45 Epistemische Logik

Die epistemische Logik versucht ein besseres Verständnis des Wissens- und des Glaubensbegriffs unter der Verwendung formallogischer Methoden zu erlangen. Die Anfänge der modernen epistemischen Logik sind in der Mitte des 20. Jahrhunderts parallel zur Entwicklung der Modallogik zu verorten. Von Wright (1951) ist wahrscheinlich die erste Arbeit dieser Tradition, die sich explizit der epistemischen Logik zuwendet, wobei Hintikkas Buch *Knowledge and Belief. An Introduction to the Logic of the Two Notions* (1962) die Grundlagen für die moderne epistemische Logik legt.

Diese Ursprünge der epistemischen Logik sind stark sprachanalytisch geprägt, so dass das Augenmerk der frühen epistemischen Logik auf die sprachliche Verwendung der Verben ›wissen‹ und ›glauben‹ gerichtet ist. Die Systeme der epistemischen Logik und deren Semantik sind gemäß dieser Tradition als Versuch zu werten, die logischen Gesetzmäßigkeiten der Verben und ihre Bedeutung wiederzugeben. In der zeitgenössischen epistemischen Logik ist die sprachanalytische Prägung nicht mehr dominant, wobei Teile dieser Tradition als Semantik der natürlichen Sprache fortgeführt werden. Auch unter Einfluss von Forschung in den Wirtschaftswissenschaften (Spieltheorie) und der Informatik wendet man sich stattdessen allgemeineren Wissens- und Glaubensbegriffen zu, die sich nicht zwangsläufig auf das Wissen oder den Glauben von Personen beziehen, sondern Wissen und Glauben auch Maschinen bzw. komplexen Systemen zuschreiben. In diesem Aufsatz werden wir Subjekte, denen Glauben bzw. Wissen zugeschrieben wird, deswegen neutral als ›Agentinnen‹ bezeichnen. Mit der Hinwendung zu allgemeineren Wissensbegriffen und dem einhergehenden Bedeutungsverlust der sprachanalytischen Tradition ist die Modellierung von epistemischen Situationen und der Informationsaustausch zwischen mehreren Agentinnen in sogenannten Multiagentensystemen in den Vordergrund der epistemischen Logik gerückt. Fagin u. a. (1995) geben einen guten ersten Ein- und Überblick in die epistemische Logik dieser Prägung.

Dieses Kapitel versucht einen allgemeinen Überblick über die verbreitetsten Systeme der epistemischen Logik sowohl für einzelne als auch für mehrere Agentinnen zu geben, die Grundlagen der Mögliche-Welten-Semantik für die epistemische Logik einzuführen und auf einige philosophische Probleme, die bei der logischen Behandlung des Wissensbegriffs auftreten, hinzuweisen. Das Ziel des Kapitels ist, den Leserinnen eine erste Einführung in die epistemische Logik zu geben, sowie auf wichtige weiterführende Diskussionen hinzuweisen. An einigen Stellen des Kapitels, insbesondere im Abschnitt »Multiagentensysteme«, wird eine gewisse Bereitschaft vorausgesetzt, sich mit etwas formalerer Terminologie auseinanderzusetzen. Jedoch können diese Stellen ohne Auswirkungen für das Verständnis des restlichen Kapitels übersprungen werden.

45.1 Grundlagen der epistemischen Logik

Genaugenommen sollte an dieser Stelle zwischen der epistemischen Logik, der Wissenslogik, und der doxastischen Logik, der Glaubenslogik, unterschieden werden. Jedoch hat es sich eingebürgert, sowohl die Wissens- als auch die Glaubenslogik als epistemische Logik zu bezeichnen. Diese Logiken erweitern die Aussagenlogik um einen Wissens- bzw. Glaubensoperator, die sich formal gesehen wie ein Modaloperator verhalten. Dementsprechend ist die aussagenlogische epistemische Sprache eine Erweiterung der aussagenlogischen Sprache um einen nicht wahrheitsfunktionalen Modaloperator, der in diesem Fall Wissensoperator genannt wird. Die intendierte Lesart des Wissensoperators K_a ist »Agentin *a* weiß, dass«. Im allgemeineren Fall wird die Sprache nicht nur um einen Wissensoperator für eine Agentin, sondern um Wissensoperatoren für mehrere Agentinnen einer endlichen Menge *I* von Agentinnen erweitert. Des Weiteren besteht die Sprache aus einer abzählbaren Menge propositionaler Variablen At $= \{p_1, p_2, \ldots\}$ die als Platzhalter für Aussagesätze verstanden werden können und den boolschen Operatoren $\neg$ für die Negation und $\wedge$ für die Konjunktion. Die restlichen boolschen Operatoren werden in der üblichen Art und Weise definiert. Die Menge der wohlgeformten Formeln kann wie folgt angegeben werden:

$$\varphi ::= p_i \,|\, \neg\varphi \,|\, (\varphi \wedge \varphi) \,|\, K_j\varphi$$

wobei $pi \in At$ mit $i \in \mathbb{N}$ und $j \in I$. Wenig überraschend ist die intendierte Lesart der Formel $K_j\phi$ als »*a* weiß, dass φ «. Bei einer doxastische Sprache, im Gegensatz zu einer epistemischen Sprache, werden die Wissensoperatoren K_j in den obigen Definitionen durch Glaubensoperator B_j (»*j* glaubt, dass«) ersetzt. In diesem Fall ist die Formel $B_j\phi$ als »*j* glaubt, dass φ« zu lesen. In diesem Abschnitt wird die epistemische Logik einzelner Agentinnen betrachtet. Das heißt, im Gegensatz zum späteren Abschnitt zu Multiagentensystemen wird jeweils nur eine Agentin betrachtet. Aus diesem Grund kann der Subskript des Wissensoperators im Folgenden weggelassen werden: Wir schreiben $K\phi$ ($B\phi$) anstelle von $K_j\phi$ ($B_j\phi$). Doch was sind die Axiome und Regeln der epistemischen Logik?

Eine grundlegende Annahme der wichtigsten Systeme der epistemischen Logik ist, dass die Agentinnen logisch kompetent sind. Das heißt, einer Agentin wird zugeschrieben, dass sie alle logischen Wahrheiten weiß und dass die Menge aller gewussten Aussagen unter *modus ponens* abgeschlossen ist. Formal bedeutet dies, dass die wichtigsten Systeme der epistemischen Logik auf das minimale modallogische System *K* aufbauen. Das System *K* besteht aus einem Axiom und einer Regel für den Operator *K* sowie einer vollständigen Axiomatisierung der Aussagenlogik in der Sprache der epistemischen Logik. Das Axiom (*K*)

$$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$$

besagt, dass Wissen unter *modus ponens* abgeschlossen ist, wohingegen die sogenannte Nezessitationregel (Nez)

$$\frac{\varphi}{K\varphi}$$

sicherstellt, dass die Axiome der klassischen Aussagenlogik sowie die etwaigen zusätzlichen Postulate der epistemischen Logik stets gewusst werden. Die Nezessitationregel ist eine Herleitungsregel, d. h., wurde ein Satz in einem System der epistemischen Logik hergeleitet, so kann darauf geschlossen werden, dass dieser Satz gewusst wird. Eine solche Regel ist von einer klassischen Einführungsregel in einem Kalkül des natürlichen Schließens zu unterscheiden. Wäre die Regel (Nez) eine solche Einführungsregel, so könnte unter der Annahme φ auf $K\varphi$ geschlossen werden. Dies hätte eine Trivialisierung des Wissensbegriffs zur Folge und ist in den Systemen der epistemischen Logik nicht möglich, da, wie erwähnt, φ hergeleitet werden muss. Die Nezessitationsregel besagt, dass eine Agentin φ beweisen muss, um Ihr Wissen um φ ergänzen zu können.

Ein entscheidender Unterschied zwischen Glauben und Wissen ist, dass Wissen faktiv ist, d. h., Wissen impliziert Wahrheit: Weiß eine Agentin, dass φ, so muss φ auch der Fall sein. In der epistemischen Logik wird die Faktivität von Wissen durch das Prinzip (T)

$$K\varphi \rightarrow \varphi$$

ausgedrückt. Glauben ist, im Gegensatz zu Wissen, nicht faktiv: Nicht alles, was eine Agentin glaubt, muss auch der Fall sein. Faktivität impliziert, in Systemen, die zumindest das System *K* enthalten, die Konsistenz der Wissensmenge der jeweiligen Agentin. Die Konsistenzannahme scheint auch für die Menge der geglaubten Sätze plausibel zu sein, vor allem weil in diesem Rahmen Agentinnen eine starke logische Kompetenz zugeschrieben wird. Da Glauben, wie bereits erwähnt, nicht faktiv ist, muss die Konsistenzannahme jedoch zu den Axiomen hinzugefügt werden. Dies geschieht mittels des modallogischen Prinzips (D):

$$B\neg\varphi \rightarrow \neg B\varphi.$$

Die bereits erwähnten Axiome der epistemischen Logik werden zumeist durch sogenannte Introspektionsprinzipien ergänzt. Dabei handelt es sich um das positive Introspektionsprinzip und das negative Introspektionsprinzip. Bei dem positiven Introspektionsprinzip, manchmal auch KK-Prinzip genannt, handelt es sich um das modallogische Prinzip (4)

$$K\varphi \rightarrow KK\varphi.$$

Das KK-Prinzip besagt, dass, falls eine Agentin weiß, dass φ, ihr dieses Wissen bekannt ist. Das negative Introspektionsprinzip besagt, dass einer Agentin auch ihr Unwissen bekannt ist. Von einem formalen Gesichtspunkt handelt es sich dabei um das modallogische Prinzip (5)

$$\neg K\varphi \rightarrow K\neg K\varphi.$$

Beide Introspektionsprinzipien sind vor allem für den Wissensoperator umstritten. In den klassischen Arbeiten zur epistemischen Logik wird das positive Introspektionsprinzip für den Wissensoperator zumeist akzeptiert, jedoch wird es innerhalb der zeitgenössischen Erkenntnistheorie und epistemischen Logik eher kritisch betrachtet. Exemplarisch sei hier Williamson (2000) genannt, der aus seiner externalistischen Position heraus mit Nachdruck gegen das KK-Prinzip argumentiert. Das negative Introspektionsprinzip wird für den Wissensoperator zumeist abgelehnt, da es in Kombination mit dem Faktivitätsprinzip für die meisten Fragestellungen einen zu stark idealisierten Wissensbegriff impliziert. Da Glauben nicht faktiv ist, kann diese Argumentation jedoch nicht auf den Glaubensoperator übertragen werden und beide Introspektionsprinzipien werden in diesem Fall zu-

M. Grajner / G. Melchior (Hg.), *Handbuch Erkenntnistheorie*, https://doi.org/10.1007/978-3-476-04632-1_45

meist akzeptiert. Für weiterführende Diskussionen der Introspektionsaxiome sei auf Hintikka (1962), Lenzen (1978) und Williamson (2000) verwiesen.

Zusammenfassend können jeweils zwei Modallogiken als die prominentesten Systeme der epistemischen bzw. der doxastischen Logik hervorgehoben werden. Wir benutzen den *Lemmoncode*, um die modallogischen Systeme zu benennen. Gemäß des Lemmoncodes erhält eine auf das modallogische System K aufbauende Logik mit den Axiomen $X_1, ..., X_r$ den Namen $KX_1...X_r$. Für die epistemische Logik sind die zwei Logiken KT4 und KT5 die wichtigsten Systeme; für die doxastische Logik sind es die Systeme KD4 und KD45. Die Nennung des modallogischen Axioms (4) entfällt im System KT5, da es unter Verwendung der Axiome (T) und (5) bewiesen werden kann. Die System KT4 und KT5 werden in der relevanten Literatur oft auch S4 bzw. S5 genannt und im weiteren Verlauf des Kapitels folgen wir dieser Praxis. Für vertiefende Diskussionen der unterschiedlichen modallogischen Systeme sei auf Hughes/Cresswell (1996) und Blackburn u. a. (2001) verwiesen.

45.2 Mögliche-Welten-Semantik für die epistemische Logik

Nachdem wir im vorhergehenden Abschnitt verschiedene Prinzipien und Systeme der epistemischen Logik vorgestellt haben, wollen wir uns nun der semantischen Interpretation dieser Prinzipien und Systeme zuwenden. Die entscheidende Frage hierbei ist, wie der Wissensoperator K zu interpretieren ist, so dass sich eine motivierte und erklärende semantische Interpretation ergibt, die mit der intuitiven Lesart des Wissensoperators übereinstimmt. Die Standardsemantik der epistemischen Logik geht weitestgehend auf Hintikka (1962) zurück und ist eine Variante der sogenannten Möglichen-Welten-Semantik der Modallogik. Die grundlegende Intuition der Semantik ist, dass eine Agentin weiß, dass φ, genau dann, wenn in allen ihren epistemischen Alternativen, d. h. in allen mit dem Kenntnisstand der Agentin kompatiblen Situationen, φ der Fall ist. Als Folge dieser Charakterisierung entsprechen die verschiedenen Systeme bzw. Prinzipien der epistemischen Logik semantisch jeweils unterschiedlichen Antworten auf die Frage, welche Situationen mit der gegenwärtigen Situation kompatibel sind. Formal wird dies durch eine sogenannte Zugänglichkeitsrelation expliziert, welche die epistemisch möglichen Alternativen einer Situation angibt. Je nach System der epistemischen Logik hat diese Relation sehr unterschiedliche Eigenschaften.

Der zentrale Begriff der Semantik ist der eines Rahmens. Ein *Rahmen* F ist ein Tupel $W = \langle W, R_1, ..., R_n \rangle$, wobei W eine nicht leere Menge von epistemischen Alternativen bzw. mögliche Welten ist und $R_i \subseteq W \times W$ für $1 \leq i \leq n$ die zuvor erwähnte *Zugänglichkeitsrelation* der Agentin i ist. Im epistemischen Rahmen besagt die Zugänglichkeitsrelation, welche epistemischen Alternativen für die Agentin vom Standpunkt der gegenwärtigen epistemischen Alternative epistemisch möglich sind. Im Folgenden werden wir wieder vereinfachend den Fall von lediglich einer Agentin betrachten und verzichten deswegen auf die Angabe eines Subskripts bei der Nennung der Zugänglichkeitsrelation. Aus einem Rahmen gewinnt man ein Modell M für die epistemische Sprache indem man den Rahmen um eine Valuationsfunktion ergänzt. Eine *Valuationsfunktion* $V{:}At \rightarrow \wp(W)$ weist jeder propositionalen Variablen der Sprache eine Menge von epistemischen Alternativen zu. Intuitiv betrachtet ist dies die Menge der epistemischen Alternativen, in denen die entsprechende Variable wahr ist. Ein Modell der Sprache ist somit ein Paar (F, V) bestehend aus einem Rahmen und einer Valuationsfunktion.

Wir können nun präzise formulieren, unter welchen Bedingungen eine Formel der Sprache in einem Model M relativ zu einer epistemischen Alternative $w \in W$ wahr ist. Formal schreiben wir $(\mathrm{M}, w) \vDash \varphi$, um mitzuteilen, dass φ in der epistemischen Alternative w und dem Modell M wahr ist.

$$(\mathrm{M}, w) \vDash \varphi \Leftrightarrow \begin{cases} w \in V(\varphi), & \text{wenn } \varphi \in \mathrm{At} \\ (\mathrm{M}, w) \nvDash \psi, & \text{wenn } \varphi \doteq \neg\psi \\ (\mathrm{M}, w) \vDash \psi \;\&\; (\mathrm{M}, w) \vDash \chi, & \text{wenn } \varphi \doteq \psi \wedge \chi \\ \forall v \in W(wRv \Rightarrow (\mathrm{M}, v) \vDash K(\psi), & \text{wenn} \varphi \doteq K\psi. \end{cases}$$

Falls eine Formel φ in einem Modell M relativ zu allen epistemischen Alternativen $w \in W$ wahr ist, so sagen wir, dass φ wahr in M ist, und schreiben $\mathrm{M} \vDash \varphi$. Sollte eine Formel unabhängig von der gewählten Valuationsfunktion wahr sein, d. h. in jedem Modell des Rahmens F, so ist φ wahr in F, geschrieben $\mathrm{F} \vDash \varphi$.

Wir können nun den anfangs erwähnten Zusammenhang zwischen den Eigenschaften der Zugänglichkeitsrelation und den Prinzipien und Systemen der epistemischen Logik explizit machen. Die unterschiedlichen epistemischen Prinzipien sind in einem Rahmen F genau dann wahr, wenn der Rahmen, d. h. die Zugänglichkeitsrelation des Rahmens, die dem Prinzip entsprechende Eigenschaft hat. So ist zum Beispiel das Faktivitätsprinzip (T) genau dann in einem Rahmen F wahr, wenn F reflexiv ist. Lediglich die Prinzipien des Systems K sind in allen Rahmen unabhängig von den spezifischen Eigenschaften der Zugänglichkeitsrelation wahr.

Bevor wir nun die den wichtigsten epistemischen Prinzipien entsprechenden Eigenschaften der Zugänglichkeitsrelation bzw. des Rahmens auflisten, führen wir zum besseren Verständnis zunächst einige dieser Eigenschaften ein. Sei $\mathrm{F} = \langle W, R \rangle$ ein Rahmen:

F ist *seriell*	$\forall w \in W \exists v \in W(wRv)$
F ist *reflexiv*	$\forall w \in W(wRw)$
F ist *transitiv*	$\forall w, v, u(wRv \wedge vRu \Rightarrow wRu)$
F ist *symmetrisch*	$\forall w, v(wRv \Rightarrow vRw)$
F ist *euklidisch*	$\forall w, v, u(wRv \wedge wRu \Rightarrow vRu)$

Der Zusammenhang zwischen den epistemischen Axiomen und den soeben aufgeführten Eigenschaften der Rahmen lässt sich nun wie folgt zusammenfassen.

Theorem Sei F ein Rahmen.

$F \vDash K\neg p \rightarrow \neg Kp \quad \Leftrightarrow \quad$ F *ist seriell;*

$F \vDash Kp \rightarrow p \quad \Leftrightarrow \quad$ F *ist reflexiv;*

$F \vDash Kp \rightarrow KKp \quad \Leftrightarrow \quad$ F *ist transitiv;*

$F \vDash p \rightarrow K\neg K\neg p \quad \Leftrightarrow \quad$ F *ist symmetrisch;*

$F \vDash \neg Kp \rightarrow K\neg Kp \quad \Leftrightarrow \quad$ F *ist euklidisch.*

Aus dem Theorem ergibt sich, dass die Systeme KD4 und KD45 semantisch der Klasse der seriellen und transitiven Rahmen bzw. der seriellen, transitiven und euklidischen Rahmen entsprechen. Die Systeme S4 und S5 entsprechen hingegen der Klasse der reflexiven und transitiven Rahmen bzw. der reflexiven und euklidischen Rahmen. Generell können sogenannte Vollständigkeitsresultate für besagte Systeme relativ zu den entsprechenden Klassen von Frames bewiesen werden. Eine System S der epistemischen Logik ist vollständig relativ zu einer Klasse C von Frames genau dann, wenn eine Formel in S genau dann bewiesen werden kann, wenn sie in jedem Frame der Klasse C wahr ist. Für eine explizite Darstellung dieser Vollständigkeits-resultate sowie weiterführende Diskussionen der Semantik sei erneut auf Blackburn u. a. (2001) verwiesen.

45.3 Multiagentensysteme

In den vorangegangenen Abschnitten wurden Systeme der epistemischen Logik für einzelne Agentinnen diskutiert: Es wurden Axiome für den Wissens- bzw. Glaubensoperator einzelner Agentinnen samt einer passenden Semantik gegeben. Was Agentinnen wissen bzw. glauben, scheint jedoch maßgeblich durch die Interaktion zwischen verschiedenen Agentinnen einer Gruppe und dem Austausch von Information innerhalb der Gruppe bestimmt zu sein. In unserer Definition der epistemischen Sprache hatten wir bereits Wissensoperatoren für eine endliche Anzahl von Agentinnen vorgesehen, uns jedoch in der folgenden Diskussion auf den Fall einer Agentin und demzufolge eines Wissensoperators beschränkt. Jetzt werden wir den allgemeineren Fall betrachten, in dem die Sprache Wissensoperatoren für eine endliche Menge $I = \{1, ..., n\}$ von Agentinnen enthält. In den Multiagentensystemen, die wir betrachten wollen, können nicht nur Aussagen über das Wissen einzelner Agentinnen getroffen werden, sondern auch über das Wissen von Gruppen $G \leq I$ von Agentinnen. In der Literatur werden zu diesem Zweck drei weitere Wissensoperatoren, die sich auf das Wissen der Gruppe beziehen, betrachtet: Der Operator E_G (*Shared Knowledge*), der Operator D_G (*Distributed Knowledge*) und schlussendlich der sogenannte *Common Knowledge*-Operator C_G. Für alle dieser Operatoren werden wir eine geeignete Zugänglichkeitsrelation unter Verwendung der Zugänglichkeitsrelationen der Agentinnen definieren, so dass die Wahrheitsbedingungen für die Operatoren des Gruppenwissens strukturell den Wahrheitsbedingungen der individuellen Wissensoperatoren ähneln. Die intendierte Lesart des Operators E_G ist ›jeder in der Gruppe weiß, dass‹. Dementsprechend wird die Zugänglichkeitsrelation des Operators als die Vereinigung über die Zugänglichkeitsrelation der einzelnen Agentinnen definiert: $wR_Ev{:}\Leftrightarrow \exists j \in G(wR_jv)$. Der Operator D_G hingegen beschreibt, was die Gruppe wissen kann, falls sie ihr Wissen kombiniert. Semantisch bedeutet dies, dass die Zugänglichkeitsrelation des Operators nur epistemische Alternativen betrachtet, die von allen Agentinnen als möglich erachtet werden. Formaler ausgedrückt ist die Zugänglichkeitsrelation des Operators R_D der Schnitt über die Zugänglichkeitsrelation der einzelnen Agentinnen der Gruppe: $wR_D v :\Leftrightarrow \forall j \in G(wR_j v)$. Der *Common Knowledge*-Operator C_G ist eine Art Verallgemeinerung des Operators E_G. Etwas ist *Common Knowledge*, d. h. Allgemeinwissen, wenn es nicht nur alle wissen, sondern auch alle wissen, dass alle es wissen, was wiederum alle wissen usw. C_G ist also eine Iteration des Operators E_G, genauer gesagt, die (unendliche) Konjunktion aller endlichen Iterationen des E_G-Operators. Dies lässt sich formal präzisieren, indem wir eine Hierarchie von *Shared Knowledge*-Operatoren definieren. Wir setzen $E^0_G\varphi := \varphi$ und $E^{k+1}_G\varphi := E_G E^k_G\varphi$. Der Operator C_G lässt sich dann wie folgt definieren:

$$C_G\varphi := \forall k \in \mathbb{N}(E^k_G\varphi).$$

Diese syntaktische Definition des *Common Knowledge*-Operators lässt sich auch semantisch nachvollziehen. Eine epistemische Alternative ist in diesem Sinne zugänglich, wenn Sie mittels der Relation R_E in endlich vielen Schritten erreicht werden kann:

$$wR_C v :\Longleftrightarrow \exists m \in \mathbb{N} \exists u_1, \ldots, u_m(wR_E u_1 \wedge u_1 R_E u_2 \wedge \ldots \wedge u_m R_E v).$$

Mittels der Zugänglichkeitsrelationen der drei verschiedenen Gruppen-Wissensoperatoren können die Wahrheitsbedingungen für die Operatoren in der üblichen Weise angegeben werden:

$$(M, w) \vDash E_G\varphi \Leftrightarrow \forall v(wR_E v \Rightarrow (M, v) \vDash \varphi)$$

$$(M, w) \vDash D_G\varphi \Leftrightarrow \forall v(wR_D v \Rightarrow (M, v) \vDash \varphi)$$

$$(M, w) \vDash C_G\varphi \Leftrightarrow \forall v(wR_C v \Rightarrow (M, v) \vDash \varphi).$$

In der Sprache der epistemischen Logik kann der Operator E_G mittels der Konjunktion aller individuellen Wissensoperatoren der Agentinnen der Gruppe $G = \{a_1, \ldots, a_l\} \subseteq I(l \leq n)$ definiert werden:

$$E_G\varphi := K_1\varphi \wedge \ldots \wedge K_l\varphi.$$

Im Gegensatz hierzu können die Operatoren D_G und C_G nicht innerhalb der epistemischen Sprache definiert werden. Für den Operator C_G mag dies verwunderlich erscheinen, da zuvor eine Definition angegeben wurde. Um jedoch die Konjunktion aller endlichen Iterationen des E_G-Operators angeben zu können, bedarf es einer unendlichen Konjunktion, die in der Sprache der epistemischen Logik nicht gebildet werden kann. Tatsächlich kann gezeigt werden, dass die Operatoren D_G und C_G der epistemischen Sprache – und damit den Systemen der epistemischen Logik – eine größere Ausdruckskraft verleihen. Für Multiagentensysteme muss deswegen die Syntax der Sprache wie folgt erweitert werden:

$$\varphi ::= p_i | \neg\varphi | (\varphi \wedge \varphi) | K_j\varphi | D_G\varphi | C_G\varphi$$

wobei $p_i \in At, j \in I \neq \emptyset$ und $G \subseteq I$.

Die Multiagentensysteme erweitern die Systeme der epistemischen Logik für einzelne Agentinnen durch Axiome für die Operatoren D_G und C_G. Sei $G = \{a_1, \ldots, a_l\}$:

(Ax-C_G) $\quad C_G\varphi \leftrightarrow E_G(\varphi \wedge C_G\varphi)$

(R-C_G) $\quad \dfrac{\varphi \rightarrow E_G(\psi \wedge \varphi)}{\varphi \rightarrow C_G\psi}$

(Ax1-D_G) $\quad D_{\{a_i\}}\varphi \leftrightarrow K_{a_i}\varphi$, für $1 \leq i \leq l$

(Ax2-D_G) $\quad D_G\varphi \rightarrow D_{G'}\varphi$, falls $G \subseteq G' \subseteq I$.

Für *Distributed Knowledge* müssen zudem noch die Axiome des individuellen Wissensoperators des entsprechenden Systems angenommen werden. Das heißt, wird der Wissensoperator K durch das modallogische System S5 charakterisiert, so müssen die modallogischen Axiome (K), (T) und (5) sowie die Regel (Nez) für den Operator D_G angenommen werden. Für die resultierenden Multiagentensysteme lassen sich dann wieder Vollständigkeitsresultate relativ zu den entsprechenden Klassen von Rahmen der Möglichen-Welten-Semantik für Multiagentensysteme angeben. Für Details sei auf Fagin u. a. (1995) verwiesen.

Während die Axiome des *Distributed Knowledge* in hohem Maße selbsterklärend sind, erscheinen die Axiome für *Common Knowledge* auf den ersten Blick weniger transparent. Jedoch ergibt sich ihre Bedeutung nach kurzer Reflexion. Das Axiom (Ax-C_G) sagt aus, dass φ Allgemeinwissen einer Gruppe ist, gdw. von allen in der Gruppe gewusst wird, dass φ der Fall und dass φ Allgemeinwissen ist. Dies scheint eine intuitive plausible Charakterisierung von *Common Knowledge* bzw. Allgemeinwissen zu sein. Das Axiom (Ax-C_G) wird auch *Fixpunktaxiom* genannt: Wenn $C_G\varphi$ für x in der Formel $E_G(\varphi \wedge x)$ eingesetzt wird, so ist die resultierende Formel logisch äquivalent zu $C_G\varphi$ selbst. Die Regel (R-C_G) wird Induktionsregel genannt: Wenn aus φ folgt, dass $\psi \wedge \varphi$ von allen Mitgliedern der Gruppe gewusst wird, so kann geschlossen werden, dass φ impliziert, dass ψ Allgemeinwissen ist. Zunächst ist zu beachten, dass die Regel wie die Nezessitationsregel eine Herleitungsregel ist, d. h. keine Einführungsregel im Sinne der Kalküle des natürlichen Schließens. Der Name der Regel ist zum besseren Verständnis der Regel hilfreich: Unter Annahme des Antezedens der Regel kann man sich per Induktion davon überzeugen, dass für alle natürlichen Zahlen n, $\varphi \rightarrow E^n_G(\psi \wedge \varphi)$ gilt und warum dies plausibel erscheint. Da die unendliche Konjunktion aller E^n_G Allgemeinwissen ist, ergibt sich die Regel (R-C_G) aus dieser Überlegung.

45.4 Logische Allwissenheit

In den bisherigen Ausführungen wurde den Agentinnen stets eine starke logische Kompetenz zugeschrieben, in der Tat eine zu starke logische Kompetenz, falls die Agentinnen keinen starken Idealisierungsbedingungen unterworfen werden und als Personen oder Maschinen verstanden werden sollen, die physischen Limitierungen unterliegen. Die logische Kompetenz der Agentinnen ergibt sich aus dem Axiom (K) und der Nezessitationsregel (Nez), die in allen Rahmen der Möglichen-Welten-Semantik gelten. Diese Prinzipien implizieren, dass, falls etwas gewusst wird, stets alle logischen Folgerungen ebenfalls gewusst werden. Das heißt, die Menge der gewussten Sätze ist stets unter logischer Folgerung abgeschlossen. Nun ist davon auszugehen, dass der renommierte Logiker Harvey Friedmann die Axiome der Robinson Arithmetik weiß: es handelt sich lediglich um sieben Axiome. Jedoch sind nicht einmal Harvey Friedman alle logischen Folgerungen dieser Axiome bekannt. So ist beispielsweise unbekannt, ob die Goldbachsche Vermutung eine Folgerung der Axiome ist, die Axiome sie widerlegen oder sie von den Axiomen unabhängig ist. Folgt die Goldbachsche Vermutung jedoch aus den Axiomen der Robinson Arithmetik, so weiß Harvey Friedman, gemäß der Systeme der epistemischen Logik, dass jede gerade Zahl, die größer ist als zwei, die Summe zweier Primzahlen ist.

Wird eine Charakterisierung des Wissensbegriffs unter geringeren Idealisierungsbedingungen angestrebt, so müssen das Axiom (K) und die Regel (Nez) aufgegeben und damit auch der Rahmen der klassischen Möglichen-Welten-Semantik verlassen werden. Die Abschwächung der logischen Abschlussbedingungen kann in unterschiedlicher Stärke erfolgen. Eine vergleichsweise geringe Abschwächung der logischen Abschlussbedingungen und somit der Idealisierungsannahmen erfolgt in der sogenannten ›Neighborhood Semantik‹ (vgl. Fagin u. a. 1995). Die Neighborhood Semantik garantiert nicht mehr den Abschluss des Wissens unter logischer Folgerung, sondern lediglich unter logischer Äquivalenz. Allerdings kollabiert die Neighborhood Semantik formal gesehen in die klassische Mögliche-Welten-Semantik, falls das Prinzip

$$K(\varphi \wedge \psi) \leftrightarrow K\varphi \wedge K\psi$$

akzeptiert und auch nur eine logische Wahrheit gewusst wird.

Der Abschluss des Wissensbegriffs unter logischer Äquivalenz ist weiterhin eine starke Idealisierungsannahme, so dass die Neighborhood Semantik auch von diesem Gesichtspunkt her nicht unbedingt befriedigend ist. Eine Alternative zur Neighborhood Semantik bietet eine Variante der klassischen Mögliche-Welten-Semantik, in der sogenannte ›unmögliche‹ Welten eingeführt werden. Unmögliche Welten sind epistemische Alternativen, in denen die Gesetze der Logik außer Kraft sind. In dieser Semantik bleiben nichts desto trotz die Begriffe der logischen Wahrheit, der logischen Äquivalenz und der logischen Folgerung unverändert, da sie relativ zu der Menge der möglichen Welten bestimmt werden. Der Wissensoperator wird jedoch auch relativ zu unmöglichen Welten evaluiert. Das heißt, Agentinnen erachten auch unmögliche Welten als epistemische Alternativen, so dass der Wissensoperator nicht zwangsläufig unter logischer Folgerung bzw. logischer Äquivalenz abgeschlossen ist. Unmögliche-Welten-Semantik ist eine hyperintensionale Semantik. Die Semantik erklärt *a weiß, dass ...* und *a glaubt, dass ...* zu *hyperintensionalen* Kontexten in denen logisch äquivalente, d. h. co-intensionale, Ausdrücke nicht mehr füreinander substituiert werden können. Generell werden in dieser Semantik keine logischen Gesetzmäßigkeiten im Skopus des Wissensoperators erhalten, was eine drastische Reaktion auf das Problem der logischen Allwissenheit darstellt (vgl. erneut Fagin u. a. 1995 für eine präzise Darstellung der Semantik).

In der epistemischen Logik und ihrer philosophischen Anwendungen sind je nach Diskussion oder Problem sowohl der idealisierte als auch der nicht oder weniger stark idealisierte Wissensbegriff opportun. Die bisher diskutierten Ansätze erlauben jedoch entweder den idealisierten oder den nicht idealisierten Wissensbegriff. Als Reaktion auf das Problem der logischen Allwissenheit unterscheiden einige Ansätze deswegen zwischen *explizitem*, nicht-idealisiertem und *implizitem*, idealisiertem Wissen. Zumeist wird hierfür die Syntax der Sprache erweitert, so dass diese Unterscheidung in der Sprache explizit gemacht werden kann. Zu nennen sind hier die *Awareness*-Semantik (vgl. Fagin u. a. 1995) sowie die *Justification*-Logik (Artemov 2008).

45.5 Quantifizierte Epistemische Logik

Bis jetzt haben wir lediglich Varianten der epistemischen Logik in aussagenlogischen Sprachen betrachtet. In diesem Abschnitt wollen wir einen kurzen Einblick in die epistemische Logik der ersten Stufe mit Identität geben. In der quantifiziert epistemischen Logik ergibt sich eine wichtige Differenzierung dieses Wissensbegriffs, die im Rahmen der aussagenlogischen epistemischen Logik nicht erfasst werden kann. Es handelt sich um die Unterscheidung zwischen Wissen *de re* und Wissen *de dicto*, die wir mittels folgendem Beispiels illustrieren wollen. Der natürlichsprachliche Satz

(b) Max glaubt, dass ein Logiker die Stelle erhalten hat.

hat zwei unterschiedliche Lesarten. Zum einen mag Max von einem bestimmten Logiker glauben, dass er die Stelle erhalten hat. Zum anderen kann der Satz so verstanden werden, dass Max überzeugt ist, dass ein beliebiger Logiker die Stelle erhalten hat. Semiformal können die Lesarten wie folgt präzisiert werden:

(b1) $\exists x(\text{Logiker}(x) \wedge \text{B}_{\text{Max}}\, \text{Stelle erhalten}(x))$.
(b2) $\text{B}_{\text{Max}} \exists x\, (\text{Logiker}(x) \wedge \text{Stelle erhalten}(x))$.

In (b1), der *de re*-Lesart, glaubt Max von einer bestimmten Person, dass Sie die Stelle erhalten hat, wohingegen in (b2), der *de dicto*-Lesart, Max lediglich eine Existenzaussage glaubt. Diese zwei Lesarten sind offensichtlich nicht äquivalent und müssen in der epistemischen Logik erster Stufe strikt unterschieden werden. Die Unterscheidung zwischen *de re* und *de dicto* hängt unmittelbar mit der referentiellen Opazität der Wissens- und Glaubenskontexte zusammen. Ein Kontext ist referentiell opak, wenn co-referentielle Terme in diesem Kontext nicht *salva veritate* füreinander substituiert werden können. Die referentielle Opazität lässt sich wie folgt illustrieren:

(a1) Max weiß, dass Carlos ein Terrorist ist.
(a2) Carlos = Ilich Ramírez Sánchez
(a3) Max weiß nicht, dass Ilich Ramírez Sánchez ein Terrorist ist.

Zumindest intuitiv betrachtet können alle drei Aussagen gleichzeitig wahr sein: So viel Max auch vom Terroristen Carlos gehört hat, so mag ihm nicht bekannt sein, das ›Carlos‹ der Alias von Ilich Ramírez Sánchez ist. Anders ausgedrückt, solange Maxens Wissen in (a3) nicht als *de re*-Wissen über Carlos, sondern *de dicto* verstanden wird, tritt, zumindest intuitiv betrachtet, kein Widerspruch auf. Wird diese Intuition ernst genommen, so ergeben sich einige Probleme für die klassische Identitätslogik sowie die referentielle Semantik, d. h. die Mögliche-Welten-Semantik für die epistemische Logik erster Stufe. Im Rahmen der Möglichen-Welten-Semantik bedeutet die Vereinbarkeit der Aussagen (a1–3), dass Max epistemische Alternativen betrachtet, in denen ›Carlos‹ und ›Ilich Ramírez Sánchez‹ nicht auf dieselbe Person Bezug nehmen. Wenn Terme aber unterschiedliche Referenten in unterschiedlichen epistemischen Alternativen haben, stellt sich die Frage, ob die Referenten auch in allen epistemischen Alternativen existieren und, allgemeiner, ob in allen epistemischen Alternativen auch dieselben Objekte existieren. Fällt die Antwort zu dieser Frage positiv aus, so kann die sogenannte *constant domain semantics* verwendet werden, in der alle epistemischen Alternativen stets den gleichen Individuenbereich haben. Jedoch kommutiert in einer solchen Semantik der Existenzquantor mit dem Wissensoperator, was, wie die Beispiele (b1) und (b2) nahelegen, der *de re*/*de dicto*-Unterscheidung nicht gerecht wird. Aus diesem Grund muss die *constant domain semantics* für die epistemische Logik erster Stufe als nicht adäquat erachtet werden. Stattdessen scheint in diesem Fall die allgemeinere, aber auch kompliziertere, *varying domain semantics*, in der der Individuenbereich von epistemischer Alternative zu epistemischer Alternative variieren kann, die bessere Wahl zu sein. Eine vollständige Axiomatisierung der Klassen von Rahmen der *varying domain semantics* erfordert jedoch zumeist die Verwendung einer freien Logik, d. h. eine Logik, in der nicht vorausgesetzt wird, dass jeder Term der Sprache ein Objekt denotiert.

Ein Rahmen F für die Mögliche-Welten-Semantik erster Stufe ist ein Tupel $\langle W,R,D,U \rangle$, wobei W eine nichtleere Menge von epistemischen Alternativen und R die epistemische Zugänglichkeitsrelation ist. $D: W \rightarrow \wp(U)$ ist eine Funktion, die jeder epistemischen Alternative einen nichtleeren Individuenbereich zuweist. Der Individuenbereich ist eine Teilmenge des Universums U, wobei wir annehmen, dass $U = \bigcup_{w \in W} D(w)$. Fordert man für alle $w,w' \in W$, dass $D(w) = D(w')$, so erhält man eine *constant domain semantics*, andernfalls handelt es sich um eine *varying domain semantics*. Ein Modell der Semantik erhält man, indem man einen Rahmen F um eine Interpretationsfunktion I ergänzt. I weist den Prädikaten und Termen eine Interpretation relativ zu einer epistemischen Alternative zu. Die Konsistenz der Aussagen (a1–3) kann in diesem Rahmen wie folgt erklärt werden. (a1) sagt aus, dass in allen (zugänglichen) epistemischen Alternativen v: $\text{I}(\textit{Carlos},v) \in \text{I}(\textit{Terrorist},v)$; (a2) sagt, dass in der gegenwärtigen epistemischen Alternative w_0: $\text{I}(\textit{Carlos},w_0) = \text{I}(\textit{I. R. Sánchez},w_0)$ und (a3) besagt, dass es eine epistemischen Alternative u gibt, so dass $\text{I}(\textit{I. R. Sánchez},u) \notin \text{I}\,(\textit{Terrorist},u)$. Das steht jedoch in keinem Widerspruch zu $\text{I}(\textit{Carlos},u) \in \text{I}(\textit{Terrorist},u)$, da unsere Semantik erlaubt, dass $\text{I}(\textit{Carlos},u) \neq \text{I}(\textit{I. R. Sánchez},u)$. In gewissem Sinne ist die Crux der Auflösung des Identitätspuzzles, dass Namen neben einer Extension auch eine Intension zugewiesen bekommen. Für die Substitution in Wissenskontexten ist es nicht ausreichend, dass Namen dieselbe Extension in einer, sei es auch die tatsächliche, epistemischen Alternative haben, sie müssen auch dieselbe Intension haben. So elegant die intensionale Behandlung von Namen erscheinen mag, so führt sie zu Schwierigkeiten im Rahmen der referentiellen Semantik. In der referentiellen Semantik werden Variablen rein extensional behandelt, sie sind lediglich Platzhalter für Objekte und nicht Intensionen. Werden Namen nun intensional behandelt, verhalten sich nicht alle Terme der Sprache gleich und die Identitätslogik von Variablen und Termen wird divergieren. Dieses Spannungsverhältnis zwischen Intensionalität und referentieller Semantik ist die zentrale Herausforderung für die Semantik der epistemischen Logik der ersten Stufe und, allgemeiner, der quantifizierten Modallogik. Eine ausführliche Diskussion der Probleme und unterschiedlichen Ansätze kann an dieser Stelle jedoch nicht geleistet werden. Wir verweisen auf die Arbeiten von Hughes/Cresswell (1996), Fitting/Mendelsohn (1998), Garson (2001) und Aloni (2005).

45.6 Ausblick

Das Ziel des Kapitels war eine Einführung in die wichtigsten Systeme einzelner Agentinnen sowie in die Multiagentensysteme der epistemischen Logik und deren Semantik. Unsere Diskussion dieser Systeme sollte deutlich gemacht haben, dass der diesen Systemen zugrunde liegende Wissensbegriff starken Idealisierungsbedingungen unterliegt, die nur schwer mit einem Wissensbegriff in Einklang zu bringen sind, der das Wissen von tatsächlichen Personen als Ausgangspunkt nimmt. Als weitere verwandte Schwäche der diskutierten Systeme erscheint, dass sie den differenzierten Wissensbegriff der verschiedenen Theorien des Wissens nicht gerecht werden. In neueren Arbeiten zur epistemischen Logik wird auf dieses Problem eingegangen, indem die epistemische Logik um zentrale Begriffe der Theorien des Wissens ergänzt wird. Stellvertretend sind hier die Arbeiten von Holliday (2015), der die epistemische Logik um Begriffe der modalen Konzeption des Wissens ergänzt, und Artemov (2008), der einen Rechtfertigungsbegriff in die epistemische Logik einführt, genannt.

Ein weiterer wichtiger Aspekt der epistemischen Logik, der in diesem Kapitel nicht betrachtet wurde, ist, wie sich Wissen durch den Erhalt neuer Informationen, z. B. durch die Interaktion mit anderen Agentinnen, verändert. Das heißt, Wissen wird nicht mehr nur statisch, zu einem bestimmten festgelegten Zeitpunkt betrachtet, sondern auch wie es sich durch den Erhalt von Informationen verändert. Solche dynamischen Ansätze verbinden Methoden der epistemischen Logik mit denen der Überzeugungsrevision und ermöglichen unter anderem alternative Lösungsansätze, wie mit den epistemischen Paradoxien, z. B. der Church-Fitch Paradoxie, umzugehen ist. Ein knapper Einblick in einige dynamische Ansätze wird in Égré (2011) gegeben, eine ausführliche Darstellung findet sich in van Ditmarsch u. a. (2008).

(Die Arbeit zu diesem Beitrag wurde durch die Europäische Kommission im Rahmen des Marie Sklodowska-Curie Individual Fellowships TREPISTEME (Projekt Nr. 703539) finanziert.)

Literatur

Aloni, Maria: Individual concepts in modal predicate logic. In: Journal of Philosophical Logic 34 (2005), 1–64.
Artemov, Sergei. N.: The Logic of Justification. In: Review of Symbolic Logic 1/4 (2008), 477–513.

Blackburn, Patrick/de Rijke, Maarten/Venema, Yde: Modal Logic. Cambridge 2001.
Égré, Paul: Epistemic logic. In: Horsten, L./Pettigrew, R. (Hg.): The Bloomsbury Companion to Philosophical Logic. London 2011, 503–542.
Fagin, Ronald/Halpern, Joseph Y./Moses, Yorav/ Vardi, Moshe Y.: Reasoning about Knowledge. Cambridge, Mass. 1995.
Fitting, Melvin/Mendelsohn, Richard L.: First-Order Modal Logic. Dordrecht 1995.
Garson, James W.: Quantification in Modal Logic. In: Gabbay/Guenther, F. (Hg.): Handbook of Philosophical Logic, Bd. 3. Dordrecht ²2001, 267–323.
Hintikka, Jaakko: Knowledge and Belief. Ithaca 1962.
Holliday, Wesley. H.: Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism. In: Journal of Philosophical Logic 44/1 (2015), 1–62.
Hughes, George Edward/Cresswell, Max: A New Introduction to Modal Logic. London 1996.
Lenzen, Wolfgang: Recent work in epistemic logic. In: Acta Philosophica Fennica 30/1 (1978), 1–219.
van Ditmarsch, Hans P./van der Hoek, Wiebe/Barteld, Kooi: Dynamic Epistemic Logic. Synthese Library Volume 337. Dordrecht 2008.
von Wright, Georg Henrik: An Essay in Modal Logic. Amsterdam 1951.
Williamson, Timothy: Knowledge and its limits. Oxford 2000.

Johannes Stern

VIII Methoden der zeitgenössischen Erkenntnistheorie