2 Aspekte entscheidungstheoretischer Grundlagen im Rahmen von Effectuation

In Sarasvathy (2009) werden entscheidungstheoretische Konzepte vorgestellt, die das Vorgehen effektuativ handelnder Entrepreneure begründen. Diese Ideen sollen im Folgenden präsentiert und diskutiert werden. Darüber hinaus werden Ansätze im Kontext von Effectuation erarbeitet, die über die bisher in der Literatur zu findenden Ausführungen hinausgehen. Sie stellen die Grundlage für das entwickelte effektuative Entscheidungsmodell dar.

2.1 Bayesianismus im Kontext von Effectuation

Die Verwendung von Methoden der wahrscheinlichkeitstheoretischen Strömung Bayesianismus durch effektuativ handelnde Entrepreneure wird in Sarasvathy (2009, S. 137–144) behandelt. Sarasvathy erläutert darin ihre Interpretation des bayesschen Wahrscheinlichkeitsbegriffs im Zuge der Bewertung von gründungsrelevanten Situationen und den Umgang mit diesen.

Bayesianismus wird in diesem Zusammenhang als Verfahren zur Steuerung von Zuständen der Natur beschrieben, die mit den eigenen Überzeugungen in Einklang gebracht werden können (Sarasvathy, 2009, S. 138). Im klassischen Sinne stellt Bayesianismus eine Möglichkeit zur Inferenz dar und wird zur Aktualisierung der persönlichen Überzeugungen im Hinblick auf die Zustände der Natur unter Verwendung gewisser Vorinformationen verwendet (Kumam et al., 2017).

Für die Handlungen von Seriengründern unter bayesschen Voraussetzungen sind zwei Interpretationen zu finden. Bezugnehmend auf klassische Ansichten wird in Sarasvathy (2009, S. 138) wie folgt argumentiert: Die Beobachtung, dass die Rate des Scheiterns von Unternehmen sehr hoch ist, erlaubt den Schluss, dass das Gründen mehrerer unabhängiger Unternehmen sinnvoll ist. Im Kontext von Effectuation kann das bayessche Theorem nach (Sarasvathy, 2009, S. 138) so interpretiert

M. Sterzel, *Effectuation entwickeln*,
https://doi.org/10.1007/978-3-658-39251-2_2

werden: Ungeachtet dessen, wie hoch die Wahrscheinlichkeit für das Scheitern von Unternehmen ist, kann der Erfolg des Entrepreneurs durch Seriengründen erhöht werden. Beide Deutungen resultieren im Seriengründen. Die Herangehensweise an die Bewertung einer Entscheidungssituation ist jedoch eine andere.

In Read et al. (2016) wird aufgezeigt, dass die Erläuterungen zu den bayesianischen Grundlagen in der Effectuation-Theorie nicht ausreichend sind und detaillierter dargestellt werden müssen. Um diesen Umstand Rechnung zu tragen, werden in den Abschnitten 2.1.1 und 2.1.2 bayessche Entscheidungsmethoden diskutiert und die Übertragung auf Effectuation behandelt.

2.1.1 Zum bayesschen Wahrscheinlichkeitsbegriff

Der Bayesianismus hat historisch gesehen eine Reihe von Interpretationen erlebt und wurde aus verschiedenen Perspektiven betrachtet. Es bildeten sich Hauptströmungen heraus, die den Bayesianismus in seiner heutigen Anschauung prägten. Hierbei sind insbesondere der Subjektive Bayesianismus, der Empirische Bayesianismus und der Logische Bayesianismus zu nennen (Corfield & Williamson, 2001).

Im Subjektiven Bayesianismus werden a-priori-Wahrscheinlichkeiten einzig und allein als Grad persönlicher und rationaler Überzeugung repräsentiert und unterliegen lediglich der Einschränkung, dass sie kohärent im Rahmen der vorliegenden Informationen sein müssen. In der Folge werden die ursprünglich getroffenen Annahmen bezüglich der a-priori-Wahrscheinlichkeit mit Hilfe hinzugewonnener Daten aktualisiert und resultieren in der a-posteriori-Wahrscheinlichkeit (de Finetti, 1974).

Der Empirische Bayesianismus stellt eine Kalibrierung des Subjektiven Bayesianismus dar. Die Grade persönlicher Überzeugung werden mit Hilfe von objektiven Häufigkeiten, sofern diese bekannt sind, ausgedrückt. Dieser Zusammenhang impliziert jedoch das Problem der Referenzklassen. Bayesianische Wahrscheinlichkeiten, nach dem Vorbild des Subjektiven Bayesianismus, beziehen sich auf einen einmaligen Fall, der mittels Sätzen oder Ereignissen formuliert wird. Häufigkeiten hingegen stützen sich auf einen übergeordneten Fall, der über Klassen von Ergebnissen definiert wird. Zu ermitteln, welche Häufigkeit mit welchem gegebenen Grad der Überzeugung kalibriert werden muss, ist nicht ohne Weiteres möglich (Ramsey, 1964). Das *Principal Principle* von D. Lewis (1980) versucht dieses Problem zu umgehen, indem es eine explizite Verbindung zwischen Graden der Überzeugung und objektiven Wahrscheinlichkeiten einmaliger Fälle postuliert.

Eine weitere Perspektive des bayesschen Wahrscheinlichkeitsbegriffs stellt der Logische Bayesianismus dar. Dieser beschreibt zum einen eine Wahrscheinlichkeit

$P(B|A)$ als den Grad, zu dem Ereignis A teilweise Ereignis B zur logischen Folge hat und zum anderen den Grad, zu dem ein rationaler Agent an das Eintreten von B glauben sollte, sofern dieser Kenntnis davon hat, dass A eingetreten ist. Dieser Ansatz suggeriert Objektivität, da zwei verschiedene Agenten mit dem selben Wissen nicht unterschiedliche Evidenzfunktionen verfolgen können, ohne sich dabei rational zu verhalten (Keynes, 1921, S. 32).

Empirischer und Logischer Bayesianismus können zum Objektiven Bayesianismus vereint werden. Zusammengefasst halten Objektive Bayesianisten das Postulat Subjektiver Bayesianisten für nicht ausreichend, dass eine Evidenzfunktion lediglich den Axiomen der Wahrscheinlichkeit genügen muss. Um als rational eingestuft zu werden, müssen weitere Bedingungen erfüllt sein (Corfield & Williamson, 2001, S. 2).

Bayesianismus hat bis heute Einzug in viele Wissenschaftsgebiete gehalten. So finden sich bayessche Erklärungsmodelle beispielsweise im Bereich der Künstlichen Intelligenz und der Wirtschaftswissenschaften wieder. Bayesianismus wird hierbei auch im Zusammenhang mit Kausalität diskutiert (Corfield & Williamson, 2001, S. 3 f.).

2.1.2 Das Bayes-Theorem

Das Bayes-Theorem bildet eine wichtige Grundlage bei der Betrachtung von Entscheidungen. Es dient somit der Modellierung von Lernprozessen. Bedingte Wahrscheinlichkeiten helfen, neue Informationen zu verarbeiten. Dadurch ist ein Agent in der Lage, bisher getroffene Entscheidungen zu überdenken. Die geistige Konstruktion der Umwelt kann angepasst und Situationen neu eingeschätzt werden (Wessler, 2012, S. 158).

Bedingte Wahrscheinlichkeiten stellen einen Zusammenhang zwischen der vor Informationserhalt vorhandenen (a-priori-) Wahrscheinlichkeit für den bestimmten Zustand S_s $(s = 1, 2, \ldots, n_S)$ und der (a-posteriori-) Wahrscheinlichkeit eben dieses Zustandes nach Eintritt des Informationsereignisses I_i $(i = 1, 2, \ldots, n_I)$ dar. Das Bayes-Theorem beruht auf bedingten Wahrscheinlichkeiten und drückt die stochastische Abhängigkeit zwischen den Zuständen und Informationsereignissen durch Wahrscheinlichkeiten aus (Laux et al., 2014, S. 304).

In der vorliegenden Arbeit werden ausschließlich diskrete Zustandsräume betrachtet, um die Komplexität der entwickelten Definitionen und Modelle auf ein Minimum zu beschränken. Zur mathematischen Beschreibung des Bayes-Theorems wird zunächst der Wahrscheinlichkeitsraum in der für diese Arbeit sinnvollen Form definiert.

Definition 2.1 (Wahrscheinlichkeitsraum)
Die Ergebnismenge eines Zufallsexperiments sei definiert als

$$\Omega := \{\omega_1, \omega_2, \ldots\}.$$

Das Ereignissystem $\mathcal{A}$ wird mit

$$\mathcal{A} := \mathcal{P}(\Omega)$$

als Potenzmenge der Ergebnismenge definiert.
Das Wahrscheinlichkeitsmaß P wird für $\mathcal{A}$ mit

$$P : \mathcal{A} \to [0, 1]$$

definiert und genüge für $A \in \mathcal{A}$ den von Kolmogoroff (1933, S. 2) eingeführten Axiomen:

1. $0 \leq P(A) \leq 1$,
2. $P(\Omega) = 1$,
3. $P\left(\bigcup_i A_i\right) = \sum_i P(A_1)$, *wenn* $A_i \cap A_j = \varnothing \; \forall \; i \neq j$

Das Tripel $(\Omega, \mathcal{A}, P)$ heißt Wahrscheinlichkeitsraum.

Es seien zudem die Zustände S_s für $s = 1, 2, \ldots, n_S$ und die Informationsereignisse S_s für $i = 1, 2, \ldots, n_I$ Ereignisse über Ω mit den Ergebnissen $\omega \in S_s$ und $\omega \in I_i$, wobei $S_s, I_i \subset \Omega$ ist.

Der Satz von Bayes kann ausgehend vom definierten Wahrscheinlichkeitsraum wie folgt dargestellt werden.

Definition 2.2 (a-priori-Wahrscheinlichkeit)
Sei $P(S_s)$ die a-priori-Wahrscheinlichkeit, dass der Zustand S_s ($s = 1, 2, \ldots, n_S$) eintritt, ohne Kenntnis vom Informationsereignis I_i ($i = 1, 2, \ldots, n_I$) zu haben.

Sei $P(I_i)$ die a-priori-Wahrscheinlichkeit für das Informationsereignis I_i ($i = 1, 2, \ldots, n_I$).

Dann heißt $P(I_i|S_s) = \frac{P(I_i \cap S_s)}{P(S_s)}$ Likelihood des Zustands S_s bezüglich des Informationsereignisses I_i unter der Voraussetzung, dass $P(S_s) \neq 0$ gilt.

Durch Kenntnis der Likelihood $P(I_i|S_s)$ und der a-priori-Wahrscheinlichkeit $P(S_s)$ ist die Berechnung der folgenden Wahrscheinlichkeiten möglich:

Definition 2.3 (a-posteriori-Wahrscheinlichkeit)
$P(S_s|I_i)$ heißt a-posteriori-Wahrscheinlichkeit für den Zustand S_s ($s = 1, 2, \ldots, n_S$) unter der Bedingung, dass I_i ($i = 1, 2, \ldots, n_I$) das Ergebnis der Beschaffung der Information ist, das heißt, dass das Informationsereignis I_i eingetreten ist.

Zur Bestimmung der Wahrscheinlichkeiten $P(I_i)$ und $P(S_s|I_i)$ gilt allgemein:

$$P(S_s \cap I_i) = P(I_i|S_s) \cdot P(S_s). \tag{2.1}$$

Hierbei ist $P(S_s \cap I_i)$ die Wahrscheinlichkeit dafür, dass sowohl das Informationsereignis I_i als auch der Zustand S_s eintritt. Offenbar gilt ebenfalls:

$$P(S_s \cap I_i) = P(S_s|I_i) \cdot P(I_i). \tag{2.2}$$

In Kombination mit Gleichung (2.1) folgt:

$$P(S_s|I_i) = \frac{P(I_i|S_s) \cdot P(S_s)}{P(I_i)}. \tag{2.3}$$

Die totale Wahrscheinlichkeit $P(I_i)$ des Informationsereignisses I_i kann folgendermaßen berechnet werden:

$$P(I_i) = \sum_{s=1}^{n_S} P(S_s \cap I_i) = \sum_{s=1}^{n_S} P(I_i|S_s) \cdot P(S_s) \ (i = 1, 2, \ldots, n_I), \tag{2.4}$$

sofern

$$\bigcup_{s=1}^{n_S} S_s = \Omega \text{ und } S_i \cap S_j = \varnothing \text{ für } i \neq j.$$

Gleichung (2.4) wird auch als *Satz der totalen Wahrscheinlichkeit* bezeichnet.

In Verbindung mit Gleichung (2.3) lässt sich der *Satz von Bayes* wie folgt beschreiben:

$$P(S_s|I_i) = \frac{P(I_i|S_s) \cdot P(S_s)}{\sum_{s=1}^{n_S} P(I_i|S_s) \cdot P(S_s)} \ (i = 1, 2, \ldots, n_I; s = 1, 2, \ldots, n_S). \tag{2.5}$$

Das Bayes-Theorem aus Gleichung (2.5) verdeutlicht, wie die a-posteriori-Wahrscheinlichkeiten $P(S_s|I_i)$ aus den a-priori-Wahrscheinlichkeiten $P(S_s)$ berechnet werden können, unter der Voraussetzung, dass die Likelihoods $P(I_i|S_s)$ gegeben sind.

Die Wahrscheinlichkeiten, die ein Agent nach Informationserhalt I_i den Zuständen $S_1, S_2, \ldots, S_{n_S}$ zuordnet, hängen, basierend auf Gleichung (2.5), von folgenden Faktoren ab:

- Festlegung der a-priori-Wahrscheinlichkeiten $P(S_1), P(S_2), \ldots P(S_{n_S})$ für die Zustände vor dem Erhalt zusätzlicher Informationen durch den Agenten
- Einschätzung der stochastischen Abhängigkeit zwischen den Zuständen S_1, $S_2, \ldots, S_{n_S}$ und den Informationsereignissen $I_1, I_2, \ldots, I_{n_I}$ vor Informationserhalt durch den Agenten in Form der bedingten Wahrscheinlichkeiten $P(I_i|S_s)$
- Tatsächliches Eintreten des Informationsereignisses I_i

Eine weitere Bedingung für den *Satz von Bayes* bezieht sich auf die Informationsereignisse $I_1, I_2, \ldots, I_{n_I}$. Diese sind nur dann prognoserelevant und damit für die Berechnung der a-posteriori-Wahrscheinlichkeiten essentiell, wenn ihre Ausprägungen stochastisch abhängig vom Zustand sind. Es ergibt sich sonst offenbar $P(S_s|I_i) = P(S_s)$. Es ist generell nur ein probabilistischer Rückschluss auf den Zustand möglich ist. Das Urteilen hinsichtlich des Zustandes verbessert sich jedoch durch das Auftreten prognoserelevanter Informationen (Laux et al., 2014, S. 305).

Aufbauend auf den vorangegangenen Definitionen wird im Folgenden eine vereinfachte Übertragung des Satzes von Bayes auf effektuatives Schließen vorgenommen. Dies ermöglicht den ursprünglich von Sarasvathy (2009, S. 138) rudimentär formulierten Vergleich von Effectuation und bayesschem Schließen anschließend zu diskutieren.

Definition 2.4 (Wahrscheinlichkeitsraum im Kontext von Effectuation)
Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ *sei konkret mit*

$$\Omega \mathrel{\hat{=}} zukünftige\ Erfolgssituationen\ der\ Unternehmen\ des\ Agenten$$

mit den Elementen $\omega_1, \omega_2, \ldots$ *bestimmt, wobei für die einzelnen Unternehmen* U_1, $U_2, \ldots$, *die mit* $+$ *(erfolgreich) oder* $-$ *(gescheitert) sein können, gilt*

$$\begin{aligned}
&\omega_1 : U_1^+ \\
&\omega_2 : U_1^- \\
&\omega_3 : U_1^+ U_2^+ \\
&\omega_4 : U_1^- U_2^+ \\
&\omega_5 : U_1^+ U_2^- \\
&\omega_6 : U_1^- U_2^- \\
&\omega_7 : U_1^+ U_2^+ U_3^+ \\
&\quad \vdots
\end{aligned}$$

Zudem seien die Ereignisse

$S_s \ldots$ *Anzahl der gegründeten Unternehmen des Agenten ist* s *für* $s = 1, 2, \ldots$ *und*

$I_i \ldots$ *Anzahl der gescheiterten Unternehmen des Agenten ist* i *für* $i = 1, 2, \ldots$

mit S_s, $I_i \subset \Omega$ *definiert.*
Offenbar gilt

$S = \Omega \setminus \{\omega_1, \omega_2\} \ldots$ *Anzahl der gegründeten Unternehmen des Agenten ist größer als 1,*

$I = \Omega \setminus \{\omega_1, \omega_3, \omega_7, \ldots\} \ldots$ *Anzahl der gescheiterten Unternehmen des Agenten ist mindestens1.*

$\mathcal{A}$ *und* P *seien gemäß Definition 2.1 festgelegt.*

Wird der *Satz von Bayes* auf das Beispiel von Sarasvathy aus Sarasvathy (2009, S. 138) angewendet, ist folgende Definition zweckmäßig:

Korollar 2.5 (Satz von Bayes im Kontext von Effectuation)
$P(I) \hat{=}$ *Wahrscheinlichkeit dafür, dass mindestens ein gegründetes Unternehmen des Agenten, in diesem Fall der Entrepreneur, scheitert.*
$P(S) \hat{=}$ *Wahrscheinlichkeit dafür, dass der Agent mehr als ein Unternehmen gründet.*

$P(I|S) \mathrel{\hat{=}}$ Wahrscheinlichkeit dafür, dass der Agent mit mindestens einem Unternehmen scheitert, unter der Annahme, dass er mehr als ein Unternehmen gründet.
$P(S|I) \mathrel{\hat{=}}$ Wahrscheinlichkeit dafür, dass der Agent mehr als ein Unternehmen gründet, unter der Bedingung, dass mindestens ein Unternehmen gescheitert ist.

Wie bereits im Abschnitt 2.1 angedeutet, würde ein Agent bzw. Serienentrepreneur im klassischen Sinne des Bayesianismus aus der Beobachtung, dass viele Unternehmen scheitern, schließen, dass es der Gründung mehrerer Unternehmen bedarf. Sarasvathy (2009, S. 138) schreibt dazu konkret:

> „I observe that the probability of firm failure is very high. Therefore I will start several firms."

Sarasvathy definiert jedoch nicht hinreichend genau, ob mit der Beobachtung die a-priori-Wahrscheinlichkeit $P(I)$ oder die bedingte Wahrscheinlichkeit $P(I|S)$ gemeint ist. Aufgrund der Schlussfolgerung, dass der Agent mehrere Unternehmen gründet, wenn er beobachtet, dass viele Unternehmen scheitern, ist anzunehmen, dass $P(S|I)$ die a-posteriori-Wahrscheinlichkeit darstellt. Diese besagt, wie wahrscheinlich es ist, dass ein Agent mehrere Unternehmen gründet, sofern er beobachtet hat, dass mindestens ein bereits von ihm gegründetes Unternehmen gescheitert ist. Folglich kann $P(I)$ nur als a-priori-Wahrscheinlichkeit verstanden werden, dass mindestens ein vom Agenten gegründetes Unternehmen scheitert.

Unter der Voraussetzung, dass neben $P(I)$ die Wahrscheinlichkeiten $P(S)$ und $P(I|S)$ bekannt sind, kann mit Hilfe des *Satz von Bayes* aus Gleichung (2.5) ermittelt werden, wie hoch die Wahrscheinlichkeit des Agenten ist, mehr als ein Unternehmen zu gründen, falls mindestens ein Unternehmen scheitert. Diese Wahrscheinlichkeit ist kleiner als 1, solange die Wahrscheinlichkeit dafür, dass der Agent mit einem Unternehmen scheitert, größer als 0 ist.

Im effektuativen Fall, argumentiert Sarasvathy (2009, S. 138 f.), wird der Agent ebenfalls mehrere Unternehmen gründen, unabhängig davon wie hoch die Wahrscheinlichkeit ist, dass Unternehmen scheitern. Da Sarasvathy davon ausgeht, dass bei einer effektuativen Interpretation des Bayesianismus der Agent die Bedingungen gestalten möchte, beispielsweise durch das Gründen mehr als eines Unternehmens, ist anzunehmen, dass der Zweck die Gründung mindestens eines erfolgreichen Unternehmens ist. Konkret schreibt Sarasvathy (2009, S. 138) dazu:

> „In the effectual interpretation, however, the entrepreneur reasons as follows: irrespective of what the probability of firm failure is, I can increase the probability of 'my' success through serial entrepreneurship."

Der Argumentation Sarasvathys folgend, versucht ein effektuativ handelnder Agent mehrere Unternehmen zu gründen, um die Wahrscheinlichkeit für seinen Erfolg zu erhöhen. Diese wird durch die bedingte a-posteriori-Wahrscheinlichkeit $P(S|I)$ ausgedrückt. Die Behauptung, dass Bayesianismus unter effektuativen Gesichtspunkten vielmehr ein Steuerungsmechanismus, als ein Inferenzmechanismus ist, kann demnach nicht bestätigt werden. In beiden Fällen werden Wahrscheinlichkeiten für das Eintreten eines bestimmten Zustandes geschätzt. Dieser wird im klassischen wie im effektuativen Fall anders definiert. Im klassischen Sinne wird die Wahrscheinlichkeit für das Eintreten des Ereignisses, dass der Agent mehr als ein Unternehmen gründet, geschätzt, unter der Bedingung, dass ein von ihm gegründetes Unternehmen gescheitert ist. Im effektuativen Beispiel wird versucht zu ermitteln, wie hoch die Wahrscheinlichkeit für den Erfolg des Unternehmens des Agenten ist, unter der Voraussetzung, dass $P(I)$ manipulierbar ist. $P(I)$ kann aus wahrscheinlichkeitstheoretischer Sicht jedoch nicht verändert werden, sondern ist lediglich beobachtbar als a-priori-Wahrscheinlichkeit (Peyrolón, 2020, S. 18). Unter Zuhilfenahme des *Satz von Bayes* aus Gleichung (2.5) kann die a-posteriori-Wahrscheinlichkeit $P(S|I)$ für das Gründen mehr als eines Unternehmens unter der Voraussetzung, dass mindestens ein bereits gegründetes Unternehmen des Agenten gescheitert ist (siehe Korollar 2.5), berechnet werden:

$$P(S|I) = \frac{P(I|S) \cdot P(S)}{P(I|S) \cdot P(S) + P(I|\overline{S}) \cdot P(\overline{S})} \tag{2.6}$$

Der Ausdruck $P(I|\overline{S})$ bildet die Wahrscheinlichkeit für das Eintreten von I unter der Bedingung, dass Ereignis S nicht eingetreten ist. $\overline{S}$ stellt das Komplement zu S dar, damit bildet der Ausdruck $P(\overline{S})$ die Wahrscheinlichkeit, dass Nicht-S eingetreten ist.

Die a-priori-Wahrscheinlichkeit $P(S)$ spielt im Kontext gründungsrelevanter Situationen eine wichtige Rolle, da sie den Grad der Überzeugung an das Eintreten eines bestimmten Zustandes S vor Betrachtung möglicher entscheidungsunterstützender Informationen angibt, die bei der Etablierung eines Unternehmens verhältnismäßig rar sind (Alvarez & Parker, 2009, S. 214 f.).

Wird das entwickelte Beispiel zum grundsätzlichen Gründungsverhalten im Kontext bayesschem Schließens auf Entscheidungssituationen im Gründungs- sowie Produkt- und Dienstleistungsentwicklungprozess erweitert, besteht die Notwendigkeit möglicherweise auftetende Ereignisse zu formulieren. Konkrete Ereignisse nach Treffen einer Entscheidung können im effektuativen Sinne veränderliche Ziele sein, die sich aus der Verwendung der zur Verfügung stehenden Mittel und dem kalkulierten leistbaren Verlust ergeben (Sarasvathy, 2009, S. 113 f.). Zur Bestimmung des Priors für das Eintreten eines Ereignisses und damit eines losen

Endes, sind in der Literatur verschiedene Methoden zu finden (Lemoine, 2019; Zondervan-Zwijnenburg et al., 2017). Die Maximum-Entropie-Methode, die in Jaynes (1957) vorgestellt wird, bietet die Möglichkeit a-posterio-Wahrscheinlichkeiten unter Ungewissheit zu modellieren. Das Verfahren beruht auf den Überlegungen zur Entropie im Rahmen der Informationstheorie (Shannon, 1948). Entropie stellt, neben ihrer üblichen Interpretation des mittleren Informationsgehaltes einer Nachricht, auch ein Maß für die Ungewissheit dar (Werner, 2008, S. 5).

Die Maximum-Entropie-Methode gilt als Verallgemeinerung des Indifferenzprinzips von Laplace. Dieses wird auch als *Prinzip vom unzureichenden Grund* bezeichnet und besagt, dass, in Abwesenheit zusätzlicher Informationen, sich gegenseitig ausschließende Ergebnisse mit einer diskreten Gleichverteilung anzusetzen sind (Kreinovich, 2008, S. 16 f.).

Zusammenfassend ist festzustellen, dass zur durch Sarasvathy (2009, S. 137–144) eingeführten Unterscheidung von bayesschem und effektuativem Schließen eine klare Definition der entscheidungsrelevanten Bestandteile benötigt wird. Mit der sich aus Definition 2.4 ergebenden Bildung des Wahrscheinlichkeitsraums und der darauf aufbauenden bayesschen Anwendung (siehe Korollar 2.5) im Kontext von Effectuation, können entscheidungstheoretische Elemente der Inferenzmechanismen transparent dargestellt werden. Ein Ansatz zur Modellierung effektuativen Entscheidens konnte damit nachgewiesen werden.

2.2 Ungewissheit

Das Thema Ungewissheit wird in Werken, die aus der Entrepreneurship-Forschung resultieren, divers diskutiert (Townsend et al., 2018, S. 564), bildet es doch die Grundlage für die Entscheidungsfindung im Prozess des Gründens (Packard et al., 2017, S. 1).

Eine rudimentäre Erklärung des Begriffs der Ungewissheit wird in Petrakis und Konstantakopoulou (2015, S. 11) geliefert. Demzufolge ist der Zeitablauf mit Veränderungen assoziert, wodurch Ungewissheit auftritt, die wiederum eine Schlüsselkomponente der Zukunft ist. Diese wiederum beinhaltet eine Kombination von Faktoren, die nicht einfach identifiziert und gesteuert werden können. Die Elemente werden *Möglichkeit*, *Gelegenheit*, *Zufall* oder *Glück* genannt.

Eine weitere Beschreibung bezeichnet Ungewissheit als eine vom Individuum wahrgenommene Unfähigkeit, etwas aufgrund des Mangels hinreichender Informationen genau vorherzusagen. Ungewissheit kann dabei noch einmal in Ungewissheit des Zustands, des Ergebnisses und der Antwort klassifiziert werden. Wobei sich die Ungewissheit des Zustandes auf den Mangel an Informationen über bestehende

Bedingungen bezieht. Ungewissheit des Ergebnisses bezieht sich auf das fehlende Wissen über den Zusammenhang von Ursache und Wirkung, während die Ungewissheit der Antwort das fehlende Wissen über die mögliche Rückmeldung des Marktes und weiterer Akteure, nachdem eine Handlung ausgeführt wurde, repräsentiert (Milliken, 1987, S. 134–143).

Ein Klassifizierungsschema von Ungewissheit im Kontext von unternehmerischen Gelegenheiten wird in Tomy und Pardede (2018, S. 615) geliefert. Ungewissheit kann entsprechend in die folgenden Ausprägungen gegliedert werden:

- Technologische Ungewissheit
- Politische Ungewissheit
- Wettbewerbsungewissheit
- Kundenungewissheit
- Ressourcenungewissheit

Technologische Ungewissheit beinhaltet die unzureichende Kenntnis über ein technologisches System oder das Vorhandensein weiterer technischer Lösungen, die dasselbe Problem adressieren (Meijer, 2008, S. 35). In der politischen Dimension beschreibt Ungewissheit das Fehlen von Informationen über das Verhalten von Regierungen, Regimen und politischen Dominanzen im Allgemeinen. Politische Faktoren haben Einfluss auf den Gestaltungsspielraum von Unternehmen (Rakesh, 2014, S. 20). Wettbewerbsungewissheit bezieht sich auf die mangelnde vollständige Kenntnis hinsichtlich der Konkurrenz und ihrer Produkte sowie Strategien, um am Markt zu bestehen (Yadav et al., 2006, S. 60). Das fehlende vollständige Wissen über die Nutzerakzeptanz und Nachfrage, bezogen auf ein Produkt oder eine Dienstleistung, werden als Kundenungewissheit klassifiziert (Gentry et al., 2013, S. 528). Ressourcenungewissheit stellt die Ungewissheit über die Verfügbarkeit von finanziellen und Human-Ressourcen dar (Meijer, 2008, S. 37).

Sarasvathy unterscheidet Ungewissheit weniger nach umweltbezogenen Gesichtspunkten, sondern vielmehr nach dem Wissen über mögliche Ergebnisse einer Entscheidung und deren zugrundeliegenden Verteilung (Sarasvathy, 2009, S. 26). Mit ihrer Taxonomie bezieht sie sich auf das Grundlagenwerk zur Ungewissheit *Risk, Uncertainty and Profit* von Knight.

Die Unterscheidung zwischen Risiko und Ungewissheit wird in Knight (1921) herausgestellt. Risiko wird hierbei als Fähigkeit charakterisiert, möglichen Umweltzuständen eine Wahrscheinlichkeitsverteilung zuzuordnen. Im Risikofall kann nicht mit Gewissheit angegeben werden, was als Nächstes passieren wird. Jedoch sind alle möglichen Umweltzustände bekannt, die eintreten können sowie die dazugehörige Wahrscheinlichkeitsverteilung (Townsend et al., 2018, S. 667).

Ungewissheit wird diesbezüglich nochmals in zwei Arten unterschieden. Im ersten Fall sind die möglichen Umweltzustände einer Handlung bekannt. Die Wahrscheinlichkeitsverteilung, der sie unterliegen, kann allerdings nicht bestimmt werden. Im zweiten Fall sind weder die von der Entscheidung des Entrepreneurs abhängigen Eintrittswahrscheinlichkeiten der Umweltzustände bekannt, noch sind die möglichen Umweltzustände bekannt, die aus der Entscheidung resultieren können (Sarasvathy, 2009, S. 26).

Die Differenzierung des begrifflichen Spektrums von Ungewissheit, welche für das Verständnis entrepreneurialer Entscheidungssituationen von zentraler Bedeutung ist, wird von Knight folgendermaßen zusammengefasst:

> „Uncertainty must be taken in a sense radically distinct from the familiar notion of risk, from which it has never been properly separated. [...] The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two is really present and operating. [...] It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all." (Knight, 1921, S. 19)

Das Treffen von Entscheidungen, in Situationen, die von knightscher Ungewissheit geprägt sind, bestimmen das Wesen von Entrepreneurship (Sarasvathy & Kotha, 2001, S. 32). Durch das Vorhandensein von Ungewissheit über zukünftige Ereignisse erhalten Unternehmer die Möglichkeit, trotz bestehender Marktgleichgewichte Gewinne zu erzielen (Blaug, 1997, S. 444). Sie können im Sinne von Schumpeter (1943, S. 83) durch „schöpferische Zerstörung" Innovationen hervorbringen.

2.3 Maschinelles Lernen im Kontext von Effectuation

Der im Abschnitt 2.1 diskutierte Zusammenhang zwischen Effectuation und bayesianischer Entscheidungstheorie sowie die Beziehung zwischen den verschiedenen Konzepten von Ungewissheit, wie sie in Abschnitt 2.2 vorgestellt wurden, bilden die Grundlage für die Entwicklung entscheidungsbezogener und lernbasierter Verfahren. Der Bereich des maschinellen Lernens, eine Unterkategorie der Künstlichen Intelligenz (Buxmann & Schmidt, 2018), umfasst eine Reihe solcher Verfahren, die von Ideen des Bayesianismus geprägt sind (Ghavamzadeh et al., 2015; Jun, 2016; Katt et al., 2019; Korb & Nicholson, 2004). Ein Ziel der Methoden des maschinellen Lernens ist es, menschliches Verhalten vorherzusagen und zu adaptieren (Plonsky et al., 2019) sowie Entscheidungen in verschiedenen Kontexten zu treffen.

Im Umfeld der Entrepreneurship-Forschung wird maschinelles Lernen in Entscheidungsprozessen vereinzelt angewendet (García et al., 2012; Haiyan, 2018). Bis auf einige Ausnahmen (Vilalta et al., 2004) lässt sich der überwiegende Teil der Ansätze des maschinellen Lernens in drei Kategorien klassifizieren (Marsland, 2014, S. 6; Murphy, 2012, S. 2):

- Supervised Learning,
- Unsupervised Learning und
- Reinforcement Learning.

Zur Lösung von Supervised-Learning-Problemen werden zusammengehörige Eingangs- und Ausgangsvariablen benötigt. Diese bilden einen Trainingsdatensatz, der aufgrund von Beobachtung eines Phänomens ermittelt wird und das gemeinsame Auftreten der Variablen repräsentiert. Werden mit x_i die Eingangsvariablen und mit t_i die Ausgangsvariablen bezeichnet, besteht ein Trainingsdatensatz aus den Tupeln (x_i, t_i) mit $i = 1, \ldots, m$ Lerndatensätzen (Marsland, 2014, S. 15 f.). Eine Hypothese $h(x)$ soll sich dem Zielvektor t möglichst genau annähern und somit erlernt werden. Mit $h(x)$ können dadurch für Werte von x auch außerhalb der Menge der Trainingsdatensätze Voraussagen getroffen werden. Um die approximierende Funktion $h(\cdot)$ zu lernen, existieren für unterschiedliche Anwendungsfälle verschiedene Verfahren, wie beispielsweise die Lineare und Logistische Regression sowie unterschiedliche Ausprägungen von Künstlichen Neuronalen Netzen (Mohri et al., 2018, S. 6 f.). Supervised Learning adressiert entsprechend Klassifizierungs- und Regressionsprobleme (Marsland, 2014, S. 6 f.). Im Kontext von Entrepreneurship gibt es Veröffentlichungen, die Supervised-Learning-Methoden verwenden, um unternehmerische Phänomene zu beschreiben und zu erklären (Luis-Rico et al., 2020; Sabahi & Parast, 2020; Tan & Koh, 1996; Zekic-Susac et al., 2013). In Unsupervised-Learning-Problemen wird die Segmentierung von Daten behandelt. Zu den Eingangswerten von x existieren zunächst keine zuordenbaren Ausgangswerte, wie es für Supervised Learning Aufgaben mit der jeweiligen Ausgangsgröße t der Fall ist. Lernalgorithmen aus diesem Bereich sind bestrebt, Muster in den Eingangswerten zu erkennen, die nicht durch Strukturlosigkeit gekennzeichnet sind. Bekannte Verfahren zur Lösung von Unsupervised Learning Problemen und zur Clusterbildung sind k-means, Principal Components Analysis und Mixture of Gaussians. Methoden des Unsupervised Learnings haben ihre Entsprechung in der Statistik als Kerndichteschätzung (Ghahramani, 2004). In der Entrepreneurship-Literatur existieren für diese Kategorie des Machine Learnings ebenfalls Anwendungen (Hemalatha & Nayaki, 2014; Nunes & Balsa, 2013; Shirur et al., 2019; Zekic-Susac et al., 2013). Reinforcement Learning (in Folge abgekürzt mit RIL) stellt die dritte

Hauptkategorie des maschinellen Lernens dar. RIL befasst sich mit dem Lernen von der Zuordnung von Zuständen zu Aktionen, indem eine Belohnungsfunktion maximiert wird. Ein Agent interagiert dabei über die Zeit mit seiner Umgebung und versucht eine Strategie zu erlernen. Dieses Konzept beruht auf dem Modell eines Markov-Entscheidungsproblems (in Folge abgekürzt mit MDP für Markov Decision Process) (Sutton & Barto, 2018, S. 3 f.). RIL findet insbesondere dann Anwendung, wenn die Steuerung eines Prozesses im Zeitverlauf erlernt werden soll (Szepesvari, 1998). Folglich findet RIL Anwendung beispielsweise im Bereich des Autonomen Fahrens, der Robotik, von Empfehlungssystemen, Chatbots, Videospielen, des Ressourcen-Managements sowie autonomer Bildung (Dhingra et al., 2017; Mandel et al., 2016; Mao et al., 2016; Theocharous et al., 2015; Yannakakis & Togelius, 2018; You et al., 2019; S. Zhang et al., 2019).

Im Kontext von Entrepreneurship wendet Haiyan (2018) RIL zur Modellierung spieltheoretischer Abläufe zwischen Investor und Entrepreneur an, um das bestmögliche Vertrauenverhältnis beider Parteien zu erlernen. Damit weist Haiyan (2018) nach, dass die grundsätzliche Anwendung von RIL Methoden auf prozessorientierte Entrepreneurship-Phänomene möglich ist. Gupta et al. (2016) nimmt Bezug auf die bisher varianztheoretische Betrachtung von Effectuation und fordert zur weiteren Theorienbildung prozessorientierte Untersuchungen. Die grundlegende prozessbezogene Natur von RIL (Szepesvari, 1998) und der Bedarf nach einer prozesstheoretischen Analyse von Effectuation (Gupta et al., 2016) begründen den Einsatz von RIL zu Modellierungszwecken. Yang und Chandra (2013) fordern ebenfalls den Einsatz agentenbasierter Modelle unter Zuhilfenahme von Methoden der Künstlichen Intelligenz zur Beschreibung entrepreneurialen, respektive effektuativen, Verhaltens.

2.3.1 Reinforcement Learning als Methode zur Lösung entscheidungstheoretischer Probleme

In RIL versucht ein Agent innerhalb eines MDP die größtmögliche Belohnung zu erreichen, indem er sich durch eine Reihe von Zuständen bewegt und Aktionen ausführt (van Otterlo & Wiering, 2012, S. 10–15). Ein MDP ist, angelehnt an Littman et al. (1995), gekennzeichnet durch ein Tupel (S, A, p, r), wobei

S... die Menge von Zuständen mit $s, s' \in S$
A... die Menge von Aktionen mit $a \in A$
p... die Transitionsfunktion und
r... die Belohnungsfunktion

darstellen.

Die Transitionsfunktion ist definiert durch $p(s'|s, a)$ und repräsentiert die Wahrscheinlichkeit vom Zustand s und Ausführung der Aktion a durch den Agenten in den neuen Zustand s' zu kommen. Für die Belohnungsfunktion ist $r : S \times A \times S \to \mathbb{R}$ mit der erwarteten unmittelbaren Belohnung $r(s', a, s)$. Betrachtet man ein MDP als zeitlichen Prozess mit Zeitpunkten $t = 1, 2, \ldots$, wobei $s_t \in S$ die Repräsentation des Zustandes der Umgebung, $a_t \in A$ die gewählte Aktion des Agenten und $r_t = r(s_{t+1}, a_t, s_t)$ die erhaltene Belohnungen zum Zeitpunkt t darstellen, gilt zudem die Markov-Eigenschaft. Diese beschreibt die Annahme, dass Zustandsübergänge lediglich vom zuletzt besuchten Zustand abhängig und unabhängig von vorhergehenden Aktionen oder Zuständen sind (Mocanu et al., 2018). So gilt beispielsweise $p(s'|s, a) = P(s_{t+1} = s'|s_t = s, a_t = a) = P(s_{t+1} = s'|s_t = s, a_t = a, s_{t-1} = \tilde{s}, a_{t-1} = \tilde{a})$.

Die weitere Beschreibung der Bestandteile eines MDP beruht auf den Ausführungen von Otterlo und Wiering (2012). Demnach ist ein Zustand s der Menge S eine einzigartige Repräsentation von Ausprägungen von Merkmalen. Die Merkmale besitzen dabei nur den Umfang, der für die Modellierung und Lösung des zu behandelnden Problems relevant ist. Beispielsweise kann die gesamte Figuren-Konfiguration auf einem Schachbrett zu einem beliebigen Zeitpunkt eines Spiels einen Zustand darstellen.

Aktionen aus der Menge A stellen Möglichkeiten dar, um von einem Zustand s in einen anderen Zustand s' zu gelangen. Die Menge der Aktionen, die innerhalb eines Zustandes ausgeführt werden können, wird mit $A(s)$ für einen bestimmten Zustand $s \in S$ notiert, wobei $A(s) \subseteq A$. Für die Transitionsfunktion p gilt zudem die Bedingung, dass für alle Zustände $s \in S$ und alle Aktionen $a \in A(s)$ $\sum_{s' \in S} p(s'|s, a) = 1$ (van Otterlo & Wiering, 2012).

RIL beinhaltet eine Reihe von Algorithmen zur Lösung des MDP. Zentrale Elemente in RIL-Problemen sind Agenten, die versuchen, innerhalb einer Umgebung durch das Erhalten von Belohnungen ein Verhalten zu erlernen. Dieser Zusammenhang lässt sich vereinfacht in Abbildung 2.1 darstellen. Mitchell (1997, S. 2) beschreibt diesbezüglich konkret, wodurch ein Lernproblem gekennzeichnet ist:

> „A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Die von Mitchell (1997) dargestellte Erfahrung E stellt im Kontext von RIL das Erhalten einer Belohnung $r(s', a, s)$ sowie die Beobachtung eines Zustandes s dar. Zur Bestimmung der Aufgabe T und des Leistungsmaßes P werden in RIL weitere Konzepte eingeführt. Dazu gehören die Policy (zu deutsch etwa Strategie)

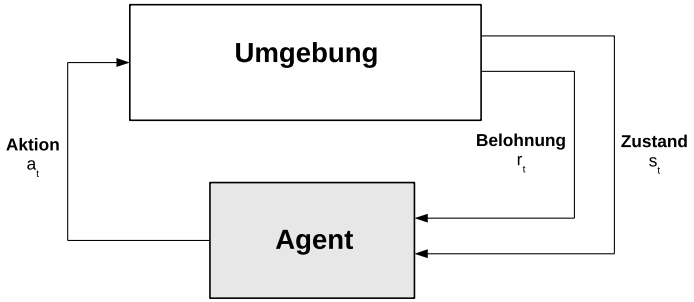

Abb. 2.1 Wechselwirkung zwischen Agent und Umgebung in einem RIL-Problem. (Modifiziert nach Amiri et al. (2018))

und die Value-Funktion (zu deutsch etwa Wertfunktion). Eine Policy π stellt in diesem Zusammenhang die Zuordnung von Zuständen $s \in S$ der Umgebung zu Aktionen $a \in A(s)$ dar. Verfolgt ein Agent eine Policy π zum Zeitpunkt t, dann ist $\pi(a|s)$ die Wahrscheinlichkeit dafür, dass $a_t = a$, wenn $s_t = s$, so dass $\pi : S \times A \rightarrow [0, 1]$. Die Policy $\pi(a|s)$ wird dann allgemein eine stochastische Policy genannt. Für den Fall, dass $\pi(a|s) = 1$, wenn zum Zeitpunkt t $a_t = a$ und $s_t = s$, und $\pi(a|s) = 0$, wenn $a_t \neq a$ und $s_t = s$, ergibt sich $\pi : S \rightarrow A$. Die Policy $\pi(s)$ wird deterministische Policy genannt. Eine Policy ist der Kern eines RIL-Agenten und bestimmt sein inhärentes Verhalten. Sutton und Barto (2018, S. 58) schreiben konkret, dass „RIL methods specify how the agent's policy is changed as a result of its experience.“. Eine Value-Funktion $v_\pi(s)$ beschreibt, welchen Gesamtbetrag an Belohnungen ein Agent über die Zeit erwarten kann, wenn er im Zustand s startet und anschließend der Policy π folgt. Damit wird die langfristige Erwünschtheit von Zuständen ausgedrückt, unter Berücksichtigung der zu erwartenden Zustände und den damit verbundenen Belohnungen. Ziel des Agenten ist es demzufolge, kumulierte Belohnungen zu maximieren, die er auf lange Sicht erhält (Sutton & Barto, 2018). Dieses Ziel entspricht der von Mitchell (1997) definierten Aufgabe T in RIL.

Die weiteren Ausführungen zu RIL-Konzepten beruhen auf einer nach Sutton und Barto (2018) angepassten Notation zur Bestimmung der (Action-)Value-Funktion und Policy. Zur Entwicklung der beiden Konzepte ist es notwendig, das Ziel des Agenten zu formalisieren. Die erwartete Gesamtbelohnung G_t kann als Funktion der Sequenz von erhaltenen Belohnungen nach Zeitpunkt t und dem finalen Zeitpunkt T mit

$$G_t \coloneqq r_{t+1} + r_{t+2} + r_{t+3} + \ldots + r_T \tag{2.7}$$

definiert werden. Die Bestimmung der Gesamtbelohnung G_t nach der Vorschrift (2.7) ist auf episodische RIL-Probleme mit einem Endzustand zum Zeitpunkt T anwendbar. T ist aufgrund der stochastischen Natur eines MDP respektive RIL-Problems eine Zufallsvariable, die von Episode zu Episode variiert.

Demgegenüber stehen Anwendungen, die fortlaufende Aufgaben beschreiben. Eine Gesamtbelohnung G_t kann dann nicht mehr wie in (2.7) bestimmt werden, da bei unendlichem Zeithorizont die zu maximierende Gesamtbelohnung G_t im Allgemeinen ebenfalls unendlich wird. Zur Lösung dieses Problems wird ein Diskontierungsfaktor γ eingeführt, für den $0 \leq \gamma \leq 1$ gilt. Ein Agent wird folglich eine Aktion $a_t \in A(s)$ so wählen, dass die erwartete diskontierte Gesamtbelohnung

$$G_t \coloneqq r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \tag{2.8}$$

maximiert wird. Belohnungen, die sich k Zeitschritte in der Zukunft befinden, werden entsprechend nur noch mit γ^{k-1} gewichtet. Für den Fall, dass $\gamma < 1$ gewählt wird, liefert die unendliche Summe aus Gleichung (2.8) einen endlichen Wert. Setzt man $\gamma = 0$, wird der Agent lediglich gewillt sein, die unmittelbare Belohnung r_{t+1} zu maximieren. Je größer γ gewählt wird, desto weitsichtiger wird der Agent im Hinblick auf die Einbeziehung künftiger Belohnungen. Darauf aufbauend kann die Gleichung aus (2.8) wie folgt zusammengefasst werden:

$$\begin{aligned} G_t \coloneqq & r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots \\ = & r_{t+1} + \gamma \left(r_{t+2} + \gamma r_{t+3} + \gamma^2 r_{t+4} + \cdots \right) \\ = & r_{t+1} + \gamma G_{t+1}. \end{aligned} \tag{2.9}$$

Zum Zwecke einer einheitlichen Notation von episodischen und fortlaufenden Aufgaben kann die Gesamtbelohnung G_t weiterhin mit

$$G_t \coloneqq \sum_{k=t+1}^{T} \gamma^{k-t-1} r_k \tag{2.10}$$

inklusive der Fälle, dass entweder $\gamma = 1$ oder $T = \infty$ gesetzt wird, formuliert werden.

Die eben erläuterten Vorschriften zur Bestimmung der Gesamtbelohnung ist zur Ermittlung der Value-Funktionen und Policy des Agenten von zentraler Bedeutung. Damit lässt sich das von Mitchell (1997) definierte Lernproblem auf RIL übertragen.

Die Unterschiede zwischen verschiedenen Policies stellen demnach einen Lerneffekt dar, wobei das Leistungsmaß P in RIL, zur Messung der Güte des Lerneffekts, im Folgenden anhand der Ausführungen von Sutton und Barto (2018) beschrieben werden soll.

Die Value-Funktion $v_\pi(s)$ eines Zustandes unter der Policy π ist die erwartete Gesamtbelohnung beim Start in Zustand s und der darauffolgenden Anwendung von π. $v_\pi(s)$ lässt sich mit

$$v_\pi(s) \coloneqq \mathbb{E}_\pi\left[G_t | s_t = s\right] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \middle| s_t = s\right] \tag{2.11}$$

formulieren. Da die Policy π sowie die Gesamtbelohnung G_t stochastisch sind, wird mit $\mathbb{E}_\pi[\cdot]$ der Erwartungswert einer Zufallsgröße bezüglich der durch π gegebenen Verteilung ausgedrückt. $v_\pi(s)$ ist eine Zufallsvariable.

Darauf aufbauend kann die erwartete Gesamtbelohnung für die Ausführung der Aktion a im Zustand s und danach der Policy π zu folgen mit $q_\pi(s,a)$ notiert werden. Formal lässt sich dies durch

$$q_\pi(s,a) \coloneqq \mathbb{E}_\pi\left[G_t | s_t = s, a_t = a\right] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \middle| s_t = s, a_t = a\right] \tag{2.12}$$

ausdrücken. Die Unterscheidung von $v_\pi(s)$ und q_π ist für die spätere Evaluierung von Lernalgorithmen von Bedeutung und kann wie folgt ausgedrückt werden:

$$v_\pi(s) = \sum_{a \in A} \pi\left(a|s\right) q_\pi(s,a). \tag{2.13}$$

Die Value-Funktion $v_\pi(s)$ kann zudem konkretisiert werden, indem die Transitionswahrscheinlichkeiten p eines MDP sowie die Verteilungsfunktion $\pi(a|s)$ in Gleichung (2.11) eingesetzt werden:

$$\begin{aligned} v_\pi(s) =& \mathbb{E}_\pi\left[G_t | s_t = s\right] \\ =& \mathbb{E}_\pi\left[r_{t+1} + \gamma G_{t+1} | s_t = s\right] \\ =& \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) \left(r(s',a,s) + \gamma \mathbb{E}_\pi[G_{t+1} | s_{t+1} = s']\right) \\ =& \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) \left(r(s',a,s) + \gamma v_\pi(s')\right) \;\forall\, s \in S \end{aligned} \tag{2.14}$$

(2.14) stellt die von Bellman (1957) formulierte rekursive Bellman-Gleichung dar. Diese ermöglicht es, für endliche MDP eine optimale Policy zu bestimmen. Endliche MDP sind dadurch gekennzeichnet, dass die dazugehörigen Zustands- und Aktionsmengen sowie die Menge der Belohnungen endlich viele Elemente enthalten.

Gesucht wird nun für jeden Zustand $s \in S$ eine Policy, die die zugehörige Value-Funktion maximiert. Diese wird als optimale Policy bezeichnet und mit π^* notiert. Es ergibt sich die optimale Value-Funktion

$$v_*(s) := \max_{\pi} v_\pi(s) \tag{2.15}$$

für alle $s \in S$.

Für die optimale Action-Value-Funktion π^* gilt analog

$$q_*(s,a) := \max_{\pi} q_\pi(s,a) \tag{2.16}$$

für alle $s \in S$ und $a \in A(s)$. Weiterhin ergibt sich der Zusammenhang zwischen q_* und v_*:

$$q_*(s,a) = \mathbb{E}\left[r_{t+1} + \gamma v_*(s_{t+1}) \middle| s_t = s, a_t = a) \right]. \tag{2.17}$$

Dieser Zusammenhang lässt sich auch wie folgt erklären:

$$\begin{aligned} v_*(s) &= \max_{a \in A(s)} q_*(s,a) \\ &= \max_{a \in A(s)} \mathbb{E}_{\pi^*}\left[G_t | s_t = s, a_t = a \right] \\ &= \max_{a \in A(s)} \mathbb{E}_{\pi^*}\left[r_{t+1} + \gamma G_{t+1} | s_t = s, a_t = a \right] \\ &= \max_{a \in A(s)} \mathbb{E}\left[r_{t+1} + \gamma v_*(s_{t+1}) | s_t = s, a_t = a \right] \\ &= \max_{a \in A(s)} \sum_{s'} p(s'|s,a) \left(r(s',a,s) + \gamma v_*(s') \right). \end{aligned} \tag{2.18}$$

Die Gleichung (2.18) stellt die Bellman-Optimalitäts-Gleichung dar. Analog gilt für die optimale Action-Value-Funktion q_*:

$$
\begin{aligned}
q_*(s,a) =& \mathbb{E}\left[r_{t+1} + \gamma v_*(s_{t+1}) | s_t = s, a_t = a\right] \\
=& \mathbb{E}\left[r_{t+1} + \gamma \max_{a' \in A(s_{t+1})} q_*(s_{t+1}, a') \middle| s_t = s, a_t = a\right] \\
=& \sum_{s'} p(s'|s,a) \left(r(s',a,s) + \gamma \max_{a' \in A(s')} q_*(s',a')\right). \qquad (2.19)
\end{aligned}
$$

Das Lösen der Optimalitätsgleichungen v_* und q_* und der damit verbundenen optimalen Policy π^* beschreibt das zentrale Lernproblem in RIL. Die beiden Funktionen v und q stellen in diesem Zusammenhang das von Mitchell (1997) definierte Leistungsmaß P dar. Nähern sich die Value-Funktionen v bzw. q den Fixpunkten v_* bzw. q_* an, ist der Agent im Begriff zu lernen. Zum Lösen der Bellman-Optimalitäts-Gleichungen existieren eine Reihe von Verfahren, die in Abschnitt 2.3.2 kurz vorgestellt werden.

2.3.2 Lösungsverfahren zur Bestimmung optimaler Policies in Reinforcement Learning

Klassische Verfahren zur Lösung von RIL-Problemen sind Algorithmen aus dem Bereich des Dynamic Programming (Busoniu et al., 2017). Verfahren dieser Art setzen voraus, dass alle Bestandteile eines MDP vollständig zur Verfügung stehen. Demnach müssen neben dem Zustandsraum, der Belohnungsfunktion und dem Aktionsraum auch die Transitionswahrscheinlichkeiten bekannt sein, welche in realen zu lösenden Problemen selten bekannt sind (Barto, 1995).

Dynamic Programming basiert auf dem Konzept der Generalisierten Policy Iteration, bei der die Policy Evaluation und die Policy Verbesserung im stetigen Wechsel stattfinden, um die Konvergenz hin zu einer optimalen Value-Funktion und optimalen Policy zu erreichen. Während der Policy Evaluation wird diesbezüglich angestrebt, eine Value-Funktion mit einer fixen Policy zu lernen. Im Fall der Policy Verbesserung wird eine Policy dahingehend angepasst, dass sie Aktionen aufnimmt, die hinsichtlich der aus der Policy Evaluation ermittelten Value-Funktion am besten sind (Sutton & Barto, 2018).

Formal kann ein Approximationsschritt der Policy Evaluation für fixierte Policy π als

$$
\begin{aligned}
v_{k+1}(s) &:= \mathbb{E}_\pi \left[r_{t+1} + \gamma v_k(s_{t+1}) | s_t = s \right] \\
&= \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) \left(r(s',a,s) + \gamma v_k(s') \right)
\end{aligned} \tag{2.20}
$$

für alle $s \in S$ ausgedrückt werden. Die Folge $\{v_k\}$ konvergiert gegen v_π für $k \to \infty$ (Bertsekas, 2011). Prozedural kann dieses Iterationsverfahren wie in Algorithmus 1 umgesetzt werden.

Algorithmus 1 Schätzung der Value-Funktion $V \approx v_\pi$

```
1: procedure ITERATIVE POLICY EVALUATION(π)
2:     θ > 0                    ▷ Schwellwert zur Bestimmung der Genauigkeit der Schätzung
3:     V(s) ← ℝ ∀s ∈ S
4:     V(s = terminal) ← 0
5:     repeat
6:         Δ ← 0
7:         for each s ∈ S do
8:             v ← V(s)
9:             V(s) ← Σ_a π(a|s) Σ_s' p(s'|s,a) [r(s',a,s) + γV(s')]
10:            Δ ← max(Δ, |v − V(s)|)
11:        end for
12:    until Δ < θ
13: end procedure
```

Die Policy Evaluation und die damit einhergehende Berechnung der Value-Funktion dient der Ermittlung besserer Policies (Mansour & Singh, 1999). Eine Policy $\pi(s)'$ ist nicht schlechter als π, wenn für alle $s \in S$

$$
q_\pi(s, \pi'(s)) \geq v_\pi(s) \tag{2.21}
$$

gilt. Eine bessere oder mindestens genauso gute Policy π' kann im Allgemeinen mittels

$$
\begin{aligned}
\pi'(s) &:= \arg\max_{a \in A(s)} q_\pi(s,a) \\
&= \arg\max_{a \in A(s)} \mathbb{E} \left[r_{t+1} + \gamma v_\pi(s_{t+1}) | s_t = s, a_t = a \right] && (2.22) \\
&= \arg\max_{a \in A(s)} \sum_{s'} p(s'|s,a) \left(r(s',a,s) + \gamma v_\pi(s') \right) && (2.23)
\end{aligned}
$$

bestimmt werden, da die Gleichungen (2.22) und (2.23) die Bedingungen, die sich aus (2.21) ergeben, erfüllen.

Dieser als Policy Iteration bezeichnete Wechsel zwischen Policy Evaluation und Policy Verbesserung kann durch das Value-Iterationsverfahren verkürzt werden und somit schneller konvergieren (Sutton & Barto, 2018). Es lässt sich für alle $s \in S$ durch

$$\begin{aligned} v_{k+1}(s) &:= \max_{a \in A(s)} \mathbb{E}\left[r_{t+1} + \gamma v_k(s_{t+1}) | s_t = s, a_t = a\right] \\ &= \max_{a \in A(s)} \sum_{s'} p(s'|s,a)\left(r(s',a,s) + \gamma v_k(s')\right) \end{aligned} \tag{2.24}$$

formalisieren.

Grundsätzlich können das Policy-Iterations- sowie Value-Iterationsverfahren als Generalisierte Policy Iteration kategorisiert werden. Die Wechselwirkungen zwischen Policy Evaluation und Verbesserung werden in Abbildung 2.2 illustriert. Nahezu alle Lernverfahren in RIL lassen sich als Generalisierte Policy Iteration beschreiben (Sutton & Barto, 2018).

Dynamic Programming wird dem Bereich der Model-Based-Methoden zugeordnet. Wie bereits erwähnt ist es bei der Verwendung diesbezüglicher Verfahren notwendig, alle Bestandteile eines MDP zu kennen (F. L. Lewis & Vrabie, 2009). Jedoch sind die Transitionswahrscheinlichkeiten sowie die Belohnungsfunk-

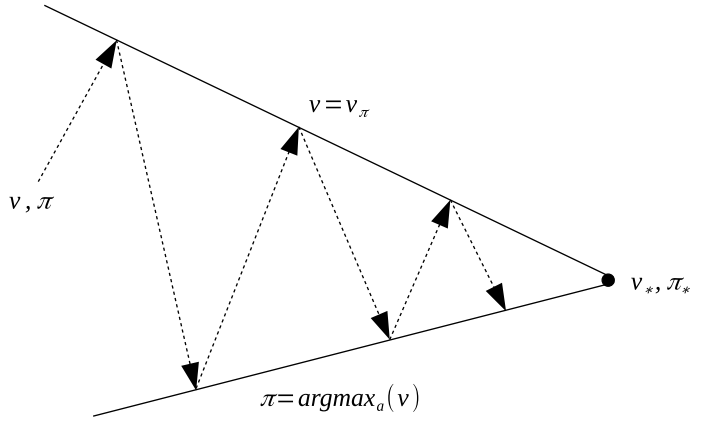

Abb. 2.2 Generalisierte Policy Iteration. (Modifiziert nach Sutton und Barto (2018))

tion nicht immer gegeben (Sutton & Barto, 2018). Zur Lösung von RIL-Problemen, bei denen kein vollständiges Wissen der Umgebung vorhanden ist, werden Model-Free-Methoden angewandt. Diese Verfahren beruhen auf stichprobenartigen Erfahrungen, die ein Agent durch sequentielle Interaktion mit seiner Umgebung erlangt (Strehl et al., 2006). Model-Free-Methoden wiederum lassen sich in On- und Off-Policy-Techniken einteilen. Dazu gehören nach Sutton und Barto (2018) beispielsweise:

- First-Visit Monte Carlo Steuerung (On-Policy)
- Importance Sampling (Off-Policy)
- Monte Carlo Vorhersage und Steuerung (Off-Policy)
- Ein-Schritt Temporal-Difference-Learning (Off-Policy)
- SARSA (On-Policy)
- Q-Learning (Off-Policy)

On-Policy-Verfahren dienen dazu, eine Policy zu verbessern, die gleichzeitig auch vom Agenten genutzt wird, um Entscheidungen bezüglich der auszuführenden Aktionen zu treffen. Bei Off-Policy-Methoden hingegen wird zwischen der Policy, die gelernt werden soll (Ziel-Policy), und der Policy, die für das Entscheidungsverhalten des Agenten verwendet wird (Verhaltens-Policy), unterschieden (Poole & Mackworth, 2017, Abschn. 11.3.6). Ziel-Policies enthalten im Allgemeinen explorative Bestandteile, während Verhaltens-Policies exploitativ hinsichtlich der bisher gelernten Policy sind. Eine konkrete Anwendung einer Off-Policy-Strategie wird in Abschnitt 4.2.2 vorgestellt.

2.3.3 Reinforcement Learning im entrepreneurialen Kontext

Um die in den Abschnitten 2.3.1 bis 2.3.2 vorgestellten RIL-Konzepte im Kontext entrepreneurialer Problemstellungen zu veranschaulichen, soll im Folgenden ein Beispiel entwickelt werden, das sich am *Recycling Robot Problem* von Connell (1989) orientiert und die Ideen von Csaszar und Levinthal (2016) zur Produktrepräsentation adaptiert.

Ein entrepreneurialer Agent hat die Aufgabe, ein Produkt erfolgreich am Markt zu etablieren. In diesem Zusammenhang stellt der Markt die Umgebung des Agenten dar. Anhand seiner zur Verfügung stehenden Mittel trifft der Agent Entscheidungen darüber, wie mit dem Produkt umgegangen werden soll. Der Zustandsraum S umfasst hierbei die Zustände `{hoch,gering}`, die sich auf die dem Agenten zur Verfügung stehenden Mittel beziehen. In jedem Zustand kann der

Agent entscheiden, ob er ein {Produkt anpassen}, ein {Produkt nicht verändern} oder ein {neues Produkt entwickeln} möchte. Im Beispiel wird die Annahme getroffen, dass der entrepreneuriale Agent bei einem Mittelbestand von {hoch} nicht daran interessiert ist, ein neues Produkt zu erstellen. Demnach ergeben sich die Aktionsmengen A(hoch)={Produkt anpassen, Produkt nicht verändern} und A(gering)={Produkt anpassen, Produkt nicht verändern, neues Produkt entwickeln}.

Der Agent erhält eine positive Belohnung, wenn sein Produkt Nachfrage am Markt erzeugt. Dies erreicht er am besten, indem er sein Produkt kontinuierlich anpasst und verbessert. Dieses Vorgehen führt jedoch zu einer Verringerung seiner zur Verfügung stehenden Mittel. Lässt der Agent das Produkt so wie es ist, verringern sich auch nicht seine Mittel. Im Falle, dass der Agent im Begriff ist sein Produkt anzupassen und ihm bei diesem Vorgang die Mittel ausgehen, ist er gezwungen, neue Mittel zu beschaffen. Tritt dies ein, wird der Agent durch Erhalt einer negativen Belohnung bestraft.

Die Wahrscheinlichkeit dafür, dass ein Agent, ausgehend von vielen zur Verfügung stehenden Mitteln und einer Produktanpassung, immer noch viele Mittel zur Verfügung hat, beträgt α; dafür dass er nach der Anpassung nur noch wenige Mittel hat, beträgt sie $1 - \alpha$. Besitzt der Agent zunächst wenige Mittel, nimmt eine Produktanpassung vor und besitzt danach immer noch wenige Mittel, ist diese Wahrscheinlichkeit β. Für den Fall, dass er mit den wenigen Mitteln eine Produktanpassung vornimmt und ihm die Mittel ausgehen, beträgt diese Wahrscheinlichkeit $1 - \beta$. Folglich ist der Agent gescheitert und er erhält eine Belohnung (Bestrafung) von $r_{\texttt{scheitern}}$. Er ist nun gezwungen neue Mittel zu beschaffen, um eine Unternehmung mit einem neuen Produkt starten zu können. Im entwickelten Beispiel beträgt entsprechend die Wahrscheinlichkeit dafür, dass der Agent einen hohen Mittelbestand hat, unter der Voraussetzung, dass er zunächst wenige Mittel hatte und ein neues Produkt entwickeln will, 1. Der Agent erhält eine Belohnung von 0. Für die erwarteten Belohnungen $r_{\texttt{Produkt anpassen}}$, $r_{\texttt{Produkt nicht verändern}}$ und $r_{\texttt{scheitern}}$ gilt $r_{\texttt{Produkt anpassen}} > r_{\texttt{Produkt nicht verändern}} > r_{\texttt{scheitern}}$.

Die Transitionswahrscheinlichkeiten sowie die erwarteten Belohnungen des beispielhaften MDP sind in Tabelle 2.1 dargestellt. Die Tabelle enthält jede Kombination von Zustandsübergängen, die aus $s \in S$ und der Aktion $a \in A(s)$ möglich sind. Für Zustandsübergänge, deren Transitionswahrscheinlichkeit $p(s'|s, a) = 0$ sind, sind keine Belohnungen definiert.

Zur Illustration der Transitionen sind zudem die Dynamiken des beispielhaften MDP in Abbildung 2.3 dargestellt. Ein großer weißer Kreis mit Text in dessen Inneren repräsentiert einen Zustandsknoten, während ein kleiner, schwarz ausgefüllter Kreis einen Aktionsknoten symbolisiert. Pfeile stellen in diesem Kontext

Tabelle 2.1 Tabellarische Darstellung eines beispielhaften MDP

s	a	s'	$p(s'\|s,a)$	$r(s',a,s)$
hoch	Produkt anpassen	hoch	α	$r_{\text{Produkt anpassen}}$
hoch	Produkt anpassen	gering	$1-\alpha$	$r_{\text{Produkt anpassen}}$
gering	Produkt anpassen	hoch	$1-\beta$	$r_{\text{scheitern}}$
gering	Produkt anpassen	gering	β	$r_{\text{Produkt anpassen}}$
hoch	Produkt nicht verändern	hoch	1	$r_{\text{Produkt nicht verändern}}$
hoch	Produkt nicht verändern	gering	0	–
gering	Produkt nicht verändern	hoch	0	–
gering	Produkt nicht verändern	gering	1	$r_{\text{Produkt nicht verändern}}$
gering	neues Produkt entwickeln	hoch	1	0
gering	neues Produkt entwickeln	gering	0	–

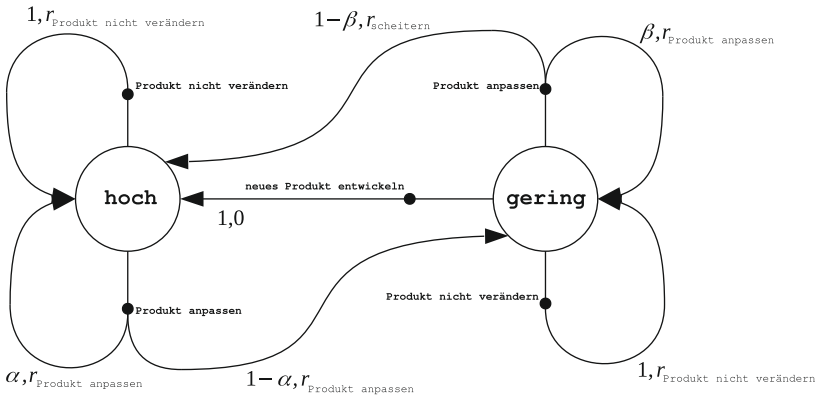

Abb. 2.3 Darstellung des beispielhaften Transitionssystems. (Modifiziert nach Sutton und Barto (2018))

die Übergänge, mit ihren Wahrscheinlichkeiten und erwarteten Belohnungen, von einem Zustandsknoten und des jeweilig gewählten Aktionsknotens in den von da aus erreichbaren Zustandsknoten dar. Zur Bestimmung der besten Policy für das dargestellte Problem wird die Bellman Optimalitätsgleichung aus 2.18 auf das entwickelte Beispiel angewendet. Die Zustände `hoch` und `gering` werden aus Gründen der Übersichtlichkeit mit `h` und `g`, die Aktionen `Produkt anpassen`, `Produkt nicht verändern` und `neues Produkt entwickeln` mit `an`, `nv` und `ne`

abgekürzt. Für die zwei Zustände h und g ergeben sich folglich die Gleichungen $v_*(\mathrm{h})$ und $v_*(\mathrm{g})$ mit

$$
\begin{aligned}
v_*(\mathrm{h}) &= \max \left\{ \begin{array}{r} p(\mathrm{h}\,|\,\mathrm{h},\mathrm{an})[r(\mathrm{h},\mathrm{an},\mathrm{h}) + \gamma v_*(\mathrm{h})] + p(\mathrm{g}\,|\,\mathrm{h},\mathrm{an})[r(\mathrm{h},\mathrm{an},\mathrm{g}) \\ +\gamma v_*(\mathrm{g})], \\ p(\mathrm{h}\,|\,\mathrm{h},\mathrm{nv})[r(\mathrm{h},\mathrm{nv},\mathrm{h}) + \gamma v_*(\mathrm{h})] + p(\mathrm{g}\,|\,\mathrm{h},\mathrm{nv})[r(\mathrm{h},\mathrm{nv},\mathrm{g}) \\ +\gamma v_*(\mathrm{g})] \end{array} \right\} \\
&= \max \left\{ \begin{array}{r} \alpha[r_{\mathrm{an}} + \gamma v_*(\mathrm{h})] + (1-\alpha)[r_{\mathrm{an}} + \gamma v_*(\mathrm{g})], \\ 1[r_{\mathrm{nv}} + \gamma v_*(\mathrm{h})] + 0[r_{\mathrm{nv}} + \gamma v_*(\mathrm{g})] \end{array} \right\} \\
&= \max \left\{ \begin{array}{r} r_{\mathrm{an}} + \gamma[\alpha v_*(\mathrm{h}) + (1-\alpha) v_*(\mathrm{g})], \\ r_{\mathrm{nv}} + \gamma v_*(\mathrm{h}) \end{array} \right\}
\end{aligned}
$$

bzw. in zusammengefasster Form

$$
v_*(\mathrm{g}) = \max \left\{ \begin{array}{r} \beta r_{\mathrm{an}} + r_{\mathtt{scheitern}}(1-\beta) + \gamma[(1-\beta) v_*(\mathrm{h}) + \beta v_*(\mathrm{g})], \\ r_{\mathrm{nv}} + \gamma v_*(\mathrm{g}), \\ \gamma v_*(\mathrm{h}) \end{array} \right\}.
$$

Es ergibt sich folglich für jede Kombination aus α, β und γ sowie r_{an}, r_{nv} und $r_{\mathtt{scheitern}}$ mit $0 \leq \alpha, \beta \leq 1$ und $0 \leq \gamma < 1$ jeweils ein Wertepaar $v_*(\mathrm{h})$ und $v_*(\mathrm{g})$, das gleichzeitig diesen beiden Gleichungen genügt.

Das entwickelte Beispiel veranschaulicht, dass RIL und dessen Funktionsweise (vgl. Abschnitte 2.3.1 und 2.3.2) als Modellierungsmethodik für eine entrepreneuriale Problemstellung anwendbar ist. Besonderheiten, die für die weitere Modellierung entrepreneurialen Entscheidens notwendig sind, basieren auf der Betrachtung der Umgebungsdynamik. Sofern dem entrepreneurialen Agenten die Transitionswahrscheinlichkeiten der Umgebung bekannt wären, wäre er in der Lage, anhand der dargestellten optimalen Value-Werte einer optimalen Policy zu folgen. Häufig sind dem Agenten diese Wahrscheinlichkeiten nicht bekannt und die Anwendung von Model-Free-Methoden kommt zum Tragen. Der Einsatz eines Verfahrens aus dieser Methodenklasse wird in Kapitel 4 thematisiert. Dabei wird der Bezug zu Abschnitt 2.2 hergestellt. Das Konzept der Ungewissheit – als zentrales Element entrepreneurialen Entscheidens – wird in Verbindung mit dem Modellierungsansatz der Model-Free-Methoden diskutiert.