzen nicht kennt, könnte glauben, dass die Extension verschieden sei, ohne dass er deswegen ein defizitäres Verständnis der beiden Begriffe hätte. Während der Inhalt ein Aspekt der Repräsentation selbst ist, ist das Repräsentierte etwas Externes. Dennoch besteht zwischen beidem eine enge Verbindung: Die Intension bestimmt die Extension (s. u.).

Die Unterscheidung zwischen Repräsentiertem und Inhalt ist deswegen wichtig, weil von der Tatsache, dass etwas repräsentiert wird, nicht darauf geschlossen werden kann, wie es repräsentiert wird (was also der Inhalt der entsprechenden Repräsentation ist). Eine Ameise und eine Wissenschaftlerin können z. B. gleichzeitig dasselbe Ameisennest repräsentieren, so dass ihre jeweiligen Repräsentationen dieselbe Extension haben. Daraus folgt aber nicht, dass beide auch die gleiche Art von Repräsentation haben, dass sie das Nest auf dieselbe Art und Weise repräsentieren, also in beiden Fällen derselbe Inhalt vorliegt: Die Ameise repräsentiert die Richtung und Entfernung ihres Nestes im Verhältnis zu ihrem derzeitigen Standort, d. h. wie weit und in welche Richtung sie gehen muss, um zum Nest zurück zu kommen: Sie benutzt eine egozentrische Reprä-

sensation (Meilinger/Vosgerau 2010). Die Wissenschaftlerin dagegen repräsentiert die Lage des Nestes möglicherweise innerhalb eines Feldes, dessen Ecken sie mit Pfosten markiert hat, um den Weg der Ameise genau bestimmen zu können: Sie repräsentiert das Nest nicht relativ zu sich selbst (egozentrisch), sondern allozentrisch, d. h. relativ zu anderen Dingen in der Welt (z. B. dem Feld bzw. den Eckpfosten des Feldes). Obwohl beide ein und dasselbe repräsentieren, nämlich das Ameisennest, unterscheiden sich die Inhalte der jeweiligen Repräsentationen deutlich.

Was beide Fälle verbindet, ist die *Möglichkeit der Fehlrepräsentation*: Sowohl die Ameise als auch die Wissenschaftlerin können den Ort des Nestes falsch repräsentieren. Diese zentrale Eigenschaft von Repräsentationen stellt eine begriffliche Notwendigkeit dar: Etwas, das nicht falsch sein kann, kann keine Repräsentation sein – weshalb wir den Motten ja keine interne Repräsentation der Lichtquelle zusprechen wollten. Fred Dretske (1981) führt für solche Fälle den Begriff der *Information* ein, der so definiert wird, dass gerade keine Fehler auftauchen können – die Motten können über den Ort einer Lichtquelle in Dretskes Sinne nicht ›fehlinformiert‹ sein. Seiner Meinung nach können Repräsentationen zwar selbst keine Information tragen, aber die Funktion haben, Information zu tragen, und da diese Funktion nicht immer erfüllt wird, entsteht die Möglichkeit der Fehlrepräsentation. Ein optischer Feuermelder hat z. B. die Funktion, Feuer anzuzeigen, d. h. Information über die Anwesenheit von Feuer zu tragen. Zu diesem Zweck enthält er eine Lichtquelle und einen Fotowiderstand, der Alarm auslöst, wenn kein Licht mehr auf ihn fällt: Dringt Rauch zwischen Lichtquelle und Fotowiderstand, wird aufgrund der Verdunkelung der Alarm ausgelöst. Da eine Verdunkelung aber auf unterschiedliche Weise ausgelöst werden kann, trägt der Alarm des Feuermelders in Dretskes Sinne keine Information über das Feuer: Er ist eine Repräsentation, die Dretske zufolge zwar die Funktion hat, Information zu tragen, dies aber, weil sie falsch sein kann, tatsächlich nicht tut.

Ganz unabhängig von den spezifischen Details von Dretskes Informationsbegriff ist an dieser Stelle entscheidend, dass ganz allgemein nicht alles, was Information trägt, eine Repräsentation ist, weil der Begriff der Repräsentation in gewisser Hinsicht anspruchsvoller ist als der der Information. Alkoholkonsum z. B. führt zu bestimmten Veränderungen im Gehirn, die damit Information über den Alkoholkonsum tragen; trotzdem repräsentieren die aus diesen Veränderungen resultierenden neuronalen

Zustände nicht den Alkoholkonsum. Ähnlich trägt z. B. die Temperatur eines Prozessors Information über die Anzahl der von ihm verarbeiteten Befehle, repräsentiert sie aber nicht. Auch in kognitiven Systemen darf also Information nicht mit Repräsentation gleichgesetzt werden. Ein Zeichen trägt Information über etwas anderes sozusagen einfach ›von sich aus‹, weshalb Information zu tragen eine zweistellige Relation zwischen dem Signal, das die Information trägt, und demjenigen, worüber es Information trägt, ist. Die Relation zwischen einer mentalen Repräsentation und dem Repräsentierten hingegen ist nicht zweistellig, denn Repräsentationen sind immer abhängig von ihrer Verwendung. Mentale Repräsentationen können immer nur von dem jeweiligen kognitiven System benutzt werden, dessen Repräsentationen sie sind. Da die Kognitionswissenschaft mentale Repräsentationen wie oben gesehen gerade zur Erklärung flexiblen Verhaltens einführt, werden die Zustände eines Systems erst dann zu repräsentationalen Zuständen, wenn sie handlungsrelevant werden können, d. h. wenn unterschiedliche repräsentationale Zustände auch zu unterschiedlichem Verhalten führen können. In die Repräsentationsrelation muss aus diesem Grund auch das System aufgenommen werden, das über den repräsentationalen Mechanismus oder die entstehenden repräsentationalen Zustände verfügt. Mentale Repräsentationen sind daher nur Repräsentationen in Bezug auf dieses eine System, das sie verwendet oder über sie verfügt. Bei mentalen Repräsentationen haben wir es also mit einer dreistelligen Relation zwischen Repräsentation, Repräsentiertem und kognitivem System zu tun.

Zusammenfassend können mentale Repräsentationen grob charakterisiert werden als interne Zustände eines Systems, die für etwas anderes stehen, fehlerhaft sein können und keine Information in Dretskes Sinne über das tragen, was sie repräsentieren. Der Inhalt einer Repräsentation, d. h. ihre Intension, die festlegt, was repräsentiert wird, muss einerseits von dem Repräsentierten, d. h. ihrer Extension, unterschieden werden, andererseits von ihrem Vehikel, das diesen Inhalt trägt. Das Repräsentierte ist verschieden von der Repräsentation, Inhalt und Vehikel sind zwei ihrer Aspekte. Diese Charakterisierung erlaubt es, einen genaueren Blick auf die Rolle von Repräsentationen in einer Theorie des Geistes zu werfen.

Welche Rolle spielen Repräsentationen in einer Theorie des Geistes?

Unter der sog. Repräsentationalen Theorie des Geistes versteht man die These, dass Kognition im Wesentlichen in der Verarbeitung von Repräsentationen besteht. Dabei geht man im Allgemeinen davon aus, dass Repräsentationen durch Wahrnehmungs- oder kognitive Prozesse aufgebaut, dann durch (weitere) kognitive Prozesse verändert und schließlich verhaltenswirksam werden (s. Kap. III.1). Repräsentationen spielen demnach nicht nur eine zentrale Rolle in der Erklärung von flexiblem Verhalten, sondern bilden auch das unverzichtbare Substrat kognitiver Prozesse. Gerade im Bereich der Künstliche-Intelligenz-Forschung (KI) sind die Debatten um eine Repräsentationale Theorie des Geistes besonders plastisch nachzuvollziehen, da dort die verschiedenen Ansätze von sehr unterschiedlichen Auffassungen zum Thema Repräsentation geprägt sind. Historisch gesehen wurzeln die Techniken der Informatik zu Wissensrepräsentation (s. Kap. IV.25), Navigation (s. Kap. II.B.2), Problemlösung (s. Kap. IV.11) und Planung (s. Kap. IV.17) in der klassischen KI (s. Kap. II.B.1, Kap. III.1) und waren in dem Sinne kognitivistisch, dass sie sich an einem rationalen, logisch formalisierbaren Intelligenzbegriff ausrichteten, der ganz ohne den Körper auskam (s. Kap. III.6). Damit gelang es bereits, Roboter zu konstruieren, die mithilfe komplexer Repräsentationen in der Lage waren, in sehr einfachen, statischen Welten zu navigieren, und einfache Aufgaben wie etwa von A nach B zu fahren meistern konnten. Eine solche rein repräsentationsorientierte Theorie körperloser Intelligenzentwicklung stößt jedoch schnell an Grenzen, wenn sie zur Beherrschung eines sich in einer dynamischen Welt real bewegenden Roboters implementiert werden soll, weil der Aufbau und die Verarbeitung einer komplexen Repräsentation u. a. zu viel Zeit beansprucht, so dass der Roboter oftmals nur über bereits ›veraltete‹ Repräsentationen verfügen würde.

Eine zentrale Kritik an der Repräsentationalen Theorie des Geistes beruht daher auf der Zurückweisung der Auffassung, dass mentale Repräsentationen integrierte und detailreiche Modelle der Welt sind, die die ganze Welt in unserem Geist ›abbilden‹, so dass sich unser Verhalten statt nach der Welt selbst nach diesem unserem Bild der Welt richten kann. Unsere Repräsentationen erweisen sich in vielen Fällen als sehr viel bruchstückhafter und aspektorientierter als diese klassische Auffassung nahelegt. Die Wüstenameise z. B. repräsentiert ihr Nest nicht

mit allen seinen Aspekten, sondern lediglich seinen Ort, und selbst diesen nur in Hinblick auf sehr wenige räumliche Eigenschaften, etwa nur in Hinblick auf die Bewegung, die sie ausführen muss, um das Nest zu erreichen. Darüber hinaus stellte sich gerade auch in der Robotik heraus, dass sich einige Probleme in der Bewegungssteuerung einfacher lösen lassen, wenn statt integrierter Repräsentationen aller Faktoren nur sehr spärliche interne Repräsentationen benutzt werden und das Verhalten stattdessen direkt über Faktoren der Umwelt gesteuert wird. Am radikalsten formulierte diesen Punkt Rodney Brooks (1991, 140): »Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems.« Erstes Ziel einer ›neuen KI‹ müsse es vielmehr sein, rein verhaltensgesteuerte Artefakte zu bauen, die sich selbstständig an veränderte Umwelten anpassen oder sich eine neue Umgebung suchen können. Wenn dies erreicht sei, so Brooks, dann werde ›der Rest‹ menschlicher Kognitions- und Intelligenzleistungen wie Kategorisierung (s. Kap. IV.9), Sprache (s. Kap. IV.20), Planung (s. Kap. IV.17), abstraktes Problemlösen (s. Kap. IV.11) usw. ganz einfach zu implementieren sein. Brooks (1990, 5) prägte für diesen Ansatz den Slogan »the world is its own best model«. Die von ihm entwickelte Subsumptionsarchitektur (s. Kap. III.7) verzichtet z. B. darauf, einen mobilen Roboter mit einer exakten Karte seiner Umwelt auszustatten, und arbeitet stattdessen mit einem einfachen Algorithmus, der auf der Basis direkter Umwelterfassung Hindernisse erkennt und die Route des Roboters beeinflusst.

Eine weitere Eigenschaft, die mentalen Repräsentationen traditionell nachgesagt wird, ist ihre Symbolhaftigkeit (diese Lesart geht v. a. auf Jerry Fodor zurück, der den Terminus ›Repräsentationale Theorie des Geistes‹ geprägt hat). Damit ist gemeint, dass jede Bedeutungseinheit durch eine Vehikeleinheit, also quasi atomar, repräsentiert wird. Eine Ameise, die ihr Nest repräsentiert, verfügt demnach über eine komplexe Repräsentation, die aus verschiedenen einfachen Symbolen (etwa aus einem Symbol für das Nest, einem Symbol für den Ort des Nestes, einem Symbol für den eigenen Ort usw.) aufgebaut ist. Allerdings ist auch an der Symbolhaftigkeit von Repräsentationen berechtigterweise Kritik geübt worden: Neuronale Netze (s. Kap. III.2) scheinen nicht symbolisch zu arbeiten, d. h. in ihnen werden semantische Einheiten nicht jeweils durch eine Einheit repräsentiert (also etwa durch ein einziges Neuron), sondern bestenfalls durch ein charakteristisches Aktivitätsmuster vieler verschiedener Neurone, wobei einzelne Neurone nicht nur an der Repräsentation einer semantischen

Einheit, sondern an einer Vielzahl solcher Einheiten beteiligt sind. Es deutet also alles darauf hin, dass in neuronalen Netzen bzw. dem Gehirn keine atomaren symbolischen Einheiten zu finden sind, die semantische Einheiten repräsentieren (oder codieren), semantische Einheiten werden dort vielmehr ›subsymbologisch‹ repräsentiert.

Eine weitaus fundamentalere Kritik an der Repräsentationalen Theorie des Geistes übt eine Richtung der Kognitionswissenschaft, die auf der mathematischen Theorie dynamischer Systeme beruht (s. Kap. III.4). Die Theorie dynamischer Systeme beschreibt mithilfe einer Menge von Differenzialgleichungen, wie sich Systeme über die Zeit hinweg verändern. Ein Beispiel für ein dynamisches System ist ein vereinfachter Raumheizungsthermostat, der mithilfe eines Thermometers die Raumtemperatur misst und die Heizleistung an die gemessene Temperatur koppelt (fällt die Raumtemperatur, erhöht der Thermostat die Heizleistung, wodurch die Raumtemperatur steigt; steigt die Raumtemperatur, verringert er die Heizleistung, wodurch die Raumtemperatur sinkt usw.). Das Verhalten der beteiligten Systemparameter (Raumtemperatur, gemessene Temperatur, Wärmeentwicklung der Heizung) unter verschiedensten Bedingungen (Außentemperatur, Raumisolierung, Genauigkeit des Thermometers) und Störeinflüssen (plötzliches Öffnen des Fensters) lässt sich sehr elegant mithilfe von Differenzialgleichungen beschreiben. Je nach Wert der Parameter kann sich das System stabil verhalten, also immer auf eine konstante Raumtemperatur zustreben, zyklisch schwanken oder sich völlig chaotisch verhalten (s. Kap. III.4). Die Grundidee der Anwendung der Theorie dynamischer Systeme auf kognitive Systeme besteht darin, dass diese nicht mehr in Isolation beschrieben, sondern ihre Umwelt sowie ihre Interaktion mit dieser Umwelt in die Beschreibung miteinbezogen werden (z. B. Thelen 1995). Das führt dazu, dass kognitive Systeme wesentlich größere Einheiten darstellen als es in der klassischen Kognitionswissenschaft üblicherweise der Fall war. Da in den zu ihrer Beschreibung verwendeten Gleichungen interne nicht mehr von externen Zuständen unterschieden werden, kann auch nicht mehr sinnvoll von Repräsentationen die Rede sein: Es gibt zunächst keinen Grund, warum ein externer Faktor in einer Gleichung in Gestalt seiner Repräsentation nochmals als interner Faktor zu wiederholen wäre. Allerdings gilt das nicht für alle Systeme: Die gemessene Temperatur kann z. B. sehr wohl als Repräsentation der Raumtemperatur aufgefasst werden, da sie eine Stellvertreterfunktion übernimmt – die Heizleistung

hängt nicht unmittelbar von der Raumtemperatur ab, sondern nur mittelbar über die gemessene Temperatur, so dass eine gewisse ›Doppelung‹ des Faktors Raumtemperatur (als tatsächliche und gemessene Raumtemperatur) in den Gleichungen durchaus sinnvoll sein kann. Für andere Parameter, etwa die Heizleistung, trifft das nicht zu, sie repräsentieren keine andere Systemkomponente. Allerdings scheinen völlig repräsentationsfreie dynamische Systeme, in denen kein Systemparameter einen anderen Systembestandteil repräsentiert, nicht auszureichen, um z. B. höhere kognitive Leistungen zu erklären: Da diese Systeme immer nur eine konkrete Interaktion mit der Umwelt abbilden können, sind kognitive Vorgänge, die keine direkte Interaktion beinhalten (und dazu gehören gerade die ›klassischen‹ kognitiven Aufgaben wie Schlussfolgern, Planen, Abwägen, Erinnern usw.), scheinbar nicht ohne Weiteres im Sinne dynamischer Systeme beschreibbar (Eliasmith 1997; van Gelder/Port 1995). Solche kognitiven Leistungen sind im Sinne von Clark/Toribio (1994) ›repräsentationshungrig‹ (*representation-hungry*; s. Kap. III.7). Außerdem ersetzen rein dynamizistische Ansätze das ursprüngliche *Explanandum* der Kognitionswissenschaft, d. h. das flexible Verhalten von kognitiven Systemen, durch ein anderes, nämlich durch das dynamische Verhalten von gekoppelten Systemen (d. h. kognitiven Systemen plus ihrer Umwelt). Anti-repräsentationalistische dynamizistische Theorien können Kognition im traditionellen Sinne also nicht befriedigend erklären. Zwar weisen sie zu Recht darauf hin, dass viele relevante Faktoren nicht explizit intern repräsentiert werden müssen, und ihre Forderungen nach ›Verkörperlichung‹ (*embodiment*) sowie Umwelt- und Situationsbezug (*situatedness*) sind in der kognitiven Robotik (s. Kap. II.B.2) inzwischen als wesentliche Konstituenten intelligenter Artefakte anerkannt (s. Kap. III.7), so dass zunehmend ein Übergang von einem repräsentations- zu einem handlungszentrierten Ansatz stattgefunden hat, dessen Verfolgung die Konstruktion von leiblichen Artefakten zwingend voraussetzt: Kognition wird damit zu einem aktiven Prozess der Weltgestaltung (s. u.). Daraus folgt jedoch nicht, dass wir in unserer Beschreibung kognitiver Systeme auf Repräsentationen vollständig verzichten können (eine sehr ausführliche Diskussion der Standpunkte derjenigen, die den Repräsentationsbegriff ablehnen und anderen, die ihn in unterschiedlicher Ausprägung anerkennen, findet sich in van Gelder (1998)).

Viele sog. Anti-Repräsentationalisten wenden sich wie gesehen gegen eine oder mehrere der ›klas-

sischen Eigenschaften von Repräsentationen (vollständig, detailreich, symbolisch). Denkt man allerdings zurück an den ursprünglichen Zweck der Einführung von Repräsentationen, so ist dort von diesen Eigenschaften gar nicht die Rede: Das ursprüngliche liberale Bild von Repräsentationen ließ vielmehr völlig offen, ob Repräsentationen symbolisch, vollständig und detailreich oder integrierte Modelle von komplexen Sachverhalten sind. Die Debatte um den Repräsentationsbegriff betrifft daher in erster Linie nicht die Existenz von Repräsentationen, sondern sagt allenfalls etwas darüber aus, welche Eigenschaften Repräsentationen nicht haben bzw. nicht unbedingt haben müssen. Um den Begriff der Repräsentation weiterhin fruchtbar in der Kognitionswissenschaft verwenden zu können, muss ein relativ liberales Verständnis von Repräsentationen vorausgesetzt werden, das Raum für ganz unterschiedliche Arten von mentalen Repräsentationen bietet. Mentalen Repräsentationen kommt in kognitionswissenschaftlichen Theorien also nach wie vor eine zentrale Rolle zu, auch wenn weniger Repräsentationen angenommen werden müssen als lange gedacht, und auch wenn viele Repräsentationen nicht alle Eigenschaften haben, die ihnen im Laufe der Entwicklung der Kognitionswissenschaft zugeschrieben wurden. Die Repräsentationale Theorie des Geistes hat klarer Weise Grenzen, aber sie muss nicht einem radikalen Anti-Repräsentationalismus weichen.

Eine weitere derartige Grenze betrifft die Erklärung von phänomenalem Erleben, also dem ›wie es ist‹ (*what it is like*), sich in einem bestimmten (bewussten) Zustand zu befinden (Nagel 1974; s. Kap. IV.4). Nach der liberalen Auffassung können mentale Repräsentationen prinzipiell bewusst oder unbewusst vorliegen. Sie erklären das Verhalten von kognitiven Systemen unabhängig davon, ob das kognitive System dabei ein phänomenales Erleben hat oder nicht (es ist z. B. völlig irrelevant, ob die Ameise etwas erlebt oder nicht). Der Inhalt einer Repräsentation legt fest, mit welchen Gegenständen interagiert wird (z. B. mit dem Nest). Einige Kognitionswissenschaftler sind davon überzeugt, dass phänomenales Erleben durch einen bestimmten ›phänomenalen Inhalt‹ oder ›Gehalt‹ erklärt werden kann. Das allerdings widerspricht der Grundidee, mit der Repräsentationen eingeführt wurden: Wir verhalten uns nicht zu unserem phänomenalen Erleben (wir interagieren nicht mit ihm), und daher kann dieses phänomenale Erleben auch nicht über den Inhalt von Repräsentationen erklärt werden (Vosgerau et al. 2008). Im Gegenteil: Jeder Inhalt

kann prinzipiell bewusst oder unbewusst vorliegen, in beiden Fällen wird damit ein Verhalten zu bestimmten Objekten erklärt. Das bedeutet natürlich nicht, dass Bewusstsein nicht zu Unterschieden im Verhalten führt, sondern nur, dass diese Unterschiede im Verhalten durch die Art der Verarbeitung der Repräsentationen erklärt werden müssen, nicht durch deren Inhalt, also einen ›phänomenalen Gehalt‹. Ein entsprechendes Erklärungsmodell ist z. B. das *global workspace model* (Dehaene/Naccache 2001; s. Kap. IV.4), das Bewusstsein von Repräsentationen auf ihre globale Verfügbarkeit im Gehirn zurückführt, was auf einer funktionalen Beschreibungsebene bedeutet, dass sie für verschiedenste kognitive Subsysteme (Schlussfolgern (s. Kap. IV.17), Gedächtnis (s. Kap. IV.7), Aufmerksamkeit (s. Kap. IV.1), Lernen (s. Kap. IV.12), Sprache (s. Kap. IV.20) usw.) verfügbar sind.

Wie hängen Repräsentationen von Umwelt und Körper ab?

Im vorangegangenen Abschnitt wurde bereits deutlich, dass verschiedene Kritiken am Repräsentationsbegriff dazu geführt haben, dass sich der Fokus der Debatte auf Aspekte verschoben hat, die einerseits außerhalb kognitiver Systeme liegen und andererseits deren körperliche Verfassung betreffen (s. Kap. III.7). Was genau ist das Verhältnis von Repräsentationen und Umwelt bzw. Körper?

In den letzten Jahren ist die Frage aufgekommen, ob die Vehikel unserer mentalen Repräsentationen nur in unseren Gehirnen oder Körpern zu finden sind (wie in der Kognitionswissenschaft traditionell angenommen wurde) oder ob sie über die Körpergrenzen hinaus reichen. Unter dem Slogan ›*extended cognition*‹ (s. Kap. III.8) wird dabei argumentiert, dass die Vehikel unserer mentalen Repräsentationen nicht nur interne Zustände (im Sinne der Grenzen des zentralen Nervensystems oder der Haut als Grenze), sondern auch externe Zustände umfassen. Die grundlegende Idee ist, dass manche externen ›Geräte‹, z. B. ein Schreibblock oder das Internet, die gleiche Funktion erfüllen können wie Gedanken, die wir uns machen, oder Dinge, die wir uns merken. Sie erfüllen also, so die Argumentation, dieselbe Rolle wie klarerweise mentale Zustände und sollten daher selbst auch als mental gelten. Im Gegensatz dazu geht der traditionelle Internalismus davon aus, dass die Vehikel mentaler Repräsentationen lediglich im Inneren des repräsentierenden Systems vorkommen, wenngleich meist völlig unklar bleibt, wie die-

ses Innere genau zu definieren ist. Was in dieser Debatte gefordert wird, ist letztlich also eine Definition des Mentalen. Unabhängig davon, was eine solche Definition leisten soll und kann, besteht eine gewisse Spannung zwischen der ursprünglichen Idee einerseits, die von der Erklärung flexiblen Verhaltens ausgeht und mit einem liberalen Verständnis von Repräsentationen einhergeht, und der Idee von *extended cognition* andererseits: Wenn externe Objekte mit in die Erklärung von Verhalten einbezogen werden, dann erklären wir nicht mehr das flexible Verhalten des Systems, das dieses Verhalten zeigt (z. B. rechne *ich*, auch wenn ich Zettel und Stift dazu brauche, und die Ameise läuft zum Nest, auch wenn sie dabei den Sonnenstand berücksichtigt), sondern das ›Verhalten‹ des größeren Systems, das sich aus diesem System und dem externen Objekt zusammensetzt (was auch immer das komplexe System *Ich+Zettel+Stift* tut, es rechnet ebenso wenig, wie das komplexe System *Ameise+Sonne* zum Nest läuft). Anhängern von *extended cognition* kann es aus diesem Grund nicht mehr darum gehen, mein Rechenverhalten oder die räumliche Orientierung der Ameise zu erklären. Auch wenn die Debatte zu Recht darauf hinweist, dass externe Objekte eine wichtigere Rolle für die Erklärung kognitiver Leistungen spielen könnten als bisher angenommen wurde, verändern auch sie, ähnlich wie die Theorie dynamischer Systeme, das ursprüngliche *Explanandum* der Kognitionswissenschaft.

In neueren Entwicklungen in der KI ist dementsprechend eine erhebliche Ausdifferenzierung bei der Konstruktion experimenteller Systeme festzustellen: Man konzentriert sich mehr und mehr darauf, Repräsentationen über mehrere Systeme zu verteilen, und darauf, dass sich Systeme gegenseitig über die eigenen direkten sensorischen Erfassungsmöglichkeiten hinaus *via* Internet mit Informationen versorgen, so dass, sofern vorab ein geeignetes Kommunikationsformat vereinbart wurde, zwischen den Systemen sowohl historische als auch aktuelle Messwerte, Erfahrungen und abgeleitetes Wissen ausgetauscht werden können. In gewisser Weise entstehen so virtuelle Repräsentationen eines gemeinsamen Raumes, die verschiedene Abstraktionsgrade aufweisen können (vgl. z. B. das Projekt *RoboEarth*: <http://www.roboearth.org/>).

Weiterhin muss geklärt werden, wie das Verhältnis von Vehikeln (die physische Entitäten und damit körperlich sind) und Inhalten (die nicht physisch sind) beschrieben werden kann. Die entsprechende Debatte ist vielschichtig, da verschiedene Aspekte ins Spiel kommen. Zum einen gibt es die metaphysi-

sche Debatte darum, wie das Verhältnis von Vehikeln und Inhalten generell zu charakterisieren ist: Hierbei geht es hauptsächlich um das sog. Leib-Seele-Problem, also um die Frage, wie mentale Eigenschaften mit den physikalischen Eigenschaften der Vehikel zusammenhängen (s. Kap. II.F.1, Kap. IV.4). Zum anderen geht es um die Frage, ob und inwiefern der Inhalt einer Repräsentation vom Vehikel abhängig ist: Die klassische Position der Kognitionswissenschaft besagt hierbei gemäß der sog. *physical symbol system hypothesis* (Newell/Simon 1976), dass die Inhalte von der jeweiligen Realisierung (also den Vehikeln) vollkommen unabhängig sind. Mentale Repräsentationen sind demnach als Symbole zu charakterisieren, die auf ganz unterschiedliche Art und Weise implementiert sein können, ohne dass die Implementierung etwas an deren Inhalt und deren (algorithmischer) Verarbeitung änderte (s. Kap. III.1).

Diese Idee wurde von verschiedensten Seiten, insbesondere auch aus Sicht der kognitiven Robotik (s. Kap. II.B.2) kritisiert. In der kognitiven Robotik als einem Teilbereich der KI versteht man unter Repräsentationen allgemein Darstellungen für unterschiedlichste, aus Sicht eines Rechner- oder Robotersystems externe und interne Entitäten und Relationen, etwa Zustände und deren Dynamik, Modelle von und für Umwelt und Raum, situative Wahrnehmungen, aber auch Muster für situierte Handlungen sowie physikalische, sprachliche oder visuelle Interaktionen. Deren Darstellungen, also die Repräsentationen, finden sich auf verschiedenen, dezentralen Verarbeitungsebenen eines informationsverarbeitenden Robotersystems und dienen bei seiner Steuerung dazu, externe und interne Stimuli in zielgerichtete Aktionen (Handlungsmuster) umzusetzen. Das zentrale Problem bei der Einführung eines solchen Konzepts von Repräsentationen als ›Vermittler‹ besteht in ihrer Entstehungsmodalität: Bislang werden Repräsentationen auf allen Ebenen der Steuerung – von der Struktur von Karten der Umgebung über Reiz-Reaktions-Schemata bis hin zu multimodalen Schnittstellen – praktisch immer von einem ›allwissenden‹ menschlichen Konstrukteur vorgegeben; neben der menschlichen Vorgabe werden höchstens noch Parameter adaptiert. Ein solcher Ansatz schließt jedoch nicht nur die Entwicklung struktureller Innovation in Verhaltensmustern aus, sondern funktioniert auch nur für ›vorgedachte‹ Eingabe-Ausgabe-Relationen. Aus diesem Grund ist es für die Entwicklung von ›intelligenten‹ Artefakten von entscheidender Bedeutung, dass Repräsentationen im Handlungs- und Umweltkontext bezogen auf die je spezifische Umwelt, Sensorik, Morphologie und Dy-

namik des Robotersystems laufend neu entstehen können. Diese Herangehensweise spiegelt sich v.a. im Paradigma des Enaktivismus wider (Maturana/Varela 1980), wonach Kognition in kontinuierlicher plastischer Koppelung eines intelligenten Wesens an seine Umwelt erfolgt und eine ständige Wechselwirkung zwischen (ganzem) Körper und Repräsentationssystem sowie zwischen Wesen und Umwelt existiert (s. Kap. III.9). Allerdings ist die systematische Untersuchung von kognitiven Architekturen für Robotersysteme, die die Entstehung von völlig neuen Repräsentationen ermöglichen, in der Robotik bislang nicht mit großem Nachdruck angegangen worden. Gelänge deren Konstruktion, dann hätte man nicht nur eine neue Grundlage für die Entwicklung intelligenten Verhaltens in Robotern auf Ebenen, die nicht von vornherein ›inprogrammiert‹ sind; dieser Erfolg hätte auch für die handlungszentrierte Kognitionstheorie erhebliche Konsequenzen, da ihre Postulate zumindest für bestimmte Domänen experimentell simuliert und validiert werden könnten.

Ein spezielleres Beispiel für eine enaktivistische Theorie stellt die Theorie visueller Wahrnehmung von O'Regan/Noë (2001) dar (s. Kap. IV.24), wonach unsere Fähigkeit zur Wahrnehmung darin besteht, die systematischen Veränderungen des Retinaabildes in Abhängigkeit von den eigenen Bewegungen zu kennen und ›vorhersagen‹ zu können. Wenn wir z. B. einen geraden Strich fokussieren, bildet dieser sich auf der Retina als gerader Strich ab. Wenn wir allerdings neben den Strich schauen, dann bildet sich derselbe Strich aufgrund der Wölbung des Augapfels nicht mehr als Strich, sondern als Kurve ab. Etwas als gerade zu sehen heißt demnach zu ›wissen‹, dass sich das zugehörige Retinaabbild mit den eigenen Bewegungen auf eine bestimmte Art und Weise ändert (s. Kap. IV.19).

Für die kognitive Robotik bedeutet dies, dass die Gestaltung der Umwelt, die körperliche Verfassung des Lebewesens und dessen Entwicklung einen verschränkten Prozess darstellen und nicht separiert werden können (s. Kap. III.7). Tatsächlich allerdings haben sowohl die ›herkömmlichen‹ kognitivistischen als auch konnektionistischen (also auf ›syb-symbolischer‹ Repräsentation basierenden) Ansätze für technische Lösungen arbeitsfähiger Artefakte bereits zu teilweise beeindruckenden Systemlösungen (z. B. in Form von humanoiden Robotern verschiedenster Ausprägung) geführt, während der Nachweis, dass die konsequente technische Umsetzung enaktivistischer Ansätze zu qualitativ ›intelligenteren‹ Robotersystemen führt als kognitivistische oder konnektionistische Ansätze, nach wie vor aussteht.

Zur stärkeren Einbeziehung des Körpers in die Betrachtung von Kognition gehört auch die weitere Entwicklung im Hinblick auf den Körper von Robotern, also den mechanischen Aufbau der Systeme selbst. Obwohl eine geeignete ›Verkörperlichung‹ als Voraussetzung für den Aufbau adäquater Repräsentationen weitgehend akzeptiert ist, sind Roboter bis heute in praktisch allen Fällen starre, elektromotorisch angetriebene Konstruktionen – also weit entfernt vom biologischen Vorbild. Damit ist es jedoch nicht möglich, Materialeigenschaften wie Nachgiebigkeit, dynamisches Verhalten usw. auszunutzen, um Verhaltensweisen möglichst treu nachzubilden (s. Kap. III.5). Genau das allerdings kann nicht nur für den Aufbau von Repräsentationen und Verhaltenskategorien entscheidend sein, sondern führt vielfach auch zu einer erheblichen Vereinfachung der Komplexität von Steuerungen. In Paul (2006) wird in diesem Zusammenhang der Begriff der ›*morphological computation*‹ eingeführt: Phänomene, die sich aus der direkten Interaktion zwischen Körper(-material) und Umwelt ergeben, müssen nicht ingenieurtechnisch modelliert, für eine Recheneinheit repräsentiert und dann über eine aktive (Motor-)Einheit in die Umwelt abgebildet werden. Unter Umständen führt dies zu ganz erheblichen Einsparungen: Die Abwärtsbewegung eines Beins muss z. B. nicht vom Roboter gesteuert werden, sondern kann einfach der Schwerkraft überlassen werden. Ein Beispiel für ein ganzes Robotersystem, welches diesem Prinzip folgt, ist der Roboter CRONOS aus dem Projekt ECCEROBOT (<http://eccerobot.org/>), der aus Gummizügen und Kunststoff aufgebaut ist und unter Ausnutzung seiner Materialeigenschaften in seinem Bewegungsverhalten dem Torso eines Menschen sehr nahe kommt, ohne dass dazu spezielle Berechnungen oder Repräsentationen nötig wären (s. Kap. III.7).

Wie können Repräsentationen individuiert werden?

Wenden wir uns nun der zentralen Frage zu, wann zwei Vorkommnisse als Vorkommnisse einer Repräsentation (eines Typs von Repräsentation) gewertet werden, d. h. der Frage, wie Repräsentationen individuiert werden. Was die Individuierung von Inhalten angeht, können die Debatten teilweise bis in die Antike zurückverfolgt werden. Es geht dabei um die Frage, was den Inhalt einer Repräsentation festlegt und damit ein Vehikel überhaupt zu einer Repräsentation eines bestimmten Inhaltstyps macht. Zu-

nächst können zwei generelle Thesen unterschieden werden: der *semantische Internalismus* und der *semantische Externalismus* (s. Kap. IV.9). Der semantische Internalismus behauptet, dass alle Faktoren, die den Inhalt einer mentalen Repräsentation bestimmen, innerhalb des repräsentierenden Systems zu finden sind, während der semantische Externalismus dafür argumentiert, dass der Inhalt mancher Repräsentationen auch von äußeren Faktoren (z. B. der Beschaffenheit der Welt oder der Sprachgemeinschaft) mitbestimmt wird. Das klassische Argument zugunsten des semantischen Externalismus beruht auf einem Gedankenexperiment zu sprachlichen Ausdrücken von Hilary Putnam und wurde von Tyler Burge (1979) auf mentale Repräsentationen übertragen. Die Grundidee ist, dass der Inhalt einer Wasser-Repräsentation (genauer: des Begriffs ›Wasser‹, der in unseren Gedanken vorkommt) auf die essenzielle Eigenschaft von Wasser referiert, nämlich auf die Eigenschaft, H_2O zu sein. Diese für das Wort ›Wasser‹ plausible Annahme wird von Burge unkritisch auf den Inhalt der mentalen Repräsentation übertragen: Stellen wir uns eine Welt vor, in der die Flüssigkeit in den Seen und Flüssen nicht H_2O , sondern XYZ ist, dann hätten die Bewohner dieser Welt folglich einen anderen Wasser-Begriff, d. h. andere Wasser-Repräsentationen, deren Inhalt nicht H_2O sondern XYZ herausgreift. Das gilt z. B. auch für einen Zwilling von mir in der XYZ-Welt, der dieselben internen Zustände hat wie ich und dasselbe Verhalten zeigt wie ich. Ob also mein interner Zustand (meine Repräsentation) tatsächlich diesen oder jenen Inhalt hat, so das Argument, hängt davon ab, wie die Welt beschaffen ist, in der wir leben (ob es in ihr H_2O oder XYZ gibt). Eine andere (ebenfalls von Burge vertretene) Version des semantischen Externalismus geht davon aus, dass der Inhalt mentaler Zustände wesentlich von der Sprache abhängt, die wir sprechen, so dass die Inhalte in Abhängigkeit davon variieren, welcher Sprachgemeinschaft wir angehören. Beide Argumente setzen aber letztendlich mehr oder weniger explizit voraus, dass der Inhalt mentaler Repräsentationen stark an den sprachlichen Ausdruck gekoppelt ist, was eine sehr starke und umstrittene Annahme ist (Newen/Vosgerau 2007).

Spezifischere Theorien der Repräsentation versuchen, die Relation zu beschreiben, die zwischen einer Repräsentation und dem Repräsentierten besteht und den Inhalt der Repräsentation festlegt. Es lassen sich grob fünf Familien von Theorien unterscheiden:

Konventionalistische Theorien. Konventionalistische Theorien gehen davon aus, dass die Verbindung zwischen Repräsentation und Repräsentiertem lediglich durch Konventionen festgelegt wird, wie es bei den meisten sprachlichen Ausdrücken der Fall zu sein scheint: dass etwa ›Tisch‹ und ›table‹ für Tische stehen, ist pure Konvention. Für mentale Repräsentationen ist dieser Ansatz allerdings wenig erfolgversprechend, da in einem einzelnen Subjekt bzw. kognitiven System keine Konventionen ›vereinbart‹ werden können.

Kausale Theorien. Kausale Theorien behaupten, dass (mentale) Repräsentationen für etwas Bestimmtes stehen, weil sie davon verursacht wurden. Das Hauptproblem dieser Theorien ist das sog. Disjunktionsproblem: Wenn nur die Verursachung den Inhalt festlegte, dann dürften ausschließlich Pferde-Pferd-Repräsentationen verursachen. Da es Fehlwahrnehmungen gibt, kann das aber nicht sein: In der Dämmerung kann ich z. B. eine Kuh für ein Pferd halten, so dass eine Kuh die Pferd-Repräsentation verursacht. Diese Verursachung müsste kausalen Theorien zufolge genauso relevant sein für die Inhaltsfestlegung, so dass der Inhalt des Begriffs ›Pferd‹ als ›Pferd oder Kuh‹, also als Disjunktion, angegeben werden müsste. Das ist aber, wie Fodor (1994, 60) es ausdrückte, »to put the case mildly, not satisfactory«, da es systematisch zu disjunktiven Inhalten führen würde, die dann keine Erklärungskraft mehr hätten.

In den Neurowissenschaften spielen kausale Theorien für den Repräsentationsbegriff allerdings eine große Rolle. Folgt man einem zweistelligen kausal-korrelativen Repräsentationsbegriff, so gelten neuronale Zustände eines kognitiven Systems dann als repräsentational, wenn sie in kausaler Weise Informationen (d. h. Information im Sinne Dretskes; s. o.) über externe Zustände oder Ereignisse enthalten, oder, anders ausgedrückt, wenn jene (die neuronalen Zustände) durch diese (die externen Zustände) erzeugt wurden. Neuronale Zustände lassen sich mit elektrophysiologischen Mitteln etwa als die Feuerate von einzelnen Neuronen oder Nervenzellen oder mittels sog. funktionell hirnbildgebender Verfahren (wie der funktionellen Magnetresonanztomografie (fMRT)) als bestimmte Aktivitätsverteilungen größerer Nervenzellpopulationen bestimmen. Diese Messergebnisse können mit externen Ereignissen korreliert werden, die dann als ihre Ursache gedeutet werden. In einem neurowissenschaftlichen Kontext kann unter einer Repräsentation daher eine neuronale Abbildung eines externen Ereignis-

nissen verstanden werden, d.h. ein bestimmter physikalischer Zustand im raumzeitlich definierten System des Gehirns. Derartige neuronale Zustände repräsentieren demzufolge externe Ereignisse, Prozesse oder Gegenstände, weil sie in einer ersten Näherung mit ihnen korrelieren (also systematisch gleichzeitig auftreten): Immer, wenn ich z.B. einen Hund sehe, wird in meinem Gehirn ein ähnlicher (Teil-)Zustand auftreten, der dann als Repräsentation des Hundes verstanden wird.

Diese Redeweise sieht sich jedoch erheblichen Schwierigkeiten gegenüber, weil die Rede von repräsentationalen Zuständen lediglich auf ihre kausale Verursachung Bezug nimmt: Von einer kausalen Repräsentation ließe sich nur dann sinnvoll sprechen, wenn empirisch erforscht wäre, welche bestimmten Eigenschaften externer Ereignisse oder Gegenstände spezifisch die entsprechenden neuronalen Zustände verursachen können, was besonders deutlich wird, wenn wir es mit einer ganzen Kette von möglichen Ursachen für einen Gehirnzustand zu tun haben. Hinzu kommt das Problem, dass die neuronale Aktivität nicht immer nur auf externe Ereignisse oder Gegenstände fokussiert ist. Die Vielzahl von neuronalen Prozessen ist vielmehr der internen Verarbeitung geschuldet, so dass ein einfacher kausal-korrelativer Repräsentationsbegriff nicht alle neuronalen Prozesse sinnvoll beschreiben kann. Darüber hinaus gibt es eine Vielzahl von Gehirnzuständen, die durch externe Ereignisse verursacht werden, diese aber nicht repräsentieren (die durch Nikotinkonsum hervorgerufenen Veränderungen im Gehirn sind z.B. keine Repräsentation des Nikotinkonsums).

Ähnlichkeitsbasierte Theorien. Der historisch gesehen wichtigste Ansatz wird durch ähnlichkeitsbasierte Theorien vertreten, die eine bestimmte Ähnlichkeit zwischen Repräsentation und Repräsentiertem für entscheidend halten. Dabei werden ganz unterschiedliche Ähnlichkeiten ins Feld geführt: Angelehnt an bildliche Ähnlichkeit ist die Idee, dass Repräsentationen die Form des Repräsentierten annehmen (z.B. bei Aristoteles). Ludwig Wittgenstein spricht von ›logischen Bildern‹, bei denen in der Repräsentation lediglich die logischen Eigenschaften des Repräsentierten abgebildet werden. In der Kognitionswissenschaft hat sich v.a. die Idee der strukturellen Ähnlichkeit durchgesetzt (Bartels 2005). Dabei werden Strukturen im mathematischen Sinne als Mengen verstanden, über die Relationen (und/oder Funktionen) definiert sind. Strukturelle Ähnlichkeit kann dann im mathematischen Sinne als Isomorphie oder Homomorphie definiert werden.

Die Idee, dass strukturerhaltende Abbildungen den Kern von Repräsentation ausmachen, spielt wiederum v.a. auch in den Neurowissenschaften eine große Rolle: Der visuelle Kortex ist weitgehend topologisch organisiert, d.h. ähnliche Eigenschaften (z.B. ähnliche Farbtöne) werden durch nahe beieinanderliegende Neuronen angezeigt, so dass die Struktur der Wirklichkeit in der Struktur des Gehirns abgebildet zu sein scheint (s. Kap. IV.24). Das Hauptproblem von ähnlichkeitsbasierten Theorien besteht darin, dass sie zwar vielleicht ein notwendiges, aber kein hinreichendes Kriterium angeben können, da Ähnlichkeitsbeziehungen normalerweise symmetrisch sind, die Repräsentationsbeziehung typischerweise aber asymmetrisch ist: Eine mentale Repräsentation steht für etwas, aber dieses etwas steht nicht für die mentale Repräsentation.

Teleosemantische Theorien. Unter der Bezeichnung ›teleosemantische Theorien‹ können Ansätze zusammengefasst werden, die einen bestimmten Zweck der Repräsentation als ausschlaggebend für ihren Inhalt ansehen. Dieser Zweck wird meist evolutionär begründet: Ein bestimmter Gehirnzustand z.B. repräsentiert Gefahr, weil er sich evolutionär als nützlich erwiesen hat, um Gefahren gezielt auszuweichen. Dabei können grob zwei Strategien unterschieden werden: Entweder wird der Zweck einer Repräsentation direkt an evolutionäre ›Ziele‹ wie Überleben und Fortpflanzung geknüpft (z.B. Millikan 1984) oder er wird etwas abstrakter als die Funktion, Information zu tragen, gefasst (z.B. Dretske 1981). Dem zweiten Ansatz liegt wie oben erwähnt die folgende Idee zugrunde: Obwohl Repräsentationen keine Information über das Repräsentierte tragen, werden sie so behandelt, als ob sie es täten – sie haben also die Funktion, Information zu tragen, auch wenn sie es nicht tun, genauso wie ein Dosenöffner die Funktion hat, Dosen zu öffnen, auch wenn er es nicht tut (s.o.). Allerdings stellt sich die Frage, inwieweit diese Funktion tatsächlich evolutionär begründet werden kann oder ob sie nicht letztlich mit der Funktion im Sinne des ersten Ansatzes zusammenfällt. Der erste Ansatz hingegen hat die Schwierigkeit, die enorme Vielzahl und Differenzierung von Inhalten mit den relativ wenigen evolutionären Zielen von Organismen zu erklären – warum und wie sollte z.B. meine Repräsentation der Jupitermonde zur Sicherung meines Überlebens beitragen?

Funktionalistische Theorien. Die ›klassischen‹ kognitionswissenschaftlichen Theorien von Repräsentation sind funktionalistische Theorien, die auf die

mathematische Automatentheorie zurückgehen. Eine bestimmte Klasse solcher Automaten, die sog. endlichen Automaten, können durch eine Übergangsfunktion beschrieben werden, die Paare von Eingaben und Ist-Zuständen auf Paare von Ausgaben und Folgezuständen abbildet und häufig in sog. Maschinentafeln tabellarisch dargestellt wird (vgl. Abb. 1).

<i>Ist-Zu-stand</i>	<i>Eingabe</i>	<i>Ausgabe</i>	<i>Folge-zustand</i>
1	50 Cent	nichts	2
1	1 Euro	Cola	1
2	50 Cent	Cola	1
2	1 Euro	Cola, 50 Cent	1

Abbildung 1: Maschinentafel eines Getränkeautomaten

Der Grundgedanke dieses Ansatzes besteht darin, dass die (internen) Zustände eines Automaten vollständig durch seine charakteristische Abbildung (Funktion) beschrieben werden können. Beispielsweise ›bedeutet‹ Zustand 2 in der Maschinentafel des Getränkeautomaten in Abbildung 1, dass bereits 50 Cent eingeworfen wurden. Dieser ›Inhalt‹ des internen Zustands ergibt sich direkt aus der Tabelle und muss nicht zusätzlich durch eine Eigenschaft des Zustands selbst festgelegt werden. Wenn kognitive Systeme auf diese Weise im Sinne der Automatentheorie beschrieben werden können, dann können auch alle ihre internen Zustände als Repräsentationen in diesem funktionalistischen Sinne gedeutet werden. Schmerzen zu haben, wäre dann etwa derjenige Zustand, der durch eine Verletzung entsteht, ein bestimmtes Verhalten (Schreien, zum Arzt gehen usw.) zur Folge hat usw.

In dieser einfachen Form hat der Ansatz v. a. das Problem, dass es unmöglich erscheint, die Eingabe und Ausgabe eines kognitiven Systems genau und in einer endlichen Tabelle zu erfassen. Aber auch etwas differenziertere Ansätze stehen vor dem Problem, die Feinkörnigkeit der Inhalte und der Verhaltensweisen (Ausgaben) zu spezifizieren: Wenn die Funktionen so feinkörnig bestimmt werden, dass alle Umstände miteinbezogen werden, dann muss letztendlich für jeden Einzelfall eine eigene Funktion eingeführt werden. Damit involvierte aber jede einzelne Episode unseres Verhaltens spezifische Repräsentationen mit ganz individuellem Inhalt, und das Ziel der Verhaltensklärung könnte nicht mehr er-

reicht werden, da jede Episode auch eine individuelle, nicht auf andere Episoden übertragbare Erklärung erhielte. Die (ungelöste) Frage lautet also, auf welcher Abstraktionsstufe wir Inhalte individuieren und wie sich die Wahl einer bestimmten Abstraktionsstufe rechtfertigen lässt.

Einige aktuelle Ansätze fügen der funktionalistischen Grundidee andere, z.B. kausale, Elemente hinzu. So gehen manche Autoren (z.B. Grush 2004) von einem kausal-korrelativ definierten Begriff der *Präsentation* aus, der die sensorischen Eingangssignale bezeichnet, die über unsere Sinnessysteme vermittelt werden und unser Gehirn erreichen. Eine Präsentation vermittelt also Information über ein externes Ereignis in der Umwelt und ist kausal mit diesem verbunden. Demgegenüber ist die *Repräsentation* als ein Modell (eine Emulation) der Umwelt zu verstehen, das gerade keine direkte kausale Verbindung zu dem jeweiligen Umweltzustand aufweist. Diese Überlegung führt unmittelbar zu einem dreistelligen Repräsentationsbegriff, der die Funktion der Repräsentation für das System miteinbezieht. Unter der Annahme eines solchen dreistelligen Begriffs beruhen die Kerneigenschaften von Repräsentationen auf der kausalen Korrelation zwischen dem repräsentierten Gegenstand, dem repräsentierenden Prozess und dem Gebrauch, den das kognitive System von repräsentationalen Zuständen macht (bzw. der Funktion, die die Repräsentation für das kognitive System erfüllt). Damit hat jede Repräsentation auch eine bestimmte Funktion, nämlich diejenige Funktion, die sie dadurch zugewiesen bekommt, dass sie für ein kognitives System handlungsrelevant werden kann (wodurch aber nicht notwendigerweise ein ›höherstufiges‹ Subjekt impliziert wird, das diese Funktionen aufruft). Der repräsentationale Gehalt eines neuronalen Zustands ist demnach nicht identisch mit externen Ereignissen oder Gegenständen, sondern vielmehr ein informativer und potenziell handlungsrelevanter Beitrag für das System. Man spricht oft auch davon, dass Repräsentationen ›funktionale Rollen‹ für ein kognitives System ausüben.

Für die Neurowissenschaften bedeutet dies, dass erst ein solcher dreistelliger Repräsentationsbegriff auch einen heuristischen Wert bekommt, weil er helfen kann, die Frage zu beantworten, wann zwei Vorkommnisse von neuronalen Zuständen zum selben Typ gehören. In den meisten Fällen der forschungspraktischen Anwendung wird der Ausdruck ›Repräsentation‹ verwendet, ohne dass zwischen der Bedeutung von Repräsentationen als einer Klasse (d. h. eines Typs) neuronaler Zustände oder Prozesse, die

durch eine funktionale Rolle oder einen repräsentationalen Gehalt bestimmt ist, und den Zuständen oder Prozessen selbst als Elementen dieser Klasse (d.h. den Vorkommnissen des Typs) klar unterschieden würde. Dahinter verbirgt sich auch ein konkretes methodisches Problem der neurowissenschaftlichen Forschungspraxis: Unter welchen Umständen ist es gerechtfertigt, räumlich oder zeitlich komplexe Muster neuronaler Aktivität als Vorkommnisse derselben Repräsentation anzusehen? Man könnte auch sagen, dass damit die Frage angeschnitten wird, wie neuronale Phänomene eigentlich als Einzelphänomene abgegrenzt werden können.

Mit einem solchen dreistelligen Repräsentationsbegriff steht auch ein Begriffsinstrument zur Verfügung, das es erlaubt, so etwas wie die funktionale Rolle für ein System konzeptuell zu erfassen. Solange nur Merkmale der neuronalen Zustände selbst (Lokalisation, Feuerrate usw.) verwendet werden können, ist konzeptuell nichts erreicht, da neuronale Zustände lediglich in Bezug auf ihre Eigenschaften zu sich selbst in Beziehung gesetzt werden können. Durch die Einführung einer funktionalen Rolle gelingt es jedoch, nach externen Kriterien zu beurteilen, ob etwa ein Unterschied in der räumlichen Verteilung eines neuronalen Aktivitätsmusters für den Status als (Vorkommnis einer bestimmten) Repräsentation relevant ist oder nicht.

Andere Ansätze verbinden die funktionalistische Grundidee mit Anleihen aus ähnlichkeitsbasierten Theorien. Ein Beispiel ist die bereits oben angesprochene Theorie strukturaler Repräsentationen (Bartels 2005), die eine funktionalistische Grundidee mit dem Isomorphieansatz kombiniert. Dabei wird eine strukturelle Ähnlichkeit zwischen Repräsentation und Repräsentiertem als integraler Bestandteil der Funktion der Repräsentation gedeutet, um so die spezifischen Probleme beider Ansätze zu vermeiden.

Eine andere Art der Verbindung unterschiedlicher Elemente stellen sog. Stufentheorien bereit. Deren Grundgedanke lautet, dass es verschiedene Arten von Repräsentationen gibt, die sich in ihrer Struktur und eventuell auch in den Bedingungen ihrer Inhaltsindividuierung unterscheiden. Gleichzeitig sollen diese Arten hierarchisch geordnet sein, so dass sie in einer onto- wie phylogenetischen Entwicklung aufeinander aufbauen. Typischerweise lassen sich die unteren Stufen als nichtbegriffliche Arten von Repräsentationen ausweisen, auf deren Grundlage dann höhere, »begriffliche« Stufen von Repräsentationen entstehen können (s. Kap. IV.4, Kap. IV.9). Auf den verschiedenen Stufen können unterschiedliche Theorien eine Rolle spielen: Bei-

spielsweise können sensorische Repräsentationen im Sinne der kausalen Theorien erfasst werden (s. Kap. IV.24), während begriffliche Repräsentationen am besten durch den Isomorphismusgedanken eingefangen werden können (z. B. Vosgerau 2009). Stufentheorien verbinden also die Frage nach der Individuierung von Repräsentationen mit der Frage, welche Arten von Repräsentationen wie unterschieden werden müssen.

Welche Arten von Repräsentationen gibt es?

Eine wichtige Debatte auf philosophischer Seite beschäftigt sich mit der Frage, ob man grundsätzlich zwei verschiedene Arten von Repräsentationen unterscheiden muss: *begriffliche* und *nichtbegriffliche* (s. Kap. IV.24). Begriffliche Repräsentationen werden meist dadurch charakterisiert, dass das Vorliegen einer solchen Repräsentation voraussetzt, dass man über die Begriffe, mithilfe derer man den Inhalt wiedergibt, selbst verfügen muss. Wenn ich mich z. B. über einen Straftzettel ärgere, dann muss ich über den Begriff des Straftzettels verfügen, d. h. ich muss wissen, was Straftzettel sind. Allerdings kann ein kleines Kind den Straftzettel sehen, ohne über den entsprechenden Begriff zu verfügen. Es hat dann, so wird argumentiert, eine nichtbegriffliche Repräsentation des Straftzettels in Form einer Wahrnehmung. Die Frage, wie genau nichtbegriffliche Repräsentationen zu bestimmen sind, wird nach wie vor intensiv diskutiert, ebenso wie die Frage, was überhaupt Begriffe sind und wie sie bestimmbar sind (s. Kap. IV.9).

In der Psychologie steht man vor dem Problem, anderen kognitiven Systemen auf der Grundlage von Beobachtungen Inhalte und Repräsentationen zuzuschreiben (s. Kap. II.F.1). Im Prinzip kann niemals sicher aus dem Verhalten innerhalb einer (experimentellen) Situation auf die zugrunde liegende(n) Repräsentation(en) geschlossen werden (Anderson 1978). Jedes Verhalten ist immer auch mit alternativen (Arten von) Repräsentationen und darauf operierenden Prozessen erklärbar. Dieses Problem lässt sich anhand der sog. *imagery debate* illustrieren (z. B. Tye 1991). In dieser Debatte ging es um die Frage, ob Menschen nur über propositionale (also satzähnliche) Repräsentationen verfügen oder zusätzlich auch noch auf Repräsentationen in anschaulichem Format zurückgreifen können. Propositionen sind satzähnlich zusammengesetzt aus Begriffen, so dass sie wahr oder falsch sein können, je nachdem, ob der durch die Proposition bezeichnete Sachverhalt be-

steht oder nicht. Dabei werden Propositionen selbst nicht als sprachliche Entitäten aufgefasst, sondern als etwas, das von einer natürlichen Sprache unabhängig ist. Eine anschauliche Repräsentation hingegen lässt sich am einfachsten als mentales Bild verstehen, z. B. als der visuelle Eindruck eines Baumes. Anschauliche Repräsentationen sind dem Repräsentierten in gewissen Aspekten ähnlich, und Teile von anschaulichen Repräsentationen, z. B. das Bild der Baumkrone, sind wiederum anschauliche Repräsentationen, die mit entsprechenden Teilen des repräsentierten Objektes korrespondieren.

Verschiedene empirische Ergebnisse wurden als Hinweise darauf gedeutet, dass wir (auch) anschauliche Repräsentationen zur Lösung bestimmter Probleme benutzen. Die Zeit, die Versuchspersonen brauchen, um ein vorgestelltes Bild abzusuchen, korrespondiert u. a. mit der Entfernung im repräsentierten Bild, wobei das repräsentierte Bild auch von einer Beschreibung gelernt werden kann (z. B. Kosslyn et al. 1978), und die mentale Rotation von Objekten dauert umso länger, je stärker diese rotiert werden (Shepard/Metzler 1971). Die Gegenposition wurde hauptsächlich von Zenon Pylyshyn (1973) verteidigt, der dafür argumentierte, dass diese und andere Phänomene auch mit rein propositionalen Repräsentationen erklärt werden können und die zugrunde liegenden mentalen Prozesse nicht notwendigerweise (wie bei anschaulichen Repräsentationen) kontinuierlich sein müssen. Die Versuchsergebnisse können in Pylyshyns Augen auch dadurch erklärt werden, dass schrittweise neue Perspektiven berechnet werden: Eine solche schrittweise Verarbeitung diskreter, propositionaler Repräsentationen würde ebenfalls zu einer erhöhten Reaktionszeit für längere Distanzen oder Rotationswinkel führen. Aus den Daten allein kann also nicht entschieden werden, ob wir bei der Vorstellung von Bildern oder der mentalen Rotation kontinuierliche, anschauliche oder vielmehr diskrete, propositionale Repräsentationen verwenden. Letztendlich kann die Frage nach dem Repräsentationsformat (anschaulich oder propositional) nur in der Zusammenschau verschiedenster Einzelbefunde aus unterschiedlichen Bereichen zufriedenstellend beantwortet werden. Als Kriterium kann dabei wohl lediglich die über alle Befunde hinweg einfachere Erklärung dienen – bei der allerdings immer Repräsentationen und die auf ihnen operierenden Prozesse gemeinsam betrachtet werden müssen. Selbstverständlich gilt das nicht nur für das angeführte Beispiel der *imagery debate*, sondern in der Kognitionspsychologie im Allgemeinen (s. Kap. II.E.1).

Trotz der erwähnten Probleme bei der empirischen Bestimmung von repräsentationalen Formaten hat sich bezüglich bestimmter Klassen von Repräsentationen ein gewisser Konsens herausgebildet – auch wenn Details dieser Unterscheidungen durchaus umstritten sind. Die grundlegendste Unterscheidung bezieht sich hierbei auf die zeitliche Stabilität der Repräsentation: Man unterscheidet z. B. sensorische Register, Kurzzeit- bzw. Arbeitsgedächtnis und Langzeitgedächtnis (s. Kap. IV.7). Im sensorischen Register werden Sensoreindrücke über die reine Reizdarbietung hinaus aufrechterhalten, wobei sich die Dauer im Bereich von Millisekunden bis Sekunden bewegt. Der Inhalt der Repräsentationen ist weitgehend an den jeweiligen Sensor gebunden, so dass visuelle, auditorische, haptische usw. Register unterschieden werden. Im Arbeitsgedächtnis hingegen verbleiben Repräsentationen über eine Dauer von Sekunden bis Minuten. Neben Repräsentationen, die v. a. der Kontrolle und Verarbeitung dienen, nimmt man auch spezielle Untersysteme an, in denen phonologische, visuelle, räumliche und episodische Inhalte getrennt voneinander repräsentiert werden. Inhalte des Langzeitgedächtnisses schließlich sind nicht prinzipiell zeitlich limitiert, können aber vergessen werden, wobei auf der Basis des Inhaltes u. a. zwischen autobiografischen Erlebnissen, Faktenwissen und motorischen Fertigkeiten unterschieden wird (s. Kap. IV.7).

Der Stellenwert von Repräsentationen in der Kognitionswissenschaft

Die Diskussionen um das richtige Verständnis von Repräsentationen werden mit Sicherheit noch lange andauern. Dabei werden sowohl die Grenzen repräsentationaler Erklärungen als auch die verschiedenen Formate von Repräsentationen eine entscheidende Rolle spielen. Anstatt sich allerdings in einzelnen Detaildebatten zu verzetteln, sollten in Zukunft wieder größere Zusammenhänge diskutiert werden, z. B. die Rolle von Repräsentationen in Erklärungen und Beschreibungen oder das Verhältnis sowie die gegenseitige Beeinflussung oder Abhängigkeit von Repräsentationen, kognitiven Prozessen, körperlichen Zuständen und der Umwelt.

Gerade in dieser Hinsicht kann die Robotik innerhalb der Kognitionswissenschaft sehr viel leisten, da sie neue Ideen aus den verschiedenen Disziplinen in Form kompletter Architekturen von kognitiven Systemen umsetzen und empirisch testen kann. Auf dem Weg zu einer umfassenden Theorie der Reprä-

sentation und des ganzheitlichen Designs von kognitiven Robotersystemen sind zwar bislang nur erste Schritte unternommen worden, es ergeben sich jedoch deutliche Fragestellungen, die schon jetzt (und umso mehr in Zukunft) relevant sind:

- *Das Verhältnis von Körper und Repräsentation* (s. Kap. III.7): Unter welchen Hardwarevoraussetzungen und -beschränkungen kann die Entwicklung welcher Repräsentationen und kognitiven Leistungen überhaupt erwartet werden?
- *Die Rolle von Repräsentationen für Verhaltensklärung*: Wie kann eine kognitive Architektur aussehen, die als Rechnerprogramm darstellbar ist und es erlaubt, handlungszentrierte Kognition auf einem Roboter zu implementieren?
- *Das Verhältnis von Umwelt und internen Repräsentationen*: Wie müssen reale Umwelten mit realen Objekten und Robotern aufgebaut werden, in welchen man die Entwicklung von kognitiven Fähigkeiten über der Zeit beobachten kann, während die Roboter mit dieser Umwelt und möglicherweise auch miteinander agieren?

Bei der Beantwortung dieser Fragen wird es auch weiterhin von zentraler Bedeutung sein, im Sinne der Kognitionswissenschaft die Ergebnisse aus unterschiedlichen Disziplinen in interdisziplinärer Zusammenarbeit zusammenzutragen und zu kombinieren. So werden für die Entwicklung zukünftiger Robotersysteme z. B. nicht nur Daten aus der Verhaltensforschung, sondern auch Erkenntnisse aus der kognitiven Neurowissenschaft (s. Kap. II.D.1) wichtig sein. Die empirischen Disziplinen insgesamt werden von der philosophischen Arbeit der Begriffsklärung und Theoriebildung profitieren, die ihrerseits wiederum fruchtbare Anstöße durch die empirische Forschung erfahren wird.

Der Begriff der Repräsentation hat in der Kognitionswissenschaft von Anfang an eine zentrale Rolle eingenommen. Trotz heftiger Kritik und vielen Auseinandersetzungen, die teilweise bis heute andauern, hat er nichts von seiner zentralen Bedeutung für die Erklärung geistiger Fähigkeiten eingebüßt – selbstverständlich hat er sich aber enorm gewandelt: von einem zunächst sehr allgemeinen Begriff zu einem engen, sehr theoriebeladenen Begriff, und von da aus wieder zurück zu einem weiten, liberalen Begriff. Dabei sind die Diskussionen in den einzelnen Disziplinen nicht immer miteinander verzahnt gewesen, so dass auch mit einiger Berechtigung davon gesprochen werden kann, dass es nicht *einen* Repräsentationsbegriff der Kognitionswissenschaft gibt, sondern verschiedene. Wenn allerdings die Kogniti-

onswissenschaft auch weiterhin als interdisziplinäres Forschungsprojekt bestehen will, dann muss wieder (und kontinuierlich) an einem gemeinsamen Grundverständnis des Repräsentationsbegriffs gearbeitet werden. In diesem Beitrag haben wir vorgeschlagen, einen liberalen Begriff von Repräsentation zur gemeinsamen Grundlage der verschiedenen Disziplinen zu machen, wonach Repräsentationen diejenigen internen Zustände sind, die wir annehmen müssen, um viele verschiedene Arten flexiblen Verhaltens zu erklären.

Literatur

- Anderson, John (1978): Arguments concerning representations for mental imagery. In: *Psychological Review* 85, 249–277.
- Bartels, Andreas (2005): *Strukturelle Repräsentation*. Paderborn.
- Brentano, Franz (1874): *Psychologie vom empirischen Standpunkt*. Leipzig.
- Brooks, Rodney (1990): Elephants don't play chess. In: *Robotics and Autonomous Systems* 6, 3–15.
- Brooks, Rodney (1991): Intelligence without representation. In: *Artificial Intelligence* 47, 139–159.
- Burge, Tyler (1979): Individualism and the mental. In: *Midwest Studies in Philosophy* 4, 73–122.
- Clark, Andy/Toribio, Josefa (1994): Doing without representing? In: *Synthese* 101, 401–431.
- Dehaene, Stanislas/Naccache, Lionel (2001): Towards a cognitive neuroscience of consciousness. In: *Cognition* 79, 1–73.
- Dretske, Fred (1981): *Knowledge and the Flow of Information*. Oxford.
- Eliasmith, Chris (1997): Computation and dynamical models of mind. In: *Minds and Machines* 7, 531–541.
- Fodor, Jerry (1994): *A Theory of Content and Other Essays*. Cambridge (Mass.).
- Grush, Rick (2004): The emulation theory of representation. In: *Behavioral and Brain Sciences* 27, 377–442.
- Kosslyn, Stephen/Ball, Thomas/Reiser, Brian (1978): Visual images preserve metric spatial information. In: *Journal of Experimental Psychology* 4, 47–60.
- Maturana, Humberto/Varela, Francisco (1980): *Autopoiesis and Cognition*. Dordrecht.
- Meilinger, Tobias/Vosgerau, Gottfried (2010): Putting egocentric and allocentric into perspective. In: Christoph Hölscher/Thomas Shipley/Marta Olivetti Belardinelli/John Bateman/Nora Newcombe (Hg.): *Spatial Cognition VII*, LNAI 6222. Heidelberg, 207–221.
- Millikan, Ruth (1984): *Language, Thought, and Other Biological Categories*. Cambridge (Mass.).
- Nagel, Thomas (1974): What is it like to be a bat? In: *Philosophical Review* 83, 435–450.
- Newell, Allen/Simon, Herbert (1976): Computer science as empirical inquiry. In: *Communications of the ACM* 19, 113–126.
- Newen, Albert/Vosgerau, Gottfried (2007): A representational account of self-knowledge. In: *Erkenntnis* 67, 337–353.
- O'Regan, Kevin/Noë, Alva (2001): A sensorimotor account

- of vision and visual consciousness. In: *Behavioral and Brain Sciences* 22, 939–973.
- Paul, Chandana (2006): Morphological computation. In: *Robotics and Autonomous Systems* 54, 619–630.
- Pylyshyn, Zenon (1973): What the mind's eye tells the mind's brain. In: *Psychological Bulletin* 80, 1–24.
- Shepard, Roger/Metzler, Jacqueline (1971): Mental rotation of three-dimensional objects. In: *Science* 171, 701–703.
- Thelen, Esther (1995): Motor development. In: *American Psychologist* 50, 79–95.
- Tolman, Edward (1948): Cognitive maps in rats and men. In: *Psychological Review* 55, 189–208.
- Tye, Michael (1991): *The Imagery Debate*. Cambridge (Mass.).
- van Gelder, Tim (1998): The dynamical hypothesis in cognitive science. In: *Behavioral and Brain Sciences* 21, 615–665.
- van Gelder, Tim/Port, Robert (1995): It's about time. In: Robert Port/Tim van Gelder (Hg.): *Mind as Motion*. Cambridge (Mass.), 1–43.
- Vosgerau, Gottfried (2009): *Mental Representation and Self-Consciousness*. Paderborn.
- Vosgerau, Gottfried/Schlicht, Tobias/Newen, Albert (2008): Orthogonality of phenomenality and content. In: *American Philosophical Quarterly* 45, 309–328.
- Wehner, Rüdiger (1999): Large-scale navigation. In: Christian Freksa/David Mark (Hg.): *Spatial Information Theory*. Berlin, 1–20.

Gottfried Vosgerau/Alois Knoll/
Tobias Meilinger/Kai Vogetley

17. Schlussfolgern, Planen und Problemlösen

Die Gebiete des Schlussfolgerns, Planens und Problemlösens betreffen essenziell das, was generell als ›höhere kognitive Fähigkeiten‹ bezeichnet wird. Die Aufgabe beim Problemlösen besteht im Allgemeinen darin, in einer Situation so zu handeln, dass ein gewisses festgelegtes Ziel erreicht wird. Allerdings würden wir eine Situation, in der es nur eine und zudem offensichtliche Möglichkeit gibt, das gewünschte Ziel zu erreichen, nicht als ein *Problem* ansehen. Zum Lösen eines Problems müssen also mehrere Alternativen zur Wahl stehen und folglich Entscheidungen getroffen werden. Damit solche Entscheidungen nachvollziehbar werden, müssen sie begründet sein: Es muss gezeigt werden, dass sie zur Lösung des Problems beitragen. Oft möchte man auch *optimale* Lösungen haben; solche liegen dann vor, wenn alle anderen Alternativen entweder das Problem nicht lösen oder zu Lösungen führen, deren Kosten (in einem sehr allgemeinen Sinne) mindestens so groß oder größer sind als die der gewählten Lösung. Um mögliche Problemlösungsalternativen zu erschließen und zwischen diesen begründet zu entscheiden, sind Schlussfolgerungsmethoden notwendig. Dies stellt die Beziehung zwischen zwei Hauptbegriffen dieses Beitrags her: *Schlussfolgern* und *Problemlösen*. *Planen*, der dritte Begriff, bezieht sich auf die Situation eines Agenten, der, um ein Ziel zu erreichen, auf Grundlage seines Handlungsinventars Folgen von Handlungen planen muss, die, falls sie in der korrekten Reihenfolge ausgeführt werden, das Erreichen des Ziels garantieren sollen.

Ein weiterer Begriff, der mit dieser Sichtweise eng verbunden ist, ist der Begriff der Rationalität. Insbesondere in den Wirtschaftswissenschaften wird häufig rationales Verhalten der verschiedenen Marktteilnehmer unterstellt (*Homo oeconomicus*) und so verstanden, dass jeder Marktteilnehmer so handelt, dass er seinen eigenen Nutzen optimiert (Franz 2004; s. Kap. III.11, Kap. V.7). Basis einer solchen Optimierung ist eine rationale (d. h. transitive und vollständige) Präferenzordnung (Mas-Colell et al. 1995). Daraus ergibt sich eine Auffassung von rationalem Handeln als (auf ein Ziel bezogen) optimalem Handeln, was wiederum bedeutet, dass ein entsprechender Optimalitätsbegriff bestimmt werden muss. Interessanterweise gibt es sehr unterschiedliche Optimalitätsbegriffe, anhand derer sich unterschiedliche Handlungsoptionen als jeweils ›optimal‹ herausstellen können: