

2 Theorie und Begriffe

Das folgende Kapitel dient der Begriffsbildung, die durch die Einführung verschiedener theoretischer und konzeptioneller Debatten und der detaillierten Herleitung ihrer Anwendung für den Kontext der vorliegenden Arbeit geleistet wird. Das Kapitel besteht aus drei Teilen: Am Anfang stehen Grundlagen aus den Bereichen Sozialtheorie, Sprach- und Medienwissenschaft zu den Begriffen *Praxis* und *Praktik* (Kap. 2.1). Daran schließt sich die ebenfalls sozialtheoretische, teilweise auf den vorherigen Begriffen aufbauende Debatte zum Verhältnis von Mensch und Maschine an (Kap. 2.2). Schließlich werden Begriffe und Konzepte aus der Medienaneignungs- bzw. Domestizierungsforschung vorgestellt und für den Kontext der Arbeit diskutiert (Kap. 2.3).

2.1 Praxeologische und medientheoretische Grundlagen

Die vorliegende Arbeit untersucht gebrauchsbasiert und auf einer empirischen Grundlage sprachliche *Praktiken* mit VUIs. Dieser Begriff wurde nicht zufällig gewählt: Die Analysen sollen sich mit Bezug auf den Praktiken-Begriff auf gleich zwei verschiedene Traditionen praxeologischer Theoriebildung stützen. Diese sollen im Folgenden zunächst kurz übersichtlich eingeführt werden, ehe sich eine genauere Betrachtung in den folgenden Unterkapiteln daran anschließt.

- (1) Erstens wird mit einem Verständnis von *sprachlichen Praktiken* im Sinne konversations- bzw. gesprächsanalytischer Forschung gearbeitet (siehe dazu Selting 2016). Sprachliche Praktiken sind in diesem Sinne wiederkehrend genutzte (Bündel von) „verbalen, vokalen und visuellen Ressourcen“ (Selting 2013: 589) bzw. sprachlich-interaktionalen Verfahren zur Lösung spezifischer kommunikativer Aufgaben. Sie stehen in einem konstitutiven Verhältnis zu den Handlungen, die sie hervorbringen (können). Von Interesse ist hier also mit Bezug auf den Gegenstand, welche Ressourcen in VUI-Dialogen⁴ zum Einsatz kommen, d. h., mit welchen sprachlich-interaktionalen Formen und Verfahren die Aufgabe eines Austauschs zwischen Mensch und Maschine bewältigt werden kann. Dabei orientiert sich die Analyse – mit leichten Anpassungen – an den von Schegloff (2006) formulierten allgemeinen Problemen der Gesprächsorganisation, entlang derer auch die Auswahl der tiefer analysierten Phänomenbereiche für dyadische VUI-Dialoge erfolgt.⁵ Dieser

4 Für eine Einordnung dieses Begriffs siehe Kap. 2.2.2.

5 Siehe die Einführung zum ersten Analyseteil in Kap. 6.1.

Praktikenbegriff wird einschließlich seiner Forschungstraditionen in Kap. 2.1.3 genauer vorgestellt.

- (2) Zweitens führen die Untersuchungen in sprachlich vermittelte *soziale Praktiken* hinein. Von besonderem Interesse sind dabei Domestizierungs- und Interface-Praktiken sowie das Gespräch als soziale Praktik,⁶ aber auch die sozialen Praktiken, die mit Smart Speakern ermöglicht und vollzogen werden (z. B. Üben, Testen, Vorführen, gemeinsam Radio Hören oder Kochen).⁷ Dabei beruhen die Analysen auf einem anders gelagerten Begriff von Praktiken: Soziale Praktiken werden als sinnhafte und als solche erkennbare Kombinationen körperlich-materieller, ko-operativer Verhaltensweisen verstanden, als „wechselseitig gefertigte gemeinsame Abläufe“ (Schüttpelz/Meyer 2017: 158). Am Vollzug von sozialen Praktiken können mehrere Menschen oder Artefakte sowie andere Einheiten beteiligt sein (vgl. Hirschauer 2016). Diese Begriffsbestimmung impliziert auch, dass einzelne Handlungen Teil von Praktiken sein können. Der linguistische Einsatz ist mithin anders ausgerichtet: Sprachliche Praktiken sind damit Bestandteil sozialer Praktiken, sie können diese (mit-)hervorbringen. Das Konstitutionsverhältnis ist im Vergleich zu sprachlichen Praktiken also genau umgekehrt: Handlungen bringen soziale Praktiken hervor. Durch die sprachlich-interaktionale Analyse kann insofern ein Zugang zu sozialen Praktiken gefunden werden.

Die Arbeit nimmt beide Traditionen auf und untersucht sowohl *sprachliche* Praktiken als auch sprachlich vermittelte *soziale* Praktiken mit Smart Speakern. Mit dem Begriff der *Praktik* ist dabei – in Abgrenzung zum Begriff der *Praxis* – in beiden Fällen auf ein übersituatives, raumzeitlich zwar verteiltes, aber gleichwohl wiederkehrendes und wiedererkennbares Element verwiesen, das Praktiken, sprachlichen wie sozialen, innewohnt und sie erst zur Praktik macht. Praxis hingegen ist der ‚Fluss der Ereignisse‘ – das situative, lokale „in einer wechselseitigen Verfertigung befindliche Geschehen“ (Schüttpelz/Meyer 2017: 158), in dem als solche erkenn- und verstehbare Praktiken zur ‚Aufführung‘ kommen können. Dieser Begriff von Praxis ist für die Analyse ebenso relevant, denn es soll auch die Frage beleuchtet werden, wer oder was an der „wechselseitigen Verfertigung“ mit welcher Rolle beteiligt ist – die Video-Interaktionsanalysen im zweiten Analy-

⁶ Siehe dazu Kap. 2.2.1.

⁷ Siehe die Analysen in Kap. 6.2.

seteil setzen genau auf dieser Ebene an. Dazu werden praxistheoretische Setzungen des Soziologen Stefan Hirschauer (2004; 2016) auf den Gegenstand der VUIs angewendet.

Der Aufbau der folgenden Unterkapitel gliedert sich insofern wie folgt: Zunächst wird der Begriff der *sozialen Praxis* (Kap. 2.1.1) nach Schüttpelz/Meyer, Goodwin und Hirschauer eingeführt. In diesem Zusammenhang wird die Praxeologie auch theoriegeschichtlich kurz verortet. Anschließend werden die Forschungstraditionen für die Termini *soziale Praktik* (Kap. 2.1.2) und *sprachliche Praktik* (Kap. 2.1.3) aufgeschlüsselt (um den Argumentationsgang zu vereinfachen, wird mit dem umfassenderen Begriff der sozialen Praktik begonnen, ehe der spezifischere Begriff der sozialen Praktik eingeführt wird). Abschließend werden diese praxeologischen Grundlagen auf den Bereich der Medientheorie angewendet, um ein praxeologisches Verständnis von Medien bzw. Interfaces darzulegen, die jeweils sich über die zugehörigen Praktiken analysieren lassen (Kap. 2.1.4).

2.1.1 Soziale Praxis und Ko-Operation

Das Feld der Praxeologie, regelmäßig auch als „Familie der Praxistheorien“ (Schäfer 2020: 13) bezeichnet, ist noch vergleichsweise jung. Sie ist vom Beginn ihrer Entwicklung an interdisziplinär (mit einer Fundierung in Soziologie bzw. Sozialtheorie und -philosophie), hat viele offene Enden und ist an ihren Rändern regelmäßig von Unschärfe mit fluiden Grenzen gekennzeichnet. Schäfer (2016: 9) versteht sie als „heterogene, aber dennoch definierbare Theoriebewegung“, ganz ähnlich Schatzki (2001: 13), der von einem „loose, but nevertheless definable movement of thought“ spricht. Diese Bewegung hat sich spätestens mit dem sog. „practical turn“ in den Sozial- und Kulturwissenschaften (Schatzki/Knorr-Cetina/Savigny 2001) als Paradigma etabliert und auch wenn Uneinigkeit darüber besteht, welche Theorien und welche Vertreter*innen der Praxistheorie zuzurechnen sind, gibt es doch erstens Denk- und Analyseperspektiven sowie zweitens theoretische Ausgangspunkte, die die verschiedenen Ausformungen der Praxistheorie einen und sie als Strömung klar erkennbar machen.

Zentraler gemeinsamer Ausgangspunkt und damit verbundenes Forschungsanliegen praxeologischer Theorie- und Begriffsarchitekturen ist der Versuch, den vermeintlichen Gegensatz von Gesellschaft und Individuum, den Dualismus von ‚Ordnen‘ und ‚Ordnung‘ und Dichotomien von Mikro- und Makroebene oder Subjekt und Objekt zu überwinden (vgl. Schäfer 2020: 18). Praxistheorie versteht sich als Antwort auf das Mikro-Makro-Problem des Strukturalismus in der Soziologie. Schatzki (2016) bringt dies mit dem Begriff der „flachen Ontologie“ auf den Punkt: In praxeologischen Anschauungsweisen gibt es nicht unterschiedliche Realitäts-

ebenen für die Mikroebene (wie Handlungen, Interaktionen und einzelne Individuen) auf der einen Seite und die Makroebene (wie Institutionen, Strukturen, Systeme u. a.) auf der anderen Seite. Es bestehen auch keine bestimmenden oder begründenden Effekte von der vermeintlich höheren Makro- auf die niedrigere Mikro-Ebene. Stattdessen existiert bloß eine einzige ontologische Ebene, in der sich die beiden Bereiche in einem „Konstitutionsverhältnis“ befinden, d. h. auseinander bestehen (vgl. Schatzki 2016: 31): Strukturen oder Systeme konstituieren sich nur *durch* Individuen, deren Handlungen und Interaktionen, und diese bilden immer auch die Strukturen oder Systeme. Praktiken sind insofern der konstitutive Ort des Sozialen, alle sozialen Phänomene sind „Konstellationen oder Aspekte von Praktiken oder in Praktiken begründet“ (Schatzki 2016: 30) und öffnen den Blick für die „beständige[n] Hervorbringungen“ (Schäfer 2020: 20) von Gesellschaft und Individuum mit einer *Prozessperspektive*.

Den praxistheoretischen Rahmen der vorliegenden Arbeit bildet eine Theorie- und Begriffsarchitektur, die auf den Setzungen von Charles Goodwin (insb. Goodwin 2018) sowie Schüttpelz/Meyer (2017) aufbaut und bereits kurz angerissen wurde: ein Verständnis von Praxis als „das in einer wechselseitigen Verfertigung befindliche Geschehen“ (Schüttpelz/Meyer 2017: 158); dies hat eine konsequente *Vollzugsperspektive* zur Folge. In etwa äquivalent dazu verhält sich der von Goodwin (2018) geprägte Begriff der „Co-Operative Action“ – beide Ansätze betonen einen fortlaufenden, hochgradig situativen und lokal entstehenden ‚Strom‘ sozialer Ereignisse. Der Begriff der Kooperation stellt dabei nicht die konsensuale Zusammenarbeit von zwei oder mehr Individuen dar, die – spieltheoretisch bzw. aus Sicht der biologischen Anthropologie gedacht – für einen bestimmten Preis ihren gemeinsamen und so auch ihren individuellen Nutzen maximieren, sondern vielmehr wird durch Ko-Operation – jetzt mit Bindestrich – wechselseitig ein Ablauf hervorgebracht (vgl. Goodwin 2018: 5–7). Diese Wechselseitigkeit entspricht im Englischen dem Konzept der „mutuality“, d. h. eine unausgesprochene, vorausgesetzte Wechselseitigkeit, auf die wir permanent angewiesen sind, um uns überhaupt mit anderen Menschen koordinieren und den ‚Strom‘ der Abläufe erzeugen zu können: „Without our permanent mutual assistance, we would not be able to speak and act at all“ (Schüttpelz/Meyer 2018: 175). Die Ziele der Beteiligten sind also nicht notwendigerweise die gleichen. Es handelt sich nicht um eine Wechselseitigkeit im Sinne einer „common“ oder – noch stärker – „joint“ action, sondern gerade das Verweigern einer „joint action“ kann Gegenstand einer Ko-Operation sein, die ebenfalls eine „mutuality“ voraussetzt. Handeln mit gemeinsamen Zielen (Interagieren) ist also eine mögliche Ausprägung, aber weder eine notwendige noch hinreichende Bedingung für Ko-Operation. Betrachten wir die Einbindung von VUIs in die soziale Praxis, ist dies eine wichtige Setzung, weil dem technischen Artefakt keine *intentionale* Wechselseitigkeit oder Zielorientie-

rung unterstellt werden muss, um Bestandteil des wechselseitig gefertigten Ablaufs zu sein.⁸ Schüttpelz/Meyer (2017) rücken von den Arbeiten Goodwins ausgehend den Praxisbegriff ins Zentrum ihrer Überlegungen und bauen darauf eine in sich konsistente Begriffsbildung auf: Praxis ist „das in einer wechselseitigen Verfertigung befindliche Geschehen“ (Schüttpelz/Meyer 2017: 158). Darauf – also auf diesen fortlaufenden ‚Fluss‘ oder ‚Strom‘ sozialer Ereignisse – wird im Folgenden als (*soziale*) *Praxis* referiert werden. Schüttpelz/Meyer (2017: 158–159) leiten darauf aufbauend die Begriffe *Kooperation* – „die wechselseitige Verfertigung gemeinsamer Ziele, Mittel oder Abläufe“ (Schüttpelz/Meyer 2017: 158) – *Interaktion*, *Praktiken*, *Handlungen*, *Routinen* und *Techniken* ab. Praxis und Kooperation werden also konzeptuell allen anderen Größen vorgelagert.

Der Begriff der *Handlung* (jedenfalls in einem klassischen Verständnis) ist bei Schüttpelz und Meyer und anderen praxeologischen Ansätzen hingegen weniger zentral. Das individuelle „Tun“ einer Handlungstheorie etwa nach Max Weber (1966[1921]) steht zur Erklärung der sozialen Ordnung in dieser Auffassung weniger im Fokus. Praxis wird in der Terminologie von Schüttpelz/Meyer (2017) und anderen praxeologischen Ansätzen überhaupt erst zu Handeln, wenn es *mit einem Anderen* und in einem bestimmten materialen Kontext vollzogen wird – und ist auch dann nur der „*Ausschnitt* [Herv. i. O.] einer Praktik, der von nur einem Körper getragen wird“ (Hirschauer 2016: 51). Goodwin (2018) entfaltet seine Idee zwar terminologisch ausgehend von „actions“, doch er nimmt außerdem an, dass diese nur dadurch entstehen, dass verschiedene allgemeingültige oder neu kombinierte, semiotisch in der spezifischen Situation immer wieder neu funktionalisierte Ressourcen miteinander in Verbindung gebracht werden und so ein neues „Ganzes“ schaffen (Goodwin 2018: 221). Erst durch diese Verbindung entsteht eine Handlung. Der *Prozess* der Verbindung ist entsprechend „co-operative action“ – also die Praxis (vgl. Habscheid et al. 2021: 27).

Für eine praxeologische Perspektive genügt es also Hirschauer (2004, 2016) zufolge nicht, das auf die einzelnen Körper zugeschnittenen bzw. von ihnen ausgehende Handeln anzusehen. Man muss die Praxis, die *co-operative action*, vielmehr als Prozess mit verschiedenen Beteiligten in den Blick nehmen. Für Hirschauer bedeutet das, dass in diesen sozialen Prozess mehrere *Partizipanden*⁹ involviert sind,

⁸ Siehe dazu auch Kap. 2.2.

⁹ Hirschauer (2004: 74–75) begründet hier explizit die Verwendung des Partizip Präsens: Dies „gehört zu den sog. Infiniten Verbformen, die anders als finite Formen in Person und Numerus unbestimmt sind. Infiniten Verben ist der Akteur egal: Etwas bleibt ‚rauchend‘ – ob als Mensch, als Colt, als Schlot oder als Vulkan“. Damit wird – wie im Folgenden näher erläutert – eine partizipatorische anstatt einer kontributorischen Perspektive fokussiert. Entsprechend wird diese

die ihren ‚Beitrag‘, ihren wahrnehmbaren Anteil, zum Vollzug der Praktik leisten, die Handlungen also auf mehrere Urheber „verteilt“ sind (Hirschauer 2016: 51). In Abgrenzung zu handlungstheoretischen Ansätzen, aber auch Gegenentwürfen wie der von Bruno Latour eingebrachten *non-human agency*¹⁰ schlägt Hirschauer (2004: 74) vor, eine „*partizipatorische* [Herv. i. O.] Perspektive auf jene Entitäten ein[zu] nehmen, die an Praxis teilhaben und in ihre Dynamik verwickelt sind“. Diese materiellen Entitäten können jedoch (a) einen sehr unterschiedlichen Aktivitätsstatus aufweisen, der zwischen proaktiven und inhibtiven sowie aktivem und passivem Handeln zu suchen ist (vgl. Hirschauer 2016: 49). Mit anderen Worten: Partizipanden am Vollzug der Praxis können eher nützlich oder hinderlich sowie explizit oder implizit daran beteiligt sein. Die für den Gegenstandsbereich der VUIs sehr interessante Pointe ist, dass außerdem (b) die Partizipanden nicht menschlich sein müssen und auch nicht in jedem Fall sind. Involvierte können auch nicht-menschliche Wesen (etwa Tiere), deren Körper, Dinge (z. B. eine Zahnbürste, eine Pfanne oder ein Smart Speaker), Dokumente oder deren Gestaltung, gegebene Umstände (z. B. das Wetter) oder räumliche „Settings“ sein. Damit ist klar, dass in dieser Hinsicht auch ein Smart Speaker mit einer synthetischen Stimme, die verbalsprachliche Laute erzeugt, ‚involviert‘ sein und z. B. bestimmte Äußerungen der menschlichen Partizipanden evozieren und andere supprimieren kann. Diese *partizipatorische* Betrachtungsweise soll auch bei der empirischen Untersuchung von Mensch-Maschine-Dialogen herangezogen werden: Sie ermöglicht es, maschinelle Artefakte differenziert als Partizipanden an der Praxis zu betrachten, ohne ihnen Subjektstatus, Intentionalität oder Handlungsmacht zuschreiben zu müssen.¹¹

2.1.2 Soziale Praktiken: Kleinste Einheit des Sozialen

Soziale Praktiken sind das, was der Soziologe Andreas Reckwitz (2003: 290) treffend als die „kleinste Einheit‘ des Sozialen“ beschrieben hat. Reckwitz stützt sich dabei auf den Sozialtheoretiker Theodore Schatzki, dessen Arbeiten auch im Folgenden leitend sein sollen und v. a. deshalb geeignet erscheinen, weil sie Rekurs auf den Sprachphilosophen Ludwig Wittgenstein nehmen und die hohe Relevanz

Schreibweise (gegenüber einer Schreibweise mit *-ant*, die eine *aktive* Rolle der Beteiligten schon voraussetzen würde) übernommen.

¹⁰ Zur Akteur-Netzwerk-Theorie (ANT), auf die an dieser Stelle nicht weiter eingegangen werden soll, siehe Latour (1996; 2007).

¹¹ Siehe dazu ausführlicher Kap. 2.2.4.

des Sprachlichen innerhalb des Gefüges sozialer Praktiken erkennen.¹² Dies erscheint vor dem Hintergrund gegenstandsadäquat, dass die Funktionsweise von VUIs dialogisch und auf gesprochene Sprache orientiert ist, sodass soziale Praktiken mit VUIs jedenfalls zu großen Teilen sprachlich vollzogen werden. Schatzki (1996: 89) spricht von sozialen Praktiken als „a temporally unfolding and spatially dispersed nexus of doings and sayings“. Damit sind soziale Einheiten beschrieben, die eine Verkettung von verteilten, verbal oder körperlich ausgeführten Handlungsschritten erfordern und durch diese konstituiert werden, z. B. Kochen, Wählen, Pflegen oder Korrigieren (vgl. Schatzki 1996: 89). An diesen können sprachliche Praktiken im Sinne von genutzten Ressourcenbündeln bzw. sprachlich-interaktionalen Verfahren einen (auch den wesentlichen) Anteil haben, insofern sie Handlungsschritte hervorbringen. Schatzki (1996: 89) beschreibt soziale Praktiken¹³ insofern als als Zusammenhang von Handlungen und Äußerungen („nexus of doings and sayings“): Diese sind miteinander erstens dadurch verbunden, dass die ausführenden Personen sie als eine Einheit verstehen (nicht notwendigerweise, dass sie von außen auch als solche beobachtbar sind – dies trifft nur auf einen Teil der Praktiken zu, s. u.). Zweitens sind sie durch eine normative Ebene in Form von Regeln, Vorgaben und Anweisungen miteinander verbunden und drittens durch eine kulturell geprägte Zielrichtung, die Schatzki (1996: 89) „teleoaffektive structure“ nennt und die die Ziele und Zwecke bestimmter Praktiken mit in den Aufbau dieser einfließen lässt. Die einzelnen Bestandteile, d. h. die Handlungen und Äußerungen, können dabei auch kausal miteinander verbunden sein. Diese Auffassung ist primär gemeint, wenn im Folgenden von *sozialen* Praktiken gesprochen wird.

Praktiken sind also einerseits konstitutiver Kontext, in dem ein bestimmtes Verhalten – in Form von körperlichem Verhalten, sprachlichen Äußerungen und durch Übung – zur Handlung werden kann (vgl. Schatzki 1996: 36). Die besondere Rolle des Sprachlichen kommt im von Ludwig Wittgenstein geprägten Begriff der „Sprachspiele“ zum Ausdruck (vgl. Wittgenstein 1990[1953]: 250; § 23). Wittgen-

12 Zur Relevanz von Schatzki für die Praxeologie siehe auch Schulz-Schaeffer (2010: 335).

13 Schatzki (1996: 90–98) unterscheidet auch noch zwei weitere Bestimmungen des englischen Begriffs „practice“: erstens verstanden als „Übung“ oder das „Einstudieren“, z. B. beim Klavierspielen „üben“ oder beim sportlichen Training. Dieser Auffassung geht Schatzki nicht weiter nach, auch wenn er sie selbst nicht für irrelevant hält und sie auch bei Goodwin in Form der Ausbildung durch Fähigkeiten und Kompetenzen eine gewichtige Rolle spielt (vgl. Goodwin 2018: 456). Der Idee des ‚Einstudierens‘ und ‚Erprobens‘ wird in der folgenden Arbeit zwar stellenweise ebenfalls Aufmerksamkeit geschenkt, allerdings bleibt dabei der Praktiken-Begriff außen vor. Zweitens wird „practice“ auch bei Schatzki (1996: 90) als „stream of activity“ bezeichnet und somit als Vollzugsbegriff eingeführt, was im Deutschen in etwa dem Terminus *Praxis* entspricht, siehe Kap. 2.1.1.

stein betont, dass Sprache in ihren Gebrauchskontext eingewoben ist und diesen (und reflexiv auch dessen Regeln) mit hervorbringt. Letzterer kann dabei sowohl lokal der Kontext einer ganz bestimmten Situation oder auch übergreifend ein bestimmter Lebensbereich (z. B. das Militär) oder anderer Zusammenhang sein – Praktiken können also auf ganz unterschiedlichen Analyseebenen beleuchtet werden. Andererseits entstehen also Praktiken auch erst aus den immer wieder so vollzogenen und für die Praktik relevanten Handlungsabläufen. Diese Handlungsabläufe unterliegen den für die Praktik spezifischen Regeln, doch da – mit Verweis auf den „unendlichen Regelregress“¹⁴ wie Wittgenstein (1990[1953]: 345–346; PU §§ 201–206) ihn herleitet – diese nicht durch weitere Regeln normiert werden können, ist es eben gerade deren Anwendung und ihr spezifischer Vollzug in einer gegebenen sozialen Situation, die sie konstituieren: „Regeln und Praktiken setzen also einander wechselseitig voraus“ (Habscheid 2016: 134).¹⁵ Das „Regelfolgen“ Wittgensteins macht deutlich, dass sich die konkreten Regelinstantiierungen, d. h. die *vollzogenen* Praktiken, durch die Deutungen der Regeln durch die Beteiligten, den situativen Kontext und andere Bedingungen (Wittgenstein 1990[1953]: 254; PU § 30), d. h. *in der Praxis* ergeben. Dabei sind Praktiken auch hier dadurch gekennzeichnet, dass der Sinn einer Praktik nicht identisch mit dem subjektiv gemeinten Sinn oder der Intention des Individuums sein muss und der Sinn so erst *in situ* entsteht (vgl. Lanwer/Coussios 2015: 128–129). Sprache stellt also mit ihren Regeln bzw. Sprachspielen einen Zugang zur Anatomie sozialer Praktiken dar – und insofern auch zu kulturellen und medialen Aspekten dessen (vgl. Meiler 2023), was sich im Hinblick auf die zu untersuchenden VUI-Dialoge aufgrund der bereits erwähnten Sprachlichkeit dieser besonders anbietet.

2.1.3 Sprachliche Praktiken: Ressourcen für Bausteine der Interaktion

Der hier ebenfalls verwendete Begriff der *sprachlichen Praktiken* liegt insofern, wie bereits eingangs dargestellt analytisch, auf einer anderen ‚Ebene‘. Sprachliche

¹⁴ Der infinite Regelregress (auch „Regelfolgen“), der auf Wittgenstein zurückgeht, besagt in Kürze zusammengefasst, dass Regeln zur Anwendung der Regel wiederum einer Regel bedürfen, wann diese Regel zu verwenden ist, und zu einer Regel zweiter Ordnung eine Regel dritter Ordnung hinzutritt, bis Regeln n-ter Ordnung aufgestellt sind, aber noch immer nicht die Anwendung der Regeln geklärt werden konnte (vgl. Hillebrandt 2014: 37; Habscheid 2016: 134); erst das tatsächliche „Befolgen“ der Regel kann dies auflösen.

¹⁵ Angewendet auf ein sprachliches Regelsystem bedeutet dies im Sinne Wittgensteins, dass „die Regeln der Sprache [...] keine natürlich gegebenen Ausgangspunkte der Praxis [sind], sie sind vielmehr deren Produkt“ (Hillebrandt 2014: 40).

Praktiken sind bei Goodwin (2018: 25) ähnlich wie etwa bei Schegloff (1997a; 2000a) als „specific grammatical practices“ bestimmt. Einbezogen werden auch Praktiken nicht-sprachlicher Art (Goodwin 2018: 46), denn es muss betont werden, dass damit keineswegs nur sprachliche Ressourcen i.e.S. erfasst werden sollen, sondern vielmehr alle in der Situation zur Verfügung stehenden sprachlichen, materialen und sozialen Mittel als Ressource zur Übermittlung von Zeichen eingesetzt werden können – man könnte hier präziser auch von „kommunikativen Praktiken“ sprechen (vgl. Habscheid 2016: 137), von denen sprachliche Praktiken ein Bestandteil sind.¹⁶ Sie ergeben sich aus der Praxis und sind deren Produkt. Goodwin versteht dabei unter Praktiken nicht einfach nur die unfunktionalisierten Ressourcen, sondern erst durch deren wiederholte, konglomerierte Nutzbarmachung werden Praktiken konstituiert. Goodwin geht dabei von einer temporalen Akkumulation einzelner Operationen aus. Genau diese Auffassung von Praktiken herrscht auch in der konversationsanalytischen Betrachtungsweise vor: Selting (2016: 28–29) spricht von „routinemäßiger Verwendung“ „verbaler, vokaler und, so weit relevant, visueller Ressourcen“. Genau dies entspricht dem Praktikenverständnis konversationsanalytischer Arbeiten (siehe etwa Schegloff 1997a) – Praktiken sind die Ressourcen-Bausteine, aus denen sich soziale Interaktion speist und von Interagierenden zusammensetzen lässt, um kommunikative Aufgaben zu bewältigen (siehe auch Schegloff 2006). Mit welchen Bausteinen die Aufgabe eines VUI-Dialogs bewältigt werden kann ist also die zentrale Frage, für die dieses Verständnis von sprachlichen Praktiken nutzbar gemacht wird.

Praktiken auf dieser Ebene dienen also dem Aufbau „soziale[r] Handlungen und Aktivitäten“ (Selting 2016: 28) in der Interaktion¹⁷ und es lässt sich noch radikaler formulieren, dass in der Interaktion alle sprachlichen und anderweitig verkörperten Ausdrücke, auch im Rückgriff auf körperliche oder materiale Ressourcen, dazu dienen, die Einheit *für das Gegenüber* verständlich zu machen – also

¹⁶ Aufgrund der sprachzentrierten Analyse bleibe ich gleichwohl bei dem Terminus „sprachliche Praktiken“, beziehe dabei aber auch multimodale Verfahren in die Analyse ein (siehe Kap. 4.4).

¹⁷ Wie hier deutlich wird, ist in dieser Auffassung, wie sie v. a. bei Selting, Schegloff und in anderen konversationsanalytischen Arbeiten (teils implizit) entwickelt wird, nicht die ‚Praxis‘ oder die ‚Ko-Operation‘ die entscheidende Bezugsgröße, sondern die Interaktion und die Verfahren, mit denen sie organisiert werden kann (vgl. Schegloff 2006: 70). Über diese Abweichung in der Fokussierung der einzelnen Domänen kann im Folgenden aber hinweggesehen werden, weil das Konzept von Interaktion, auf das Schegloff sich bezieht, dem Konzept der Ko-Operation bzw. der Praxis nicht unähnlich ist – auch er geht nicht von Subjekten aus, sondern betont die Wechselseitigkeit und den „common ground“, von dem aus eine Interaktion überhaupt erst möglich ist (vgl. Schegloff 2006: 89).

das Konzept des Recipient-Design (vgl. Sacks/Schegloff/Jefferson 1974: 727) Anwendung findet. Diese Ressourcen werden jedoch nur dann eine Praktik, wenn sie für einen routinisierten und wiederholt auftretenden Zweck eingesetzt werden, so z. B. der Einsatz bestimmter Partikeln zur Initiierung von Reparaturen (Schegloff 1997a). Selting (2016: 29) formuliert dieses Verhältnis zwischen Ressourcen, Praktik und Handlung so: „Durch die Verwendung einer für eine bestimmte Handlung designierten Praktik wird diese bestimmte Handlung ausgeführt, deren Erkennen ermöglicht oder deren Interpretation nahegelegt“. Dafür kommen bestimmte „Ressourcen oder Ressourcenbündel, d. h. sprachliche und nicht-sprachliche Formate“ zum Einsatz. Werden diese Ressourcen nun aber nicht für den routinisierten Zweck eingesetzt, sondern anderweitig funktionalisiert, sind sie eben nicht mehr eine „für eine bestimmte Handlung designierte[.] Praktik“. ¹⁸ Dabei ist die Bedeutung einer Praktik dennoch nicht *a priori* gesetzt: Praktiken sind „kontextsensitiv“ und können „flexibel an die je spezifischen Gegebenheiten angepasst werden“ (Selting 2016: 29). Diese Flexibilität ist im Hinblick auf VUI-Dialoge zu prüfen, die mit vorgegebenen Dialogstrukturen arbeiten und insofern bestimmte Eingaben an bestimmten konversationellen Stellen erwarten – die Flexibilität könnte an dieser Stelle v. a. seitens der Nutzer*innen bestehen, sich an den präskribierten Verlauf anzupassen (vgl. Habscheid 2022: 191).

Selting (2013) unterscheidet zwischen „verbalen“, „vokalen“ und „visuellen“ Ressourcen, die beim Vollzug von Praktiken zum Einsatz kommen. Mit Blick auf die „interaktionale Infrastruktur“ (Schegloff 2006) ist darüber hinaus nach den interaktionalen Merkmalen zu fragen: Konversationelle Strategien für die sequenzielle Organisation, die Rederechtsverteilung oder für Reparaturen sowie der Aufbau, das Zustandekommen und das Beenden der Interaktion sollen in den Blick genommen werden (siehe auch Schegloff/Sacks 1973). Diese sprachlich-interaktionalen Merkmale sind nicht zufällig genau die Ressourcen bzw. Eigenschaften der gesprochenen Sprache, an denen auch die Forschung zu Spracherkennung und VUIs ansetzt, um ihre angebotenen Technologien zu verbessern (vgl. Kap. 3). Es ist die Kombination sehr kleiner Einheiten von weniger als einer Sekunde, die mit Blick auf Tonhöhe, Sprechtempo, Sprechpausen und Zögerungen das synthetische Erzeugen

¹⁸ Wie Selting (2016: 50) ausführt, ist dies analog zum Verhältnis von „Kontextualisierungshinweis“ und „Kontextualisierungsverfahren“ in der Kontextualisierungstheorie nach Gumperz (1982) zu betrachten: Während die *Hinweise* („Contextualization Clues“ bei Gumperz) lediglich die Ressourcen sind, sind es die *Verfahren*, die einen bestimmten Kontext konstituieren, indem sie Bezüge zum Hintergrundwissen der Gesprächsbeteiligten herstellen (vgl. Auer 1986: 24). Auffällig ist, dass sich zwar Gumperz (1982) noch überwiegend auf Sprachliches bezog, während in der deutschsprachigen Weiterentwicklung von Auer (1986) bereits Ressourcen wie Blick, Gestik, Pausen u. a. einbezogen werden; siehe auch Kap. 4.4.

einer dem Menschlichen ähnlichen Mündlichkeit erst ermöglicht. Will man also, mit einer Engführung des Erkenntnisinteresses auf das Sprachliche, herausfinden, welche sprachlichen Praktiken im Kontext des Gebrauchs von VUIs zum Einsatz kommen, ist eine sequenzielle Analyse¹⁹ der einzelnen Handlungsschritte im Sinne der Konversationsanalyse und der Interaktionalen Linguistik durchaus sehr hilfreich. Selting/Couper-Kuhlen (2001a: 280) nennen die „Sprachanalyse und -synthese in elektronischen Dialogsystemen“ explizit als möglichen Ertrag der Interaktionalen Linguistik.

2.1.4 Medien- und Interface-Praktiken

Ein Ansatz zur Untersuchung von Medien, der von den Praktiken ausgeht, mit denen sie verfertigt werden – mithin *Medienpraktiken* – hat in den letzten Jahren eine erhöhte Aufmerksamkeit erfahren (siehe etwa Schüttpelz/Gießmann 2015; Dang-Anh et al. 2017; Schüttpelz/Meyer 2017; Meiler 2019; Schüttpelz et al. 2021). Diese Auffassung ist nicht nur eine Folge davon, dass der sogenannte „practice turn“ auch die Medienwissenschaften erreicht hat (vgl. Schüttpelz 2021), sondern auch eine Reaktion darauf, dass Medien sich spätestens seit der bereits von Kittler (1996: 131) beobachteten Konvergenz von Medien in einem „Supermedium Computer“ sich nicht ohne ihre sozialen Voraussetzungen beschreiben lassen. Entsprechend müssten, so die Argumentation, die Praktiken, mit denen Medien verfertigt werden, am Ausgangspunkt der Analyse stehen (vgl. Schüttpelz/Gießmann 2015: 8–9). Medien bringen in diesem Sinne also Medienpraktiken hervor und sind andersherum, in Anlehnung an das ethnomethodologische Programm Harold Garfinkels und die dort entworfene Idee von Reflexivität (vgl. Garfinkel 1967: 7–9; siehe auch Bergmann/Meyer 2021b: 39–42) zugleich das Ergebnis von Medienpraktiken (vgl. Schüttpelz/Gießmann 2015: 35–36): Medien konstituieren

¹⁹ Dabei soll die von Goodwin herausgestellte und insbesondere von Schüttpelz/Meyer (2018: 174, 183) pointierte ‚Abkehr‘ von der (vermeintlichen) „Reduktion“ der sozialen Ereignisse bzw. der Interaktion auf „chained (strategic) moves of individuals“ bzw. „aggregations of individual acts that are sequentially performed“ nicht übersehen werden: Praktiken zeigen sich in dieser Auffassung „simultaneously in the midst of single actions, rather than sequentially“ (Goodwin 2018: 13). Eine zu starke Engführung auf die sequenzielle Organisation verstellt Goodwin (2018: 41) zufolge den Blick für die relevanten Bestandteile des Vollzugs der Praktiken. Doch trotz der ganzheitlichen Betrachtungsweise sind es für die Erkenntnisinteressen an der Schnittstelle von Medienlinguistik und Gesprächsforschung gerade die sequenziellen Vollzüge, in denen die sprachlichen Praktiken zum Einsatz kommen, analytisch beobachtbar werden und auch einen analytischen Zugriff auf soziale Praktiken ermöglichen. Weder Goodwin noch Schüttpelz/Meyer negieren diese Inkrementalität von sprachlicher Entfaltung.

sich in einem reflexiven Wechselspiel immer wieder neu als Kooperationsbedingungen, hervorgebracht als das, was kooperativ durch Medienpraktiken als Kooperationsbedingungen verfertigt wird. Diese Kooperationsbedingungen müssen freilich nicht notwendigerweise digitale sein, auch wenn durch immer weiter in alle Lebensbereiche eingewobene Digitalität und Vernetzung durchaus spezifische Bedingungen entstehen (vgl. Kaerlein 2020: 46). Besonders in Bezug auf digitale Medien wird also ein Spannungsverhältnis sichtbar zwischen sich konventionalisierenden Medienpraktiken und ihrer fortlaufenden Aktualisierung durch den Vollzug.

Smart Speaker können in verschiedener Hinsicht als Kooperationsbedingungen verstanden werden. Für die Anwender*innen sind sie gleich in mehrfacher Hinsicht Bedingung für die Kooperation: einerseits für die Kooperation mit den Systemanbietern, die Sprachsteuerung durch das Übersetzen von Schallwellen in diskrete Zeichen (und zurück) anbieten und dabei zum Teil opake Datenverarbeitungs- und -verwertungsprozesse vollziehen,²⁰ andererseits mit Drittanbietern, die u. a. gestreamte Medieninhalte (Musik, Podcasts, Nachrichten oder auch interaktive Formate wie Spiele) bereitstellen. Nutzer*innen kooperieren also (mehr oder weniger wissentlich) mit den Anbietern der Systeme sowie mit Plattformen zur Bereitstellung von Inhalten.²¹ Untersuchungen zeigen, dass die Anwender*innen die hintergründig ablaufenden Prozesse dieser Kooperation zu großen Teilen nicht kennen können oder wollen – diese Kooperationsform ist damit ein geradezu typischer Fall von „Kooperation ohne Konsens“ nach Susan Leigh Star (1993) (vgl. Waldecker/Hector/Hoffmann 2024). Stationäre Sprachassistentensysteme sind zudem – für die Nutzer*innen nur vermittelt einseh- und konfigurierbar – Kooperationsbedingung für die Kooperation zwischen verschiedenen Plattformen, u. a. zwischen den Diensteanbietern der Sprachsteuerung und denen zur Bereitstellung der Inhalte, ebenso wie Werbetreibenden.²² Sie sind somit Bestandteil ganz unterschiedlicher Medienpraktiken, durch die sie zugleich immer wieder neu hervorgebracht werden.

Medienpraktiken konstituieren und vollziehen sich allerdings nicht im luftleeren Raum. Sie werden durch *Interfaces* konturiert – ein Begriff, der hier als wegweisend für die theoretische Fundierung der Analysen eingeführt werden soll und zunehmend das Interesse der medienwissenschaftlichen Forschung weckt; als Belege dafür sei auf die Verweise bei Distelmeyer (2020: 63) sowie auf

²⁰ Siehe dazu Kap. 3.2.2.

²¹ Siehe Kap. 3.2.3.

²² Für eine genauere Betrachtung siehe Kap. 3.2.2 und 3.2.3 sowie Hector (2025).

das derzeit entstehende Handbuch zur Einführung in die Interface Studies (siehe Ernst et al. i. V.) verwiesen. Interfaces waren in medienwissenschaftlichen Untersuchungen die längste Zeit ein eher diffuser und unscharfer Begriff (vgl. Wirth 2016: 17), doch durch theoretische Arbeit am Interface-Begriff²³ scheint sich eine Perspektive zu etablieren, die Interfaces nicht als neutrale Objekte sieht. Die Interface-Forschung hat dazu verschiedene Thesen formuliert, die – unterschiedlich nuanciert – auf einen „prozessualen“ (Kaerlein 2020: 50), man kann, je nach Fokus, auch sagen: relationalen oder praxeologischen Begriff von Interfaces abstellen.²⁴ Interfaces entstehen demzufolge erst dort, wo zwei unterschiedliche Einheiten aufeinandertreffen. Sie sind mit Kaerlein (2020: 50) „dasjenige, was erst in der Begegnung heterogener Entitäten Kontur gewinnt, aber die Beziehung zwischen den beteiligten Größen konstitutiv reguliert“. Diese pointierte Zusammenfassung der gegenwärtig vorherrschenden Position zu Interfaces baut auf Theoriebildungen der Medienwissenschaften bzw. der Software Studies auf, so postuliert etwa Galloway (2012: 33) „[A]n interface is not a thing, an interface is always an effect“, Hookway (2014: ix) beschreibt Interfaces als „a relation with technology rather than as a technology in itself“. Distelmeyer (2020: 63) weist darauf hin, dass Interfaces auch in der Begriffsgeschichte mit einer Fokussierung auf „Prozesse und Effekte konkreter Konstellationen“ verstanden wurden; ein solches Verständnis ist auch im Hinblick auf den Wortursprung konsequent, der im 19. Jahrhundert bei den chemischen Untersuchungen James Thomsons zu finden ist. Bei Thomson entsteht – vereinfacht gesagt – ein Interface beim *Kontakt* zweier Flüssigkeiten mit einer geteilten Grenzfläche (vgl. Hookway 2014: 59–60): Das Interface ist also nicht einfach schon gegeben, sondern kommt erst als Grenzfläche zweier ungleicher Ein-

²³ Für eine Übersicht über die Entwicklungslinien des Interface-Begriffs siehe Wirth (2016: 18–21) und ausführlich Wirth (2019: 68–81; 2025); der Begriff der *Interaktion* bzw. der *Interaktivität* hat sich demzufolge zwar ebenfalls etablieren können – bis hin zur Teildisziplin der Human-Computer-Interaction –, bleibt jedoch im Hinblick auf sein heuristisches Potenzial unbestimmt (siehe auch Manovich 2001) und bringt zudem weitere begriffliche Schwierigkeiten mit sich; er soll dementsprechend nicht für den Austausch zwischen Mensch und Computer verwendet werden (siehe auch Kap. 2.2).

²⁴ Cramer/Fuller (2008: 149) unterscheiden in einer Interface-Typologisierung fünf verschiedene Arten von Interfaces: User-Hardware, Hardware-Hardware, Hardware-Software, Software-Software und User-Software. Auch wenn im Hinblick auf den zu untersuchenden Gegenstand der stationären Sprachassistenzsysteme eine eingehende Betrachtung aller diese Typen lohnen würde, konzentrieren sich die Ausführungen im Folgenden ausschließlich auf User-Software-Relationen (trotz der Kritik von Cramer/Fuller (2008: 149), dass es gerade diese Beziehung ist, die in medienwissenschaftlichen Untersuchungen fälschlicherweise häufig als die einzige Form von Interfaces betrachtet wird).

heiten zustande, wenn diese in den Austausch treten. Konsequenterweise gehen neuere Ansätze (etwa Lipp/Dickel 2022) von einem „Interfacing“ aus, d. h. von den „manifold practices and activities that are needed to establish, maintain, and repair the human/machine interface“ (Lipp/Dickel 2022: 429) – und beziehen somit neben den Nutzer*innen auch Perspektiven von Ingenieur*innen oder Entwickler*innen ein. Interfaces sind entsprechend „the contingent, temporarily stabilized *products* [Herv. i. O.] of ongoing practices of interfacing“ (Lipp/Dickel 2022: 432). Diese Praktiken zeichnen sich, so Lipp/Dickel (2022: 439), dadurch aus, dass sie *gleichzeitig* menschliche und nicht-menschliche Einheiten sowohl verbinden als auch trennen.²⁵ Sie bestimmen „contact points and corridors that allow to ‚bridge‘ those differences“ (Lipp/Dickel 2022: 429).

Damit ist auf die regulierenden, steuernden und leitenden Eigenschaften von Interfaces verwiesen, die bei aller Situationsbezogenheit der Analyse nicht aus dem Blick geraten dürfen. Interfaces sind entscheidend für die Beziehungen zwischen Menschen und Computern (vgl. Ernst/Bächle 2020: 418): durch Interfaces „stehen nur bestimmte Handlungs- und Kommunikationsweisen zur Auswahl, die dadurch die Interaktion von menschlichen Akteuren mit Computersystemen nachhaltig prägen“ (Ernst/Bächle 2020: 418). Dadurch werden bestimmte Praktiken inhihiert, zu anderen wird geradezu eingeladen, was auch konkrete Auswirkungen auf den Sprachgebrauch hat. Interfaces tragen also aktiv zur Verortung bestimmter Medienpraktiken auf dem Proaktiv-Inhibitiv-Spektrum nach Hirschauer (2016: 49) bei: „interfaces both engender and condition human/machine relations“ (Lipp/Dickel 2022: 429; vgl. auch Kaerlein 2020: 51). Damit wird auch noch einmal deutlich, dass Interfaces keine Nullstelle sind, sondern sich gerade dadurch auszeichnen, dass sie (Macht)asymmetrien zwischen zwei ungleichen Einheiten miteinander vermitteln (vgl. Cramer/Fuller 2008: 150). Computer-Interfaces wirken einerseits ermächtigend auf Nutzer*innen, weil diese in die Lage versetzt werden, etwas zu bestimmen – zugleich wirken sie andererseits ebenfalls auf die Nutzer*innen sowie die ausgeführten Praktiken regulierend (vgl. Distelmeyer 2018: 29). Von Interfaces geht also – wie Distelmeyer (2020) detailliert am Begriff des „Leitens“ illustriert – ein Machtpotenzial aus; konkret wirkt dies auf die Nutzer*innen z. B. in vorstrukturierten Oberflächen, die bestimmte Praktiken wahrscheinlicher machen als andere – z. B., im Zusammenhang mit grafischen Interfaces, die Zustimmung zu den Datenschutz-Bestimmungen wahrscheinlicher als deren Ablehnung oder das Verweilen auf einer Internetseite wahrscheinlicher als deren schnelles Verlassen (vgl. Kaerlein 2020: 52).

25 Siehe auch Kap. 3.1.

Eine Konzentration auf die Betrachtung von Interfaces erlaubt, so argumentiert Kaerlein (2020: 49–52), eine Vermittlung des Forschungsinteresses an Medienpraktiken mit dem an Infrastrukturen. Während der konstitutive Zusammenhang zwischen beiden Größen, die sich wechselseitig hervorbringen, in den letzten Jahren immer wieder betont und deutlich herausgearbeitet werden konnte (vgl. insbesondere Gießmann 2018; aber auch Couldry/Hepp 2021; Burkhardt et al. 2022), entstand in der Forschung zu diesen beiden Themenfeldern zugleich eine Lage, die Kaerlein (2020: 47–48) als „Diskrepanzverhältnis“ zwischen den situierten Praktiken der Anwender*innen auf der einen und den raumzeitlich davon distanzierten Praktiken der Datenaus- und verwertung auf der anderen beschreibt.²⁶ Kaerlein (2020: 48) wirbt für ein „post-situationales“, wenngleich nicht weniger praxeologisches Medienverständnis, das nicht nur das Geschehen *in* einer Situation, sondern auch vor, nach und zwischen Situationen einbeziehen kann. Dies sei, so Kaerlein weiter, über die Beforschung von Interfaces zu leisten:

Interfaces strukturieren oder formatieren einerseits Praktiken durch ihre eingeschriebenen Affordanzen (a), und andererseits haben Praktiken nur dann eine infrastrukturierende Wirkung, d. h. eine kontinuierliche Einschreibung zur Folge, wenn sie von Interfaces geleitet und vermittelt werden (b). (Kaerlein 2020: 51)

Interfaces bzw. Interface-Praktiken zu untersuchen hilft dabei, dieses Wechselverhältnis nicht nur zu benennen, sondern es beschreiben und kritisieren zu können – denn entlang der Interfaces lassen sich Praktiken Schritt für Schritt als Infrastrukturen erkennen (vgl. Kaerlein 2020: 54). Sprachliche Praktiken mit VUIs im Kontext ihrer Domestizierung zu beschreiben (und ihre Verwebung mit der sozialen Praxis der Nutzer*innen), ist insofern auch ein Forschungsvorhaben, das Kaerleins (2020: 54) Aufruf „Follow the interfaces“²⁷ ernst nimmt und daran aus Sicht der Nutzer*innen ansetzt (siehe auch Hector 2025).

Wie Christoph Ernst (2017) herausstellt und im Rahmen einer Analyse exemplifiziert, entsteht im Austausch mit Interfaces *implizites Wissen* über diese Interfaces, das sich im Alltag der Nutzer*innen verankert.²⁸ Mit diesem impliziten

²⁶ Zu Datenverwertung bei der Nutzung von Smart Speakern siehe Kap. 3.2.6.

²⁷ Dieser Ausdruck ist eine Referenz auf die ANT-Forschung, die als ‚Motto‘ ein „Follow the actors“ ausgerufen hatte (vgl. Latour 1987).

²⁸ Der Begriff des Impliziten Wissens (*tacit knowledge*) hat seinen Ursprung bei Polanyi (1985 [1966]) und wurde bereits dort auf sprachliche Praktiken zwischen explizierbarem Regelwissen und implizitem Regelfolgen (im Sinne Wittgensteins, s. o.) angewendet (siehe auch Sharrock 2012; Loenhoff 2015; 2019).

Wissen schreiben sich Computer, ihre Operationen und auch die großflächige Datenverwertung und die computationale Vernetzung in die sprachlichen und sozialen Praktiken der Nutzer*innen ein (vgl. Ernst 2017: 111). User-Interfaces können wesentlicher Alltagsbestandteil der Anwender*innen sein bzw. werden und dabei implizites Wissen konstituieren, argumentiert auch Wirth (2016: 32): „A user interface can be a space of inhabitation, a comfort zone, which we tacitly know to use and which offers ways of operating and ‚handling‘, that appear to be ‚ready-to-hand‘ to the user“ – damit verbunden ist der Vorschlag, den alltäglichen Gebrauch in die Konzeptionierung des Interface-Begriffs als Medienpraktiken einzubeziehen (vgl. Wirth 2016: 33). Die sprachlichen Praktiken zur Bedienung von VUIs, ihre Ausbildung und Verfestigung, sind Ausdruck dieser sozialen Praktiken und machen „Übergänge zwischen situierter Kognition und dem impliziten Regelwissen von Praktiken“ (Ernst 2017: 101) sichtbar, die entlang der Interfaces vollzogen werden. Wenn sich Nutzer*innen dabei an etablierten, z. B. gesprächsorganisatorischen, Praktiken orientieren, die ebenfalls als implizites Wissen vorliegen, entstehen im Interface-Gebrauch epistemologische Kontinuitäten, deren Erscheinungsformen einer Untersuchung zugänglich sind (vgl. Ernst 2017: 109).

Mit anderen Worten: Einerseits ist zu erwarten, dass Nutzer*innen im ‚Interfacing‘ Bezug auf als implizites Wissen vorliegende Prinzipien der Gesprächsorganisation nehmen und dies durch sprachliche Praktiken zum Ausdruck bringen. Andererseits haben Nutzer*innen auch Vorstellungen von den Funktionsweisen der Systeme („conceptual model“),²⁹ die sie beim Eintritt in den Dialog mit den Interfaces berücksichtigen – dies zeigen Merkle/Hector (2025) am Beispiel wiederholter Reparaturversuche einer VUI-Nutzerin. Zwar besteht mit der gewählten Ausrichtung dieser Arbeit kein direkter Zugriff auf die „situierter Kognition“ der Nutzer*innen und ihr implizites Wissen. Aber: Es lässt sich über eine detaillierte Untersuchung sprachlicher Praktiken als Medienpraktiken zur Bedienung von VUIs verfolgen, wie sich VUIs in die sprachlichen und sozialen Praktiken zu injizieren suchen. In diesem Ausloten von Möglichkeiten und Grenzen symbolischer Operationen mit den Interfaces entstehen in einem Prozess der Domestizierung Medien als Gewohnheiten, Erwartungen, Vorstellungen, Möglichkeitsbedingungen für zukünftigen Austausch, von denen die sprachlichen Praktiken zeugen.

Die primär in den Medienwissenschaften beheimatete Interface-Forschung war und ist größtenteils auf die Untersuchung visueller Effekte und ihrer Ge-

²⁹ Mit dem Begriff *conceptual model* werden zusammenhängende Annahmen bzw. Vorstellungen über Möglichkeiten des Systems und die daraus folgenden möglichen Umgangsweisen mit einem Interface beschrieben, die im Design entlang von Möglichkeiten, Limitierungen und Verknüpfungen konzipiert und von den Nutzer*innen im Umgang mit den Interfaces (implizit mit-) entwickelt werden (vgl. Norman 1988: 12; siehe auch Ernst 2017: 100).

schichte konzentriert, teilweise unter dem Einbezug von Haptik (vgl. Borbach 2019). Die Betrachtung auch akustischer Interfaces ist noch neu und bislang sind nur wenige Arbeiten zu ihrer Untersuchung erschienen (vgl. Schoon/Volmar 2012; Borbach 2018; 2019; Borbach et al. 2025). Dabei muss betont werden, dass die mündliche Steuerung maschineller Anwendungen nicht *prinzipiell* neu ist, so haben z. B. Tonwahl- sowie Sprachdialogsysteme (etwa für vorstrukturierende Abfragen im Callcenter) eine in die Mitte des 20. Jahrhunderts zurückreichende Kulturgeschichte, die als Wegbereiter für die natürlichsprachige Bedienung von Interfaces verstanden werden kann (vgl. Volmar 2019), wobei sich eine Entwicklung von stark prästrukturierten Tasten- und Stimmein- und ausgaben hin zu zunehmend offenen, gesprächsähnlichen und mit Natural Language Processing verarbeiteten Dialogsystemen nachzeichnen lässt (vgl. Volmar 2019: 71–72).

Die Geschichte visueller Interfaces weist Parallelen auf: So setzen sich für visuelle Interfaces in den frühen 80er-Jahren „direct manipulation“-Interfaces (Shneiderman 1983) durch. Sie ermöglichen die Bedienung von Computern über visuelle Repräsentationen statt über symbolische Eingaben (z. B. das Icon eines Druckers für die Druckfunktion statt der zuvor notwendigen Tastatur-Eingabe ‚Strg + P‘). Shneiderman (1983: 66) erklärt den Durchbruch damit, dass mit den visuellen Repräsentationen der computationalen Objekte semantisches Wissen verbunden ist, das die Nutzer*innen sich leichter aneignen und anwenden können, als das zuvor notwendige syntaktische Wissen über die Gestaltung der Eingaben: „Dealing with the re-presentations of objects may be more ‚natural‘ and closer to innate human capabilities“, so Shneidermans (1983: 66) These.

In der Historie akustischer Interfaces vollzieht sich mit dem Entstehen von Voice Assistants eine ähnliche Entwicklung, die weg von prästrukturierten, symbolischen Eingaben und hin zu ‚natürlichsprachigen‘, gesprächsähnlichen Ein- und Ausgaben für die durchzuführenden Operationen führt. Diese Entwicklung steht im Einklang mit der zunehmenden Tendenz in der HCI und im Interface-Design zu Natural User Interfaces (NUIs), die zunehmend den gesamten Körper einbezieht – Stefan Rieger (2019: 190) spricht von einer „Naturalisierung“ sowie von einer „Rückkehr des Sensorischen [...]“, die auf nicht weniger als auf die Auflösung vormals getrennter Bereiche abzielt“. In der HCI-Forschung wird dieser Wiedereinbezug menschlicher Sinne und die Orientierung an menschlichen Kapazitäten und Sinneskonfigurationen als ‚natürlich‘ konzeptionalisiert, zugleich steht dieser Begriff in der Kritik (vgl. Norman 2010). Indes ist noch einmal herauszustellen, dass Nutzer*innen sich in der Bedienung akustischer Interfaces nicht nur an Ähnlichkeiten des Dialogs zu zwischenmenschlichen Interaktionen orientieren, sondern auch an den (unterstellten) technologischen Möglichkeiten und Grenzen der Interfaces (vgl. Ernst 2017) und ihren Sprachgebrauch daran anpassen (vgl. Habscheid 2022: 191). Insofern verfertigen die Nutzer*innen zu Teilen

auch symbolische Eingaben, die von der Gesprächsähnlichkeit wenig übrig lassen (vgl. Merkle/Hector 2025) – eine Folge von Recipient-Design im Interfacing.

Ein letzter Aspekt: Smart Speaker sind als stationäre Lautsprecher konfiguriert, sodass der Zugriff nicht nur in dyadischen Dialogsituationen möglich ist, sondern die Interface-Operationen prinzipiell auch anderen Anwesenden in Mehrparteien-Konstellationen zugänglich sind. Das unterscheidet Smart Speaker etwa auch von mit Sprachassistentz-Software ausgestatteten Smartphones, deren Sehfläche Nutzer*innen ‚standardmäßig‘ als privat behandeln (vgl. Oloff 2019). Damit erzeugen Smart Speaker durch ihre Interfaces eine Spannung zwischen Gemeinschafts- und Individualmedium (für diese Unterscheidung siehe Röser/Peil 2014), auch vor dem Hintergrund personalisierter Benutzer*innen-Konten und der möglichen Privatheit von Kommunikationen. VUI-Dialoge sind damit hörbar für andere Anwesende – und somit auch für die Beteiligung oder Kommentierung dieser. In der akustischen Wahrnehmung der Anwesenden ‚konkurrieren‘ VUI-Dialoge mit laufenden Gesprächen und treten mit diesen unweigerlich in eine Beziehung, was sich auch aufgrund ihrer oberflächlichen (mehr oder minder großen) Ähnlichkeit mit Gesprächen ergibt. Insofern können neben dyadischen Dialogen auch die interaktiven Dynamiken in Gesprächen durch die Interface-Operationen beeinflusst und in die Medienpraktiken einbezogen werden – in den Analysen werden entsprechend beide Konstellationen beleuchtet.

2.1.5 Zwischenfazit Praxeologie und Medientheorie

Die hier aus Gründen der terminologischen Klarheit zur Begriffsbestimmung aufgefächerten Begriffe sind empirisch eng miteinander verwoben, wie auch in den Analysen deutlich wird. Wie nicht nur die Arbeiten Charles Goodwins oder die Studien von Michael Tomasello (2005; 2009) zeigen, besteht eine enge Verbindung zwischen der menschlichen Kooperativität und dem Ursprung und der Entwicklung von Sprache(n). Sprache wiederum ist ein Mittel, das im Prozess der praktischen Verfertigung sozialer Ordnung eingesetzt wird. Die gebrauchsbasierte Linguistik kann also von einer sozialtheoretischen Unterfütterung profitieren, weil sprachliche Formen im Kontext ihrer zentralen Rolle in der reflexiven Hervorbringung sozialer Gegebenheiten verständlich werden. Andersherum haben sprachliche Formen wesentlichen Anteil an sozialen Praktiken und ihrer Entstehung, sodass praxeologische Forschungsarbeiten von linguistischen Betrachtungen Erkenntnisse über die Anatomie sozialer Praktiken erwarten dürfen. Dies gilt ganz besonders für die Untersuchung von Medientechnologien, die mit gesprochener Sprache arbeiten und insofern gleichermaßen sowohl sprachliche als auch soziale Praktiken potenziell rekonfigurieren oder hervorbringen.

Verdeutlicht werden sollten zwei wesentliche Unterscheidungen:

- (1) Während die *Praxis*, „das in einer wechselseitigen Verfertigung befindliche Geschehen“ (Schüttpelz/Meyer 2017: 158) selbst immer situativ, spezifisch in ihrer Situation, in ihren gegebenen raumzeitlichen Umständen und unter Berücksichtigung der konkreten Partizipanden, die das gemeinsame Geschehen wechselseitig verfertigen, neu entsteht und damit einmalig ist, sind demgegenüber *Praktiken* niemals ausschließlich situiert, sondern entstehen erst durch ihre übersituative Verteiltheit. Praktiken sind mit Hirschauer (2016: 46) „situationsübergreifende, analytische Einheiten“, deren „kollektiver Sinn“ (Hirschauer 2016: 56) sich nicht ohne kulturelles Hintergrundwissen erschließen lässt. Die übersituative Verteiltheit von Praktiken in den Blick zu nehmen, ist gerade das, was verschiedene praxeologische Ansätze miteinander verbindet (vgl. Hillebrandt 2014: 58).
- (2) Während die Untersuchung sprachlicher Praktiken auf der Ebene des Sprachgebrauchs und der dabei genutzten „verbalen, vokalen und visuellen Ressourcen“ (Selting 2013: 589) ansetzt, ist die Analyse sozialer Praktiken auf, wenn man so will, ‚größere Einheiten‘ fokussiert. Für den Gegenstand der Smart Speaker sind dies z. B. Üben, Testen und Vorführen, außerdem soll die soziale Praktik des Gesprächs im Zentrum des Interesses stehen. Die Konstitutionsverhältnisse sind jeweils umgekehrt: Sprachliche Praktiken konstituieren Handlungen, miteinander verbundene und verteilte Handlungsschritte können soziale Praktiken konstituieren. Sprachliche Praktiken können daran aber beteiligt sein und wie unter Rückgriff auf Ludwig Wittgenstein und das Konzept der „Sprachspiele“ deutlich wurde, sind sie sogar wesentlich an deren Hervorbringung beteiligt.

Sprachwissenschaftlich können also beide Domänen im Rahmen von Video-Interaktionsanalysen untersucht werden: Einerseits werden die Ressourcenbündel aufgedeckt, die für sprachliche Praktiken mit VUIs relevant sind. Andererseits werden über die Analyse des Sprachlichen auch die sozialen Praktiken und deren Anatomie verständlich. Die Analyse von Gesprächen mit ihren Beteiligungsstrukturen und materiellen Partizipanden kann dabei auch Aufschluss darüber geben, welchen (sozialen) Stellenwert ein Smart Speaker haben kann. Nicht zuletzt soll die Analyse auch zeigen, in welcher Beziehung der Sprachgebrauch zu den technischen Präkonfigurationen der Interfaces steht.

Linguistisch knüpft die Arbeit also sowohl an eine Perspektive aus der Gesprächsforschung als auch an eine bisher eher medienlinguistisch fruchtbar gemachte Sichtweise an. Letztere manifestiert sich erst in neuerer Zeit: Als ‚neue Medien‘ zur Jahrtausendwende zunehmend in den Blick interaktionsorientierter Sprachwissenschaft kamen, etablierten sich in der Linguistik zunächst das Kon-

zept der Kommunikationsformen, mit dem Medien v. a. von ihren technologischen und institutionellen Eigenschaften her in den Blick genommen wurden (vgl. Dürscheid 2005; Brock/Schildhauer 2017), sowie kulturlinguistische Ansätze, die mit einem ‚weiteren‘ Medienverständnis arbeiteten (vgl. Schneider 2008; Linz 2016; Schneider 2017).³⁰ Eine sozialtheoretische Perspektive, die nicht bereits von den Konventionen des Mediengebrauchs ausgeht, sondern den Gebrauch – die *Praxis* – in den Mittelpunkt stellt (und damit eine materielle, lokale und temporale Dynamik³¹ in den Blick nimmt, die mit anderen Ansätzen nicht gleichermaßen eingefangen werden kann), etabliert sich in der Medienlinguistik zunehmend (vgl. Deppermann/Feilke/Linke 2016; Meiler 2019; 2021; Habscheid et al. 2021: 24–29; Waldecker/Hector 2023); einem solchen Ansatz soll hier gefolgt werden.³²

Mit dieser Ausrichtung kann also zusammengefasst erstens Gegenstand der Untersuchung werden, wie die Nutzer*innen durch eine ‚Anhäufung‘ von sprachlichen und mithin sozialen Praktiken selbst das Gerät in der Praxis aneignen und es domestizieren³³ und wie sie zu kompetenten Nutzer*innen werden, und wie die sozialen Praktiken insbesondere im Verhältnis zu ihren Interfaces (als Interface-Praktiken) gestaltet sind. So kann in den Blick kommen, wie das Interface dabei als materieller Partizipand an konversationellen Vollzügen ‚beteiligt‘ wird und welche Verfahren dazu eingesetzt werden. Mit einer Analyse sprachlicher Praktiken im Gebrauch können zweitens sprachliche Praktiken entschlüsselt werden, die in Dialogen mit VUIs zum Einsatz kommen können. Der Charakter der Analyse lässt nicht zu, einen abschließend zu verstehenden ‚Katalog‘ von sprachlichen Praktiken aufzustellen, was angesichts der hohen Flexibilität und Kreativität

30 Eine vertiefte Diskussion zum Medienbegriff in unterschiedlichen Teilen der Sprachwissenschaft soll an dieser Stelle unterbleiben, siehe dazu neben den genannten Belegstellen etwa Habscheid (2020; 2024) sowie Habscheid et al. (2021: 24–29).

31 Siehe Meiler (2021).

32 Sprachtheoretiker*innen wie Ludwig Wittgenstein oder Karl Bühler sowie einzelne Autor*innen wie Charles Goodwin oder einzelne Verwendungsweisen des Praktiken-Begriffs etwa bei Fiehler et al. (2004) sollen dabei nicht übersehen werden; im wissenschaftsgeschichtlichen Überblick siehe Meiler (2019). Gloning (2016: 460) merkt zur zunehmenden Etablierung des Praktikenbegriffs als ‚Innovation‘ in der Sprachwissenschaft kritisch an, dass die Linguistik in ihrem Denken schon lange Ideen des Gebrauchs und der Praxis von Sprache verankert und „nicht nur aus de Saussure und Chomsky“ bestehe und bestand; er nennt zahlreiche weitere Arbeiten, in denen sich diese Tradition zeige. Er argumentiert allerdings in Abgrenzung zu Pennycook (2010: 6), der einen Verzicht auf derartige Perspektiven mit der Verfestigung der Sprachwissenschaft und dem „Kampf“ um die Anerkennung als eigene Disziplin begründet – und wendet ein, dass im Forschungsdiskurs Arbeiten, die solche Perspektiven bereits früher eingenommen hatten, nicht genügend Berücksichtigung fänden. Diesen wird durch die Ankunft des „practice turn“ in der Linguistik jedenfalls nun zunehmend Aufmerksamkeit zuteil.

33 Zur Bildung dieser beiden Begriffe siehe Kap. 2.3.

tät beim Vollzug sprachlicher Praktiken auch gar nicht sinnvoll wäre. Sprachliche Praktiken machen aber die Transformation bestehender bzw. die Emergenz sprachlicher Praktiken an der Oberfläche sichtbar und lassen insofern Aussagen über das Verhältnis von Sprachgebrauch und Medien bzw. Interfaces zu – ob schon solche Veränderungen nicht einfach „ablesbar“ sind, sondern einer methodisch vorgenommenen Analyse bedürfen (vgl. Deppermann 2001: 59–60). Die Analyse sprachlicher Praktiken ist insofern auch geeignet, um die Veränderung und Neubildung von Praktiken durch bzw. mit VUIs zu erfassen.

2.2 Zum Verhältnis von Mensch und Maschine

Auf diesem praxeologischen Fundament sollen Formen des zwischenmenschlichen Austauschs sowie des Austauschs zwischen Mensch und Maschine sozialtheoretisch konzeptualisiert und begrifflich gefasst werden. Dabei beginnt das Kapitel mit einer Klärung der Termini ‚Interaktion‘ und ‚Gespräch‘ (Kap. 2.2.1), die als zwischenmenschliche Austausch-Situationen begriffen werden. Daran anschließend kann der Austausch zwischen Mensch und Maschine als ‚Dialog‘ gefasst, davon abgegrenzt und unter Bezug auf bisherige Forschung aus diesem Bereich genauer charakterisiert werden (Kap. 2.2.2 und 2.2.3), um die daraus abgeleitete Terminologie in den hieran anschließenden Teilen der Arbeiten sauber verwenden zu können, auch mit Blick auf die methodologische Verortung im Bereich der ethnomethodologischen Konversationsanalyse bzw. der ethnografischen Gesprächsanalyse.³⁴ Das Kapitel begegnet darüber hinaus auch soziologischen Debatten zur Handlungsträgerschaft bzw. Agency von Maschinen und greift dabei die praxeologischen Grundlagen wieder auf, um ein Verständnis von Mensch-Maschine-Verhältnissen zu entwickeln, auf das in den Analysen rekurriert werden kann (Kap. 2.2.4). Die Ausführungen zum theoretischen Hintergrund dieser Begriffe sollen insofern auf das beschränkt bleiben, was für das Verständnis ihres Gebrauchs in der Arbeit notwendig ist; ein Beitrag zu theoretischen Grundsatzfragen, die sich an allen drei Begriffen (Interaktion, Gespräch, Dialog) aufhängen, liegt nicht im Interesse dieser Arbeit. Die Notwendigkeit einer Klärung und genaueren Charakterisierung ergibt sich gleichwohl aus dem Untersuchungsgegenstand der Voice User Interfaces (VUIs) und der Frage, wie der Verlauf und das Produkt eines Austauschs zwischen Anwender*innen und VUIs begrifflich und konzeptionell gefasst werden können.

³⁴ Siehe Kap. 4.

2.2.1 Interaktion und Gespräch

Soziale Interaktionen sollen im Anschluss an den Soziologen Erving Goffman (1983: 2) als „environments in which two or more individuals are physically in one another's response presence“ begriffen werden. In fokussierten Interaktionen gehört dazu bei Goffman weiterhin ein geteilter Aufmerksamkeitsfokus, eine geteilte Wahrnehmung und die wechselseitige Wahrnehmung dieser Wahrnehmung (vgl. Goffman 1983: 3) sowie eine koordinierte Ausrichtung von Handlungen am Gegenüber. Damit folgt die vorliegende Arbeit den Ausführungen von Antonia Krummheuer (2010: 13–30), in denen unter Rückgriff auf die Studien Goffmans eine vergleichsweise eng gefasste Auffassung von sozialer Interaktion entwickelt wird. Krummheuer schlägt dies in ihrer Analyse zur Abgrenzung von Austausch-situationen mit einem „Embodied Conversational Agent“³⁵ vor und entwickelt für Letzteres den Begriff des „hybriden Austauschs“ (Krummheuer 2010: 5), um zu vermeiden, den Austausch zwischen menschlichen und technischen Entitäten *a priori* als soziale Interaktion zu bezeichnen. Krummheuer (2010: 323) kommt in ihrer Untersuchung zu dem Schluss, dass im hybriden Austausch zwar durchaus vorübergehend Eigenschaften von Interaktion aufzufinden sind, diese aber nicht kontinuierlich aufrecht erhalten bleiben und der technisch präfigurierte Charakter des Austauschs im Vordergrund steht: „Er [der Austausch, T.H.] ist somit weder eine soziale Interaktion noch ein rein instrumenteller Umgang mit einem technischen Artefakt. Der Austausch changiert vielmehr zwischen den Polen“ (Krummheuer 2010: 323). Die Nutzer*innen schreiben also weder sich noch dem technischen Artefakt konsequent zu, Interaktionspartner*innen zu sein.

Die Domäne der Interaktion ist also menschlichen Akteur*innen vorbehalten – diese Perspektive wird auch in der vorliegenden Arbeit übernommen – und bezeichnet lediglich solche Situationen, in denen mindestens zwei menschliche Beteiligte ko-präsent sind, einander wahrnehmen, diese Wahrnehmung wahrnehmen, und miteinander kommunizieren. Dabei teilen sie einen Aufmerksamkeitsfokus und richten ihre Handlungen aneinander aus (vgl. Goffman 1983: 2–3). Absichtlich wird hier insofern mit Goffman ein stärker an Subjekten orientiertes Interaktionsverständnis gewählt als etwa in den Arbeiten Garfinkels, der von der Interaktion ausgehend vielmehr ein Interesse an deren praktischer Hervorbringung und kontinuierlicher Konstitution entwickelte (vgl. vom Lehn 2021: 300), die auch in den folgenden Analysen im Zentrum stehen.³⁶ Diese

³⁵ Siehe Kap. 3.1.2.

³⁶ Siehe auch Kap. 4.1.

beiden Aspekte – Interaktion auf der einen und ihre praktische Hervorbringung auf der anderen Seite – sollen allerdings in der vorliegenden Arbeit zunächst *analytisch* getrennt werden, um zwischenmenschliche, soziale, sprachliche Interaktion und den Austausch mit einem VUI voneinander abzugrenzen und sprachliche Praktiken aufdecken zu können, die spezifisch für Letzteres sind. Damit soll gerade *nicht* ausgeschlossen sein, dass auch der Austausch zwischen Mensch und VUI phasenweise interaktionsähnliche Eigenschaften hat – vielmehr geht diese Festlegung gerade davon aus, dass hier auf der Phänomenebene eine Unschärfe besteht. Ähnlich wie in Krummheuers Arbeit kann dies aber nicht vor der Analyse festgelegt werden, da ja u. a. die sprachlichen Praktiken Gegenstand der Untersuchung sein sollen, mit denen VUIs an der Praxis ‚beteiligt‘ werden. Insofern scheint es sinnvoll, wenigstens einen Begriff für den rein zwischenmenschlichen, an Subjekten und Handlungen orientierten Austausch vorzubehalten.

Mit dem Begriff *Gespräch* wird hingegen im Folgenden eine soziale Praktik bezeichnet – die des konzentrierten Austauschs von Redebeiträgen und deren Organisation. Mit Goffman (1979: 6)³⁷ kann unterschieden werden zwischen „talk“ und „a talk“ – dem „Reden“ und dem „Gespräch“. Ein Gespräch benötigt zu seiner Konstitution ein irgendwie gestaltetes „Reden“, wobei dies nicht notwendigerweise verbalsprachlich sein muss – man denke etwa an Gebärdensprache. Hingegen entsteht nicht durch jedes Reden schon ein Gespräch: Reden kann nämlich auch Teil einer anderen Praktik sein, „one in which nonlinguistic events may have the floor“ (Goffman 1979: 15), etwa bei Verkaufssituationen, Begegnungen zwischen Gästen und Servicepersonal in einem Restaurant oder ärztlichen Untersuchungen. Umgangssprachlich ist dann zwar häufig auch von Gesprächen die Rede, obschon die vollzogene Praktik nicht primär ein Gespräch darstellt. Sehr viele Alltagspraktiken haben gesprächsartige Anteile, sind aber darum noch nicht oder nicht in erster Linie ein Gespräch. Im Folgenden sollen als Gespräch v. a. solche Situationen bezeichnet sein, in denen die Durchführung dessen im Zentrum steht. Solche Situationen können je nach sozialem Kontext und möglicher Zweckbindung der Gespräche sehr unterschiedlich gestaltet sein, wozu die Forschung zu kommunikativen Gattungen (Günthner/Knoblach 1994) einen großen Beitrag leisten konnte, auch im interkulturellen Vergleich (vgl. Günthner/Linke 2006).

37 Die im Folgenden verwendeten deutschen Begriffe sind der Übersetzung von Knoblach/Leuenberger/Schnettler (2017) entnommen.

Zugleich wird methodologisch bei der Untersuchung von Gesprächen als soziale Praktik in der vorliegenden Arbeit an die Gesprächsforschung nach Deppermann (2008) angeknüpft. Gespräche setzen sich demzufolge zusammen aus gemeinsam zu verfertigenen Handlungen: Die „Interaktivität“ und der „Handlungscharakter“ sind mit Deppermann (2000: 98) zwei der konstituierenden Prinzipien von Gesprächen. Zusammen mit drei weiteren Eigenschaften – ihrer Methodizität, ihrer Sequenzialität und ihrer Reflexivität – stellen sie die ‚Anatomie‘ von Gesprächen dar (hierfür und für das Folgende vgl. Deppermann 2000: 98). Die *Methodizität* verweist auf den mehr oder weniger hohen Grad an methodischer Routinisierung bei der Durchführung von Gesprächen. Die *Sequenzialität* ist als grundlegende Eigenschaft durch die Zeitlichkeit bedingt; sie gilt insbesondere für die verbalsprachlichen Anteile des Gesprächs und steht vor der Herausforderung, diese Anteile nicht überzufokussieren und die Gleichzeitigkeit anderer Bestandteile des Gesprächs nicht zu übersehen (vgl. Deppermann/Schmitt 2007: 27–29).³⁸ Die *Reflexivität* ist die wesentliche übersituative Eigenschaft von Gesprächen: Dadurch, dass die Gesprächsteilnehmenden ihr Gespräch in einem Kontext auf eine bestimmte Weise führen, „bestätigen [sie] stets aufs Neue die Wirklichkeit und Gültigkeit der betreffenden Kontexte“ (Deppermann 2000: 98; siehe auch Garfinkel 1967: 1) – und machen sie so übertragbar auch auf andere Situationen.

2.2.2 Dialog zwischen Mensch und Maschine

VUIs ermöglichen den Austausch zwischen Mensch und Maschine über dialogisch gestaltete verbale und non-verbale Signale. Der Austausch zwischen menschlichen und maschinellen Einheiten kann jedoch – das wird aus den obigen Ausführungen deutlich – nicht mit zwischenmenschlicher Interaktion oder einem Gespräch gleichgesetzt werden – zu verschieden sind die in einen Austausch tretenden Entitäten (vgl. Pfänder/Wagner 2009: 92; Krummheuer 2010). Für den ‚Austausch‘ mit den VUIs soll daher der Begriff *Dialog* verwendet werden, auch wenn diesem Begriff aus Sicht der empirischen Linguistik durchaus skeptisch begegnet wurde (vgl. Imo 2016: 351; Auer 2016). Dialog steht dabei nicht in (etymologisch ohnehin fragwürdiger) Opposition zu Monolog oder Polylog (vgl. Linell 2009: 3), auch soll mit dem Begriff nicht die praxeologische Analyse auf Sprachliches oder gar ‚Vernünftiges‘ reduziert werden, wovor Auer (2016: 358) warnt.³⁹ Vielmehr wird versucht, mit dem Dialogbegriff einen weiter gefassten, komplementär zu

³⁸ Siehe auch Kap. 4.4.

³⁹ Für die Methodologie der Arbeit siehe Kap. 4.

Interaktion verstandenen Begriff für den Austausch zwischen Mensch und Maschine zu entwickeln. Michail Bachtin grenzt in seinen sprachphilosophischen Arbeiten ein ‚enges‘ von einem ‚weiten‘ Verständnis von Dialog ab: Das enge Verständnis, der „real-life-dialogue“ (Bachtin 1986: 75), sind Gattungen des alltäglichen Sprechens (vgl. Velmezova 2015: 277). Linell (2009: 4) bezeichnet dies in der Hinführung zu seiner auf Bachtin gründenden Dialogtheorie als „[t]he most down-to-earth meaning of ‚dialogue‘“, als „direct interactive encounter between two or more, mutually co-present individuals ...“. Genau mit dieser Bedeutung von Dialog soll hier *nicht* gearbeitet werden.

Stattdessen wird ‚Dialog‘ in einem weiten, abstrakten Sinn verstanden (vgl. Linell 2009: 5). Bei Bachtin ist dieser weite Sinn von Dialog ein sehr breit gefasstes, nahezu ausuferndes Konzept, das er auf unterschiedliche Lebensbereiche und Analysefelder angewendet hat – darunter bewusstseinspsychologische, religiöse, kulturorientierte, ontologische und ethische bzw. philosophische Fragestellungen (vgl. Velmezova 2015: 277). Dabei steht im Mittelpunkt eine grundlegende Orientierung auf etwas Anderes oder eine*n Andere*n – die „other-orientation“ (Linell 2009: 13). Alle Diskursbeiträge zeichnen sich durch eine grundlegende Responsivität (auf irgendwelche anderen, kontextuellen Bezugspunkte) und Antizipation (möglicher Folgebezüge) aus – ohne, dass dabei geklärt sein muss oder sichtbar ist, wer oder was *genau* die andere Einheit ist und ohne, dass dies Menschen vorbehalten wäre. So haben auch (technische) Artefakte durch in sie eingeschriebene Eigenschaften responsive und antizipative Bedeutungspotenziale, die wiederum in Relation zu den Nutzer*innen der Artefakte treten (vgl. Linell 2009: 346).⁴⁰ Linell (2009: 350) empfiehlt explizit für rezente Entwicklungen im Bereich Künstlicher Intelligenz, dialogorientierte Ansätze zu nutzen und Dialoge (im weiteren Sinne) zwischen Nutzer*innen, Hard- und Software zu betrachten.

So mag der Dialogbegriff bei Bachtin einerseits teils eklektisch, schwierig zu operationalisieren und uneindeutig sein. Bachtin-Exegese ist, wie auch Auer (2016: 357) andeutet, ein ebenso umfangreiches wie, auch aufgrund der sprachlichen Barrieren zum Russischen, schwieriges Unterfangen, das hier nicht in der Tiefe verfolgt werden soll. Andererseits lädt der weitreichende Dialogbegriff, wie ihn im Anschluss daran auch Linell beschreibt, geradezu dazu ein, für noch unbe-

⁴⁰ Linell (2009: 346) führt zur weiteren Auseinandersetzung das Konzept der Affordanzen ein, mit dem Beziehungen beschrieben werden können zwischen (a) den Nutzungsszenarien, für die Artefakte vorgesehen sind, (b) dem, was sie tatsächlich an Nutzungsweisen „anbieten“, und (c) den darauf aufbauenden Praktiken, die Nutzer*innen damit vollziehen; eine vertiefte Auseinandersetzung damit führt an dieser Stelle zu weit und ist für die späteren Analysen nicht von Relevanz. Zum Affordanzkonzept siehe Gibson (1977), Zillien (2008; 2019) sowie Pentzold/Bischof (2019).

stimmte Kategorien des Austauschs verwendet zu werden, denen – schon durch ihre Interfaces – unbestreitbar ein hoher Grad an Dialogizität innewohnt und über die durch eine dialogische Analysehaltung Erkenntnisse gewonnen werden können. Auch Fritz/Hundsnurscher (1994a: XIV) verwenden den Begriff in dieser Weise und wenden ihn explizit auf den „Dialog“ zwischen Mensch und Maschine“ an – als deviante Form zu zwischenmenschlichen Interaktionen.

Es soll in diesem Sinne also von einem ‚Dialog‘ im weiteren Sinne gesprochen werden, wenn die Nutzer*innen und die VUIs in einen Austausch treten (*VUI-Dialog*). Dieser Dialog sieht zwar durch die Spezifik von VUIs an der Oberfläche einer Interaktion nach Goffman (oder einem Dialog im engeren Sinne nach Bachtin) durchaus ähnlich, wie wir noch sehen werden – und gerade das macht ihn als linguistischen Untersuchungsgegenstand interessant – soll damit aber nicht verwechselt werden. VUI-Dialoge sind Bezugnahmen auf technische Artefakte, wie sie auch für die nicht-stimmliche, sondern z. B. auf Maus und Tastatur basierende Bedienung von PCs angewendet werden könnten. Interfaces konstituieren sich – wie im vorherigen Kapitel dargestellt – prozessual an und in diesen Dialogen, d. h. im Sprechen und Hören; Interfaces sind „der Ort eines ‚Dialogs‘ zwischen Mensch und Computertechnologie“ (Ernst/Bächle 2020: 419) – ein Ort, der nur dann entsteht, wenn dieser Dialog auch stattfindet.

Vor dem Hintergrund, dass die genaue Funktionsweise von Smart Speakern und damit auch VUIs den Nutzer*innen zu großen Teilen opak bleibt und nicht offensichtlich ist, mit wem genau gesprochen wird und wohin das Gesprochene übertragen wird, scheint eine Betrachtung als Interaktion, die ja gerade die wechselseitige Orientierung am Gegenüber und die Wahrnehmungswahrnehmung zum Ausgangspunkt hat, auch theoretisch unpassend. Eine Betrachtung hingegen, die die grundlegende Responsivität und Antizipation als Eigenschaften der Äußerungen beider Parteien setzt, nicht aber die zeitlichen oder sozialen Bezugspunkte *a priori* festlegt, die vielmehr auch unklar, vielschichtig und diffus sein können, scheint für die Konzeptualisierung dieser Art von Äußerungen⁴¹ bestens geeignet.

Wie an den bisherigen Ausführungen deutlich wird, sind die Konzepte Interaktion, Gespräch und (VUI-)Dialog nur analytisch zu trennen. Aus Perspektive der Teilnehmenden vollziehen sie sich in einer einzigen sozialen Praxis, in der nicht nur die einzelnen Konzepte sich gleichzeitig vollziehen, sondern auch übergänglich sind. War ein Austausch zwischen Ärztin und Patient zunächst eine ärztliche Untersuchung, in der das Reden keine gesprächshafte Funktion hatte, kann der

41 Wie Velmezova (2015: 287–288) ausführt, sind es für Bachtin gerade die Äußerungen („utterances“), die in einer genaueren dialogischen Untersuchung vielversprechend sind.

daran anschließende Austausch durchaus eine spezifische Form des Gesprächs annehmen, z. B. ein Beratungs- oder Aufklärungsgespräch, das dann auch in einer veränderten körperlichen Konstellation stattfindet: Die Ärztin wird sich möglicherweise bemühen, einen besseren Blickkontakt zum Patienten herzustellen, ggf. ist auch entsprechendes Mobiliar vorhanden, das den Austausch rahmt und als Interaktionsressource dienen kann. Je nach Gesprächstyp sind also die körperlichen und lokalen Rahmenbedingungen anders. Zu beiden Teilen dieser exemplarischen sozialen Situation gehört außerdem (unterschiedlich ausgeformte) Interaktion.

Gerade daran zeigt sich aber im Hinblick auf die Fragestellungen die Sinnhaftigkeit einer analytischen Trennung: Der Charakter des sprachlichen Austauschs mit stationären VUIs in den häuslichen Alltagsumgebungen der Nutzer*innen ist erst durch wenige Arbeiten empirisch beleuchtet worden, die noch diskutiert werden (vgl. Porcheron et al. 2018; Beneteau et al. 2019; Habscheid 2022; Hector 2022; 2025; Habscheid/Hector/Hrncal 2023; 2025; Reeves/Porcheron 2023). In diesen Arbeiten – und auch in anderen Studien zu ‚artverwandten‘ Geräten, u. a. bei Krummheuer (2010) oder Pitsch (2015) – wird erstens deutlich, dass der Austausch an der Oberfläche Ähnlichkeit zu menschlichen Gesprächen hat – er vollzieht sich primär verbalsprachlich, handlungsorientiert, sequenziell und methodisch und erfüllt damit schon einige der weiter oben referierten Eigenschaften nach Deppermann (2000: 98–99). Zweitens vollzieht er sich parallel und potenziell überlagernd zu zwischenmenschlichen Gesprächen. Da gerade dieser Austausch, sein Verhältnis zu menschlichen Gesprächen und die ihn verfertigenden sozialen Praktiken Gegenstand der Untersuchung sein sollen, bietet es sich zum einen an, die Unbestimmtheit des Austauschs auch begrifflich zu fassen und von VUI-Dialogen (im Bachtinschen, weiten Sinn) zu sprechen. Es ist insofern das Anliegen auch dieses Begriffsinstrumentariums, von einer zu starren konzeptionellen Festlegung über den Austausch mit VUIs Abstand zu nehmen. Zum anderen soll eine möglichst große konzeptionell-begriffliche Sicherheit für die parallel vollzogene soziale Praxis geschaffen werden.

2.2.3 Charakteristika von Mensch-Maschine-Dialogen

Auch wenn mit ‚Dialog‘ also ein Begriff für den Austausch zwischen Mensch und Maschine gefunden ist, muss dieser insbesondere für die in späteren Kapiteln folgenden Analysen zum Gesprächsbeteiligten-Status und zur praktischen Einbindung in interaktive Prozesse der sozialen Praxis genauer bestimmt und charakterisiert werden. Dazu sollen zunächst die Arbeiten der Anthropologin Lucy Suchman als Pionierin auf dem Feld der Mensch-Maschine-Interaktion und der

interdisziplinären Betrachtung zwischen Computer- und Sozialwissenschaften ausgewertet werden, die auch methodologisch bereichernd sind. Unter besonderer Berücksichtigung der Analysen der Soziologin Antonia Krummheuer (aus dem Feld der Wissenschafts- und Techniksoziologie) sowie des sozialtheoretischen Entwurfs nach Stefan Hirschauer werden anschließend Konzepte zur Kommunikation zwischen Mensch und Maschine und zur ‚Beteiligung‘ von Maschinen am praktischen Geschehen vorgestellt und auf die Fragestellungen der vorliegenden Arbeit angewendet.

Suchman unternahm in ihrer Arbeit „Plans and Situated Actions“ (1987) als eine der ersten Forscher*innen den Versuch, zwischen Informatik und Sozialwissenschaften zu vermitteln (vgl. Suchman 2007: 4–5).⁴² Sie ging dabei der Frage nach, wann Kategorien, Eigenschaften und Zuschreibungen von „Mensch“ und „Maschine“ relevant werden und wie diese Unterschiede in bestimmten Situationen aufgeführt (englisch „enacted“) werden – und welche Konsequenzen dies diskursiv und materiell hat (vgl. Suchman 2007: 2). Mehr noch als die erste Ausgabe ihrer Arbeit von 1987 prüft sie in der zweiten Auflage: Wie werden die Grenzen gezogen und *in situ* von den Beteiligten relevant gesetzt? Suchman betont, dass sie kategorial-theoretische Debatten über das Sein und das Wesen von Maschinen und Menschen ablehnt und dieser mit empirischer Evidenz begegnen will – dieser Position schließt sich die vorliegende Arbeit an. In ihren Arbeiten geht Suchman mit Mitteln vor, die denen der ethnomethodologischen Konversationsanalyse ähnlich sind (vgl. Suchman 2007: 10–11). Suchman (1990) stellt heraus, dass in verschiedenen kommunikativen Arrangements auch verschiedene kommunikative Ressourcen im Gebrauch sind. Der Austausch zwischen Mensch und Maschine lässt sich insofern als eine Extremform von ressourcenlimitierter Kommunikation fassen, auf den die gleichen Methoden angewendet werden können – d. h. bei Lucy Suchman ethnomethodologisch-konversationsanalytische Verfahren (vgl. Suchman 1990: 43). Über diesen methodologischen ‚Trick‘ gelingt es Suchman, die Unterschiede zwischen „human interaction and machine operation“ aufzuzeigen und sich von „a priori identification“ (Suchman 1990: 46) der Eigenschaften zu lösen.

Durch ihre detaillierten Analysen konnte sie zeigen, dass die Ko-Konstruktion des Austauschs zwischen Mensch und Maschine wesentlich davon geprägt war, dass Asymmetrien zwischen den beiden Entitäten herrschten: Während die

⁴² Eher theoretisch ausgerichtete Versuche einer Vermittlung begannen bereits in den frühen 80er-Jahren (siehe etwa Friedrich 1980), doch entbehrten sie überwiegend empirischer Untersuchungen und fokussierten auch nicht auf die interaktionistischen Aspekte des Mensch-Maschine-Austauschs. Die ebenfalls zentralen Arbeiten von Agre/Chapman (1987) werden im weiteren Verlauf der Ausführungen noch diskutiert.

Anwender*innen keinen Zugang zu den in die Maschine eingebauten Skripten und damit zu deren ‚Erwartungen‘ hatten, kann die Maschine den größten Teil menschlicher Handlungen nicht wahrnehmen: Für die Ko-Konstruktion sind lediglich Eingaben relevant, die den Status der Maschine verändern können. Menschliches Handeln ist aber immer – selbst, wenn es vorgeplant ist – situativ und selbst das Planen selbst ist situiert. Planen sowie auch ‚das Handeln an dem Plan orientiert zu begreifen‘, werden als von den Beteiligten gemeinsam hervorbrachte Leistungen verstanden, die Garfinkel/Sacks (1979: 135) folgend, sprachlich „beobachtbar“ und „berichtbar“ (mithin „accountable“) gemacht werden (vgl. Suchman 2007: 13).⁴³ Das Handeln als geplantes Handeln zu verstehen und zu kennzeichnen, „accountable“ zu machen, ist insofern eine reflexive Sinnzuschreibung nach Garfinkel (1967: 1).

Suchman (1987) zeigt in ihren Analysen von Erstnutzungen eines Hilfesystems zur Bedienung eines Fotokopierers (siehe Suchman 2007: 126–168), dass die fehlende Fähigkeit zur reflexiven Hervorbringung und gerade die menschliche Fähigkeit zur Reflexivität und deren Anwendung Blindstellen bei der Maschine verursacht:

The new user of a system, however, is engaged in ongoing, situated inquiries regarding an appropriate next action. Although the instructions of the expert help system are designed in anticipation of the user's inquiries, problems arise from the user's ability to move easily between a simple request for a next action [...] and embedded requests for clarification of the actions described within a procedure. (Suchman 2007: 167)

Das Design einer Maschine ‚erzwingt‘, um diese Blindstellen und fehlende Reflexivität auszugleichen, eine bestimmte Reihenfolge für den Ablauf von Handlungen und erzeugt dadurch den Eindruck einer maschinellen Antwort – und „enables the appearance of instructions occasioned by the user's actions“ (Suchman 1990: 31). Die menschlichen Handlungen sind also Aufführungen in einem vorstrukturierten Plan, der den Austausch überhaupt ermöglicht – die *tatsächliche* Aufführung allerdings bleibt der Maschine weiterhin verborgen, und somit z. B. alle anderen soziomateriellen Vorgänge, Einstellungen, unterschiedliche Auffassungen, die die Nutzer*innen von der Planstruktur haben, Unterbrechungen und Störungen.

⁴³ Suchman (2007: 71) verweist auch auf die Arbeiten G. H. Meads zum Verhältnis von situierter Aktivität und geplanter Aktivität, wobei sie betont, dass der geplanten Aktivität schon die Vorstellung von der Situierung innewohnt. Diese sozialtheoretische Diskussion soll hier nicht weiter verfolgt werden.

Es zeigt sich also eine Differenz, die die Informationswissenschaftler Philip Agre und David Chapman⁴⁴ herausgearbeitet haben: „plans-as-program“ und „plans-as-communication“. Während die erste Perspektive v. a. die Ausführung von determinierten Handlungen oder Schritten in einer miteinander verbundenen Abfolge beschreibt, ist die zweite nicht-deterministisch. Sie geht von einer fortlaufenden Improvisation bei der Befolgung von Plänen aus, die als eine Ressource verstanden werden können, die bei der Bewältigung sozialer Aufgaben zum Einsatz kommt (vgl. Agre/Chapman 1987: 17). Ein Hauptanliegen, das Suchman, Agre und Chapman teilten, war die Vermittlung dieser Unterscheidung in die Sozialwissenschaften, und mehr noch in die Informatik hinein, in Verbindung mit daraus abgeleiteten Konsequenzen für das User-Design, um die Grenzen der ‚Planbarkeit‘ sozialer Handlungen einerseits, aber auch die Geplantheit der maschinellen Prozesse andererseits in eine Verbesserung der Systeme einfließen zu lassen.

Suchman, Agre und Chapman begreifen auf diese Weise den Austausch zwischen Mensch und Maschine als asymmetrischen Dialog, in dem unterschiedliche Konzepte vom *praktischen* Ablauf von Handlungen und unterschiedliche Ausdrucksressourcen zu Missverständnissen, Störungen und Abbrüchen führen können und andererseits Verfahren entwickelt werden, um diese Grenzen zu identifizieren und zu überwinden. Die dabei genutzte Vorgehensweise der ethnomethodologischen Konversationsanalyse soll auch in der vorliegenden Arbeit zum Einsatz kommen.⁴⁵ Eine Folge von Äußerungen wird sequenziell analysiert und es wird dabei an der Oberfläche zunächst nicht zwischen menschlichen und maschinellen Äußerungen unterschieden – die Situation wird als Ganzes betrachtet und alle Ausdrucksressourcen werden einbezogen. Die Erkenntnisse von Suchman über Asymmetrien und fehlende Ressourcen bzw. Wahrnehmung sowie die Vorstellungen von „plan-as-program“ und „plan-as-communication“ werden herangezogen, um interaktive Abläufe aus verschiedenen Perspektiven zu analysieren, Beteiligungsstrukturen herauszuarbeiten und Störungsquellen zu identifizieren.

Die Analyse soll dabei auch die Ergebnisse von Krummheuer einbeziehen und ein darauf sowie auf die Arbeiten von Suchman aufbauendes Verständnis

⁴⁴ Wie Suchman (2007: 13–14) beschreibt, haben sich Suchman, Agre und Chapman insbesondere in den 1980er-Jahren wechselseitig in ihrer Forschung beeinflusst; sie waren zur gleichen Zeit an der Erforschung der Einbettung technologischer Systeme in soziale Zusammenhänge interessiert – Agre und Chapman aus informationswissenschaftlicher Perspektive, Suchman aus anthropologischer. Agre entwickelte aufbauend auf dem Austausch seine Konzeption von „Critical Technical Practice“ (Agre 1997; siehe auch Hind/Seitz 2021).

⁴⁵ Siehe Kap. 4.

von VUI-Dialogen entwickeln. Ähnlich wie Suchman sucht auch Krummheuer (2010: 6) „keine ontologische Debatte darüber, ob Technik handeln kann oder nicht“. Stattdessen werden strikt empirisch Situationen analysiert, in denen Nutzer*innen mit einem virtuellen Agenten in Kontakt treten. Stellt Suchman v. a. die Asymmetrien heraus, die sich in der Kommunikation mit Maschinen ergeben, betont Krummheuer (2010: 10; 323–325) den *hybriden* und *ambigen* Charakter des Mensch-Maschine-Dialogs. Während Hybridität im menschlichen Alltag Krummheuer (2010: 323) zufolge ein gängiges Phänomen ist, das bei allen menschlichen Anwendungen von technischen oder dinglichen Artefakten entsteht, wird es im Rahmen der Austauschsituation zwischen Menschen und dialogischen Maschinen (wie in Krummheuers Beispiel dem virtuellen Agenten Max) relevant gesetzt: So betone einerseits der Agent regelmäßig sowohl seine Eigenschaft als technisches Artefakt als auch seine Rolle als sozialer Interaktionspartner, andererseits orientierten sich auch die Nutzer*innen auf den ‚Zwischenstatus‘ des Geräts. Dieses weist sowohl menschliche Eigenschaften auf (z. B. die grundlegende Sprachfähigkeit, Mimik und Gestik oder visuelle Repräsentation durch einen dem Menschen nachempfundenen Körper auf einem Bildschirm), wie auch maschinelle Charakteristika (z. B. eine stark funktionalistische Syntax oder die physische Beschaffenheit nicht aus ‚Fleisch und Blut‘, sondern aus Plastik, Draht und anderen Materialien zum Aufbau von Computern): „Der hybride Austausch bezeichnet somit die wechselseitige technisch vermittelte Koordination von Aktivitäten zwischen zwei Welten und zwei Wesen“ (Krummheuer 2010: 105). Dieses ‚Zusammenlaufen‘ von Welten bzw. Eigenschaften, die zuvor unterschiedlichen Domänen angehörten (Menschen und Computern) und nun nicht mehr getrennt voneinander auftreten, wird sowohl durch die Nutzer*innen wie auch durch die Maschine selbst (bzw. durch die einprogrammierten Dialog-Bestandteile) immer wieder Thema im Austausch mit dem virtuellen Agenten: „Die Teilnehmer zeigen eine Orientierung am hybriden Status des Austausch und markieren diese Hybridität als Besonderheit“ (Krummheuer 2010: 325).

Gleichzeitig zeichnet sich der Austausch Krummheuer (2010: 324–325) zufolge durch eine unentschiedene Ambiguität aus. So wird der als hybride Einheit wahrgenommene Dialogpartner nicht konsequent unter dem Aspekt der Einheit betrachtet, sondern immer wieder werden auch die Unterschiedlichkeiten zwischen dem virtuellen Agenten und den Nutzer*innen betont. Gerade dadurch trete Krummheuer zufolge die Hybridität des Austauschs immer wieder hervor und werde immer wieder relevant gesetzt. Nicht nur die Nutzer*innen, sondern auch der virtuelle Agent stellt die Unterschiede häufig heraus und trägt insofern zu der ‚unentschiedenen Unentschiedenheit‘ mit bei, die die regelmäßige Thematisierung der Hybridität zur Folge hat. Dies zeigt sich besonders deutlich in von Krummheuer analysierten Erst-Begegnungen zwischen unerfahrenen Nutzer*innen und

dem virtuellen Agenten im Rahmen einer Veranstaltung. In diesen Situationen treten häufig Test-Situationen auf, in denen der virtuelle Agent als „fremdes und andersartiges Gegenüber“ angesprochen wird, dieser sich aber auch „selbst als ein ungewohntes Gegenüber dar[stellt]“ (Krummheuer 2010: 124).

Außerdem bestätigt Krummheuer (2010: 105) die Unterschiede in der Dialogführung, die Planbasiertheit des virtuellen Agenten im Unterschied zu Nutzer*innen, die es „gewohnt [sind], Gespräche situativ und interaktiv mit seinem Gegenüber auszuhandeln“. Diese Unterschiede führen zu einer „Kontextblindheit“ (vgl. Krummheuer 2010: 109–112), gerade wenn mehrere Nutzer*innen am Austausch mit dem virtuellen Agenten beteiligt sind. Unter Rückgriff auf die theoretische Rahmung von Goffman (1973) zeigt Krummheuer, dass das ‚Publikum‘, d. h. umstehende Personen, die im Rahmen der Veranstaltung den Austausch mit der Maschine verfolgen, wesentlich zur Teilnehmendenstruktur beiträgt und dem virtuellen Agenten – unentschieden – einerseits die Rolle eines „andersartigen“ Gesprächsbeteiligten zuschreibt, ihn gleichzeitig aber im Sinne Goffmans als „non-person“ behandelt, über welche in deren Anwesenheit ohne soziale Sanktionierung gesprochen werden kann. Auch anhand der Analyse ritueller Klammern (z. B. Begrüßungen und Verabschiedungen) zeigt sich, dass durch die vorgegebene Programmstruktur teilweise „der Eindruck eines wechselseitigen Aushandlungsprozesses [entsteht]“ (Krummheuer 2010: 160), der allerdings gerade durch seine Vorstrukturiertheit Limitierungen im Aktionsradius der Maschine verdeutlicht. Dadurch werden Deutungsrahmen für Situationen, Beziehungen und Äußerungen im Verlauf des Gesprächs brüchig und inkonsistent. Durch Nutzer*innen, die das Gerät aktiv an seine Grenzen bringen wollen und es herausfordern und testen, wird die technische Seite stärker betont als durch Nutzer*innen, die sich den Äußerungen des Geräts anpassen (vgl. Krummheuer 2010: 175).

Die Struktur des hybriden Austauschs bezeichnet Krummheuer (2010: 229) als „aufgebrochen“: Während im Zentrum eine „Abfolge von Aktivitäten“ steht, „in denen eine Handlung des Nutzers zu einer Veränderung des Zustands der Maschine [...] führt“, ist diese Aktivitätenfolge eingebettet in Umgebungshandlungen, die mit Arminen (2005) als „pre-“ und „post-monitoring of the action“ charakterisiert werden können, die aber von der Maschine nicht einbezogen werden (so auch Suchman 1987). Krummheuer weitet diese ‚Aufgebrochenheit‘ auch auf die sequenzielle Verknüpfung der Beiträge auf. Insbesondere nach Paarsequenzen, bei denen der zweite Beitrag von der Maschine produziert wurde, folge häufig kein dritter Zug, der eine Verstehenssicherung dokumentieren oder Korrekturen bzw. Reparaturen mit sich bringen würde (vgl. Krummheuer 2010: 229). Krummheuer entwickelt auf Basis dessen ein Modell für den hybriden Austausch, das zusammengefasst hervorhebt, dass „keine Interaktion zwischen System und einem isolierten Nutzer stattfindet“ (Krummheuer 2010: 231–232). Stattdessen werden

verschiedene Teilnahme- und Erwartungsstrukturen sowie die verschiedenen Vorgehensweisen zur Konstruktion von Sinn bei Mensch und Maschine abbildet.

2.2.4 Beteiligung maschineller Entitäten an der Praxis

Teil der praktisch-situativen Aushandlung des Charakters der Kommunikation ist neben dem Gesprächsbeteiligtenstatus auch die Zuschreibung von Handlungsträgerschaft („Agency“) – ein bereits mehrfach angerissener Themenbereich. Wurde die Diskussion um den Akteursstatus technischer Entitäten zunächst in Philosophie und Anthropologie geführt, trug sie in den 80er-Jahren wesentlich mit zur Begründung der Techniksoziologie bei. In dieser Diskussion treten deterministische Auffassungen von Technik mehr und mehr zurück, stattdessen werden Positionen dominant, die ein Verschwimmen klassischer ontologischer Grenzen von Technik und Sozialem annehmen (vgl. Rammert 1993: 10). Die ausschließliche Zuordnung von Handlungsfähigkeit zum Menschen scheint nicht mehr generell gültig zu sein (vgl. Rammert/Schulz-Schaeffer 2002a).

Die Akteur-Netzwerk-Theorie (ANT) schlägt in dieser Debatte vor, auch bei nicht-menschlichen Akteur*innen („actants“) eines Netzwerks Handlungsträgerschaft zu sehen (vgl. Latour 2005: 70–72) und Handlungen als Knotenpunkte des Netzwerks zu begreifen, in denen Handlungen von „Hybrid-Akteuren“ vollzogen werden – ein solcher werde z. B. beim Schusswaffengebrauch aus „Waffe und Schütze gebildet“ (Latour 2019: 428). Intention als Voraussetzung für Handlungen wird dabei abgelehnt: „An actant can literally be anything provided it is granted to be the source of an action“ (Latour 1996: 373). Krummheuer (2010: 44) führt in diesem Zusammenhang eine Arbeit Bruno Latours an, die die Durchführung von Experimenten zur Untersuchung von Hefe beschreibt – Latour konzeptualisiert dabei neben dem durchführenden Forscher auch die Hefe als Aktant (siehe Latour 2002). Die Ursprünge der ANT, die u. a. in den Laboratory Studies zu finden sind (siehe Lynch 1993; Knorr-Cetina 1995), bahnen zwar eine solche Auffassung an, indem sie Subjekten gleichermaßen wie Objekten, menschlichen und technischen Materialien eine Eingebundenheit in die vollzogenen Handlungen attestiert (vgl. Knorr-Cetina 1995). Auch die darauf aufbauenden Workplace Studies und das Forschungsfeld der Computer-Supported Cooperative Work (CSCW) verweisen auf diese praktische Dimension des Zusammenspiels von menschlicher und maschineller ‚Beteiligung‘ am Geschehen (vgl. Luff/Hindmarsh/Heath 2000; Hughes 2001; vom Lehn 2018). Die ANT jedoch hat sich hier radikaler positioniert und hat die Aufhebung einer ontologischen Scheidelinie zwischen sozialen und technischen Abläufen noch weitergetrieben. Diese sozialtheoretisch avancierte und

radikale Idee wurde für ihre Ambiguität und ihren unklaren konzeptuellen Status (siehe etwa Jones 1996: 296) ebenso kritisiert wie für die Ausblendung wesentlicher Differenzen in der Handlungsfähigkeit von Mensch und Maschine.⁴⁶ Erstens trage die Perspektive der ANT zu einer Diffusion des Akteursbegriffs bei, der die Analysen „intelligenter“ Technik erschwere (vgl. Braun-Thürmann 2002: 63), zweitens blende sie wesentliche Unterschiede im Handlungspotenzial von Menschen und technischen Akteuren aus.

Auch in der vorliegenden Arbeit soll mit einer weniger kategorischen Vorstellung der Verteilung von Agency gearbeitet werden. Dies soll v. a. dadurch gelingen, dass keine (letztlich auch in der ANT vertretene) Akteursperspektive eingenommen, sondern die Praxis in den Vordergrund gerückt wird.⁴⁷ Konsequenterweise muss in dieser auch die Verteilung von Agency als Gegenstand einer fortlaufenden Aushandlung betrachtet werden (vgl. Habscheid/Hector/Hrncal 2023; Waldecker/Hector/Hoffmann 2024). Dazu soll noch einmal auf die Konzeption Hirschauers (2004; 2016) eingegangen werden.⁴⁸ In dieser wird eine Auffassung vertreten, die die Praktiken in den Mittelpunkt rückt, an denen Körper und Dinge – auch technologische Artefakte – teilhaben können („partizipatorisch“), zu denen sie aber nicht immer („kontributorisch“) einen *aktiven* Beitrag leisten (Hirschauer 2004: 75). Dinge und insofern auch technische Artefakte wie Smart Speaker sind in dieser Sichtweise „materielle Partizipanden des Tuns“. Sie können nach Rammert/Schulz-Schaeffer (2002b: 56) als Teil eines „verteilten Handelns“ (man könnte mit Verweis auf spätere Debatten auch sagen: als Teil von Praxis) begriffen werden.⁴⁹ Technologische Artefakte können also am Vollzug der Praxis ‚beteiligt sein‘, in den Prozess der Ausführung oder als Ressource in die Ko-Operation und die entstehende Interaktion eingebunden werden und so auf verschiedene Weise *partizipieren*. Wie diese Partizipation gestaltet ist, ist Teil der fortlaufenden Praxis und wird unter allen Beteiligten ausgehandelt.

Denkbar wäre auch gewesen, zur Frage der Beteiligung von VUIs an Gesprächen (insbesondere in Mehrparteien-Konstellationen) mit Goffmans (1979) im *Footing*-Konzept integrierter Kategorisierung von *ratified participants*, *unaddressed recipients* und *bystanders* zu arbeiten – dieses Konzept wurde etwa von Matthewson (1992) produktiv auf pseudokommunikative Äußerungen von TV-Zuschauer*innen

⁴⁶ Für einen Überblick über wesentliche Strömungen dieser Debatte siehe Braun-Thürmann (2002: 35–65); siehe auch Rammert/Schulz-Schaeffer (2002b).

⁴⁷ Zu den praxeologischen Grundlagen siehe Kap. 2.1.

⁴⁸ Diese wurde bereits als praxeologische Grundlage eingeführt (siehe Kap. 2.1.1) und wird hier im Licht der Mensch-Maschine-Relation nur in aller Kürze aufgegriffen.

⁴⁹ Für eine Perspektivierung dieser Diskussion für die Analyse von Embodied Conversational Agents siehe Krummheuer (2010: 41–48).

angewendet, die sich an Fernsehfiguren gerichtet haben (siehe auch Alač et al. 2020: 28–29). Allerdings bringt ein solches Vorgehen auch Schwierigkeiten mit sich – v. a., weil VUIs eher nicht als *participants* im Sinne Goffmans konzeptualisiert werden können. Im Folgenden soll kurz diskutiert werden, wie weit ein solcher Ansatz reichen könnte und welche Probleme sich daraus ergäben.

Goffman (1979: 7–10) unterscheidet zwischen Teilnehmenden einer sozialen Begegnung, die als solche anerkannt sind und insofern z. B. in die rituelle Rahmung der Situation eingebunden werden (*ratified participants*). Diese können im Verlauf einer Unterhaltung direkt adressiert sein, können aber auch zuhörend am Gespräch teilnehmen ohne primär Angesprochene zu werden (*unaddressed recipients*). Davon abzugrenzen sind Individuen, die nicht an einer sozialen Begegnung teilnehmen, aber dennoch anwesend sind und die Unterhaltung zu Teilen mitverfolgen können (*bystanders*). Unternimmt man den Versuch, mit einem solchen Konzept dem Beteiligungsstatus einer Maschine habhaft zu werden, böte es sich zunächst an, die beteiligten Menschen einer Interaktion als *ratified participants* oder *unaddressed participants* zu verstehen, während der Smart Speaker überwiegend *bystander* zu sein scheint: Er ist in Hörweite, wird aber nur im Rahmen von „cross-play“ (Goffman 1979: 9) – d. h. einer über die Grenzen der sozialen Begegnung hinweg vollzogenen Konversation – angesprochen. Der Umstand, dass Smart Speaker sowohl von Nutzer*innen als auch von Wissenschaftler*innen sowie im journalistischen Diskurs fortgesetzt als „lauschende“ oder „zuhörende“ Geräte bezeichnet werden (vgl. Lau/Zimmerman/Schaub 2018; Pins/Alizadeh 2021; Lind 2021: 363–364; Waldecker/Volmar 2022: 171; Krotz 2022: 26), scheint diese Konzeptualisierung zu bestätigen.⁵⁰ In der Konzeptualisierung von *bystanders* bei Goffman (1979: 8) ist auch ein mögliches Abhören mitgedacht („[...] they may surreptitiously exploit the accessibility they find they have, thus qualifying as eavesdroppers, here not dissimilar to those who secretly listen in on conversations electronically“), was vor dem Hintergrund opaker Verwertung der aufgezeichneten Daten (vgl. Crawford/Joler 2018)⁵¹ auf den ersten Blick ebenfalls stimmig erscheint. Allerdings muss dazu gesagt werden, dass *eavesdropping* bei Goffman nicht als „Abhören“ im Sinne einer Überwachung zu verstehen ist, sondern vielmehr als fokussiertes, sinnverarbeitendes Zuhören, während die bei Smart Speakern gängigen Abhörpraktiken der Firmen viel schematischer ablaufen (siehe auch Hector 2025).

⁵⁰ Lind (2021: 362) wertet dazu korpuslinguistisch auf Basis des Deutschen Referenzkorpus (DeReKo) Belegstellen aus der medialen Berichterstattung für das Lexem ‚Alexa‘ aus und analysiert die ‚Alexa‘ zugeschriebenen semantischen Rollen. Mit Blick auf „hören“, „horchen“, „lauschen“, und „mithören“ entstehe durch die Verwendung dieser Begriffe eine „agensähnliche Rolle“ – Lind (2021: 364) verweist dabei auf den vergleichsweise hohen Intentionalitätsgrad des Zuhörens.

⁵¹ Siehe außerdem Kap. 3.2.6.

Bei genauerem Hinsehen erweist sich eine solche Konzeptionierung allerdings auch in anderer Hinsicht als ungeeignet. So scheint sie nicht nur wenig aufschlussreich zu sein, sondern auch problematisch im Hinblick auf die theoretischen Implikationen des Konzepts für die Untersuchung. Wenig Aufschluss über die Einbindung in die soziale Praxis und den Beteiligungsstatus gibt die Anwendung dieses Konzepts, weil sich bloß erneut bestätigen würde, was spätestens seit den Arbeiten von Suchman (2007) bekannt ist und sich auch in den hier vorgelegten Analysen noch einmal bestätigen wird: Maschinen sind keine interaktional gleichberechtigten Teilnehmenden an sozialen Interaktionen – die Folgen ihrer limitierten Wahrnehmungs- und Ausdrucksressourcen und grundlegend andere Prozessierung verunmöglichen dies. Mit einer Betrachtung nach Goffman käme man insofern zu dem Ergebnis, dass VUIs mal wie *ratified participants* behandelt werden, mal wie *bystanders* – zudem wäre ein schnelles Oszillieren zwischen diesen beiden Kategorien beobachtbar.⁵² Problematisch wäre die Anwendung des Konzepts in einer praxeologisch ausgerichteten Untersuchung, weil sie den VUIs Subjekthaftigkeit zuschreiben würde – und auch wenn Smart Speakern ein gewisser Grad an ‚Eigensinn‘ innezuwohnen scheint und Anwender*innen stellenweise formal Subjekthaftigkeit ‚zuschreiben‘, führt eine solche Unterscheidung konzeptionell zu einer Wiedereinführung der Dichotomie von Subjekt und Objekt, die ja gerade mit einer praxeologischen Perspektive und einer flachen Ontologie nach Schatzki (2016) suspendiert werden sollte (vgl. Schäfer 2020: 18). Insofern soll dieses Konzept, obwohl es möglicherweise naheliegend erscheint, nicht für die Analysen der vorliegenden Arbeit herangezogen werden – die zuvor eingeführten praxeologischen Perspektiven bilden hingegen einen konzeptuellen Rahmen, um sowohl sprachlich-interaktionale Praktiken in dyadischen VUI-Dialogen als auch soziale Praktiken, in denen VUIs ‚beteiligt‘ sind, analysieren zu können, ohne VUIs dabei essentialistisch von Menschen unterscheiden oder diesen implizit Subjekthaftigkeit zuschreiben zu müssen.

2.2.5 Zwischenfazit zum Verhältnis von Mensch und Maschine

Mit Krummheuer (2010) wurde die Auffassung eingeführt, dass es sich beim Austausch zwischen Menschen und Maschinen nicht um *soziale Interaktion* handelt, der auch in den folgenden Ausführungen gefolgt wird. Darauf aufbauend kann in der vorliegenden Arbeit der Austausch als *Dialog* zwischen ungleichen Entitäten konzeptualisiert werden. Die ‚Menschlichkeit‘ des maschinellen Gegenübers wird

⁵² Siehe dazu v. a. die Analysen in Kap. 6.2.

insofern situativ und abhängig von der Teilnehmendenkonstellation, vom sozio-kulturellen Kontext und den materiellen Gegebenheiten immer wieder neu ausgehandelt. Daran partizipieren maschinelle Entitäten im Vollzug der Praxis als „materielle Partizipanden des Tuns“ (Hirschauer 2004) – die Arbeit affirmiert insofern konzeptionell eine praxistheoretische Perspektive, grenzt sich dabei aber von der kurz andiskutierten Positionierung der ANT ab. Agency wird in diesem Sinne eine verhandelte und von den Beteiligten zugeschriebene, empirische Größe und es ist Aufgabe auch der vorliegenden Arbeit, die Aushandlungs- und Zuschreibungsprozesse näher zu analysieren (insoweit sie in sprachliche Praktiken eingebunden sind bzw. darin zum Ausdruck kommen). Dies gilt besonders vor dem Hintergrund des potenziellen Aufbaus von Erwartungen an die Umwelt, v. a. in Form konditioneller Relevanzen in den sequenziell verlaufenden Dialogen mit dem VUI. Auch Fälle von Initiierungen oder durch das VUI ausgelöste Expansionen eines VUI-Dialogs können entsprechende Erwartungen aufbauen. Gerade bezüglich dieser kann vermutet werden, dass sie als Form der eigenständigen Handlung behandelt werden. Den Geräten könnte insofern eine Handlungsträgerschaft und damit auch das Potenzial zur Beteiligung an anderen sozialen Praktiken wie Gesprächen zugeschrieben werden. Der Status der Gesprächsbeteiligung von VUIs ist insofern ebenfalls Gegenstand eines fortlaufenden Aushandlungsprozesses, der mit sprachlichen Praktiken hergestellt oder verworfen wird, wobei das *Gespräch* selbst als soziale Praktik diskutiert wurde. Dies soll im Rahmen der Analyse genauer beleuchtet werden.

2.3 Aneignung und Domestizierung

Neue Medientechnologien in der Wohnumgebung eines Haushalts treten in einen mehrschrittigen und teilweise reziproken Prozess mit den Haushaltsmitgliedern ein. In diesem Prozess machen die Nutzer*innen sich mit den Funktionen der Technologien vertraut, testen ihre Grenzen, weben sie in ihre Alltagspraktiken ein und *domestizieren* die Geräte auf diese Weise. Dieser Prozess vollzieht sich nicht zuletzt in den sprachlichen Praktiken der Nutzer*innen, insbesondere, wenn es um VUIs geht, die ohnehin durch die Produktion verbalsprachlicher und dialogischer Ausdrücke gesteuert werden. Unterschiedliche Strömungen im Schnittbereich von Kommunikations- und Medienwissenschaft, Soziologie und Sprachwissenschaft haben diesen Prozess genauer untersucht und Konzepte sowie Begriffe hervorgebracht, an die die vorliegende Arbeit anknüpfen will: Im Analyseteil werden sprachliche Praktiken herausgearbeitet und das Verhältnis von VUIs zu Gesprächen der Nutzer*innen untersucht. Beides ist Bestandteil der praktischen

Domestizierung bzw. Aneignung der Geräte durch ihre Anwender*innen und die Einhegung der Medientechnologien in die Praktiken der Haushaltsgemeinschaft.

Der Begriff der Domestizierung verweist bereits darauf, dass Medientechnologien in diesem Zusammenhang als etwas beschrieben werden, das von außen kommt, potenziell ‚wild‘ ist und gezähmt werden muss – und zugleich die Wohnumgebung mit der Außenwelt in teilweise opaken Prozessen verknüpft und zugänglich macht (vgl. Waldecker/Hector 2023). Im Anschluss an den Ethnologen Hermann Bausinger (1984) steht dabei die Widersprüchlichkeit des (häuslichen) Alltagslebens im Mittelpunkt: Dieses steht zwischen konservativer Veränderungsfeindlichkeit, die dem Individuum bisweilen zur Last wird, auf der einen Seite, sowie Rückzugsort und Vertrauensbasis auf der anderen Seite. Bausinger arbeitet ein in dieser Spannung stehendes Merkmal des Alltagslebens heraus, auf das die Begründer des Domestizierungsansatzes – im Rückgriff auf Anthony Giddens – mit „ontological security“ verweisen: das Vertrauen in die Konstanz der eigenen Identität, der Kontinuität des Lebens und der direkten Umgebung und insofern auch in die Beständigkeit des Haushalts (vgl. Giddens 2003[1984]: 50). In diese ontologische Sicherheit sind Medien als spezifische Form der Technologie eingewoben: Sie sind verknüpft mit alltäglichen Ritualen und auch wenn die ontologische Sicherheit zunächst durch die Medientechnologien hinterfragt wird (vgl. Silverstone/Hirsch 1992: 17), verlieren sie durch das Eingewoben-Werden ihre Bedrohlichkeit (vgl. Bausinger 1984: 349–350) – sie werden domestiziert. Mit diesem Prozess hat sich der Domestizierungsansatz in unterschiedlichen Projekten und Facetten beschäftigt (vgl. Hartmann 2013a: 39–52).

Zugleich sind im Bereich der Medienrezeptionsforschung ähnliche Fragestellungen untersucht worden. Dabei wurde, ebenfalls in den 90er-Jahren, zunehmend der Begriff der (kommunikativen) *Aneignung* in verschiedenen Disziplinen gebraucht – eine Übersicht über verschiedene Verwendungsweisen liefert Faber (2001: 27). Dieser Begriff betont dabei die aktive Rolle, die Nutzer*innen von Medientechnologien bei der Medienrezeption haben: Etwa Fernsehzuschauer*innen sind nicht einfach passive Konsument*innen, sondern vielmehr durch kognitive, emotionale, kommunikative und somit transformatorische Prozesse aktiv an der Rezeption beteiligt (vgl. Faber 2001: 28; siehe auch Ayaß 1993: 38–39). Diese Prozesse stellen eine natürliche Grenze des Einflusses von Medieninstitutionen und -konzernen dar: Partizipieren die Nutzer*innen nicht oder anders als vorhergesehen an den (von Institutionen bzw. Konzernen geplanten) Medienpraktiken, limitiert dies deren Macht, was für Fernsehen und andere Massenmedien (vgl. Keppler 1993) ebenso gilt wie für datengetriebene Medientechnologien wie Smart Speaker, wenngleich hier eine hintergründige Datenverarbeitung und -verwertung das Bild noch komplexer werden lässt (vgl. Habscheid et al. 2021; Waldecker 2022; Waldecker/Hector/Hoffmann 2024). Für Smart Speaker ist die

aktive Rolle der Nutzenden jedenfalls an der Oberfläche offensichtlich, weil der bidirektionale Charakter des Interfaces dialogische Eingaben erfordert. Gerade die Aneignungsforschung hat sich mit den verbalsprachlichen Äußerungen der Nutzer*innen beschäftigt, die Ausdruck und Vollzug einer Aneignung durch die Nutzer*innen sind: Durch sie binden Nutzer*innen die Medientechnologien und die von außen kommenden Inhalte an die Lebenswelt an und die Rezipient*innen erhalten „die Möglichkeit, [der] eigenen Beobachtung Relevanz zu verleihen“ (Faber 2001: 34). Eine Betrachtung durch die Linse der angewandten Sprachwissenschaft in Verbindung mit (mikro)soziologischen Ansätzen hat sich also als fruchtbar erwiesen, wie insbesondere die Ergebnisse des Projekts „Über Fernsehen sprechen: Die kommunikative Aneignung von Fernsehen in alltäglichen Kontexten“ zeigen (vgl. Holly/Püschel 1993a; Holly/Püschel/Bergmann 2001). Zuerst sollen also Ergebnisse aus dem Bereich der sprachwissenschaftlich-mikrosoziologischen Medienaneignungsforschung zusammengefasst werden, um sie für die Analyse von VUI-Dialogen anzuwenden. Im Anschluss wird das Domestizierungskonzept einschließlich einer Betrachtung seiner Anwendung und Kritik eingeführt.

2.3.1 Kommunikative Aneignung von Medien

Der Aneignungsbegriff hat sich in unterschiedlichen mit der Medienforschung befassten Disziplinen etabliert, um eine aktivere Rolle von ‚Rezipient*innen‘ bzw. ‚Nutzer*innen‘ unterschiedlicher Medientechnologien und -inhalte zu betonen (vgl. Hepp 2005: 67). Im deutschsprachigen Raum hat sich die Untersuchung dieser Medienaneignungsprozesse v. a. an der Schnittstelle von mediensoziologisch und sprachwissenschaftlich ausgerichteter Medienforschung entfaltet (vgl. Hartmann 2009b: 306) und im Kontext der Fernsehforschung, etwa im bereits erwähnten DFG-Projekt „Über Fernsehen sprechen: Die kommunikative Aneignung von Fernsehen in alltäglichen Kontexten“ in Chemnitz, Gießen und Trier, besondere Beachtung erfahren (siehe Holly/Püschel 1993a; Holly/Püschel/Bergmann 2001). In diesem Verständnis ist mit dem Begriff ‚kommunikative Aneignung‘ v. a. auf kommunikative Prozesse während der bzw. im direkten Anschluss an die Medienrezeption abgestellt worden. Ausgangspunkt war der Versuch eines Neuansatzes in der Medienrezeptionsforschung, der von reziproken Verhältnissen zwischen Zuschauer*innen und ‚Medientexten‘ ausgegangen ist (vgl. Holly 2001b: 17).

Für die Aneignungsforschung lassen sich unterschiedliche Forschungstraditionen bestimmen, auf denen Arbeiten aus diesem Bereich aufbauen. Die vorliegende Arbeit bezieht sich v. a. auf Arbeiten aus drei relevanten Strängen, die sich auch teilweise überlappen (im Überblick siehe Holly/Püschel 1993b: 8; Holly

2001b: 17–18): Erstens hat sie, wie die Domestizierungsforschung, Wurzeln in den British Cultural Studies (siehe etwa Hepp 1998), zweitens in einer wissenssoziologischen und damit verbundenen ethnomethodologisch-konversationsanalytischen Tradition (z. B. Keppler 1993; 1996; Ayaß 1993; 2001; Bergmann/Ulmer 1993) sowie drittens eine gesprächslinguistische Ausprägung (etwa Baldauf 2002). In den Cultural Studies wurde der Begriff in einem weiten Sinn gebraucht, um zu beschreiben, dass sich „Menschen eingebettet in kulturelle Kontexte in einem aktiven Prozess Medieninhalte zu Eigen machen“ (Hepp 2005: 69). In dieser weiten Verwendung sei, so Hepp weiter, jede Mediennutzung auch ein Medienaneignungsprozess. Im engeren Sinn hingegen, etwa bei Charlton (1993: 13) ausgeführt, wird mit der Aneignung auf eine spezifische Rezeptionsphase verwiesen (vgl. Winter 1995: 31; siehe auch Faber 2001: 27–28), in der Nutzer*innen eine produktive Auseinandersetzung bei und nach der Rezeption vollziehen. Die Aneignung kann, wie in der Arbeit von Charlton (1993: 13) aus einer sozialpsychologischen Perspektive, sogar den „Intradialog“ umfassen. Als *kommunikative* Aneignung werden hingegen spezifisch sprachliche bzw. interaktive Praktiken begriffen, wie dies etwa Püschel (1993) oder Hepp (1998) fokussieren. Gesprächspraktiken und die Rezeption von neuen (Medien)inhalten sind auch außerhalb der Fernsehsituation eng miteinander verwoben – durch sie verbinden die Nutzer*innen bzw. Rezipient*innen die Inhalte mit ihren Erfahrungen und Lebenswelten, kommen zur gemeinsamen Aushandlung von Bewertungen und wirken vergemeinschaftend und für Institutionen konstituierend (vgl. Ayaß 2012: 2–3), z. B. in Theaterpausengesprächen (vgl. Hrncal 2020: 288–289) oder bei der schulischen Diskussion von TV-Serien (vgl. Weiser-Zurmühlen 2021: 309). Püschel (1993: 115) unterscheidet am Beispiel des Fernsehens zwischen primären und sekundären Thematisierungen: Primäre Thematisierungen sind dabei das „Sprechen über Fernsehereignisse unmittelbar in der Rezeptionssituation“ (Püschel 1993: 115), während sekundäre Thematisierungen von der Rezeption unabhängiges, nachträgliches Sprechen darüber umfasst. Die Arbeit von Hepp (1998) zur Rolle von Gesprächen über das Fernsehen zeigt ebenfalls auf, dass die Aneignung sich zu einem erheblichen Teil konversationell vollzieht und Gespräche als „Katalysator“ (Hepp 1998: 209) für die Medienaneignung wirken: Sie helfen dabei, „gesehene Bezüge und Zusammenhänge zu intersubjektivieren, sich gegenseitig zu orientieren oder gemeinsame Lesarten von Fernsehsendungen auszuhandeln“ (Hepp 1998: 209–210).

Püschel (1993: 118) arbeitet außerdem heraus, dass sich in den primären Thematisierungen das Sprechen nicht als Gespräch im engeren Sinne darstellt, sondern eine weitaus größere Toleranz für Schweigephasen besteht und die Konzentration auf eine andere Aktivität – eben das Fernsehen – vordergründig sein kann; die Teilnehmenden befinden sich insofern in einem „open state of talk“ (Goffman 1981: 107), in dem Teilnehmende zwar berechtigt zu Äußerungen sind,

aber dazu und nicht einmal zu Reaktionen auf andere Äußerungen ‚sozial verpflichtet‘ sind (vgl. Goffman 1981: 134–135). Fernsehbegleitendes Sprechen ist dabei mit Verweis auf Henne/Rehbock (2001: 31) als „empraktisches Sprechen“ konzeptualisiert worden, das in die Ausübung einer anderen, außersprachlichen Handlung eingebunden ist bzw. dieser dient (vgl. Holly/Baldauf 2001: 45).⁵³ Konzeptionell an die Idee des ‚Empraktischen‘, an Goffman und an die genannte Arbeit von Püschel (1993) sowie an Studien von Brünner (1987) und Fiehler (1993) zum Sprechen als Nebenkommunikation anschließend benennt Baldauf (2001: 65–68) unterschiedliche Erscheinungsformen des fernsehbegleitenden Sprechens: „Freistehende Äußerungen“ sowie „Gesprächinseln“. Sie beschreibt anschließend deren kommunikative Strukturen, zu denen neben Interaktionen unter den Fernsehzuschauer*innen auch solche gehören, in denen die Zuschauer*innen die Fernsehfiguren zu adressieren scheinen.

Damit ist ein weiteres Phänomen angesprochen, das im Rahmen der Medienaneignungsforschung beleuchtet wurde: der Dialog zwischen Medieninhalten (z. B. Fernsehfiguren) und Rezipient*innen. Dazu wurde das ursprünglich von Horton/Wohl (1956) entwickelte Konzept der parasozialen Interaktion aufgegriffen und weiterentwickelt. Wie Ayaß (1993: 36) darstellt, hat das Konzept v. a. mit Bezug auf die „Schein-Reziprozität massenmedialer Inhalte“ Anwendung gefunden und ist dabei erneut v. a. in den Dienst solcher Argumentationen gestellt worden, die eine von passiven Medienrezipient*innen bzw. -konsument*innen ausgehen. Dies hat gerade nicht dazu beigetragen, Rezipient*innen als aktive Handelnde zu konzeptionalisieren – obwohl dies im ursprünglichen Konzept nach Horton/Wohl durchaus verankert gewesen sei. „Schein-Reziprozität“ hingegen, so argumentiert Ayaß (1993: 35–36), könne durch viele verschiedene Formen hergestellt werden, die aber alle als soziales Handeln verstanden werden müssten: „Sogar der einsame Zapper ‚handelt‘“ (Ayaß 1993: 36) – und ist insofern nicht in scheinbar (d. h. *para*-) soziale, sondern in soziale Vollzüge eingebettet (siehe auch Mikos 1996). Der Begriff *para*-sozial sei insofern irreführend – was hingegen eben nicht gegeben sei, sei eine ‚echte‘ Reziprozität und folglich ‚echte‘ Interaktivität, dafür aber kommunikative Praktiken, mit denen Schein-Reziprozität und insofern scheinbare Interaktivität hergestellt werde. Ayaß (1993: 36) schlägt konsequenterweise vor, mit dem Begriff der *Para*-Interaktion zu arbeiten, was mittlerweile trotz weiterer begrifflicher Debatten wohl als etabliert gelten kann (vgl. Luginbühl/Meer 2022).

⁵³ Der Begriff des „Empraktischen“ geht auf Karl Bühler zurück, der damit den „Einbau des Sprechens in anderes sinnvolles Verhalten“ (Bühler 1965[1934]: 52) bezeichnet; mit Verweis auf Kap. 2.1.3 unterbleibt eine ausführliche Diskussion an dieser Stelle.

Zudem führen Holly/Baldauf (2001: 44) den Begriff der Pseudokommunikation ein, den sie enggeführt für an Medienfiguren adressierte Äußerungen von Rezipient*innen beschreiben – Sprechhandlungen, in denen Rezipient*innen „reagieren, als könnten die Fernsehakteure sie hören und verstehen“ (Holly/Baldauf 2001: 59) – während Para-Interaktion das gesamte Spektrum von auf scheinbarer Reziprozität beruhenden Handlungen abdeckt und dabei häufig v. a. auf das Handeln der Medienfiguren fokussiert wurde (vgl. Luginbühl/Meer 2022). Pseudokommunikation soll hier also als eine mögliche Ausprägung von Para-Interaktion verstanden werden. Erkenntnisse aus dem bereits erwähnten Projekt zum fernsehbegleitenden Sprechen zeigen, dass pseudokommunikative Äußerungen nicht als „der Versuch, mit Fernsehakteuren zu kommunizieren“ (Baldauf 2001: 72) missverstanden werden dürfen. Stattdessen hat diese Form der ‚Beteiligung‘ im fernsehbegleitenden Sprechen vielfältige Funktionen, etwa solche „emotionaler Aneignung“ (Faber 2001: 36), aber auch „implizite Stellungnahme[n]“ (Baldauf 2001: 72) und Aspekte der Verarbeitung, der Gruppendynamik und der „positiven Selbstdarstellung“ (Klemm 2001: 107) spielen eine wichtige Rolle. Genau dieses Spannungsfeld will die Analyse von ohnehin gesprächsähnlichen VUI-Dialogen sowie ihre Einbindung in Gesprächspraktiken genauer beleuchten.

VUI-Dialoge können Merkmale von Para-Interaktion bzw. Pseudokommunikation aufweisen. Sowohl die VUIs wie auch die Anwender*innen produzieren durch die Stimmein- und ausgaben, in denen sie sich wechselseitig (implizit und explizit) adressieren. Allerdings besteht dabei auch ein entscheidender Unterschied im Gegensatz zu den bisherigen Begriffsbedingungen: Das VUI ermöglicht tatsächlich einen Rückkanal und weist eine Form von Dialogizität auf. Damit ist der Austausch zur Bedienung der Geräte nicht nur *scheinbar* reziprok, sondern tatsächlich von einer gewissen Wechselseitigkeit geprägt – wenngleich nach wie vor nicht von Interaktion im engeren Sinne gesprochen werden kann.⁵⁴ Bei VUIs ist die Para-Interaktion also von einer ganz anders gelagerten Operativität, bei Radio, Fernsehen und anderen massenmedialen Inhalten besteht dieser Kanal hingegen nicht. Als Medienfigur tritt außerdem spezifisch die „Persona“ des VUI durch die synthetisierte Stimme auf. Teilweise produzieren die Nutzer*innen von VUIs zwar auch Äußerungen, die zwar formal an den Smart Speaker bzw. das VUI adressiert sind, die Geräte wurden zuvor allerdings gar nicht aktiviert und können die Eingaben insofern gar nicht verarbeiten. Diese lateralen Adressierungen des VUI, die Funktionen außerhalb des VUI-Dialogs übernehmen, könnten möglicherweise als pseudokommunikative Praktiken verstanden werden. Das Konzept kann aber – aufgrund des synthetischen und funktionalen

⁵⁴ Siehe Kap. 2.2.1.

Charakters des Gegenübers und auch, weil durch eine erneute Aktivierung potenziell ein Kanal hergestellt werden könnte – nicht ‚einfach‘ übertragen werden. Vielmehr sollen diese sprachlichen Praktiken einer genauen Analyse unterzogen werden, u. a. um sie als Teil des Domestizierungsprozesses verständlich zu machen. Auf den theoretischen Hintergrund zu diesem Bereich wird im folgenden Abschnitt eingegangen.

2.3.2 Zum Domestizierungsansatz

Die Metapher der Domestizierung betont v. a. die Beziehungsarbeit zwischen dem „Öffentlichen“, den für die (noch näher zu erörternde) „ontological security“ (Giddens 2003[1984]: 50) bedrohlichen, zu zähmenden und wilden Medien auf der einen Seite und dem „Privaten“, der Wohnumgebung und der sozialen Konfiguration des Haushalts auf der anderen Seite. Gerade in den Texten der frühen britischen Domestizierungsforschung wird dies deutlich: Es ist der Anspruch der Autoren, ein Modell zu entwickeln „aimed towards understanding the nature of the relationship between private households and public worlds and the role of communication and information technologies in that relationship“ (Silverstone/Hirsch/Morley 1992: 13). Der Domestizierungsbegriff wie er in einer Arbeitsdefinition bei der Kommunikationswissenschaftlerin Maren Hartmann (2013a: 146) ausbuchstabiert ist, fokussiert stärker Prozesse der Aneignung von Medientechnologien *innerhalb* der Haushaltsgemeinschaft:

Die Domestizierung von Medien und Technologien umfasst den dynamischen und reziproken Prozess des Medien-Häuslich-und-Zu-eigen-Machens. In diesem Prozess werden den technischen Konsum-Objekten und dazugehörigen Inhalten anhand sozialer Aushandlungen Bedeutungen zugewiesen (u. a. als Teil der sogenannten moralischen Ökonomie). Kernkontext der Aushandlungen ist der Alltag der Mediennutzer, der Routinen, räumliche Anordnungen und ähnliches beinhaltet. Medien werden dabei als besondere technische Objekte definiert, da sie Objekt und Inhalt gleichermaßen umfassen. (Hartmann 2013: 146)

Hier scheint die Vorstellung von Medien als Objekte und Inhalte, die in die häusliche Umgebung und seine sozialen Strukturen eingegliedert und dort reziprok „zu eigen gemacht“ werden, zu dominieren. Gerade datenintensive Medientechnologien wie Smart Speaker tragen allerdings zugleich Inhalte aus der privaten Wohnumgebung nach außen (vgl. Waldecker/Hector 2023: 6). Dies macht eine erneute Öffnung des Blicks auch für die Dynamik zwischen Haushalt und Öffentlichkeit relevant und fordert dichotome Grenzziehungen zunehmend heraus (vgl. Waldecker/Hector 2023: 6).

Hartmann (2020: 48) unterscheidet zwischen einer britischen und einer norwegischen bzw. „nordischen“ Strömung des Domestizierungsansatzes, die sich unabhängig voneinander entwickelt haben und sich in einigen Punkten voneinander unterscheiden. Die Entwicklung des britischen Ansatzes war eingebettet in Arbeiten am von der Wissenschaftsförderung Großbritanniens geförderten „Programme on Information and Communication Technologies“ (PICT) und insbesondere in das darin integrierte und von Roger Silverstone geleitete Projekt „The Household Uses of Information and Communication Technologies“ (HICT). Das inhaltliche Zentrum des Ansatzes bildet die moralische Ökonomiestruktur des Haushalts (*moral economy of the household*), die als Ausgangspunkt der erwähnten ‚ontologischen Sicherheit‘ steht und in die Medientechnologien eingebunden und in ihr von den Haushaltsmitgliedern domestiziert werden und damit wechselseitig eine aktive Rolle in der Gestaltung von Medientechnologien und -inhalten einnehmen (vgl. Silverstone/Hirsch/Morley 1992: 17). Von diesem Zentrum aus entwickelt der Ansatz eine Reihe von Konzepten, sowohl zur Aneignung der Technologien wie auch zur Vermittlung dieser mit privaten und öffentlichen Sphären, die sich als heuristisch sehr produktiv herausgestellt haben (vgl. Berker et al. 2006: 4; Hartmann 2009a).

Ebenfalls in den 1990er-Jahren gewann an der norwegischen Universität Trondheim ein in der Technologie- und Wissenschaftsforschung gründender Forschungsstrang an Kontur, dessen erste Publikationen teilweise noch vor den britischen Studien veröffentlicht wurden und welcher die im erwähnten Projektzusammenhang des HICT entwickelten Konzepte auch nicht einbezog. Die Untersuchungen aus Norwegen sind weiter gefasst: Sie beschränken sich nicht auf den Haushalt und auch nicht auf Medientechnologien, außerdem betrachten sie die Zähmung stärker als gesellschaftlichen denn als individuellen Prozess (vgl. Sørensen 1994: 42; siehe auch Hartmann 2020: 48–49). Sie ähneln dennoch den britischen Domestizierungsstudien, erstens terminologisch, weil sie auch den Begriff der Domestizierung nutzen und als Technologieaneignung prozesshaft beschreiben (vgl. Sørensen 1994), andererseits, weil die norwegischen Studien ebenfalls empirisch – v. a. ethnografisch – arbeiten, den Fokus auf die Praktiken der Nutzer*innen legen und mit diesem Instrumentarium technische Aneignungsprozesse zu fassen vermögen. Hartmann (2020: 49) zufolge war diese Orientierung auf die Nutzer*innen sogar der Initialzündung für den weiteren Ausbau dieses Forschungsstrangs. Die von Hartmann (2020) als ‚nordisch‘ bezeichnete Domestizierungsschule, zu der etwa auch eine finnische Strömung aus dem Bereich der *design research* gerechnet werden kann,⁵⁵ zeichnet

55 Für einen Überblick über Arbeiten in diesem Feld und eine Diskussion zur Klassifizierung der Ansätze als „nordisch“ siehe Hartmann (2020: 50–52).

sich neben der Nutzer*innen-Zentrierung durch eine größere Offenheit aus, sowohl methodologisch als auch im Hinblick auf die untersuchten Technologien. So wurden von Aune (2007) z. B. Technologien der privaten Stromversorgung in den Blick genommen, zu diesen Studien zählen aber auch Untersuchungen zu Autos (Sørensen 1994), Glasfaserkabeln (Löfström 2007) oder E-Government-Anwendungen (Østergaard Madsen/Kræmmergaard 2015). Daran wird deutlich: Der norwegische bzw. nordische Ansatz ist insbesondere für eine Anwendung außerhalb des ‚klassischen‘ Haushalts offener.

Zunehmend ist die Frage gestellt worden, was einen Haushalt charakterisiert: Das „Domestische“ kann nicht nur durch das gemeinsame Bewohnen einer physischen Gebäudeeinheit bedingt sein, sondern der domestische Raum ist auch in den frühen Texten zur Domestisierungsforschung als Raum für ontologische Sicherheit im Sinne Giddens‘ und Vertrauen betrachtet worden, der sich – gerade durch die zunehmende Mobilität der Medientechnologien – von den physischen Wohneinheiten lösen kann (vgl. Silverstone 2006: 234); die Wohneinheit ist dann nur noch Ausgangspunkt und Ressourcenlager für die von dort ausgehende Mobilität (vgl. Bakardjieva 2006: 68–70). Zugleich kann die domestische Sphäre durch Medientechnologien auch in ein Café, in den Stadtpark oder in einen Jugend-Treff verlagert werden. Es lösen sich so naive Dichotomien zwischen Inner- und Außerhäuslichkeit, Privatheit und Öffentlichkeit auf, deren Dimensionsgrenzen nicht an Wohnungsmauern, sondern vielmehr an Praktiken der Bewohner*innen gekoppelt sind, die in diesen hausen und dynamische Beziehungen zwischen diesen Domänen konstituieren. Damit ist auf das Alltagsleben der Nutzer*innen verwiesen, das mit Bakardjieva (2006: 70) die Fokussierung der frühen Arbeiten auf den *Medienkonsum* aufricht und damit unterschiedliche Analyseperspektiven zulässt: Ihr zufolge ist der Begriff des Konsums zu sehr zugeschnitten auf eine Betrachtung aus Design- und Ökonomie-Perspektiven. Demgegenüber erlaubt der Alltagsbegriff, den sie in der Phänomenologie bei Schütz/Luckmann (2003) verwurzelt, eine umfassendere Betrachtung der Realitäten und darin vollzogenen Praktiken der Mediennutzung und -anwendung.⁵⁶

Diese Weiterentwicklungen haben in der britischen Entwicklungslinie erst in den frühen 2000er-Jahren stattgefunden, während sie in den norwegischen Texten schon von Beginn an impliziert waren (vgl. Sørensen 1994). Konsequenterweise herrscht darin auch von Anfang an eine größere methodologische Flexibilität (vgl. Hartmann 2020: 52). Obschon bereits die britischen Ausgangstexte keine methodologische Starre postulieren, sind gerade die Anwendungen des Domestisierungsansatzes um die 2000er von großen methodologischen Unterschieden (in-

⁵⁶ Siehe auch Kap. 4.

nerhalb des qualitativen Spektrums) und von Anpassungen ans Feld geprägt, so dass hier nur schwer ein Methoden-Nexus zu bestimmen ist. Die Studien aus Norwegen und Finnland tendieren zudem dazu, die den Technologien zugrundeliegenden Infrastrukturen stärker zu berücksichtigen und sich eher der Verknüpfung von Erfindung, Design, Umsetzung und Nutzung bestimmter Technologien zu widmen.

Die Grundzüge des Domestizierungsansatzes (d. h. die Dimensionen der Domestizierung, die „doppelte“ bzw. „dreifache“ Artikulation der Medien und die Idee der moralischen Ökonomie des Haushalts) können helfen, sprachliche Praktiken einzuordnen, soziotechnische Ordnungen an der sprachlichen Oberfläche verständlich zu machen und so Beteiligungen, Bewertungen und Statuszuschreibungen zu Smart Speakern, auch innerhalb der Haushaltsgemeinschaft, zu verstehen. Das Modell stellte sich im Rahmen einer Reflexion als „to be particularly useful for the empirical in-depth analysis of daily media consumption“ (Berker et al. 2006: 4) heraus. Bei der Untersuchung von Smart Speakern fallen der physisch verfasste Haushalt, d. h. die ummauerte Wohnumgebung mit ihren Bewohner*innen, mit den dort vollzogenen Alltagspraktiken und den von den gemeinsamen Bewohner*innen geteilten Wertvorstellungen und Positionen, auf den ersten Blick noch zusammen – Smart Speaker sind nicht mobil, sie werden nicht mitgenommen und in wechselnden physischen Zusammenhängen eingesetzt, sondern sind klar dem ‚zu Hause‘ in seiner physischen und sozialen Verfasstheit zugeordnet.⁵⁷ Erst eine Analyse mit Hilfe eines auf diese Domäne zugeschnittenen Modells kann kontrastiv zeigen, wo Außerhäuslichkeit und verschwimmende Dichotomien auch für die Nutzer*innen eine Rolle spielen und die „continuous negotiation“ von „inside und outside“ (Silverstone 2006: 233) sichtbar machen. Zweifelsohne ist diese strukturell nicht nur bei der über vernetzte Cloud-Systeme prozessierten Sprachverarbeitung relevant, sondern auch z. B. bei mobiler Smart-Home-Steuerung, der Anwendung von Geodaten für Wettervorhersagen oder der Verbindung von Streaming-Diensten mit den Smart Speakern. Doch erst durch den Gebrauch der auf das ‚zu Hause‘ zugeschnittenen Konzepte können die Aspekte besonders gut herausgestellt werden, in denen die Nutzer*innen selbst in ihren Domestizierungspraktiken die Grenzverschiebungen relevant machen und sie in ihren alltagspraktischen Umgang mit den Geräten und in den Austausch über die Geräte und damit verbundene Anwendungen einfließen lassen. Es können so wirtschaftliche und gesellschaftliche Prozesse verstanden werden – „through the messi-

⁵⁷ Damit soll nicht ausgeschlossen werden, dass auch in (teil-)öffentlichen Umgebungen und z. B. in beruflichen Kontexten Smart Speaker zum Einsatz kommen können; diese sind aber nicht Bestandteil der vorliegenden Untersuchung.

ness of domestic practices in everyday life“ (Livingstone 2007: 17). Dies gilt umso mehr, wenn diese Praktiken als Teil sozialer, räumlicher und somit interaktiver Ordnungen mit einer praxeologischen Brille an der sprachlichen Oberfläche beobachtet und verstanden werden sollen und muss insofern der erste Schritt als Bestandsaufnahme sein. Anschließend können die Befunde als Prozesse mit ökonomischer und gesellschaftlicher Tragweite ausgewertet werden – Letzteres ist jedoch nicht im Rahmen dieser Arbeit zu leisten.

Wesentlich zur Popularität des Domestizierungsansatzes beigetragen hat die Instruktivität der prozesshaft formulierten Dimensionen der Domestizierung. Diese wurden erstmals von Morley/Silverstone (1990) formuliert, im Folgenden aber immer wieder angepasst und lange nicht einheitlich gefasst (vgl. etwa Silverstone/Hirsch 1992; Silverstone/Haddon 1996; Berker et al. 2006; Hynes/Romes 2006; Sørensen 2006). Diese sollen nachfolgend kurz charakterisiert werden. Die Dimensionen sind allerdings nicht die einzige und möglicherweise nicht einmal die wichtigste Stütze des Domestizierungsansatzes: Vielmehr funktionieren sie nur flankiert durch die Vorstellung von der „doppelten“ bzw. „dreifachen Artikulation der Medien“ sowie durch das, was oben schon als Sphäre der ‚ontologischen Sicherheit‘ beschrieben wurde und konzeptuell bei Silverstone/Hirsch/Morley (1992) als „moral economy of the household“ gefasst ist. Diese beiden Konzept-Säulen charakterisieren die beiden Einheiten, die durch den Domestizierungsprozess miteinander vermittelt werden: die Verfasstheit des Haushalts auf der einen und die Verfasstheit der Medientechnologien auf der anderen Seite. Diese Charakterisierungen sind für die heuristische Anwendung des Ansatzes zentral, weil auch diese sich an dem Anspruch messen lassen müssen, *in situ* beobachtet zu werden.

2.3.2.1 Dimensionen der Domestizierung

Auch wenn sich die verschiedenen Zuschnitte der Dimensionen gerade in den frühen Arbeiten unterscheiden, lassen sich nach Hartmann (2013a: 21–24) vier Bereiche ausmachen: (a) Kommodifizierung bzw. Imagination, (b) Aneignung – mit den Unterkategorien Objektifizierung und Eingliederung – und (c) Umwandlung. Sie sind allerdings nicht in ihrer zunächst naheliegenden Temporalität zu lesen,⁵⁸ sondern vollziehen sich parallel und beeinflussen sich reziprok. Gleichwohl „ist eine gewisse Linearität nicht zu leugnen“ (Hartmann 2009a: 307). Die drei genannten Dimensionen greifen ineinander und zeichnen ein ganzheitliches Bild des Domestizierungsprozesses.

⁵⁸ In der ursprünglichen Konzeptualisierung der Dimensionen – noch als Phasen – waren diese durchaus linear zu verstehen (vgl. Silverstone/Hirsch/Morley 1992: 20–21), wovon jedoch später als Reaktion auf deutliche Kritik abgerückt wurde (vgl. etwa Silverstone/Haddon 1996).

Zur Dimension der Kommodifizierung (*commodification*) zählen alle Aspekte, die noch vor dem Kauf der untersuchten Medientechnologien liegen. Sie können insofern bei der Betrachtung der Domestizierungspraktiken nur mittelbar beobachtet werden: Überlegungen der Hersteller zu Design, Vermarktung und Vertrieb spielen hier eine besondere Rolle. Der für diese Dimension wesentliche Aspekt der Imagination bezieht sich primär auf die Vorstellungen der Designer*innen, Hersteller*innen und Produktmanager*innen von ihren Produkten und deren Nutzungsszenarien bei den Anwender*innen sowie die Platzierung auf dem Markt (vgl. Silverstone/Haddon 1996: 44–46). In der vorliegenden Studie wird diese Seite der Imagination nur gestreift, weil keine Daten dazu erhoben werden. Zugriff besteht allerdings auf Imaginationen der Anwender*innen, wenn diese mit den ersten visuellen, haptischen und (em)praktischen Eindrücken von den Geräten umgehen und diese in der Interaktion verbalisieren: Erwartungen und Einstellungen, die beim Entpacken und der ersten Inbetriebnahme geäußert werden und praktisch in den ersten Anwendungsversuchen zum Ausdruck kommen, werden an der sprachlichen Oberfläche sichtbar und können entsprechend analysiert werden. So entsteht retrospektiv ein Bild von den Imaginationen im Kontrast zu den tatsächlichen realweltlichen Erfahrungen mit den Geräten und ihren Einbindungen in die jeweiligen technologischen Gefüge der Haushalte.

In der zweiten Dimension, der Aneignung (*appropriation*)⁵⁹ wird unterschieden zwischen der Objektifizierung (*objectification*) und der Eingliederung oder ‚Veralltäglichung‘ (*incorporation*). Während bei der Objektifizierung der räumliche Aspekt im Vordergrund steht, ist in der Eingliederung die Zeitlichkeit, die „injection of media technological practices into the temporal patterns of domestic life“ (Silverstone 2006: 235), die entscheidende Größe. Silverstone (2006: 234–235) bezeichnet diese zusammen als Strategien der Domestizierung – Kommodifizierung und Umwandlung begreift er demgegenüber als Rahmung der Aneignung. Hartmann (2013a: 22–23) bereitet die Strategien später als zwei ‚Unterdimensionen‘ der Aneignung auf, wie sie auch hier verstanden werden sollen.⁶⁰ Zur Objektifizierung zählen Praktiken der Platzierung (vgl. Röser 2007: 21) und insofern die

59 Die bei Hynes/Romes (2006) dargestellte Terminologie, in der die Dimension der *appropriation* den Umgang mit Imaginativen zu fassen versucht, zeigt einerseits noch einmal deutlich die Verschränkung der verschiedenen Dimensionen miteinander, ist andererseits aber auch in anderen Arbeiten üblich, in denen die Kommodifizierung mit der Imagination deutlich von der Aneignung abgegrenzt wird; Hynes/Romes (2006: 127) beziehen sich hier insbesondere auf die erste Darlegung der Dimensionen von Silverstone/Hirsch/Morley (1992), in der die Dimension Kommodifizierung noch nicht enthalten war und v. a. in der Aneignung impliziert wurde.

60 Siehe auch Hartmann (2009b: 306).

Einbettung in die materielle, räumliche Umgebung des Haushalts. Die Dimension trägt wesentlich dem Umstand Rechnung, dass die Aneignung von Medientechnologien nicht losgelöst vom sie umgebenden physischen Haushalt stattfindet – und dass Objekte immer in bereits vorkonfigurierten räumlichen Umgebungen erscheinen (Silverstone/Hirsch/Morley 1992: 20). Dies ist keineswegs nur eine Frage der Ästhetik der Inneneinrichtung: Interaktive Verfahren und Ordnungen sind wechselseitig von der räumlichen Umgebung, in der sie stattfinden, geprägt und prägen die räumliche Umgebung (vgl. Schmitt/Hausendorf 2016); räumliche Arrangements können als interaktive Ressource fungieren und ein Faktor bei der inkrementellen Produktion von Teilnehmer*innen-Strukturen sein (vgl. Hausendorf/Mondada/Schmitt 2012). Sie geben insofern Aufschluss über die bereits weiter oben adressierte Frage der Handlungsträgerschaft von Technologien⁶¹ und helfen, Konfigurationen von Gesprächs,beteiligung‘ zu verstehen. Die Dimension der Objektifizierung vollzieht sich insofern nicht nur unter dem Aspekt der Platzierung und der materialen Arrangements, sondern auch in ihrer Anwendung und Nutzung *in* diesen Arrangements (vgl. Silverstone/Hirsch/Morley 1992: 20). Unter Einbezug von Mimik und Gestik ergeben sich analytische Räume, in denen der Prozess der Domestizierung mit Blick auf den Gesprächsbeteiligtenstatus der Geräte beschrieben und verstanden werden kann. Gerade vor dem Hintergrund der primär verbalen Steuerung dieser Technologien lassen sich so über die Analyse von Rederechtsaushandlungen (*conversational floor*) und von Adressierung von Evaluations- und Aufzeigeleistungen Erkenntnisse zu diesem Prozess gewinnen.

Dies überlappt sich mit dem Prozess der Eingliederung (*incorporation*), in dem das Gerät nicht nur in die räumlichen Arrangements im Haushalt, sondern auch in dessen etablierten Ablauf sozialer Praktiken, d. h. dessen Routinen (vgl. Schüttpelz/Meyer 2017: 158),⁶² eingebunden wird, die ja ebenfalls räumliche Arrangements prägen und von diesen geprägt werden (vgl. Silverstone/Hirsch/Morley 1992: 21). Die Medientechnologien werden dabei mit einer funktionalen Perspektive betrachtet:

Technologies are functional. They may be bought with other features in mind and indeed serve other cultural purposes in appropriation. They may indeed become functional in ways somewhat removed from the intentions of designers or marketers. They may have functions. Functions may change or disappear altogether [...]. To become functional a technology has to find a place within the moral economy of the household, specifically in terms of its incorporation into the routines of daily life. (Silverstone/Hirsch/Morley 1992: 21)

⁶¹ Siehe Kap. 2.2.4.

⁶² Siehe Kap. 2.1.1 zu Praxis und Ko-Operation.

Zu den Untersuchungsgegenständen zählen insofern einerseits die emergenten Routinen in der Anwendung des Geräts selbst, die durch neue sprachliche Praktiken sichtbar und verstehbar werden. Sprachliche Formen der tatsächlich beobachtbaren Kommandos an den Smart Speaker sind hier, gerade in ihrem Zusammenhang zu bestimmten Nutzungskontexten und -szenarien, von Interesse, ebenso wie die sich möglicherweise verändernde Gestaltung dieser im zeitlichen Verlauf oder vor Publikum. Analysiert werden vor diesem Hintergrund auch Reparatur-, Lern- und Anpassungsprozesse, die auch über die Annahmen der Nutzer*innen über die Geräte Aufschluss geben. Andererseits zählen dazu auch Fragen der Einbettung von Smart-Speaker-Anwendungen in laufende soziale Interaktionen und der Gesprächsorganisation, d. h. die Eingliederung in sprachlich hervorgebrachte interaktionale Ordnungen und damit verbundene Routinen, die diese herstellen. Diese können auch Bestandteil sozialer Ordnungen innerhalb des Haushalts sein, in denen z. B. ‚Nutzungshoheit‘, Besitzverhältnisse und situative Aushandlungen über die Gerätenutzung wie Porcheron et al. (2018) sie nachvollziehen, zum Ausdruck kommen.

Als Umwandlung (*conversion*) wird schließlich die Dimension der Domestizierung bezeichnet, die – in der ursprünglichen Formulierung der ‚Gründerväter‘ des Ansatzes – abhebt auf „the relationship between the household and the outside world“ (Silverstone/Hirsch/Morley 1992: 22). Wie bereits ausgeführt, ist diese Dichotomie zwischen Haushalt und „outside world“ überkommen und kann so nicht aufrecht erhalten werden, wie – als eine der stärksten Kritikerinnen dieser Konzeptualisierung – Bakardjieva (2005; 2006) ausführt. Silverstone (2006: 234) nennt in einer Konzeptionsreflexion verschiedene Aspekte: Rückanbindung an die Designer*innen, Vorführung, Darstellung und Bewertung im sozialen Umfeld, ggf. auch vor Besucher*innen, den daran anschließenden auch öffentlichen Diskurs und kulturelle Widerstände und Veränderungen. Hartmann (2013a: 23) fasst die Aspekte als „das Nach-außen-Tragen der Nutzung“ zusammen, und auch wenn erneut mit der Erwähnung eines „Außen“ ein unklares „Innen“ konstruiert wird, ist doch die Formulierung eingängig: Es sind die Momente, die als Teil des Domestizierungsprozesses auftreten, in denen sich die Haushaltsmitglieder nicht nur miteinander, in der eigenen Wohnumgebung und (vermeintlich) exklusiv mit den Medientechnologien und den dahinterstehenden Infrastrukturen befassen, sondern in denen sie diese mit anderen teilen, vorführen, bewerten, und somit zur Bildung eines gesellschaftlichen Diskurses beitragen sowie wechselseitig zur Überprüfung eigener Bewertungen gelangen. „Consumption is never a private matter“, schreibt dazu Silverstone (2006: 234) und auch wenn die Fokussierung auf den Medienkonsum abgelegt wurde, lässt sich auch über die Anwendung und alltagspraktische Nutzung von Medientechnologien sagen, dass diese keine pri-

vate Angelegenheit ist – dies gilt bei Smart Speakern noch viel mehr vor dem Hintergrund des Spannungsverhältnisses von privaten und öffentlichen Domänen.⁶³

2.3.2.2 Dreifache Artikulation: Medien als Objekte, Rituale und Inhalte

Die „doppelte Artikulation“ der Medientechnologien, der zweite vorzustellende Grundpfeiler des Domestizierungsansatzes neben den beschriebenen Dimensionen, ist seit den frühen Publikationen der britischen Domestizierungsforschung in diversen Arbeiten präsent (vgl. etwa Morley/Silverstone 1990: 36). Die doppelte Artikulation ist in Anlehnung an das sprachwissenschaftliche bzw. semiotische Verständnis von doppelter Artikulation zu begreifen: als Unterscheidung zwischen bedeutungs**tragenden** und bedeutungs**unterscheidenden** Einheiten. Mit dem von Silverstone/Hirsch/Morley (1991) hergeleiteten Verweis auf den Linguisten André Martinet und dessen Werk zur „double articulation linguistique“ (1949) wird der Begriff zugänglich: In André Martinets strukturfunktionalistischer Semiotik liegt die Bedeutung eines Wortes eben sowohl in den einzelnen Phonemen als signifikative Einheiten (erste Artikulation), wie auch in den durch sie gebildeten distinktiven Einheiten (zweite Artikulation). Dem entspricht – auch wenn diese Übertragung anfällig ist für Kritik (vgl. Hartmann 2006: 85) – das Gerät als Objekt auf der Ebene der ‚ersten Artikulation‘ und die vermittelten Inhalte als ‚zweite Artikulation‘ (vgl. Silverstone/Haddon 1996: 62).⁶⁴ Die beiden Artikulationen sind in der Semiotik Martinets zwei strukturell unterschiedliche, aber voneinander abhängige Ebenen des Sprachgebrauchs. Diese Interdependenz zwischen medientechnologischem Objekt und den vermittelten Inhalten, die zwar strukturell ver-

⁶³ Siehe Kap. 3.2.6.

⁶⁴ Bei Martinet (1963[1960]: 21–24) ist die *erste* Artikulation (bzw. *Gliederung*) die signifikative und die *zweite* die distinktive. Es überrascht, dass schon bei Morley/Silverstone (1990: 36) der ersten Artikulation, der morphematischen Ebene, die Medientechnologien als Objekt und der zweiten, phonematischen Ebene die Inhalte zugeordnet wurden. Zwar werden die Objekte als Bedeutungsträger verstanden, was ihre Zuordnung zur morphematischen Ebene plausibel macht, doch der Vergleich der Inhalte mit der semiotisch bedeutungsunterscheidenden Ebene ist vor dem Hintergrund der Arbitrarität der bedeutungsunterscheidenden Elemente und ihrer (prinzipiell beliebigen) Kombinierbarkeit nicht nachvollziehbar und legt den Schluss nahe, dass die Zuordnung vielmehr andersherum hätte erfolgen müssen: Die Objekte sind bedeutungs**unterscheidend** und bringen Inhalte auditiv oder visuell zum Ausdruck, erst die Inhalte selbst sind bedeutungs**tragend** (so auch bei Röser 2007: 22). Möglicherweise liegt der Ursprung bei Silverstone (1990: 178) selbst: Er schreibt „it [u. a. Fernsehen, T.H.] is both meaningful in itself and it is the transmitter of meanings“. Diese Passage könnte so interpretiert werden, dass die Objekte Bedeutungsträger sind, während die Bedeutung der Inhalte in diesen selbst liegt. Dieses Verständnis scheint zeichentheoretisch gesehen schief, setzt sich aber explizit in den weiteren Arbeiten zum Modell (etwa Hartmann 2006: 87) fort.

schiedene Ebenen der Mediennutzung darstellen, aber gleichzeitig nicht ohne einander zu fassen sind, stellt den Kernaspekt der Anwendung des semiotischen Konzepts im Domestizierungskontext dar (vgl. Hartmann 2006: 85–86; 2009a: 309).

Die erste Artikulation der Medien, verstanden als materielles, physisches Objekt, als gekaufte, gegenständliche Ware, ist eingängig. Sie verweist einerseits auf die bereits vorgestellte Dimension der Objektifizierung: Smart Speaker präsentieren sich als physischer, zunächst zu entpackender Artefakt mit Farbe und Form und rufen als solche ästhetische Bewertungen und Einstellungen hervor; ihre materiale Beschaffenheit wird z. B. durch die Anwendung von Touch-Technologien oder Lichtringen auch semiotisch genutzt und durch ihre Lokalisierung in der Wohnumgebung rufen sie interaktional-räumliche Orientierungen und Schemata auf. Durch die physische Präsenz des Geräts erhält es in Interaktionsordnungen ein körperäquivalentes ‚Dasein‘ und ist folglich auch körperlich ‚adressierbar‘: Ressourcen von Blick, Mimik, Gestik und Körperhaltung könnten von Interaktionsbeteiligten etwa genutzt werden, um das eigene Sprechen als ‚an einen Smart Speaker gerichtet‘ verständlich zu machen.

Demgegenüber ist die Interpretation dessen, was unter der ‚zweiten Artikulation‘ zu verstehen ist, schwankend und bis in die heutigen Arbeiten hinein unstet. Wie oben schon zum Zwecke der Veranschaulichung geschehen, bietet es sich auf den ersten Blick an, den ‚Objekten‘ kontrastiv die ‚Inhalte‘ gegenüberzustellen, die sie als Bedeutungsträger vermitteln; diese naheliegende Lesart ist auch in etlichen Studien zur Anwendung gekommen (vgl. etwa Röser 2007). Dies lässt sich auch mit den Ausführungen von Silverstone (1994: 123) begründen, der die Annahme vorschlägt, dass „the cultural value of such a machine as a television lies both in its meaning as an object [...] and in its content [...]“, sodass der Begriff des „Inhalts“ bereits gesetzt zu sein scheint. Zuvor geht Silverstone (1994: 123) jedoch auch auf eine andere Basis für die ‚zweite Artikulation‘ ein: „Television is a medium and its communications – its programmes, narratives, rhetorics and genres provide the basis for its second articulation“. So sind verschiedentlich auch nicht direkt die vermittelten Inhalte, sondern vielmehr die *Inhaltsgestaltung*, die ‚Software‘ oder die vollzogenen Narrative, die Sendepläne oder die „Programm-Narration“ (Hartmann 2013a: 26) herangezogen worden (vgl. auch Hartmann 2006: 87) – ausgelöst sicherlich auch durch eine Formulierung bei Silverstone/Haddon (1996: 74), in der die doppelte Artikulation – anders als bei Silverstone (1994) – explizit nicht als ‚Inhalt‘ („specific ‚programme‘ messages“), sondern als „symbolic environment to which the technology gives access“ verstanden wird – und, so muss konstatiert werden, in den folgenden Domestizierungsstudien und -bezügen, immer wieder unterschiedlich und teilweise diffus rezipiert wird.

Eine mögliche Interpretation ist, dass Silverstone und Haddon eine analytische Trennlinie zwischen der Objekt- und der Inhaltsseite aufstellen, ohne damit

die Thesen etwa Marshall McLuhans zu unterlaufen und eine singularistische Perspektive auf einzelne Inhalte bzw. Texte zu postulieren, die ohne Ansehen ihrer medialen Vermittlung interpretiert werden. Darüber hinaus scheint auch weder in den Arbeiten der frühen britischen Vertreter noch in den Weiterentwicklungen und Anwendungen des Domestizierungsansatzes ein starkes Interesse an einer Auseinandersetzung mit den einzelnen Inhalten in ihrer jeweiligen konkreten Ausprägung (einer konkreten Sendung oder Anwendung und deren visuelle, auditive Ressourcen) zu bestehen, sondern in der Tendenz werden unter der zweiten Artikulation vielmehr die Medientechnologien neben ihrer Eigenschaft als Objekt als *Potenzial* verstanden, überhaupt Inhalte zu vermitteln. Der Fokus liegt darauf, welcher Art diese Inhalte sein können, wie sie zwischen domestischer Sphäre und Außenwelt ‚vermitteln‘ können, wie sie programmatisch, narrativ, rhetorisch gestaltet, aufbereitet sowie durch die Objekte geprägt sind und im Haushalt rezipiert werden. Diese Sichtweise bezieht sich durchaus auf die Inhaltsseite der Medientechnologien, betrachtet aber weniger einzelne Kommunikate, sondern vielmehr den Charakter der potenziell vermittelbaren Inhalte und ihren historischen, gesellschaftlichen Kontext.

Der Vorschlag von Hartmann (2006: 96–97; 2013a: 117–123), die doppelte Artikulation auf eine dreifache Artikulation auszuweiten, in der Medien als Objekte, Inhalte und Kontexte bzw. Rituale verstanden werden, kann als Versuch gelesen werden, diese Undeutlichkeiten aufzuheben. Wie Hartmann allerdings selbst bemerkt, bleiben Unklarheiten in Bezug auf die Begriffe ‚Inhalt‘ und ‚Kontext‘: Unter Inhalten können weiterhin sowohl eine engere Auslegung (im Sinne einzelner Kommunikate) als auch eine weitere Auslegung als inhaltlicher Kontext (oder beides) verstanden werden; ferner ist der Begriff ‚Kontext‘ stellenweise unklar, da er sowohl für den konkreten sozialen, alltagspraktischen Kontext der Medienutzung als auch für den weiteren Kontext im Sinne eines gesellschaftlichen Bezugs der Inhalte und deren Formate zu gebrauchen wäre (siehe auch Livingstone 2007). Auf Hartmanns Arbeiten aufbauend legen allerdings Courtois/Verdegem/Marez (2013) eine sehr brauchbare Heuristik vor: Sie unterscheiden zwischen Medien als Objekten, als Texten und als „immediate sociospatial context“ (Courtois/Verdegem/Marez 2013: 424). Am Rande dieser Triangulation platzieren sie den ‚Kontext‘ im gesellschaftlichen Sinne, der zwar für alle drei ‚Artikulationen‘ der Medien eine Rolle spielt, aber nicht selbst eine darstellen soll (vgl. Courtois/Verdegem/Marez 2013: 425).

Mit dieser ‚dreifachen Artikulation‘ ausgestattet entsteht eine größere heuristische Klarheit, die auch Smart Speaker der Analyse in diesem Sinne zugänglich macht: Neben der bereits angerissenen Ebene von Smart Speakern als ‚Objekte‘ (s. o.) können so die von Smart Speakern in die Interaktion eingebrachten ‚Inhalte‘ umrissen werden – darunter fallen dann z. B. Nachrichten, Musikstreaming

oder interaktive Angebote.⁶⁵ Dazu in Beziehung zu setzen sind drittens der Kontext und die wechselseitige und ritualisierte Herstellung dessen – einerseits etwa durch die Anwendung oder Nicht-Anwendung von Begrüßungsformeln, das Betreiben von Face-Work (Goffman 1955; Brown/Levinson 1987) im Dialog mit dem Smart Speaker, andererseits durch die Einbettung des Smart Speakers in spezifische soziale Situationen und nicht zuletzt durch die abgerufenen Inhalte. Der gesellschaftliche Kontext (z. B. zu Fragen der Datenverwertung, des Privatheits- und Datenschutzdiskurses ebenso wie zu Genderfragen oder zur Verteilung von Arbeit und Lohn)⁶⁶ kann in diesem Rahmen analytisch abgegrenzt werden und wird nicht explizit beachtet, ist aber *in situ* nicht von den tatsächlichen Alltagspraktiken zu trennen.

2.3.2.3 Moralische Ökonomie des Haushalts

Einen dritten Stützpfeiler des Domestizierungsmodells bildet die „moral economy of the household“. Diesen Teil des Konzepts bauen Silverstone/Hirsch/Morley (1992), wie auch die Fokussierung auf den Haushalt, auf anthropologischen und historischen Studien auf, die den Haushalt als nicht nur als wirtschaftlich zusammengehörige Wohneinheit mit geteilten Mahlzeiten betrachten, sondern die Haushalte auch als „part of a transactional system of economic and social relations within the formal or more objective economy and society of the public sphere“ (Silverstone/Hirsch/Morley 1992: 14) sehen. Haushalte sind also eine Einheit innerhalb eines privaten, informellen, transaktionalen ‚Subsystems‘ der Bewohner*innen – ein *moralisches* Austauschsystem, das auch deswegen ‚Ökonomie‘ genannt wurde, weil so die Relation zur öffentlichen Ökonomie hervorgehoben wird (vgl. Hartmann 2009a: 309).

Ausgetauscht und ‚verhandelt‘ werden moralische Einheiten: Wertvorstellungen, Einsichten, Haltungen, Geschmäcker und Bewertungen über die den Haushalt umgebende Umwelt (vgl. Silverstone/Hirsch 1992: 16). Diese bilden über ein gemeinsames Erleben und Erfahren der Welt und daraus entstehende Narrative eine Häuslichkeit – einen stabilen ‚moralischen Kreislauf‘, der einerseits dynamisch ist und durch neue Erfahrungen und Erlebnisse in Schwung gehalten wird und über die Zeit auch Veränderungen an Haltungen und Wertvorstellungen hervorrufen kann. Andererseits ist die moralische Ökonomie zugleich Basis für die stabile und veränderungsresistente ontologische Sicherheit im Sinne Giddens‘ (2003[1984]: 50), die bereits Gegenstand war (s. o.) und die – Haushalt und Häuslichkeit zusammengefasst – ein ‚zu Hause‘ ausbilden (vgl. Bakardjieva 2005: 69–72).

⁶⁵ Siehe Kap. 3.2.3.

⁶⁶ Siehe Kap. 3.2.4 bis 3.2.6.

Die moralische Ökonomie bildet eine Integrität, die den Haushalt gegen neue Einflüsse verteidigen oder diese inkorporieren kann und ihn in Bezug zur Öffentlichkeit als Einheit entstehen lässt. Wird dieses häufig als sensibel beschriebenes Gefüge um noch nicht domestizierte Medientechnologien ergänzt – d. h. neue Verbindungen zur ‚Außenwelt‘ geschaffen, mit evtl. unbekannten Formaten und Inhalten – muss die moralische Ökonomie des Haushalts diese in ihren Kreislauf eingliedern, um die Stabilität aufrecht zu erhalten und ihrer Neigung zur Konstanz und zur konservativen Behandlung von Veränderungen beizukommen. Dies geschieht im Rahmen der Domestizierungspraktiken, mit denen die Haushaltsmitglieder die Medien durch Eingliederungs- und Anpassungsleistungen ‚domestizieren‘.

Die ‚Einheit‘ des Haushalts ist in diesem Modell auch als öffentliche, wirtschaftliche Einheit zu begreifen. Die frühen Domestizierungsstudien arbeiteten sich viel am Begriff des ‚Konsums‘ von Medien ab. Sie reihten sich damit ein in eine seit den 1980er-Jahren entstandene Strömung der Medienforschung, die Konsument*innen von Medien nicht als passive, schwammartig aufsaugende ‚Verbraucher‘ von Medien verstanden wissen wollte, sondern diesen eine eigene Handlungsfähigkeit und Autonomie zuschrieb.⁶⁷ Diese Rolle der Haushalte bleibt auch in der späteren Domestizierungsforschung erhalten – „Households, through their consumptions of goods and services, become actively engaged with the formal economy“ (Ward 2006: 148) –, wird allerdings im weiteren Verlauf der Entwicklung auch expandiert und von der Engführung auf die ökonomische Perspektive gelöst. Es wird dann nicht mehr nur der ökonomische Bezug zwischen Haushalt und Umwelt betrachtet, sondern generell die Zugehörigkeit des individuellen Haushalts zu öffentlichen Diskursen, an denen er als Einheit und mit seinen einzelnen Bewohner*innen beteiligt ist. Diese Expansion wird schon von Silverstone (2006: 238) in seiner Konzeptreflexion angelegt:

In one sense the notion of the moral economy is naively empirical. It asks the questions in what ways, if at all, households or families create for themselves private and personal cultures, which have consequences for the way in which the anonymous, homogenizing technologies and services of public and commercial life, are used and valued. (Silverstone 2006: 238)

Noch wichtiger ist aber für die analytische Arbeit der darauf folgende Teil, mit dem deutlich wird, dass der formulierte empirische Anspruch auch tatsächlich in der Untersuchung von Praxis einzulösen ist:

⁶⁷ Siehe auch Kap. 2.3.1.

[...] from a lateral perspective, the question arises too of how we can relate an understanding of patterns of information and communication technology use, resistance, participation and the rest to what we can understand. (Silverstone 2006: 238)

Es sind genau diese Muster der Nutzung, der Widerstände und Beteiligungsstrukturen (und mögliche andere Kategorien, die bereits angesprochen waren), mit Hilfe derer Aufschluss über Wertvorstellungen, entstehende Narrative und Haltungen gegeben werden kann. Deren Konstitution und Weiterentwicklung vollziehen sich in der alltäglichen Nutzung von Medientechnologien: Relevante Geschichten werden geteilt und zu Narrationen aufgebaut, Evaluationen ausgedrückt und verhandelt, Wertvorstellungen implizit und explizit in Praxis vollzogen und geäußert.

Smart Speaker können für die moralische Ökonomie des Haushalts und für dessen ontologische Sicherheit in besonderem Maße eine ‚Bedrohung‘ sein. Dazu trägt zum einen die Verwendung interaktionsähnlicher, gesprochener Sprache als Interface bei. Zum anderen entsteht durch eine teilweise unklare Datenverarbeitung der eigenen Stimme, durch die Anknüpfung an kommerzielle Dienste und Online-Shopping sowie durch die konstante Aufzeichnung der Wohnumgebung ein virulenter Datenschutz-Diskurs um Smart Speaker und andere Smart-Home-Anwendungen (siehe etwa Pridmore/Mols 2020; Mols 2021; Waldecker 2022; Hector et al. 2023).⁶⁸ In diesem positionieren sich die Anwender*innen von Smart Speakern und bauen mittels ihrer Nutzungspraxis und durch ihre eigenen praktischen Erfahrungen und Erlebnisse mit den Geräten sowie den Austausch mit anderen Haushaltsmitgliedern, eigene Bewertungen auf, die Teil von Wertvorstellungen werden können und in die moralische Ökonomie des Haushalts einfließen. Dieser ist (auf die eine oder andere Weise, selbst *ex negativo*) Bestandteil des öffentlichen Diskurses. Insofern sind Evaluationen der Nutzungspraxis mit Bezug auf den Datenschutz, Positionierungen der Haushaltsmitglieder aber auch – gerade im Hinblick auf die Domestizierungsdimension der ‚Umwandlung‘ – der Austausch mit Besucher*innen und die dabei entstehende Entwicklung eigener Positionen, für die Beschreibung der moralischen Ökonomie des Haushalts von besonderer Relevanz.

2.3.2.4 Kritik des Domestizierungsansatzes

Der Domestizierungsansatz ist unter verschiedenen Gesichtspunkten kritisiert worden,⁶⁹ denen teilweise mit der spezifischen Methodologie der vorliegenden Arbeit, teilweise mit im Gegenstandsbereich der Smart Speaker selbst liegenden Argumenten begegnet werden soll. So betonen Arbeiten, in denen die Dimensio-

⁶⁸ Siehe Kap. 3.2.6.

⁶⁹ Für eine Übersicht siehe Hartmann (2013a: 101–117).

nen des Domestizierungsansatzes überwiegend selbst angewendet werden, dass diese nicht linear zu begreifen sind und arbeiten sich an einer vermeintlichen Linearität von Kommodifizierung, Aneignung und Umwandlung ab; gleichzeitig verarbeitet eine Reihe von Arbeiten diese Dimensionen in der Tat zu einer linearen Abfolge. Die Kritik muss sich hier auf die Arbeiten beziehen, die von einer temporalen Abfolge der Dimensionen ausgehen. Als Umgang damit schlägt Ling (2004: 32) vor, den Ansatz immer wieder in der Empirie zu überprüfen und festzustellen, ob spezifisch temporale Muster auffällig sind. Ähnliches gilt für die immer wieder diskutierte Frage nach der Handlungsmacht der Nutzer*innen: Mit dem Begriff der ‚Domestizierung‘ scheint sich – so die Kritik – die Medienaneignung auf einen Prozess des Zähmens zu beschränken, der die reziproken Mechanismen dieses Vorgangs vernachlässigen könnte und die Distribution von Agency aus dem Blick verliert.⁷⁰ Dabei sind sowohl eine Überbetonung der Agency der Medientechnologien auf der einen (siehe etwa Feenberg 1999: 107; Helle-Valle/Slettemeås 2008) wie auch eine Überbetonung der Agency der Nutzer*innen auf der anderen Seite (vgl. Bassett 2007: 65–72) kritisiert worden. Für die vorliegende Arbeit bedeutet dies: Erstens muss auch dieser Kritikgegenstand Frage einer empirischen Überprüfung bleiben: Relevant sind die Strategien, mit denen Nutzer*innen *in situ* Handlungsträgerschaft aushandeln und dabei die beteiligten Geräte einbeziehen (vgl. Habscheid/Hector/Hrncal 2023). Zweitens wird mit dem bereits eingeführten praxeologischen Konzept der „materiellen Partizipanden des Tuns“ von Hirschauer (2004; 2016)⁷¹ eine praxeologische Perspektive eingenommen, die eine empirische Betrachtung dieser Fragestellung frei von abstrakten Festlegungen ermöglicht.

Bassett (2007: 63–67) ist es auch, die die bereits mehrfach formulierte Kritik, dass die soziale Praxis der Domestizierung nicht hinreichend Ausgangspunkt der Domestizierungsforschung ist, zugespitzt formuliert und die These aufstellt, dass die Betrachtung der *Routinen* des Alltags die Betrachtung der sozialen Praxis verdränge. Diese Kritik muss in zweierlei Hinsicht zurückgewiesen werden: Einerseits, weil Silverstone (2006: 231) kurz zuvor richtigerweise betont: „Domestication is practice“. Die Praxis der Domestizierung von Medientechnologien nachzuvollziehen und verstehbar zu machen, ist – jedenfalls in der vorliegenden Arbeit – gerade Gegenstand der Untersuchung. Zweitens sind, insbesondere in einem praxeologischen Verständnis, wie es hier skizziert wurde, insbesondere die Routinen des Alltags soziale Praxis (vgl. Schüttpelz/Meyer 2017: 158). Der Alltagsbegriff war ebenfalls

⁷⁰ Waldecker/Hector/Hoffmann (2024) sowie Habscheid/Hector/Hrncal (2023) diskutieren die Agency der Nutzer*innen von Smart Speakern vor dem Hintergrund der Domestizierungsprozesse.

⁷¹ Siehe dazu Kap. 2.1 bzw. 2.2.

Gegenstand vereinzelter Kritik: Die Aneignung neuartiger Medientechnologien und deren Einbettung falle eben *nicht* unter die Alltagspraxis (vgl. Ling 2004: 32) und im Speziellen nicht die erste Inbetriebnahme. Es kann an dieser Stelle nicht ausführlich auf Theorien des Alltags eingegangen werden, doch sprechen erstens die Argumente einer Phänomenologie des Alltags wie sie Schütz/Luckmann (2003) skizzieren durchaus dafür, dass auch solche Situationen zum „Alltagsleben“ zählen (‘alltätlich’ ist in diesem Sinne nicht mit ‘gewöhnlich’ gleichzusetzen) und zweitens wird abseits der begrifflichen Fragestellungen in der vorliegenden Untersuchung durch eine zweite Erhebungsphase mehrere Monate nach der ersten Inbetriebnahme der Smart Speaker sichergestellt, dass auch eine routiniertere und ‘gewöhnlichere’ Nutzung, die sich in übrige Alltagsabläufe der Haushalte einbettet, in den Blick genommen werden kann.⁷²

Der letzte Kritikpunkt, auf den hier (noch einmal zusammenhängend) eingegangen werden soll, betrifft die Frage nach dem *Wo* der Alltagsroutinen. Es wurde zurecht eingewendet, dass der Alltag keineswegs nur in der Wohnumgebung der Nutzer*innen stattfindet, sondern auch außerhalb. In diesem Zusammenhang wurde das Modell – wie bereits kurz erwähnt – auch für eine Dichotomie von „innen“ und „außen“ kritisiert; dieser Kritik ist, wie ebenfalls bereits ausgeführt, schon konstruktiv begegnet worden (vgl. Bakardjieva 2006; Hartmann 2013b; 2019). Darüber hinaus trifft sie im Falle der Anwendung auf den Gegenstandsbereich der Smart Speaker nur bedingt zu: Da diese Geräte für die stationäre Nutzung konzipiert sind und sich insbesondere die hier im Fokus stehende verbale Steuerung auch nahezu ausschließlich in der Wohnumgebung vollzieht, ist eine Domestizierungspraxis von Smart Speakern außerhalb dessen nur schwer vorstellbar – diese müsste mit der Außerhaussteuerung etwa von Smart-Home-Geräten verbunden sein, die dann aber in eine größere Medienökologie eingebettet wären. Das soll noch einmal nicht heißen, dass nicht gerade Smart Speaker durch ihre Funktionsweise zur Auflösung von Dichotomien von Öffentlichkeit und Privatheit beitragen – im Gegenteil (vgl. Waldecker 2022). Wie diese Auflösungsprozesse interaktiv verhandelt werden und im Prozess der Domestizierung – gerade in der Dimension der Umwandlung und in der Beschaffenheit der moralischen Ökonomie des Haushalts – von den Beteiligten relevant gesetzt werden, wäre aber im Rahmen empirischer Analysen zu zeigen. Auch sind Steuerungen von Smart-Home-Anwendungen über mobile Apps möglich, die auf eine verteilte Infrastruktur zurückgreifen, in die auch der Smart Speaker eingebunden ist (vgl. Goulden 2019; Strüver 2023a). Doch die Nutzungspraxis mit Smart Speakern, deren Domestizierung hier im Vordergrund steht, sind an die Wohnumgebung ge-

⁷² Siehe Kap. 5.

koppelt, sodass im Rahmen der Datenerhebung eine Fokussierung auf diese sinnvoll scheint. Inwiefern Außerhäusliches dann für den Domestizierungsprozess relevant wird, ist als empirisch zu beantwortende Frage mit in die Analysen zu tragen. Dieser vollzieht sich in den sprachlichen Praktiken der Nutzer*innen, mit denen Smart Speaker auf der einen Seite genutzt und zugleich in die kommunikativen Konfigurationen der Haushalte eingebettet werden.

2.3.3 Zwischenfazit Aneignung und Domestizierung

Wendet man die dargestellten, bisherigen Untersuchungen zur kommunikativen Aneignung und Domestizierung auf Smart Speaker an, tritt die sprachliche Dimension der Domestizierungs- bzw. Aneignungsprozesse auch aufgrund der Akustik als spezifisches Charakteristikum in der Bedienung dieser Interfaces besonders hervor (vgl. Merkle/Hector 2025; Habscheid/Hector/Hrncal 2025). Der Domestizierungsansatz wird dabei mit den Arbeiten zur sprachlichen Aneignung von Medientechnologien kombiniert und erweitert: Erstens vollzieht sich der in den Dimensionen und Konzepten abgebildete Domestizierungsprozess in den sozialen Praktiken der Haushaltsmitglieder und ist insofern an der sprachlichen Oberfläche beschreib- und verstehbar. Die sprachlichen Praktiken der Haushaltsmitglieder ziehen sich durch alle Dimensionen des Domestizierungsansatzes. Sie sind der Treibstoff, mit dem sie die Geräte in ihre interaktiven Ordnungen einbinden, sich zu ihnen positionieren und mit denen sie innerhalb der moralischen Ökonomie des Haushalts Smart Speaker als Teil des Wertesystems verhandeln und über den sie überhaupt zu entsprechenden Einsichten und individuellen Werturteilen gelangen. Wie immer wieder betont wird, nimmt der Domestizierungsansatz schon von seiner Verwurzelung her nicht nur die Praktiken innerhalb der häuslichen Umgebung selbst in den Blick, sondern ist auch an Prozessen an der Schnittstelle von Öffentlichkeit und Privatheit interessiert. Das macht ihn für die Analyse von Smart Speakern besonders interessant, da diese Geräte Teil einer fortlaufenden, diskursiven Aushandlung von Grenzen zwischen den beiden Domänen sind: Schon durch ihre Funktionsweise und ihre Datenverarbeitung und -verwertung machen sie Aspekte des Privaten zu Aspekten von (teils opaken) Öffentlichkeiten und zu Teilen eines sich radikal verändernden wirtschaftlichen Systems. Das ubiquitäre Konzept des „Überwachungskapitalismus“ (Zuboff 2018)⁷³ wird auf diese Weise einer empirischen Überprüfung auf Seiten der Anwender*innen unter-

⁷³ Siehe Kap. 3.2.6.

zogen. Durch die empirischen Methoden der angewandten Sprachwissenschaft⁷⁴ wird das Domestizierungsmodell um eine Analyseebene erweitert, in der die von den Nutzer*innen selbst relevant gesetzten Aspekte in den Mittelpunkt rücken.

Damit wird zweitens der Versuch unternommen, das Modell methodologisch auszubauen: In den bisherigen Untersuchungen zur Domestizierung standen v. a. Methoden der qualitativen Sozialforschung und insbesondere der Ethnografie im Mittelpunkt. Davon wird durch die Anwendung einer ethnomethodologisch-praxeologischen Gesprächsanalyse, wie sie in den folgenden Analysen vorzunehmen ist, zwar nicht radikal abgewichen; allerdings sind Methoden im Kernbereich der Gesprächsanalyse bisher v. a. im Rahmen der sprachwissenschaftlichen Aneignungsforschung angewendet worden und nicht so sehr in einer integrierten Betrachtungsweise mit dem in der Kommunikations- und Medienwissenschaft beheimateten Domestizierungsansatz zum Einsatz gekommen. Der darin liegenden Spannung, dass die klassischerweise verfolgten Erkenntnisinteressen des Domestizierungsansatzes nicht unbedingt deckungsgleich mit denen der empirischen Linguistik zu sein scheinen, wird in zweierlei Hinsicht begegnet: Erstens hat es in der angewandten Sprachwissenschaft durchaus bereits Tradition, linguistische Methoden und Analysen in den Dienst der Beantwortung von Forschungsfragen aus anderen Teildisziplinen zu stellen (vgl. etwa Meer/Pick 2019). Treibend für die Analyse ist nicht das Interesse an den linguistischen Strukturen selbst, sondern das Augenmerk liegt auf den Zusammenhängen von sprachlichen Praktiken und kommunikativer Funktion. Zweitens profitiert von dieser Verbindung insbesondere eine Medienlinguistik, die sich zur Aufgabe macht, den Zusammenhang von Medien und sprachlichen Formen aufzudecken. Gerade technologische Artefakte wie Smart Speaker, die konstitutiv einen sprachlichen Dialog mit ihnen evozieren, bedürfen einer Form der Medienrezeptionsforschung, die deren dialogische Anatomie in den Mittelpunkt stellt und die Medienaneignung entlang der sprachlichen Praktiken beleuchtet.

⁷⁴ Siehe dazu Kap. 4 und 5.