

Universität Bielefeld
Fakultät für Geschichtswissenschaft, Philosophie und Theologie
Abteilung Philosophie

Falsche Überzeugungen verstehen

Begriffliche und methodische Überlegungen
zur Erforschung der alltagspsychologischen Praxis

Dissertation von

Alexander Hendrik Mäder

Gutachter:

Prof. Dr. Martin Carrier

Prof. Dr. Ansgar Beckermann

eingereicht am 16. September 2002

verteidigt am 19. September 2003

Vorbemerkung

Der Hauptteil dieser Arbeit ist am Institut für Wissenschafts- und Technikforschung der Universität Bielefeld entstanden, wo ich von 1999 bis 2002 Stipendiat der Deutschen Forschungsgemeinschaft war. Ich danke Matthias Adam, Ralf Busse, Michael Schütte und Torsten Wilholt für ihre fachliche und moralische Unterstützung in dieser Zeit. Meinem Doktorvater Martin Carrier danke ich dafür, diese wissenschaftstheoretische Studie begleitet und gefördert zu haben.

Berlin, 15. März 2005

Inhalt

Einleitung: Die Erforschung der alltagspsychologischen Praxis	6
Zwei Erklärungsmodelle für den Umgang mit falsche Überzeugungen · Ziel der Arbeit · Aufbau der Arbeit	
I Das Theorienmodell und der Überzeugungsbegriff	
Vorbemerkung: Zur Analyse der Testanforderungen	11
Der <i>False Belief Test</i> · Kompetenz und Performanz · Varianten und Analoga	
1 Das kognitive Theorienmodell	17
1.1 Alltagspsychologische Theorien anwenden	18
Alltagspsychologische Platitüden · Die Analogie zur Linguistik	
1.2 Empirische Prüfung des Theorienmodells	23
Theorien- vs. Simulationsmodell: eine Scheindebatte? · Kognitive Architektur	
2 Überblick über die empirische Forschung	28
2.1 Der Alterseffekt im <i>False Belief Test</i>	28
Die Robustheit des Alterseffekts · Selbstzuschreibungen falscher Überzeugungen · Vereinfachte Varianten des <i>False Belief Tests</i>	
2.2 Die Theorien zwei- und dreijähriger Kinder	35
Die Wahrnehmungs-Wunsch-Theorie zweijähriger Kinder · Ein frühes Verständnis falscher Überzeugungen? · Das Verständnis des epistemischen Zugangs	
2.3 Überzeugung, Wahrnehmung und Repräsentation	43
Perspektivität und Erscheinung · Kippbilder und veraltete Fotos	
3 Der Erwerb des Überzeugungsbegriffs	47
3.1 Die Überzeugungsrelation verstehen	47
Perners Erklärung des Alterseffekts · Die Unterscheidung zwischen Fehlern und <i>Pretend Play</i>	
3.2 Die Bestimmung von Überzeugungsgehalten	52
Wimmers Erklärung des Alterseffekts · Epistemischer Zugang und Überzeugungsbegriff	

3.3	Eine repräsentationale Theorie des Geistes?	57
	Merkmale des theoretischen Wandels · Die Entwicklung einer repräsentationalen Theorie des Geistes	
	Fazit: Unerklärte Daten und gesuchte Indikatoren	61
II	Die Diskussion um das Simulationsmodell	
	Vorbemerkung: Was leistet die Alltagspsychologie?	64
	Zwei Formen der Simulation · Hybridmodelle · Die Herausforderung	
4	Die Grundidee des Simulationsmodells	70
4.1	Faktische und hypothetische Simulation	70
	Das Verhältnis der beiden Simulationstypen · Die Drei-Schritt-Analyse des Simulationsprozesses · Erklärungen durch Simulation	
4.2	Die Grenzen der Simulation	75
	Die Zuschreibung des Simulationsergebnisses · Das Erfassen psychologischer Unterschiede · Die Ähnlichkeitsbedingung	
5	Simulation als "Mitdenken"	80
5.1	Ein Argument gegen das Theorienmodell	81
	Holismus epistemischer Verknüpfungen · Theorien über epistemische Verknüpfungen · Das Hintergrundwissen heranziehen	
5.2	Reichweite und Tiefe der Alltagspsychologie	86
	Eine verdeckte Prämisse · Reaktion des Theorienmodells · Was wird simuliert?	
6	Versuche einer empirischen Unterscheidung	90
6.1	Das Kriterium der kognitiven Durchdringbarkeit	90
	Falsche alltagspsychologische Theorien · Die Fehlbarkeit der Simulation · Unbestimmte alltagspsychologische Theorien	
6.2	Hinweise zur Theorieanwendung	94
	Das Hervorheben relevanter Faktoren · Die Vorhersage von Schätzungen · Spielraum bei der Interpretation · Die Rolle alltagspsychologischer Intuitionen	
	Fazit: Zum Nachweis von Prozessmerkmalen	100

III Die kognitiven Grundlagen der Simulation

	Vorbemerkung: Simulation und <i>Pretend Play</i>	102
	Das Rollenspiel · Normales <i>Pretend Play</i> · Der Erleichterungseffekt im kontrafaktischen Schließen	
7	Die flexible Abgrenzung hypothetischer Überzeugungen	110
	7.1 Ähnlichkeit und Abgrenzung	110
	Metarepräsentationen und <i>Pretend</i> -Zustände · Die logische Form von <i>Pretend</i> -Zuständen	
	7.2 Die Fähigkeiten autistischer Kinder	115
	Die Fähigkeit zum kontrafaktischen Schließen · Leslies Erklärung des FBT-Alterseffekts	
	7.3 Kontrafaktisches Schließen	120
	Modifizierte Schlussfolgerungsprozesse · Kontrafaktisches Schließen und <i>Pretend Play</i>	
8	Empirische Indikatoren für Theorien- und Simulationsmodell	125
	8.1 Der Umgang mit kontrafaktischen Prämissen	125
	Vertieftsein in hypothetische Szenarien · Der Erleichterungseffekt im kontrafaktischen Schließen · Unverständliche Prämissen	
	8.2 Zwei neue Varianten des <i>False Belief Tests</i>	130
	Der P-FBT und das Simulationsmodell · Ausschluss des Theorienmodells	
	Fazit: Die Simulation falscher Überzeugungen	136
	Zusammenfassung: Muster in der kognitiven Entwicklung	138
	Literaturangaben	143

Einleitung: Die Erforschung der alltagspsychologischen Praxis

Auf den ersten Blick sind psychologische Überlegungen im Alltag oft nur daran zu erkennen, dass triviale, tiefgründige oder auch unplausible psychologische Weisheiten zitiert werden, um Handlungen zu erklären oder sie vorherzusagen. Die interessanten alltagspsychologischen Leistungen sitzen jedoch tiefer. Jemanden zu täuschen setzt zum Beispiel voraus, dass man einige Bedingungen dafür kennt, wie Überzeugungen ausgebildet werden, um mit diesem Wissen die andere Person zu motivieren, eine falsche Überzeugung auszubilden. Um solche grundlegenden Leistungen soll es in dieser Arbeit gehen, und die zentrale Frage dabei lautet: Welche kognitiven Mechanismen sind geeignet, um diese Leistungen zu erklären? Die Antwort liegt meines Erachtens nicht auf der Hand.

Der Hintergrund für meine Untersuchung ist der folgende: In den letzten zwanzig Jahren wurden zwei grundsätzlich verschiedene kognitionswissenschaftliche Theorien der alltagspsychologischen Praxis diskutiert.¹ Das Theorienmodell (im Englischen: *theory-theory*) postuliert Theorien über psychologische Zusammenhänge, die Menschen in ihrer alltagspsychologischen Praxis anwenden. Dem Simulationsmodell (*simulation-theory*) zufolge versetzen sich Menschen in ihrer Einbildung in die Situation anderer Personen und vollziehen in diesem Rahmen deren Gedankengänge nach.² Zur Frage, welches Modell besser gestützt ist, sind zahlreiche empirische und theoretische, psychologische und philosophische Studien erschienen. Dennoch erscheint die Forschungslandschaft recht unübersichtlich; mir ist keine Position bekannt, die allen einschlägigen Daten und Argumenten gerecht wird. Daher will ich in dieser Arbeit versu-

¹ Eine erste Anmerkung zur Terminologie: Da 'Alltagspsychologie' auch als Psychologie *des Alltagslebens* verstanden werden kann, ziehe ich ihr die Bezeichnung 'alltagspsychologische Praxis' vor. Ein weiterer Grund für diese Wahl ist, dass in der philosophischen Literatur mit 'Alltagspsychologie' (im Englischen: *folk psychology*) auch die implizite Theorie bezeichnet wird, die der alltagspsychologischen Praxis (möglicherweise) zugrunde liegt. Mit 'alltagspsychologischer Praxis' werde ich jedoch nur die Praxis des alltagspsychologischen Erklärens und Vorhersagens von Handlungen und des Zuschreibens geistiger Zustände bezeichnen.

² Die Übersetzungen 'Theorienmodell' und 'Simulationsmodell' stammen von mir. Da es so gut wie keine deutsche Literatur zu diesem Thema gibt, habe ich mir diese freie Übersetzung erlaubt, ohne damit jedoch eine inhaltliche Aussage verbinden zu wollen.

chen, die begrifflichen und methodischen Voraussetzungen einer Entscheidung zwischen beiden Modellen zu klären.

Zwei Erklärungsmodelle für den Umgang mit falschen Überzeugungen

Wie der Titel der Arbeit bereits andeutet, werde ich mich vor allem mit dem Verständnis falscher Überzeugungen befassen. Alle anderen Aspekte der alltagspsychologischen Praxis lasse ich außen vor, weswegen ich auch auf eine systematische Darstellung der – vermutlich sehr heterogenen – alltagspsychologischen Praxis verzichte. Die zentrale Frage lautet also, welche kognitiven Mechanismen dafür verantwortlich sind, dass Menschen die gelegentlich falschen Überzeugungen Anderer bestimmen können (und auch vorhersagen können, welche Schlüsse sie aus diesen Überzeugungen ziehen und für welche Handlungen sie sich auf dieser Grundlage entscheiden werden). Es geht also kurz gesagt darum, welche kognitiven Mechanismen den Umgang mit den falschen Überzeugungen Anderer steuern und damit die kognitive Grundlage des Verständnisses falscher Überzeugungen darstellen.

Wie sehen die Erklärungen im Groben aus, die die beiden Modelle für diesen Aspekt der alltagspsychologischen Praxis anbieten? Das Theorienmodell behauptet, dass Menschen Theorien darüber besitzen, wie Überzeugungen kausal mit anderen geistigen Zuständen, Umweltreizen und Handlungen zusammenhängen, und dass sie diese Theorien auch *anwenden*, wenn es in der alltagspsychologischen Praxis um Überzeugungen geht. Dabei ist es unerheblich, ob Menschen die Theorien bewusst sind; man geht davon aus, dass auch die Anwendung *impliziter* Theorien die alltagspsychologische Praxis steuern kann – wie es zum Beispiel auch für den Bereich der Alltagsphysik behauptet wird.

Das Simulationsmodell spezifiziert hingegen einen übergeordneten kognitiven Mechanismus, der diejenigen kognitiven Mechanismen ansteuert, die für das Verarbeiten der *eigenen* Überzeugungen zuständig sind. Er versorgt sie mit eingebildetem Input, der die Situation der anderen Person widerspiegelt, lässt sie daraufhin ungestört ablaufen und verarbeitet das Ergebnis des simulierten Gedankengangs schließlich weiter zu einer alltagspsychologischen Erklärung oder Vorhersage. Soll etwa vorhergesagt werden, welche Schlussfolgerung die andere Person aus einer falschen Überzeugung ziehen wird, stellt man sich vor, man unterhalte diese falsche Überzeugung selbst, und zieht Schlüsse daraus wie man auch sonst Schlüsse ziehen würde. Statt auf Theorien greift man da-

mit indirekt auf psychologische Regularitäten zurück, die durch die eigenen kognitiven Mechanismen instantiiert werden, um das Simulationsergebnis richtig einzuschränken.³

Ziel der Arbeit

Die Forschung zum Verständnis falscher Überzeugungen sowie zur Entscheidung zwischen Theorien- und Simulationsmodell ist sehr umfangreich und unübersichtlich. Das liegt vor allem daran, dass neben der alltagspsychologischen Praxis auch verwandte kognitive Fähigkeiten untersucht werden. Insbesondere wird eine enge Beziehung zwischen dem Umgang mit falschen Überzeugungen und dem Schließen aus kontrafaktischen Prämissen vermutet. Einschlägig ist aus Sicht des Simulationsmodells zudem die Einbildung, d.h. die Fähigkeit zum Aufbauen und Ausarbeiten hypothetischer Szenarien.⁴ Da diese Fähigkeiten zu unterschiedlichen Zeitpunkten in der kindlichen Entwicklung auftreten, gilt es ein Entwicklungsmuster zu erklären.

Mein eigener Vorschlag dazu ist dieser: Die Einbildung bzw. das *Pretend Play* stellt die für den Umgang mit falschen Überzeugungen Anderer und mit kontrafaktischen Prämissen entscheidende kognitive Grundlage dar. Man verarbeitet falsche Überzeugungen und kontrafaktische Prämissen, indem man sich ein Szenario vorstellt, in dem diese Überzeugungen oder Prämissen wahr sind. Dieser Zusammenhang passt gut zum beobachteten Entwicklungsmuster: Über ein Verständnis des *Pretend Play* verfügen Kinder nämlich früher als über die Fähigkeit zum kontrafaktischen Schließen und zum Umgang

³ Ich möchte der verbreiteten Auffassung entgegentreten, das Simulationsmodell bestreite die Existenz oder die Anwendung alltagspsychologischer Theorien. Denn Menschen verfügen offensichtlich über zahlreiche alltagspsychologische Theorien (etwa über Partnerschaften), die sie im Alltag auch anwenden (z.B. wenn sie Freunden Ratschläge erteilen). Ebenso ist es mit dem Simulationsmodell verträglich, dass Menschen Theorien erwerben, die die bisher durchgeführten Simulationen ersetzen. Das Simulationsmodell behauptet vielmehr, dass es die kognitiven Mechanismen zur Simulation gibt und dass sie in der alltagspsychologischen Praxis eine wichtige Rolle spielen. Das Simulationsmodell wäre schon dann eine interessante Alternative zum Theorienmodell, wenn es Aspekte der alltagspsychologischen Praxis erklären könnte, die das Theorienmodell nicht (oder zu diesem Zeitpunkt in der kognitiven Entwicklung *noch* nicht) erklären kann.

⁴ Statt der Einbildung im Allgemeinen werde ich in dieser Arbeit das So-tun-als-ob (im Englischen: *Pretend Play*) untersuchen. Es gibt natürlich Unterschiede zwischen diesen Fähigkeiten. Zum Beispiel ist es möglich, so zu tun als sei man ein Huhn, indem man mit angewinkelten Armen auf und ab schlägt, ohne sich dabei vorzustellen, man sei ein Huhn. In dieser Arbeit soll nur der Einbildungsaspekt des *Pretend Play* eine Rolle spielen. Das typische Beispiel wird sein: so zu tun, als sei ein leeres Glas voll.

mit falschen Überzeugungen, die beide etwa gleichzeitig im Alter von vier Jahren erstmals auftreten. Allerdings bleibt noch der Faktor zu spezifizieren, der dafür verantwortlich ist, dass Kinder nicht schon dadurch mit falschen Überzeugungen umgehen können, dass sie das *Pretend Play* beherrschen.

Dieser Vorschlag ist meiner Ansicht nach im Kern ein Simulationsmodell, da das *Pretend Play* wesentliche Merkmale des Simulierens erfüllt. Das Ergebnis dieser Arbeit wird also die These sein, dass das Simulationsmodell zumindest den alltagspsychologischen Umgang mit falschen Überzeugungen gut erklärt. Gerade in diesem Bereich, so werde ich versuchen zu zeigen, hat das Theorienmodell Schwierigkeiten, da die zugrunde liegende alltagspsychologische Theorie sehr komplexe Beziehungen von Überzeugungen untereinander erfassen müsste. Am Ende der Arbeit werde ich ein experimentelles Design entwickeln, mit dem sich das Simulationsmodell für den Bereich der Überzeugungen überprüfen lässt.

Aufbau der Arbeit

Der Aufbau der Arbeit orientiert sich nicht in erster Linie an den skizzierten argumentativen Zusammenhängen, sondern vielmehr an der üblichen thematischen Einteilung. Dieses Vorgehen halte ich angesichts der Fülle des Materials für unumgänglich. Bei der Wahl der Reihenfolge habe ich aber versucht, Querverweise und Exkurse in der Darstellung möglichst zu vermeiden. Darüber hinaus habe ich jedem der drei Teile eine eigene Einleitung gegeben, in der ich versuche, die übergeordneten argumentativen Linien aufzuzeigen.

Im ersten Teil stelle ich die wesentlichen Daten und Theorien der *Theory-of-Mind*-Forschung vor (2. und 3. Kapitel). Dabei wird es vor allem um den Alterseffekt im sogenannten *False Belief Test* gehen. Ich werde kritisch prüfen, ob sich dieser Entwicklungsschritt durch einen Wandel in der zugrunde liegenden alltagspsychologischen Theorie erklären lässt. Dieser Diskussion werde ich einige allgemeine Bemerkungen zum Theorienmodell voranstellen (1. Kapitel).

Im zweiten Teil stelle ich die eigentliche Debatte zwischen Theorien- und Simulationsmodell vor. Nach einigen allgemeinen Bemerkungen zum Simulationsmodell (4. Kapitel) werde ich auf begriffliche und empirische Argumente in dieser Debatte eingehen (5. und 6. Kapitel). Ein wichtiges Ergebnis wird sein, dass die beiden Modelle keine direkte

empirische Überprüfung erlauben, weil der Spielraum bei der Interpretation der Daten zu groß ist. Vielmehr muss auf Ergebnisse der *Theory-of-Mind*-Forschung zurückgegriffen werden, um die möglichen Interpretationen einzuschränken.

Diesen Ansatz werde ich im dritten Teil verfolgen. Dort werde ich zunächst auf einige Versuche eingehen, die kognitiven Grundlagen der Simulation genauer zu beschreiben (7. Kapitel). Dabei werde ich auf die o.g. Zusammenhänge zwischen Alltagspsychologie, *Pretend Play* und kontrafaktischen Schließen zu sprechen kommen. Im letzten Kapitel werde ich schließlich das angekündigte Simulationsmodell und ein experimentelles Design zur Überprüfung vorstellen.

I Das Theorienmodell und der Überzeugungsbegriff

Vorbemerkung: Zur Analyse der Testanforderungen

Das Theorienmodell, das ich in dieser Arbeit diskutieren werde, besagt, dass Menschen alltagspsychologische Theorien anwenden, wenn sie Alltagspsychologie betreiben.⁵ Es erklärt die Vorhersage und Erklärung von Handlungen sowie die Zuschreibung geistiger Zustände, indem es einen dafür zuständigen *kognitiven Mechanismus* der Theorieanwendung postuliert. In diesem ersten Teil der Arbeit werde ich die Daten vorstellen, auf die das Theorienmodell meist gestützt wird, und werde untersuchen, wie gut es die alltagspsychologischen Fähigkeiten erklärt. Das skeptische Ergebnis, zu dem ich gelangen werde, soll die Formulierung des Simulationsmodells motivieren, einer Alternative zum Theorienmodell, die ich im zweiten Teil der Arbeit vorstelle.

Die Daten stammen aus Experimenten, in denen einzelne alltagspsychologische Fähigkeiten von Kindern getestet werden – vor allem die Fähigkeit, anderen Menschen *falsche Überzeugungen zuzuschreiben* und damit deren realitätsunangemessene Handlungen vorherzusagen. Diese Fähigkeit ist von besonderem Interesse, weil sie ein echtes Verständnis geistiger Vorgänge anzuzeigen scheint. Denn falsche Überzeugungen können untypische, neuartige Handlungen hervorrufen, die sich nicht unter Rückgriff auf einfache Verhaltensregularitäten vorhersagen lassen; ihre Vorhersage verlangt ein Verständnis des kausalen Zusammenhangs zwischen Überzeugungen und Handlungen. Wenn Kinder also realitätsunangemessene Handlungen vorhersagen können – so der Ansatz –, zeigen sie damit ihr Verständnis dieses Zusammenhangs.

In der Forschungsliteratur, die ich in den folgenden drei Kapiteln diskutieren werde, wird dieses Verständnis als *theoretisches* Verständnis charakterisiert. Der Vorteil von Theorien liegt auf der Hand: Sie erlauben die systematische Beschreibung disparat erscheinender Phänomene sowie die Vorhersage vergleichsweise seltener Effekte. Daher

⁵ Mit 'Theorie' sind in dieser Arbeit meist die Theorien gemeint, deren Kenntnis Alltagspsychologen zugeschrieben wird. Theorien *über* alltagspsychologische Theorien und ihre Anwendung heißen 'Erklärungen', 'Positionen' oder 'Modelle', ohne dass ich mit dieser Wortwahl einen anderen Zweck verfolgen würde, als den Text lesbarer zu gestalten.

ist es *prima facie* plausibel, eine Fähigkeit wie die, realitätsunangemessene Handlungen vorherzusagen, durch die Kenntnis einer Theorie zu erklären, die das kausale Profil von Überzeugungen in wesentlichen Zügen erfasst.

Der False Belief Test

Um zu belegen, dass Menschen tatsächlich eine Theorie dieser Art anwenden, wenn sie Anderen falsche Überzeugungen zuschreiben, wird oft der *False Belief Test* (kurz: FBT) herangezogen. Dieser Test ist ein wichtiges Werkzeug im Forschungsprogramm 'Theory of Mind', das sich dem Aufdecken der alltagspsychologischen Theorien von Kindern widmet. Das Forschungsinteresse richtet sich vor allem auf Kinder, weil das theoretische Verständnis Erwachsener ungleich komplexer und damit einer Analyse nur schwer zugänglich ist. Wie sich zeigen wird, sind schon die alltagspsychologischen Leistungen drei- und vierjähriger Kinder so voraussetzungsreich, dass sie nur indirekt über mehrere Indizien zu erklären sind.

Im FBT müssen Kinder die Handlung einer Person vorhersagen, der eine entscheidende Information vorenthalten wurde. Typischerweise wird Kindern ein Szenario vorgeführt, in dem eine Protagonistin abwesend ist, während ein Objekt ihrer Begierde versteckt wird. Die Kinder werden anschließend gebeten, vorherzusagen, wo die Protagonistin nach dem Objekt suchen werde. Entscheidend ist, ob die Kinder bei ihrer Vorhersage berücksichtigen können, dass die Protagonistin nicht wissen kann, wo das Objekt versteckt worden ist, und entsprechend antworten, dass sie es dort suchen werde, wo es vorher war. Mit dem Test wird also geprüft, ob Kinder verstehen, dass Handlungen auch auf Überzeugungen beruhen können, die inzwischen überholt und damit falsch geworden sind.

Der überraschende Befund des FBT ist ein *Alterseffekt*: Im Alter von vier Jahren sind Kinder auf einmal dazu in der Lage, realitätsunangemessene Handlungen aufgrund falscher Überzeugungen richtig vorherzusagen. Bis zu diesem Alter geben Kinder indes an, dass die Protagonistin das Objekt ihrer Begierde in seinem gegenwärtigen Versteck suchen werde. Sie begehen also systematisch einen Fehler. Da es sich um einen robusten Befund handelt, ist es möglich, auf die genauen Altersangaben zu verzichten. Ich werde in dieser Arbeit stets von drei- und vierjährigen Kindern sprechen und damit diejenigen Kinder meinen, die beim FBT noch durchfallen bzw. die den FBT bestehen.

Der Alterseffekt wird oft durch den Erwerb eines alltagspsychologischen *Begriffs* erklärt, zum Beispiel durch den Erwerb des Überzeugungsbegriffs. Der Überzeugungsbegriff wird dabei als theoretischer Begriff aufgefasst, der durch eine Reihe alltagspsychologischer Theorien definiert wird. Vierjährige Kinder fassen Überzeugungen demzufolge als Typ geistiger Zustände auf, die auf eine charakteristische Weise mit Umweltreizen, Wünschen und Handlungen kausal verbunden sind.

Das Problem, die Theorien vierjähriger Kinder auszubuchstabieren, wird in dieser Arbeit nicht meine primäre Aufgabe sein. Vielmehr möchte ich der Frage nachgehen, wie begründet werden kann, dass Kinder die (wie auch immer gearteten) Theorien *anwenden*, um die Aufgabe im FBT zu lösen.⁶ Ein starker Grund dafür ist der beobachtete Alterseffekt: Es hat den Anschein, als würden Kinder im Alter von vier Jahren ihre alltagspsychologische Kompetenz so sprunghaft ausweiten, wie man es vom Erwerb theoretischer Begriffe kennt. Es gibt daher – so die Vertreter des Theorienmodells – keine bessere Erklärung als die, dass vierjährige Kinder eine geeignete alltagspsychologische Theorie lernen und auch anwenden.⁷ An verschiedenen Stellen wird es in dieser Arbeit trotzdem nötig sein, konkrete alltagspsychologische Theorien zu diskutieren (vor allem in den Kapiteln 3 und 6).

Kompetenz und Performanz

Im ersten Teil der Arbeit will ich das Theorienmodell zurückweisen. Der FBT misst meines Erachtens nicht den Erwerb einer alltagspsychologischen Theorie; der Alterseffekt stützt daher auch nicht die These, dass alltagspsychologische Theorien angewendet werden, um den FBT zu lösen. Dreijährige Kinder besitzen meines Erachtens schon die *Kompetenz*, realitätsunangemessene Handlungen vorherzusagen, sie können diese

⁶ Im Abschnitt 1.1 werde ich diesen Erklärungsanspruch ausführlicher diskutieren und als 'kognitives Theorienmodell' bezeichnen. Denn es ist durchaus möglich, Theorien zu formulieren, mit deren Hilfe sich die beobachteten alltagspsychologischen Erklärungen und Vorhersagen ableiten lassen, ohne dass damit der Anspruch verbunden wäre, dass diese Theorien im kognitiven Haushalt von Alltagspsychologen vorkommen.

⁷ Auch aus der selektiven Störung einzelner Fähigkeiten lassen sich Hinweise über die kognitiven Grundlagen der Alltagspsychologie gewinnen. Hier gilt das Interesse vor allem den Leistungen autistischer Kinder, denen der FBT große Schwierigkeiten bereitet. Auf diesen Punkt werde ich im Abschnitt 7.2 zu sprechen kommen.

Kompetenz aber noch nicht ausüben. In der *Theory-of-Mind*-Forschung spricht man in diesem Fall von einem *Performanzdefizit*; Fortschritte in der kognitiven Entwicklung können auch Fortschritte in der Performanz sein.⁸

Vertreter dieser Position ziehen die Möglichkeit in Betracht, dass nur solche Aspekte des FBT Schwierigkeiten für Dreijährige darstellen, die unabhängig von ihren alltagspsychologischen Fähigkeiten sind. Zum Beispiel wäre denkbar, dass die Testfragen von Kindern sprachlich noch nicht richtig verstanden werden, im Prinzip aber richtig beantwortet werden könnten. Erklärungen diesen Typs werden 'Early-Competence-Accounts' genannt; die Erklärung durch den Erwerb eines Überzeugungsbegriffs gehört hingegen zum Typ 'Conceptual-Change-Account'.⁹ Ich zähle meine Position zu den *Early-Competence-Accounts*, werde mich aber im dritten Teil dieser Arbeit von den vorherrschenden Positionen dieses Typs abgrenzen.

Im folgenden möchte ich die methodischen Handgriffe erläutern, mit denen die verschiedenen denkbaren Theorienmodelle und ihre Alternativen gestützt werden können. In jedem Fall muss der Faktor bestimmt werden, der für das Bestehen des FBT *entscheidend* ist. Entscheidend ist der Faktor, der Kindern trotz ansonsten hinreichender Kompetenz noch fehlt, um den Test zu bestehen. Um zwischen *Early-Competence-* und *Conceptual-Change-Accounts* zu entscheiden, muss geklärt werden, ob der entscheidende Faktor zur alltagspsychologischen Kompetenz gehört.

Varianten und Analoga

Der erste Schritt in der Suche nach entscheidenden Faktoren ist die *Analyse der Testanforderungen*. Aus einer möglichst genauen Beschreibung dessen, was zum Bestehen des FBT (oder eines anderen *Theory-of-Mind*-Tests) nötig ist, werden Hypothesen darüber aufgestellt, worin die Schwierigkeiten liegen könnten, die Kinder haben, die beim Test noch durchfallen. In einem zweiten Schritt werden dann abgewandelte Tests durchgeführt, die leichter zu bestehen sein sollten, weil in ihnen der vermeintlich ent-

⁸ *Erworben* oder *gelernt* werden in meiner Terminologie nur Kompetenzen, nicht aber Fähigkeiten. Der Erwerb einer Kompetenz kann das erstmalige *Auftreten* einer Fähigkeit erklären.

⁹ Die gelegentlich diskutierte Frage, ob einige alltagspsychologische Kompetenzen angeboren sind, wird in dieser Arbeit keine Rolle spielen. Im Vordergrund werden die alltagspsychologischen Leistungen drei- und vierjähriger Kinder stehen.

scheidende Faktor nicht mehr benötigt wird. Der empirische Nachweis des entscheidenden Faktors erfordert also eine experimentelle Manipulation, die spezifisch die Leistungen derjenigen Kinder steigert, die den Test noch nicht bestehen. Im Fall der *Early-Competence-Accounts* ist dieses Vorgehen leicht zu verstehen.¹⁰ Wellman, Cross & Watson (2001) fassen es in ihrer Meta-Analyse von FBT-Studien so zusammen:

Empirical confirmation of early competence accounts requires that there be some version of the target task that indeed demonstrates enhanced performance by young children, when task limitations have been eliminated or reduced. (Wellman, Cross & Watson 2001, 672)

Vertreter eines *Conceptual-Change-Accounts* müssen hingegen einen Test konstruieren, der dem FBT zwar ähnelt, aber ein *nicht-psychologisches Problem* betrifft und damit den Überzeugungsbegriff (oder eine bestimmte alltagspsychologische Theorie) nicht voraussetzt. Solche Tests möchte ich 'Analoga' des FBT nennen – im Unterschied zu den oben erwähnten *Varianten*, die dem FBT bis auf ein nicht-psychologisches Detail ähneln. Würden Analoga von dreijährigen Kindern bestanden, wäre das ein Hinweis darauf, dass der Überzeugungsbegriff diejenige Ressource darstellt, die ihnen noch fehlt, um den FBT zu bestehen.

Im bekanntesten Test dieser Art nimmt ein Fotoapparat die Rolle der unwissenden Protagonistin ein; das Foto eines inzwischen veränderten Objekts spielt die Rolle einer veralteten Überzeugung. Doch der experimentelle Befund widerspricht dem Theorienmodell: Die nicht-psychologischen FBT-Analoga liefern einen ähnlichen Alterseffekt wie der FBT selbst. Man spricht von einer 'Korrelation' der Alterseffekte und meint damit das gleichzeitige Auftreten der beiden entscheidenden Faktoren in der kognitiven Entwicklung des Kindes.¹¹

Statt nachzuweisen, dass der entscheidende Faktor des FBT psychologischer Natur ist, können Vertreter eines *Conceptual-Change-Accounts* auch versuchen, nicht-alltagspsychologische Kandidaten für den entscheidenden Faktor *auszuschließen*. Es kann zum

¹⁰ Eine weitere, wichtige Bedingung lautet natürlich: *Early-Competence-Accounts* sollten auch die falschen Antworten dreijähriger Kinder erklären können, da diese nicht nur auf ein fehlendes, sondern auf ein *systematisch falsches* Verständnis hinweisen. Darauf gehe ich im dritten Teil der Arbeit ein.

¹¹ Die hier vorgestellte Korrelation wird von einigen Autoren dennoch als Stütze des Theorienmodells gewertet, da sie – so die Erklärung – einen umfassenden Wandel in der kindlichen Theorie des Geistes anzeigen: den Wandel hin zu einem Verständnis von *Repräsentationen*. Im dritten Kapitel werde ich darlegen, dass diese Erklärung zahlreiche Probleme mit sich bringt. Einwände empirischer Natur gegen den Korrelationsbefund werde ich im Abschnitt 2.3 vorstellen.

Beispiel mit Hilfe von Kontrollfragen gezeigt werden, dass dreijährige Kinder die Testfrage sprachlich verstehen und dass sie sich daran erinnern können, wo das Objekt der Begierde der Protagonistin war, bevor es versteckt wurde. In den Worten von Wellman, Cross & Watson lautet die Empfehlung:

Conceptual change accounts also require specific empirical findings. To begin with, a task is needed that plausibly assesses a target conceptual understanding [...]. Confidence in both good and poor performance on the task must additionally be bolstered by correct judgments on control tasks or on control questions that demonstrate memory for key information and grasp of the task format. (Wellman, Cross & Watson 2001, 672)

Beide Strategien – der Nachweis und der Ausschluss entscheidender Faktoren – werden dadurch erschwert, dass die Analysen der Testanforderungen oft keine eindeutige Auszeichnung des entscheidenden Faktors erlauben. Auch die eigens konstruierten Varianten und Analoga eines Tests können mehrere Interpretationen zulassen und damit auf mehrere Kandidaten für den entscheidenden Faktor hinweisen. Dementsprechend gibt es in der Forschungsliteratur eine Fülle von Erklärungsansätzen. Es scheint oft nur eine Sache des Einfallsreichtums zu sein, um neue Aspekte der Testanforderungen zu finden und diese für die Schwierigkeiten jüngerer Kinder verantwortlich zu machen. Eine erfolgreiche Erklärung des FBT-Alterseffekts wird sich daher auf die Daten mehrerer Tests stützen müssen.

Im zweiten Kapitel werde ich die Daten vorstellen, auf die sich die Forschungsdebatte wesentlich stützt. Dort werde ich auch das Ergebnis der Meta-Analyse von Wellman, Cross & Watson ausführen, dass ein entscheidender Faktor noch nicht gefunden wurde – und damit auch den *Early-Competence-Accounts* ein wichtiger Beleg fehlt. Daher bietet es sich an, im dritten Kapitel zunächst verschiedene *Conceptual-Change-Accounts* vorzustellen, darunter die beiden bereits erwähnten Positionen, dass Kinder im Alter von vier Jahren den Überzeugungs- bzw. den Repräsentationsbegriff erwerben. Zu den *Early-Competence-Accounts* werde ich erst im dritten Teil kommen, nachdem ich das Simulationsmodell eingeführt habe, auf dem mein eigener Vorschlag beruht.

Im ersten Kapitel werde ich zuvor das kognitive Theorienmodell genauer erläutern und den für meine Zwecke strittigen Punkt mit dem Simulationsmodell herausarbeiten.

1 Das kognitive Theorienmodell

Die beiden Modelle der Alltagspsychologie, die ich in dieser Arbeit untersuche, antworten auf die Frage, welche kognitiven *Prozesse* ablaufen, wenn Menschen Anderen Überzeugungen zuschreiben oder deren Handlungen vorhersagen. Das Theorienmodell behauptet, dass Menschen alltagspsychologische Theorien verwenden, um zu alltagspsychologischen Urteilen zu gelangen.¹² Es kommt in der Debatte zwischen Theorien- und Simulationsmodell nicht darauf an, *welche* Theorien verarbeitet werden (solange es psychologische sind), sondern vielmehr *dass* Theorien angewendet werden (und nicht etwa ein kognitiver Prozess der anderen Person simuliert wird). Ich werde daher im folgenden vom 'kognitiven Theorienmodell' sprechen, um zu verdeutlichen, dass es mir um die kognitiven Grundlagen der Alltagspsychologie geht.

Im Abschnitt 1.1 werde ich einen weiteren Typ von Theorienmodellen vorstellen, die alltagspsychologische Urteile mit Hilfe konkreter alltagspsychologischer Theorien ableiten wollen. Obwohl es mir letzten Endes nicht darum geht, solche Theorien zu spezifizieren, werden sie im Laufe dieser Arbeit noch wichtig werden. Einige Autoren haben in diesem Zusammenhang versucht, das Spektrum möglicher alltagspsychologischer Theorien zu begrenzen. Ihnen zufolge ist es unplausibel anzunehmen, dass Menschen unbewusst mit psychologischen Begriffen und Theorien operieren, die ihnen nicht bewusst zugänglich sind. Ich werde mich am Ende des ersten Abschnitts dieser Kritik anschließen. Sie bildet die Grundlage für die Interpretationen der empirischen Befunde, die ich im zweiten und dritten Kapitel vorstellen werde.

Im Abschnitt 1.2 werde ich mich mit den möglichen Indikatoren für das Theorienmodell befassen. Dabei werde ich gegen die Position von Martin Davies argumentieren, dass es prinzipiell problematisch sei, empirisch zwischen Theorien- und Simulationsmodell zu unterscheiden. Ich werde versuchen zu zeigen, dass dazu mehrere Strategien zur Verfügung stehen. Obwohl es ein Ergebnis dieser Arbeit sein wird, dass einige dieser Strategien bislang nicht zum Erfolg geführt haben, möchte ich nicht den Eindruck erwecken, es sei von vorneherein aussichtslos, sie zu verfolgen.

¹² Alltagspsychologische Urteile sind alle alltagspsychologischen Einzelleistungen: Erklärungen und Vorhersagen von Handlungen sowie Zuschreibungen geistiger Zustände.

1.1 Alltagspsychologische Theorien anwenden

Wie könnte der kognitive Mechanismus der Theorieanwendung beschaffen sein? Denkbar wäre ein Mechanismus für die Handhabung allgemeiner Sätze, der zur Lösung alltagspsychologischer Probleme auf alltagspsychologische Theorien zurückgreift, die im Langzeitgedächtnis gespeichert sind. Zur Handhabung allgemeiner Sätze zählen das Abstrahieren von Regularitäten aus wiederholten Beobachtungen, das Systematisieren der Beobachtungen durch die Postulierung nicht direkt beobachtbarer Entitäten und das Ableiten von Hypothesen aus Theorien. Es wird in diesem Zusammenhang oft das Schlagwort vom Kind als naivem Wissenschaftler verwendet, denn das Theorienmodell zieht Parallelen zwischen Alltagspraxis und den theoretischen Ableitungen in der Wissenschaft.

Ein einfaches Beispiel einer theoretischen Ableitung könnte die Anwendung der Theorie des praktischen Syllogismus sein:

- (A1) 1. Prämisse: Person *A* wünscht sich, dass *p*.
- (A2) 2. Prämisse: Person *A* glaubt, dass sie ihren Wunsch erfüllen kann, indem sie die Handlung *H* ausführt.
- (A3) Anwendung der Theorie: Personen, die glauben, dass sie über Ausführung einer bestimmten Handlung einen ihrer Wünsche erfüllen können, führen diese Handlung – *ceteris paribus* – aus.
- (A4) Vorhersage: Person *A* wird *H* ausführen.

Das Beispiel ist natürlich nicht vollständig. Die Vorhersage beruht darauf, dass die Prämissen eine Situation beschreiben, auf die sich eine einzelne Theorie anwenden lässt. Doch das tun sie nicht ohne weiteres. Zum einen sind die geistigen Zustände, auf die sich die Prämissen beziehen, nicht direkt beobachtbar und müssen aus beobachtbaren Stimuli und Reaktionen erschlossen werden. In der vollständigen alltagspsychologischen Ableitung müssten demnach auch die psychologischen Prämissen (A1) und (A2) aus nicht-psychologischen Beobachtungen abgeleitet werden. Zum anderen müssten die Bedingungen der *ceteris-paribus*-Klausel überprüft werden. Dabei könnte es notwendig werden, eine längere Kausalkette geistiger Zustände zu postulieren, die zu der Handlung *H* führt: einen Entscheidungsprozess.

Fodor (1987) hat diesen Punkt stark gemacht: In der Regel resultierten alltagspsychologische Urteile nicht direkt aus der Anwendung einer Theorie auf eine Situation.

[S]ubsumption under platitudes is *not* the typical form of commonsense psychological explanation. Rather, when such explanations are made explicit, they are frequently seen to exhibit the 'deductive

structure' that is so characteristic of explanation in real science. There are two parts to this: the theory's underlying generalizations are defined over unobservables, and they lead to its predictions by iterating and interacting rather than by being directly instantiated. (Fodor 1987, 7, Hervorhebung im Original)

Einigen Autoren erscheinen solche Ableitungen zu umfangreich, als dass man sie tatsächlich ausführen könnte. Vor allem die Überprüfung der *ceteris-paribus*-Klausel birgt die Gefahr, dass die Ableitung unabsehbar langwierig wird. Angesichts der Fülle möglicher kausaler Verknüpfungen zwischen geistigen Zuständen mag es so aussehen, als überfordere die Überprüfung aller relevanten Verknüpfungen die kognitiven Fähigkeiten eines Menschen. Goldman (1993) hebt zum Beispiel hervor, dass schon die Selbstzuschreibung einer Überzeugung nach dem Theorienmodell die Prüfung verlange, ob man auch den Implikationen der in Frage stehenden Überzeugung zustimmen würde. Es seien also auch Zusammenhänge mit *möglichen* geistigen Zuständen und möglichen Handlungen zu überprüfen.

A given belief may causally interact with quite a large number of other belief tokens and desire tokens. To type-identify that belief, it looks as if the subject must track its relations to each of these other internal states, their relations to further states, and so on until each path terminates in an input or output. When subjunctive properties are added to the picture the task becomes unbounded, because there is an infinity of possible beliefs and desires. (Goldman 1993, 19)

Bedenken dieser Art sind es, die die Formulierung alternativer Modelle motivieren, die ohne Rückgriff auf alltagspsychologische Theorien auskommen. Das Simulationsmodell ist ein solches Modell; ich werde diese Bedenken daher erst im zweiten Teil dieser Arbeit weiter ausführen. Ich kann aber schon jetzt andeuten, dass ich für eine *Arbeitsteilung* zwischen Theorien- und Simulationsmodell plädieren werde: Simulationen eignen sich vor allem, um den alltagspsychologischen Umgang mit Schlussfolgerungen zu erklären. Denn es sind Überzeugungen, deren vielfältige Beziehungen untereinander die Anwendung von Theorien erschweren. Im Hinblick auf andere geistige Zustände wie Wahrnehmungen oder Wünsche halte ich den Einwand Goldmans für weniger schlagkräftig; in diesen Fällen sind handhabbare alltagspsychologische Theorien denkbar.

Alltagspsychologische Platitüden

Stich & Ravenscroft (1994) nennen Erklärungen des o.g. Typs 'interne Theorienmodelle' und grenzen sie von *externen* Modellen ab, die das explizierbare alltagspsychologische Wissen in Form einer Theorie systematisieren.

External accounts either collect or systematize the intuitively recognizable generalizations of common-sense psychology, while internal accounts focus on the cognitive mechanism that underlies our ability to have those intuitions, to predict behavior, etc. (Stich & Ravenscroft 1994, 463)

Ich möchte nicht für oder gegen externe Theorienmodelle argumentieren. Sie sind meines Erachtens mit dem Simulationsmodell, das ich im zweiten Teil dieser Arbeit vorstellen werde, kompatibel. Die Streitfrage zwischen Theorien- und Simulationsmodell betrifft nur die Frage, ob Menschen alltagspsychologische Theorien anwenden. Ich möchte dennoch kurz auf externe Theorienmodelle eingehen, da sie in den folgenden Kapiteln noch eine Rolle spielen werden.

Als Kronzeugen für externe Theorienmodelle führen Stich & Ravenscroft Lewis (1972) an, einem Vertreter des *Commonsense*-Funktionalismus'. Dieser Position zufolge bezeichnen psychologische Begriffe Typen geistige Zustände, die durch ihre funktionale Rolle individuiert werden. Lewis hat eine Methode entwickelt, mit der sich die theoretische *Definition* der Begriffe ermitteln lässt; sie werden durch psychologische *Platitüden* definiert, die allgemein geteilt werden. Diese Platitüden beschreiben kausale und begriffliche Relationen zwischen geistigen Zuständen, ihren Ursachen und Wirkungen. Sie umfassen keine Volksweisheiten der Art 'Gleich und gleich gesellt sich gern' oder 'Gebranntes Kind scheut das Feuer', denn von diesen Weisheiten wissen wir, dass sie falsch sind. Sie stellen vielmehr den alltagspsychologischen Wissensschatz dar, den Menschen im Alltag explizieren können.

Es ist zu vermuten, dass es viele solcher Platitüden gibt, und dass es viel Mühe bedeuten würde, sie zu explizieren. Des Weiteren ist zu vermuten, dass alle geistigen Zustände in verschiedenen Platitüden auftauchen und auf diese Weise miteinander verknüpft werden. Überzeugungen können zum Beispiel durch eine Fülle disparater Reize hervorgerufen werden, und führen auch nur im Zusammenhang mit vielen weiteren Überzeugungen zu bestimmten Handlungen. Es ergibt sich demnach ein dichtes Netz von Platitüden, das Lewis als Theorie auffasst.

Ich möchte alltagspsychologische Theorien jedoch nicht auf Platitüden einschränken. Vielmehr halte ich es für sinnvoll, alle alltagspsychologischen Urteile in die Systematisierung einzubeziehen. Externe Theorienmodelle behaupten demnach, dass sich auch alltagspsychologische Vorhersagen und Erklärungen als Ableitungen mit Hilfe alltagspsychologischer Theorien rekonstruieren lassen. Im Unterschied zu internen (in meiner Terminologie: kognitiven) Theorienmodellen machen sie jedoch keine Aussage

darüber, ob Menschen diese Theorien in ihrer alltagspsychologischen Praxis anwenden. Sie sind nicht, wie Stich & Ravenscoft es ausdrücken, Teil des zugrunde liegenden kognitiven Mechanismus'.

Diese Ausweitung des Datenbestands für die Systematisierungen externer Theorienmodelle hat Lewis in einem späteren Aufsatz selbst gefordert:

We have a very extensive shared understanding of how we work mentally. Think of it as a theory: folk psychology. It is common knowledge among us; but it is tacit, as our grammatical knowledge is. We can tell which particular predictions and explanations conform to its principles, but we cannot expound those principles systematically. (*Pace* Lewis, 1972, p. 256, eliciting the general principles of folk psychology is no mere matter of gathering platitudes.) (Lewis 1994, 416)

Lewis scheint es für möglich zu halten, dass die explizierbaren Platitüden nicht ausreichen, um das volle funktionale Profil geistiger Zustände zu definieren. Mit dem Appell an die Linguistik versucht er, auch solche Elemente der alltagspsychologischen Theorie einzubeziehen, die uns im Alltag nicht bewusst sind. Doch mit der Analogie zur impliziten Grammatik geht er – wie ich nun ausführen möchte – zu weit.

Die Analogie zur Linguistik

Die alltagspsychologische Praxis wird auch von anderen Vertretern des Theorienmodells mit der Sprachproduktion verglichen: Menschen würden in demselben Sinn alltagspsychologische Theorien besitzen wie sie über eine implizite Grammatik ihrer Muttersprache verfügen. Denn das Geschäft der Linguistik könne als Versuch aufgefasst werden, Menschen eine grammatische "Theorie" zuzuschreiben, aus der sich ihre sprachlichen Äußerungen ergeben. Grundlage für diese Behauptung ist der (angestrebte) Befund, dass mit Hilfe der grammatischen Theorie genau die Sätze produziert werden können, die von Muttersprachlern als grammatisch korrekt eingestuft werden. Analog versuchen Vertreter externer Theorienmodelle, mit Hilfe alltagspsychologischer Theorien Urteile zu erzeugen, die Alltagspsychologen für plausibel halten.

Einige Autoren haben die Analogie zwischen alltagspsychologischer und grammatischer Theorie jedoch kritisiert. Ihnen zufolge gibt es einen wichtigen Unterschied: Die alltagspsychologischen Theorien sind dadurch eingeschränkt, dass sie mit den Begriffen formuliert sein müssen, die Menschen im Alltag gebrauchen und verstehen. Linguisten haben hingegen freie Hand bei der Wahl der Begriffe, mit der sie die implizite Grammatik formulieren. Ich möchte drei Argumente dieser Kritiker anführen.

Heal (1995) zufolge ist nicht ohne weiteres einzusehen, warum es einen Bruch zwischen dem expliziten und dem impliziten Teil der alltagspsychologischen Theorie geben sollte. Vielmehr wäre ein solcher Bruch erklärungsbedürftig.

This supposed theory [of the psychological by which we do our predicting of each other] is taken to be constituted by two components: first, the explicit remarks we can muster, and second, (probably) a further unconscious component – a hidden part of the cognitive iceberg. And it would be odd if these two parts differed radically in their format or in the kinds of characterizations they offered. (Heal 1995, 38)

Nach Stone & Davies (1996) hinkt die Analogie zur Linguistik, weil die alltagspsychologische Praxis selbst mit alltagspsychologischen Begriffen operiert, während in der Sprachpraxis kaum linguistisches Vokabular auftaucht. Den Autoren zufolge müssen Theorienmodelle auch den Gebrauch alltagspsychologischer Begriffe (etwa bei der Explizierung alltagspsychologischer Platitüden) erklären. Im Unterschied dazu muss eine implizite Grammatik lediglich das Zustandekommen grammatisch korrekter Sätze erklären.

In the case of linguistics there is something to be said for the idea that the content of the tacit knowledge does not have to be conceptualised by the subject whose tacit knowledge it is. Ordinary language users do not grasp the concepts of linguistic theory, and so their tacit knowledge of linguistic principles does not constitute their mastery of those concepts. But in the case of folk psychology, ordinary practitioners do possess the concepts that will, presumably, be central in the tacitly known theory. (Stone & Davies 1996, 124)

Schließlich argumentiert Botterill (1996) gegen die Analogie, indem er an das Gefühl der Sicherheit appelliert, das Menschen bei ihren alltagspsychologischen Urteilen haben. Selbst wenn wir nur einige Bruchstücke einer alltagspsychologischen Erklärung vorbrächten, komme uns die Erklärung in der Regel explanatorisch vor. Dieses Gefühl lasse sich nur dadurch erklären, dass die Zusammenhänge dieser Bruchstücke im Prinzip expliziert werden können, und daher auch mit den bekannten Begriffen formuliert sein müssen. In der Sprachproduktion gibt es hingegen keine solche Erklärungen.

We [...] can explain why people did what they did in terms of what they hoped to achieve and what they thought they might get by so acting. But why should citing the psychological antecedents of action in this way appear to us to be explanatory, unless we have some sort of awareness of the principles involved? (Botterill 1996, 114)

Diese Argumente sind meines Erachtens nicht von der Hand zu weisen. Externe Theorienmodelle müssen sich auf das bekannte alltagspsychologische Vokabular beschränken, wenn sie alltagspsychologische Urteile systematisieren. Und auch Vertreter interner Theorienmodelle sollten von der Möglichkeit Abstand nehmen, dass Menschen impli-

zite alltagspsychologische Theorien anwenden, die mit unbekanntem Begriffen formuliert sind. Im Grunde orientiert sich die *Theory-of-Mind*-Forschung an dieser Beschränkung. Doch einige der postulierten alltagspsychologischen Theorien, die ich im dritten Kapitel diskutieren werde, enthalten neuartige (und zum Teil auch schwer verständliche) psychologische Begriffe.

1.2 Empirische Prüfung des Theorienmodells

Bevor ich erläutere, wie im Rahmen der Kognitionswissenschaft ein Mechanismus der Anwendung alltagspsychologischer Theorien nachgewiesen werden kann, möchte ich darauf eingehen, wie kognitive Erklärungen grundsätzlich funktionieren. Das kognitive Theorienmodell und das Simulationsmodell postulieren – wie andere kognitionswissenschaftliche Theorien auch – so genannte Wissensstrukturen. Stich & Nichols (1992) erläutern das Prinzip solcher Erklärungen wie folgt:

[T]he dominant explanatory strategy [in contemporary cognitive science] proceeds by positing an internally represented 'knowledge structure' – typically a body of rules or principles or propositions – which serves to guide the execution of the capacity to be explained. These rules or principles or propositions are often described as the agent's 'theory' of the domain in question. (Stich & Nichols 1992, 35)

Statt von 'Wissensstrukturen' wird oft auch von 'tacit knowledge' gesprochen – ein Begriff, der auf Fodor (1968) zurückgeht. Fodor argumentiert in diesem Aufsatz wie folgt: Wenn es gelingt, mit einem Computer menschliches Verhalten (also z.B. die alltagspsychologische Praxis) zu simulieren, darf man das Programm des Computers den Menschen als *tacit knowledge* zuschreiben. Diese Zuschreibung rechtfertigt er durch ein Schlussprinzip von ähnlichen *Wirkungen* (Verhalten des Computers und der Menschen) auf ähnliche *Ursachen* (Programm und *tacit knowledge*). Mit anderen Worten: Da das Computerprogramm die Kausalkette beschreibt, die zum Verhalten des Computers führt, und Menschen dasselbe Verhalten zeigen, beschreibt das Programm auch die Kausalkette, die zum menschlichen Verhalten führt.

If machines and organisms can produce behaviors of the same type and if descriptions of machine computations in terms of the rules, instructions, etc., that they employ are true descriptions of the etiology of their output, then the principles that licenses inferences from like effects to like causes must license us to infer that the tacit knowledge of organisms is represented by the programs of the machines that simulate their behavior. (Fodor 1968, 640)

Solche Beschreibungen von Kausalketten, die im simulierenden Computer zum gewünschten Verhalten führen, hält Fodor für sinnvolle psychologische Erklärungen. Es bleibe zwar fraglich, wann eine psychologische Erklärung *vollständig* sei, doch diese Frage hänge daran, welche kognitiven Prozesse man für fundamental, d.h. nicht weiter analysierbar halte.

A completed psychological theory must provide systems of instructions to account for the forms of behavior available to an organism, and it must do so in a way that makes reference to no unanalyzed psychological processes. (Fodor 1968, 629)

Wenn also Theorien- und Simulationsmodelle konkurrierende Erklärungen derselben Fähigkeit sein sollen, müssen sie zwei verschiedene Wissensstrukturen angeben, aus denen sich das beobachtete alltagspsychologische Verhalten ableiten lässt. Doch schon der Begriff 'Theorie' im obigen Zitat von Stich & Nichols deutet an, dass die Unterscheidung dieser Wissensstrukturen schwierig sein könnte. Auf den ersten Blick hat es den Anschein, als lasse sich *jede* Wissensstruktur als Theorie beschreiben.

Diesen Punkt möchte ich an folgendem Beispiel diskutieren: Dem Simulationsmodell zufolge liegt der alltagspsychologischen Praxis die Verwendung der eigenen kognitiven Mechanismen (etwa des Entscheidungsmechanismus') zum Zweck der Simulation zugrunde. Wenn aber etwa ein Entscheidungsprozess nach dem Simulationsmodell durch die Anwendung des eigenen Entscheidungsmechanismus' simuliert wird, und das Funktionieren des eigenen Entscheidungsmechanismus' – der kognitionswissenschaftlichen Erklärungsstrategie gemäß – durch eine Entscheidungstheorie erklärt wird, dann würde auch nach dem Simulationsmodell bei der Vorhersage oder Erklärung einer Entscheidung eine Entscheidungstheorie angewendet.

Da sie diesen Schluss für absurd halten, fassen Stich & Nichols das Simulationsmodell auch als Kritik an der vorherrschenden Erklärungsstrategie der Kognitionswissenschaft auf:

If these [advocates of simulation theory] are right, [...] the dominant explanatory strategy in cognitive science, the strategy that appeals to internally represented knowledge structures, will be shown to be mistaken in at least one crucial corner of our mental lives. (Stich & Nichols 1992, 36f.)

Doch ich sehe das Simulationsmodell durchaus im Rahmen der Kognitionswissenschaft. Ich möchte daher entgegnen: Sowohl Theorien- als auch Simulationsmodell postulieren jeweils eine *zweite* Wissensstruktur neben der genannten Entscheidungstheorie. Das kognitive Theorienmodell postuliert eine alltagspsychologische Theorie darüber, wie

der Entscheidungsmechanismus funktioniert. Es gibt also zwei Theorien: eine implizite, nach der man entscheidet, und eine potenziell explizite, die beschreibt, wie man sich typische Entscheidungen vorstellt. Das Simulationsmodell postuliert hingegen eine Wissensstruktur, die das Verwenden kognitiver Mechanismen zum Zweck der Simulation beschreibt. Diese Replik möchte ich im folgenden ausführen. Doch bevor ich dazu komme, muss ich das skeptische Argument deutlicher herausarbeiten. Dazu orientiere ich mich an Davies (1994). Am Ende dieses Abschnitts werde ich skizzieren, welche Möglichkeiten es gibt, zwischen den beiden "zweiten" Wissensstrukturen empirisch zu unterscheiden.

Theorien- vs. Simulationsmodell: eine Scheindebatte?

Davies (1994) spricht von einem drohenden Zusammenfallen des Theorien- und des Simulationsmodells. Ausgangspunkt seiner Argumentation ist Goldmans (1989) frühes Simulationsmodell, das "Prozess-Ähnlichkeit" zwischen Simulationsprozess und simuliertem Originalprozess verlangt: Simulationen sind nur dann erfolgreich, wenn der simulierte Prozess dem simulierten Originalprozess hinreichend ähnlich ist, d.h. wenn er – in Davies' Worten – den "Konturen" des Originalprozesses folgt.

An dieser Stelle wendet Davies ein, dass die Ähnlichkeitsbedingung erlaube, Simulationsprozesse durch eine Wissensstruktur à la Theorienmodell zu beschreiben. Denn eine Ableitung mit geeigneten Theorien könnte (auf einem Computer implementiert) dem Originalprozess hinreichend ähnlich sein. Der Grad der Ähnlichkeit orientiert sich schließlich an den beobachtbaren Reiz-Reaktions-Mustern, und eine *gute* Theorie zeichnet sich gerade dadurch aus, Reize und Reaktionen richtig zu verknüpfen. Eine Simulation ist daher von der Anwendung einer guten Theorie empirisch nicht zu unterscheiden.

The crucial point to notice is that the processing in the simulator could *also* follow the contours of the derivational structure of a proof of a conclusion about, say, an intention to act from premises about, say, the agent's beliefs and desires – a proof cast in a psychological theory. [...] The threat of collapse ensues because it is just this idea of matching of structure between a causal process and a derivational process that is used in some accounts of what it is for a cognitive processing system to embody tacit knowledge of a particular theory of the task domain. Roughly speaking, a component processing mechanism embodies tacit knowledge of a particular rule or axiom if it plays a role in mediating causally between representational states that is structurally analogous to the role that the rule or axiom itself plays in mediating derivationally between premises and conclusions [...]. (Davies 1994, 115, Hervorhebung im Original)

Es macht also Davies zufolge im Ergebnis (d.h. in den alltagspsychologischen Urteilen) keinen Unterschied, ob eine gute alltagspsychologische Theorie angewendet wird oder ein kognitiver Mechanismus simuliert wird, der durch ebendiese Theorie beschrieben wird. Nach Fodors Methodologie scheint es daher auf den ersten Blick nicht möglich zu sein, Erklärungsmodelle zu formulieren, die die Anwendung von Theorien ausschließen.

Kognitive Architektur

Gegen diese pessimistische Argumentationslinie zur Unterscheidbarkeit der Modelltypen möchte ich mich nun wenden.¹³ Meines Erachtens macht es für die Erklärung der alltagspsychologischen Praxis durchaus einen Unterschied, ob kognitive Mechanismen simuliert oder Theorien angewendet werden. Das kognitive (interne) Theorienmodell behauptet nämlich, dass Menschen alltagspsychologische Theorien so lernen und anwenden wie sie es bei anderen Theorien auch tun. Eine "Theorie", die einen Entscheidungsmechanismus beschreibt, ist aber von einer anderen Sorte. Sie kann zum Beispiel mit Begriffen formuliert werden, die Menschen im Alltag nicht verwenden (ähnlich den grammatischen "Theorien" in der Linguistik). Vor allem aber ist es keine Theorie, die Eingang in die kognitiven Mechanismen findet, mit denen Menschen "theoretisieren" – ebenso wenig wie die impliziten physikalischen Theorien, mit deren Hilfe Menschen ihre Bewegungen koordinieren, ihr Verständnis physikalischer Zusammenhänge beeinflussen.

Simulationsmodelle formulieren eine Alternative zu den Theorienmodellen, da sie einen Mechanismus beschreiben, der untergeordnete kognitive Mechanismen zum Zweck der Simulation ansteuert. Es ist sogar der Fall denkbar, in dem eine Person ein alltagspsychologisches Urteil einer anderen Person vorhersagt, indem sie ihren eigenen Alltagspsychologie-Mechanismus als Modell verwendet. Es ist in diesem Fall nicht ausgeschlossen, dass ein Mechanismus der Theorieanwendung simuliert wird.

Es sind demnach Aspekte der *kognitiven Architektur*, die den Gehalt des kognitiven Theorienmodells und des Simulationsmodells ausmachen. Wie lässt sich nachweisen, welche kognitive Architektur der alltagspsychologischen Praxis zugrunde liegt? Man

¹³ Auf Davies' eigene Lösung dieses Problems werde ich erst im Abschnitt 7.1 eingehen.

könnte beispielsweise prüfen, ob Menschen mit Defiziten im theoretischen Denken auch Schwierigkeiten mit der Alltagspsychologie haben. Das würde die These stützen, dass sie auch in diesem Bereich (mehr schlecht als recht) Theorien anwenden. Ebenso wäre es denkbar zu prüfen, ob Menschen mit Defiziten in der Kontrolle ihrer kognitiven Mechanismen Schwierigkeiten mit der Alltagspsychologie haben. Das würde die These stützen, dass ihr übergeordneter Simulationsmechanismus seine Arbeit nur unzureichend verrichtet.

Doch mit solchen Vorschlägen entferne ich mich weit von der vorliegenden Forschungsliteratur. Ich wähle daher einen anderen Zugang. Mein Ansatzpunkt ist der, dass man der Alltagspsychologie nicht zu viel zutrauen sollte. Die pessimistische Haltung von Davies, Stich und Nichols fußt meines Erachtens auf der falschen Prämisse, dass die Theorieanwendung bzw. die Simulation perfekt funktionieren und das Verhalten Anderer zuverlässig und korrekt vorhersagt oder erklärt. In diesem Fall wären die beiden Modelle tatsächlich empirisch kaum zu unterscheiden.

Doch die Alltagspsychologie ist nicht perfekt. Stich & Nichols (1992) bieten sogar selbst einen empirischen Indikator an, der darauf beruht, dass Theorieanwendung und Simulation unterschiedliche Erfolgsbedingungen haben: Die Theorieanwendung ist erfolgreich, wenn die Theorien wahr sind; die Simulation hängt hingegen von der Ähnlichkeit zwischen Simulations- und Originalprozess ab. Es lassen sich Situationen konstruieren, in denen nur eine der beiden Bedingungen erfüllt ist, so dass sich aus der Trefferquote alltagspsychologischer Urteile auf den zugrunde liegenden Mechanismus schließen lässt. Diesen und andere Indikatoren werde ich im Abschnitt 6.1 diskutieren.

In den folgenden beiden Kapiteln werde ich mich jedoch erst einmal mit den alltagspsychologischen Leistungen von drei- und vierjährigen Kindern befassen. Ihre Kompetenz ist noch weit davon entfernt, perfekt zu sein, so dass es im Prinzip möglich ist, an ihrer kognitiven Entwicklung das eine oder andere über die kognitive Architektur abzulesen.

2 Überblick über die empirische Forschung

Die empirische Forschung, die ich nun vorstellen werde, ist der Analyse des kindlichen Verständnisses des Geistes gewidmet. Im Mittelpunkt steht dabei der *False Belief Test* (oder: *False Belief Task*, kurz: FBT), der die Fähigkeit prüft, anderen Menschen falsche Überzeugungen zuzuschreiben. Da dieser prominente Test vor allem bei drei- und vierjährigen Kindern interessante Ergebnisse liefert, werde ich mich auf diese Altersstufe konzentrieren.

Um eine kleine Übersicht zu geben: Ich werde zunächst den FBT und einige seiner vereinfachten Varianten vorstellen (Abschnitt 2.1). Diese Tests messen das Aufkommen einzelner alltagspsychologischer Fähigkeiten. Der üblicherweise gemessene Alterseffekt liegt bei vier Jahren; jüngere Kinder fallen bei den Tests noch durch. Auch die vereinfachten Varianten reproduzieren im Wesentlichen den Alterseffekt.

Im Abschnitt 2.2 werde ich die relevanten Leistungen *dreijähriger* Kinder vorstellen, die wichtig sind, da sie als Kandidaten zur Erklärung des Alterseffekts ausscheiden. Das besondere Interesse wird der Fähigkeit gelten, aus dem fehlenden epistemischen Zugang einer Person zu etwas auf ihr Nicht-Wissen darüber zu schließen. Im Abschnitt 2.3 werde ich dann noch einige verwandte kognitive Fähigkeiten vorstellen, die in der Forschungsliteratur oft herangezogen werden, um ein Entwicklungsmuster zu beschreiben, das über das Verständnis falscher Überzeugungen hinausgeht.

2.1 Der Alterseffekt im *False Belief Test*

Das prominenteste experimentelle Design der *Theory-of-Mind*-Forschung ist der *False Belief Test*, den Wimmer & Perner (1983) bekannt machten.¹⁴ Dieser Test stellt fest, ob Kinder in der Lage sind, anderen Menschen andere Überzeugungen zuzuschreiben als ihre eigenen – ob sie Anderen also (ihrer Meinung nach) *falsche* Überzeugungen zu-

¹⁴ Das experimentelle Design wurde in der veröffentlichten Diskussion von Premack & Woodruff (1978) vorgeschlagen. In dieser Diskussion ging es um die Frage, wie man prüfen könne, ob Schimpansen über ein alltagspsychologisches Verständnis verfügen. Johnson & Maratsos (1977) hatten jedoch bereits mit einem ähnlichen Design den Alterseffekt gemessen.

schreiben können. In der inzwischen als "Standard" bekannten Version des FBT wird Kindern eine Szene mit Puppen vorgespielt: Ein Protagonist namens Maxi legt eine Tafel Schokolade in den Küchenschrank und geht hinaus zum Spielen. In seiner Abwesenheit nimmt die Mutter die Schokolade aus dem Schrank und legt sie in eine Schublade.¹⁵ Nach einer Weile erscheint Maxi wieder in der Küche. Getestet wird, ob die Kinder die folgende Frage richtig beantworten: "Maxi kann sich noch daran erinnern, wohin er die Schokolade getan hat. Wo wird er jetzt nach der Schokolade schauen?"

Das Ergebnis: Im Alter von etwa vier Jahren setzt die Fähigkeit, die Frage richtig zu beantworten und damit den FBT zu bestehen, plötzlich ein. Jüngere Kinder geben in der Regel die falsche Antwort, obwohl sie Kontrollfragen nach dem gegenwärtigen und früheren Ort der Schokolade richtig beantworten. Sie können sich also an alle relevanten Details des Szenarios erinnern. Dass die Leistung dreijähriger Kinder deutlich unter der Trefferquote des Ratens von 50% liegt, weist darauf hin, dass auch sie die Aufgabe systematisch angehen; sie sind in diesem Test nicht einfach verwirrt oder ratlos.

Um einen Eindruck von den typischen Größen des Alterseffekts zu geben, ziehe ich eine neuere Meta-Analyse von FBT-Studien heran, die Wellman, Cross & Watson durchgeführt haben. (Ich werde jedoch im folgenden auf detaillierte Altersangaben verzichten und nur von den Leistungen Drei- und Vierjähriger sprechen.)

At younger ages – essentially 41 months (3 years, 5 months) and younger – children performed below chance, making the classic false-belief error. At older ages – essentially 48 months (4 years) and older – they performed above chance, significantly correct. (Wellman, Cross & Watson 2001, 663)

Es zeigte sich schnell, dass der FBT spezifisch den Umgang mit *falschen* Überzeugungen misst – und nicht etwa den Umgang mit Überzeugungen allgemein. Wellman & Bartsch (1988) stellten dreijährigen Kindern Fragen der folgenden Art:

Jane sucht ihre Katze. Die Katze könnte sich in der Garage oder in der Küche versteckt haben. Jane glaubt, dass sich ihre Katze in der Küche versteckt hat. Wo wird Jane nach der Katze suchen?

Die guten Leistungen Dreijähriger in diesem Test zeigen, dass ihnen die kausale Verknüpfung zwischen Handlungen und Überzeugungen vertraut ist. Die Schwierigkeiten tauchen erst auf, wenn die Testfrage wie folgt abgeändert wird:

Jane sucht ihre Katze. Die Katze ist in der Garage. Jane glaubt aber, dass die Katze in der Küche ist. Wo wird Jane nach ihrer Katze suchen?

¹⁵ Dieses experimentelle Design wird daher auch 'unexpected change task' oder 'location task' genannt.

In diesem Fall antworten dreijährige Kinder, dass Jane ihre Katze in der Garage suchen werde.

Mit ihrem Janes-Katzen-Ansatz konnten Wellman & Bartsch auch eine weitere nahe liegende Interpretation der Leistung Dreijähriger im FBT ausschließen: Die Kinder wenden keine egozentrische Strategie der Überzeugungszuschreibung an, d.h. sie schreiben Anderen nicht immer ihre eigene Überzeugung zu. Um das zu zeigen, baten die Autoren die Kinder zunächst um ihre eigene Meinung dazu, wo Janes Katze zu finden sei. Sie schrieben Jane daraufhin eine andere Meinung zu und baten die Kinder um die Vorhersage der Handlung. Auch dreijährige Kinder konnten richtig vorhersagen, wo Jane nach ihrer Katze suchen werde, obwohl sie selbst woanders gesucht hätten.

Die Robustheit des Alterseffekts

In der ersten Studie von Wimmer & Perner wurde nicht gefragt, wo Maxi die Schokolade *vermute*; die Kinder sollten nicht nur eine falsche Überzeugung zuschreiben, sondern aufgrund der falschen Überzeugung eine passende *Handlung* vorhersagen. Doch das spielt keine Rolle. Im Lauf der Jahre wurde der FBT in vielen Variationen durchgeführt, deren Befunde sich nicht wesentlich vom Standardtest unterscheiden. Ein bedeutendes Ergebnis der Meta-Analyse von Wellman, Cross & Watson (2001) ist, dass eine ganze Reihe dieser Variationen keinen Einfluss auf den Alterseffekt haben. Es macht zum Beispiel keinen Unterschied, ob Kinder angeben müssen, was Maxi *denkt*, was er *weiß*, was er *sagen* würde oder wo er *suchen* wird. Für jeden dieser Fragetypen ergibt sich derselbe Alterseffekt.

Der Test kann auch so abgewandelt werden, dass Kinder angeben müssen, was eine andere Person über den Inhalt einer Schachtel glaube, wenn dieser Person kein Einblick in die Schachtel gewährt wurde. Kindern wird etwa gezeigt, dass eine Schachtel Smarties wider Erwarten keine Smarties, sondern einen Bleistift enthält.¹⁶ Dreijährige Kinder geben durchweg an, dass auch andere Personen, die nicht in die Schachtel schauen konnten, wissen, dass sie einen Bleistift enthält.

¹⁶ Dieses experimentelle Design wird daher auch 'unexpected content task' genannt.

Die Autoren der Meta-Analyse fügen der Liste irrelevanter Faktoren noch weitere hinzu: So macht es zum Beispiel ebenso wenig aus, ob Maxi von einer Puppe oder einem Schauspieler repräsentiert wird, und ob es um eine Tafel Schokolade geht oder nur um das Bild einer solchen Tafel.

In addition, the "medium" in which the false-belief task is presented has no significant effect. That is, it makes no difference if the protagonist is presented as a real person, a puppet, a doll, a pictured storybook character, or a videotaped person. Similarly, as long as a concrete target object is present at the time the false-belief question is asked, it makes no difference whether the object is a real item (a piece of edible chocolate), a toy (a small toy car), a picture of an object (a drawing of a piece of chocolate), and so forth. (Wellman, Cross & Watson 2001, 664)

Viele dieser Variationen waren untersucht worden, um das FBT-Szenario jüngeren Kindern in einer Form nahe zu bringen, die ihren Alltagserfahrungen besser entspricht. Das Ausbleiben des Effekts weist darauf hin, dass die Schwierigkeiten dreijähriger Kinder nicht dadurch zu erklären sind, dass das FBT-Szenario zu alltagsfern ist.

Selbstzuschreibungen falscher Überzeugungen

Auch bei Selbstzuschreibungen ergibt sich ein dem FBT vergleichbarer Alterseffekt. Im o.g. Smarties-Test werden Kinder nach dem Blick in die Schachtel gefragt, was sie selbst über deren Inhalt glaubten, *bevor* sie hineingesehen hatten. Gopnik & Astington (1988) stellten fest, dass dreijährige Kinder angeben, ihre Meinung habe sich nicht geändert. Sie hätten auch vorher geglaubt, die Schachtel enthalte einen Bleistift. Jüngere Kinder sind demnach nicht in der Lage, ihre eigenen früheren falschen Überzeugungen anzuerkennen.

Wimmer & Hartl (1991, 2. Experiment) konnten diesen Alterseffekt angeblich zum Verschwinden bringen: Wenn dreijährigen Kindern eine Dose Smarties gezeigt wird, die tatsächlich Smarties enthält, und der Inhalt der Dose erst später ausgetauscht wird, können diese Kinder ihre frühere Überzeugung über den Inhalt der Dose richtig angeben. Im Gegensatz zu Gopnik & Astington (1988) wurden die Kinder in diesem Test jedoch nicht nach einer *falschen* Überzeugung gefragt. Mitchell (1996, 127f.) zieht daher aus den guten Leistungen der Dreijährigen in Wimmer & Hartls Test den Schluss, dass sie nur mit falschen Überzeugungen Schwierigkeiten haben.

Trotz des parallelen Alterseffekts ist meines Erachtens nicht klar, ob Selbst- und Fremdzuschreibung im selben Kontext diskutiert werden dürfen. Es ist bekannt, dass

auch erwachsene Menschen nur einen schlechten Einblick in ihre geistigen Prozesse haben. In vielen psychologischen Experimenten konnten Einstellungsänderungen induziert werden, die von den Versuchspersonen abgestritten wurden.¹⁷ Ich halte es daher nicht für ausgeschlossen, dass Kinder einen noch schlechteren Einblick in ihre geistigen Vorgänge haben als erwachsene Versuchspersonen. Selbst wenn sie über die begrifflichen Voraussetzungen verfügten, die Möglichkeit eigener falscher Überzeugungen zuzulassen, wäre es denkbar, dass sie von dieser Überzeugung nichts wissen, weil sie – trotz ihrer Überraschung über den Bleistift – introspektiv keine Überzeugungsänderung feststellen. Ich möchte daher nicht davon ausgehen, dass Gopnik & Astingtons Test dieselbe kognitive Fähigkeit misst wie der FBT, und ihren Befund daher aus der weiteren Diskussion ausklammern.

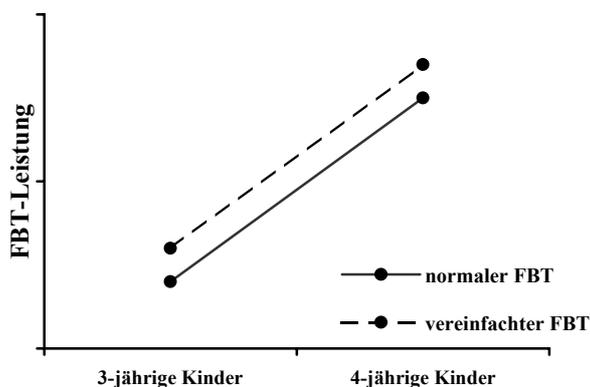
Vereinfachte Varianten des False Belief Tests

Wie bereits in der Vorbemerkung zum ersten Teil erwähnt, versuchen Vertreter von *Early-Competence-Accounts* die alltagspsychologische Kompetenz jüngerer Kinder zum Bestehen des FBT dadurch nachzuweisen, dass diese leichtere Varianten des FBT bestehen. Einige dieser erfolglosen Varianten habe ich schon genannt. Eine weitere ist ein Test, in dem der Protagonist eine falsche Überzeugung besitzt, weil er *getäuscht* worden ist.

In der Studie von Chandler, Fritz & Hala (1989) erhielten die Kinder die Gelegenheit, einen Spielpartner darüber zu täuschen, in welchem von vier Kästchen ein Schatz versteckt ist. Das Verstecken übernahm eine Puppe, die beim Gang zu den Kästchen Fußspuren aus Tinte hinterließ. Die Kinder waren in der Lage, zu Täuschungszwecken die Fußspuren zum richtigen Kästchen abzuwischen oder gar falsche Fußspuren zu legen, bevor ihr Spielpartner in den Raum geführt wurde, um den Schatz zu suchen. Die Autoren schließen daraus, dass bereits dreijährige Kinder über die Kompetenz verfügen, den FBT zu bestehen, weil ihre Fähigkeit zur Täuschung zeigt, dass sie auf falsche Überzeugungen Anderer hinwirken können. Irrelevante Aspekte der Aufgabenstellung würden jedoch die erste erfolgreiche Ausübung dieser Kompetenz im Standard-FBT um ein Jahr verzögern.

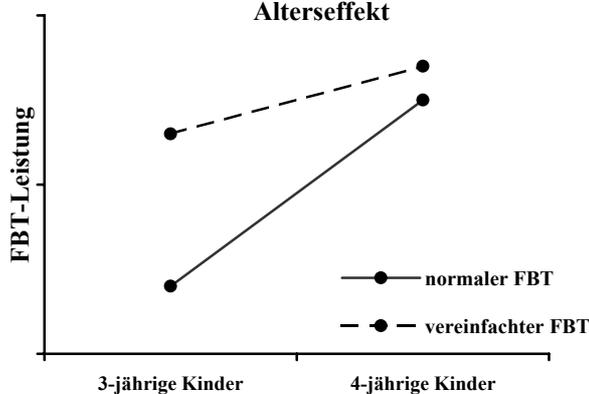
¹⁷ Die bekannteste und inzwischen klassische Studie dürfte Nisbett & Wilson (1977) sein.

Abb. 1: Alterseffekt wird vorgezogen



Teilnahme des Kinds an der Geschichte. Diese Faktoren beeinflussen den Alterseffekt, indem sie ihn ein paar Monate vorziehen. Doch sie *interagieren* nicht mit ihm, da sie die Steigung der Geraden nicht verändern. Abb. 1 veranschaulicht eine solche *Erleichterung* des FBT.

Abb. 2: Vereinfachung interagiert mit Alterseffekt



Wellman, Cross & Watson (2001) erkennen diesen Befund an, werten ihn jedoch anders. Es gebe eine Reihe von Merkmalen des experimentellen Designs, die die durchschnittliche Quote richtiger Antworten leicht erhöhen. Dazu gehören die Einbettung der Aufgabe in ein Täuschungsszenario und die aktive

Abb. 2 zeigt hingegen einen Effekt, der mit dem Alterseffekt interagiert. Die im Schaubild dargestellte Interaktion ist eine, die von besonderer theoretischer Relevanz ist: Ein Faktor, der spezifisch die Leistungen dreijähriger Kinder erhöht – also der Kinder, die den FBT in seiner normalen Form noch nicht bestehen –

könnte den Alterseffekt erklären. Denn er würde auf die *entscheidende* Schwierigkeit hinweisen, die dreijährige Kinder davon abhält, den normalen FBT zu bestehen. Doch ein solcher Faktor ist Wellman, Cross & Watson zufolge noch nicht gefunden worden. Eine weitere Möglichkeit, den FBT zu vereinfachen, besteht darin, Kinder auf die relevanten Überzeugungen aufmerksam zu machen. Mitchell & Lacohee (1991) ließen Kinder zum Beispiel eine explizite Repräsentation (ein Bild) ihrer ersten Behauptung über den Inhalt der Smarties-Dose auswählen. Wenn Kinder später mit ihrer Wahl des Bildes

konfrontiert wurden, gaben sie ihre frühere Überzeugung richtig an.¹⁸ Die Meta-Analyse zeigt jedoch, dass diese Aspekte der Aufgabenstellung sowohl die Leistungen jüngerer als auch die Leistungen älterer Kinder fördern. Es sieht daher zwar so aus, als handle es sich um Aspekte, die den FBT zusätzlich erschweren. Doch es kann nicht gefolgert werden, dass diese Aspekte zum Bestehen des FBT entscheidend sind und damit den Alterseffekt erklären. Denn sie steigern nicht spezifisch die Leistungen jüngerer Kinder.

Our findings show that several task manipulations do increase young children's performance: framing the task in terms of deception or trickery, involving the child in actively making the key transformation, and highlighting the salience of the protagonist's mental state or reducing the salience of the contrasting real-world state of affairs, all help young children to perform better. Early competence accounts require more than improved performance, however. [...] [S]uch accounts require that relevant task manipulations differentially enhance young children's performance. (Wellman, Cross & Watson 2001, 672)

In der Meta-Analyse interagierte ein einziger Faktor mit dem Alterseffekt: der Hinweis auf die zeitliche Dimension des Problems durch die Frage "Wenn Maxi zurückkommt, wo wird er *zuerst* nach der Schokolade suchen?" Die Interaktion war jedoch gegenläufig: Der Faktor erleichterte den FBT *nur für ältere* Kinder – also solche, die ihn in der Regel schon bestehen. Für diesen Effekt gibt es bislang keine Erklärung; ich werde ihn daher als Besonderheit stehen lassen.

Im nächsten Abschnitt werde ich noch weitere mögliche Belege für ein frühes Verständnis falscher Überzeugungen bei Dreijährigen vorstellen. Doch diese Belege stammen von Tests, die nicht mit denselben Testfragen arbeiten wie der FBT, und sind daher für *Early-Competence-Accounts* weniger geeignet. Vielmehr sind sie nötig, um *Conceptual-Change-Accounts* zu etablieren. Denn wer die These vertritt, dass erst Vierjährige über einen bestimmten theoretischen Begriff verfügen, muss zeigen können, dass dieser Begriff zur Erklärung der Leistungen jüngerer Kinder entbehrlich ist.

2.2 Die Theorien zwei- und dreijähriger Kinder

Auch bevor sie den FBT bestehen, können Kinder bereits unter Rückgriff auf geistige Zustände Handlungen erklären und vorhersagen. Doch man sagt, dass ihr Repertoire an

¹⁸ Einen entsprechenden Test für Fremdzuschreibungen, in dem Kinder eine explizite Repräsentation der Überzeugung des Protagonisten wählen, gibt es meines Wissens nicht.

alltagspsychologischen Begriffen noch eingeschränkt oder ihr Verständnis dieser Zustände nicht ausgereift sei. Die Untersuchung der Leistungen zwei- und dreijähriger Kinder ist für die Interpretation des FBT-Alterseffekts wichtig, weil die Fähigkeiten jüngerer Kinder als Kandidaten zur Erklärung des Alterseffekts ausgeschlossen werden können. Denn sie reichen offenbar nicht aus, um den Test zu bestehen, und sind damit nicht der entscheidende Faktor, der den Erwerb der FBT-Löse-Kompetenz erklärt.

Ich werde zunächst einige Fähigkeiten jüngerer Kinder aus dem "Umfeld" des FBT vorstellen und dann auf die meines Erachtens wichtige Frage zu sprechen kommen, ob dreijährige Kinder bereits über eine Theorie des epistemischen Zugangs verfügen. Diese Theorie würde es ihnen ermöglichen zu begreifen, dass der Protagonist nicht wissen kann, dass die Schokolade verlegt worden ist, weil er nicht Zeuge dieses Ereignisses war. Diese Frage wird bei der Interpretation der empirischen Befunde im dritten Kapitel eine bedeutende Rolle spielen.

Die Wahrnehmungs-Wunsch-Theorie zweijähriger Kinder

Das alltagspsychologische Verständnis von *zweijährigen* Kindern wird von Gopnik & Wellman (1992, 1994) beschrieben als Theorie von Wünschen (*desires*) und Wahrnehmungen (*perceptions*), die zusammenwirken, um Handlungen hervorzurufen. Zum Beispiel sagt die Theorie voraus, dass jemand, der ein Objekt haben will und es sieht, danach greifen wird.¹⁹ Doch die geistigen Zustände, die diese frühe alltagspsychologische Theorie postuliert, sind in einer noch zu klärenden Weise eng an Stimuli und Reaktionen geknüpft:

Desires at first are conceived simply as something like drives towards objects (Wellman & Woolley, 1990). Perceptions are at first understood simply as awareness of or visual contact with objects (Flavell, 1988). In neither case need the child conceive of a complex propositional or representational relationship between these mental states and the world. Instead, these very young children seem to treat desire and perception as simple causal links between the mind and the world. Given that an agent desires an object, the agent will act to obtain it. Given that an object is within a viewer's line of sight, the viewer will see it. (Gopnik & Wellman 1994, 265)

¹⁹ Meltzoff (1995) konnte beispielsweise zeigen, dass schon 18 Monate alte Kinder erfolglose Handlungen imitieren können – spontan und in der richtigen Ausführung. Das zeigt, dass auch sehr junge Kinder die Absichten Anderer verstehen.

In dem zitierten Experiment von Wellman & Wooley (1990) sollten Kinder das Suchverhalten von anderen Kindern vorhersagen. Manipuliert wurde, was die anderen Kinder bereits gefunden hatten: den gewünschten Gegenstand, einen anderen attraktiven Gegenstand oder nichts. Geprüft wurde, ob die Kinder begreifen, dass das andere Kind mit dem Suchen nicht aufhört, solange es nicht das gewünschte Objekt gefunden hat. Kinder meistern diesen Test schon vor ihrem dritten Geburtstag.

In dem Aufsatz von Flavell (1988) werden Studien zitiert, denen zufolge Kinder bereits mit zwei Jahren einiges über den Sehprozess verstehen. So sind sie bereits zu den folgenden Handlungen in der Lage:

They can produce or engender the seeing of x [i.e. the visual target] by O [i.e. another person] by pointing to or verbally designating x , by getting O to open his or her eyes and face x , by moving or reorienting x so that it is in O 's line of sight, and by repositioning either a [i.e. any large object interposed between O and x such as to block O 's vision of x] or x so that a no longer blocks O 's seeing of x . (Flavell 1988, 248)

Ebenso sind sie in der Lage, analoge Handlungen auszuführen, um O 's Sehen von x zu verhindern, und sie können feststellen, ob ein bestimmter Versuchsaufbau O in die Lage versetzt, x zu sehen. Dieses Verständnis visueller Wahrnehmung ist sicher eine Voraussetzung dafür, den FBT zu bestehen, da es unter anderem darauf ankommt, dass der Protagonist nicht sieht, dass das Objekt an einem anderen Ort versteckt wird.

Doch es hat den Anschein, als seien die geistigen Zustände, die Zweijährige anderen Menschen zuschreiben, an die Umstände des Augenblicks gebunden. Flavell (1988) spricht deshalb von "kognitiven Verbindungen" und meint damit dasselbe, was Gopnik & Wellman mit "einfachen kausalen Verknüpfungen" bezeichnen:

Young children [by 2 to 3 years of age] know that they can become cognitively connected to something by seeing it; they also understand that they may not see it – that they may not be connected to it in this way at a given moment. (Flavell 1988, 244f.)

Zweijährige Kinder verstehen demnach, dass die Welt auf den Geist einwirkt (und so Wahrnehmungen erzeugt) und dass der Geist auf die Welt einwirkt (indem er Handlungen hervorbringt). Doch sie begreifen in diesem Alter noch nicht, dass Menschen *verschiedene Perspektiven* auf ein Ding haben können – und auf dieser Grundlage handeln. Von einigen Autoren wird vermutet, dass diese fortgeschrittene Fähigkeit mit dem Überzeugungs- oder Repräsentationsbegriff zusammenhängt. Im dritten Kapitel werde ich meine Kritik an dieser Position ausführen. Meines Erachtens benötigt man kein Ver-

ständnis der Perspektivität der Wahrnehmung, um den FBT zu bestehen. Die in diesem Zusammenhang einschlägigen Daten werde ich im nächsten Abschnitt vorstellen.

Ein frühes Verständnis falscher Überzeugungen?

Ich möchte nun auf drei Tests zu sprechen kommen, die ein *frühes* Verständnis falscher Überzeugungen anzeigen (d.h. auf ein Verständnis vor dem vierten Geburtstag). Diese Tests messen das Verständnis jedoch auf eine andere Weise als der FBT. Da die Tests auf keine interessante Weise miteinander zusammenhängen, werde ich sie einfach der Reihe nach aufzählen.

1. Bartsch & Wellman (1989) konnten zeigen, dass dreijährige Kinder weniger Schwierigkeiten damit haben, Handlungen aufgrund falscher Überzeugungen zu *erklären*, als sie auf dieser Grundlage *vorherzusagen*. Wenn ihnen eine besonders gekennzeichnete, aber leere Pflasterdose und eine neutrale Dose mit Pflastern gezeigt wird, können sie richtig erklären, warum jemand, der sich verletzt hat, die Pflasterdose aufmacht: "Er dachte, da sind Pflaster drin." Diese Erklärung liefern sie, obwohl sie wissen, dass nur die unmarkierte Dose Pflaster enthält. Die meisten dieser Kinder fallen jedoch beim FBT noch durch.

Dieser Befund legt die Interpretation nahe, dass Kinder bereits mit drei Jahren im Prinzip eine Vorstellung von der Existenz und dem kausalen Profil falscher Überzeugungen haben – doch nur, wenn man sie mit der Nase darauf stößt. Die Erklärung der Autoren lautete auch zunächst: Dreijährige Kinder verfügen bereits über den Überzeugungsbe-griff, können dieses Verständnis aber nicht immer richtig anwenden, da ihnen das Zusammenspiel der verschiedenen geistigen Zustände noch nicht ganz klar ist. Zum Beispiel überschätzen sie noch das Streben nach Wunscherfüllung: Menschen werden zunächst in diejenige Dose schauen, die das enthält, was sie haben wollen. Wird diese Handlungsalternative durch die vollzogene Handlung ausgeschlossen, bleibt den Kindern nichts anderes übrig, als ihr Verständnis von Überzeugungen anzuwenden.

[F]or 3-year-olds at any rate, the false belief prediction tasks present a conflict between reasoning about what would satisfy the actor's desire [...] and reasoning in terms of the actor's beliefs. Faced with this conflict, 3-year-olds predict according to what would satisfy the actor's desires. [...] Why this mode of reasoning wins out, we believe, concerns the young child's strong allegiance to the maxim that in general people act so as to satisfy their desires. (Bartsch & Wellman 1989, 963)

Wellman (1990) hat diese Interpretation allerdings widerrufen. Als Grund gibt er an, dass dreijährige Kinder auch dann Schwierigkeiten mit falschen Überzeugungen haben, wenn keine Wünsche im Spiel sind – zum Beispiel wenn sie lediglich ihre eigenen früheren Überzeugungen angeben sollen (siehe das Smarties-Experiment von Gopnik & Astington 1988 aus dem letzten Abschnitt). Damit werde der Vorschlag von Bartsch & Wellman entkräftet, das Streben nach Wunscherfüllung würde das Verständnis von Überzeugungen "aushebeln".

Moses & Flavell (1990) entdeckten zudem, dass das vermeintliche frühe Verständnis von Überzeugungen nicht so robust ist wie es sein sollte. Obwohl dreijährige Kinder eine Handlung unter Rückgriff auf falsche Überzeugungen erklären können, scheitern sie an Folgefragen. Auch nachdem sie die Handlung einer Person durch eine falsche Überzeugung erklärt haben, geben sie den Gehalt der Überzeugung dieser Person falsch an. Dieses Ergebnis nährt den Verdacht, dass diese Kinder noch kein ernsthaftes Verständnis von Überzeugungen erworben haben. Ich halte den Befund von Bartsch & Wellman daher nicht für belastbar und werde keine Schlüsse daraus ziehen.

2. In einer neueren Studie konnten Happé & Loth (2002) zeigen, dass dreijährige Kinder in der Lage sind, die falschen Überzeugungen einer anderen Person zu berücksichtigen, wenn sie den Bezug bei Namensgebungen rekonstruieren. In einem FBT-ähnlichen Szenario gibt der Protagonist einem in einer Kiste verborgenen Objekt einen Namen. Wenn das Objekt vor der Namensgebung und in Abwesenheit des Protagonisten ausgetauscht worden ist, berücksichtigen bereits dreijährige Kinder, dass der Protagonist nicht das Objekt in der Kiste benannt hat, sondern das Objekt, das vorher in der Kiste war. Sie sind anschließend in der Lage, das neu benannte Objekt aus einer Menge Objekte auszuwählen. Dieser Befund passt zu keinem der Theorienmodelle, die ich im nächsten Kapitel diskutieren werde; ich lasse ihn daher vorerst unerklärt stehen. Ich werde das Ergebnis dieser Studie aber im weiteren Verlauf der Arbeit als Motiv dafür nehmen, nach einem frühen Verständnis falscher Überzeugungen zu suchen.

3. Das gleiche gilt für den folgenden Befund. Clements & Perner (1994) änderten den Versuchsablauf im FBT wie folgt: Bevor die Versuchsleiter den Kindern die entscheidende Frage nach dem Suchverhalten des Protagonisten stellten, fragten sie rhetorisch: "I wonder where he's going to look?" Eine Auswertung von Videofilmen des Experiments ergab, dass Kinder ab dem Alter von 2 Jahren und 11 Monaten auf diese Frage hin spontan auf das alte, inzwischen leere Behältnis schauen. In einer Kontroll-

untersuchung, in der der Protagonist vor seiner Rückkehr über den neuen Ort des Objekts informiert wurde, zeigten sie dieses Verhalten nicht.

Perner (1995, 2000) interpretiert diesen Befund als Nachweis eines aufkommenden, aber noch "impliziten" Verständnisses von Überzeugungen:

I would concede that this demonstrates earlier understanding of false belief than hitherto thought. However, as a tentative interpretation, this understanding seems to be in a different "implicit" form, inaccessible at the level of verbal communication. (Perner 1995, 262)

Perner weist vor allem darauf hin, dass es auch in der Studie von Clements & Perner einen deutlichen Alterseffekt gibt: Weniger als 20% der Kinder aus der Gruppe der 2;5- bis 2;10-jährigen sahen im kritischen Moment in die richtige Richtung, während dies fast 80% der 2;11- bis 3;2-jährigen taten. Der normale FBT misst demnach den Zeitpunkt in der kognitiven Entwicklung, in dem das Verständnis von Überzeugungen "explizit" wird, d.h. – nach Perners Erläuterung – in kommunikatives Verhalten wie etwa das Beantworten von Testfragen einfließen kann. Das Verständnis werde hingegen einige Monate vorher erworben. Dennoch bleibe der Alterseffekt bestehen.

Garnham & Perner (2001) konnten den o.g. Effekt replizieren und zudem die alternative Erklärung ausschließen, dass den dreijährigen Kindern noch das Gefühl der Sicherheit fehle, um ihrer spontanen Reaktion Glauben zu schenken. Dazu änderten sie den Versuchsaufbau so, dass der Protagonist – je nach seiner Überzeugung über den Standort des Objekts – auf einer von zwei Rutschen aus dem Hintergrund ins einsehbare Szenario zurückrutscht. Die Kinder wurden gebeten, schnell eine Matte unter eine der Rutschen zu legen, um den Protagonisten aufzufangen. Da das Platzieren einer Matte das Kind (im Gegensatz zum bloßen Hinschauen) klar auf eine Vorhersage festlegt, kann dieses Verhalten als explizite Antwort gewertet werden.

Auch bei dieser Aufgabe schnitten dreijährige Kinder gut ab. Ihre Leistungen wurden nur dann geschmälert, wenn sie vorher ausdrücklich gefragt wurden, wo sie die Matte platzieren würden. Dieser Befund deutet darauf hin, dass die *verbale Explikation* der Antwort eine Schwierigkeit für Dreijährige darstellt – diese Erklärung werde ich im Verlauf der Arbeit noch genauer ausführen. Ich glaube, dass sie zur richtigen Lösung führt.

Das Verständnis des epistemischen Zugangs

Oben habe ich erwähnt, dass dreijährige Kinder bereits verstehen, dass der Protagonist im FBT nicht sieht, dass das Objekt verlegt wird. An dieser Stelle möchte ich auf die Frage eingehen, ob sie auch die Verknüpfung zwischen (visuellen) Wahrnehmungen und Überzeugungen verstehen, d.h. ob sie bereits über eine entsprechende Theorie der *Überzeugungsausbildung* verfügen. Mit Hilfe einer solchen Theorie könnten sie zumindest schließen, dass der Protagonist *nicht weiß*, wo die Schokolade nun versteckt ist. Würde sich zeigen, dass dreijährige Kinder bereits über eine solche Theorie verfügen, wäre der Besitz dieser Theorie nicht der entscheidende Faktor zum Bestehen des FBT. In der Literatur wird die zur Überzeugungsausbildung nötige Beziehung oft als 'informationeller' oder 'epistemischer Zugang' (*epistemic access*, kurz: EZ) bezeichnet.²⁰ Die Theorie lässt sich so formulieren:

(EZ) Sehen (und vielleicht auch: gesagt bekommen o.ä.), dass *p*, führt zur Ausbildung der Überzeugung, dass *p*.

Im folgenden möchte ich eine Reihe empirischer Befunde durchgehen, die meines Erachtens nahe legen, dass dreijährige Kinder bereits über (EZ) verfügen. Doch die Datlage ist nicht klar genug, um dieser These großes Gewicht zu verleihen. Zunächst stelle ich einige Befunde vor, die kein einheitliches Bild ergeben:

- Hogrefe, Wimmer & Perner (1986) sowie Perner & Wimmer (1988) konnten zwar zeigen, dass dreijährige Kinder entsprechende Fragen zu ihrem Verständnis von (EZ) überraschend richtig beantworten können. Doch dieser Effekt ging nicht über die 50%-Marke hinaus, d.h. er lässt die Interpretation zu, dass die Kinder *nur zufällig* richtig lagen.²¹

²⁰ Siehe zum Beispiel Wimmer, Hogrefe & Sodian (1988). Die Autoren versuchen allerdings, den FBT durch den Erwerb von (EZ) zu erklären. Siehe dazu den Abschnitt 3.2. – Man mag gegen (EZ) einwenden, dass die Theorie in einer Weise formuliert ist, die Dreijährige nicht verstehen können. Doch es ist möglich, sie umgangssprachlich zu formulieren – etwa in der Art: 'Man weiß nur, was man gesehen hat.' In der Forschungsliteratur ist die Theorie auch als 'Seeing-is-believing-rule' bekannt. Das deckt sich zwar nicht mit der kausalen Verknüpfung zwischen Wahrnehmung und Überzeugung, die ich für (EZ) gewählt habe. Doch im Hinblick auf die Ambiguitäten bei der Interpretation der experimentellen Befunde lässt sich (EZ) meines Erachtens nicht genauer festlegen.

²¹ Auf diesen so genannten 'baseline effect' machte erst Perner (1995, 253) aufmerksam.

- Roth & Leslie (1998, 1. Experiment) erhielten dagegen ein deutliches Ergebnis: Die meisten der dreijährigen Kinder gaben an, dass ein Freund, der draußen wartete, nicht wisse, dass der Inhalt einer Schachtel (etwa eines bekannten Spiels) ausgetauscht worden ist.
- Gopnik & Graf (1988) wiederum zeigten, dass dreijährige Kinder Schwierigkeiten damit haben, die Quellen ihrer aktuellen Überzeugungen anzugeben. Ihnen wurde gesagt oder gezeigt, was sich in verschiedenen Schubladen befand. Die Kinder sollten daraufhin wiederholen, was sich in den Schubladen befand, und angeben, auf welche Weise sie davon erfahren hatten. Obwohl die Leistungen dreijähriger Kinder über der Ratewahrscheinlichkeit lagen, stellten die Autoren einen signifikanten Alterseffekt fest.
- Povinelli & deBlois (1992) ließen Kinder raten, unter welchem Hütchen eine Überraschung versteckt sei. In der entscheidenden Phase des Experiments deuteten sie den Kindern das richtige Hütchen an. Obwohl alle Kinder unter dieses Hütchen schauten, konnten nur die Vierjährigen begründen, warum sie dies getan hatten. Dreijährige Kinder konnten zudem nicht richtig zwischen zwei Personen diskriminieren, die beide auf ein Hütchen zeigten – von denen aber nur eine gesehen hatte, wo die Überraschung versteckt worden war.
- Pillow (1989) konnte schließlich zeigen, dass dreijährige Kinder richtig angeben können, dass eine Person, die in eine Kiste geschaut hat, ihre Inhalt kennt, nicht aber eine Person, die lediglich auf der Kiste steht.

Angesichts der uneinheitlichen Befunde ist denkbar, dass bislang unbekannte Testanforderungen für die unterschiedlichen Leistungen verantwortlich sind.²² Eine Studie spricht jedoch deutlich für ein frühes Verfügen über (EZ) – wenn es auch auf die non-verbale Ebene beschränkt zu sein scheint. O'Neill (1996) beobachtete spielende Kinder, deren Lieblingsspielzeug unerreichbar auf das oberste Regal gelegt wurde. Es war jeweils die Mutter oder der Vater des Kindes anwesend, so dass es für die Kinder nahe lag, ihre Eltern zu bitten, das Spielzeug wieder herunterzuholen. Es zeigt sich, dass bereits dreijährige Kinder bei ihrer Bitte den epistemischen Status ihrer Eltern berücksichtigten:

²² Ich habe zwar den Eindruck, dass die negativen Resultate von Gopnik und Povinelli auf komplizierte Fragestellungen zurückzuführen sind, die dreijährige Kinder überfordern. Schließlich ist es eine Sache, (EZ) anzuwenden und das Ergebnis als Antwort zu nennen, und eine andere, diesen Prozess zu rekapitulieren. Doch auf diese Diskussion möchte ich mich nicht einlassen.

Waren die Eltern nicht im Raum, als das Spielzeug ins Regal gestellt wurde, zeigten die Kinder häufiger auf das gewünschte Spielzeug und nannten es auch häufiger beim Namen.

Darüber hinaus scheint (EZ) für den FBT nicht entscheidend zu sein. Um das zu zeigen, weist man dreijährige Kinder ausdrücklich darauf hin, dass der Protagonist nicht weiß, dass die Schokolade verlegt worden ist. Zum Beispiel kann der Protagonist auf dem Spielplatz seine Schwester treffen und ihr erzählen, wo sich die Schokolade seiner Meinung nach befindet. Damit schafft man eine Variante des FBT, die die Anwendung von (EZ) nicht mehr erfordert. Sollten dreijährige Kinder trotzdem Schwierigkeiten mit einer solchen Variante des FBT haben, kann ihre Schwierigkeit nicht in der Anwendung von (EZ) liegen. Perner & Wimmer (1988) konnten zeigen, dass dreijährige Kinder auch diesen Test nicht bestehen, so dass das Verständnis des epistemischen Zugangs für den FBT nicht entscheidend sein kann. Ein ähnliches Ergebnis erhielten Wellman & Bartsch (1988) mit einer ihrer Fragen nach Janes Katze: Wenn den Kindern gesagt wird, die Katze befinde sich in der Garage, Jane glaube aber, sie befinde sich Küche, sind die Kinder nicht in der Lage vorherzusagen, dass Jane in der Küche nach der Katze suchen werde (siehe Abschnitt 2.1).

Ein wichtiger Befund spricht jedoch *gegen* ein frühes Verständnis von (EZ): die durchweg falschen Antworten Dreijähriger im FBT. Wenn sie über (EZ) verfügten, müssten sie in der Lage sein zu bestimmen, dass der Protagonist nicht weiß, wo sich die Schokolade jetzt befindet. In diesem Fall sollten sie um eine Antwort auf die Frage, wo er nach der Schokolade suchen werde, verlegen sein und möglicherweise raten, d.h. eine durchschnittliche Leistung um 50% liefern. Doch die FBT-Studien zeigen, dass die meisten der dreijährigen Kinder antworten, dass der Protagonist die Schokolade an ihrem neuen Ort suchen werde. Dieser Befund ist auf den ersten Blick mit dem Verständnis von (EZ) unvereinbar. Im dritten Teil der Arbeit werde ich verschiedene Ansätze diskutieren, um dieses Problem zu lösen.

2.3 Überzeugung, Wahrnehmung und Repräsentation

In der Interpretation der empirischen Befunde, die ich im folgenden Kapitel vorstellen werde, spielen auch weitere kognitive Fähigkeiten eine Rolle. Dazu gehört das Ver-

ständnis der visuellen Wahrnehmung und die Unterscheidung zwischen Sein und Schein. Die in diesem Zusammenhang einschlägigen Befunde möchte ich nun darstellen. Außerdem werde ich auf die Frage eingehen, ob dreijährige Kinder dieselben Schwierigkeiten mit veralteten Fotos haben wie mit veralteten Überzeugungen. Diese Befunde werden in der Forschungsliteratur dazu verwendet, um ein Muster der kognitiven Entwicklung des Kindes herauszuarbeiten.

Perspektivität und Erscheinung

Im letzten Abschnitt habe ich bereits das Verständnis zweijähriger Kinder über die Bedingungen der visuellen Wahrnehmung erwähnt. Grob gesprochen verstehen sie, dass ein visueller Kontakt bestehen muss, damit etwas wahrgenommen werden kann. Dieses Verständnis wird später um das Verständnis der *Perspektivität* visueller Wahrnehmung erweitert: Im Alter von vier Jahren beginnen Kinder zu verstehen, dass Menschen Dinge aus verschiedenen räumlichen Perspektiven betrachten und demzufolge verschiedene visuelle Wahrnehmungen des Dings haben können (Flavell 1988). Zum Beispiel geben erst vierjährige Kinder richtig an, dass eine andere Person, die ihnen gegenüber sitzt, die *Rückseite* eines Objekts sieht, das zwischen ihnen platziert ist.²³

Da der Entwicklungsschritt parallel zum Alterseffekt im FBT zu verlaufen scheint, werden Zusammenhänge zwischen den beiden Fähigkeiten vermutet. Ein Verständnis der Perspektivität wird zwar für den FBT nicht benötigt, da der Protagonist nur hinschauen müsste, um zu sehen, dass die Schokolade verlegt wird. Doch es ist möglich, dass das Verständnis der Perspektivität dieselben begrifflichen Ressourcen benötigt wie das Verständnis falscher Überzeugungen. Man spricht bei gleichzeitig aufkommenden Fähigkeiten auch von einer "Korrelation" der Effekte.

Ein Zusammenhang wird auch mit einer weiteren Fähigkeit vermutet: der Unterscheidung zwischen Schein und Wirklichkeit (*appearance-reality distinction*). Einen klassischer Test dieser Unterscheidungsfähigkeit haben Flavell et al. (1983) entwickelt. Kindern wird zum Beispiel ein Schwamm gezeigt, der aussieht wie ein Stein. Obwohl die Kinder ihn zunächst für einen Stein halten, können sie, nachdem sie ihn angefasst ha-

²³ Dieses erweiterte Verständnis visueller Wahrnehmung wird in der Literatur als 'Level 2' bezeichnet und vom oben erwähnten *Level-1*-Verständnis unterschieden.

ben, nicht mehr sagen, dass er *nur aussehe* wie ein Stein. Es sei ein Schwamm, der auch aussehe wie ein Schwamm (oder umgekehrt: es sei ein Stein, der auch so aussehe wie ein Stein). Ebenso sind sie nicht in der Lage, korrekt anzugeben, den Schwamm vorher für einen Stein gehalten zu haben.

In der Forschungsliteratur werden diese Tests als nicht-psychologische *Analoga* zum FBT aufgefasst. In einem späteren Überblicksartikel fasst Flavell zusammen:

[T]he distinction between perceptual appearance and reality is conceptually very similar both to the distinction between false belief and reality and to the Level 2 distinction between two different perceptual appearances resulting from different observer perspectives. Consistent with this fact, there is some correlational evidence that these distinctions tend to develop together; that is, young children who perform well (or poorly) on appearance-reality tasks also tend to perform well (or poorly) on false-belief tasks, Level 2 visual perspective-taking tasks [...]. (Flavell 1999, 36)

Doch in einer neueren Studie konnten Sapp et al. (2000, 1. Experiment) die Unterscheidungsfähigkeit schon bei jüngeren Kindern nachweisen. Die Autoren bateten Kinder, ein Objekt aus mehreren auszusuchen, wobei sie variierten, ob sie nach einem bestimmten Objekt X fragten oder (zum Zweck eines Fotos) nach einem Objekt, das nur so aussieht wie X. Fast alle dreijährigen Kinder konnten in beiden Fällen das richtige Objekt auswählen. Es ist daher zumindest zweifelhaft, ob ihre Schwierigkeiten in Flavells ursprünglichem Test auf ein fehlendes Verständnis der Relation *So-aussehen-wie* zurückzuführen sind. Da ich noch weitere Studien zu diesem Punkt der korrelierten Fähigkeiten diskutieren werde, stelle ich eine abschließende Bewertung vorerst zurück.

Kippbilder und veraltete Fotos

Eine Möglichkeit, den Zusammenhang zwischen den drei Testtypen zu formulieren, ist diese: Die Tests messen das aufkommende Verständnis des *interpretativen Charakters* der Überzeugungsbildung. Im Alter von vier Jahren beginnen Kinder zu verstehen, dass Menschen die Welt nicht nur so repräsentieren wie sie ist, sondern dass verschiedene Faktoren – der epistemische Zugang, die räumliche Perspektive oder ein irreführendes Äußeres – ihre Überzeugungsbildung beeinflussen. Doch nach Carpendale & Chandler (1996) sollte die Beschreibung 'interpretativ' für Prozesse reserviert bleiben, in denen interne Faktoren (wie das Hintergrundwissen) die Überzeugungsbildung beeinflussen.

Den Autoren zufolge zeigen Kinder zum Beispiel erst im Grundschulalter erstmals ein Verständnis von der *Interpretierbarkeit* (mehrdeutiger) visueller Reize. Die Autoren legten Kindern zweideutige Bilder vor (wie den bekannten Entenhasen) und vergewisserten sich, dass die Kinder die Bilder auf beide Weisen sehen können. Daraufhin baten sie sie vorherzusagen, auf welche Weise eine andere Person das Bild sehen würde. Einige der vier- und fünfjährigen Kinder geben auf diese Frage hin nur eine *bestimmte* Interpretation des Bilds an. Andere wollen sich zwar nicht festlegen, können dafür aber keine angemessene Begründung liefern. Erst im Schulalter (ab sechs bis sieben Jahre) sind sie dazu in der Lage, beide Interpretationen zu nennen oder die Antwort mit einer angemessenen Begründung zu verweigern. Die Rede vom aufkommenden Verständnis des interpretativen Charakters der Überzeugungsbildung ist daher unpassend, da sie einen weiteren Entwicklungsschritt verdeckt – nämlich den Schritt hin zum Verständnis des *konstruktiven* Charakters der Überzeugungsbildung.

Eine andere Möglichkeit, den Zusammenhang zwischen FBT und den zwei anderen vorgestellten Tests zu formulieren, ist diese: Die Tests messen das Aufkommen einer *repräsentationalen* Theorie des Geistes. Dieser Erklärung zufolge beginnen Kinder im Alter von vier Jahren zu verstehen, dass Menschen die Welt nicht notwendigerweise so repräsentieren wie sie ist. In diesem Zusammenhang ist der folgende Befund einschlägig.

Zaitchik (1991) testete Kinder mit diesem experimentellen Design: Die Kinder sehen, wie ein Polaroid-Foto von einem Gegenstand (einer Person auf einer Matte) aufgenommen wird. Während sich das Foto entwickelt, ändert sich die Situation (eine andere Person nimmt auf der Matte Platz). Erst vierjährige Kinder können richtig vorhersagen, welche Person auf dem Foto zu sehen sein wird. Dieser Test wird von vielen ebenfalls als nicht-psychologisches Analogon zum FBT aufgefasst: Statt um eine veraltete Überzeugung geht es hier um eine veraltete fotografische Repräsentation. Die Korrelation der beiden Alterseffekte stützt die Vermutung, dass Kinder im Alter von vier Jahren ein Verständnis von Repräsentationen psychologischer wie nicht-psychologischer Natur erwerben.

Zwei Studien legen jedoch nahe, dass das Verständnis von Fotos vom Verständnis von Überzeugungen unabhängig ist. Perner et al. (Ms.) entdeckten, dass der Foto-Test und der FBT Kindern im Durchschnitt zwar ähnlich große Schwierigkeiten bereitet, dass jedoch viele Kinder nur einen der beiden Tests bestehen. Die Autoren fanden keine si-

gnifikante Korrelation zwischen Zaitchiks Test und dem FBT. Das Resultat einer Studie von Leslie & Thaiss (1992) mit autistischen Kindern deutet in dieselbe Richtung: Autistische Kinder, die in der Regel besondere Schwierigkeiten mit dem FBT haben, können mit Fotos (und Karten) sogar besser umgehen als normale vierjährige Kinder.

3 Der Erwerb des Überzeugungsbegriffs

Wenn mehrere FBT-Varianten vergleichbare Ergebnisse (d.h. einen vergleichbaren Alterseffekt) liefern, verringert sich die Wahrscheinlichkeit, dass es sich um Effekte der Aufgabenstellung handelt. Wellman, Cross & Watson argumentieren daher, dass mit dem FBT ein Fortschritt im Verständnis von Überzeugungen gemessen werde:

That children's false-belief judgments are systematically unrelated to such task variations increases the likelihood that their judgments reflect robust, deep-seated conceptions of human actions, rather than task-specific responses provoked by the special features of one set of materials or questions. Specifically, the irrelevance of these task and procedural variations increases the likelihood that children's performance is systematically dependent on the one thing that does not vary, namely their conception of belief states. (Wellman, Cross & Watson 2001, 671)

Eine prominente Erklärung dieser Art – diejenige Pernalers – werde ich exemplarisch im Abschnitt 3.1 ausführen. Ich möchte aber noch zwei weitere Erklärungen vom Typ *Conceptual-Change-Account* diskutieren, also Positionen, denen zufolge der FBT den Erwerb eines theoretischen Begriffs misst. Zum einen werde ich im zweiten Abschnitt die Erklärung Wimmers vorstellen, derzufolge das Verfügen über den Überzeugungsbegriff an das Verfügen einer Theorie der Überzeugungsbildung geknüpft ist. Vierjährige Kinder bestehen den FBT seiner Ansicht nach, erst nachdem sie eine solche Theorie erlernt haben. Zum anderen werde ich im dritten Abschnitt die Position diskutieren, derzufolge das Muster der kognitiven Entwicklung auf den schrittweisen Erwerb einer repräsentationalen Theorie des Geistes hinweist.

Ich werde alle diese Positionen ablehnen; meines Erachtens gibt es keine guten Gründe für die Behauptung, dass der FBT den Erwerb eines theoretischen Begriffs misst. Dieses Ergebnis bedeutet, dass dem kognitiven Theorienmodell im Hinblick auf den Bereich der falschen Überzeugungen ein wichtiger Beleg abhanden kommt.

3.1 Die Überzeugungsrelation verstehen

Was weiß ein Kind, wenn es – Pernaler zufolge – über den Überzeugungsbegriff verfügt? Wesentlich für den Überzeugungsbegriff ist die Überzeugungsrelation zwischen Personen und Propositionen. In der psychologischen Forschungsliteratur wird häufig davon

gesprochen, dass Menschen Propositionen *semantisch bewerten*.²⁴ Einem Kind wird das Verständnis dieser Relation zugeschrieben, wenn es versteht, dass Menschen auch von falschen Propositionen überzeugt sein können. Allgemeiner gesprochen kann einem Kind das Verständnis von Überzeugungen als propositionalen Einstellungen zugeschrieben werden, wenn es versteht, dass Menschen in verschiedenen Relationen zu einer Proposition stehen können. Grundlage für Perner's Theorienmodell ist daher die Feststellung:

Ein Kind verfügt über den Überzeugungsbegriff genau dann, wenn es versteht, dass zwei Menschen inkompatible Überzeugungen haben können.

Woran erkennt man dieses Verständnis? In der Forschungsliteratur wird diese Verknüpfung mit einem empirischen Indikator verwendet:

Ein Kind versteht, dass zwei Menschen inkompatible Überzeugungen haben können, wenn es in der Lage ist, anderen Menschen (aus seiner Sicht) falsche Überzeugungen zuzuschreiben.

Dieser Indikator eignet sich besser als der, zwei Menschen inkompatible Überzeugungen zuzuschreiben, weil nicht jede inkompatible Zuschreibung ein tieferes Verständnis offenbart. Ein Kind kann zum Beispiel zwei anderen Menschen inkompatible Überzeugungen zuschreiben, ohne den Widerspruch der beiden Einstellungen zu bemerken. Im Fall des Widerspruchs mit einer *eigenen* Überzeugung sollte dies aber ausgeschlossen sein.

Perner's Erklärung des Alterseffekts

Der wesentliche Punkt, den Kinder lernen müssen, um den FBT zu bestehen, ist nach Perner die Tatsache, dass Überzeugungen eine bestimmte *Einstellung* zu einer Proposition sind (die Proposition wird für wahr gehalten). Dreijährige Kinder halten natürlich bestimmte Propositionen für wahr und andere für falsch, doch sie verstehen Perner zufolge noch nicht, *dass* Menschen Propositionen für wahr oder falsch halten. Dieses Ver-

²⁴ Vierjährigen Kindern einen Überzeugungsbegriff zuzuschreiben setzt die Zuschreibung aller Begriffe voraus, die zur Explikation des Überzeugungsbegriffs nötig sind – vor allem *Proposition* und *semantische Bewertung*. Leslie (2000, 213f.) glaubt daher, dass Kindern mit dem Überzeugungsbegriff eine zu große begriffliche Kompetenz zugeschrieben werde, und scheint das an der Zahl der Begriffe festzumachen. Doch er begründet nicht, warum sich begriffliche Kompetenz in dieser Weise quantifizieren lassen sollte.

ständnis verknüpft Perner mit dem Überzeugungsbegriff: Ohne den Überzeugungsbegriff zu besitzen können Kinder nicht die Möglichkeit in Betracht ziehen, dass andere Menschen Propositionen für wahr halten, die sie selbst für falsch halten.

Im FBT müssen Kinder jedoch verstehen, dass Maxi für falsch hält, dass die Schokolade an ihrem neuen Ort ist, und stattdessen für wahr hält, dass die Schokolade immer noch an ihrem alten Ort ist – und das, obwohl sie selbst die beiden Propositionen anders bewerten. Da das Bestehen des FBT demnach das Verständnis der Möglichkeit von Überzeugungen anzeigt, die mit der eigenen inkompatibel sind, kann es durch den Besitz des Überzeugungsbegriffs erklärt werden.

My developmental claim about the intellectual change around 4 years is that the younger children have procedural knowledge of evaluating propositions against the world. They can assign truth values; they can decide whether the proposition is true or false but they do not understand that propositions are being evaluated. Consequently, they cannot conceive of the possibility that a proposition could be evaluated differently, e.g. a false proposition being evaluated as true as in the case of a false belief. The children who pass the false belief task have achieved this level of understanding. (Perner 2000, 376)

Eine Aufgabe aller *Conceptual-Change-Accounts* ist es nun zu zeigen, dass der Besitz des Überzeugungsbegriffs nicht schon zur Erklärung von Fähigkeiten benötigt wird, die *früher* in der kognitiven Entwicklung des Kindes auftreten. Bei diesem Nachweis steht die Fähigkeit zum spielerischen So-tun-als-ob (*Pretend Play*) im Vordergrund, die Kinder schon im Alter von zwei Jahren gut beherrschen.²⁵ Schon in diesem Alter verstehen Kinder, wenn Andere zum Beispiel so tun als sei eine Banane ein Telefon. Es kann daher davon ausgegangen werden, dass sie verstehen, dass Menschen auf der Grundlage falscher Überzeugungen handeln können. Perner muss diese Fähigkeit jedoch erklären, ohne ihnen ein Verständnis der semantischen Bewertung von Propositionen zuzuschreiben.

Dazu führt er einen Vorläufer des Überzeugungsbegriffs ein, dessen Besitz er zwei- und dreijährigen Kindern zuschreibt. Perner, Baker & Hutton (1994) nennen diesen Vorläuferbegriff 'Preief' (zusammengesetzt aus 'pretence' und 'belief'). Die geistigen Zustände, die er umfasst, sind ebenfalls propositionale Einstellungen, doch die Einstellung ist kein Für-wahr-halten. Perner unterlässt es, die Einstellung genauer zu charakterisieren, doch ich glaube, dass man einen Eindruck von seiner Idee bekommen kann.

²⁵ Erklärungen des *Pretend Play* diskutiere ich im dritten Teil der Arbeit und stelle die Fähigkeit dort auch genauer vor.

Zwei- und dreijährige Kinder kennen Perner zufolge zwei Arten von Handlungen: ernsthaft und bloß vorgegebene. Sie gehen mit einem Unterschied im Gehalt des zugrunde liegenden *Prelieds* einher: Ernsthaften Überzeugungen liegen *wahre Prelieds*, dem So-tun-als-ob hingegen *falsche Prelieds* zugrunde.²⁶ Um diese semantische Bewertung vorzunehmen, müssen Kinder nicht verstehen, dass sie sie vornehmen.

Da Maxi in FBT-Szenario nun offenbar nicht so tut als ob, sondern ernsthaft nach der Schokolade suchen will, können zwei- und dreijährige Kinder Maxis Verhalten nur aufgrund des *wahren Prelieds*, dass die Schokolade am neuen Ort ist, vorhersagen.

[T]he younger child can differentiate between pretend using something (banana) as a telephone and really using something as a telephone because the child can represent the difference in terms of the truth of the propositions on which these behaviours are based [...] In other words, what allows the young child to make this distinction is the ability to distinguish what is real (which propositions are true) from what is false (which propositions are false). Although this requires the child to be able to *evaluate* propositions, it does not require understanding *that* propositions are being evaluated and that difference people might evaluate one and the same proposition differently. This further ability is, however, required for distinguishing pretence from mistaken actions based on false belief [...]. (Perner 1995, 246, Hervorhebungen im Original)

Die Unterscheidung zwischen Fehlern und Pretend Play

Perner testet die folgende Implikation seiner Erklärung. Da in der vermeintlichen Theorie zwei- und dreijähriger Kinder ernsthafte Handlungen an wahre *Prelieds* gebunden sind, sind ernsthafte, aber realitätsunangemessene Handlungen – sprich: Fehler – begrifflich ausgeschlossen. Jüngere Kinder können also nicht zwischen Fehlern und *Pretend Play* unterscheiden.

Um diese Hypothese zu prüfen, fragten Perner, Baker & Hutton (1994) Kinder, ob die Protagonisten in verschiedenen Szenarien tatsächlich denken würden, dass *p*, oder nur so tun würden als ob. In einem Szenario ging es z.B. um die Frage, ob ein anderes Kind, das eine Karotte in einen Käfig hielt, ein Kaninchen füttert oder nur so tut als füttere es.

²⁶ Da der Besitz des *Prelied*-Begriffs kein Verständnis der semantischen Bewertung einschließt, scheinen diese propositionale Einstellungen von einem Typ zu sein, der zwischen einer Person und *allen* Propositionen besteht. Wenn jüngere Kinder demnach mit überzeugungsartigen Zuständen operieren, dann schreiben sie anderen Menschen zu jeder Proposition einen *Prelied* mit diesem Gehalt zu und unterscheiden lediglich zwischen *Prelieds* mit wahrem und solchen mit falschem Gehalt. Diesem Erklärungsmodell zufolge kommt mit dem Erwerb des Überzeugungsbegriffs eine dritte Kategorie ins Spiel: die des Nicht-Wissens.

In beiden Fällen war zwar kein Kaninchen im Stall, doch es wurde manipuliert, ob der Protagonist davon wusste. Die Autoren konnten zeigen, dass dreijährige Kinder in der Regel angeben, dass auch jemand, der noch glaubt, das Kaninchen sei im Stall, nur so tut, als würde er das Kaninchen füttern.

Dieses Ergebnis stützt die Hypothese, dass dreijährige Kinder Handlungen aufgrund falscher Überzeugungen als *Pretend Play* auffassen. Damit wird Perner zufolge eine Unterscheidung von zwei Einstellungen (Für-wahr- bzw. Für-falsch-Halten) überflüssig; es reicht die Unterscheidung von zwei Wahrheitswerten (Wahr- bzw. Falsch-sein) aus, um die Leistungen zwei- und dreijähriger Kinder zu erklären.

This confirms the results from our conceptual analysis that pretending, which children can do long before their fourth birthday, does not require a clear differentiation between belief and pretence. It only requires a clear differentiation between reality-adequate (acting-as-is) and reality-inadequate action (acting-as-if P). (Perner, Baker & Hutton 1994, 275)

Trotz dieses Belegs für die *Prelief*-These möchte ich im nächsten Abschnitt aus theoretischen Erwägungen gegen Perners Erklärung des Alterseffekts argumentieren. An dieser Stelle formuliere ich bereits zwei Bedenken. Zum einen ist der Nachweis von Perner, Baker & Hutton nicht überraschend, da schon vorher zu vermuten war, dass Kinder, die beim FBT durchfallen, Schwierigkeiten haben werden, Fehler als solche zu benennen. Wenn ihnen das So-tun-als-ob als eine von zwei Antwortmöglichkeiten geboten wird, könnte es sein, dass sie diese Alternative einfach als die plausiblere wählen.

Zum anderen ist Perner bislang den zweiten, nahe liegenden Beleg schuldig geblieben: den Nachweis, dass dem *Pretend Play* immer falsche *Preliefs* zugrundeliegen. Diese Implikation seiner Theorie ließe sich dadurch prüfen, dass dreijährigen Kindern gesagt wird, dass Maxi im FBT (oder in einem geeigneten anderen Szenario) *nur so tue*, als wolle er nach der Schokolade suchen. In diesem Fall sollten Kinder bei der Vorhersage des Verhaltens auf einen falschen *Prelief* zurückgreifen – z.B. auf den *Prelief*, dass die Schokolade immer noch an ihrem alten Ort ist.²⁷ In diesem Zusammenhang ist aber die Beobachtung Leslies (1994, 225) von Interesse, dass jüngere Kinder auch wahre *Pretend*-Überzeugungen zulassen: Man kann so tun, als sei eine leere Tasse leer – etwa dann, wenn eine leere, aber angeblich volle Tasse ausgeschüttet wurde. Auch jüngere

²⁷ Es ist denkbar, dass Kinder in einem solchen Experiment keine klare Antwort geben, da nicht vorhersehbar ist, auf welchen falschen *Prelief* sich jemand beim *Pretend Play* stützen wird. Dennoch würde dieser Befund von den systematisch falschen Antworten Dreijähriger im Standard-FBT abweichen.

Kinder können zwischen einer solchen leeren Tasse und einer leer gebliebenen Tasse unterscheiden.²⁸

3.2 Die Bestimmung von Überzeugungsgehalten

Mit dem Verständnis der *Möglichkeit* falscher Überzeugungen ist die Fähigkeit, den FBT zu bestehen, natürlich nicht vollständig erklärt. Um richtig vorherzusagen, wo Maxi nach der Schokolade suchen wird, müssen Kinder den *spezifischen Gehalt* Maxis Überzeugung über den Standort der Schokolade bestimmen. In diesem Abschnitt werde ich untersuchen, ob hierin der entscheidende Faktor für den FBT liegt.

Es sind mehrere Strategien denkbar, mit deren Hilfe Überzeugungsgehalte bestimmt werden können; ich möchte zunächst zwei einfache vorstellen. Die erste wird durch den Befund von Bartsch & Wellman (1989) motiviert, dass dreijährige Kinder realitätsunangemessene Handlungen unter Verweis auf falsche Überzeugungen erklären. Es scheint möglich, dass Kinder den Überzeugungsgehalt aus der vollzogenen Handlung ablesen. Die zweite Möglichkeit wird durch den Befund von Mitchell & Lacohee (1991) nahe gelegt, dass dreijährige Kinder ihre frühere falsche Überzeugung richtig angeben, wenn sie mit einer expliziten Repräsentation (einem Bild mit einer Gedankenblase) die-

²⁸ Zwei Bemerkungen zu Perner's Terminologie: Er spricht gelegentlich davon, dass das (von Nelson Goodman stammende) *Repräsentieren-als* für den FBT entscheidend sei. Um den Test zu bestehen, müssen Kinder nach Perner verstehen, dass sie und der Protagonist die Schokolade *als an verschiedenen Orten seiend* repräsentieren. Doch nicht nur Überzeugungen, auch das *Pretend Play* kann in dieser Weise beschrieben werden: Bereits dreijährige Kinder sind in der Lage, einen Sandhaufen *als einen Kuchen* zu repräsentieren. Daher sehe ich nicht, wieso Kinder erst im Alter von vier Jahren das Repräsentieren-als verstehen. Zudem unterscheidet Perner zwischen einer alltagspsychologischen Theorie *propositionaler Einstellungen* und einer Theorie *mentaler Repräsentationen*. Dabei hat er Fodors und Fields *Repräsentationale Theorie des Geistes* im Sinn, nach der mentale Repräsentationen die kausal wirksamen Agenten in kognitiven Prozessen sind. Perner scheint propositionale Einstellungen im Unterschied zu mentalen Repräsentationen als theoretisch konstruierte Dinge aufzufassen, die kein kausales Profil besitzen – vermutlich, weil sie abstrakt sind. Doch ich sehe nicht, welchen Vorteil es für die Charakterisierung des Verständnisses Vierjähriger es bringen soll, die Frage der physikalischen Realisierung geistiger Zustände ins Spiel zu bringen. Meines Erachtens kann den Kindern ohne weiteres ein Verständnis des kausalen Profils propositionaler Einstellungen zugeschrieben werden. Das wäre auch eher mit der Bedingung aus Abschnitt 1.1 in Einklang zu bringen, dass Kinder die ihnen zugeschriebenen alltagspsychologischen Theorien auch verstehen können sollten. Meines Erachtens sind diese beiden Rückgriffe auf philosophische Terminologie ungeeignet, um die kindliche Theorie des Geistes zu charakterisieren, weil sie keine zusätzliche Klarheit schaffen.

ser Überzeugung konfrontiert werden. Es scheint möglich, dass Kinder Überzeugungsinhalte aus expliziten Repräsentationen von Überzeugungen ablesen können.

Doch beide Alternativen eignen sich nicht für den Umgang mit dem Standard-Szenario im FBT. Um die veraltete Überzeugung des Protagonisten zu bestimmen, müssen komplexere Strategien angewendet werden. Denn seine Überzeugung ist nicht explizit repräsentiert und kann auch nicht aus einer Handlung abgeleitet werden. Und der o.g. Befund aus dem Janes-Katze-Experiment von Wellman & Bartsch (1988) weist darauf hin, dass dreijährige Kindern unter Umständen auch die ausdrücklich genannte falsche Überzeugung des Protagonisten in ihrer Vorhersage nicht berücksichtigen können.

In diesem Abschnitt werde ich eine für den FBT geeignetere Strategie der Überzeugungsbestimmung diskutieren, die mit dem Verständnis des *epistemischen Zugangs* in der Überzeugungsbildung zusammenhängt (siehe dazu auch Abschnitt 2.2). Perner zufolge ist eine Theorie des epistemischen Zugangs für die Erklärung des FBT-Alterseffekts unerheblich; ich werde im folgenden daher zunächst auf die entgegengesetzte Position Wimmers eingehen. Am Ende der Diskussion werde ich auf der Grundlage dieser Diskussion ein Argument gegen Perners Erklärung des Alterseffekts aus dem letzten Abschnitt formulieren.

Wimmers Erklärung des Alterseffekts

Wimmer & Gschaidner (2000) gehen davon aus, dass erst vierjährige Kinder über die Theorie des epistemischen Zugangs (EZ) verfügen, und behaupten, dass der Erwerb von (EZ) den Alterseffekt im FBT erkläre. Sie stützen diese Behauptung auf einen begrifflichen und einen empirischen Punkt. Der begriffliche Punkt ist dieser: Epistemische Zugänge sind nicht nur kausal verantwortlich für die Ausbildung von Überzeugungen, sondern liefern auch *gute Gründe* dafür, diese Überzeugungen zu besitzen.²⁹ Die Auto-

²⁹ Diesen Punkt formulieren sie sehr eigenwillig und ohne Rücksicht auf etablierte philosophische Terminologie: "Note that there is an intrinsic connection between 'taken a false thought to be reality' and 'having good reasons for that'. If you cannot come up with reasons that are good in the sense that they fully determined what you were thinking about the critical fact, then you cannot claim to have experienced the thought to be reality." (Wimmer & Gschaidner 2000, 255) Damit wollen die Autoren sagen: Solange man nicht in der Lage ist, hinreichende Gründe für eine (frühere, falsche) Überzeugung anzugeben, kann man nicht behaupten, diese Überzeugung unterhalten zu haben. Um eine Überzeugung zu unterhalten, darf man demnach nicht bloß geraten haben.

ren meinen, dass Kinder den FBT erst bestehen können, wenn sie verstehen, dass der Protagonist seine falsche Überzeugung durch einen geeigneten epistemischen Zugang begründen kann.

Diese begriffliche Analyse verknüpfen sie mit einem empirischen Befund aus Wimmer & Hartl (1991, 1. Experiment). Wimmer & Hartl konnten zeigen, dass nur vierjährige Kinder *begründen* können, warum sie zu einem früheren Zeitpunkt eine falsche Überzeugung unterhielten. Dreijährige Kinder können solche Begründungen auch dann nicht angeben, wenn sie mehrfach auf ihre frühere falsche Überzeugung hingewiesen werden und diesen Hinweis schließlich auch als korrekt anerkennen. Etwa ein Drittel der dreijährigen Kinder verweigerte die Zustimmung zu der Behauptung, es habe früher eine falsche Überzeugung unterhalten, sogar ganz (wodurch sich eine Begründung erübrigte). Das Ergebnis dieser Studie ist demnach, dass es kaum Kinder gibt, die falsche Überzeugungen zwar zuschreiben, diese nicht aber begründen können.

Aus dieser Korrelation des Auftretens der Begründungs-Fähigkeit und des FBT-Alterseffekts lässt sich meines Erachtens aber nicht schließen, dass der Erwerb von (EZ) für den FBT entscheidend ist. Denn der FBT ist komplizierter als es Wimmers Erklärungsmodell vermuten lässt. Der Protagonist könnte seine Überzeugung über den Standort der Schokolade zwar unter Rückgriff auf die Theorie (EZ) begründen. Doch aus Sicht des Kindes stellt sich die Situation anders dar; hier reicht die Anwendung von (EZ) nicht aus, um dem Protagonisten die richtige Überzeugung zuzuschreiben (und diese Zuschreibung auch begründen zu können). Das möchte ich nun erläutern.

Da die Abwesenheit des Protagonisten das hervorstechende Merkmal des FBT-Szenarios darstellt, ist es plausibel anzunehmen, dass Kinder mit Hilfe von (EZ) zunächst zur Feststellung gelangen, dass der Protagonist *nicht weiß*, dass sich die Schokolade an einem neuen Ort befindet. Doch diese Feststellung des Nicht-Wissens des Protagonisten reicht zum Bestehen des Tests nicht aus, da damit nicht geklärt ist, was der Protagonist *stattdessen* glaubt. Um letzteres zu ermitteln, muss das Kind verstehen, dass ein *zeitlich zurückliegender* epistemischer Zugang zur aktuellen Überzeugung des Protagonisten geführt hat. Doch der Besitz der Theorie (EZ) – wie ich sie im Abschnitt 2.2 formuliert habe – ermöglicht dieses Verständnis nicht. Denn es ist nicht ohne weiteres klar, auf welche Wahrnehmung des Protagonisten (EZ) angewendet werden soll.

Vielmehr muss (EZ) zu diesem Zweck ergänzt werden; benötigt wird eine Art *Aktualisierungstheorie* für Überzeugungen. Sie würde es Kindern ermöglichen, in der epi-

stemischen Geschichte des Protagonisten zurückzugehen, um den relevanten epistemischen Zugang zur Schokolade zu finden und damit den relevanten Überzeugungsgehalt zu ermitteln.³⁰ Die Theorie könnte beispielsweise so formuliert sein:

(EZ*) Sehen (usw.), dass p , führt zur Ausbildung der Überzeugung, dass p . Diese Überzeugung bleibt bestehen, solange man nicht sieht (usw.), dass q , und q mit p nicht kompatibel ist.

Mit dieser Theorie im Hinterkopf könnte ein Kind ungefähr wie folgt schlussfolgern: "Da Maxi nicht gesehen hat, dass die Schokolade an einen anderen Ort gelegt wurde, hat er seine Überzeugung über den Ort der Schokolade nicht aktualisiert – vorausgesetzt er hatte überhaupt eine. Die Voraussetzung ist erfüllt, da Maxi die Schokolade zuvor an einem anderen Ort gesehen hat. Er glaubt daher immer noch, dass die Schokolade dort ist, und würde bestreiten, dass sie sich an ihrem neuen Ort befindet."³¹ (EZ*) befähigt demnach das Kind zu begreifen, dass die Wahrnehmung, die Schokolade sei am Ort B, bei Maxi dazu geführt hätte, dass er seine Überzeugung, die Schokolade sei am Ort A, revidiert.

Zusammengefasst lautet mein Einwand gegen Wimmer: Obwohl der Protagonist im FBT seine Überzeugung gerade *nicht* revidiert, kann es den Kindern nicht verborgen bleiben, dass in diesem Szenario inkompatible Überzeugungen im Spiel sind, die auf verschiedene epistemische Kontakte zurückzuführen sind. Das lässt sich nur mit einer Aktualisierungstheorie von Überzeugungen erfassen.

Das führt nun zu der Frage, ob der Erwerb einer Theorie wie (EZ*) für das Bestehen des FBT entscheidend sein könnte? Für eine klare Antwort reichen die Daten m.E. noch nicht aus. Da es sich bei (EZ*) um eine komplexe Theorie handelt, ist zu vermuten, dass sie *nicht in einem Schritt* erworben wird. Das könnte die uneinheitliche Datenlage, die ich im Abschnitt 2.2 zum Thema epistemischer Zugang vorgestellt habe, erklären helfen. Möglicherweise verfügen dreijährige Kinder bereits über (EZ) und begreifen beispielsweise, dass jemand, der während eines Ereignisses nicht anwesend ist, von diesem Ereignis nichts wissen kann. Mit dieser Theorie geraten sie jedoch in Schwierigkeiten, sobald zusätzliche inkompatible Überzeugungen ins Spiel kommen (etwa ihre eigene

³⁰ Dieser Vorschlag taucht in der Forschungsliteratur meines Wissens nicht auf.

³¹ Natürlich sind weitere, leicht abgewandelte Varianten sowohl von (EZ*) als auch von dem skizzierten Gedankengang des Kindes denkbar. Mir geht es an dieser Stelle aber nicht um die Details, sondern um den allgemeinen Punkt, dass jede dieser Varianten ein Verständnis des Meinungswandels impliziert.

Überzeugung derjenigen des Protagonisten widerspricht). Daher fallen sie beim FBT und einigen anderen Tests zum epistemischen Zugang noch durch.

Doch die Robustheit des FBT-Alterseffekts lässt an diesem Entwicklungsmodell Zweifel aufkommen. Dreijährige Kinder scheinen sich in ihren falschen Antworten zu sicher sein, als dass man ihnen den Besitz einer Vorläufertheorie zuschreiben könnte. Sollten sie bereits über (EZ) verfügen, wäre zu erwarten, dass sie erkennen, dass der Protagonist nicht wissen kann, dass sich die Schokolade inzwischen an einem neuen Ort befindet. Doch nichts in ihren Antworten lässt auf eine solche Erkenntnis schließen.

Daher bleibt zwar die Möglichkeit bestehen, dass der Erwerb von (EZ*) für den FBT entscheidend ist. Doch meines Wissens hat bislang niemand diese These vertreten. Ich werde daher im dritten Teil dieser Arbeit eine Alternative zum (EZ*)-Erwerb formulieren und hoffen, dass diese Alternative überzeugt, weil sie vierjährigen Kindern keinen drastischen theoretischen Wandel abverlangt.

Epistemischer Zugang und Überzeugungsbegriff

Die Diskussion von Wimmers Position bringt mich zu einem theoretischen Einwand gegen Perner's Erklärung des FBT-Alterseffekts aus dem Abschnitt 3.1: Es ist möglich, dass Kinder den Überzeugungsbegriff besitzen, ohne zugleich in der Lage zu sein, die falschen Überzeugungen Anderer zu bestimmen. Denn schon die Fähigkeit, mit Hilfe der Theorie (EZ) das Nicht-Wissen des Protagonisten zu ermitteln, impliziert den Überzeugungsbegriff, da dem Protagonisten damit eine bestimmte Überzeugung *abgesprochen* wird. Mit dem *Prelief*-Begriff ist das – wie im letzten Abschnitt ausgeführt – nicht möglich; dass Kinder sich einer *Prelief*-Zuschreibung enthalten, hat Perner in seinem Erklärungsmodell nicht vorgesehen. Wenn es also stimmt, dass (EZ) zum Bestehen des FBT nicht ausreicht, dann reicht auch der Überzeugungsbegriff nicht aus.

In diesem Zusammenhang ist interessant zu bemerken, dass Wimmer & Hartl (1991) ihren o.g. Befund als Hinweis darauf werten, dass der Überzeugungsbegriff das Verständnis der Rolle des epistemischen Zugangs in der Überzeugungsausbildung impliziert. Ihnen zufolge misst der FBT sowohl den Erwerb des Überzeugungsbegriffs als auch den Erwerb einer Theorie der Überzeugungsausbildung.

One relevant finding was that most children who were able to identify their prior expectation as false belief were also able to respond correctly to the more demanding justification question, that is, they could

explain how they got their false belief. This finding suggests that informational causation is entailed by the child's concept of belief. (Wimmer & Hartl 1991, 135)

Doch der Überzeugungsbegriff – wie er im letzten Abschnitt erläutert wurde – impliziert keine bestimmte Theorie der Überzeugungsbildung. Vielmehr verfügen Kinder über den Überzeugungsbegriff, sobald sie *irgendeine* Strategie kennen, die es erlaubt, einer anderen Person eine aus ihrer Sicht falsche Überzeugung zuzuschreiben.

3.3 Eine repräsentationale Theorie des Geistes?

In diesem Abschnitt möchte ich die Position untersuchen, die den Erwerb des Überzeugungsbegriffs als Teil eines umfassenderen Wandels in der kindlichen Theorie des Geistes beschreibt. Wie bereits im zweiten Kapitel deutlich geworden ist, gibt es eine Reihe korrelierender Alterseffekte, d.h. eine Reihe von Fähigkeiten, die gleichzeitig in der kognitiven Entwicklung des Kindes auftreten. Dazu zählen neben der Zuschreibung falscher Überzeugungen vor allem das Verständnis der Perspektivität der visuellen Wahrnehmung und die Fähigkeit zur Unterscheidung zwischen Sein und Schein.

Dieses Entwicklungsmuster wird von einigen Autoren analog zum Theorienwandel in der Wissenschaft interpretiert. Dieser Ansatz vom Kind als naiven Wissenschaftler ist kritisiert worden, weil er eine zu einheitliche Entwicklung verschieden veranlagter und unterschiedlich geförderter Kinder postuliere (Carruthers 1996, 23). Doch ich sehe in den vorliegenden experimentellen Befunden trotz einer großen Streuung der individuellen alltagspsychologischen Leistungen genug Gemeinsamkeiten in der Entwicklung, um von einem Muster zu sprechen. Ich werde das postulierte Entwicklungsmuster jedoch aus anderen Gründen zurückweisen.

Merkmale des theoretischen Wandels

Gopnik & Wellman (1992, 1994) zählen *dynamische* Eigenschaften von Theorien auf, die sich ihrer Ansicht nach in der Entwicklung der alltagspsychologischen Praxis wieder finden. Das für ihre Argumentation wichtigste Merkmal ist das Konstruieren und Ausprobieren neuer Hypothesen im Licht beobachteter Anomalien.

We propose that [...] dynamic features should be apparent in children's transition from one theory to a later one, and specifically from one view of the mind to another. Children should ignore certain kinds of counterevidence initially. Later, they might account for them by auxiliary hypotheses, only invoked to deal with particular kinds of counterevidence. Furthermore, children might first use the new theoretical ideas in limited contexts, contexts closely related to the conceptual structure of the earlier theory. Only after the theory change is complete, will they finally reorganize their knowledge so that new theoretical entities play a certain role. (Gopnik & Wellman 1994, 263f.)

In der Wissenschaft bleibt eine neue Theorie zunächst auf den erfolgreich behandelten Phänomenbereich der Vorgängertheorie und die zusätzlich beobachteten Anomalien beschränkt. Erst später wird die größere Reichweite der neuen Theorie erforscht. Gopnik & Wellman leiten daraus zwei Vorhersagen für die Entwicklung der alltagspsychologischen Theorien ab. Beide Vorhersagen betreffen das kritische Alter von drei Jahren und gehen davon aus, dass die neue Theorie eine *Theorie von Repräsentationen* ist. Diese Theorie – so die erste Vorhersage – werde zunächst nur auf die bekannten geistigen Zustände, d.h. auf Wünsche und Wahrnehmungen, angewandt.³² Zudem gebe es einen gewissen Widerstand, die neue Theorie überhaupt anzuwenden: Kinder tun dies zunächst nur – so die zweite Vorhersage –, wenn sie mit der Nase darauf gestoßen werden.

[I]n a transitional period the crucial idea of the new theory may appear as an auxiliary hypothesis couched in the vocabulary of the original theory, or be used in order to deal with particularly salient types of counter-evidence, but may not be widely applied. There is evidence for both these phenomena in the period from three to four. Children seem to us to initially develop the idea of misrepresentation in familiar contexts like those of desire and perception, without extending the idea more generally. They also initially apply the idea only when they are forced to by counter-evidence. (Gopnik & Wellman 1992, 156f.)

Als Belege für die zweite Vorhersage, dass Kinder ihre neue Theorie zunächst nur bei salienten Anomalien einsetzen, zitieren Gopnik und Wellman die oben vorgestellten Studien von Bartsch & Wellman (1989) und Mitchell & Lacohee (1991). In beiden Studien zeigten bereits dreijährige Kinder etwas von dem Verständnis, das nötig ist, um den FBT zu bestehen: Sie sind in der Lage, Handlungen aufgrund falscher Überzeugungen zu erklären, und sie geben ihre eigenen früheren falschen Überzeugungen zu, wenn sie mit einem selbst ausgewählten Bild konfrontiert werden, das ihre frühere Überzeugung repräsentiert. Gopnik & Wellman interpretieren diese Daten wie folgt: Kinder schreiben sich selbst und Anderen nur dann falsche Überzeugungen zu, wenn sie ausdrücklich darauf hingewiesen werden, dass die wahren Alternativen nicht mehr in Betracht kommen.

³² Diese alltagspsychologische Theorie Zweijähriger habe ich im Abschnitt 2.2 vorgestellt.

Diese zweite Vorhersage ist meines Erachtens nicht spezifisch für einen theoretischen Wandel. Ein Übergang, wie ihn Gopnik & Wellman beschreiben, wäre auch mit dem Simulationsmodell kompatibel. Zum Beispiel könnte es sein, dass dreijährige Kinder erst allmählich lernen, geistige Prozesse zu simulieren, die stark von ihren eigenen abweichen. Dieses ungeübte und aufwendige Verfahren versuchen sie jedoch zu vermeiden und wenden es deshalb nur in solchen Fällen an, in denen die schlichte Übertragung ihrer eigenen geistigen Zustände offensichtlich zu falschen Ergebnissen führt.

Die Entwicklung einer repräsentationalen Theorie des Geistes

Daher wende ich mich der ersten Vorhersage der Autoren zu: Kinder lernen nach und nach die repräsentationale Natur verschiedener geistiger Zustände. Als Beleg dafür, dass Kinder bereits die repräsentationale Natur von Wünschen und Wahrnehmungen verstehen, bevor sie die Repräsentationstheorie auch auf Überzeugungen ausdehnen, zitieren Gopnik & Wellman die Studie von Gopnik & Slaughter (1991). Dieser Untersuchung zufolge verstehen dreijährige Kinder bereits die repräsentationale Natur fast aller geistigen Zustände – mit Ausnahme von Überzeugungen.

Was heißt es, die repräsentationale Natur eines geistigen Zustands zu verstehen? Gopnik & Wellman haben nur einige Aspekte dessen im Blick, was einen repräsentationalen geistigen Zustand ausmacht. Sie betonen, dass Kinder lernen müssten, dass Überzeugungen nicht notwendigerweise mit den repräsentierten Sachverhalten kovariieren.

What we want and see (by and large) is not the thing itself but the thing as represented. Understanding some aspects of desire and perception requires this sort of representational understanding. When we are satiated with something we no longer desire it, but the object itself has not changed. When different types of people have different tastes or values, their desires differ but the objects of desire remain the same [...]. There is evidence that these representational aspects of desire are understood earlier than equivalently representational aspects of belief [...]. (Gopnik & Wellman 1992, 157)

Dementsprechend zeigten Gopnik & Slaughter (1991), dass dreijährige Kinder frühere, bereits erfüllte Wünsche korrekt angeben können; ebenso können sie auch dann noch sagen, was sie gesehen haben, nachdem das Objekt aus ihrem Blickfeld verschwunden ist. Sie können jedoch nicht sagen, was sie geglaubt haben, nachdem sich der relevante Aspekt der Welt geändert hat.

Mit diesem Befund belegen die Autorinnen sicherlich, dass sich Überzeugungen deutlich von anderen propositionalen Einstellungen unterscheiden. Insbesondere bestätigen

sie, dass veraltete Überzeugungen im Gegensatz zu veralteten Wünschen und Wahrnehmungen ein Problem für jüngere Kinder darstellen. Doch ich glaube nicht, dass dieser Unterschied viel mit der repräsentationalen Natur der geistigen Zustände zu tun hat. Denn um beispielsweise anzugeben, was man gesehen hat, bevor man die Augen schloss, genügt es, sich daran zu *erinnern*, was zu diesem Zeitpunkt im Blickfeld war. Dazu muss man nicht verstehen, dass der Gehalt dieser Wahrnehmung einen Ausschnitt der Welt auf eine bestimmte Weise repräsentiert.

Ein besserer Test des Verständnisses der repräsentationalen Natur geistiger Zustände wäre daher zu prüfen, ob verschiedene Menschen verschiedene Einstellungen zu derselben Proposition haben können. Einen solchen Test gibt es schon für das Sehen: Flavell konnte zeigen, dass Kinder erst im Alter von vier Jahren verstehen, dass Menschen ein Objekt aus verschiedenen Perspektiven unterschiedlich sehen – und damit unterschiedlich repräsentieren – können (siehe Abschnitt 2.3). Dieser Befund spricht gegen die These von Gopnik & Wellman, dass dreijährige Kinder die repräsentationale Natur aller geistigen Zustände (bis auf Überzeugungen) verstehen. Daher zweifle ich an der Richtigkeit der von ihnen skizzierten kognitiven Entwicklung des Kindes.

Als letzten Punkt möchte ich das *Pretend Play* ansprechen: Müssen *Pretend*-Überzeugungen von zweijährigen Kindern nicht als Repräsentationen betrachtet werden? Gopnik & Wellman gestehen zu, dass *Pretend*-Zustände als repräsentationale Zustände aufgefasst werden müssten, doch sie geben zu bedenken, dass dreijährige Kinder diese Zustände als kausal unabhängig von der Welt auffassen würden.

While [by their third birthday] children see desires as states that modify the world, and perceptions as states that are modified by the world, pretenses, images and dreams, on their view, bear no causal relation to the world at all. It is possible that postulating these states, which are representational but divorced from reality, also plays a role in the eventual development of the full representational theory. (Gopnik & Wellman 1992, 152)

Diese Charakterisierung von *Pretend*-Überzeugungen macht meines Erachtens aber nicht klar, worin der Unterschied zu normalen, nicht-repräsentationalen Überzeugungen bestehen soll.

Fazit: Unerklärte Daten und gesuchte Indikatoren

Ziel des ersten Teils war zu untersuchen, wie gut kognitive Theorienmodelle alltagspsychologische Fähigkeiten erklären können. Ich habe mich – der Forschungsliteratur folgend – auf die Fähigkeit konzentriert, realitätsunangemessene Handlungen aufgrund falscher Überzeugungen vorherzusagen. Diese Fähigkeit wird vornehmlich mit dem *False Belief Test* gemessen. Ich bin zu dem Ergebnis gekommen, dass weder die Daten noch die theoretischen Überlegungen, die in diesem Zusammenhang vorgebracht werden, eine Erklärung dieser Fähigkeit im Sinn eines kognitiven Theorienmodells stützen. Das heißt, dass meines Erachtens bislang keine guten Gründe für die Behauptung vorliegen, vierjährige Kinder würden die Testfrage im FBT nur mit Hilfe alltagspsychologischer Theorien beantworten. Ich möchte nun kurz den Weg skizzieren, auf dem ich zu diesem Fazit gelangt bin.

Im dritten Kapitel habe ich mich mit der Frage befasst, ob der Alterseffekt im FBT auf einen Wandel in der impliziten kindlichen Theorie des Geistes hinweist. Ich habe drei Positionen untersucht, die den Alterseffekt auf den Erwerb konkreter alltagspsychologischer Theorien oder Begriffe zurückführen. Ihr Hauptproblem ist meines Erachtens, dass sie die Leistungen Dreijähriger nicht zufrieden stellend erklären können. Zwei Beobachtungen stehen dabei im Vordergrund: das *Pretend Play* und das systematisch falsche Antworten im FBT.

Zunächst zum *Pretend Play*: Wenn dreijährige Kinder so tun als sei eine Banane ein Telefon, ist es meines Erachtens plausibel anzunehmen, dass sie verstehen, dass sie eine falsche Überzeugung im Rahmen eines *Pretend*-Szenarios für wahr halten. Es sind zwar andere Rekonstruktionen der kindlichen Theorie des Geistes denkbar (wie Perners Alternative der *Prelieds*), doch im Vergleich dazu wirken sie bemüht. Zudem reicht ein Verständnis von Überzeugungen als propositionale Einstellung oder geistige Repräsentation nicht aus, um den FBT zu bestehen, da es keine Hilfe bei der Bestimmung der relevanten Überzeugung des Protagonisten bietet. Zu wissen, dass Menschen inkompatible Einstellungen zur selben Proposition haben können, sagt einem nicht, welche Einstellung sie in einem konkreten Fall haben. Daher lassen diese Modelle die Frage offen, ob zum Bestehen des FBT nicht noch weitere theoretische Ressourcen erworben werden müssen.

Der entscheidende Faktor zum Bestehen des FBT muss zwar nicht zugleich ein hinreichender sein, so dass mein letzter Einwand nicht grundsätzlich gegen Perner's Theorie spricht. Doch der Überzeugungs- und der Repräsentationsbegriff erscheinen im Rahmen der Theorienmodelle zu dünn, um erst im Alter von vier Jahren eine Rolle zu spielen. Den Typen geistiger Zustände, die sie bezeichnen, fehlt das kausale Profil. Dabei ist die Einführung neuartiger theoretischer Begriffe nicht der einzige denkbare theoretische Wandel; auch die Erkenntnis neuer kausaler Zusammenhänge stellt einen Fortschritt dar – mitunter einen dramatischen, der dem robusten Alterseffekt im FBT gerecht werden kann.

Ein solcher Fortschritt wäre der Erwerb einer Aktualisierungstheorie für Überzeugungen. Eine solche Theorie namens (EZ*) habe ich im Abschnitt 3.2 vorgestellt. Ich habe dort dafür argumentiert, dass Kindern beim Lösen der FBT-Aufgabe auffallen muss, dass der Protagonist seine Überzeugung über den Standort der Schokolade nicht aktualisiert hat und daher eine veraltete Überzeugung unterhält. Der Erwerb dieser Theorie passt aber nicht zu den systematisch falschen Antworten Dreijähriger, da diese Theorie zu komplex erscheint, um als Ganze innerhalb kurzer Zeit erworben zu werden. Plausibler wäre es, wenn dreijährige bereits über eine Vorläufertheorie verfügten. Doch wie auch immer diese geartet wäre, müsste sie doch verhindern, dass Dreijährige so deutlich zur falschen Antwort tendieren. Daher glaube ich, dass der Alterseffekt im FBT noch nicht hinreichend erklärt worden ist, um den Schluss auf einen theoretischen Wandel und damit auf ein kognitives Theorienmodell zu rechtfertigen.

Die falscher Antworten Dreijähriger im FBT könnten auch auf ein *Performanzdefizit* hinweisen. Einen Kandidaten dafür habe ich im Abschnitt 2.2 bereits vorgestellt: die verbale Explikation des alltagspsychologischen Urteils. Es ist denkbar, dass Dreijährige bereits die richtige Überzeugung des Protagonisten bestimmen, diese aber nicht als Antwort angeben können. Eine solche Position werde ich im dritten Teil dieser Arbeit ausführen.

Im zweiten Teil werde ich jedoch erst einmal das Simulationsmodell genauer vorstellen und im 5. und 6. Kapitel auf theoretische Überlegungen und empirische Studien zur Unterscheidbarkeit von Theorien- und Simulationsmodell zu sprechen kommen. Dabei werden einige der Merkmale der jeweiligen kognitiven Architektur im Vordergrund stehen, die ich im Abschnitt 1.2 angedeutet habe. Ich werde allerdings dafür argumentieren, dass sich diese Merkmale nicht unmittelbar in empirische Tests übersetzen las-

sen. Zwischen Theorien- und Simulationsmodell kann aus praktischen Gründen nur innerhalb gut erforschter Kontexte unterschieden werden. Im dritten Teil der Arbeit werde ich daher auf die *Theory-of-Mind*-Forschung aus diesem ersten Teil zurückkommen, die einen solchen Kontext bietet.

II Die Diskussion um das Simulationsmodell

Vorbemerkung: Was leistet die Alltagspsychologie?

Mitte der 80er-Jahre formulierten mehrere Autoren unabhängig voneinander das Simulationsmodell als Alternative zum Theorienmodell. Als Gründungsdokumente des Simulationsmodells gelten: Heal (1986), Gordon (1986) und Goldman (1989); seltener genannt werden Ripstein (1987), Johnson (1988) und Harris (1991). Die erste größere Diskussion des Simulationsmodells wurde 1992 in einer Sonderausgabe der Zeitschrift *Mind & Language* dokumentiert (wiederabgedruckt in: Davies & Stone 1995a).³³

Mit den neuen Modellen verfolgten die Autoren recht unterschiedliche Ziele; ich werde sie daher in dieser Arbeit nicht alle berücksichtigen. Ausgangspunkt der Debatte, die mich interessiert, war der Verdacht, dass die geistigen Vorgänge, mit denen Menschen alltagspsychologisch umgehen, nicht hinreichend durch eine Theorie und schon gar nicht durch eine Alltagstheorie erfasst werden können. Die alltagspsychologischen Theorien, die Menschen anwenden können, sind dieser Position zufolge *zu grob*, um den subtilen alltagspsychologischen Differenzierungen, zu denen Menschen fähig sind, Rechnung zu tragen. Bevor ich mich jedoch mit diesem Argument genauer auseinandersetze, möchte ich auf eine trivial erscheinende Form der Simulation hinweisen.

Zwei Formen der Simulation

Ebenso wie niemand bestreitet, dass Menschen über einige alltagspsychologische Theorien verfügen, bestreitet auch niemand, dass Menschen gelegentlich die kognitiven Prozesse Anderer simulieren. Wenn ich zum Beispiel eine (einfache) Rechenaufgabe löse, kann ich den kognitiven Prozess, der mich zur Lösung der Aufgabe bringt, als Simulation von kognitiven Prozessen ansehen, die Andere zur Lösung der Aufgabe bringen. Das ist möglich, weil ich davon ausgehen kann, dass Menschen einfache Rechenaufgaben ähnlich lösen. Es ist nicht einmal nötig, sich selbst als Modell zu verwenden: Ich kann

³³ Kurz darauf erschienen sind die Sammelbände Davies & Stone (1995b) und Carruthers & Smith (1996).

auch das, was im Kopf einer dritten Person A vorgeht, wenn sie sich in der Situation X befindet, als Simulation dessen auffassen, was im Kopf von Person B vorgeht, wenn sie sich in der Situation X befindet.

Wichtig ist in allen diesen Fällen lediglich, dass die kognitiven Mechanismen der simulierenden und simulierten Person ähnlich sind. Heal (2000) hat darauf aufmerksam gemacht, dass diese Bedingung die Reichweite der Simulation deutlich einschränkt. Insbesondere bei höheren kognitiven Leistungen dürften Kompetenzunterschiede Simulationen aussichtslos machen. Heals Beispiel ist das eines Schachexperten, der die Züge eines Anfängers vorhersagen soll. In diesem Fall werde die Ähnlichkeitsbedingung verletzt und eine Simulation daher erfolglos.³⁴

Substanziellere Simulationsmodelle behaupten, dass Menschen auch zu Simulationen *in ihrer Einbildung* fähig sind. Diesen Modellen zufolge kann man – indem man sich vorstellt, in der Situation einer anderen Person zu sein – an seinen eigenen Reaktionen ablesen, was der anderen Person durch den Kopf geht. Diese Grundidee wird häufig in Analogie zur Simulation des Verhaltens von Flugzeugen durch kleinere Flugzeugmodelle im Windkanal erläutert. Versuche mit Modellen im Windkanal sind deshalb nützlich, weil sie Vorhersagen über das Verhalten von Flugzeugen erlauben, die sich nur schlecht oder gar nicht aus den bekannten aerodynamischen Theorien ableiten lassen. Diesen Vorteil könnten auch die eigenen kognitiven Mechanismen bieten – sofern sie denen der simulierten Person ähnlich sind.

Um seine eigenen kognitiven Mechanismen als Modell für die Mechanismen Anderer nutzen zu können, muss es einen Mechanismus geben, der dem Experimentieren im Windkanal entspricht: ein Mechanismus, der andere kognitive Mechanismen zum Zweck der Simulation ansteuert, mit geeignetem Input versorgt und das Ergebnis zu alltagspsychologischen Urteilen weiterverarbeitet. Dieser übergeordnete Mechanismus darf jedoch das Funktionieren der von ihm angesteuerten Mechanismen nicht wesentlich beeinflussen, weil sonst Gefahr bestünde, dass die Ähnlichkeitsbedingung verletzt wird. Welcher kognitive Mechanismus diese Aufgaben erfüllen kann, ist beim gegen-

³⁴ Es ist interessant zu bemerken, dass in der Debatte um das Simulationsmodell hauptsächlich Beispiele aus dem Repertoire Erwachsener diskutiert werden. Die kognitiven Leistungen von *Kindern*, die in der *Theory-of-Mind*-Forschung untersucht werden, spielen kaum eine Rolle. Ich werde mich im dritten Teil der Arbeit jedoch um eine Simulations-Erklärung der Fähigkeit bemühen, den *False Belief Test* zu bestehen.

wärtigen Forschungsstand völlig offen. Doch auch ohne die Grundlagen des Ansteuerns der eigenen kognitiven Mechanismen zu kennen, werde ich in diesem Teil genug Diskussionsstoff haben, da auch die Forschungsdebatte ohne Kenntnis der kognitiven Grundlagen der Simulation vorangeschritten ist.

Hybridmodelle

Zu Beginn der Debatte ist das Simulationsmodell oft so verstanden worden, dass es die Relevanz theoretischer Ressourcen für die alltagspsychologische Praxis bestreite. Gegen das so verstandene Simulationsmodell wurde argumentiert, dass für jede Simulation *Anfangsbedingungen* bestimmt werden müssten. Zu den Anfangsbedingungen gehöre die Bestimmung relevanter *psychologischer* Unterschiede zwischen der anderen Person und sich selbst. Simulationen von Entscheidungsprozessen sollten zum Beispiel von den Wünschen der simulierten Person abhängen. Es sei jedoch unklar, wie die Wünsche der anderen Person bestimmt werden können, ohne dabei alltagspsychologische Theorien anzuwenden. In ähnlicher Weise wurde argumentiert, dass auch die *Übertragung* des Simulationsergebnisses auf die andere Person (in Form einer Zuschreibung, Vorhersage oder Erklärung) die Anwendung theoretischer Begriffe erfordere.

Diese Einwände werde ich im vierten Kapitel vortragen. Sie treffen meines Erachtens zu, doch sie gehen am Kern des Simulationsmodells vorbei. Denn das Simulationsmodell kann sich auf die Behauptung beschränken, dass Simulationen von den (theoretisch bestimmten) Anfangsbedingungen zum (theoretisch weiterzuverarbeitenden) Ergebnis führen. Nur für diesen Zwischenschritt wird bestritten, dass theoretische Ressourcen eine Rolle spielen, weil gerade in diesem Schritt die entsprechenden eigenen kognitiven Mechanismen angewendet werden. Das Simulationsmodell konkurriert also nur mit einem Theorienmodell, das behauptet, der *gesamte* kognitive Prozess bis hin zum alltagspsychologischen Urteil lasse sich durch die Anwendung alltagspsychologischer Theorien erklären. Meine Position lässt sich auch als Hybridmodell beschreiben, da Theorie und Simulation zusammenwirken müssen, um ein alltagspsychologisches Urteil zustande zu bringen.

Mit 'Hybridmodell' kann auch gemeint sein, dass Theorien- und Simulationsmodell unterschiedliche Bereiche der alltagspsychologischen Praxis erklären. Ich halte auch eine solche Form der Arbeitsteilung für plausibel. Meines Erachtens eignet sich das Simula-

tionsmodell vor allem zur Erklärung von Fällen, in denen es auf das Hintergrundwissen ankommt. Da ich mich in dieser Arbeit auf den Umgang mit falschen Überzeugungen konzentrieren möchte (und diese spezielle Leistung im Sinn des Simulationsmodells erklären will), werde ich nicht viel über andere Bereiche der alltagspsychologischen Praxis sagen.

Die Herausforderung

Goldman (1989) zufolge lässt sich introspektiv den Eindruck gewinnen, dass man gelegentlich simuliert:

The simulation idea has obvious initial attractions. Introspectively, it seems as if we often try to predict other's behavior – or predict their (mental) choices – by imagining ourselves in their shoes and determining what we would choose to do. (Goldman 1989, 169)

Dieser Eindruck reicht natürlich nicht aus, um das Simulationsmodell zu stützen. Denn auch in anderen Wissensbereichen kann man den Eindruck haben, kein theoretisches Wissen anzuwenden, obwohl es klar ist, dass die kognitive Leistung entscheidend von theoretischem Wissen abhängt. Eine bekannte Formulierung dieses Punkts bietet Dennett (1987):

If I make believe I am a suspension bridge and wonder what I will do when the wind blows, what "comes to mind" in my make-believe state depends on how sophisticated my knowledge is of the physics and engineering of suspension bridges. Why should my making believe I have your beliefs be any different? In both cases, knowledge of the imitated object is needed to drive the make-believe "simulation," and the knowledge must be organized into something rather like a theory. (Dennett 1987, 100f.)

Damit wird das Simulationsmodell herausgefordert: Wie wird ausgeschlossen, dass das vermeintliche Simulieren in der Einbildung durch die Anwendung alltagspsychologischer Theorien gesteuert wird? Diese Steuerung muss der simulierenden Person nicht bewusst sein. Dennett appelliert an physikalische Transformationen in der Einbildung. Neigt man in seiner Einbildung ein Glas Wasser zur Seite, wird die kausale Entwicklung in der Einbildung davon abhängen, was man über das Verhalten von Flüssigkeiten weiß. Dem Theorienmodell zufolge verhält es sich in der Alltagspsychologie genauso: Beleidigt man in seiner Einbildung eine andere Person, wird die Reaktion dieser Person in der Einbildung davon abhängen, was man über das Verhalten von beleidigten Menschen weiß.

Doch im Fall von Überzeugungen – so die Behauptung von Heal, die ich im fünften Kapitel vorstellen werde – besteht die Möglichkeit, direkt über die Gehalte der Überzeugungen nachzudenken, die man simuliert. Das ist Heal zufolge auch die effizientere Strategie, da die alltagspsychologische Praxis zu differenziert ist, um von alltagspsychologischen Theorien erfasst zu werden. Es ist schließlich plausibel anzunehmen, dass Theorien für den Alltagsgebrauch nicht beliebig komplex sein dürfen.

Als Maßstab für die Differenziertheit der Alltagspsychologie verwendet Heal den Vergleich mit der wissenschaftlichen Psychologie. Man neigt dazu, die Alltagspsychologie für besonders differenziert zu halten, wenn sie Erklärungen und Vorhersagen liefert, an denen die wissenschaftliche Psychologie scheitert. Heal nennt als Beispiel alltägliche Schlussfolgerungen, an denen sich die Forschung zur künstlichen Intelligenz (KI) seit Jahrzehnten die Zähne ausbeißt. Das Problem der KI-Forschung ist das Heranziehen relevanter Elemente des Hintergrundwissens: Es ist nicht klar, wie Menschen unter Zeitdruck bestimmen, welche Wissens Elemente zur Beantwortung einer bestimmten Frage relevant sind. Es sei daher zu vermuten, so Heal, dass Menschen über keine entsprechend komplexe alltagspsychologische Theorie verfügen, die die Vorhersage von Schlussfolgerungen ermöglicht. Meines Erachtens funktioniert dieses Argument nicht ohne weiteres. Denn es ist nicht von vornherein ersichtlich, dass Menschen in der Lage sind die Schlussfolgerungen vorherzusagen, die die KI-Forscher zu beschreiben versuchen. Möglicherweise ist die Alltagspsychologie in diesem Bereich nicht so gut, wie es Heal voraussetzt.

Die empirischen Studien, auf die ich im sechsten Kapitel zu sprechen komme, zeigen, dass es im Prinzip möglich ist, Heals Erwiderung auf Dennett zu überprüfen. In den Experimenten werden allgemeine Merkmale des Simulationsprozesses und des Prozesses der Theorieanwendung mit empirischen Indikatoren verknüpft. Zum Beispiel lässt sich vermuten, dass sich die Anwendung von Theorien durch geeignete Hinweise verbessern lässt. Verbessern sich nun die alltagspsychologischen Leistungen von Menschen, denen geeignete Hinweise gegeben wurden, kann darauf geschlossen werden, dass sie Theorien angewendet haben. Versuche dieser Art scheitern meines Erachtens jedoch daran, spezifische Hypothesen für konkrete experimentelle Situationen zu formulieren. Im Experiment wird in der Regel die Qualität alltagspsychologischer Erklärungen und Vorhersagen erhoben. Selbst wenn man die allgemeine Verknüpfung der Prozessmerkmale mit einer Veränderung in der Qualität akzeptiert, schließt man damit

nicht aus, dass auch andere Faktoren auf die Qualität einwirken. Zum Beispiel lässt das Theorienmodell die Falschheit alltagspsychologischer Theorien zu, so dass ein vermeintlich geeigneter Hinweis ins Leere laufen kann, wenn die Versuchspersonen nicht über die entsprechende Theorie verfügen.

Für das Simulationsmodell gelten ähnliche Einschränkungen: Es erlaubt zum Beispiel die gelegentliche Anwendung von Theorien; es erlaubt sogar, dass die Anwendung falscher Theorien ein richtiges Simulationsergebnis "kontaminiert". Daher lässt sich beispielsweise aus dem Befund, dass Menschen oft ihr eigenes Verhalten nicht richtig vorhersagen können (obwohl in diesem Fall die Ähnlichkeitsbedingung erfüllt ist), nicht ohne weiteres schließen, dass sie nicht simulieren würden. Denn es könnte sein, dass ein wenig schmeichelhaftes Simulationsergebnis intern zensiert wird.

Ich werde dafür argumentieren, dass die Verknüpfung von Prozessmerkmalen mit empirischen Indikatoren nur gelingt, wenn klar ist, was die Alltagspsychologie leistet, d.h. wenn davon ausgegangen werden kann, dass eine bestimmte alltagspsychologische Theorie angewendet bzw. der ein bestimmter kognitiver Mechanismus simuliert wird. Da solche Begründungen bislang fehlen, werde ich schließen, dass die Debatte um das Simulationsmodell noch unentschieden ist. Wie im ersten Teil jedoch angekündigt, bieten sich die Ergebnisse der *Theory-of-Mind*-Forschung an, um den Interpretationsspielraum einzuschränken. Darauf werde ich im dritten Teil eingehen.

4 Die Grundidee des Simulationsmodells

4.1 Faktische und hypothetische Simulation

Der Ansatz, dass es die eigenen geistigen Regularitäten sind, die die alltagspsychologische Praxis leiten, muss um eine Strategie erweitert werden, mit deren Hilfe die eigenen geistigen Regularitäten zur Vorhersage oder Erklärung ausgebeutet werden können.

Die denkbar einfachste (und unumstrittene) Strategie ist: sich *faktisch* in die Situation der anderen Person zu begeben und die Hoffnung zu hegen, dass alle Menschen in dieser spezifischen Situation ähnlich denken, fühlen und handeln. Will man etwa vorhersagen, wie sich ein Mensch im Löwenkäfig bewegen wird, kann man selbst in einen Löwenkäfig steigen und seine eigene Reaktion als Grundlage für eine Vorhersage verwenden. Hierbei handelt es sich um den Typ trivialer Simulationen, den ich in der Vorbemerkung zum zweiten Teil erwähnt habe.

Diese Strategie hat offensichtliche Nachteile: Sie ist aufwändig, womöglich gefährlich, und auf die Fälle begrenzt, in denen psychologische Unterschiede zwischen den Menschen keine Rolle spielen. Denn die eigenen Reaktionen sind nur dann eine gute Grundlage für eine Vorhersage, wenn die relevanten psychologischen Faktoren bei simulierender und simulierter Person ähnlich sind. Es ist zum Beispiel schwer, auf diese Weise das Verhalten eines Löwenbändigers im Käfig richtig vorherzusagen, wenn man selbst unerfahren und ängstlich ist. Dennoch wird diese Strategie gelegentlich als trivialer Grenzfall einer umfassenderen Simulationsfähigkeit angesehen, da auch sie dem Ansatz Rechnung trage, die eigenen kognitiven Mechanismen zu nutzen.

Stich & Nichols (1997, 305f.) machen aber darauf aufmerksam, dass eine umfassendere Simulationsstrategie zusätzliche kognitive Elemente enthalten kann, die einen Bruch zwischen beiden Strategien darstellen. Gordon (1986) bietet eine solche umfassendere Strategie an: sich *hypothetisch*, d.h. in seiner Einbildung, in die Situation einer anderen Person hineinzusetzen und ihre geistigen Prozesse nachzuvollziehen. Stellt man sich selbst in der Situation einer anderen Person vor, kann man darauf hoffen, mit den in diesem imaginativen Kontext ablaufenden eigenen geistigen Prozessen die geistigen Prozesse dieser Person zu simulieren. Auch diese Strategie setzt voraus, dass die geisti-

gen Prozesse bei simulierender und simulierter Person ähnlich ablaufen. Gegenüber der einfacheren Strategie bietet sie den Vorteil, dass – in den Grenzen der Einbildungskraft – psychologisch relevante Unterschiede berücksichtigt werden können. Man kann sich zum Beispiel vorstellen, andere Wünsche zu haben oder eine andere soziale Rolle auszufüllen, und unter diesen Prämissen bestimmte Entscheidungen nachvollziehen. Um griffigere Namen zu haben, werde ich die einfachere Strategie 'faktische Simulation' und die umfassendere Strategie 'hypothetische Simulation' nennen.

Schon sehr junge Kinder zeigen im spielerischen So-tun-als-ob (*Pretend Play*) die Fähigkeit, imaginative Kontexte zu schaffen. Gordon baut die Simulationsstrategie auf dieser Fähigkeit auf. Für das Beispiel von Entscheidungsprozessen beschreibt er den Ablauf so: Die simulierende Person tut so als sei sie selbst die andere Person und als sei sie in deren Situation. Auf Grundlage dieser hypothetisch angenommenen Prämissen ('Ich bin die Person X. Ich bin in der Situation S_X .) durchläuft sie dann den anstehenden Entscheidungsprozess für sich selbst.

[...] I might predict, for example, what I would do if, right now, the screen of the word processor I am working on were to go blank; or what I would do if I were now to hear footsteps coming from the basement. To simulate the appropriate practical reasoning I can engage in a kind of *pretend-play*: pretend that the indicated conditions *actually obtain*, with all other conditions remaining (so far as is logically possible and physically probable) as they presently stand; then – continuing the make-believe – try to 'make up my mind' what to do given these (modified) conditions. (Gordon 1986, 160, Hervorhebungen im Original)

Der Entscheidungsprozess wird durch das Hintergrundwissen der simulierenden Person und durch ihre sonstigen Präferenzen eingeschränkt. Oft wird in diesem Zusammenhang von *Default*-Einstellungen gesprochen: Sofern es über den Inhalt eines geistigen Zustands der anderen Person, der in einer Simulation relevant wird, keine Informationen gibt, setzen simulierende Menschen den Inhalt ihres eigenen entsprechenden Zustands ein. Solange man zum Beispiel nicht weiß, dass eine Person in den letzten Wochen keine Nachrichten gelesen oder gehört hat, geht man davon aus, dass sie über die wichtigsten Ereignisse informiert ist. Im Unterschied zum normalen Entscheiden wird das Ergebnis der Simulation jedoch nicht als eigene Entscheidung, sondern als *Vorhersage* der Entscheidung eines Anderen gewertet. Der Entscheidungsprozess läuft also ohne die übliche Kopplung an die handlungssteuernden Mechanismen ab; man sagt auch, er laufe *off-line*.

Das Verhältnis der beiden Simulationstypen

Die zusätzlichen Elemente der hypothetischen Simulation, von denen Stich & Nichols (1997) sprechen, sind zum einen die Fähigkeit, geeignete imaginative Kontexte zu schaffen, und die Forderung, dass die eigenen kognitiven Mechanismen in diesen Kontexten genauso ablaufen wie sonst auch. Es ist unklar, ob das *Pretend Play* diese Bedingungen erfüllt. Goldman zumindest ist skeptisch, und ich teile seine Ansicht.

Pretend beliefs and desires may differ in intrinsic as well as relational properties from natural, non-pretend beliefs and desires. But there must be important respects of resemblance if the simulation heuristic is to succeed in making accurate attributions and predictions. It is an open, empirical question, however, just what the points of similarity and dissimilarity may be. (Goldman 2000b, 186)

Im Zusammenhang mit Überzeugungen glaube ich jedoch, dass es unplausibel wäre zu behaupten, dass es zwei Typen von Schlussfolgerungsmechanismen gibt: einen für wahre und einen für falsche Prämissen. Stattdessen ist eher zu vermuten, dass die Schlussfolgerungsmechanismen beliebige Prämissen (ohne Ansehen ihrer Wahrheit) auf ihre Konsequenzen hin untersuchen. Über die jeweilige Einstellung zu den Prämissen, die für die eigentliche Schlussfolgerung nicht so wichtig ist, wird vermutlich separat Buch geführt. Diese Möglichkeit zieht auch Heal (1998) in Betracht – und grenzt es vom *Pretend-Play-Ansatz* Gordons ab:

[W]orking out relations of relevance, entailment, contradiction and the like between representations is one thing and done by one bit of mental machinery, while applying or withholding attitudes is a different matter, handled by some other part of the mental machinery. If this were so, then there would be no need for a person to generate a 'pretend' belief to input to her inference mechanism when she thinks about another's beliefs. [...] All she needs to do is input a content, perhaps tagging it as the content of the other's belief, note the output, and tag that too as the content of a probable further belief of the other." (Heal 1998, 494)

Im dritten Teil dieser Arbeit werde ich näher darauf eingehen, warum ich es für falsch halte, wie Heal zwischen dem *Pretend Play* und anderen Formen des kontrafaktischen Schließens zu unterscheiden. Es ist nicht klar, warum nicht auch *Pretend-Überzeugungen* von Heals zwei "mentalen Maschinen" verarbeitet werden sollten. Schließlich legt sich Gordon nicht auf eine kognitive Erklärung des *Pretend Play* fest. Wichtig ist an dieser Stelle, dass die Postulierung geeigneter kognitiver Mechanismen eine *Diskontinuität* zwischen beiden Simulationsstrategien darstellt, so dass faktische Simulationen nicht als trivialer Grenzfall hypothetischer Simulationen aufgefasst werden können. Hypothetische Simulationen beruhen auf der Annahme, dass sich kognitive

Mechanismen im Rahmen einer Simulation durch dieselben Faktoren auf dieselbe Weise beeinflussen lassen wie im normalen Funktionieren. Sollte diese Annahme falsch sein, gäbe es keine hypothetischen Simulationen. Die Zuverlässigkeit faktischer Simulationen hängt hingegen nur von der Ähnlichkeit der Anfangsbedingungen und der allgemeinen kognitiven Regularitäten ab.

Außerdem sind faktische Simulationen nicht auf das Ausnutzen der *eigenen* geistigen Ressourcen beschränkt; sie lassen sich auch auf das Denken und Handeln *Dritter* erweitern. Daraus wird ersichtlich, dass der Simulationsprozess bzw. dessen Resultat bei faktischen Simulationen *erst im Nachhinein* als Simulation aufgefasst werden muss. Hypothetische Simulationen werden jedoch als Simulationen angegangen und als Simulationen durchgeführt. Es ist daher klar, dass die Bestätigung faktischer Simulationen von keinem großen theoretischen Interesse sein kann. Ich werde mich daher im folgenden mit hypothetischen Simulationen befassen.

Die Drei-Schritt-Analyse des Simulationsprozesses

Um die folgende Diskussion zu erleichtern, möchte ich das Schema eines dreischrittigen Simulationsprozesses vorstellen. Für hypothetische Simulationen, die zu alltagspsychologischen *Vorhersagen von Entscheidungen* führen, sieht dieses Schema so aus:

1. Bestimmung der relevanten Merkmale der Situation, in der sich die andere Person befindet: *Ich stelle fest, dass die Situation der anderen Person durch die Merkmale S_1, \dots, S_n gekennzeichnet ist, und die Person sich anschickt, eine Entscheidung zu treffen.*
2. Übernahme dieser Merkmale als Randbedingungen eines eigenen kognitiven Prozesses, und Durchlaufen dieses Prozesses: *Ich tue so als seien die Merkmale S_1, \dots, S_n für mich gegeben, und versuche, auf dieser Grundlage eine eigene Entscheidung zu treffen. Dabei gelange ich zum Ergebnis E .*
3. Verwertung des Ergebnisses zu alltagspsychologischen Zwecken: *Ich stelle fest, dass ich mich für E entscheiden würde, wenn S_1, \dots, S_n für mich gegeben wären. Ich erwarte daher (oder sage explizit voraus), dass die andere Person ebenso reagieren wird.*

Nun noch das Schema einer hypothetische Simulation, die zur *Vorhersage einer Schlussfolgerung* führt:

1. Bestimmung der relevanten Merkmale der Situation, in der sich die andere Person befindet: *Ich stelle fest, dass die andere Person die Prämissen P_1, \dots, P_n glaubt und sich anschickt, vor diesem Hintergrund die Frage F zu beantworten.*
2. Übernahme dieser Merkmale als Randbedingungen eines eigenen kognitiven Prozesses, und Durchlaufen dieses Prozesses: *Ich tue so als würde ich die Prämissen P_1, \dots, P_n für wahr halten, und versuche, auf dieser Grundlage die Frage F zu beantworten. Dabei gelange ich zum Ergebnis E .*
3. Verwertung des Ergebnisses zu alltagspsychologischen Zwecken: *Ich stelle fest, dass ich mit E auf die Frage F antworten würde, wenn ich P_1, \dots, P_n glauben würde. Ich erwarte daher (oder sage explizit voraus), dass die andere Person ebenso reagieren wird.*

Erklärungen durch Simulation

Lässt sich der dritte Schritt auch für *Erklärungen* umformulieren? Die Standardantwort auf diese Frage fasst Erklärungen als Vorhersagen des gegenwärtigen Stands der Dinge aus vergangenen Situationen auf. Alltagspsychologische Erklärungen sind diesem Ansatz zufolge nachträgliche Vorhersagen. Um die richtige Erklärung zu finden, muss man unter Umständen in einer Reihe von Simulationen ausprobieren, welche Anfangsbedingungen zum richtigen Ergebnis führen. Dieser experimentelle Charakter von Simulationen wird von vielen Autoren betont.

Ein bekannter Einwand Churchlands (1988) betrifft jedoch gerade diesen Punkt: Simulationen können seiner Meinung nach kein Verständnis ("explanatory understanding") für psychologische Zusammenhänge wecken. Das werde bei Erklärungen aber erwartet. Alltagspsychologische Erklärungen dürften daher nicht in einzelnen nachträglichen Vorhersagen bestehen, da Vorhersagen durch Simulation kein solches Verständnis erzeugten.

Ich glaube, dass Churchlands Erwartungen an alltagspsychologische Erklärungen übertrieben sind. In den meisten Fällen genügt Menschen das Gefühl, mit ihrer Erklärung

richtig zu liegen. Dieses Gefühl kann auch dadurch erzeugt werden, dass der simulierte Gedankengang (oder zumindest dessen Ergebnis) plausibel und stimmig erscheint. Es reicht – mit anderen Worten – aus, einen erklärenden Umstand in der Geschichte der Person zu benennen, von dem man per Simulation zum *Explanandum* gelangt. Gordon (1992, 29) spricht von der "relativen Attraktivität" der eigenen Schlussfolgerungen und Entscheidungen, die sich auf die Vorhersage übertrage. Es scheint nicht nötig zu sein, dass Menschen in alltagspsychologischen Erklärungen an systematische Zusammenhänge appellieren.

4.2 Die Grenzen der Simulation

Von vielen Autoren ist kritisiert worden, dass der oben skizzierte Simulationsprozess nicht ganz auf die Anwendung alltagspsychologischer Theorien verzichten kann. Im dritten Schritt wird ein Schluss vom Ergebnis des simulierten kognitiven Prozesses auf die Vorhersage des Verhaltens der anderen Person vollzogen. Mit diesem möglichen Einfallstor für alltagspsychologische Theorien möchte ich mich zuerst beschäftigen; auf das Erfassen der Anfangsbedingungen S_1, \dots, S_n werde ich danach zu sprechen kommen. In der Diskussion wird es auch darum gehen, in welchen Fällen die Ähnlichkeitsbedingung als erfüllt betrachtet werden kann.

Die Zuschreibung des Simulationsergebnisses

Das Einfallstor im dritten Schritt ist dieses: Um den Schluss auf die andere Person ziehen zu können, reicht es nicht aus, im Rahmen einer Simulation zu einem Ergebnis zu kommen; dieses Ergebnis muss der anderen Person *zugeschrieben* werden. Dazu muss unter anderem zwischen der eigenen und der anderen Person unterschieden werden. Des Weiteren muss geprüft werden, ob der Schluss gerechtfertigt ist – ob zum Beispiel die Ähnlichkeitsforderung erfüllt ist. Schließlich muss das Simulationsergebnis typindividualisiert werden, um eine Zuschreibung der Art 'Person A ist im geistigen Zustand Z' zu ermöglichen. Die dazu nötigen Begriffe sind aller Voraussicht nach theoretische Begriffe; der Schluss auf die andere Person erfordert daher die Anwendung alltagspsychologischer Theorien.

Gordon und Goldman haben in zahlreichen Aufsätzen versucht zu zeigen, wie das Simulationsmodell diesem Einwand entgegen gehen kann.³⁵ Ich ziehe es jedoch vor, diesen Einwand gelten zu lassen. Denn meines Erachtens bietet die Grundidee des Simulierens selbst keine Antwort darauf; das Simulationsmodell müsste erweitert werden, um die Anwendung alltagspsychologischer Theorien im dritten Schritt auszuschließen. Das möchte ich kurz erläutern.

Ich habe das Simulationsmodell als Theorie des Zusammenspiels kognitiver Mechanismen dargestellt: Es postuliert einen übergeordneten Mechanismus, der andere Mechanismen ansteuert und mit Input versorgt (1. Simulationsschritt) und den Output dieser Mechanismen weiterverarbeitet (3. Simulationsschritt). Die Grundidee des Simulationsmodells ist, dass die Regularitäten des eigenen Geistes ausgenutzt werden, um alltagspsychologische Urteile hinreichend zu bestimmen. Sie betrifft jedoch nicht die Arbeitsweise des übergeordneten Mechanismus'. Die einzige Bedingung, die das Simulationsmodell diesem Mechanismus aufbürdet, ist, dass er die Arbeitsweise der von ihm angesteuerten Mechanismen nicht stört, um die Ähnlichkeit zwischen Simulationsprozess und simuliertem Originalprozess zu gewährleisten.

Ein Beispiel: Bei der Vorhersage einer Entscheidung kann das Simulationsmodell – dem o.g. Einwand zufolge – nicht auf die Anwendung der Begriffe Absicht, Handlung und Person verzichten, mit denen das Ergebnis des simulierten Entscheidungsprozesses erfasst und übertragen wird. Doch es ist eine Sache, alltagspsychologische Begriffe zu verwenden, um etwa die Motivation zu einer Handlung zu beschreiben, und eine andere Sache, Entscheidungsprozesse theoretisch zu erfassen. Auf letztere Theorie – in diesem Beispiel eine Entscheidungstheorie – kann das Simulationsmodell durchaus verzichten. Den dritten Schritt werde ich daher im folgenden vernachlässigen.

Das Erfassen psychologischer Unterschiede

Wie sieht es mit den möglicherweise notwendigen theoretischen Ressourcen im ersten Simulationsschritt aus? Beim Erfassen der Anfangsbedingungen zum Zweck der Simulation müssen unter Umständen *psychologische Unterschiede* zwischen der simulieren-

³⁵ Zuletzt in Goldman (2000) und Gordon (1995, 2000). Auch auf den Einwand, der den ersten Simulationsschritt betrifft, haben die beiden Autoren reagiert.

den und der simulierten Person festgestellt werden. Auch hier bietet sich ein Einfallstor für alltagspsychologische Theorien, da Begriffe benötigt werden, um diese Unterschiede feststellen zu können. Und auch hier möchte ich den Kritikern ihren Punkt zugestehen; einige theoretische Ressourcen im ersten Simulationsschritt disqualifizieren des Simulationsmodell ebenso wenig wie theoretische Ressourcen im oben diskutierten dritten Schritt.

In diesem Punkt stimme ich mit der Simulationstheoretikerin Heal (1995) überein. Ihr zufolge können Simulationen zwar mit dem eigenen Hintergrundwissen als *Default* für die Überzeugungen der anderen Person beginnen, müssen aber in Fällen psychologischer Unterschiede auf theoretische Ressourcen zurückgreifen:

At its simplest the initial premise could be just the view that the other thinks what I do. [...] Such an initial hypothesis cannot stand for long. It rapidly gets modified in various ways, in the light of views about what people can or cannot see from various angles, or are not likely to have learned and so forth. But these modifying strategies are not plausibly thought of on a simulationist model. On the contrary, they will involve the sorts of basic, but still theoretical, information about sense organs etc. [...] and may, indeed, involve quite complicated schemes and calculations. (Heal 1995, 49f.)

Wilkerson (2001) macht jedoch auf eine Gefahr aufmerksam, die mit diesem Zugeständnis an das Theorienmodell verbunden sein könnte: Es ist zu befürchten, dass mit dem Festlegen der Anfangsbedingungen einer Simulation schon ein großer Teil der alltagspsychologischen Leistung vollbracht ist. Wilkerson hält den ersten Simulationsschritt daher für wichtiger als den zweiten:

[B]efore either simulation processes or theoretical reasoning begins, a characterization of [the situation] must be found, and this characterization will determine, *far more than actual simulating or theorizing*, the prediction or explanation generated, because this explanation or prediction must base itself upon this characterization. (Wilkerson 2001, 149, meine Hervorhebung)

Wilkerson diskutiert als Beispiel den Fall, in dem er im Auto eines Freundes mitfährt, als ein Lämpchen auf dem Armaturenbrett aufleuchtet und einen drohenden Motorschaden anzeigt. Er sagt voraus, dass sich sein Freund darüber ärgern werde, da er beruflich auf sein Auto angewiesen ist. Das Simulationsmodell, so Wilkerson, könne nicht erklären, wieso er das Aufleuchten des Lämpchens für ein Ereignis halte, auf das sein Freund reagieren wird. Denn diese Charakterisierung der Anfangsbedingungen sei so gut wie jede andere. Um zur richtigen Verhaltensvorhersage zu gelangen, sei dieser Schritt aber schon der entscheidende.

In seiner Argumentation unterschätzt Wilkerson jedoch das Potenzial des Simulationsmodells. Es reicht für sein Beispiel aus, dem Blick des Freundes auf das Armaturenbrett zu folgen, um zu verstehen, dass sich ein ärgerliches Problem anbahnt. Denn bei ausreichendem Hintergrundwissen fasst man das Aufleuchten des Lämpchens selbst als ärgerliches Problem auf. Die Auswahl der psychologisch relevanten Charakteristika der Situation kann hier schon zum zweiten Simulationsschritt zählen. Wilkerson scheint dies für unmöglich zu halten:

[It is] absurd to suggest that I have somehow attributed a near infinite set of beliefs and desires to my friend, at each instant, in order to notice that her poorly running car will be important. (Wilkerson 2001, 146)

Doch hier hat er Unrecht: Es ist in Simulationen nicht nötig, der anderen Person die vielen Überzeugungen explizit zuzuschreiben, die während der Simulation aus dem eigenen Hintergrundwissen herangezogen werden. Beim Simulieren eines Gedankengangs müssen die Zwischenzustände ebenso wenig der anderen Person zugeschrieben werden wie man beim normalen Vollziehen eines Gedankengangs die Zwischenzustände sich selbst zuschreibt. Ich glaube daher, dass Wilkersons Beispiel keine Gefahr für das Simulationsmodell darstellt, sondern gerade eine seiner Stärken dokumentiert (nämlich das Bestimmen geistiger Zustände ohne Kenntnis der einschlägigen psychologischen Regularitäten).

Die Ähnlichkeitsbedingung

Schließlich möchte ich auf die Forderung eingehen, der Simulationsprozess müsse dem simulierten Originalprozess ähnlich sein. Ist es plausibel, diese Forderung im Normalfall als erfüllt anzusehen? Heal (2000) diskutiert den Fall, in dem man den nächsten Zug eines Schachspielers vorhersagen soll, der deutlich schlechter spielt als man selbst. Da man sich nicht dümmer stellen könne als man sei, sagt Heal, müsse diesem Kompetenzunterschied *im Nachhinein* Rechnung getragen werden. Es hat den Anschein, als lasse Heal zu diesem Zweck auch die Anwendung einer alltagspsychologischen Theorie zu.

[M]y strategy is to use my capacity on the chess situation as presented to generate a rich appreciation of the options available and then to cull from this range those which I judge to be too advanced, leaving only some limited range of obviously attractive moves as ones the inexperienced player might choose. (Heal 2000, 17)

Die unterschiedlichen Schachspieler nimmt Heal als Beispiel, um zu zeigen, dass ein Analogieschluss, der sich auf die Ähnlichkeit der kognitiven Fähigkeiten stützt, oft nicht möglich ist. Sie beschreibt eine solche Simulation wie folgt:

I am to perform a kind of experiment on myself in which I set up my own mental state to be as like that of the other as I can manage and then engage in thinking. It is the upshot of this which entitles me to the judgement 'If I were less experienced and faced with this chess position then I would come to believe so and so.' But this story about what goes on does not make sense in the case in question. With the inexperienced chess player, in order to recreate his thinking with this kind of closeness what I would need to do is shut down part of my mind which generates and intelligently appraises the strategies which he is not yet capable of generating and appraising. But how am I thus to render myself temporarily uninventive or uncritical? Shutting down capacities is not something we can do at will or by engaging in some imaginative enterprise [...]. (Heal 2000, 17)

Heal sieht demnach nur eine geringe Reichweite der hypothetischen Simulation; sie bestreitet für viele interessante Fälle, dass es möglich sei, so zu tun als würden die psychologischen Anfangsbedingungen für einen selbst gelten. Denn es sei nicht möglich, als Bedingung für eine Simulation zu berücksichtigen, dass bestimmte kognitive Mechanismen einzusetzen und andere Mechanismen auszuschalten seien.

Ich halte Heals Punkt für plausibel. Es kann gut sein, dass das Simulieren nur in ausgewählten Bereichen der Alltagspsychologie erfolgreich ist. Nicht nur im Schach-Beispiel ist es möglich, dass Menschen eine alltagspsychologische Theorie anwenden (die beispielsweise Auskunft darüber gibt, welche Züge der Schachfiguren zu welchem Kompetenzniveau passen). Auch im Zusammenhang mit dem Lebenspartner (z.B.: "Könnte ich sie eifersüchtig machen, würde ihr wieder bewusst, was sie an mir hat."), dem Vorgesetzten ("Ich muss auf eine ruhige Minute warten, um ihm meinen Vorschlag zu unterbreiten.") oder dem Volk ("Mit einer Personaldebatte könnte ich von meiner inhaltlichen Orientierungslosigkeit ablenken.") spielen alltagspsychologische Theorien vermutlich eine wichtige Rolle. Meiner Meinung nach kommt Simulation vor allem dann zum Tragen, wenn es um Überzeugungen geht. Dieser Punkt ist Thema des nächsten Kapitels.

Ein Problem wird allerdings in der folgenden Diskussion die Frage darstellen, ob das kontrafaktische Schließen dem normalen Schlussfolgern hinreichend ähnlich ist. Da zu diesem Punkt aber nicht genügend Daten vorliegen, muss ich diese Frage unbeantwortet lassen. Das Simulationsmodell, mit dem ich im dritten Teil den Umgang mit falschen Überzeugungen erkläre, könnte daher daran scheitern, dass empirische Studien zeigen, dass die Ähnlichkeitsbedingung nicht erfüllt ist.

5 Simulation als "Mitdenken"

Von Heals Variante des Simulationsmodells habe ich bereits einiges vorgestellt. Ich möchte nun im Detail auf ihr Argument gegen das Theorienmodell (und *für* ihr Simulationsmodell) zu sprechen kommen. Sie entwickelt es am Beispiel der Vorhersagen von Schlussfolgerungen Anderer. Da bei Schlussfolgerungen oft Hintergrundwissen herangezogen wird, muss dieses Hintergrundwissen auch in Vorhersagen von Schlussfolgerungen einfließen. Doch das kann Heal zufolge nur im Rahmen einer Simulation geschehen; alltagspsychologische Theorien, die diesen komplexen Prozess beschreiben und damit zu solchen Vorhersagen befähigen, seien nicht möglich.

Heal bezeichnet ihr Argument als a priori; sie sei keine Kognitionswissenschaftlerin.³⁶ Doch ich werde versuchen zu zeigen, dass ihr Argument auf einer verdeckten und ungedeckten empirischen Annahme beruht, von der bislang nicht klar ist, wie sie gestützt werden soll. Heal traut der Alltagspsychologie meines Erachtens zu viel zu. Ich werde im zweiten Abschnitt, nachdem ich im ersten ihr Argument vorgestellt habe, die Frage aufwerfen, wie gut Menschen die Schlussfolgerungen Anderer vorhersagen können. Von der Antwort auf diese Frage hängt letztlich ab, was Theorien- und Simulationsmodell erklären müssen – und ob die Leistungen tatsächlich so komplex sind, dass sie nicht theoretisch erfasst werden können. Eine systematische Taxierung der Qualität alltagspsychologischer Leistungen gibt es meines Wissens bislang nicht (bis auf einen eng umgrenzten Bereich: das kindliche Verständnis falscher Überzeugungen).

Im dritten Teil werde ich diesen Faden wieder aufnehmen. Ich werde dort von einer schwachen Variante von Heals Argument Gebrauch machen: Ich werde dafür argumentieren, dass das theoretische Erfassen von Schlussfolgerungen unnötig kompliziert wäre im Vergleich zur Simulation. Dies werde ich im Übrigen nicht auf die Vorhersage von Schlussfolgerungen beziehen, sondern auf das *Pretend Play*, bei dem ebenfalls unter hypothetischem Vorzeichen Schlüsse gezogen werden.

³⁶ Als a priori bezeichnet Heal Überzeugungen, die im Zentrum unseres Überzeugungssystems stehen, und daher nur schwer zu falsifizieren sind (siehe Heal 1998, 480). Sie argumentiert also nicht rein logisch oder begrifflich, sondern appelliert an Intuitionen, die wir nicht bereit wären aufzugeben.

5.1 Ein Argument gegen das Theorienmodell

Heal macht die plausible Annahme, dass Schlussfolgerungen wesentlich von den propositionalen Gehalten der geistigen Zustände abhängen, und dass alltagspsychologische Vorhersagen von Schlussfolgerungen dies berücksichtigen. Es reiche für alltagspsychologische Zwecke nicht aus, allein die logische Form der Wünsche und Überzeugungen zu betrachten. Diese Prämisse möchte ich anhand einiger Beispiele erläutern:

- (1) X liegt links von Y. \Rightarrow Y liegt rechts von X.
- (2) A ist als erste im Ziel. \Rightarrow A ist Siegerin.
- (3) Über dem Wald steigt Rauch auf. \Rightarrow Im Wald brennt ein Feuer.
- (4) Die Lunte wird angezündet. \Rightarrow Gleich explodiert die Bombe.
- (5) A betritt seine Wohnung. \Rightarrow A legt seinen Mantel ab.

Diese Beispiele stellen Schlussfolgerungen unterschiedlichen Typs dar, die sich nicht allein über die logische Form erklären lassen. Auf welche nicht-logischen Aspekte hebt Heal nun ab? Welche Aspekte dieser Beispiele spielen eine besondere Rolle in der alltagspsychologischen Vorhersage von Schlussfolgerungen und stellen daher *Explananda* für Theorien- und Simulationsmodell dar? Heal selbst verwendet in einem frühen Aufsatz ein irreführendes Beispiel:

[S]uppose I wish to predict what John will think of the new jacket; will he think it garish? Suppose further that I know that John believes the jacket to be scarlet and that he thinks all bright colours to be garish. I will, of course, expect him to think the jacket garish. [...] [That prediction] needs [...] to contain information about possession of the particular concepts 'scarlet' and 'bright colour', namely that anyone who possesses them (probably) knows that scarlet is a bright colour. (Heal 1995, 39)

In diesem Beispiel geht es um die Bedeutung nicht-psychologischer Begriffe wie *Scharlachrot* – und zwar um zentrale Aspekte ihrer Bedeutung. Die Vorhersage – gegeben das Wissen um Johns Meinung zu grellen Farben – hängt daher davon ab, ob die vorhersagende Person weiß, dass Scharlachrot (bei Kleidungsstücken) eine grelle Farbe ist. Möglicherweise zählt dieses Wissen aber schon zur Minimalkompetenz des Begriffsbesitzes. Die oben aufgeführten Beispiele (1) und (2) sind vom selben Typ: Menschen, die diese Schlüsse nicht ziehen können, verfügen vermutlich nicht über die Begriffe *Links-von* bzw. *Siegerin*.

Diese Interpretation ihres Beispiels kann Heal nicht beabsichtigt haben. Denn in diesen Fällen hängt die Fähigkeit eines Menschen, die Schlussfolgerung vorherzusagen, in erster Linie davon ab, ob er das Vokabular der Prämissen versteht. Diese Fähigkeit liegt

meines Erachtens außerhalb der Alltagspsychologie und muss daher auch nicht erklärt werden.

Holismus epistemischer Verknüpfungen

In einem späteren Aufsatz gibt Heal das folgende Beispiel eines Arztes, der einen königlichen Patienten mit Masern untersucht. An diesem Beispiel möchte ich ihr Argument erläutern.

[S]uppose I am a doctor investigating whether a patient has measles. I have a great deal of information about symptoms and the results of physiological tests. These are obviously relevant. But it may seem equally obvious that the further information I possess that Henry VII of England was a Tudor could not be relevant. However, we need to insert only a few further beliefs (not outrageously bizarre) into my belief set to put a connection in place. Perhaps I believe that measles has an unusual variant in which the presence of a certain gene, common in the Tudor family, leads to its running a non-standard course. And perhaps my patient has boasted of his royal ancestry. (Heal 1996a, 79f.)

An diesem Beispiel wird deutlich, dass es um Verknüpfungen geht, die empirisches Wissen repräsentieren. Heal charakterisiert sie als Wissen über *kausale* und *evidenzielle* Zusammenhänge. Ein Beispiel für einen evidenziellen Zusammenhang habe ich mit (3) gegeben; (4) betrifft einen kausalen Zusammenhang. Ebenso wird die Stoßrichtung Heals deutlich: Die Verknüpfungen können auch zwischen Überzeugungsgehalten aus sehr entfernten Wissensbereichen bestehen. Beim Schluss des Arztes auf eine Diagnose können im Prinzip Überzeugungen eine Rolle spielen, die zu einem davon so abgelegenen Wissensbereich wie der Dynastie der Tudors gehören. Mit Beobachtungen dieser Art begründet Heal ihre Ansicht, dass die evidenziellen und kausalen (zusammenfassend: *epistemischen*) Verknüpfungen ein *holistisches Netz* ergeben.

[T]he central factor underlying [epistemological holism] is the potential complex interconnectedness of things, *both causally and evidentially*. The world may present itself to us much of the time as more or less isolated subsystems, further features of which can be inferred on the basis of information about only the current or preceding states of that subsystem. But each such subsystem is, we believe, embedded in a wider spatio-temporal framework which may *impinge on it or provide clues about it*. (Heal 1996a, 80, meine Hervorhebungen)

Es wirken also zwei Faktoren zusammen, um diesen Holismus zu erzeugen: Einerseits ist die Welt komplex und voller Überraschungen, andererseits versuchen wir, unser empirisches Wissen in eine kohärente Weltsicht zu integrieren. Wenn wir Schlüsse ziehen, wirkt sich dieser Holismus dahingehend aus, dass eine Vielzahl von Überzeugungsge-

halten potenziell relevant wird: diejenigen, die mit den Prämissen epistemisch verknüpft sind. Welche zusätzlichen Informationen tatsächlich aus dem Gedächtnis abgerufen werden müssen, hängt natürlich von der Frage ab, die es zu beantworten gilt.

Vor einem Missverständnis möchte ich noch warnen: Es geht nicht darum, Schlüsse vorherzusagen, die eine Person nur ziehen kann, weil sie über ein besonderes Wissen verfügt. Es muss um empirisches Wissen gehen, das als allgemein bekannt vorausgesetzt werden kann. Der Schluss aus Beispiel (5) könnte zwar dadurch gerechtfertigt sein, dass A regelmäßig seinen Mantel ablegt, wenn er seine Wohnung betritt. Doch die Kenntnis einer solchen Gewohnheit wird in der alltagspsychologischen Praxis nur berücksichtigt, wenn sie *explizit* als Prämisse in die Schlussfolgerung eingeht. Das Heranziehen idiosynkratischen Wissens zählt nicht zu den *Explananda* von Theorien- und Simulationsmodell.

Theorien über epistemische Verknüpfungen

Wie erklären Theorien- und Simulationsmodell die Berücksichtigung des relevanten Hintergrundwissens in der Vorhersage von Schlussfolgerungen? Hier sieht Heal einen klaren Vorteil des Simulationsmodells. Im Rahmen von Simulationen würden Menschen *direkt* über die Prämissen und die Fragestellung nachdenken und dabei das Hintergrundwissen heranziehen, das sie auch heranzögen, wenn sie die Antwort für sich selbst suchten. Um diesen Punkt hervorzuheben, rückt Heal von der Bezeichnung 'Simulation' ab und spricht davon, dass man mit der anderen Person "mitdenken" müsse (*co-cognize with the other*).

Co-cognition is just a fancy name for the everyday notion of thinking about the same subject matter [...]. Those who co-cognize exercise the same underlying multifaceted ability to deal with some subject matter. (Heal 1998, 483)

Nach dem Theorienmodell müssen hingegen alltagspsychologische Theorien Anweisungen darüber geben, wie aus bestimmten Überzeugungen andere Überzeugungen abgeleitet werden können; das Theorienmodell ist Heal zufolge dazu verpflichtet, eine alltagspsychologische Theorie des normalen Schlussfolgerns zu postulieren. Wenn aber in Schlüssen epistemische Verknüpfungen berücksichtigt werden müssen, werde diese

Theorie auch Anweisungen darüber geben müssen, wie das für eine Fragestellung relevante Hintergrundwissen bestimmt wird.³⁷

For strong theory-theory to be defensible we need to make plausible the idea that we actually grasp (even if only tacitly) some principles about how particular contentful thoughts are connected, principles which have nothing to do with connections between the states of affairs these thoughts are about [...]. (Heal 1998, 487)

Heals Punkt lässt sich so erläutern: Dem Theorienmodell zufolge wenden Menschen bei der Vorhersage von Schlussfolgerungen alltagspsychologische Theorien über Schlussfolgerungen an – und damit auch Theorien über die Verknüpfung von Überzeugungsgehalten. Bei der Vorhersage von Schlussfolgerungen im Sinn des Theorienmodells ist es nicht möglich, auf die epistemischen Verknüpfungen zwischen den Elementen des eigenen Hintergrundwissens zurückzugreifen. Denn man denkt über die *Prämissen* Anderer nach und nicht direkt über den *Gehalt* dieser Überzeugungen. Daher ist das Theorienmodell dazu verpflichtet, eine alltagspsychologische Theorie zu postulieren, die das holistische Netz kausaler und evidenzieller Verknüpfungen von Überzeugungsgehalten erfasst. Eine solche Theorie zu postulieren, ist aber unplausibel.

Heal sieht eine ähnlich schwierige Situation für das Theorienmodell wie für die Forschung zur künstlichen Intelligenz (kurz: KI), die das so genannte Rahmenproblem (*frame problem*) kennt. Ein Ziel der KI-Forschung ist es, die intelligente Verwendung empirischen Wissens bei der Lösung von Problemen zu erklären. Um ein Beispiel aus Dennett (1984) zu nennen: Es geht darum Mechanismen zu finden, die nicht den Fehler begehen, ein Streichholz anzuzünden, um zu schauen, ob noch Benzin im Tank ist. Innerhalb akzeptabler Zeiträume müssen dafür nicht alle, aber doch die relevanten Konsequenzen der Handlungsalternative Streichholz-anzünden überprüft werden. Eine entsprechende Theorie über diese Mechanismen fehlt bis heute.

The difficulty [Artificial Intelligence] researchers have run into is that of finding a format for coding knowledge and questions, and a way in which a machine can process its knowledge in the light of its questions, so as to enable it to come up with the required answer rather than coming up with some one of the vast number of other true but irrelevant claims which its knowledge base and inference rules license. (Heal 1996a, 81)

³⁷ Heal betrachtet eine Variante des Theorienmodells, die sie (beispielsweise im folgenden Zitat) "stark" nennt. Das soll nicht bedeuten, dass das Modell abseitig wäre, sondern dass es sich dabei um den Kern des Theorienmodells handelt.

Das Rahmenproblem ist laut Heal aber gerade das Problem, das das Theorienmodell lösen müsste: eine alltagspsychologische Theorie zu postulieren, die angibt, welche Elemente des Hintergrundwissens zur Lösung der gestellten Frage relevant sind. Heals Argument schließt mit der rhetorischen Frage: Wenn es der KI-Forschung noch nicht gelungen ist, dieses Problem zu lösen, welchen Grund haben wir dann anzunehmen, dass fast alle Menschen über eine implizite alltagspsychologische Theorie verfügen, die dieses Problem löst?

Das Hintergrundwissen heranziehen

Stich und Nichols, zwei Vertreter des Theorienmodells, gegen die sich Heal ausdrücklich wendet, haben kritisiert, dass Heal ein übertrieben starkes Theorienmodell angreife. Heals Punkt, das Theorienmodell müsse theoretischen Ersatz für die epistemischen Verknüpfungen bieten, die im normalen Denken berücksichtigt werden, treffe nicht den Kern des Theorienmodells.

[W]e have continued to be quite explicit that on our view people's understanding of other minds is not independent of their own beliefs and inferences. [...] We even claimed that it's difficult to elaborate any 'theory-theory' story about mindreading that doesn't involve tapping into one's own beliefs about various subject matters. [...] If we are the most likely suspects for being strong theory-theorists, then we doubt that there are any strong theory-theorists out there. (Nichols & Stich 1998, 510)

Die Formulierung 'tapping into one's belief systems' ist meines Erachtens zu unspezifisch, um als Einwand gegen Heal zu taugen. Sie kann in einem schwachen Sinn verwendet werden, der nichts weiter bedeutet als: Hintergrundwissen in die Vorhersage einfließen lassen. Es wäre tatsächlich übertrieben, dem Theorienmodell die Behauptung unterzuschieben, Menschen würden bei alltagspsychologischen Vorhersagen ihr Hintergrundwissen beiseite lassen. Zum Beispiel müssen sie die alltagspsychologischen Theorien, die für eine bestimmte Vorhersage geeignet sind, aus ihrem Hintergrundwissen heranziehen. Und sie müssen die Prämissen der Schlussfolgerung verstehen (was z.B. das Wissen, dass scharlachrote Kleidungsstücke grell sind, einschließt). Das Theorienmodell ist sicher nicht verpflichtet, diese kognitiven Leistungen zu erklären.

Doch das ist nicht der Punkt. Heal geht es vielmehr darum, dass auch solche Elemente des Hintergrundwissens herangezogen werden müssen, die nicht nahe liegen. Für diese Fälle, behauptet sie, müsse das Theorienmodell eine geeignete Relevanztheorie postulieren. Stütze man sich aber bei der Erklärung der Vorhersage von Schlussfolgerungen

darauf, dass die vorhersagende Person das Hintergrundwissen heranzieht, das sie selbst für relevant hält, formuliere man ein Simulationsmodell.

5.2 Reichweite und Tiefe der Alltagspsychologie

Obwohl ich Heals Prämissen akzeptiere, zweifle ich an ihrer Konklusion. Denn ich glaube, dass noch eine weitere verdeckte Prämisse im Spiel ist, die die Leistungsfähigkeit der Alltagspsychologie betrifft, und die sich nicht leicht begründen lässt.

Eine verdeckte Prämisse

Heal stellt meines Erachtens zu hohe Ansprüche an unsere alltagspsychologische Fähigkeiten. Ich halte es hingegen für möglich, dass alltagspsychologische Vorhersagen weniger differenziert sind als das normale Denken selbst. Heal argumentiert zwar für einen Holismus der epistemischen Verknüpfungen zwischen Überzeugungsgehalten, nicht aber dafür, dass wir alle diese Verknüpfungen in alltagspsychologischen Vorhersagen auch berücksichtigen. Heal scheint die Möglichkeit, dass sich ein Schlussfolgerungs- und ein Vorhersageprozess nicht decken, für absurd zu halten:

Thinking about thinking cannot be easier or simpler than the first-level thinking itself. If it were, then, absurdly, I could short cut my own laborious intellectual endeavours. I could simply feed into my psychological theory the specification of my own initial information and problem [...] and use the theory to predict the solution I shall come up with. (Heal 1995, 36)

Doch es ist meines Erachtens durchaus möglich, seine eigenen, unmittelbar anstehenden Schlussfolgerungen und Entscheidungen vorherzusagen – und sich dadurch einige Mühe zu ersparen. Der Grund dafür ist, dass Menschen auf verschiedenen Wegen zu ihren Entscheidungen kommen können, und einige dieser Entscheidungsprozesse recht oberflächlich sind. Zum Beispiel kann ein Sachbearbeiter Anträge auf der Grundlage einiger weniger Kriterien ablehnen, ohne sich ausführlich mit der Materie zu beschäftigen. Es könnte sein, dass er aus seinen bisherigen Entscheidungen die Theorie abstrahiert hat, dass es ihm nur auf eben diese Kriterien ankommt. Es wäre aber auch möglich, dass er einfach aus Bequemlichkeit eine gründliche Entscheidung durch eine oberflächliche simuliert.

Der Vorteil dieser Abkürzung liegt auf der Hand, der Preis ebenso: Die Qualität der Vorhersage (und in diesem Fall auch der Entscheidung) dürfte in der Regel leiden. Es ist daher zu vermuten, dass solche Abkürzungen selten gewählt werden, da die eigenen Entscheidungen (im Unterschied zu Vorhersagen von Entscheidungen Anderer) zu wichtig sind. Ich halte es dennoch für möglich, dass Heal ein *Explanandum* für Theorien- und Simulationsmodelle formuliert, das nicht mit der alltagspsychologischen Realität übereinstimmt. Denn der Punkt gilt allgemein: Alltagspsychologische Vorhersagen und Erklärungen sind oft nicht so wichtig, als dass man ihren Erfolg sicher stellen müsste. Und bei Erklärungen genügt es oft, eine plausible zu finden. Daher könnte eine genaue (bislang noch nicht durchgeführte) Untersuchung ergeben, dass alltagspsychologische Urteile oft in die Irre gehen. Wenn aber im Fall der Vorhersage von Schlussfolgerungen nicht alles relevante Hintergrundwissen herangezogen werden muss, sondern möglicherweise nur die nahe liegenden Elemente daraus, dann bietet sich dem Theorienmodell eine Möglichkeit, Heals rhetorische Frage aus dem letzten Abschnitt zu beantworten.

Reaktion des Theorienmodells

Wenn Heals verdeckte Prämisse, alltagspsychologische Vorhersagen von Schlüssen seien perfekt, geschwächt wird, ergibt sich eine Möglichkeit für das Theorienmodell, Plausibilität zurückzugewinnen. Der Ansatz dafür ist dieser: Vielleicht ist sind alltagspsychologische Vorhersagen auf das Hintergrundwissen desjenigen Wissensbereichs beschränkt, der durch die Prämissen vorgegeben wird. Da das Wissen innerhalb eines Wissensbereichs stark systematisiert ist (vielleicht sogar durch eine Alltagstheorie dieses Bereichs), dürfte es leichter sein, die relevanten Elemente des Hintergrundwissens auszumachen.

Ein Beispiel ist der Fall des Tudor-Patienten mit Masern. Es ist durchaus denkbar, dass das spezifische Gen der Tudor-Familienmitglieder Teil der medizinischen Theorie über Masern wäre, wenn bekannt wäre, dass es den Krankheitsverlauf auf eine typische Weise beeinflusst. In einer Theorie sind die darin vorkommenden theoretischen Konstrukte jedoch explizit miteinander verknüpft, so dass es keiner umfangreichen alltagspsychologischen Theorie bedarf, um die relevanten Überzeugungen zu bestimmen. Wäre die Rolle dieses Gens bekannt, müssten Ärzte bei der Masern-Diagnose das Vorhandensein

dieses Gens prüfen. Müssten sie die Diagnose eines Kollegen *vorhersagen*, würden sie ihr Hintergrundwissen zu Hilfe nehmen, soweit es die Diagnose von Masern betrifft. Die alltagspsychologische Theorie, die diesen Schritt vermittelt, müsste lediglich lauten: 'Berücksichtige bei der Vorhersage der Ausübung standardisierter Verfahren dein Wissen der Verfahrensstandards.'

Doch in dem Beispiel hält der diagnostizierende Arzt nicht das Vorhandensein des Gens, sondern die Familienzugehörigkeit des Patienten für relevant. Wie kommt diese Verknüpfung zustande? Hier gibt es zwei Möglichkeiten gibt: Entweder ist es ein fester Bestandteil der Theorie, dass es sich bei dem speziellen Gen um ein Gen der Tudor-Familie handelt. In diesem Fall würden diagnostizierende wie vorhersagende Ärzte in der Diagnose nach der Familienzugehörigkeit des Patienten fragen – der Fall wäre dem oben geschilderten ähnlich.

Oder es handelt sich um zusätzliches, nicht-medizinisches Wissen, demzufolge das spezifische Gen zufällig und nur *unter anderem* in der Tudor-Familie häufig auftritt. Aber auch in diesem Fall wären diagnostizierende wie vorhersagende Ärzte in einer vergleichbaren Situation: Beide müssten erst darauf kommen, dass die Familienzugehörigkeit für die Diagnose relevant ist. Das geschieht für beide Ärzte durch das Befolgen der Regel, alle verfügbaren Informationen auf ihre Relevanz hin zu überprüfen. Sollte der vorhersagende Arzt aber nicht auf den Zusammenhang kommen, dann wäre das kein Problem des Theorienmodells, sondern vielmehr eins der Alltagspsychologie: Die Vorhersagefähigkeit wäre in diesem Fall, der einiges an Kreativität bei der Diagnose erfordert, an ihre Grenzen gelangt.

Was wird simuliert?

Um nicht den Eindruck zu erwecken, Simulationen seien immer perfekt, möchte ich noch eine Bemerkung anschließen. Es gibt, wie bereits erwähnt, oft mehrere Strategien, ein Problem zu lösen. Man kann z.B. bei Entscheidungen die Alternativen im Detail bewerten; man kann sich aber auch nach einigen wenigen Kriterien schnell entscheiden. Wenn es aber mehrere Strategien gibt, ein Problem zu lösen, dann gibt es mehrere Strategien, die simuliert werden können. Das Simulationsmodell könnte also behaupten, dass gelegentlich andere Mechanismen zur Simulation verwendet werden als bei der simulierten Person, um zu demselben Ergebnis (etwa einer Entscheidung) zu gelangen.

Sollte dabei die Ähnlichkeitsbedingung nicht erfüllt sein, wäre die resultierende Vorhersage falsch.

Ich halte daher fest, dass Heals apriorisches Argument nicht ohne weiteres funktioniert. Solange unklar ist, wie gut Schlussfolgerungen im Alltag vorhergesagt werden, bleibt auch offen, was genau Theorien- und Simulationsmodell eigentlich erklären müssen. Folgt man meiner Vermutung, dass der Alltagspsychologie in diesem Bereich Grenzen gesetzt sind, haben sowohl Vertreter des Theorien- als auch des Simulationsmodell einen gewissen Spielraum beim Ausformulieren ihrer genauen Erklärung.

Ich halte aber auch fest, dass Heals einen wichtigen Vorteil des Simulationsmodells herausarbeitet: Selbst wenn es *möglich* wäre, alltagspsychologische Theorien zu postulieren, mit deren Hilfe relevantes Hintergrundwissen bestimmt werden kann, wäre die Erklärung des Simulationsmodells immer noch einfacher und damit *plausibler*. Es wird aber noch zu klären sein, ob das auch für Vorhersagen von Schlussfolgerungen gilt, die von *falschen* Prämissen ausgehen. Auf den ersten Blick scheint dies zu Widersprüchen mit dem Hintergrundwissen zu führen. Über die Maßnahmen, die diese Widersprüche verhindern sollen, werde ich im dritten Teil zu sprechen kommen.

6 Versuche einer empirischen Unterscheidung

In diesem Kapitel komme ich zu einer Reihe von Experimenten, in denen Theorien- und Simulationsmodell direkt miteinander verglichen wurden. Im ersten Teil der Arbeit habe ich bereits erläutert, dass es in solchen Studien um die Unterscheidung von zwei Prozesstypen geht: Werden alltagspsychologische Theorien angewendet oder läuft ein kognitiver Prozess im Rahmen einer Simulation ab? In der Forschungsliteratur sind einige Merkmale diskutiert worden, die spezifisch für einen der beiden Prozesstypen sind. In diesem Kapitel werde ich untersuchen, ob es gelungen ist, diese Merkmale nachzuweisen.

In den Experimenten, die ich vorstellen werde, wird stets ein Faktor manipuliert, der sich – je nach Erklärungsmodell – unterschiedlich auf die Qualität alltagspsychologischer Urteile auswirken sollte. Wird die alltagspsychologische Vorhersage durch die Manipulation erleichtert, spricht das für eines der Modelle, wird sie durch die Manipulation erschwert, spricht das für das andere. Diese Hypothesen werde ich für jede Studie zurückweisen. Obwohl ich der Analyse zustimme, dass die ausgewählten Merkmale jeweils spezifisch für die Theorieanwendung oder die Simulation sind, werde ich dagegen argumentieren, dass sich daraus für konkrete Versuchsbedingungen Aussagen über die Qualität der alltagspsychologischen Urteile ableiten lassen.

6.1 Das Kriterium der kognitiven Durchdringbarkeit

Das erste Merkmal zur Unterscheidung von Theorien- und Simulationsmodell wurde von Stich & Nichols (1992) vorgeschlagen und erhielt von ihnen auch einen Namen: das Kriterium der *kognitiven Durchdringbarkeit*.

Falsche alltagspsychologische Theorien

Als 'kognitiv durchdringbar' bezeichnen Stich und Nichols kognitive Prozesse, deren Verlauf von der Kenntnis alltagspsychologischer Theorien abhängt. Nach dem Theorienmodell hängen alltagspsychologische Vorhersagen und Erklärungen klarerweise von

der Kenntnis der entsprechenden Theorien ab. Menschen, die über bessere Theorien verfügen, werden zu besseren Vorhersagen und Erklärungen imstande sein. Nach dem Simulationsmodell sollten diese kognitiven Prozesse jedoch von der Kenntnis alltagspsychologischer Theorien unberührt bleiben; Simulationsprozesse sind daher nicht kognitiv durchdringbar (in Stich & Nichols Sinn). Ihre Vorhersagequalität hängt stattdessen davon ab, ob es sich bei dem simulierten kognitiven Mechanismus um einen Mechanismus handelt, der bei allen Menschen auf die gleiche Weise abläuft. Diese Analyse verknüpfen Stich & Nichols wie folgt mit einem empirischen Indikator:

If there is some quirk in the human decision making system, something quite unknown to most people that leads the system to behave in an unexpected way under certain circumstances, the accuracy of predictions based on simulations should not be adversely affected. [...] Just the opposite is true for predictions that rely on a theory. [...] Theory based predictions are sensitive to what we know and don't know about the laws that govern the system [...]. (Stich & Nichols 1992, 67)

Stich & Nichols gehen im Sinn des Theorienmodells davon aus, dass unser Repertoire alltagspsychologischer Theorien nicht alle psychologischen Regularitäten abdeckt und die alltagspsychologische Praxis daher in einigen Fällen zu falschen Vorhersagen und Erklärungen führen müsste. Das dürfte für solche Fälle zu erwarten sein, die im Alltag selten vorkommen bzw. deren Vorhersage und Erklärung im Alltag irrelevant sind. Wird ein solcher Fall gefunden, in dem zugleich ein kognitiver Mechanismus zum Tragen kommt, der bei allen Menschen gleich abläuft, ließe sich prüfen, ob Vorhersagen typischer Reaktionen über die Anwendung einer Theorie oder über eine Simulation erreicht werden: Ist die Vorhersage falsch, lag es an der fehlenden oder falschen Theorie; ist die Vorhersage richtig, lag es an der Ähnlichkeit des simulierenden und simulierten kognitiven Mechanismus'.

Beispiele für solche Fälle gibt es viele; man denke nur an die kontraintuitiven Ergebnisse der psychologischen Forschung. Allein die Tatsache, dass es sich um kontraintuitive Regularitäten handelt, zeigt an, dass wir dieses Verhalten nicht vorhersagen würden (und deshalb das Theorienmodell dem Simulationsmodell vorzuziehen ist). Stich & Nichols untersuchen zu diesem Zweck den so genannten Langer-Effekt. Dieser Effekt betrifft den Wert, den Versuchspersonen einem Los zuschreiben: Wenn sie sich das Los aussuchen dürfen, besitzt es für sie einen höheren Wert als wenn sie es zufällig zugeteilt bekommen. Das zeigt sich darin, dass sie ihr selbst ausgewähltes Los nur zu einem übertrieben hohen Preis zurückverkaufen. Dieses Verhalten wird von Versuchspersonen nicht vorhergesagt (Nichols et al. 1996, 51f.). Kühberger et al. (1995) bezweifeln zwar

die Reliabilität dieses Effekts, doch es dürfte nicht schwierig sein, andere Beispiele zu finden, in denen typische Effekte nicht vorhergesagt werden.

Die Fehlbarkeit der Simulation

Der Ansatz von Stich & Nichols wurde von mehreren Autoren kritisiert. Es wurde zunächst bezweifelt, dass die Versuchsanordnung eine angemessene Simulation zulässt, da sich die Situation den Versuchspersonen, die den Preis vorhersagen sollen, anders darstelle, als denen, die ihn für ihr eigenes Los festlegen. Nichols et al. (1996) tragen dieser Kritik dadurch Rechnung, dass sie den vorhersagenden Versuchspersonen ein Video zeigten und alle Materialien vorlegten, die die Situation der Personen beschreibt, deren Verhalten es vorherzusagen gilt.

Doch es wurde weiter Kritik geübt von Heal (1996b). Ihr zufolge sind Vertreter des Simulationsmodells auch bei Erfüllung der Ähnlichkeitsbedingung nicht zu perfekten Vorhersageergebnissen verpflichtet. Auch wenn der Simulationsprozess dem simulierten Prozess ähnlich ist, kann er anders in den kognitiven Haushalt eingebettet sein. Sie zählt eine Reihe von Wegen auf, in denen Überzeugungen Simulationen beeinflussen können; Simulationen sind ihr zufolge kognitiv durchdringbar. Sowohl im ersten als auch im dritten Schritt des Simulationsprozesses können Fehler auftreten, die trotz Erfüllung der Ähnlichkeitsbedingung zu falschen Vorhersagen führen.³⁸

Den Langer-Effekt kommentiert Heal zum Beispiel wie folgt: Es ist denkbar, dass eine einfache Rationalitätstheorie dazu führt, dass man gar keine angemessene Simulation beginnt. Man erkennt, dass ein Lotto-Ticket nicht an Wert gewinnt, nur weil man es selbst aus einer Lostrommel zieht, und neigt daher zur Annahme, dass Menschen beim Verkaufen des Loses nur den ursprünglichen Preis und die Gewinnchancen berücksichtigen werden. Eine dadurch eingegrenzte und fehlgeleitete Simulation schließt aber nicht aus, dass eine erfolgreiche Simulation möglich gewesen wäre. Vielleicht würden Menschen mit einer anderen Theorie im Hinterkopf (etwa zur Rolle von Emotionen) eine umfassendere und realistischere Simulation einleiten.

³⁸ Auch im zweiten Simulationsschritt sind Fehler denkbar. Man kann zum Beispiel die Simulation nicht ernst genug nehmen und sich nur unvollständig in die Lage der simulierten Person hineinversetzen. Doch das sind keine Beispiele für kognitive Durchdringbarkeit der Simulation.

If so the simulationist can say that the simulation did not run properly because the simulator did not take on board a wide enough range of considerations and did not make him or herself open to the possibility of these less clear-cut kinds of rationality. [...] And why is this so? Perhaps because of the influence of some (tacit) theoretical assumptions [...]. (Heal 1996b, 64f.)

Ich füge hinzu, dass es auch denkbar wäre, dass solche Überlegungen zur Rationalität im Nachhinein ein korrektes Simulationsergebnis als unglaubwürdig (weil irrational) disqualifizieren und "überschreiben" können.

Stich & Nichols (1997) haben schließlich eingestanden, dass das Kriterium der kognitiven Durchdringbarkeit keine eindeutig zu interpretierenden Daten liefert:

On our view, Heal's conjectures about the ways in which a predictor's beliefs might affect the outcome of a decision prediction subserved by a [simulation] are all very plausible. What they show, we think, is that it was a mistake to claim that [simulation] is not cognitively penetrable. Decision predictions can fail as the result of cognitive penetration, even when the predictor and the target are psychologically similar, and the predictor is in an excellent position to generate pretend beliefs and desires similar to the target's. (Stich & Nichols 1997, 315)

Ich werde daher dieses Kriterium nicht ausführlicher diskutieren, sondern im nächsten Abschnitt auf einen neueren Vorschlag eingehen.

Unbestimmte alltagspsychologische Theorien

Bevor ich dazu komme, möchte ich aber noch kurz einen weiteren Kritikpunkt anbringen: Stich & Nichols machen unbegründete Annahmen über alltagspsychologische Theorien. Um aus dem Theorienmodell eine Hypothese für den Ausgang der einzelnen Experimente zu gewinnen, muss angenommen werden, dass die Versuchspersonen über keine oder eine falsche Theorie zum Stimulus verfügen. Die Autoren begründen ihre Annahme zwar damit, dass es sich um Situationen handelt, die im Alltag selten vorkommen oder irrelevant sind. Doch diese Einschätzung kann täuschen.

Es wäre zum Beispiel möglich, dass die Versuchspersonen im Langer-Experiment eine Theorie anwenden, nach der es Gegenstände gibt, zu denen Menschen ein – für andere nicht immer nachvollziehbares – liebevolles Verhältnis entwickeln. Das geschieht oft genug mit Autos, Büchern, Armbanduhren und der eigenen Kaffeetasse, um plausibel erscheinen zu lassen, dass Menschen über diese Theorie verfügen. In diesem Fall wäre nach dem Theorienmodell zu erwarten, dass die Versuchspersonen den Effekt *richtig* vorhersagen – und die Unterscheidbarkeit vom Simulationsmodell wäre dahin. Doch

wie kann vorab entschieden werden, über welche alltagspsychologische Theorie Menschen verfügen?

6.2 Hinweise zur Theorieanwendung

Perner et al. (1999) haben ein experimentelles Design entwickelt, das auf dem folgenden plausiblen Merkmal der Theorieanwendung aufbaut: Theorien lassen sich leichter anwenden, wenn die theoretisch relevanten Faktoren hervorgehoben werden. Das Hervorheben der relevanten Faktoren ist nichts anderes als das Geben von Tipps. Ich halte diesen Ansatz für viel versprechend und die ersten damit gewonnenen Befunde für aussagekräftig, werde das Prinzip jedoch ebenfalls kritisieren.

Das Hervorheben relevanter Faktoren

Das Merkmal der Theorieanwendung verknüpfen die Autoren wie folgt mit einem empirischen Indikator: Legt man Versuchspersonen *zwei* Situationen zur gleichzeitigen Verhaltensvorhersage vor, die sich gerade hinsichtlich des relevanten Faktors unterscheiden, macht man diesen Faktor für die Versuchspersonen salient. Nach dem Theorienmodell sollte das dazu führen, dass der Effekt dieses Faktors (im Durchschnitt) deutlicher vorhergesagt wird als wenn die Versuchspersonen nur die Reaktion in einer dieser Situationen vorhersagen sollen.

Dieser Ansatz ermöglicht auch die Ableitung einer entgegengesetzten Hypothese aus dem Simulationsmodell: Die Simulation sollte durch die gleichzeitige Präsentation von zwei Situationen *erschwert* werden, da zu vermuten ist, dass Menschen Schwierigkeiten dabei haben, sich hintereinander in zwei ähnliche Situationen hineinzusetzen. Die Qualität ihrer Vorhersagen sollte bei gleichzeitiger Präsentation demnach sinken.

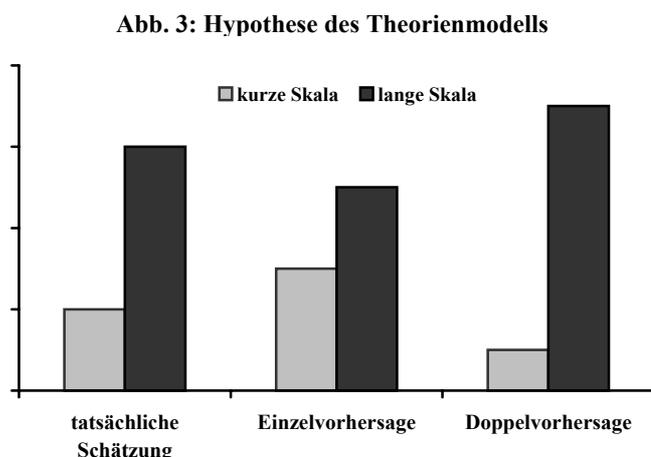
[S]imulation should work better the more similar the task conditions for the simulator are to the conditions of the simulated person. By juxtaposing the contrasting conditions, prediction might deteriorate. [...] With some plausibility one can make the opposite prediction for the theory user. A theory is easier to apply if the critical factor is being highlighted. That is exactly what juxtaposition does [...]. (Perner et al. 1999, 60)

Als Beispiel wählten Perner und seine Mitarbeiter die Vorhersage von Schätzungen (*Estimation Tasks*). Menschen lassen sich bei Schätzungen unbekannter Werte von den

Antwortmöglichkeiten beeinflussen (der so genannte *Response Bias*). Eine der Aufgaben lautete: Wie schnell ist ein gutes Rennpferd? Die Antworten sollten auf einer Skala angekreuzt werden, die in einem Fall von 20 bis 90 km/h und im zweiten Fall von 20 bis 120 km/h ging. Die durchschnittliche Schätzung betrug für die kürzere Skala 54 km/h und für die längere Skala 67 km/h – ein signifikanter Unterschied.

Die Vorhersage von Schätzungen

Einige der Versuchspersonen wurden nun gebeten, die typische Schätzung für *eine* der Skalen vorherzusagen, während andere gebeten wurden, die typischen Schätzungen für *beide* Skalen vorherzusagen. Es zeigte sich, dass die Versuchspersonen in der ersten Gruppe einen kleineren Effekt vorhersagten (Unterschied von 6 km/h) als die Versuchspersonen der zweiten Gruppe, die die Schätzungen auf beiden Skalen vorhersagten (Unterschied von 14 km/h).



Dieses Ergebnis spricht für das Theorienmodell, da das Hervorheben des relevanten Faktors (Länge der Skala) die Anwendung einer Theorie, derzufolge sich Menschen bei Schätzungen von der Länge der Skala beeinflussen lassen, vereinfacht wird.³⁹ Die Hypothese des Theorienmodells für die zu

messenden Werte (Durchschnitt der Schätzungen) ist in Abb. 3 illustriert.

Ein weiteres Experiment von Perner et al. (1999) spricht jedoch für das Simulationsmodell. Der Effekt ist dieser: In einigen Fällen liegen Menschen mit ihren Schätzungen völlig falsch, wenn ihnen keine Skala vorgegeben wird, die ihre Schätzung einschränken würde. Versuchspersonen schätzten zum Beispiel die Fläche Brasiliens auf 560.000 km², wenn sie um eine freie Schätzung gebeten wurden. Auf einer Skala von 2 bis 14

³⁹ Es reicht nicht aus, in der Doppel-Vorhersage-Bedingung den Effekt ungefähr vorherzusagen. Vielmehr sollte der Effekt in diesem Fall *stärker* als in der Einzel-Vorhersage-Bedingung vorhergesagt werden. Denn nur so wird der Erleichterungseffekt deutlich.

Millionen km² war ihre durchschnittliche Schätzung hingegen deutlich höher: 9,4 Millionen km².

Diesen Effekt konnten Versuchspersonen vorhersagen, wenn sie *nur eine* der beiden Schätzungen vorhersagen sollten. Wurden sie gleichzeitig um zwei Schätzungen gebeten (eine freie und eine skalengebundene), gaben sie auch im Fall der freien Schätzung einen ähnlich hohen Wert wie in der skalengebundenen Schätzung an. Das legt den Schluss nahe, dass *simuliert* wurde, da die Skala der einen Antwort-Situation die Vorhersage für *beide* Antwort-Situationen beeinflusst hat. Sobald die Versuchspersonen die Skala sehen, ist ihre Vorhersage schon so sehr eingeschränkt, dass sie sich die drastisch abweichende Schätzung in der freien-Antwort-Situation nicht mehr vorstellen können. Die Autoren der Studie sprechen hier von einem *Carry-Over*-Effekt der skalengebundenen Vorhersage auf die freie Vorhersage.

Spielraum bei der Interpretation

Es fällt auf, dass diese Experimente alle mit Erwachsenen durchgeführt wurden. Dabei bestreitet niemand, dass Erwachsene über alltagspsychologische Theorien verfügen. Ich halte es daher für plausibel, aber uninteressant, dass Perner und seine Kollegen zum Beispiel aus dem stärkeren Effekt in der Doppel-Vorhersage-Bedingung schließen, dass durch Hervorheben des relevanten Faktors die Anwendung einer Theorie gefördert wurde. Selbst wenn Kinder vornehmlich simulieren sollten, ist davon auszugehen, dass ihnen nach und nach gewisse Regularitäten auffallen werden, die sie in Form einer Theorie repräsentieren können. Erwachsene sollten daher – unabhängig davon, ob das Theorien- oder das Simulationsmodell "primär" ist – über eine ganze Reihe alltagspsychologischer Theorien verfügen.

Wie sieht es aber mit dem Simulationsmodell aus; stützt der *Carry-Over*-Effekt das Simulationsmodell? Ich denke schon; allerdings deutet er nur auf eine *faktische* Simulation hin, deren Existenz ebenfalls niemand bestreitet (siehe dazu Abschnitt 4.1). Die Versuchspersonen haben keinerlei Anhaltspunkte über die Menschen, deren Schätzung sie vorhersagen sollen; insbesondere liegen ihnen keine Informationen über zu berücksichtigende psychologische Unterschiede zwischen sich selbst und den zu simulierenden Personen vor. Es ist daher anzunehmen, dass sie sich einfach selbst fragen, welche Schätzung angemessen wäre.

Zudem möchte ich einen grundsätzlichen Einwand am Unterscheidungskriterium anbringen: Die Autoren können durch das Hervorheben bestimmter Faktoren nur die Anwendung von Theorien über diese Faktoren erleichtern. Sie unterlassen es jedoch, *andere* alltagspsychologische Theorien in Betracht zu ziehen, die die Versuchspersonen angewendet haben könnten. Mit anderen Worten: Das Hervorheben des relevanten Faktors fördert zwar im Prinzip die Anwendung einer Theorie über den tatsächlichen Effekt, diese Förderung muss aber nicht von Erfolg gekrönt sein. Das scheinen die Autoren der Studie aber vorauszusetzen: Wie sonst könnten sie behaupten, dass die erleichterte Theorienanwendung zu einer *besseren* Vorhersage des Effekts führe?

Versuchspersonen könnten meines Erachtens auch *falsche* Theorien über den in Frage kommenden Effekt anwenden. Zum Beispiel könnten die Versuchspersonen nach einigem Nachdenken darüber, ob die Skala einen Einfluss auf die Schätzung ausübt, zum Ergebnis gelangen, dass sie das nicht tut. In diesem Fall würden sie für beide Antwort-Situationen eine ähnliche Schätzung voraussagen – und das Ergebnis wäre von dem einer Simulation nicht zu unterscheiden. Es könnte auch sein, dass sie über *gar keine* geeignete Theorie verfügen und vermuten, dass es in dem Experiment darauf ankommt, den relevanten Faktor zu bemerken und in die Vorhersage einfließen zu lassen. Obwohl sie in diesem Fall eine Theorie anwenden, würde man es nicht als Bestätigung des Theorienmodells werten, da die Theorie *ad hoc* gebildet wurde und womöglich nicht typisch ist für die alltagspsychologische Praxis.

Das Verfahren der erleichterten Theorienanwendung von Perner et al. greift also nur, wenn die alltagspsychologische Theorie der Versuchspersonen *bekannt* ist, so dass abgeschätzt werden kann, in welche Richtung die Änderung der Vorhersage geht, wenn der relevante Faktor hervorgehoben wird. Doch wie kann eine solche Annahme begründet werden? Soweit ich sehe, hat man die alltagspsychologischen Theorien Erwachsener bislang nur sehr bruchstückhaft explizieren können.

Die Rolle alltagspsychologischer Intuitionen

Perner & Kühberger (2003) wandten das experimentelle Design auf Lotteriespiele an. In den neuen Experimenten ging es darum, dass Menschen das Risiko eher scheuen, wenn es um Gewinne geht – bei drohenden Verlusten aber das Risiko suchen. Stellt man Versuchspersonen beispielsweise vor die Wahl, 10 Euro zu gewinnen oder mit einer

50%igen Wahrscheinlichkeit 20 Euro zu gewinnen, wählen die meisten die sicheren 10. Ändert man das *Vorzeichen* der Gewinne, kehrt sich die Präferenz um: Die Mehrheit der Versuchspersonen wählt die riskante Alternative, in der man mit einer Wahrscheinlichkeit von 50% nichts verliert.⁴⁰

Werden Versuchspersonen nun gebeten, für eine der beiden Bedingungen eine Vorhersage abzugeben, replizieren sie den Effekt. Werden sie gleichzeitig um Vorhersagen für *beide* Bedingungen gebeten, fällt der vorhergesagte Effekt jedoch deutlich *geringer* aus. Gemäß der bisherigen Interpretation spricht das für das Simulationsmodell, da angenommen wird, dass die gleichzeitige Präsentation von zwei Situationen die Fähigkeit beeinträchtigt, sich in beide angemessen hineinzusetzen.

In einem zweiten Experiment ging es um den *Größeneffekt*: Menschen sind risikoscheuer, wenn es um größere (positive) Beträge geht, und wählen häufiger die sichere Variante, wenn es um viel Geld geht. Dieser Effekt wird jedoch nur dann vorhergesagt, wenn den Versuchspersonen beide Situationen gleichzeitig vorgelegt werden. Einzelvorhersagen ergeben nur einen geringen Größeneffekt. Gemäß der bisherigen Interpretation spricht das für das Theorienmodell.

Wie lässt sich der Unterschied im Ergebnis der beiden Experimente deuten? Die Autoren sehen den entscheidenden Unterschied darin, dass vielen Menschen der Größeneffekt plausibel, der Vorzeicheneffekt (auch *Framing*-Effekt genannt) hingegen irrational vorkommt. Einen Effekt als rational oder plausibel einzuschätzen, ist ihrer Meinung nach auf das Verfügen über eine entsprechende alltagspsychologische Theorie zurückzuführen. Sie vermuten also, dass das Gefühl, es gebe einen kausalen Zusammenhang zwischen einem Stimulus und einer Reaktion, auf eine zugrundeliegende alltagspsychologische Theorie dieses Zusammenhangs verweist.

[I]ntuitively one has the impression that the payoff-size effect is not very surprising and that people might even have an explicit theory about it. Hence we might expect the prediction experiment to show the use of theory. In contrast, the framing effect is puzzling. [...] Since people find the effect puzzling they seem to have no, at least, not a conscious theory about it. So, the prediction data might indicate the use of simulation. (Perner & Kühberger 2003, 158)

Doch m.E. geht es nicht darum, ob den Versuchspersonen *die wahre Theorie* plausibel vorkommt, denn sie wissen ja nicht, welchen Effekt sie vorhersagen sollen. Es ist

⁴⁰ Wenn es um Verluste geht, erhalten die Versuchspersonen zu Beginn des Experiments einen bestimmten Betrag, den sie – im schlimmsten Fall – komplett verlieren.

durchaus denkbar, dass z.B. das Ergebnis im ersten Experiment (Vorzeichen-Effekt) auf die Anwendung einer alltagspsychologischen Theorie zurückzuführen ist, derzufolge das Vorzeichen *keine* Rolle spielt. Wenn man sich überlegt, worin der Unterschied der Präsentation als Gewinn und als Verlust besteht, kann man zu der Überzeugung gelangen, dass es keinen gibt. Diese Theorie erschien vermutlich den meisten Versuchspersonen plausibel.

Fazit: Zum Nachweis von Prozessmerkmalen

In diesem Teil der Arbeit habe ich eine Reihe überzeugender Studien und Argumente vorgestellt. Perner und seinen Kollegen ist es gelungen, einige Hinweise sowohl auf die Anwendung alltagspsychologischer Theorien als auch auf Simulationen zu gewinnen. Und Heal weist darauf hin, dass das Simulationsmodell zumindest in einem Bereich (der Vorhersage von Schlussfolgerungen) *prima facie* einfacher und damit plausibler ist als das Theorienmodell.

Dennoch habe ich die Prinzipien, die diesen Studien zugrunde liegen, kritisiert. Ich habe versucht zu zeigen, dass es bislang nicht gelungen ist, auf diese Weise eine Entscheidung zwischen Theorien- und Simulationsmodell herbeizuführen. Der Grund dafür ist, dass solche Nachweise spezifische Annahmen darüber erfordern, welche alltagspsychologische Theorie angewendet bzw. welcher kognitive Mechanismus simuliert wird.

Perner und seine Kollegen haben Merkmale untersucht, die für die Theorieanwendung bzw. die Simulation spezifisch sind. Diese Faktoren wirken sich aber nur mittels einer bestimmten Theorie bzw. eines bestimmten Mechanismus' auf die gemessene Leistung aus. Solange diese jedoch nicht bestimmt werden können, weisen Verbesserungen bzw. Verschlechterungen der Qualität alltagspsychologischer Urteile nicht eindeutig auf das Theorien- oder das Simulationsmodell hin. Beide Modelle lassen zuviel Spielraum bei der Interpretation der Daten. So können Perner und seine Kollegen beispielsweise nur feststellen, dass das Hervorheben eines relevanten Faktors die Anwendung einer Theorie fördert, die eine Aussage über diesen Faktor macht. Doch es bleibt offen, über *welche* Theorie die Versuchspersonen verfügen, so dass nicht klar ist, ob die alltagspsychologische Vorhersage durch das Hervorheben des relevanten Faktors verbessert oder verschlechtert wird.

Die Effekte, die die Autoren messen konnten, können zwar klar dem Theorien- bzw. Simulationsmodell zugeordnet werden. Doch ich habe, was diese speziellen Befunde betrifft, die Vermutung geäußert, dass es sich um keine interessanten alltagspsychologischen Phänomene handelt. Es ist zum Beispiel denkbar, dass die Theorien von den Versuchspersonen während des Experiments erfunden wurden, oder dass ihnen mangels einer Theorie nichts übrig blieb als sich selbst im Rahmen einer faktischen Simulation zum Modell zu machen.

Auch bei der Diskussion von Heals Argument gegen das Theorienmodell bin ich zu einem ähnlichen Ergebnis gelangt: Sie kann zwar feststellen, dass einige kognitive Leistungen (z.B. Schlussfolgerungen) zu komplex sind, als dass man hoffen könnte, sie mit Hilfe einer alltagspsychologischen Theorie vorherzusagen. Doch es bleibt offen, ob Menschen zu diesen Vorhersagen überhaupt in der Lage sind, so dass nicht klar ist, genau welche Leistung Theorien- und Simulationsmodell erklären müssen.

Gibt es eine Möglichkeit, die Annahmen über verwendete Theorien und simulierte Mechanismen besser zu begründen? Mein Vorschlag ist, die Merkmale der beiden Prozessstypen anhand eines gut untersuchten alltagspsychologischen Tests nachzuweisen. Hierzu bietet sich der *False Belief Test* an, den ich im ersten Teil der Arbeit diskutiert habe. Der Vorteil wäre, dass die Anforderungen *dieses* Tests so weit analysiert sind, dass viele mögliche Erklärungen bereits ausgeschlossen werden konnten. Der Nachteil wäre, dass geeignete Prozessmerkmale gefunden werden müssten, die in diesem alltagspsychologischen Test manipuliert werden können. Im nächsten Teil dieser Arbeit werde ich einen Vorschlag ausarbeiten, der beiden Anforderungen gerecht wird.

III Die kognitiven Grundlagen der Simulation

Vorbemerkung: Simulation und *Pretend Play*

Im zweiten Teil dieser Arbeit habe ich dafür argumentiert, dass die bisherigen Versuche, zwischen Theorien- und Simulationsmodell zu unterscheiden, unbegründete Annahmen darüber machen, welche alltagspsychologischen Theorien im Einzelfall angewendet bzw. welche kognitiven Mechanismen simuliert werden. Die beiden Modelle lassen daher keine konkreten Hypothesen darüber zu, wie sich die experimentelle Manipulation bestimmter Faktoren im Einzelfall auf die Qualität alltagspsychologischer Vorhersagen und Erklärungen auswirkt. In diesem dritten Teil werde ich untersuchen, ob die Befunde der *Theory-of-Mind*-Forschung aus dem zweiten und dritten Kapitel dabei helfen können, den Spielraum bei der Interpretation einzuschränken.

Diesen Ansatz werde ich vor allem für das Simulationsmodell ausführen; das Ziel soll sein, ein Verfahren zum Nachweis von Simulationsprozessen zu entwickeln. Um das Theorienmodell werde ich mich nur am Rande kümmern. Denn ich werde mich darauf beschränken zu zeigen, dass der Umgang mit Überzeugungen durch das Simulationsmodell gut erklärt wird. Ich werde nicht ausschließen, dass sich auch alternative Theorienmodelle konstruieren lassen. Doch ich stütze mich auf Heals Argument aus dem fünften Kapitel, demzufolge das Theorienmodell im Hinblick auf Überzeugungen komplizierter ist.

Das besondere Interesse am Simulationsmodell lässt sich dadurch rechtfertigen, dass die Existenz von hypothetischen Simulationen (im Unterschied zur Existenz alltagspsychologischer Theorien) umstritten ist und faktische Simulationen nur für triviale Fälle geeignet scheinen.⁴¹ Wenn nachgewiesen werden könnte, dass zumindest einer interessanten alltagspsychologischen Fähigkeit ein Simulationsprozess zugrunde liegt, wäre das ein wichtiger Anhaltspunkt in der Debatte zwischen Theorien- und Simulationsmodell. Die Debatte könnte dann von der Diskussion um die prinzipielle Unterscheidbarkeit der beiden Modelle zu konkreteren Fragen der Erklärung einzelner alltagspsychologischer Fähigkeiten fortschreiten.

⁴¹ Die beiden Simulationstypen habe ich im vierten Kapitel vorgestellt.

Als zu untersuchende alltagspsychologische Fähigkeit werde ich – wegen der Menge verfügbarer Daten und wegen der Bedeutung dieser Fähigkeit für die alltagspsychologische Praxis – das Verständnis falscher Überzeugungen wählen. Im achten Kapitel werde ich dafür argumentieren, dass das Simulationsmodell eine gute Erklärung dafür liefert, wie vierjährige Kinder anderen Menschen falsche Überzeugungen zuschreiben.

Bevor ich zu dieser Behauptung komme, werde ich im siebten Kapitel jedoch eine Reihe von Vorschlägen aus der Forschungsliteratur durchgehen, die den Zusammenhang von Simulation, kontrafaktischem Schließen und dem spielerischen So-tun-als-ob (im Englischen: *Pretend Play*, kurz: PP) betreffen. Ich glaube, dass einige dieser Vorschläge einen wahren Kern haben. Ich werde aber auch dafür argumentieren, dass jeder von ihnen eines der vielen *Explananda*, die es in dieser Debatte gibt, übersieht. Ich möchte nun einen ersten Überblick über diese kognitiven Fähigkeiten geben.

Das Rollenspiel

Für das Simulationsmodell ist eine Form des PP von besonderem Interesse: das Rollenspiel (*Role Play*). Schon junge Kinder übernehmen die Rollen anderer Personen, schreiben ihren Puppen solche Rollen zu oder denken sich gar fiktive Freunde aus.⁴² Dazu ändern sie ihren Tonfall, nehmen andere Namen an und betrachten die Welt aus einer anderen Perspektive. Ein Kind etwa, das Mutter spielt, kann einer Puppe verbieten, auf dem Plastik-Pony zu reiten, obwohl es selbst sehr gerne zum Ponyhof geht. Junge Kinder sind in der Lage, sich in andere Rollen bzw. Personen hineinzusetzen und sich dabei so zu verhalten, dass ihre angenommene Identität auch nach Außen sichtbar wird. Welche kognitiven Fähigkeiten sind nötig, um fremde Rollen angemessen übernehmen zu können? Die Rollen von Mutter oder Vater sind mit gewissen Pflichten, Einstellungen und zu bewältigenden Aufgaben verknüpft. Das Kind kann zum Beispiel eine Mutter spielen, die ihr Baby ins Bett bringt, später aber entdeckt, dass ihr Baby heimlich aus dem Bett gekrabbelt ist, und auf diese böse Überraschung hin mit dem Baby schimpft.

⁴² Nach Taylor (1998) ist diese Form des PP weit verbreitet. Kinder mit unsichtbaren Freunden scheinen sich dabei hinsichtlich ihrer sonstigen Persönlichkeitseigenschaften nicht von Kindern zu unterscheiden, die keinen solchen Freund haben. Taylor & Carlson (1997) konnten sogar zeigen, dass Kinder mit unsichtbaren Freunden bessere Leistungen im FBT erbringen.

Dazu muss das Kind hypothetische Wünsche und Überzeugungen unterhalten: etwa die Überzeugung, dass das Baby im Bett liegt und schläft.

Die hypothetischen Wünsche und Überzeugungen muss das Kind von seinen normalen Wunsch- und Überzeugungssystemen *abgrenzen*, damit es nicht zum Konflikt hypothetischer Wünsche und Überzeugungen mit inkompatiblen ernsthaften Wünschen und Überzeugungen kommt. Die Abgrenzung darf wiederum nicht strikt sein, sondern muss "durchlässig" oder "flexibel" gehandhabt werden, um das Heranziehen von Hintergrundwissen zu ermöglichen: Aus der Annahme, das Baby liege im Bett und schlafe, kann das Kind ableiten, dass das Baby voraussichtlich die ganze Nacht im Bett liegen und schlafen wird. Schließlich wird das Kind auf der Grundlage der angenommenen Wünsche und Überzeugungen auch Schlüsse und Entscheidungen vollziehen. So wird es zum Beispiel entscheiden müssen, wie es auf das überraschende Verhalten des Babys reagieren soll.

Das Entwickeln eines abgegrenzten hypothetischen Szenarios und das Sich-hineinversetzen in eine andere Rolle erinnert an das Simulationsmodell. Es ist denkbar, dass Kinder beim Rollenspiel die kognitiven Mechanismen anwenden, die auch die Bildung ihrer ernsthaften Wünsche und Überzeugungen steuern. Das Rollenspiel wird auch 'on-line simulation' genannt, da das Simulationsergebnis (im Gegensatz zur *off-line simulation* der alltagspsychologischen Praxis) tatsächlich ausgeführt wird. Es könnte also sein, dass dieselben kognitiven Mechanismen, die alltagspsychologische Simulationen ermöglichen, auch das Rollenspiel steuern.

Lägen nun über das Rollenspiel *unabhängige* Erkenntnisse vor, die diese Erklärung stützen würden, wären simulationsartige kognitive Prozesse zumindest im Bereich des PP nachgewiesen. Das würde die Hoffnung nähren, dass solche Prozesse auch der alltagspsychologischen Praxis zugrunde liegen. Diese Ansicht vertreten auch Nichols & Stich (2000):

[I]f the simulation account of pretend play is right, it would mean that the basic mechanisms posited by simulation theorists have to be in place. And if that is true it might lend considerable plausibility to simulation theorists' claim that mindreading also depends on a process of simulation [...]. (Nichols & Stich 2000, 116)

Es ist meines Erachtens aber unrealistisch, auf eine einfache Bestätigung des Simulationsmodells des Rollenspiels zu hoffen. Denn beim Rollenspiel ergibt sich dasselbe Problem wie bei der Erklärung der alltagspsychologischen Praxis: Es ist möglich, dass

Kinder auf ihre alltagspsychologischen Theorien zurückgreifen, um weitere Überzeugungen und Wünsche innerhalb ihrer neuen Rolle zu bestimmen. Bei der Wahl einer angemessenen Reaktion auf das Verhalten des Babys kann das Kind zum Beispiel auf eine Theorie darüber zurückgreifen, wie Eltern reagieren, wenn ihre Erwartungen enttäuscht werden.

Harris unternimmt zwar einen Versuch, gegen das Theorienmodell des Rollenspiels zu argumentieren, der an Heals Argument aus dem fünften Kapitel erinnert: Das Rollenspiel junger Kinder sei so flexibel und reichhaltig, dass es sich nicht unter Rückgriff auf typische Verhaltensmuster bzw. Verhaltensscripts erklären lasse. Vielmehr müsse davon ausgegangen werden, dass Kinder ihre normalen Entscheidungsmechanismen anwenden, um ihre Rollen mit Leben zu füllen. Mit anderen Worten: Harris wirft die Frage auf, ob sich komplexes Rollenverhalten theoretisch erfassen lässt.

By way of illustration, suppose that the child is playing at pirates. [...] For example, the pirate may have found some treasure; in addition he may have enemies who might steal it. Even in this relatively simple situation, there is no standard 'pirate' script that specifies what to do next. However, the process of simulation can help the child to construct a suitable plan. (Harris 2000, 35)

Doch mit diesem Beispiel überschätzt Harris meines Erachtens die Flexibilität des kindlichen Rollenspiels und unterschätzt zugleich die Mächtigkeit von Theorien. Selbst unter der – aus Sicht von Kindern sicher zweifelhaften – Annahme, dass das Vergraben von Schätzen nicht zum typischen Verhalten eines Piraten gehört, gehört nicht viel dazu die alltagspsychologische Theorie zu postulieren, dass Menschen Dinge verstecken, wenn die Gefahr des Diebstahls droht. Zu diesem Ergebnis kann man demnach auch gelangen, ohne zu simulieren.

Harris scheint bei Theorien nur an Verhaltensroutinen in einem eng umgrenzten Bereich zu denken. Doch es sind auch Theorien möglich, die verschiedene Verhaltensweisen aus disparaten Bereichen systematisch beschreiben. Diese Alternative nimmt Harris meines Erachtens nicht ernst. Ich werde seinen Ansatz daher nicht weiter verfolgen; stattdessen werde ich davon ausgehen, dass die Chancen, das Simulationsmodell des Rollenspiels direkt zu bestätigen, nicht besser stehen als für das Simulationsmodell der Alltagspsychologie. Eine theoretische Verknüpfung zwischen Rollenspiel und Simulationsmodell ist für meine Zwecke daher ungeeignet.

Normales Pretend Play

Das Rollenspiel ist nicht die einzige Form des PP. Ich wende mich nun einer anderen Form des PP zu, die ich 'normales PP' nennen möchte. Damit meine ich die allgemeine Fähigkeit, mit fiktiven *Pretend*-Szenarien umzugehen, die durch eine Reihe nicht-psychologischer Stipulationen konstituiert werden. Zweijährige Kinder haben zum Beispiel keine Schwierigkeiten damit zu verstehen, was *Naughty Teddy* macht, wenn er ein leeres Glas in die Hand nimmt und es über dem Kopf einer anderen Puppe umdreht.⁴³ Auch in diesem Fall müssen Kinder ihre widersprechenden Überzeugungen ausblenden und auf ihr Hintergrundwissen zurückgreifen, um auszuarbeiten, was die *Pretend*-Stipulationen implizieren bzw. um kausale Prozesse innerhalb des *Pretend*-Szenarios zu verfolgen. Harris beschreibt diese Fähigkeit so:

[Pretend play] emerges in late infancy, around the same time as language. By that time, the child has gained a limited but functional understanding of the way the world works. Children exploit a good deal of that already established conceptual knowledge in their pretend play. In particular, they begin to deploy their causal understanding of everyday transformations. (Harris 2000, 27)

Im Unterschied zum Rollenspiel geht es jedoch nur um physikalische Zustände, nicht aber um Wünsche und Entscheidungen. Es ist daher kein *alltagspsychologisches* Wissen erforderlich – und damit auch keine Erklärung im Sinn des Theorienmodells der Alltagspsychologie nötig. Vielmehr ist es plausibel anzunehmen, dass dem PP dieselben kognitiven Mechanismen zugrunde liegen, die auch das normale Schließen unter Heranziehen des relevanten Hintergrundwissens steuern (anstatt einen zweiten Schlussfolgerungsmechanismus zum Zweck des PP zu postulieren).

Eine solche Theorie der kognitiven Grundlagen des PP bieten Nichols & Stich (2000). Sie postulieren eine besondere kognitive "Schachtel", die *Possible World Box*, in der die *Pretend*-Stipulationen enthalten sind und in die alle sonstigen relevanten Überzeugungen "hineinkopiert" werden. Diese Schachtel zu spezifizieren soll heißen, den darin enthaltenen geistigen Zuständen ein besonderes kausales Profil zuzuschreiben: Sie werden etwas anders behandelt als normale Überzeugungen. Genauer gesagt werden die Überzeugungen in der *Pretend*-Schachtel vom Hintergrundwissen abgegrenzt, aber innerhalb der Schachtel im Großen und Ganzen nach denselben Regeln verarbeitet wie auch die

⁴³ Eine gründliche Untersuchung des kindlichen PP's bieten Harris & Kavanaugh (1993).

Elemente der normalen Überzeugungsschachtel. Das macht die Erklärung von Nichols & Stich zu einer Erklärung im Sinn des Simulationsmodells. Insbesondere postulieren die Autoren einen *UpDater*, der normalerweise bereits vorhandene Überzeugungen im Licht neuer Informationen aktualisiert. Im Rahmen des PP werden die Stipulationen des *Pretend*-Szenarios als neue Informationen betrachtet, und der *UpDater* berechnet ihre Implikationen (er malt gewissermaßen das Szenario aus).

The basic idea is that when pretense is initiated, the UpDater is called into service. It treats the contents of the Possible World Box in much the same way that it would treat the contents of the Belief Box when a new belief is added, though in the [Possible World Box] it is the pretense premise that plays the role of the new belief. (Nichols & Stich 2000, 124)

Diese Theorie ist sicher ausbaufähig, da die Unterschiede in der Verarbeitung normaler und hypothetischer Überzeugungen nur zum Teil spezifiziert werden. Doch sie ist so weit ausgearbeitet, dass sie die theoretische Verknüpfung der alltagspsychologischen Praxis mit dem normalen PP erlaubt. Das ist viel versprechend, da über das PP (wie sich noch zeigen wird) unabhängige Erkenntnisse vorliegen. Mein Ziel wird es sein zu zeigen, dass das normale PP die kognitive Grundlage (zumindest eines Teils) der alltagspsychologischen Praxis darstellt. Das würde dem Simulationsmodell der Alltagspsychologie eine gewisse Anfangs-Plausibilität verleihen.

Der Erleichterungseffekt im kontrafaktischen Schließen

Um dieses Ziel zu erreichen, werde ich noch eine dritte Kompetenz ins Spiel bringen: das kontrafaktische Schließen. Es hat sich gezeigt, dass zwei- und dreijährige Kinder große Schwierigkeiten mit kontrafaktischen Syllogismen haben. Bei Aufgaben der Art:

Alle Fische leben in den Bäumen.
Peter ist ein Fisch.
Lebt Peter im Wasser oder in den Bäumen?

greifen junge Kinder meist auf ihre ernsthaften Überzeugungen zurück und vernachlässigen die kontrafaktische Prämisse (der so genannte *realist intrusion error*).⁴⁴ Sie sind

⁴⁴ Perner (2000) hat darauf hingewiesen, dass dreijährige Kinder nur ein Problem mit kontrafaktischen Prämissen haben, nicht aber mit hypothetischen Prämissen im Allgemeinen. Im Hinblick auf ein Szenario, in dem man mit dem Zug zum Ort A und mit dem Bus zum Ort B gelangen kann, können auch dreijährige Kinder korrekt angeben, wohin Peter gelangen würde, wenn er den Zug nähme.

demnach nicht ratlos, sondern geben systematisch die falsche Antwort. Diese Schwierigkeiten verschwinden erst im Alter von vier Jahren.

Die Schwierigkeiten jüngerer Kinder mit kontrafaktischen Syllogismen stehen auf den ersten Blick im Widerspruch zu ihren sonst erstaunlichen Fähigkeiten zum PP: Warum können sie so tun, als lebten die Fische in den Bäumen, und zugleich bei der o.g. Syllogismusfrage scheitern? Ich glaube, dass bereits dreijährige Kinder über die Kompetenz verfügen, die kontrafaktischen Fragen zu beantworten. Es ist die *Art der Aufgabenstellung*, die sie davon abhält, diese Kompetenz zu nutzen. Zu dieser Behauptung motiviert mich ein *Erleichterungseffekt*, den Harris und Dias nachgewiesen haben. Die Autoren konnten Dreijährigen dadurch helfen, kontrafaktische Syllogismen richtig zu lösen, dass sie die Kinder aufforderten, die Prämissen als *Pretend*-Stipulationen aufzufassen. Im Abschnitt 8.1 werde ich diesen Effekt ausführlich vorstellen.

Wenn nun der Hinweis auf ein *Pretend*-Szenario das kontrafaktische Schließen erleichtert, dann ist auch denkbar, dass er den Umgang mit den falschen Überzeugungen Anderer fördert. Denn es ist möglich, dass die Beantwortung der kontrafaktischen Frage

(K-Frage) Wo wäre die Schokolade jetzt, wenn sie nicht verlegt worden wäre?

den Kern des *False Belief Tests* (kurz: FBT) ausmacht. Um das zu verdeutlichen, erinnere ich an die Theorie (EZ), die ich im Abschnitt 2.2 vorgestellt habe. Dieser Theorie wird zur Ausbildung von Überzeugungen ein "epistemischer Zugang" benötigt. Mit Hilfe dieser Theorie können Kinder im FBT erkennen, dass der Protagonist nicht über den gegenwärtigen Standort eines bestimmten Objekts informiert ist. Wie aber geht es dann weiter?

Das Simulationsmodell, das ich in diesem Teil der Arbeit vorschlage, erklärt den nächsten Schritt wie folgt: Die Kinder versetzen sich in die Lage, in der sie selbst nicht wissen, dass das Objekt an einen neuen Ort gelegt worden ist. Auf dieser Grundlage bestimmen sie die Antwort auf die K-Frage.⁴⁵ Die Antwort fassen sie dann als Vorhersage der Schlussfolgerung des Protagonisten auf.

⁴⁵ Im Hinblick auf das dreischrittige Schema einer Simulation, das ich im Abschnitt 4.1 vorgestellt habe, möchte ich betonen, dass es sich hierbei um den *zweiten*, entscheidenden Schritt handelt (also den Schritt, in dem die eigenen geistigen Ressourcen ausgenutzt werden). Der Punkt, dass die Beantwortung der K-Frage noch keine Antwort auf die eigentliche Testfrage nach dem Verhalten des Protagonisten bietet, ist ohne Belang. Denn es ist durchaus möglich, die Beantwortung der Testfrage in mehrere Teilschritte zu zerlegen, die einzeln zu erklären sind.

Typischerweise können erst vierjährige Kinder die K-Frage beantworten; man sagt, die Fähigkeit "korreliere" mit der, den FBT zu bestehen. Riggs et al. (1998) fanden sogar eine sehr deutliche Korrelation.⁴⁶ Daher gibt es einen Anfangsverdacht, dass der Erleichterungseffekt auch beim FBT wirken könnte. Sollte sich diese Erwartung bestätigen, würde das für einen *Early-Competence-Account* sprechen, demzufolge Kinder schon vor ihrem vierten Geburtstag die Kompetenz besitzen, den FBT zu bestehen.

⁴⁶ Tests zum kontrafaktischen Schließen werden in der Forschungsliteratur daher auch als nicht-psychologische *Analoga* des FBT behandelt. Diesen Punkt habe ich bei der Vorstellung der anderen FBT-*Analoga* im Abschnitt 2.3 ausgespart.

7 Die flexible Abgrenzung hypothetischer Überzeugungen

In diesem Kapitel werde ich mich mit Positionen aus der Forschungsliteratur befassen, die die Natur von *Pretend Play* (kurz: PP) und kontrafaktischem Schließen – sowie den Zusammenhang dieser Kompetenzen – betreffen. Ich hoffe, mit dieser Diskussion einige Grundlagen für meine eigene Position zu legen, die ich im nächsten Kapitel entwickeln werde.

7.1 Ähnlichkeit und Abgrenzung

Für den Erfolg von Simulationen muss die Bedingung erfüllt sein, dass der Simulationsprozess dem simulierten Originalprozess *ähnlich* ist. Allerdings müssen Simulationszustände auch vom sonstigen kognitiven Haushalt in geeigneter Weise *abgegrenzt* werden. Im folgenden werde ich Versuche vorstellen, mit Hilfe einiger Kunstgriffe der Logik zu beschreiben, wie Überzeugungen beschaffen sein müssten, um beide Bedingungen gleichzeitig zu erfüllen. Einige dieser Vorschläge gehen meines Erachtens in die Irre, andere sind nur wenig gehaltvoll.

Metarepräsentationen und Pretend-Zustände

Die erste Vorschlag geht auf Davies (1994) zurück. Nach Davies' Meinung werden in theoretischen Ableitungen nach dem Theorienmodell die relevanten geistigen Zustände als propositionale Einstellungen aufgefasst, so dass der kognitive Prozess hin zu einer alltagspsychologischen Erklärung oder Vorhersage eine Abfolge von *metarepräsentationalen* Überzeugungen mit dem Gehalt 'Person P glaubt / wünscht / etc., dass *p*' darstellt. Der Prozess lässt sich als Abfolge von Überzeugungen \ddot{U} über bestimmte geistige Zustände $Z_i(p_i)$ veranschaulichen:

$$(Th) \quad \ddot{U} [Z_1(p_1)] \dots \ddot{U} [Z_2(p_2)] \dots \ddot{U} [Z_3(p_3)] \dots \ddot{U} [Z_4(p_4)] \dots$$

Die Verarbeitung solcher Zustände setzt ein theoretisches Wissen darüber voraus, wie die Operatoren 'glaubt, dass' oder 'wünscht, dass' untereinander und mit Umweltreizen

und Handlungen zusammenhängen. Die durch Punkte angedeuteten Übergänge sind theoretisch vermittelt.

Das Simulationsmodell kann Davies zufolge auf die Einbettung der propositionalen Einstellungen durch intensionale Operatoren verzichten. Es ersetzt die Abfolge metarepräsentationaler Zustände durch eine Abfolge entsprechender *Pretend*-Zustände $P-Z_i(p_i)$. Dadurch erübrigen sich psychologische Theorien, die die Verarbeitung der Zustände steuern.

(Sim) $P-Z_1(p_1) \dots P-Z_2(p_2) \dots P-Z_3(p_3) \dots P-Z_4(p_4) \dots$

Das klingt auf den ersten Blick nach einer plausiblen Rekonstruktion der beiden Modelle. Doch ich bezweifle, dass sie ihren Zweck erfüllt. Denn es ist nicht ohne weiteres klar, dass ein Prozess des Typs (Sim) dem simulierten Originalprozess

(Orig) $Z_1(p_1) \dots Z_2(p_2) \dots Z_3(p_3) \dots Z_4(p_4) \dots$

ähnlich ist. Davies begründet die Ähnlichkeit damit, dass ein Zustand des Typs $P-Z(p)$ denselben propositionalen Gehalt hat wie $Z(p)$.⁴⁷

The [simulation] alternative that I have in mind is a process in which the simulator [...] imagines believing that p and desiring that q . [...] [T]he states of the simulator [...] are 'pretend belief' and 'pretend desire' states, whose contents are simply that p or that q . [...] [P]rocessing mechanisms that mediate amongst states with such contents are not going to be embodiments of tacit knowledge of the principles of a psychological theory. (Davies 1994, 117)

Doch Gleichheit des propositionalen Gehalts garantiert nicht Gleichheit der Verarbeitung: Ein Wunsch, dass p , wird zum Beispiel anders verarbeitet als die Überzeugung, dass p . Propositionale Einstellungen sind komplexe Gebilde – eben *Einstellungen* zu propositionalen Gehalten. Die Regeln ihrer Verarbeitung umfassen nicht allein die Regeln zur Verarbeitung des propositionalen Gehalts. Davies kann daher nicht ausschließen, dass *Pretend*-Überzeugungen doch anders verarbeitet werden als ihre normalen Gegenstücke. Sein Vorschlag (Sim) zur Rekonstruktion des Simulationsprozesses bleibt damit unbegründet.

Dieses Problem greift Perner (1996) auf und unterscheidet zwei Arten der Simulation: Neben der von Davies favorisierten Möglichkeit, jeden einzelnen geistigen Zustand des simulierten kognitiven Prozesses als *Pretend*-Zustand auszuzeichnen, könnten alle Zwi-

⁴⁷ Mit dieser Rekonstruktion der beiden Modelle glaubt Davies im übrigen, seinem eigenen Argument für das Zusammenfallen von Theorien- und Simulationsmodell zu entkommen (siehe Abschnitt 1.2).

schenschritte eines simulierten Gedankengangs in einen *Pretend-Kontext* **P** eingebettet sein. Diese zweite Variante des Simulationsmodells lässt sich so veranschaulichen:

$$(Sim^*) \quad \mathbf{P} [Z_1(p_1) \dots Z_2(p_2) \dots Z_3(p_3) \dots Z_4(p_4) \dots]$$

In diesem Fall wird Perner zufolge deutlicher, dass sich Simulations- und Originalprozess ähneln; er kommentiert (Sim*) wie folgt:

Starting from an imagined (instead of perceived) situation it retraces the transitions among our actual beliefs, feelings, wants, etc. within the scope of off-line (pretend) execution. So it is for this reason, that our simulation recaptures the causal connections of the simulated person's mental states via the causal connections of our own corresponding mechanism [...] Whereas, in Davies' version [...] the requirement that the sequence of pretend-states [...] bears a correlation to the sequence of the real situation [...] would be purely coincidental [...]. (Perner 1996, 93)

Diese Rekonstruktion ist sicher plausibler als (Sim*), da sie eine Verdopplung geistiger Zustandstypen (in eine normale und hypothetische Variante) vermeidet und der Gedankengang innerhalb des *Pretend-Kontextes* dem simulierten Prozess (Orig) ähnelt. Allerdings wird im Gegenzug der Begriff des *Pretend-Kontextes* eingeführt, der nicht weiter erläutert wird. Es ist denkbar, dass dieser Kontext das kausale Profil der darin verarbeiteten geistigen Zustände so stark ändert, dass man sie dennoch als neuen Typ in Davies' Sinn betrachten muss. Beim derzeitigen Stand der Forschung ist diese Frage wohl nicht zu klären, so dass keine der beiden Rekonstruktionen weiterhilft.

Die logische Form von Pretend-Zuständen

Daher gehe ich nun auf den Versuch ein, die *Pretend-Zustände* aus Davies' (Sim) selbst als Metarepräsentationen zu analysieren – und somit den vermeintlichen Simulationsprozess in die Nähe eines Theorienmodells zu rücken. Diese irritierende Position geht auf Leslie (1987) zurück und wurde von ihm in mehreren Aufsätzen ausgearbeitet. Seine Grundidee ist, dass *Pretend-Überzeugungen* eine *andere Semantik* benötigen als normale Überzeugungen.

Sowohl bei der Substitution von Objekten (Stipulation: 'Diese Banane ist ein Telefon.') als auch bei kontrafaktischen Prädizierungen ('Diese [leere] Tasse ist voll.') und beim Umgang mit bloß vorgestellten Objekten ('Unter meinem Bett hält sich ein Monster versteckt.') gebe es sonst Schwierigkeiten mit dem Bezug von Namen und Kennzeichnungen sowie den Wahrheitsbedingungen der Stipulationen, sagt Leslie. Es sei daher

nicht möglich, *Pretend*-Überzeugungen als einfache Überzeugungen (in seiner Terminologie: primäre Repräsentationen) aufzufassen. Im Fall der Objektsubstitution würde zum Beispiel der Begriff *Telefon* mit einer wahrgenommenen Banane "referentiell verknüpft". Solche Verknüpfungen seien bei normalen Überzeugungen bedeutungskonstitutiv, d.h. im Fall des PP würde der Begriff *Telefon* etwas anderes bedeuten.⁴⁸ Leslies Lösung besteht darin, *Pretend*-Überzeugungen logisch von den entsprechenden primären Repräsentationen abzukoppeln (er nennt es "decoupling" oder "quarantining").

Pretense affects the normal reference, truth, and existence relations of the representations it uses. These relations become highly deviant. Any primary representational system affected would quickly be undermined by arbitrary meaning changes. To prevent this, pretend representations must somehow be marked off, or "quarantined," from primary representations. (Leslie 1987, 415).

Nach Leslie übernehmen *intensionale Operatoren* das Abkoppeln. Die drohende Bedeutungsänderung im Fall der Objektsubstitution lasse sich dadurch verhindern, dass ein intensionaler Operatoren einen opaken Kontext schaffe. Es ist zwar nicht ganz klar, um welche Operatoren es Leslie geht. Während er in früheren Aufsätzen von *Metarepräsentationen* spricht – also von Repräsentationen, die Repräsentationsrelationen und ihre Relata repräsentieren –, führt er in späteren Aufsätzen "M-Repräsentationen" ein, die (unter anderem) eine "informationelle Relation" von Personen zu Propositionen repräsentieren. In seiner Notation sieht das so aus: Wenn ich so tue als sei eine Banane ein Telefon, dann unterhalte ich eine Überzeugung mit dem Gehalt

I pretend "this banana is a telephone".

Es wird offenbar eine *Pretense*-Beziehung repräsentiert, die einen intensionalen Kontext schafft, der durch die Anführungszeichen angedeutet wird. Ich schlage vor, Leslies Notation wie folgt zu übersetzen: Wenn ich so tue als sei eine Banane ein Telefon, dann unterhalte ich eine Überzeugung der Form

⁴⁸ Im Original erläutert Leslie diesen Punkt an dem Beispiel, so zu tun als sei eine Muschel eine Tasse: "An item (CUP) in the symbolic code may represent a class of cups by having a *reference linkage* with percepts of the right sort [...]. Thus, having such a symbolic item, the infant could now recognize a member of a category of *cups* and distinguish them from noncups (e.g., *shells*). Pretending that a shell is a cup would involve establishing a reference link between the symbolic item CUP and an item \$shell\$ in the sensory code. However, if the reference link between symbolic CUP and sensory \$shell\$ has the same status as that between symbolic CUP and sensory \$cup\$, then CUP will change its meaning. CUP now refers to \$shell\$ as well as to \$cup\$." (Leslie 1987, 415, Hervorhebungen im Original)

Ich tue so als glaube ich, dass die Banane ein Telefon ist.

Wenn die propositionalen Gehalte dieser Überzeugungen so in intensionale Kontexte eingebettet werden, lassen sich Leslie zufolge die befürchteten semantischen Schwierigkeiten vermeiden.⁴⁹

Einige Autoren haben gegen diese Theorie eingewendet, dass sie nur ein Scheinproblem löse und daher überflüssig sei. Ich teile diese Kritik. Die Schwäche von Leslies Theorie wird m.E. offenbar, wenn man betrachtet, wie er Perner's Erklärung des FBT-Alterseffekts kritisiert.⁵⁰ Perner's Theorie, sagt Leslie, zeichne Überzeugungen nur dadurch aus, dass sie eine Überzeugungsrelation zu Propositionen darstellen. Wünsche bestünden aber ebenfalls in einer bestimmten Relation zu Propositionen; Leslie spricht in *beiden* Fällen von semantischer Bewertung. Die *Art* der semantische Bewertung sei jedoch von Zustandstyp zu Zustandstyp jeweils eine andere.

What differs from case to case is the 'mode' of evaluation. The obvious temptation, then, is to add a simple qualification: Sally semantically evaluates *with respect to believing* a [proposition]. Certainly, this will do the job. By the same token, simply replacing 'semantically evaluates' with 'believes' will also do the job [...]. Unfortunately, this makes the theory-theory circular." (Leslie 2000, 215, Hervorhebung im Original)

Doch Perner charakterisiert die Überzeugungsrelation als ein *Für-wahr-halten* – eine Charakterisierung, die sie von allen anderen Einstellungsrelationen zwischen Personen und Propositionen unterscheidet. Warum reicht Leslie das nicht aus? Es scheint, als glaube er, dass propositionale Einstellungen nur anhand der intensionalen Operatoren unterscheidbar seien, die sie *enthalten*. Entsprechend postuliert seine Theorie keine *Pretend*-Überzeugungen mit besonderem kausalen Profil; er fasst *Pretend*-Überzeugungen vielmehr als *normale* Überzeugungen *über* einen *Pretend*-Zustand auf. Diese Theorie der Typindividuierung geistiger Zustände müsste sich verallgemeinern lassen. Zum Beispiel müssten auch normale Überzeugungen dadurch von anderen propositionalen Einstellungen unterschieden werden, dass sie den Überzeugungsbegriff

⁴⁹ Leslie scheint diese Operatoren als einfache, zusätzliche Kennzeichnung geistiger Zustände zu verstehen, die "logisch neutral" bleiben. Aus den abgekoppelten Überzeugungen lassen sich seiner Ansicht nach normale Schlussfolgerungen ziehen: "[W]e may say that the inference operates within the decoupled level." (Leslie 1994, 220)

⁵⁰ Zur kritisierten Erklärung Perner's siehe Abschnitt 3.1. Perner zufolge bestehen erst vierjährige Kinder den FBT, weil sie in diesem Alter den Überzeugungsbegriff erwerben.

enthalten. Eine Überzeugung, dass p , wäre nach Leslie zu analysieren als Überzeugung mit dem Gehalt

Person A glaubt, dass p .

Sonst bestünde die Gefahr, dass sie mit dem Wunsch, dass p , verwechselt wird. Diese absurde Konsequenz ziehen Nichols & Stich (2000) aus Leslies Theorie:

[I]t would be impossible to believe *anything* without having the concept of belief, and impossible to desire *anything* without having the concept of desire. (Nichols & Stich 2000, 141, Hervorhebungen im Original)

Ich bezweifle daher, dass sich über die logische Struktur hypothetischer Zustände klären lässt, wie sie zugleich vom normalen kognitiven Haushalt abgegrenzt und doch ähnlich verarbeitet werden wie ihre normalen Gegenstücke.

7.2 Die Fähigkeiten autistischer Kinder

Es gibt noch einen weiteren Aspekt Leslies Theorie, den ich gesondert diskutieren möchte. Er behauptet, dass es einen kognitiven Mechanismus gebe, der spezifisch für die Abgrenzung von *Pretend*-Überzeugungen verantwortlich ist (also Metarepräsentationen erzeugt). Diesen Aspekt seiner Theorie stützt er auf empirische Befunde, die ich in der Arbeit noch nicht vorgestellt habe. Seine Theorie eignet sich daher, um die in diesem Zusammenhang einschlägigen Daten vorzustellen. Sie betreffen hauptsächlich die kognitiven Leistungen und Defizite *autistischer* Kinder.

Autistische Kinder können nicht-psychologische Analoga des FBT gut lösen. Zum Beispiel verstehen sie, dass Landkarten oder Fotos veraltet sein können und daher die gegenwärtige Situation falsch repräsentieren (Leslie & Thaiss 1992). Sie fallen aber selbst lange nach ihrem vierten Geburtstag noch beim FBT durch. Die alltagspsychologischen Defizite autistischer Kinder im FBT wurden richtungweisend von Baron-Cohen, Leslie & Frith (1985) beschrieben. Einer Meta-Analyse von Happé (1995) zufolge bestehen (mit steigendem Alter) nur 15% bis 45% autistischer Kinder den FBT.⁵¹

⁵¹ Eine Übersicht über die aktuelle Forschung gibt die Aufsatzsammlung Baron-Cohen, Tager-Flusberg & Cohen (2000) sowie die Monographie von Baron-Cohen (1995).

Leslie erklärt dieses Defizit durch die Beschädigung des Abgrenzungsmechanismus'. Autistische Kinder können keine metarepräsentationalen Zustände ausbilden und scheitern daher an allen alltagspsychologischen Aufgaben. Zu dieser Erklärung passt auch der Befund, dass autistische Kinder Schwierigkeiten mit dem PP haben. Sie sind zwar im Prinzip in der Lage, einfaches PP auszuführen, wenn man sie dazu drängt, doch spontan tritt es in ihrem Verhalten nur selten auf. Das ist Leslie zufolge auf denselben beschädigten Mechanismus zurückzuführen: Autistische Kinder können die dazu benötigten hypothetischen Überzeugungsgehalte nicht angemessen repräsentieren.

Die Fähigkeit zum kontrafaktischen Schließen

Die Gefahr bei solchen Erklärungen korrelativer Befunde ist die, dass sie nicht hinreichend gegen Alternativen abgesichert werden. Leslie muss ausschließen, dass es sonstige Fähigkeiten autistischer Kinder gibt, die eine Abgrenzung in seinem Sinn erfordern. Einen Kandidaten dafür gibt es: die guten Fähigkeiten autistischer Kinder mit Schlussfolgerungen aus kontrafaktischen Prämissen.

Scott, Baron-Cohen & Leslie (1999) verwendeten die Syllogismen, die ich in der Vorbemerkung zu diesem Teil der Arbeit erwähnt habe, z.B.: 'Alle Fische leben in den Bäumen. Peter ist ein Fisch. Lebt Peter in den Bäumen?' Autistische Kinder sind bei solchen Aufgaben sogar besser als ihre normalen Altersgenossen.⁵² Peterson & Bowler (2000) prüften zudem autistische Kinder mit der (nicht-psychologischen) K-Frage: 'Wenn die Schokolade nicht verlegt worden wäre, wo wäre sie jetzt?' Diese Frage können autistische Kinder beantworten, obwohl sie beim FBT durchfallen.

Leslies Reaktion auf diese Befunde lautet: Kontrafaktisches Schließen ist keine alltagspsychologische Leistung und hat daher mit dem Abgrenzungsmechanismus nichts zu tun. Die Fähigkeit autistischer Kinder zum kontrafaktischen Schließen führt er vielmehr auf das Funktionieren eines anderen kognitiven Mechanismus' zurück, den er *Se-*

⁵² Da autistische Kinder ihren normalen Altersgenossen oft auch in der sonstigen kognitiven Entwicklung hinterherhinken, wird bei Vergleichen nicht das tatsächliche, sondern das *mentale* Alter der Kinder zugrunde gelegt. Das mentale Alter wird im Hinblick auf *verbale* Fähigkeiten durch Wortschatz- und Grammatiktests überprüft. (Untersuchungen mit nicht-autistischen, aber lernschwachen Kindern haben ergeben, dass diese den FBT im verbalen Alter von vier Jahren bestehen. Die alltagspsychologischen Defizite autistischer Kinder lassen sich daher nicht auf ihre allgemeinen kognitiven Defizite zurückführen.)

lection Processor nennt. Dieser Mechanismus grenzt kontrafaktische Prämissen und Konklusionen vom Hintergrundwissen ab, indem er widersprechende Überzeugungen ausblendet.

The [Selection Processor] hypothesis [locates] the competition [...] between two possible belief contents, one of which represents physical reality, the other not. (Roth & Leslie 1998, 5)

Roth & Leslie (1998, 2. Experiment) legten eine Murmel in einen von zwei unterschiedlich aussehenden Körben und versteckten diese Körbe hinter einem Schirm. Daraufhin stellten sie zwei ähnlich aussehende Körbe auf den Tisch und legten eine Murmel in den jeweils anderen Korb. Die Kinder wurden nun aufgefordert, die richtige Position der Murmel *hinter* dem Schirm anzugeben, und mussten dazu die Situation, die ihnen vor Augen ist, ausblenden. Keine leichte Aufgabe: Sie wird erst im Alter von vier Jahren beherrscht. Aber Autisten schneiden hier nicht schlechter ab als andere Kinder. Auch wenn es in diesem Experiment nicht um die Unterdrückung einer widersprechenden Überzeugung, sondern um die Unterdrückung einer ähnlichen Überzeugung geht, wird m.E. deutlich, welche Aufgabe der *Selection Processor* erfüllt.⁵³ Ich bezweifle jedoch, dass Leslies zwei Mechanismen das Entwicklungsmuster gut erklären.

Auch das normale PP – möchte ich einwenden – erfordert die Ausblendung salienter Aspekte. Ein Kind, das so tut als sei eine leere Tasse voll, muss seine Überzeugung über das ausblenden, was ihm vor Augen ist und mit der *Pretend*-Stipulation im Konflikt steht: die leere Tasse. In diesem Bereich sind autistische Kinder jedoch viel schlechter als beim kontrafaktischen Schließen. Für Leslie scheint sich dieses Problem nicht zu stellen; er geht meines Wissens in keinem seiner Aufsätze explizit darauf ein. Das Abgrenzen von *Pretend*-Überzeugungen ist seiner Theorie zufolge etwas anderes als das Ausblenden inkompatibler Überzeugungen.

Doch meines Erachtens sind das PP und das kontrafaktische Schließen keine voneinander unabhängigen Fähigkeiten. In beiden Fällen geht es um Überzeugungsgehalte, die nicht die Realität repräsentieren, und die daher eine besondere Rolle im kognitiven Haushalt spielen müssen. Warum in einem Fall (PP) eine eigene Semantik nötig sei –

⁵³ Genauer werde ich Leslies Mechanismen weiter unten in diesem Abschnitt vorstellen, wenn ich auf seine Erklärung des FBT-Alterseffekts zu sprechen komme. Für meine Kritik am *Selection Processor* reicht diese dürftige Skizze aus.

wie es Leslie behauptet – und im anderen Fall (kontrafaktisches Schließen) nicht, bleibt offen.

In der Forschungsliteratur sind weitere Vorschläge Erklärung des beschriebenen Entwicklungsmusters gemacht worden. Peterson & Bowler (2000) schlagen zum Beispiel vor, der entscheidende Unterschied zwischen FBT und ihrem K-Frage-Test liege darin, dass im FBT die K-Frage nicht explizit gestellt werde. Die Kinder müssten sich selbst klar machen, dass aus Sicht des Protagonisten das Objekt nicht verletzt worden ist. Da autistische Kinder für ihre mangelnde Kreativität und ihr stereotypes Verhalten bekannt sind, sei es möglich, dass sie am FBT scheitern, weil sie nicht auf die Idee kommen, sich die K-Frage zu stellen.

[False belief] tasks [...] require spontaneous generation of counterfactual suppositions. The fact that such tasks are harder for children with autism suggests that their failure may be linked to impaired spontaneous generativity. (Peterson & Bowler 2000, 403)

Doch diese Erklärung klingt zu dürftig. Wieso sollte mangelnde Kreativität dafür verantwortlich sein, dass sich autistische Kinder die K-Frage im FBT nicht stellen? Es ist auch denkbar, dass autistische Kinder das kontrafaktische Antezedens der K-Frage nicht begreifen, weil ihnen das nötige alltagspsychologische Wissen fehlt.

Ich zweifle allerdings an der These, dass Autisten keine alltagspsychologische Theorie ausbilden können, weil m.E. nicht klar ist, warum sie gerade in *diesem* (und nicht in irgendeinem anderen) Wissensbereich Schwierigkeiten haben. Meine eigene Spekulation geht dahin, dass Autisten einen engen Aufmerksamkeitsfokus haben. Daher bereitet es ihnen keine Probleme, ihr Hintergrundwissen beiseite zu lassen, während sie eine kontrafaktische Aufgabe lösen. Die Widersprüche fallen in ihrem Denken nicht ins Gewicht. Doch es fällt ihnen schwer, ihr Hintergrundwissen und ein hypothetisches Szenario *gleichzeitig* im Blick zu haben – wie es beispielsweise beim PP und auch beim FBT nötig ist. Daher ist es für sie eine besondere Herausforderung, den Wissensstand anderer Menschen zu berücksichtigen. Doch diese Spekulation möchte ich hier nicht weiter treiben, sondern mich stattdessen im nächsten Abschnitt mit den kontrafaktischen Schlüssen *normaler* Kinder befassen. Dadurch will ich versuchen zu klären, wie das kontrafaktische Schließen, der FBT und das PP zusammenhängen.

Leslies Erklärung des FBT-Alterseffekts

Der Vollständigkeit halber möchte ich noch erwähnen, dass Leslie im Rahmen seiner Theorie auch den Alterseffekt im FBT erklärt. Nach Scholl & Leslie (1999) ist die Entwicklung (Reifung) des *Selection Processors* für Fortschritte in der Entwicklung der alltagspsychologischen Fähigkeiten verantwortlich. Sie erklärt, warum erst Vierjährige den FBT bestehen. Leslie zählt seine Position daher zu den *Early-Competence-Accounts*: Junge Kinder verstehen bereits, dass es falsche Überzeugungen geben kann, weil ihr Abgrenzungsmechanismus schon funktioniert (etwa für das PP Zweijähriger), doch sie sind bis zu einem bestimmten Alter, nicht in der Lage, dieses Verständnis zu zeigen.

The standard false-belief task [...] places (at least) two demands on the 3-year-old: (a) the metarepresentation must be computed, and (b) the correct content of the belief must be selected. [...] [T]he latter [is a job] for [the Selection Processor]. [...] Basically, then, passing the standard false-belief task requires both a conceptual competence and a performance capability, and failure might simply reflect an immature [Selection Processor]. (Scholl & Leslie 1999, 147f.)

Leslie stützt seine Inhibitions-Theorie auf ein begriffliches und ein empirisches Argument. Das begriffliche Argument baut darauf auf, dass es Sinn macht, anderen Menschen wahre Überzeugungen zuzuschreiben. Damit begründet Leslie die Existenz einer *Default-Heuristik*, anderen Menschen – sofern nichts dagegen spricht – wahre Überzeugungen zuzuschreiben. Es ist die Aufgabe des *Selection Processors*, die Anwendung dieser Heuristik zu unterbinden.

There is a conceptual basis for the existence of this pre-potent response. *Normatively* beliefs are true: this is what beliefs are "for"; they are "for" accurately describing the world; a belief is "useful" to an agent only to the extent it is true; in short, beliefs "ought" to be true. (Leslie 1994, 228f., Hervorhebung im Original)

Damit scheint Leslie auf ein Argument abzielen, dass über die bloße Feststellung, Überzeugungen seien häufig wahr, hinausgeht. Vermutlich meint er, dass sinnvolles Handeln nicht möglich wäre, wenn nicht ein Großteil unserer Überzeugungen wahr wäre. Aus diesem Grund ist es seiner Ansicht nach sinnvoll, dass junge Kinder in der Regel davon ausgehen, dass Menschen wahre Überzeugungen haben. Erst im Alter von vier Jahren lernen sie, dass es wichtige Ausnahmen gibt (vor allem Fehler und Täuschungsversuche).

Doch Leslie schätzt die Bedeutung dieses Arguments falsch ein. Was es meines Erachtens zeigt ist, dass man nicht überrascht zu sein braucht, dass Kinder erst so spät lernen, mit falschen Überzeugungen umzugehen. Schließlich spielen falsche Überzeugungen in unserem kognitiven Leben – von vergleichsweise spektakulären Ausnahmen abgesehen – nur eine untergeordnete Rolle. Was das Argument aber nicht zeigt, ist dass junge Kinder eine Heuristik anwenden, nach der alle Überzeugungen wahr sind. Vielmehr beschreibt es das zu erklärende Phänomen. *Jede* Erklärung der alltagspsychologischen Fähigkeit, mit Überzeugungen umzugehen, muss darüber Auskunft geben, warum Kinder und Erwachsene anderen Menschen in der Regel wahre Überzeugungen zuschreiben. Leslies Erklärung, dass sie eine Heuristik mit der gewünschten Wirkung anwenden, ist unbefriedigend.⁵⁴

7.3 Kontrafaktisches Schließen

Peterson & Riggs (1999) haben den FBT als kontrafaktisches Problem charakterisiert und auf diese Weise den Alterseffekt im Sinn eines Simulationsmodells erklärt. In diesem Abschnitt werde ich ihre Erklärung diskutieren. Ich glaube, dass die Autoren sowohl mit ihrer Rekonstruktion des FBT als auch mit der Erklärung seines Alterseffekts auf der richtigen Spur sind. Doch ich werde am Ende des Abschnitts auf einen problematischen Punkt aufmerksam machen.

Modifizierte Schlussfolgerungsprozesse

Als kognitive Grundlage der Simulation schlagen Peterson & Riggs einen Prozess vor, den sie 'modified derivation' nennen. Die Idee ist die, bei der Vorhersage von Schlussfolgerungen hypothetische Prämissen als Modifikationen des normalen Schlussfolge-

⁵⁴ Fodor (1992) vertritt eine ähnliche Position über die anzuwendende Heuristik, begründet sie aber besser: Die Heuristik anzuwenden sei kognitiv *weniger aufwändig* und eigne sich daher vor allem für kognitiv noch nicht allzu belastbare junge Kinder. (Fodor fügt eine Bedingung hinzu: Die Heuristik werde nur angewendet, solange sie eine *eindeutige* Vorhersage des Verhaltens liefere; mehrdeutige Vorhersagen würden auch jüngere Kinder dazu bewegen, ihre ganze alltagspsychologische Kompetenz zum Tragen zu bringen. Wimmer & Weichbold (1994) gelangten bei einer experimentellen Überprüfung dieser These jedoch zu einem negativen Ergebnis.)

rungsprozesses zu betrachten. Die Prämissen bleiben dabei vom allgemeinen Überzeugungssystem abgegrenzt, indem sie nur in einzelne Schlussfolgerungen einfließen. In Fällen, in denen eine zusätzliche Überzeugung berücksichtigt werden muss, ist die erforderliche Modifikation einfach: Die Überzeugung wird – für diese Schlussfolgerung – als zusätzliche Prämisse betrachtet. In Fällen, in denen jedoch eine Überzeugung *ausgeblendet* werden muss – wie zum Beispiel im FBT – wird diese Prämisse während der Schlussfolgerung *unterdrückt*.

We [...] consider a strategy, which we call *modified derivation (MD)*, which requires not actual updates to our database but modifications of the process of answer-derivation. In order to effect the subtraction of a fact F, we impose the *modification-instruction* 'ignore(F)' [...]. (Peterson & Riggs 1999, 82, Hervorhebungen im Original)

Der Abgrenzungsmechanismus besteht also darin, dass einzelne Überzeugungen besondere Berücksichtigung erfahren und als Prämisse eingeführt bzw. ausgeblendet werden; es handelt sich um eine Art von Selbstkontrolle. Der sonstige Schlussfolgerungsprozess bleibt davon unberührt, d.h. es werden die üblichen Schlussfolgerungsmechanismen angewendet und es wird das relevante Hintergrundwissen herangezogen. Peterson & Riggs sagen zwar nicht, wie sich der hypothetische Status der Prämissen auf die Konklusion "vererbt". Doch ich nehme an, dass die für die Modifikation erforderliche Selbstkontrolle ein hohes Maß an Aufmerksamkeit mit sich bringt, so dass die ganze Ableitung als ein besonderes Unterfangen ausgezeichnet ist.⁵⁵

Das Verfahren der modifizierten Schlussfolgerungen möchte ich anhand eines der Beispiele der Autoren erläutern und dabei auf die Frage zu sprechen kommen, inwiefern es eine Erklärung im Sinn des Simulationsmodells ist.

Consider a case in which Maria parks her car outside her place of work at time t_1 , and then goes into work. Maria is aware of the *base-fact* that she parked her car outside her place of work at time t_1 , and she is also aware of the background *norm* that things generally stay where they are put. [...] Thus if, at time t_3 , we ask the question 'Where is your car?', [...] in the absence of evidence to the contrary, it is reasonable to give this *default answer*, 'Outside work'. (Peterson & Riggs 1999, 83, Hervorhebungen im Original)

⁵⁵ Der Zusammenhang alltagspsychologischer Fähigkeiten zur allgemeinen Selbstkontrolle ist in den letzten Jahren stärker ins Blickfeld der Forschung gerückt. Carlson & Moses (2001) konnten zum Beispiel zeigen, dass eine gute Fähigkeit der Unterdrückungskontrolle (*inhibitory control*) ein früheres Bestehen des FBT ermöglicht. Die Fähigkeit zur Unterdrückung wird zum Beispiel durch Spiele getestet, in denen Kinder die Bewegungen des Spielleiters nachahmen sollen – aber in einer entgegengesetzten Ausführung. Auf diese Forschung werde ich im folgenden nicht eingehen, da dies die Ergebnisse dieser Arbeit (soweit ich sehe) nicht verändern würde. Ich weise jedoch auf den Überblicksartikel von Perner & Lang (2000) hin.

Wenn nun zum Zeitpunkt t_2 Marias Auto abgeschleppt worden wäre, und Person A dies gesehen hätte, Maria jedoch nicht – wie wäre dann zu erklären, wie Person A darauf kommt, dass Maria immer noch an ihrer alten Überzeugung über den Standort ihres Autos festhält? Peterson & Riggs zufolge wendet Person A zunächst eine alltagspsychologische *Theorie* an, um zu ermitteln, dass Maria nicht weiß, dass ihr Auto abgeschleppt worden ist. Als Theorie kommt die Theorie des epistemischen Zugangs in Frage, die ich im Abschnitt 2.2 vorgestellt habe.

Anschließend – und das ist der *Simulationsteil* der Erklärung – stellt sich Person A die Frage, wo das Auto stehen würde, wenn man – wie Maria – nicht weiß, dass es abgeschleppt worden ist. Zur Beantwortung dieser Frage zieht man sowohl sein Wissen darüber heran, dass Maria ihr Auto auf dem Parkplatz abgestellt hat (die Autoren nennen das eine "Basis-Tatsache"), und das Wissen darüber, dass die Dinge im Allgemeinen dort bleiben, wo man sie hinstellt ("Norm"). Auf diese Weise vermeiden Peterson & Riggs, eine Aktualisierungstheorie von Überzeugungen zu postulieren, die ich im Abschnitt 3.2 angeführt habe. Person A muss keine *psychologische* Theorie darüber anwenden, dass sich Marias Überzeugungen *ceteris paribus* nicht ändern, sondern kann auf die *physikalische* Theorie zurückgreifen, dass sich die Dinge *ceteris paribus* nicht ändern.

Den Alterseffekt im FBT erklären Peterson & Riggs nun damit, dass dreijährige Kinder die Unterdrückung ihrer wahren Überzeugung über den neuen Ort der Schokolade nicht implementieren oder nicht durchhalten können. Sie beantworten die Testfrage daher unter Rückgriff auf ihre Überzeugung über den aktuellen Stand der Dinge.

Here [i.e. in the failed false belief test] MD has been attempted, but has failed, since the modification-instruction [i.e. ignore (The chocolate was moved to the cupboard at time t_2)] has not been sustained. As a result we get the answer appropriate to the given database without the modification-instruction. (Peterson & Riggs 1999, 90)

Ihre Erklärung stützen die Autoren mit einer hohen Korrelation zwischen dem Bestehen des FBT und dem Bestehen eines nicht-psychologischer *Analogons* zum FBT: dem Beantworten der K-Frage 'Wo wäre die Schokolade, wenn sie nicht verlegt worden wäre?'. Es scheint demnach nur die Unterdrückung der Überzeugung relevant zu sein, nicht aber die Tatsache, dass es sich um einen geistigen Zustand handelt, der unterdrückt wird. Damit widersprechen Peterson & Riggs den Vertretern des Theorienmodells.

The MD Hypothesis has the implication that the presence or absence of the mentalistic component in the [False Belief] Task will not substantially affect children's success at the age in question. The 'theory of mind hypothesis', in contrast, has the opposite implication, since the mentalistic component is the thing which the participant's theory of mind is required to deal with. (Peterson & Riggs 1999, 93)

Ich möchte nun untersuchen, ob das Aufkommen der Fähigkeit, Überzeugungen gezielt auszublenden, die Alterseffekte beim FBT und beim kontrafaktischen Schließen erklärt. Ich werde dafür argumentieren, dass das Ausblenden nicht die einzige und auch nicht die entscheidende kognitive Grundlage zum Bestehen des FBT darstellen kann.

Kontrafaktisches Schließen und Pretend Play

Das Problem dieser Erklärung ist, dass die angeblich im Alter von vier Jahren erworbene Kompetenz möglicherweise schon zur Erklärung bereits früher auftretender Fähigkeiten notwendig ist. Einen in diesem Sinn problematischen Kandidaten gibt es: Die Fähigkeit, Schlussfolgerungsprozesse zu modifizieren, scheint schon zur Ausübung des normalen PP nötig zu sein. Kausale Schlüsse innerhalb eines *Pretend*-Szenarios lassen sich meines Erachtens gut als kontrafaktische Schlüsse beschreiben. Um zu verstehen, was passiert, wenn *Naughty Teddy* ein leeres Glas über dem Kopf einer anderen Puppe umdreht, muss man schließen können, was der Fall wäre, wenn *Naughty Teddy* ein volles Glas umdrehen würde.

Die Autoren machen zwar einen Unterschied zwischen ihrer Simulationsstrategie und dem PP aus. Im PP gehe es darum, Objekten oder Personen Rollen zuzuordnen, die sie eigentlich nicht innehaben: etwa wenn eine Banane als Telefon betrachtet wird. Diese Zuordnungen würden einem *anderen Zweck* dienen als die Annahme kontrafaktischer Prämissen in Schlussfolgerungen: dem Erzählen einer Geschichte. Dieser Unterschied begründet Peterson & Riggs zufolge, dass es sich beim kontrafaktischen Schließen und PP um verschiedene Fähigkeiten handelt, die unterschiedliche kognitive Erklärungen erfordern. Doch das Zuordnen von Rollen ist nicht die einzige Form des PP, die junge Kinder beherrschen. Zweijährige Kinder können auch kausale Abfolgen innerhalb eines *Pretend*-Szenarios nachvollziehen. Das Ausblenden von Überzeugungen (bzw. die Modifikation der Ableitung) kann daher nicht die letzte Hürde sein, die Kinder nehmen müssen, um den FBT zu bestehen.

Zum Schluss dieses Kapitels möchte ich noch eine neuere Theorie von Nichols & Stich (2003) vorstellen, die erklärt, wie Kinder im FBT ihr relevantes Hintergrundwissen her-

anziehen. Dieser Theorie zufolge führen Vierjährige gewissermaßen Buch über die Überzeugungen des Protagonisten. Die Autoren nennen das Buch 'Possible World Box Model'. Wenn sie erkennen, dass der Protagonist nicht sieht, dass die Schokolade an einen anderen Ort gelegt wird, nehmen sie in ihrem Buch keine Änderung vor. Wenn sie dann danach gefragt werden, wo der Protagonist die Schokolade vermutet, müssen sie daher in ihrem Buch bloß die richtige Antwort nachschlagen.

[T]he child does not transfer the belief that the chocolate is in location 2 to the PWB model, and the child's prediction is based on the pre-existing PWB model which includes the belief that the chocolate is in location 1. (Nichols & Stich 2003, 89)

Der Preis für diese Erklärung der Fähigkeit, den FBT zu lösen, ist meines Erachtens hoch: Da Kinder die Frage des Versuchsleiters nicht vorhersehen können, müssen sie für jeden Menschen, den sie beobachten, ein solches Buch führen. Viel einfacher wäre ein Verfahren, mit dem sich *ad hoc* die Überzeugung des Protagonisten ermitteln lässt. Ein solches Verfahren werde ich im nächsten Kapitel entwickeln.

8 Empirische Indikatoren für das Simulationsmodell

8.1 Der Umgang mit kontrafaktischen Prämissen

Bei der Diskussion im letzten Kapitel deutlich geworden, dass eine Erklärung für den folgenden Befund gesucht wird: Wieso können Kinder erst im Alter von vier Jahren kontrafaktisch schließen, wenn doch die nötige Kompetenz mit falschen Prämissen umzugehen, bereits mit dem normalen PP im Alter von zwei Jahren vorliegt. Kommt mit vier Jahren eine weitere Kompetenz ins Spiel oder werden lediglich Performanzschwierigkeiten überwunden? In diesem Kapitel werde ich weitere Daten zu diesem Thema vorstellen, deren Erklärung die Grundlage für das Experiment darstellen wird, das ich im dritten Abschnitt entwickeln werde. Die Lösung wird einen Schlüssel für den Nachweis von alltagspsychologischen Simulationsprozessen liefern.

Vertieftsein in hypothetische Szenarien

Die erste Studie, die ich vorstellen möchte, wurde von Perner und Feichtinger durchgeführt (berichtet in: Perner 2000, 391). Die Autoren konnten zeigen, dass junge Kinder durchaus in der Lage sind, mit einer bestimmten Sorte hypothetischer Schlüsse umzugehen: mit hypothetischen Blicken in die Zukunft. Kinder wurden in ein Szenario eingeführt, in dem von verschiedenen kombinierten Bahn- und Busstationen aus verschiedene Ziele zu erreichen waren. Fast alle der drei- bis fünfjährigen Versuchspersonen konnten richtig vorhersagen, wo eine Person landen würde, wenn sie von einer bestimmten Station den Bus oder die Bahn nähme. Doch die Leistungen dreijähriger Kinder fielen bei Fragen der Art 'Wenn Person A vom Ort X nicht den Bus, sondern die Bahn genommen hätte, wohin wäre sie gelangt?' deutlich ab. Richtige Antworten auf solche Fragen korrelierten zudem mit dem Bestehen des FBT.

Nach Perner ist das folgende Merkmal der Aufgabenstellung für den Effekt verantwortlich: die Klarheit des Verweises auf die reale oder auf eine fiktionale Welt. Im Fall des Blicks in die Zukunft ist klar, dass eine Entwicklung in einer fiktionalen Welt zu bestimmen ist, die durch eine Stipulation konstituiert wird. Fordert man Kinder auf, eine Entwicklung vorauszusagen, werden sie von ihrer Repräsentation der realen Welt aus-

gehen, die Stipulation hinzufügen und ansonsten ihr Hintergrundwissen heranziehen, um die Aufgabe zu lösen. Da Aussagen über die Zukunft nicht mit Überzeugungen über den gegenwärtigen Stand der Dinge in Konflikt geraten können, sind Missverständnisse darüber, worauf sich die Aufgabenstellung bezieht, ausgeschlossen.

Worin die Schwierigkeit bei einem *unklaren* Verweis bestehen kann, lässt sich am Beispiel des normalen PP erläutern. Im PP stoßen fiktionale und reale Welt aufeinander: Eine reale Banane wird als Telefon benutzt oder der vorgestellte Inhalt einer leeren Tasse ausgeschüttet. Stellt man zum *Pretend*-Szenario Fragen – etwa nach dem Zustand eines Objekts – verstehen auch junge Kinder in der Regel, dass sich die Frage auf den Als-ob-Zustand des Objekts bezieht. Perner erklärt diesen Befund dadurch, dass sie die Fragen als Fragen zum *Pretend*-Szenario verstünden, solange sie in das Szenario *vertieft* seien. Bezugnehmend auf die Untersuchungen von Harris & Kavanaugh (1993) zu den Fähigkeiten junger Kinder, kausale Abfolgen innerhalb eines *Pretend*-Szenarios zu verstehen, schreibt er:

The trick in Harris and Kavanaugh's procedure is to keep the child engrossed in the pretend world so that the natural interpretation of the critical question is within the pretence. The child's ability to keep reference within the pretence is, indeed, volatile. Harris and Kavanaugh (1993) had to let the child get actively involved in spelling out the pretend consequences (e.g. "Teddy made the chocolate all ...?") so that the critical contrast question (Is the chocolate now wet or dry?) appears to be asked and is answered within the pretend world. (Perner 2000, 393, meine Hervorhebung)

Doch das Vertieftsein in ein *Pretend*-Szenario ist leicht gestört. Lillard & Flavell (1992, 2. Studie) stellten fest, dass junge Kinder unter bestimmten Umständen nicht richtig angeben können, was sich eine Person gerade eingebildet hat – nämlich dann, wenn das PP nicht von angemessenem *Pretend*-Verhalten begleitet wird. In der Studie wurde Kindern eine Puppe vorgeführt, von der ihnen gesagt wurde, sie tue so als sei eine Flasche Saft im Regal. Anschließend sahen die Kinder, dass nur ein Teddy im Regal saß. Solange die Puppe ihre Stipulation nicht durch entsprechendes Verhalten untermauerte (z.B. mit einem leeren Glas zum Regal ging), konnten sie die Stipulation anschließend nicht mehr richtig wiedergeben. Ohne das begleitende *Pretend*-Verhalten, argumentiert Perner, geht den Kindern leichter der Faden verloren, so dass sie die anschließende Frage nach der *Pretend*-Überzeugung der Puppe nicht mehr als Frage nach einer Überzeugung innerhalb des *Pretend*-Szenarios verstehen können.

Bei kontrafaktischen Schlüssen haben jüngere Kinder nach Perner ähnliche Schwierigkeiten, den Bezug der Aufgabenstellung zu verstehen. Auch hier müssen Kinder zwi-

schen der realen Welt und einem hypothetischen Szenario unterscheiden. Im Experiment von Perner und Feichtinger sollten dreijährige Kinder bestimmen, wo eine Person *jetzt* wäre, wenn sie statt dem Bus die Bahn genommen hätte. Entscheidend für die Leistung des Kindes sei, diese Frage als eine Frage nach dem *hypothetischen* Aufenthaltsort der Person aufzufassen – obwohl nach dem *gegenwärtigen* Aufenthaltsort gefragt wurde, was als Frage nach dem *realen* Aufenthaltsort missverstanden werden kann.

You first have to think back to a past state of the real world: Peter at station A. Construct a counterfactual world of him taking the train instead of the bus. [...] Then comes the question, referring to Peter's present location [i.e.: "Where would Peter be now?"]. How does the child know that it is the counterfactual location? (Perner 2000, 393)

Während Kinder beim PP meist in das Szenario vertieft sind, und dies durch explizite Stipulationen und gegenseitige Aufforderungen zum Spielen untermauert wird, fehlt ihnen diese Hilfestellung beim kontrafaktischen Schließen. Erst im Alter von vier Jahren gelingt es ihnen, den Bezug der Frage zum hypothetischen Szenario herzustellen. Im folgenden werde ich diese Analyse ausbauen, um zu erklären, warum dreijährige Kinder beim FBT durchfallen.

Der Erleichterungseffekt im kontrafaktischen Schließen

Wie ich bereits in der Vorbemerkung zu diesem Teil der Arbeit erwähnt habe, scheitern dreijährige Kinder an kontrafaktischen Syllogismen der Art 'Alle Fische leben in den Bäumen. Peter ist ein Fisch. Lebt Peter in den Bäumen?' Stattdessen greifen sie auf ihr einschlägiges Hintergrundwissen zurück, wenn sie solche Frage beantworten. Dias & Harris (1988, 1990) konnten jedoch zeigen, dass die Schwierigkeiten dreijähriger Kinder verschwinden, wenn man dem Syllogismus die Aufforderung voranstellt, sich bildlich vorzustellen, dass Fische in den Bäumen leben. Denselben Effekt erreicht man auch durch Verändern des Tonfalls oder durch den Hinweis, es handele sich um Informationen über einen fremden Planeten. Scott et al. (1999) konnten den Effekt replizieren.⁵⁶

Dias & Harris erklären diesen Erleichterungseffekt folgendermaßen: Da die verschiedenen Einleitungsphrasen alle darauf abzielen, einen imaginären Kontext zu schaffen (d.h. ein *Pretend*-Szenario aufzubauen), und alle dieselbe Wirkung haben, weist das auf eine

⁵⁶ Die Autoren erhielten den Effekt jedoch nur für normale, nicht aber für autistische Kinder. Siehe dazu Abschnitt 7.2.

besondere Einstellung zu kontrafaktischen Prämissen hin: Das PP scheint dreijährigen Kindern zu ermöglichen, die kontrafaktischen Prämissen bei der Beantwortung der Frage zu berücksichtigen. Wenn die kontrafaktischen Prämissen *als hypothetische Stipulationen* präsentiert werden, fällt es jüngeren Kindern offenbar leichter – werden sie hingegen neutral, d.h. *als normale Information* präsentiert, fällt es ihnen schwerer, die Frage darauf zu beziehen. Es sieht also so aus, als verfügten dreijährige Kinder zwar über die Kompetenz kontrafaktische Prämissen zu verarbeiten. Sie können jedoch noch nicht alle Informationen selbständig als kontrafaktische Prämissen einstufen, die eine besondere Berücksichtigung erfordern. Erleichtert der Versuchsleiter ihnen die Aufgabe, die kontrafaktischen Prämissen als *Pretend*-Stipulationen aufzufassen, kommt ihre Kompetenz jedoch zum Vorschein. Die Kompetenz, um die es dabei geht, ist die Fähigkeit zum normalen PP: Kinder nehmen *Pretend*-Stipulationen ernst und blenden ihre widersprechenden wahren Überzeugungen aus.

A plausible interpretation is that these various make-believe cues, notwithstanding their heterogeneity, all converge in producing a similar psychological stance. They prompt children to treat the problems as descriptions of an imaginary world; this imaginary world contains creatures and events that violate their everyday knowledge – fishes that live in trees and cats that bark. Yet precisely because children have knowingly entered an imaginary world, empirical considerations are set aside and do not interfere with their acceptance of the initial premise. (Harris 2000, 105f.)

Der Erleichterungseffekt zeigt nach dieser Interpretation eine Performanzschwierigkeit auf; vierjährige Kinder erwerben keine zusätzliche Kompetenz für kontrafaktisches Schließen, sondern vielmehr die Fähigkeit, den bereits vorhandenen Mechanismus zu anderen Zwecken als dem normalen PP einzusetzen.

Der Interpretation von Dias und Harris möchte ich mich anschließen. Bevor ich diesen Erklärungsansatz für meine Zwecke weiterverfolge, möchte ich auf die Kritik eingehen, die Harris in späteren Arbeiten an diesem Ansatz geübt hat. Ein erster Grund für seine Kritik ist, dass der Erleichterungseffekt zu stabil sei, um sich auf das Vertieftsein in ein flüchtiges *Pretend*-Szenario zurückführen zu lassen. Leavers & Harris (1999, 2000) konnten zeigen, dass sich die Leistungssteigerung bei syllogistischen Aufgaben auch zwei bis drei Wochen nach dem ersten Test noch messen lässt, obwohl die kontrafaktischen Prämissen in der zweiten Testphase *neutral* präsentiert werden.

It makes little sense to claim that make-believe cues prompt children to adopt a pretend stance, and that they hold onto that stance or attitude until they meet the interviewer again. Conceivably, the second encounter with the interviewer might somehow reactivate a pretend stance. For example, children might

think of the interviewer as a kind of storyteller or mythmaker, but this interpretation is post-hoc, at best. (Harris 2000, 108f.)

Harris schlägt als mögliche Erklärung vor, dass sich Kinder anhand von Situationsmerkmalen (zum Beispiel: derselbe Versuchsleiter) daran erinnern, dass sie in dieser Situation ihre *Pretend*-Einstellung einnehmen können. Kinder kennen schließlich eine Reihe solcher Situationsmerkmale: Beim Hören von Märchen und im Theater kommt ihre *Pretend*-Einstellung auch ohne explizite Aufforderung zum Tragen. Harris selbst bezeichnet diese Erklärung als *post hoc*, doch ich teile den damit verbundenen Pessimismus nicht. Es muss nicht viel dazu gehören, Kinder in ein *Pretend*-Szenario zu involvieren.

Unverständliche Prämissen

Doch Leever & Harris werfen noch ein weiteres Problem für die ursprüngliche Erklärung des Erleichterungseffekts auf: Durch die Aufforderung zum PP lässt sich auch die Leistung bei Aufgaben erhöhen, deren Prämissen die Kinder *nicht verstehen*. Eine der verwendeten Prämissen lautete: "Alle Pongodaps haben Streifen auf ihrem Rücken." Dieser Befund ist problematisch, weil es in den genannten Fällen keine Überzeugungen gibt, die der kontrafaktischen Prämisse widersprechen und die ausgeblendet werden müssten. Es ist daher nicht ohne weiteres einzusehen, wie das PP dabei helfen könnte, die kontrafaktischen Prämissen zu verarbeiten.

Die alternative Erklärung, die Leever & Harris vorschlagen, lautet: Erst im Alter von vier Jahren verstehen Kinder die Aufgabenstellung als eine besondere *Dialogform*, in der die Prämissen besonders zu berücksichtigen sind. Außerhalb dieses kommunikativen Kontextes neigen jüngere Kinder dazu, ungewöhnliche oder unverständliche Prämissen zu ignorieren.

[L]ooking back at the various make-believe cues that we employed – the use of dramatic intonation, the reference to a far-off planet, the instruction to use imagery to make a mental picture of the initial premise – all these cues were likely to prompt children to treat the initial premise as an important starting point – a stipulation or premise that needed to be accepted as true, however temporarily. By contrast, when the interviewer began with a false or unfamiliar statement presented in a matter-of-fact tone, children may not have realized that the statement was an important presupposition for what was to follow, and failed to integrate it with the premises that followed. (Harris 2000, 111)

Leevers & Harris (1999) konnten dementsprechend auch zeigen, dass auch der alleinige Hinweis, über die erste Prämisse *nachzudenken*, denselben (stabilen) Effekt hat wie die Präsentation als *Pretend*-Stipulation. Es reicht also aus, dreijährige Kinder dazu aufzufordern, die Prämisse *als solche* zu verwenden, um den Erleichterungseffekt zu erhalten. Die Performanzschwierigkeiten jüngerer Kinder lassen sich demnach darauf zurückführen, dass sie die kontrafaktische Prämisse nicht für relevant erachten und daher für den zu ziehenden Schluss nicht nutzen.

Wie verhält sich diese Erklärung zu Perner's Vorschlag aus dem letzten Abschnitt, dass dreijährige Kinder den Bezug der Frage missverstehen? Meines Erachtens lassen sich beide Erklärungen gut kombinieren: Dreijährige Kinder verstehen nicht, dass sich die Testfrage auf ein *Pretend*-Szenario bezieht, weil sie die Prämisse nicht als Stipulation aufgefasst und daher auch kein *Pretend*-Szenario aufgebaut haben. Sie beachten Stipulationen nur, wenn sie ausdrücklich als wichtige Prämisse präsentiert werden. In solchen Fällen sind sie allerdings in der Lage, ein entsprechendes Szenario aufzubauen, auf das sie die Testfrage schließlich beziehen können.

In der bisherigen Diskussion habe ich den FBT-Alterseffekt ausgeklammert. Im folgenden Abschnitt werde ich aber der Frage nachgehen, ob die Fähigkeit Vierjähriger, bestimmte Informationen selbständig als *Pretend*-Stipulationen aufzufassen, für den FBT-Alterseffekt entscheidend sein könnte. Ich werde einen Test vorschlagen, der mir geeignet erscheint, eine Antwort auf diese Frage zu liefern.

8.2 Zwei neue Varianten des *False Belief Tests*

Für den letzten Abschnitt der Arbeit habe ich einen potenziellen Nachweis von Simulationsprozessen im Zusammenhang mit der Zuschreibung falscher Überzeugungen in Aussicht gestellt. Die Grundidee ist die folgende: Das kontrafaktische Schließen und der FBT hängen so stark zusammen, dass der Erleichterungseffekt auch im Umgang mit falschen Überzeugungen zu beobachten ist. Sollte ich mit dieser Behauptung Recht haben, dann würde das die Bedeutung des PP – und damit des Simulierens – für die Alltagspsychologie unterstreichen.

Bevor ich den entsprechenden Test vorstelle, möchte ich den Zusammenhang zwischen kontrafaktischem Schließen und dem FBT verdeutlichen. Zu diesem Zweck greife ich

auf die Analyse von Peterson & Riggs (1999) zurück. Der FBT ist dadurch zu lösen, dass man sich die kontrafaktische Frage

(K-Frage) Wo wäre die Schokolade jetzt, wenn sie nicht verlegt worden wäre?

zum Zweck einer Simulation des Protagonisten stellt.⁵⁷ Denn die Beantwortung dieser Frage führt zur Zuschreibung der falschen Überzeugung und stellt vermutlich den schwierigsten Teil des FBT dar. Bereits Dreijährige können auf der Grundlage einer Überzeugung das richtige Verhalten vorhersagen (siehe die Janes-Katzen-Experimente aus dem Abschnitt 2.1); möglicherweise verfügen sie sogar über eine Theorie der Überzeugungsbildung, die ihnen die

(K-Prämisse) Die Schokolade ist nicht verlegt worden.

als Ausgangspunkt einer Simulation des Protagonisten liefert (siehe Abschnitt 2.2). Zudem korrelieren die Alterseffekte beim kontrafaktischen Schließen und FBT.

Ich möchte nun eine Variante des FBT vorstellen: den P-FBT (für: *Pretend False Belief Test*). Mit diesem Test lässt sich meines Erachtens untersuchen, ob sich die Leistungen dreijähriger Kinder dadurch verbessern lassen, dass sie aufgefordert werden, die für die K-Frage relevante kontrafaktische Prämisse *als Prämisse ernst zu nehmen*. Dazu müsste man eine der Einleitungssphrasen von Dias & Harris aus dem letzten Abschnitt verwenden. Beispielsweise könnte man die Kinder dazu auffordern, so zu tun als wüssten sie nicht, dass die Schokolade verlegt worden ist. Besser wäre noch sie aufzufordern, so zu tun als hätten sie nicht *gesehen*, dass die Schokolade verlegt worden ist. Denn diese Fragestellung nimmt ihnen nicht die Aufgabe ab festzustellen, dass der Protagonist nicht weiß, dass die Schokolade verlegt worden ist. Vielmehr nimmt sie ihnen lediglich die Aufgabe ab festzustellen, dass der Protagonist nicht gesehen hat, dass die Schokolade verlegt worden ist. Doch das können sie bereits im Alter von zwei Jahren (siehe die Befunde zum *Level-1*-Verständnis visueller Wahrnehmung im Abschnitt 2.2). Im Hinblick auf die Studie von Leevers & Harris (1999) kann man auch versuchen, die Kinder

⁵⁷ Siehe dazu Abschnitt 7.3. Im Unterschied zu den Autoren erkläre ich den Alterseffekt jedoch anders. Peterson & Riggs haben einen Prozess spezifiziert (den sie 'modified derivation' nennen), der ihrer Meinung nach erst im Alter von vier Jahren beherrscht wird. Ich behaupte dagegen, dass dieser Prozess bereits zur Ausübung des PP benötigt wird. Ich werde daher eine Performanzschwierigkeit für das Scheitern Dreijähriger im FBT verantwortlich machen.

lediglich aufzufordern, die K-Prämisse bei der Beantwortung der Testfrage zu *berücksichtigen*.

Ich schlage also eine *Testreihe* vor, in der auf verschiedene Weise versucht wird, dreijährige Kinder zur Beachtung der K-Prämisse anzuregen. Wenn die Erklärung aus dem letzten Abschnitt richtig ist, dann ist es denkbar, dass dreijährige Kinder dadurch befähigt werden, die Testfrage nach der Überzeugung des Protagonisten auf das durch die K-Prämisse konstituierte Szenario zu beziehen.

Sobel & Lillard (2001, 1. Experiment) haben bereits einen Test durchgeführt, der als P-FBT aufgefasst werden könnte. Sie führten das FBT-Szenario mit bekannten *Fantasiefiguren* als Protagonisten aus (mit Elmo und dem Krümelmonster). Dadurch sollten die Kinder dazu angeregt werden, ein hypothetisches Szenario aufzubauen. Doch Sobel & Lillard fanden keinen Unterschied in den Leistungen im FBT und ihrer Version des P-FBT. Die Autoren vermuten, dass eine Fantasiefigur nicht ausreicht, um einen *Pretend*-Kontext zu erzeugen. Meines Erachtens scheitert das Experiment jedoch daran, dass der *falsche* Kontext hervorgehoben wurde. Die Kinder wurden zwar dazu angeregt, alle Fragen innerhalb des hypothetischen Szenarios zu beantworten, in dem das Krümelmonster existiert. Doch sie wurden nicht dazu angeregt, die Fragen innerhalb des hypothetischen Szenarios zu beantworten, in dem der veraltete Wissensstand des Krümelmonsters noch aktuell ist.

Der P-FBT und das Simulationsmodell

Im folgenden will ich nun zeigen, warum ein positiver Effekt – d.h. eine Erleichterung des FBT für Dreijährige – auf einen zugrunde liegenden *Simulationsprozess* hinweisen würde. Im P-FBT wird die K-Prämisse mit dem Ziel hervorgehoben, dass Kinder ein entsprechendes Szenario aufbauen bzw. ihr bereits bestehendes Szenario aufrecht zu halten, das die Überzeugungen des Protagonisten repräsentiert. Damit wäre der erste Schritt der Simulation, das Feststellen der Rahmenbedingungen, geschafft. Im zweiten, entscheidenden Schritt müssen die Kinder herausfinden wo sie (in der Rolle des Protagonisten) nach der Schokolade suchen würden. Dazu müssen sie sich die Frage stellen, wo die Schokolade ist. Durch die besondere Aufforderung des Versuchsleiters wird ihnen nahe gelegt, diese Frage *innerhalb* des hypothetischen Szenarios zu stellen. Sie fragen sich also, wo die Schokolade wäre, wenn sie nicht verlegt worden wäre. Die Be-

antwortung dieser Frage kann als Simulation des Prozesses aufgefasst werden, mit dem der Protagonist für sich die Frage beantwortet, wo die Schokolade ist.⁵⁸ Die Antwort auf die K-Frage verwenden die Kinder dann im dritten Simulationsschritt, um das Verhalten des Protagonisten vorherzusagen.

Wenn die Aufforderung im P-FBT es dreijährigen Kindern nun tatsächlich erleichtern würde, das Verhalten des Protagonisten richtig vorherzusagen, wäre das meines Erachtens ein Beleg dafür, dass sie das soeben beschriebene Verfahren verwenden. Denn eine alternative Erklärung wird in der Forschungsliteratur nicht angeboten. Ein positiver Befund im P-FBT würde daher die These stützen, dass schon dreijährige Kinder die Kompetenz besitzen, den Entscheidungsprozess des Protagonisten (wo nach der Schokolade zu suchen ist) zu simulieren, indem sie sich (innerhalb eines hypothetischen Szenarios, das die Überzeugungen des Protagonisten berücksichtigt) die Frage stellen, wo nach der Schokolade zu suchen ist.

Man könnte nun gegen diese Erklärung einwenden, dass sie den Anschein erwecke, als könnten Kinder nur dann simulieren, wenn sie durch besondere Umstände dazu angeregt werden. Daher schließe die Erklärung nicht aus, dass vierjährige Kinder trotz allem die Fähigkeit erwerben, den FBT mit Hilfe einer alltagspsychologischen Theorie zu lösen. Den zweiten Teil des Einwands halte nicht für problematisch. Selbst wenn vierjährige Kinder Überzeugungen theoretisch erfassen sollten, könnte Simulation immer noch das sein, was in der Entwicklung des Kindes *zuerst* auftritt. Zudem besteht die Möglichkeit zu argumentieren, dass Simulationen die Daten und damit die Grundlage für die Theoriebildung liefern, und daher *notwendig früher* auftreten.

Doch was, wenn es stimmen sollte, dass Kinder nur dann simulieren, wenn man sie dazu besonders auffordert? In diesem Fall wäre zwar nachgewiesen, dass es Simulationsprozesse gibt, aber nicht, dass sie in der alltagspsychologischen Praxis eine wesentliche Rolle spielen. Vielmehr könnte man den Eindruck bekommen, dass sie nur unter künstlichen Bedingungen zum Vorschein kommen. Um diesen Eindruck zu zerstreuen, möchte ich auf die Befunde von Clements & Perner (1994) sowie Happé & Loth (2002) mit nonverbalen Indikatoren zurückgreifen, die ich im zweiten Kapitel vorgestellt habe. Clements & Perner konnten zeigen, dass dreijährige Kinder im ersten Augenblick – in einer Kunstpause vor der eigentlichen Testfrage – in Richtung des *alten* Orts der Scho-

⁵⁸ Wegen der kontrafaktischen Prämisse handelt es sich um einen Fall von hypothetischer Simulation.

kolade blicken, bevor sie antworten, dass der Protagonist die Schokolade am neuen Ort suchen werde. Happé & Loth konnten zeigen, dass dreijährige Kinder die Bezugsabsicht eines Protagonisten erkennen, der ein Objekt benennt, das ohne sein Wissen gegen ein anderes ausgetauscht worden ist.

Diese Experimente sind m.E. so konstruiert, dass sie den Erleichterungseffekt nicht benötigen: Dreijährige identifizieren sich mit dem Protagonisten – so wie sie es im Puppentheater und beim Fernsehgucken immer machen. Sie sind zum Zeitpunkt der Testfrage also in ein hypothetisches Szenario vertieft, das die Überzeugungen des Protagonisten repräsentiert. Das Vertieftsein erlaubt es ihnen, das Nicht-Wissen des Protagonisten zu berücksichtigen und die richtige Erwartung bezüglich seines Such-Verhaltens auszubilden. Doch das PP Dreijähriger kann leicht gestört werden (siehe Abschnitt 8.1). Ich vermute daher, dass die Testfrage des Versuchsleiters ausreicht, um dreijährige Kinder aus dem hypothetischen Szenario herauszureißen. Da ihnen auf diese Weise der Faden verloren geht, beantworten sie die Testfrage so wie sie Fragen üblicherweise beantworten: unter Rückgriff auf ihr relevantes Hintergrundwissen. Die guten Leistungen Dreijähriger in den o.g. Experimenten (mit nonverbalen, indirekten Indikatoren) zeigen meines Erachtens, dass dreijährige Kinder bereits simulieren können; ihr Scheitern im FBT lässt sich darauf zurückführen, dass die Testfrage ihre Simulation unterbricht.

Ausschluss des Theorienmodells

Das Theorienmodell könnte ein positives Ergebnis im P-FBT nicht erklären. Denn es sieht nicht so aus, als habe der Erleichterungseffekt, der den P-FBT motiviert, etwas mit alltagspsychologischen Theorien zu tun. Dennoch wäre ein empirischer Befund denkbar, der belegt, dass das Hervorheben der K-Prämisse nicht auch die Anwendung einer alltagspsychologischen Theorie erleichtert. Ich schlage daher eine zweite Variante des FBT vor: den T-FBT (*Twin False Belief Test*).

Grundlage dieses Tests ist das experimentelle Design von Perner et al. (1999), das ich im sechsten Kapitel diskutiert habe. Ihr Ansatz war, die Anwendung von Theorien durch Hervorheben des relevanten Faktors zu erleichtern. Das Hervorheben geschieht durch das gleichzeitige Präsentieren von zwei Testaufgaben, die sich gerade hinsichtlich des relevanten Faktors unterscheiden. Perner et al. haben diesen Ansatz nur mit Erwach-

senen verfolgt, da das gleichzeitige Präsentieren von zwei Aufgaben für Kinder zu schwierig sein könnte.

Doch Robinson & Mitchell (1994) bieten eine elegante Lösung, um dieses Problem zu umgehen: Sie präsentieren Kindern ein einziges Szenario, ersetzen den Protagonisten jedoch durch ein Paar Zwillinge. Der T-FBT könnte demnach so aussehen: Einer von den Zwillingen sieht, wie die Schokolade an einen anderen Ort gelegt wird, während der Andere draußen spielt. Sollte dieses Arrangement dreijährige Kinder dazu befähigen, richtig zu bestimmen, wo der ignorante Zwilling nach der Schokolade suchen wird, wäre das ein Hinweis auf das Vorliegen einer alltagspsychologischen Theorie. (Im entscheidenden Augenblick sollte nur der suchende Protagonist im Raum sein, damit der Zwilling ihm nicht bei der Suche helfen kann.)

Sollten Dreijährige in diesem Test gute Leistungen zeigen, müsste ich zugeben, dass es zwei Wege gibt, den FBT zu lösen. Da Erwachsene über geeignete alltagspsychologische Theorien verfügen, um den FBT zu bestehen, ist durchaus zu erwarten, dass Kinder den T-FBT irgendwann bestehen. Doch ich vermute, dass sie ihn deutlich *nach* dem P-FBT bestehen werden. Zu dieser Vermutung bringt mich Heals Argument gegen das Theorienmodell, das ich im fünften Kapitel vorgestellt habe. Demzufolge kann das Simulationsmodell den Umgang mit den Überzeugungen einfacher erklären als ein Theorienmodell, da die kognitive Verarbeitung von Überzeugungen recht kompliziert ist.

Fazit: Die Simulation falscher Überzeugungen

Im Laufe dieser Arbeit habe ich einige Versprechungen bezüglich des achten Kapitels gemacht. Ich habe versprochen, in diesem Kapitel einen *Early-Competence-Account* zu formulieren, der den FBT-Alterseffekt im Sinn des Simulationsmodells erklärt. Darüber hinaus habe ich angekündigt, eine empirische Prüfung des Simulationsmodells zu entwickeln, die die Analyse der Testanforderungen mit der Analyse der Prozessmerkmale verbindet und damit den Schwierigkeiten der im ersten und zweiten Teil diskutierten Experimente entgeht. Ich möchte nun zeigen, dass ich diese Versprechen erfüllt habe.

Das Simulationsmodell, das ich im achten Kapitel entwickelt habe, erklärt den Alterseffekt im FBT durch einen Entwicklungsschritt *außerhalb* der Alltagspsychologie. Dreijährige Kinder scheitern am FBT, weil sie nicht in der Lage sind, die Testfrage auf das hypothetische Szenario zu beziehen, das die Perspektive des Protagonisten repräsentiert. Sie verfügen zwar schon über die Kompetenz, den FBT zu bestehen, da sie hypothetische Szenarien aufbauen und Fragen dazu richtig beantworten können, solange sie in das Szenario vertieft sind. Dieses Vertieftsein ist bei dreijährigen Kindern aber noch an Hilfestellungen von außen gebunden; erst im Alter von vier Jahren beginnen sie, den Bezug von Fragen selbständig zu handhaben und diese Einschränkung auf das Vertieftsein zu überwinden.

Dreijährige Kinder lassen sich durch die Frage des Versuchsleiters aus ihrer Simulation herausreißen und beziehen die Frage systematisch falsch auf die Realität, da ihnen der Zugang zum Szenario anschließend nicht mehr möglich ist. (In diese Richtung weisen auch Studien, in denen nonverbale Indikatoren gemessen werden; hier kommen Dreijährige zu besseren Ergebnissen als im FBT.) Das Simulationsmodell kann daher als *Early-Competence-Account* angesehen werden, weil es Performanzschwierigkeiten für das späte Bestehen des FBT verantwortlich macht.

Dieses Simulationsmodell lässt sich dadurch testen, dass das Vertieftsein in das hypothetische Szenario extern angeregt wird, um den FBT zu erleichtern. Bestünden Dreijährige diesen Test, wäre das ein Beleg für die These, dass die externe Anregung der entscheidende Faktor für das Bestehen des Tests ist. Dieser Test entgeht den Schwierigkeiten, die im sechsten Kapitel besprochen wurden, da er ein Prozessmerkmal mit einer Analyse der Testanforderungen im FBT verknüpft; aus dem Simulationsmodell lässt sich eine klare Hypothese ableiten. Die Analyse der Testanforderungen (im 3. und 7.

Kapitel) hat mich darauf gebracht, einen Schritt zur Lösung des FBT besonders zu betrachten: der Schritt von der Feststellung des Nicht-Wissens des Protagonisten bezüglich des gegenwärtigen Stands der Dinge hin zur Ermittlung, was er stattdessen glaubt. Die anderen Teilschritte zur Lösung der FBT-Aufgabe (d.h. die Bestimmung des Nicht-Wissens des Protagonisten und die Ableitung einer Handlungsvorhersage aus dem Simulationsergebnis) gestehe ich dem Theorienmodell zu.

Der entscheidende Zwischenschritt lässt sich gut als Lösung einer kontrafaktischen Frage auffassen: Wo wäre die Schokolade, wenn sie nicht verlegt worden wäre? Ich habe daher nicht nach einem Prozessmerkmal der Simulation selbst, sondern nach einem Merkmal kontrafaktischen Schließens gesucht. Es bot sich der Erleichterungseffekt an. Werden Dreijährige aufgefordert, eine kontrafaktische Prämisse besonders zu beachten, sind sie in der Lage, korrekte Schlüsse daraus zu ziehen. Dieser Effekt – so meine Erwartung – sollte sich auch im Zusammenhang mit dem FBT zeigen.

Zwei Befunde kann ich allerdings nicht im Detail erklären: die alltagspsychologischen Defizite autistischer Kinder (Abschnitt 7.2) sowie die Fähigkeit zum PP. Doch das halte ich für keinen schwer wiegenden Einwand. Zum einen fassen neuere Erklärungsmodelle des Autismus' auch Anomalien in Betracht, die nichts mit dem Thema Alltagspsychologie zu tun haben. Zum Beispiel wird vermutet, dass autistische Kinder neue Informationen nicht in der gleichen Weise in ihr Hintergrundwissen integrieren wie normale Kinder; sie behandeln Informationen allgemein recht isoliert (dieses Phänomen heißt 'weak central coherence'). Es ist also durchaus möglich, dass die alltagspsychologischen Defizite autistischer Kinder auf eine andere kognitive Besonderheit zurückgeführt werden.

Zum anderen genügt für meine Zwecke eine oberflächliche Analyse des PP. Wichtig ist nur, dass Überzeugungen vom Hintergrundwissen in geeigneter Weise abgegrenzt werden müssen, und dass diese Leistung auch bei kontrafaktischen Schlüssen und dem FBT eine Rolle spielt. Solange dies nicht durch eine alltagspsychologische Theorie bewerkstelligt wird – und ein solcher Vorschlag findet sich nicht in der Forschungsliteratur – kann m.E. von einer simulationsartigen Erklärung ausgegangen werden, obwohl die kognitiven Grundlagen noch im Dunkeln liegen.

Zusammenfassung: Was misst der *False Belief Test*?

Die Kernfrage dieser Arbeit war, welche kognitiven Mechanismen den Umgang mit falschen Überzeugungen im Rahmen der Alltagspsychologie erklären. Zur Debatte standen die Anwendung alltagspsychologischer Theorien (Theorienmodell, siehe Abschnitt 1.1) und das Ausnutzen der eigenen kognitiven Mechanismen zum Zweck der Simulation (Simulationsmodell, siehe Abschnitt 4.1). Obwohl in vielen Bereichen der Alltagspsychologie klar ist, dass theoretisiert bzw. simuliert wird, war die Erklärung in diesem Fall *prima facie* offen.

Um zwischen den beiden Modellen zu entscheiden, sind verschiedene Experimente durchgeführt worden. Sie machen sich z.B. die unterschiedlichen Erfolgsbedingungen der beiden Modelle zunutze: Theorien müssen wahr, simulierte Prozesse müssen dem Originalprozess ähnlich sein. In den Experimenten wurden Situationen konstruiert, in denen nur eine der beiden Bedingungen erfüllt ist, so dass sich aus Erfolg oder Misserfolg einer alltagspsychologischen Vorhersage oder Erklärung auf den zugrunde liegenden Mechanismus schließen ließ. Doch ich habe dafür argumentiert, dass in diesen Situationen die Anwendung *verschiedener* Theorien sowie das Simulieren *verschiedener* Mechanismen möglich ist.⁵⁹ Das verhindert eine klare Interpretation der Befunde (siehe Abschnitte 6.1 und 6.2).

Ich bin daher dazu übergegangen, mich mit den Ergebnissen der *Theory-of-Mind*-Forschung zu befassen, da sie eine bessere Eingrenzung der verwendeten Theorien und Mechanismen erlauben. In diesem Zusammenhang erwies sich der Alterseffekt im *False Belief Test* (kurz: FBT) als zusätzlich zu erfüllende Bedingung: Das gesuchte Erklärungsmodell muss auch Auskunft darüber geben, inwiefern sich die zugrunde liegende Kompetenz im Alter von vier Jahren weiter entwickelt.

Bei der Analyse der Testanforderungen des FBT ist ein Aspekt der Aufgabenstellung aufgefallen: Man muss von der Feststellung, dass der Protagonist nicht weiß, wo sich das Objekt gegenwärtig befindet, zur Feststellung gelangen, dass der Protagonist glaubt,

⁵⁹ Der Kern dieses Arguments ist die Feststellung, dass die Alltagspsychologie vielfältiger ist als oft angenommen wird. Ebenso wie Menschen meist verschiedene Möglichkeiten haben, zu einer Entscheidung oder Schlussfolgerung zu gelangen, stehen Alltagspsychologen verschiedene kognitive Prozesse zur Verfügung, die sie theoretisch erfassen bzw. simulieren können.

dass sich das Objekt dort befindet, wo es vorher war. Auf diesen Schritt habe ich meine Untersuchung konzentriert.⁶⁰ Die Kernfrage der Arbeit stellte sich nun so dar: Welche kognitiven Mechanismen erklären den Schluss auf die veraltete Überzeugung des Protagonisten? Dem Theorienmodell zufolge muss es eine *Aktualisierungstheorie* für Überzeugungen sein, derzufolge Menschen ihre Überzeugungen im Licht inkompatibler Informationen revidieren – ansonsten aber ihre Einstellung beibehalten (siehe Abschnitt 3.2). Nach dem Simulationsmodell muss man hingegen die kontrafaktische Prämisse, dass das Objekt nicht verlegt worden ist, annehmen und sich fragen, wo das Objekt unter dieser Bedingung wäre (siehe Abschnitt 7.3). Im Rahmen der Simulation wird also ein kontrafaktischer Schluss vollzogen.⁶¹

Neben dem FBT und dem kontrafaktischen Schließen rückte schließlich noch eine dritte Fähigkeit in den Fokus der Aufmerksamkeit: das *Pretend Play* (kurz: PP). Denn um zu bestimmen, was geschieht, wenn man eine leere Tasse umdreht, während man so tut als sei sie voll, muss man ebenfalls einen kontrafaktischen Schluss vollziehen: Man fragt sich, was geschehen würde, wenn die Tasse voll wäre.⁶² Das PP erhöht die Anforderungen an das gesuchte Erklärungsmodell noch einmal, denn es handelt sich um eine Kompetenz, über die schon Zweijährige verfügen. Es ist also das folgende *Entwicklungsmuster* zu erklären: Der Umgang mit falschen Überzeugungen Anderer und der mit kontrafaktischen Prämissen entwickeln sich gleichzeitig etwa zwei Jahre nach dem ersten Auftreten des PP.

Mir ist kein Modell bekannt, das dieses Muster vollständig erklärt. Es werden stets eine oder zwei der Kompetenzen ausgespart und für unabhängig erklärt. D.h. die Vorschläge in der Forschungsliteratur fallen jeweils in eine von drei Kategorien:

⁶⁰ Die anderen beiden Aspekte habe ich dem Theorienmodell zugestanden. Die Feststellung der Ignoranz bezüglich des gegenwärtigen Standorts des Objekts kann durch eine einfache Theorie der Überzeugungsbildung gedeckt werden. Es gibt Anzeichen dafür, dass diese Theorie bereits von Dreijährigen angewendet werden kann (siehe Abschnitt 2.2). Gleiches gilt für die anschließende Ableitung einer Handlungsvorhersage aus der veralteten Überzeugung des Protagonisten (siehe Abschnitt 2.1).

⁶¹ Beim Beantworten kontrafaktischer Schlüsse ergibt sich übrigens ein ähnlicher Alterseffekt wie im FBT; nur wenige Kinder zeigen eine der beiden Fähigkeiten, aber nicht die andere.

⁶² In der Forschungsliteratur wird der Zusammenhang der drei Kompetenzen oft so ausgedrückt: Es wird behauptet, dass in jedem dieser Fälle eine hypothetische Überzeugung angenommen werden muss (das Nicht-Wissen des Protagonisten, die Prämisse eines kontrafaktischen Syllogismus', die *Pretend-Stipulation*), die ähnlich verarbeitet wird wie eine normale Überzeugung – aber doch in gewisser Weise vom Hintergrundwissen abgegrenzt. Wie das genau geschieht, konnte bislang nicht geklärt werden.

1. Das PP wird für nicht-alltagspsychologisch erklärt, so dass nur der gemeinsame Alterseffekt von FBT und kontrafaktischem Schließen zu erklären ist. Solche Erklärungsmodelle zeichnen einen charakteristischen Aspekt des kontrafaktischen Schließens als entscheidenden Faktor aus, der dreijährigen Kindern noch fehlt, um die Aufgaben zu meistern. Z.B. kann das Ausblenden der hypothetischen Prämisse zum Problem erklärt werden (ein Beispiel habe ich im Abschnitt 7.3 vorgestellt). Aber auch Theorienmodelle fallen in diese Kategorie, die die Alterseffekte im FBT und beim kontrafaktischen Schließen auf den Erwerb einer alltagspsychologischen Theorie zurückführen (siehe dazu die Abschnitte 3.1 und 8.1).
2. Das kontrafaktische Schließen wird für nicht-alltagspsychologisch erklärt, so dass nur der Schritt vom PP zum Bestehen des FBT erklärt werden muss (ein Beispiel habe ich im Abschnitt 7.2 vorgestellt).
3. Sowohl das PP als auch das kontrafaktische Schließen werden für nicht-alltagspsychologisch erklärt, so dass nur noch der Alterseffekt im FBT übrig bleibt (siehe die Beispiele in den Abschnitten 3.2 und 3.3).

Modelle des Typs 1 werden dadurch gestützt, dass zweijährige Kinder bereits so tun können als ob, aber weder mit falschen Überzeugungen noch mit kontrafaktischen Prämissen richtig umgehen können. Doch der Unterschied zwischen den letzteren Kompetenzen und dem PP konnte bislang nicht verständlich gemacht werden. Es besteht daher die Möglichkeit, dass das spätere Auftreten des kontrafaktischen Schließens und der Fähigkeit, den FBT zu bestehen, auf eine Schwierigkeit auf der Performanzebene zurückzuführen ist.

Modelle des Typs 2 werden durch die Fähigkeiten autistischer Kinder gestützt: Sie können gut kontrafaktisch schließen, scheitern aber im FBT und tun sich mit dem PP schwer. Doch es hat sich gezeigt, dass sich auch diese Gruppierung der Kompetenzen theoretisch nicht verständlich machen lässt. Es besteht daher die Möglichkeit, dass die Besonderheiten autistischer Kinder auf andere Ursachen zurückzuführen sind.

Vertreter von Modellen des Typs 3 heben hingegen die Besonderheit des FBT hervor und bestreiten jeden Zusammenhang mit Kompetenzen, die nicht alltagspsychologischer Natur sind. Doch dagegen sprechen m.E. zwei Gründe. Zum einen schneiden Dreijährige in einigen Varianten des FBT, die keine explizite Antwort erforderten, gut ab. Das nährt den Verdacht, dass die Kinder bereits über die nötige Kompetenz im Umgang mit falschen Überzeugungen verfügen (siehe Abschnitt 2.2). Zum anderen wäre eine kohä-

rente Erklärung der drei Fähigkeiten angesichts des vermuteten inneren Zusammenhangs erstrebenswert.

Die von mir favorisierte Erklärung fasst das PP als kognitive Grundlage sowohl des FBT als auch des kontrafaktischen Schließens auf. Dieser Ansatz ist einfacher als der des Theorienmodells, das u.a. die Anwendung einer Aktualisierungstheorie von Überzeugungen postuliert. Denn bei der Anwendung einer solchen Theorie müssen auch Beziehungen *zwischen* Überzeugungen theoretisch erfasst werden. Zum Beispiel muss die Theorie dem Kind verständlich machen, dass der Protagonist seine Überzeugung revidiert hätte, wenn er gesehen hätte, dass das Objekt verlegt wird (siehe dazu das im fünften Kapitel diskutierte Argument). Das Simulationsmodell muss lediglich voraussetzen, dass sich das Kind an den früheren Standort des Objekts erinnert und einen kontrafaktischen Schluss vollziehen kann.⁶³

Als Performanzschwierigkeit, die für das späte Auftreten des Umgangs mit falschen Überzeugungen und kontrafaktischen Prämissen verantwortlich ist, kommt ein Aspekt des Unterhaltens von hypothetischen Szenarien in Betracht: Dreijährige haben Schwierigkeiten, ein hypothetisches Szenario ohne externe Anregung aufrecht zu erhalten. Im PP unterstreichen sie und ihre Spielpartner den hypothetischen Charakter durch allerlei Handlungen und gegenseitige Aufforderungen zum Spiel. Sobald die Kinder jedoch aus dem Szenario herausgerissen werden, sind sie nicht mehr in der Lage, Fragen der Versuchsleiter als Fragen zum Szenario aufzufassen. Stattdessen beziehen sie jede Frage auf die Realität. So kommt es, dass sie die Prämissen kontrafaktischer Syllogismen und das Nicht-Wissen des Protagonisten im FBT ignorieren, wenn sie die Testfrage beantworten (siehe dazu Abschnitt 8.1).

Eine experimentelle Bestätigung dieses Modells halte ich für möglich (siehe Abschnitt 8.2). Denn es hat sich gezeigt, dass auch Dreijährige kontrafaktisch schließen können, wenn man sie dazu anregt, ein hypothetisches Szenario aufzubauen. Analog dazu könnte man versuchen, dreijährige Kinder im FBT dazu anzuregen, ihr hypothetisches

⁶³ Vor dem Hintergrund dieses Arguments wirkt es seltsam, dass Vertreter des Theorienmodells so stark auf die Erklärung des FBT gesetzt haben. Andere Typen geistiger Zustände lassen sich vermutlich einfacher theoretisch erfassen und eignen sich daher besser für das Theorienmodell. Das Motiv für Theorienmodelle für den Bereich der Überzeugungen war, dass Theorien auch disparate Phänomene systematisch beschreiben können. Doch es scheint, dass realitätsunangemessene Handlungen aufgrund falscher Überzeugungen disparater sind als es die theoretischen Kompetenzen von Alltagspsychologen erlauben.

Szenario aufrecht zu erhalten. Ich erwarte, dass dies ihre Leistungen erhöht. Auf der Grundlage eines solchen Befunds ließe sich die o.g. Performanzschwierigkeit als entscheidender Faktor auszeichnen und ein *Early-Competence-Account* etablieren, demzufolge die Kompetenz zum Umgang mit falschen Überzeugungen bereits mit dem PP im Alter von zwei Jahren angelegt ist.

Bislang steht ein unabhängiger Beleg noch aus, und es konnten auch nicht alle Alternativen dieses Erklärungsmodells definitiv ausgeschlossen werden. Doch aus drei Gründen wäre eine Prüfung meines Simulationsmodells sinnvoll: Es erklärt erstens das beobachtete Entwicklungsmuster *vollständig*. Zweitens lässt sich der erwartete Befund in meiner Variante des FBT *eindeutig* als Stütze des Simulationsmodells interpretieren. Und drittens wird eine *zentrale* alltagspsychologische Leistung durch einen zugrunde liegenden Simulationsprozess erklärt, die bislang von Vertretern des Theorienmodells beansprucht wurde. Selbst wenn sich das Simulationsmodell nicht auf andere kognitive Fähigkeiten erweitern ließe, hätten Simulationen damit einen festen Platz in einer Theorie der interessanten alltagspsychologischen Leistungen.

Literaturangaben

- Astington, Janet Wilde, Harris, Paul L., Olson, David R. (eds.) (1988), *Developing theories of mind*, Cambridge: University Press.
- Baron-Cohen, Simon (1995), *Mindblindness: an essay on autism and theory of mind*, Cambridge, MA: MIT Press.
- Baron-Cohen, Simon, Leslie, Alan M. & Frith, Uta (1985), "Does the autistic child have a 'theory of mind'?", *Cognition*, 21, 37-46.
- Baron-Cohen, Simon, Tager-Flusberg, Helen, Cohen, Donald J. (2000), *Understanding other minds. Perspectives from developmental cognitive neuroscience*, 2. Auflage, Oxford: University Press.
- Bartsch, Karen & Wellman, Henry M. (1989), "Young children's attribution of action to beliefs and desires", *Child Development*, 60, 946-964.
- Botterill, George (1996), "Folk psychology and theoretical status", in: Carruthers & Smith (1996), 105-118.
- Carlson, Stephanie M. & Moses, Louis J. (2001), "Individual differences in inhibitory control and children's theory of mind", *Child Development*, 72, 1032-1053.
- Carpendale, Jeremy I. & Chandler, Michael J. (1996), "On the distinction between false belief understanding and subscribing to an interpretive theory of mind", *Child Development*, 67, 1686-1706.
- Carruthers, Peter (1996), "Simulation and self-knowledge: a defence of theory-theory", in: Carruthers & Smith (1996), 22-38.
- Carruthers, Peter & Smith, Peter K. (eds.) (1996), *Theories of theories of mind*, Cambridge: Cambridge University Press.
- Chandler, Michael, Fritz, A.S. & Hala, S. (1989), "Small scale deceit: deception as a marker of 2-, 3-, and 4-year-olds' early theories of mind", *Child Development*, 60, 1263-1277.
- Churchland, Paul M. (1988), "Folk psychology and the explanation of human behaviour", *The Aristotelian Society Supplement*, 62, 209-221.
- Clements, Wendy A. & Perner, Josef (1994), "Implicit understanding of belief", *Cognitive Development*, 9, 377-397.
- Davies, Martin (1994), "The mental simulation debate", in: Peacocke, Christopher (ed.), *Objectivity, Simulation and the unity of consciousness: current issues in the philosophy of mind* (Proceedings of the British Academy, vol. 83), Oxford: Oxford University Press, 99-127.
- Davies, Martin & Stone, Tony (eds.) (1995a), *Folk psychology: the theory of mind debate*, Oxford: Blackwell. [Wiederabdruck einer Sonderausgabe von: *Mind & Language*, 7 (1992)]
- Davies, Martin & Stone, Tony (eds.) (1995b), *Mental simulation: evaluations and applications*, Oxford: Blackwell.
- Dennett, Daniel C. (1984), "Cognitive wheels: the frame problem of AI", in: Hookway, Christopher (ed.), *Minds, machines and evolution. Philosophical studies*, Cambridge: Cambridge University Press, 129-151.
- Dennett, Daniel C. (1987), *The intentional stance*, Cambridge, MA: Bradford.
- Dias, M. & Harris, Paul L. (1988), "The effect of make-believe play on deductive reasoning", *British Journal of Developmental Psychology*, 6, 207-221.

- Dias, M. & Harris, Paul L. (1990), "The influence of the imagination on reasoning by young children", *British Journal of Developmental Psychology*, 8, 305-318.
- Flavell, John H. (1988), "The development of children's knowledge about the mind: from cognitive connections to mental representations", in: Astington et al. (1988), 244-267.
- Flavell, John H., Flavell, E.R. & Green, F.L. (1983), "Development of the appearance-reality distinction", *Cognitive Psychology*, 15, 95-120.
- Fodor, Jerry A. (1968), "The appeal to tacit knowledge in psychological explanation", *Journal of Philosophy*, 65, 627-640.
- Fodor, Jerry A. (1987), *Psychosemantics. The problem of meaning in the philosophy of mind*, Cambridge, MA: MIT Press.
- Fodor, Jerry A. (1992), "A theory of the child's theory of mind", *Cognition*, 44, 283-296.
- Garnham, Wendy A. & Perner, Josef (2001), "Actions really do speak louder than words - but only implicitly: young children's understanding of false belief in action", *British Journal of Developmental Psychology*, 19, 413-432.
- Goldman, Alvin I. (1989), "Interpretation psychologized", *Mind & Language*, 4, 161-185.
- Goldman, Alvin I. (1993), "The psychology of folk psychology", *Behavioral and Brain Sciences*, 16, 15-28.
- Goldman, Alvin I. (2000a), "Folk psychology and mental concepts", *Proto-Sociology*, 14, 4-25.
- Gopnik, Alison & Astington, Janet Wilde (1988), "Children's understanding of representational change, and its relation to the understanding of false belief and the appearance-reality distinction", *Child Development*, 59, 26-37.
- Gopnik, Alison & Graf, Peter (1988), "Knowing how you know: young children's ability to identify and remember the source of their beliefs", *Child Development*, 59, 1366-1371.
- Gopnik, Alison & Slaughter, Virginia (1991), "Young children's understanding of changes in their mental states", *Child Development*, 62, 98-110.
- Gopnik, Alison & Wellman, Henry M. (1992), "Why the child's theory of mind really is a theory", *Mind & Language*, 7, 145-171.
- Gopnik, Alison & Wellman, Henry M. (1994), "The theory theory", in: Hirschfeld, Lawrence A. & Gelman, Susan A. (eds.), *Mapping the mind. Domain specificity in cognition and culture*, Cambridge: University Press, 257-293.
- Gordon, Robert M. (1986), "Folk psychology as simulation", *Mind & Language*, 1, 158-171.
- Gordon, Robert M. (1992), "The simulation theory: objections and misconceptions", *Mind & Language*, 7, 11-34.
- Gordon, Robert M. (1996), "'Radical' simulationism", in: Carruthers & Smith (1996), 11-21.
- Gordon, Robert M. (2000), "Simulation and the explanation of action", in: Kögler, Hans Herbert & Stueber, Karsten R. (eds.), *Empathy and agency: the problem of understanding in the human sciences*, Boulder: Westview Press, 62-82.
- Happé, Francesca G.E. (1995), "The role of age and verbal ability in the theory of mind task performance of subjects with autism", *Child Development*, 66, 843-855.
- Happé, Francesca G.E. & Loth, Eva (2002), "'Theory of mind' and tracking speakers' intentions", *Mind & Language*, 17, 24-36.

- Harris, Paul L. (1991), "The work of the imagination", in: Whiten, Andrew (ed.), *Natural theories of mind: evolution, development, and simulation of everyday mindreading*, Cambridge, MA: Blackwell, 283-304.
- Harris, Paul L. (2000), *The work of the imagination*, Oxford: Blackwell.
- Harris, Paul L. & Kavanaugh, Robert D. (1993), "Young children's understanding of pretense", *Monographs of the Society for Research in Child Development*, 58(1).
- Heal, Jane (1986), "Replication and functionalism", in: Butterfield, Jeremy (ed.), *Language, mind and logic*, Cambridge: Cambridge University Press, 135-150.
- Heal, Jane (1995), "How to think about thinking", in: Davies & Stone (1995b), 33-52.
- Heal, Jane (1996a), "Simulation, theory, and content", in: Carruthers & Smith (1996), 75-89.
- Heal, Jane (1996b), "Simulation and cognitive penetrability", *Mind & Language*, 11, 44-67.
- Heal, Jane (1998), "Co-cognition and off-line simulation: two ways of understanding the simulation approach", *Mind & Language*, 13, 477-498.
- Heal, Jane (2000), "Other minds, rationality and analogy", *Aristotelian Society Supplement*, 74, 1-19.
- Hogrefe, Jürgen, Wimmer, Heinz & Perner, Josef (1986), "Ignorance versus false belief: a developmental lag in attribution of epistemic states", *Child Development*, 57, 567-582.
- Johnson, Carl Nils (1988), "Theory of mind and the structure of conscious experience", in: Astington et al. (1988), 47-63.
- Johnson, Carl Nils & Maratsos, M.P. (1977), "Early comprehension of mental verbs: think and know", *Child Development*, 48, 1743-1747.
- Kühberger, Anton, Perner, Josef, Schulte, Michael & Leingruber, Robert (1995), "Choice or no choice: is the longer effect evidence against simulation?", *Mind & Language*, 10, 423-436.
- Leevers, H.J. & Harris, Paul L. (1999), "Persisting effects of instruction on young children's syllogistic reasoning with incongruent and abstract premises", *Thinking and Reasoning*, 5, 145-173.
- Leslie, Alan M. (1987), "Pretense and representation: the origins of 'theory of mind'", *Psychological Review*, 94, 412-426.
- Leslie, Alan M. (1994), "*Pretending and believing* : issues in the theory of ToMM", *Cognition*, 50, 211-238.
- Leslie, Alan M. (2000), "How to acquire a representational theory of mind", in: Sperber, Dan (ed.), *Meta-representations. A multidisciplinary perspective*, Oxford: University Press, 197-223.
- Leslie, Alan M. & Thaiss, L. (1992), "Domain specificity in conceptual development: neuropsychological evidence from autism", *Cognition*, 43, 225-251.
- Lewis, David (1972), "Psychophysical and theoretical identifications", *Australasian Journal of Philosophy*, 50, 249-258.
- Lewis, David (1994), "Reduction of mind", in: Guttenplan, Samuel (ed.), *A companion to the philosophy of mind*, Oxford: Blackwell, 412-431.
- Lillard, Angeline S. & Flavell, John H. (1992), "Young children's understanding of different mental states", *Developmental Psychology*, 28, 626-634.
- Meltzoff, Andrew N. (1995), "Understanding intentions of others: re-enactment of intended acts by 18-month-old children", *Developmental Psychology*, 31, 838-850.
- Mitchell, Peter (1996), *Acquiring a conception of mind. A review of psychological research and theory*, Hove: Psychology Press.

- Mitchell, Peter & Laco  e, Hazel (1991), "Children's early understanding of false belief", *Cognition*, 39, 107-127.
- Mitchell, Peter & Riggs, Kevin John (eds.) (2000), *Children's reasoning and the mind*, Hove: Psychology Press.
- Moses, Louis J. & Flavell, John H. (1990), "Inferring false beliefs from actions and reactions", *Child Development*, 61, 929-945.
- Nichols, Shaun & Stich, Stephen (2000), "A cognitive theory of pretense", *Cognition*, 74, 115-147.
- Nichols, Shaun & Stich, Stephen (2003), *Mindreading. An integrated account of pretence, self-awareness, and understanding other minds*, Oxford: Clarendon Press.
- Nichols, Shaun, Stich, Stephen, Leslie, Alan & Klein, David (1996), "Varieties of off-line simulation", in: Carruthers & Smith (1996), 39-74.
- Nisbett, Richard E. & Wilson, Timothy DeCamp (1977), "Telling more than we can know: verbal reports on mental processes", *Psychological Review*, 84, 231-259.
- O'Neill, Daniela K. (1996), "Two-year-old children's sensitivity to a parent's knowledge state when making requests", *Child Development*, 67, 659-677.
- Perner, Josef (1995), "The many faces of belief: reflections on Fodor's and the child's theory of mind", *Cognition*, 57, 241-269.
- Perner, Josef (1996), "Simulation as explicitation of predication-implicit knowledge about the mind: arguments for a simulation-theory mix", in: Carruthers & Smith (1996), 90-104.
- Perner, Josef (2000), "About + Belief + Counterfactual", in: Mitchell & Riggs (2000), 367-401.
- Perner, Josef, Baker, Sarah & Hutton, Deborah (1994), "*Prelief*: the conceptual origins of belief and pretence", in: Lewis, Charlie & Mitchell, Peter (eds.), *Children's early understanding of mind*, Hove: Psychology Press, 261-293.
- Perner, Josef, Gschaider, Andreas, K  hberger, Anton & Schrofner, Siegfried (1999), "Predicting others through simulation or by theory? A method to decide", *Mind & Language*, 14, 57-79.
- Perner, Josef & K  hberger, Anton (2003), "Putting philosophy to work by making simulation theory testable: the case of endowment", in: Kanzian, Ch., QUITTERER, J. & Runggaldier, E. (eds.), *Persons. An interdisciplinary approach*, Wien:   bv-hpt Verlagsgesellschaft, 101-116.
- Perner, Josef & Lang, Birgit (2000), "Theory of mind and executive function: Is there a developmental relationship?", in: Baron-Cohen et al. (2000).
- Perner, Josef, Leekam, S. R., Myers, D., Davis, S., & Odgers, N. (Ms.), "Misrepresentation and referential confusion: Children's difficulty with false beliefs and outdated photographs", unver  ffentliches Manuskript.
- Perner, Josef & Wimmer, Heinz (1988), "Misinformation and unexpected change: testing the development of epistemic-state attribution", *Psychological Research*, 50, 191-197.
- Peterson, Donald M. & Bowler, Dermot M. (2000), "Counterfactual reasoning and false belief understanding in children with autism", *Autism*, 4, 391-405.
- Peterson, Donald M. & Riggs, Kevin J. (1999), "Adaptive modelling and mindreading", *Mind & Language*, 14, 80-112.
- Pillow, B.H. (1989), "Early understanding of perception as a source of knowledge", *Journal of Experimental Child Psychology*, 47, 116-129.

- Povinelly, Daniel J. & deBlois, Sandra (1992), "Young children's (*Homo sapiens*) understanding of knowledge formation in themselves and others", *Journal of Comparative Psychology*, 106, 228-238.
- Premack, David & Woodruff, Guy (1978), "Does the chimpanzee have a theory of mind?", *Behavioral and Brain Sciences*, 4, 515-526.
- Riggs, Kevin J., Peterson, Donald M., Robinson, Elizabeth J. & Mitchell, Peter (1998), "Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality?", *Cognitive Development*, 13, 73-90.
- Ripstein, Arthur (1987), "Explanation and empathy", *Review of Metaphysics*, 40, 465-482.
- Robinson, Elizabeth J. & Mitchell, Peter (1994), "Children's judgements of ignorance on the basis of absence of experience", *British Journal of Developmental Psychology*, 12, 113-129.
- Roth, Daniel & Leslie, Alan M. (1998), "Solving belief problems: towards a task analysis", *Cognition*, 66, 1-31.
- Sapp, Felicity, Lee, Kang & Muir, Darwin (2000), "Three-year-olds' difficulty with the appearance-reality distinction: is it real or is it apparent?", *Developmental Psychology*, 36, 547-560.
- Scholl, Brian J. & Leslie, Alan M. (1999), "Modularity, development and 'theory of mind'", *Mind & Language*, 14, 131-153.
- Scott, Fiona J., Baron-Cohen, Simon, Leslie, Alan (1999), "'If pigs could fly': a test of counterfactual reasoning and pretence in children with autism", *British Journal of Developmental Psychology*, 17, 349-362.
- Sobel, David M. & Lillard, Angeline S. (2001), "The impact of fantasy and action on young children's understanding of pretence", *British Journal of Developmental Psychology*, 19, 85-98.
- Stich, Stephen P. & Nichols, Shaun (1992), "Folk psychology: simulation or tacit theory?", *Mind & Language*, 7, 35-71.
- Stich, Stephen P. & Nichols, Shaun (1997), "Cognitive penetrability, rationality and restricted simulation", *Mind & Language*, 12, 297-326.
- Stich, Stephen P. & Ravenscroft, Ian (1994), "What is folk psychology?", *Cognition*, 50, 447-468.
- Stone, Tony & Davies, Martin (1996), "The mental simulation debate: a progress report", in: Carruthers & Smith (1996), 119-137.
- Taylor, Marjorie (1998), *Imaginary companions and the children who create them*, New York: Oxford.
- Taylor, Marjorie & Carlson, Stephanie M. (1997), "The relation between individual differences in fantasy and theory of mind", *Child Development*, 68, 436-455.
- Wellmann, Henry M. (1990), *The child's theory of mind*, Cambridge, MA: MIT Press.
- Wellman, Henry M. & Bartsch, Karen (1988), "Young children's reasoning about beliefs", *Cognition*, 30, 239-277.
- Wellman, Henry M., Cross, David & Watson, Julianne (2001), "Meta-analysis of theory-of-mind development: the truth about false belief", *Child Development*, 72, 655-684.
- Wellman, Henry M. & Wooley, J.D. (1990), "From simple desires to ordinary beliefs: the early development of everyday psychology", *Cognition*, 35, 245-275.
- Wilkerson, William S. (2001), "Simulation, theory, and the frame problem: the interpretive moment", *Philosophical Psychology*, 14, 141-153.
- Wimmer, Heinz & Gschaider, Andreas (2000), "Children's understanding of belief: Why is it important to understand what happened?", in: Mitchell & Riggs (2000), 253-266.

- Wimmer, Heinz & Hartl, M. (1991), "The Cartesian view and the theory of mind: developmental evidence from understanding false belief in self and other", *British Journal of Developmental Psychology*, 9, 125-138.
- Wimmer, Heinz, Hogrefe, Jürgen & Sodian, Beate (1988), "A second stage in children's conception of mental life: understanding informational accesses as origins of knowledge and belief", in: Astington, Harris & Olson (1988), 173-192.
- Wimmer, Heinz & Perner, Josef (1983), "Beliefs about beliefs: representing and constraining function of wrong beliefs in young children's understanding of deception", *Cognition*, 13, 103-128.
- Wimmer, Heinz & Weichbold, Viktor (1994), "Children's theory of mind: Fodor's heuristics examined", *Cognition*, 53, 45-57.
- Zaitchik, Deborah (1991), "Is only seeing really believing? Sources of the true false belief in the false belief task", *Cognitive Development*, 6, 91-103.