

Henning Mayer, Florian Muhle und Indra Bock  
**Whiteboxing MAX**

Zur äußeren und inneren Interaktionsarchitektur  
eines virtuellen Agenten

## Einleitung

Im Zentrum dieses Sammelbandes stehen „das Versiegeln und Öffnen von Black Boxes“. Damit befassen sich die hier versammelten Beiträge mit einer zentralen Problemstellung der interdisziplinären Wissenschafts- und Technikforschung.<sup>1</sup> Denn in deren Tradition – ob sozialkonstruktivistisch<sup>2</sup> oder ‚material-semiotisch‘<sup>3</sup> – geht es vornehmlich darum, die sog. Black Boxes der Wissenschaft und der Technik zu öffnen. Als Black Boxes können mit Bruno Latour die funktionierenden bzw. ‚wahren‘ Resultate der Technikentwicklung und wissenschaftlichen Forschung betrachtet werden, denen nicht mehr anzusehen ist, welche Arbeit und kontingenten Entscheidungen in sie eingeflossen sind.<sup>4</sup>

Entsprechend scheint es naheliegend Black Boxes als Formen geronnener Macht zu begreifen und mit dem Begriff des Dispositives, wie er bei Foucault zu finden ist, eng zu führen, wie es als Möglichkeit auch auf der diesem Band zugrundeliegenden Tagung diskutiert wurde. Für entsprechende Projekte gibt es bereits einige Vorbilder, die insbesondere auf Übereinstimmungen des Dispositiv-Begriffes mit dem material-semiotischen Konzept der Akteur-Netzwerke hinweisen.<sup>5</sup>

Aber auch wenn ein solcher Zugang plausibel ist und gewinnbringende Einsichten verspricht, scheint er uns dennoch zu sehr den Untersuchungsschwerpunkt auf Machtanalyse und – mit Blick auf technische Artefakte – das Funktionieren von Technik zu lenken. Denn während eine solche Perspektive im Konzept der Black Box bereits angelegt ist, könnte diese durch den Rückgriff auf den Dispositivbegriff gleichsam festgeschrieben werden. Eine Gefahr hierbei scheint uns zu sein, technische Artefakte a priori als Ergebnis strategischer und machtvoller Konstruktionsprozesse zu begreifen und deren Funktionieren und Stabilität zu sehr vorauszusetzen, was möglicherweise mit zu wenig analytischer Sensibilität für Zufälle, Inkonsistenzen und ‚schwache Performance‘ von Technik einhergeht.<sup>6</sup>

---

1 Vgl. Pinch 1992, 489.

2 Vgl. Pinch/Bijker 1984; Pinch 1992.

3 Vgl. Law 2013 [2009].

4 Vgl. Latour 1999, 304.

5 Vgl. Matthewman 2013; Seier 2013; Gnosa 2018.

6 Ein ähnliches Problem hat Susan Leigh Star (Star 1990) bei frühen Studien der Akteur-Netzwerk-

Darum möchten wir vorschlagen, sich bei der Analyse technischer Artefakte weniger an der Machtanalytik Foucaults zu orientieren, sondern sich Methodologien zuzuwenden, die mit weniger starken Begriffen arbeiten und damit verbunden einen ergebnisoffeneren Blick erlauben. Eine vielversprechende Kandidatin hierfür scheint uns die kommunikationstheoretisch und ethnomethodologisch inspirierte *Interaktionsarchitekturanalyse* zu sein, wie sie etwa von Hausendorf, Schmitt und Kesselheim vorgeschlagen wird.<sup>7</sup> Dies möchten wir im Folgenden anhand einer Analyse der (äußeren und inneren) Interaktionsarchitektur des verkörperten Agenten MAX plausibilisieren, der sich als Ausstellungsobjekt im *Heinz-Nixdorf-Museumsforum* (HNF) in Paderborn befindet und Museumsbesucher\*innen dazu einlädt, mit ihm zu ‚interagieren‘.<sup>8</sup> Konkret werden wir kurz die Interaktionsarchitekturanalyse und deren von uns vorgenommene Erweiterung vorstellen (Abschnitt 2), um uns dann sowohl der ‚äußeren‘ als auch der ‚inneren‘ Interaktionsarchitektur der Black Box MAX zuzuwenden. Gemeint sind damit die sichtbare ‚äußere‘ Interfacegestaltung auf der einen Seite (Abschnitt 3) und die Systemarchitektur von MAX auf der anderen (inneren) Seite (Abschnitt 4). Deren Analyse wird verschränkt mit der Untersuchung des Beginns einer ‚Interaktion‘ eines Museumsbesuchers mit MAX, um auf diese Weise zu zeigen, dass und wie die Interaktionsarchitekturanalyse gewinnbringende Einsichten für das Verständnis von (Verläufen der) Mensch-Maschine-Interaktion produziert und auf diese Weise zugleich einen relevanten Beitrag zur Analyse technischer Artefakte liefert.

## Interaktionsarchitektur(en)

Das Konzept der Interaktionsarchitektur ist in den vergangenen Jahren im Kontext ethnomethodologisch und kommunikationstheoretisch inspirierter Forschung entstanden, die sich – wie andere sozialtheoretische Ansätze im Gefolge von Material, Practice und Spatial Turn auch – zunehmend für die Bedeutung von Materialität und Verkörperung für soziale Sinnproduktion interessiert.<sup>9</sup> Das Besondere des Konzeptes der Interaktionsarchitektur liegt u. E. darin, dass es erstens die Beachtung der materiellen Umwelt mit einem operativen Verständnis von Sozialität verbindet und zweitens in der Tradition der Ethnomethodologie Wert darauf legt, Analysen und Einsichten ‚from the data

---

Theorie gesehen, denen sie zuschrieb „managementzentriert und sogar militärischen Charakters“ (Law 2013 [2009], 35) zu sein.

<sup>7</sup> Hausendorf/Schmitt/ Kesselheim 2016.

<sup>8</sup> Die folgenden Überlegungen und Analysen entstammen dem von der DFG geförderten Forschungsprojekt ‚Kommunikation an den Grenzen der Sozialität‘ ([www.gepris.dfg.de/gepris/projekt/31892483](http://www.gepris.dfg.de/gepris/projekt/31892483)). Wir sind sowohl dem HNF – und hier insbesondere Dr. Stefan Stein – als auch Prof. Dr. Ipke Wachsmuth und Dr. Thies Pfeiffer zu großem Dank verpflichtet.

<sup>9</sup> Vgl. Hausendorf/Schmitt/Kesselheim 2016.

themselves‘ zu entwickeln.<sup>10</sup> Dies bedeutet einerseits, dass materielle Phänomene nicht als solche interessieren, sondern nur mit Blick darauf, wie sie a) bestimmte Formen der Kommunikation nahelegen und b) selbst im Zuge von Kommunikationsprozessen relevant gemacht werden. Andererseits geht mit der datengetriebenen Analyseperspektive einher, sich den interessierenden Phänomenen nicht kategoriengeleitet zu nähern (bspw. machtanalytisch), sondern weitgehend (ergebnis-)offen.

Die Kernidee der Interaktionsarchitekturanalyse besteht darin, räumliche und materielle Erscheinungsformen als „Manifestation von Erwartungen“<sup>11</sup> zu interpretieren, die bestimmte „Bewegungen, Wahrnehmungen und Handlungen unter Anwesenden nahelegen“,<sup>12</sup> um daran anschließend zu betrachten, wie von Interagierenden an diese Erwartungen angeschlossen wird. Genau in diesem Sinne begreifen Hausendorf und Kesselheim Interaktionsarchitektur in einer an Luhmann angelehnten Formulierung als „das physische Substrat der Kommunikation mit und durch Architektur, das als Anreiz verstanden werden kann, eine Unterscheidung von Information und Mitteilung zu machen und so den Raum zu ‚verstehen‘“.<sup>13</sup>

Damit verbunden ist die Vorstellung, dass je spezifische Anordnungen und Gestaltungsweisen von Materialien und Gegenständen im Raum von (Inter-)Agierenden als interaktive Ressourcen genutzt werden können, die Hinweise darauf liefern, wie sie sich benutzen, betrachten, anhören, beschreiten, bewegen oder bedienen lassen.<sup>14</sup> Entscheidend ist hierbei, dass entsprechende Hinweise mögliche Anschlüsse nicht determinieren, sondern „strukturell überschüssig“<sup>15</sup> sind und eher ein „Universum“<sup>16</sup> an Anschlussmöglichkeiten auffächern. Zudem muss damit gerechnet werden, dass entsprechende Hinweise auch widersprüchlich oder verwirrend sein können. Daher ist es auch wichtig, empirische Analysen strikt gegenstandsgebunden durchzuführen und sich weder von vorgegebenen analytischen Kategorien noch von der eigenen „sozialräumlichen Kompetenz“<sup>17</sup> zu voreiligen Deutungen verleiten zu lassen. Stattdessen gilt es – ähnlich wie dies auch in der Methodologie der objektiven Hermeneutik vorgesehen ist<sup>18</sup> – verschiedene ‚Lesarten‘ der interessierenden Interaktionsarchitektur zu entwickeln und deren „Implikationsreichtum im Detail“<sup>19</sup> zu rekonstruieren.

---

**10** Hausendorf/Schmitt 2016c, 166.

**11** Hausendorf/Schmitt 2016a, 39.

**12** Hausendorf/Schmitt 2016a, 34.

**13** Hausendorf/Kesselheim 2016b, 63.

**14** Vgl. Hausendorf/Schmitt 2016a, 34. Entsprechende Raumordnungen sind immer auch mit institutionalisierten Nutzungsweisen verknüpft. Einerseits gibt es demnach typische Raumanordnungen bspw. von Hör- oder Kirchensälen, andererseits etablierte Raumnutzungsmuster, derer sich Menschen vertrautheits- und wissensabhängig bedienen können.

**15** Hausendorf/Schmitt 2016c, 172.

**16** Hausendorf/Schmitt 2016c, 172.

**17** Hausendorf/Schmitt 2016c, 172.

**18** Hausendorf und Schmitt weisen selbst auch auf die Ähnlichkeit der eigenen Methodologie mit dem Analyseverfahren der Objektiven Hermeneutik hin (Hausendorf/Schmitt 2016c, 172).

**19** Hausendorf/Schmitt 2016c, 176.

Konkret schlagen Hausendorf und Schmitt vor, die Analyse der Interaktionsarchitektur als ersten Teil einer umfassenden Interaktionsanalyse durchzuführen und hierzu auf Standbilder von Videoaufzeichnungen (möglichst ohne interagierende Personen), aber (wenn nötig und vorhanden) auch auf weitere Daten wie Fotos, Zeichnungen, Pläne oder Grundrisse der interessierenden Phänomene zurückzugreifen.<sup>20</sup> Diese gilt es dann in zweierlei Hinsicht aufzuschlüsseln: Zum einen gilt es die im untersuchten Gegenstand simultan vorfindbaren Aspekte in ihren Details, aber auch im Verhältnis der Details zueinander auf ihre Implikationen hin abzusuchen. Zu diesem Teil der Rekonstruktion gehört also insbesondere „die Frage der Gleichzeitigkeit und des Verhältnisses von sehr unterschiedlichen Aspekten“,<sup>21</sup> wodurch auch (Un-)Stimmigkeiten in der Interaktionsarchitektur zu Tage treten. Zum anderen soll die untersuchte ‚fixierte‘ Gestalt aber auch in ihrer sequenziellen Einbettung in den Blick genommen werden. Hierbei geht es entsprechend des Anliegens der Interaktionsarchitekturanalyse darum zu klären, was aufgrund des Eingefrorenen, im Standbild Sichtbaren im Anschluss zu erwarten ist,<sup>22</sup> um in einer anschließenden Untersuchung sich tatsächlich vollziehender Interaktion(en) zu rekonstruieren, wie hieran von Seiten Interagierenden angeschlossen wird.

So wie es von seinen Vertreter\*innen konzipiert wurde, bezieht sich das Konzept der Interaktionsarchitektur, vor allem auf den gebauten und gestalteten Raum, in dem sich Interagierende bewegen. U. E. lässt es sich aber auch zur Analyse konkreter technischer Artefakte und ihrer ‚Interfaces‘ sowie zur Analyse der ‚inneren‘ Architektur dieser Artefakte, namentlich ihrer Programmierung fruchtbar machen. Wir schlagen also vor, das Konzept der Interaktionsarchitektur als allgemeinen analytischen Rahmen zu begreifen, der für die

1. Analyse der räumlichen Einbettung eines Artefaktes,
2. Analyse der äußeren Gestalt bzw. des Interfaces eines Artefaktes und
3. Analyse der internen Funktionsweise des Artefaktes

genutzt und – im Sinne der Gegenstandsangemessenheit von Methoden<sup>23</sup> – jeweils spezifisch angepasst werden kann.

Mit Blick auf die Analyse der Interfacegestalt(ung) fällt eine solche Anpassung verhältnismäßig leicht. Denn hier kann – wie auch Hausendorf und Kesselheim<sup>24</sup> erwähnen – das Konzept der Interaktionsarchitektur mit Überlegungen zu „Affordances“ von Kommunikationstechnologien kurzgeschlossen werden, die seit einigen Jahren im Kontext der Ethnomethodologie<sup>25</sup> aber auch darüber hinaus<sup>26</sup> in den Blick

<sup>20</sup> Hausendorf/Schmitt 2016c, 177.

<sup>21</sup> Hausendorf/Schmitt 2016c, 176.

<sup>22</sup> Hausendorf/Schmitt 2016c, 176.

<sup>23</sup> Vgl. Garfinkel/Wieder 1992.

<sup>24</sup> Vgl. Hausendorf/Kesselheim 2016b, 79.

<sup>25</sup> Vgl. Hutchby 2001; Arminen et al. 2016.

<sup>26</sup> Vgl. Zillien 2009.

nehmen, in welcher Form die materiale Gestaltung von Kommunikationstechnologien Möglichkeiten mediatisierter Interaktion öffnet – aber auch beschränkt.<sup>27</sup> Der Unterschied zwischen ‚klassischer‘ Interaktionsarchitekturanalyse und der Analyse medialer Affordanzen besteht u. E. lediglich in den divergenten Forschungsgegenständen (gebauter und gestalteter Raum vs. technische Artefakte). Zudem zielt die Analyse medialer Affordanzen dezidiert auf Aspekte wie Benutzbarkeit und Bedienbarkeit und ist damit spezifischer als die Analyse der räumlichen Einbettung eines Artefaktes, bei der es u. a. auch um Fragen der Begehbarkeit und Verweilbarkeit geht.<sup>28</sup>

Etwas schwieriger verhält es sich mit der ‚inneren‘ Funktionsweise, die im Fall des verkörperten Agenten MAX von den Entwickler\*innen tatsächlich auch als „Systemarchitektur“ bezeichnet wird.<sup>29</sup> Denn anders als beim gestalteten Raum und der Interfacegestalt(ung) ist die Systemarchitektur für diejenigen, die mit MAX ‚interagieren‘ nicht direkt ersichtlich, sondern bleibt eine Black Box, die nur indirekt über den Output des Systems Hinweise auf ihre Funktionsweise liefert. Entsprechend gibt die Systemarchitektur auch keine sichtbaren Benutzbarkeits- und Bedienbarkeitshinweise.

Dennoch trägt sie selbstverständlich dazu bei, Interaktionsmöglichkeiten zu eröffnen und einzuschränken, weshalb sie als weiterer Bestandteil der Interaktionsarchitektur begriffen werden kann. Analysen der ‚inneren‘ Interaktionsarchitektur technischer Artefakte komplementieren somit Untersuchungen der kommunikativen Implikationen der ‚äußeren‘ Aspekte der Interaktionsarchitektur und ermöglichen so ein umfassendes Verständnis der Mensch-Maschine-Kommunikation (und ihrer Probleme). Um entsprechende Untersuchungen durchzuführen, bedarf es aber anderer Daten als bei der Analyse von Interfaces und gestaltetem Raum. Denn Fotografien und Standbilder führen hier nicht weiter, dafür aber Informationen über die Systemarchitektur, die etwa von MAX’ Entwicklern in Publikationen zugänglich gemacht werden sowie selbstverständlich der konkrete Programmcode, anhand dessen ganz im Sinne des Affordanzkonzeptes rekonstruiert werden kann, „what is made possible and facilitated, and what is made difficult and inhibited“.<sup>30</sup>

Dies sollen die folgenden Abschnitte zeigen und dabei zugleich das analytische Potential des Konzeptes der Interaktionsarchitektur belegen, indem sie in systematischer Weise Analyse der ‚äußeren‘ und ‚inneren‘ Architektur des Agenten MAX mit der Analyse sich tatsächlich vollziehender Interaktion verschränken. Da wir den Fokus in unserem Beitrag auf Interface und Programmierung legen, verzichten wir auf die Analyse der Einbettung von MAX in den gestalteten Raum und beginnen mit der Analyse der Interface-Gestaltung.

---

<sup>27</sup> Das Konzept der Affordances geht auf den Wahrnehmungspsychologen James J. Gibson zurück (Gibson 1966), hat aber von den genannten Autor\*innen eine soziologische Rahmung erfahren (vgl. insb. Zillien 2009).

<sup>28</sup> Vgl. Hausendorf/Schmitt 2016a, 34.

<sup>29</sup> Vgl. Gesellensetter et al. 2005.

<sup>30</sup> Bearne/Kress 2001, 9.1.

## Analyse des Interfaces von MAX

MAX ist ein sogenannter verkörperter konversationeller Agent, der ursprünglich an der Universität Bielefeld mit dem Ziel entwickelt wurde, „die Grundlagen kommunikativer Intelligenz“<sup>31</sup> zu erforschen und zugleich ein „möglichst funktionstüchtiges, überzeugendes System zu bauen“,<sup>32</sup> welches es Menschen erlaubt, weitgehend „natürlich mit einer Maschine [...] kommunizieren zu können“.<sup>33</sup> Seit dem Jahr 2004 befindet sich eine Version von MAX als Ausstellungsobjekt in der Dauerausstellung des HNF und lässt sich dort von Museumsbesucher\*innen betrachten und ausprobieren.<sup>34</sup>

### Interaktionsarchitektur des Interfaces

Um das Interface von MAX zu analysieren, verwenden wir eine Fotografie, die das Ausstellungsobjekt in seiner Gesamtheit zeigt sowie (weiter unten) zwei Fotografien, welche Detailaufnahmen der zum Objekt gehörenden „Objektkennungen“ darstellen.

Die erste Fotografie (Abb. 1) zeigt die Präsentation des Agenten zu einem Zeitpunkt, an dem noch niemand mit ihm ‚interagiert hat‘. Sie dokumentiert damit gleichsam den ‚Urzustand‘ des Systems, wie er morgens nach ‚Hochfahren‘ des Systems existiert.

Nimmt man gemäß des heuristischen Verfahrens der Interaktionsarchitekturanalyse in einem ersten Analyseschritt die simultan erkennbaren Aspekte des Ausstellungsobjektes MAX in den Blick, fällt sofort auf, dass es sich bei MAX im engeren Sinne nicht um ein Objekt handelt, sondern um ein ganzes ‚Artefaktarrangement‘. Dieses besteht aus

1. der Leinwand, auf der einerseits der Agent zu sehen ist, andererseits aber auch ein zweizeiliges Textfeld
2. dem Stehtisch mit eingelassener Tastatur
3. einer Kamera, die rechts neben der Leinwand ungefähr auf Kopfhöhe erwachsener Besucher\*innen positioniert ist
4. einem Tisch unterhalb der Leinwand
5. zwei Schildern, von denen sich eines auf dem Stehtisch und eines auf dem Tisch unterhalb der Leinwand befindet

---

<sup>31</sup> Wachsmuth 2010, 139.

<sup>32</sup> Wachsmuth 2010, 139.

<sup>33</sup> Wachsmuth 2010, 136.

<sup>34</sup> War MAX lange Teil des Ausstellungsbereiches „Künstliche Intelligenz und Robotik“, ist er seit Herbst 2018 im neu gestalteten Bereich „Mensch, Roboter!“ zu sehen und hat seitdem sein Äußeres leicht verändert. Unsere Analyse bezieht sich jedoch auf den ‚alten‘ MAX, da unsere Datenerhebungen vor der Neugestaltung des Ausstellungsbereiches erfolgte.



**Abb. 1:** Präsentation des Agenten MAX im Heinz Nixdorf MuseumsForum (HNF).

Der Agent in seiner verkörperten Gestalt ist also nur ein Teil eines komplexen Interfaces, das sich aus mehreren Artefakten zusammensetzt. Als Exponat wird dieses Interface vor allem dadurch präsentiert, dass es nicht nur aus Tastatur, Bildschirm bzw. Leinwand und Kamera besteht, sondern darüber hinaus auch mit zwei Schildern versehen ist. Selbst wenn diese sich aus der gewählten Entfernung nicht lesen lassen, lassen sie doch im Rahmen<sup>35</sup> von Ausstellungskommunikation erwarten, dass es sich um museumstypische „Objektkennungen“<sup>36</sup> handelt, die weitere Informationen über das Objekt enthalten und es genau hierdurch als Exponat präsentieren, das betrachtet und (mit Hilfe weiterer Informationen) verstanden werden kann.<sup>37</sup>

Dass es sich bei dem so inszenierten Exponat nicht einfach um ein ‚klassisches‘ Ausstellungsobjekt handelt, macht insbesondere das Zusammenspiel von Leinwand, Tisch mit Tastatur und Kamera deutlich. Denn dieses markiert Ähnlichkeiten mit bekannten Settings von Videokonferenzen (z. B. per Skype), in denen sich die Beteiligten per Kamera wechselseitig auf den eigenen Bildschirm holen, um miteinander zu sprechen, parallel aber auch die Möglichkeit haben, Textnachrichten auszutauschen,

<sup>35</sup> Nach Goffman 1980 [1974].

<sup>36</sup> Hausendorf/Kesselheim 2016b, 76.

<sup>37</sup> Vgl. Hausendorf/Kesselheim 2016b, 76. Der Tisch unterhalb der Leinwand dient in diesem Kontext u. E. als funktionales Äquivalent zur Glasvitrine. Seine Funktion ist darin zu sehen, dafür zu sorgen, dass die Leinwand nicht berührt wird. Denn seine Materialität sorgt dafür, dass niemand in den Bereich unter der Leinwand treten kann und schützt diese dadurch vor Berührungen.

wie dies auch durch das in zwei Zeilen geteilte weiße Feld unterhalb des sichtbaren Agenten angedeutet wird. Dargestellt wird somit die Möglichkeit, mit dem Agenten MAX, der bereits ‚online‘ zu sein scheint, in eine mediatisierte Interaktion nach dem Vorbild der Videokonferenz einzutreten. Entsprechend wird MAX auch nicht als ein typisches interaktives Objekt präsentiert, das bedienbar wäre,<sup>38</sup> sondern als potentielles Gegenüber mit dem man medienvermittelt – je nach Wahl – mündlich oder schriftlich interagieren kann.

Der Unterschied zu einer ‚normalen‘ Konferenz scheint vor allem darin zu liegen, dass im Vergleich zur Nutzung am eigenen Computer der Bildschirm durch eine große und weithin sichtbare Leinwand (und eine große Kamera) ersetzt ist. Das Artefaktarrangement zeigt damit an, dass mögliche Interaktionen nicht privat bleiben, sondern öffentlich sicht- und vermutlich auch hörbar werden. Personen, die in Erwägung ziehen, in Interaktion mit MAX zu treten, müssen also mit Zaungästen rechnen. ‚Interaktionen‘ mit MAX werden dementsprechend als öffentliche Dialoge gerahmt, was Auswirkungen darauf hat, was und wie etwas kommuniziert werden kann.<sup>39</sup>

Die über-lebensgroße Darstellung der Verkörperung des Agenten sowie seine zugewandte Körperausrichtung und sein Blick, der vermeintlich auf die näherkommenen Besucher\*innen gerichtet ist, erzeugen zudem eine spezifische Form der Präsenz des Agenten, die den Eindruck einer Face-to-Face-Situation statt einer medienvermittelten Situation hervorruft.

Was bedeuten diese Überlegungen nun für die Frage, was von Seiten der Museumsbesucher\*innen zu erwarten ist, die sich dem Exponat MAX nähern? Zunächst einmal ergeben sich zwei grundsätzlich unterschiedliche Optionen. So lässt sich einerseits daran anschließen, dass MAX – wie andere Artefakte im Museum auch – als Exponat präsentiert wird. Dies würde bedeuten, ihn lediglich zu betrachten und ggf. näherzutreten, um sich die Schilder durchzulesen und anschließend zum nächsten Exponat weiter zu gehen.<sup>40</sup> Andererseits lässt sich aber auch an das zugleich präsentierte Angebot anschließen, ihn als potentielles Gegenüber zu behandeln. In diesem Fall wäre naheliegend, sich an den Tisch mit der Tastatur und damit auch ins Sichtfeld der Kamera zu begeben, um anschließend entweder schriftlich oder mündlich einen Gesprächseinstieg zu wagen.<sup>41</sup> Weitgehend unbestimmt ist dagegen, worüber man mit MAX eigentlich kommunizieren soll. Denn einen konkreten Anlass gibt es ja zunächst nicht, wodurch jenseits der Tatsache, dass mit Zaungästen zu rechnen ist, kaum Strukturvorgaben für den Austausch existieren.

Haben sich Museumsbesucher\*innen aber dafür entschieden, einen Kontaktversuch zu wagen und sind hierzu tatsächlich an den Tisch mit der Tastatur getreten,

<sup>38</sup> Vgl. hierzu Heath/vom Lehn 2008.

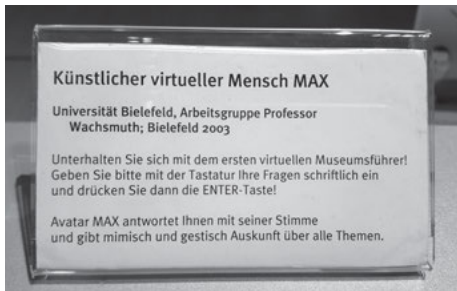
<sup>39</sup> Vgl. Goffman 2005, 50.

<sup>40</sup> Wie sich das Wandeln von einem zum anderen Kunstobjekt vollzieht, hat Dirk vom Lehn (2006) sehr schön anhand der ‚Kunst der Kunstbetrachtung‘ rekonstruiert.

<sup>41</sup> Selbstverständlich lassen sich diese Alternativen auch kombinieren.



erhalten sie durch die nun lesbaren Schilder erste Hinweise zu möglichen Gesprächsthemen – sofern sie diese beachten. So gibt das Schild auf dem Stehtisch wie erwartet tatsächlich Auskunft über das Exponat, liefert darüber hinaus aber auch Hinweise bezüglich thematischer Interaktionsmöglichkeiten. So wird der hier als „künstlicher virtueller Mensch“ vorgestellte Agent, zugleich als „erster virtueller Museumsführer“ eingeführt, dem man Fragen stellen kann und der „Auskunft über alle Themen“ (Abb. 2) gibt. Das zusätzliche Schild auf dem Tisch unter der Leinwand entpuppt sich demgegenüber als Werbung für ein Buch des „Vaters“ von MAX, dem Bielefelder Professor Ipke Wachsmuth (Abb. 3).



**Abb. 2:** Hinweisschild zur Bedienung von MAX.



**Abb. 3:** Hinweis auf ein Buch des Entwicklers von MAX.

Werden diese Schilder, die als „Objektkennungen“ bereits aus der Entfernung ihre Lesbarkeit anzeigen, tatsächlich betrachtet, lassen sie sich als interaktive Ressourcen verwenden, um einen Einstieg ins Gespräch zu finden. So fordert der Text des auf dem Stehtisch befindlichen Schildes die Lesenden explizit dazu auf, Fragen an den ‚Museumsführer‘ MAX zu richten. Nimmt man diese Funktionsbezeichnung sowie die Informationen auf dem Werbeschild ernst, bietet es sich an, Fragen zu formulieren, die sich auf das Museum und die darin befindliche(n) Ausstellung(en) beziehen, und sich spezifischer noch um das Thema Künstliche Intelligenz und vielleicht auch MAX selbst und seinen ‚Vater‘ drehen könnten.

Doch das kleine Hinweisschild auf dem Stehtisch gibt nicht nur Hinweise für einen thematischen Einstieg. Darüber hinaus weist es auch auf eine technische Besonderheit des Interfaces hin, die sich aus dessen Gestaltung erst einmal nicht intuitiv

erschließt. Denn anders, als auf den ersten Blick nahegelegt, ist es offensichtlich nicht möglich mit dem Agenten mündlich zu kommunizieren, bzw. zwischen mündlicher und schriftlicher Kommunikation zu wechseln. Stattdessen müssen User\*innen ihre Fragen schriftlich eingeben, während MAX ausschließlich verbal Auskunft gibt. Zugleich folgt hieraus – ohne dass dies explizit gemacht würde – dass Eingaben, die in dem weißen Feld stehen, nur von menschlichen Nutzer\*innen stammen können. Während in Videokonferenzen üblicherweise die Beiträge beider Seiten stehen bleiben und nur die untere Zeile exklusiv für die (Erstellung der) eigenen Beiträge reserviert ist, ist dies hier also entgegen der weiter oben am Vorbild der Videokonferenz formulierten Erwartungen nicht der Fall.

Dies kann deshalb folgenreich sein, weil im laufenden Betrieb stets die jeweils letzte Äußerung, die über die Tastatur eingegeben wurde, in der oberen Zeile stehen bleibt. Die Darstellung des Artefaktarrangements im unbenutzten Zustand, wie sie sich in Abb. 1 findet, stellt sich somit als Ausnahme im alltäglichen Betrieb des Museums heraus. Die meisten User\*innen finden bereits eingegebenen Text vor, der somit ähnlich wie die Hinweisschilder als Ressource genutzt werden kann, um einen Anfang für die ‚Interaktion‘ zu finden. Je nach Inhalt liefern die ‚stehen gebliebenen‘ Textbeiträge vorheriger Besucher\*innen Hinweise auf Themen und Formen für einen Austausch mit MAX.

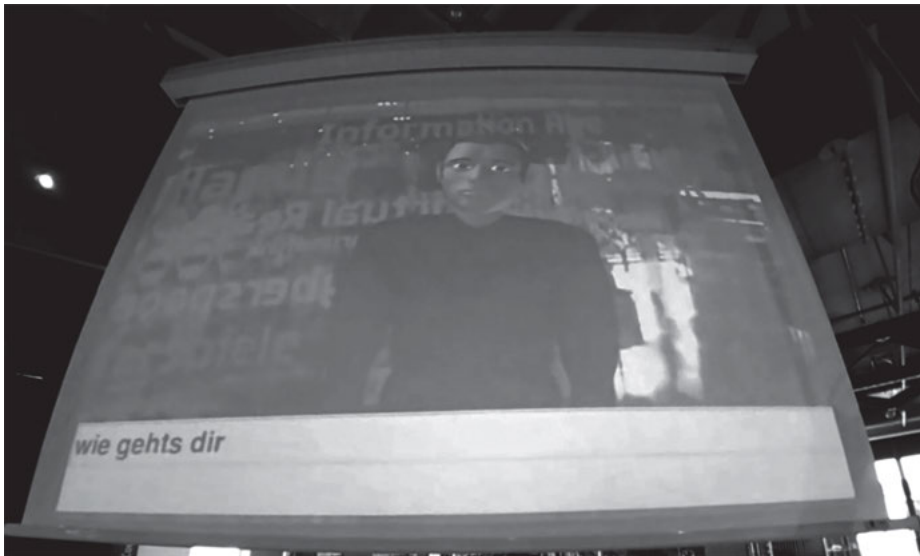
Je nachdem, ob Museumsbesucher\*innen nun die ergänzenden Hinweise auf den Schildern zur Kenntnis genommen und auch deren Relevanz hinsichtlich der technischen Bedingungen des Austauschs begriffen haben, ergeben sich hieraus unterschiedliche Konsequenzen für die Möglichkeiten des Anfangs einer ‚Interaktion‘ mit MAX. Wurden die Informationen zur Kenntnis genommen, steht zu erwarten, dass die Besucher\*innen – zumindest wenn sie gemäß der durch das Schild auf dem Stehtisch zum Ausdruck gebrachten präferierten Nutzung agieren – in der Folge zur Tastatur greifen und eine Frage an MAX formulieren, die dieser als Museumsführer beantworten kann. Sollten sie die Informationen jedoch nicht zur Kenntnis genommen und nachvollzogen haben, scheinen Schwierigkeiten der Kontaktaufnahme vorprogrammiert. Nicht nur bleibt unklar, worüber man sich mit dem Agenten austauschen kann, auch besteht die Möglichkeit, dass nicht zur Tastatur gegriffen wird, sondern ein Versuch der verbalen Kontaktaufnahme startet. Zudem kann bereits im Textfeld stehender Text auch kontrafaktisch, aber aus Perspektive der Nutzer\*innen plausibel, als Beitrag gedeutet werden, der von MAX produziert und an sie adressiert wurde.

Genau dies trifft auch für den Fall zu, dem wir uns im Weiteren widmen wollen. Dieser zeigt, dass und wie durch die Differenz von durch die Interface-Gestaltung nahegelegter intuitiver Bedienbarkeit und tatsächlicher Bedienbarkeit Probleme erzeugt werden, welche den Einstieg in eine ‚Interaktion‘ mit MAX erschweren. Zumindest implizit wird anhand der Fallanalyse auch ein weiteres Problem der Interface-Gestaltung deutlich (vgl. Abschnitt 4.2). Denn die Kamera, die neben der Leinwand montiert ist, dient MAX gar nicht dazu, seine Umwelt wahrzunehmen. Stattdessen erfüllt sie die Funktion auf Wunsch der Museumsbesucher\*innen Fotos aufzunehmen.

Während der laufenden Interaktion ist MAX dagegen ‚blind‘ und registriert lediglich schriftliche Eingaben, die er als Indikator für die Anwesenheit eines Gegenübers verarbeitet.

## Interaktionsanalyse

Die Ausgangssituation in diesem empirischen Fall sieht so aus, dass in dem Moment, in dem der Museumsbesucher sich MAX nähert, bereits die Worte „wie gehts dir“ in der oberen Zeile des Textfeldes auf der Leinwand stehen. Dies zeigt Abb. 4, die wir als Standbild einer Videoaufnahme entnommen haben, mit der die interessierende Begegnung aufgezeichnet wurde.



**Abb. 4:** Standbild der Leinwand vor Hinzutreten des Museumsbesuchers.

De facto handelt es sich bei den Worten „wie gehts dir“ um die letzte Äußerung, die zuvor von einer anderen Person, die nun nicht mehr vor Ort ist, per Tastatur eingegeben und abgeschickt wurde, was vermutlich darauf hinweist, dass die vorherige ‚Interaktion‘ nicht über die Begrüßungssequenz hinaus gekommen ist und dann abgebrochen wurde.<sup>42</sup>

---

<sup>42</sup> Unsere Untersuchungen im HNF weisen darauf hin, dass ein nicht unbeträchtlicher Teil der ‚Interaktionen‘ mit MAX – gemessen am Vorbild der zwischenmenschlichen Interaktion – nicht ordnungsgemäß zu Ende geführt, sondern vorher abgebrochen wird.

Ohne das nötige Kontextwissen über die tatsächliche Bedienbarkeit von MAX lassen sich diese Worte aber wie beschrieben durchaus plausibel als schriftlich realisierter Beitrag vonseiten des Agenten auf der Leinwand interpretieren. Und genau dies geschieht im vorliegenden Fall, wie das folgende Transkript zeigt. Dem Besucher haben wir den Namen ‚Didi‘ gegeben. In dem Moment, in dem das Transkript einsetzt, bewegt er sich in Richtung des Ausgangs aus dem Ausstellungsbereich, in dem MAX präsentiert wird. MAX selbst befindet sich links des Ausgangs, während auf der rechten Seite ein Roboterarm zu sehen ist, der als „Zeichenroboter Vincent“ auf Wunsch Porträts der Museumsbesucher\*innen anfertigt. Auf Letzteren ist zunächst auch Didis Aufmerksamkeit gerichtet, bis er nach links schaut und MAX erblickt. In diesem Moment setzt das Transkript ein, das einer Partiturdarstellung folgt und Tabellenform annimmt. Die simultan produzierten Aktivitäten sind untereinander in getrennten Zeilen notiert. Die Sequenzialität des multimodalen Geschehens wird deutlich, wenn die Tabellen von links nach rechts gelesen werden. Notiert sind in den Zeilen verschiedene Modalitäten des interaktiven Geschehens. Für Didi sind sowohl seine verbalen Äußerungen (Didi\_v) als auch seine Blickrichtung (Didi\_blick) und seine Bewegungen (Didi\_move) notiert, außerdem seine Eingaben auf der Tastatur (Didi\_tipp). Zudem gibt es eine Zeile, welche den Text notiert, der in der ersten Zeile des Textfeldes auf der Leinwand steht (Leinwand\_Text) sowie eine Zeile, die notiert, was MAX spricht (MAX\_v). Aus Gründen der Übersichtlichkeit sind in den jeweiligen Tabellen immer nur diejenigen Zeilen aufgeführt, die für die Analyse jeweils relevant sind.

Tab. 1: Ereignisabfolge der Interaktionen, Sequenzen 1–8.

	1	2	3	4	5	6	7	8
<b>Didi_v</b>				hallo MAX.			(2,7")	
<b>Didi_blick</b>	MAX	Stehtisch vor MAX		MAX			Stehtisch	MAX
<b>Didi_move</b>				wendet sich nach links und geht auf MAX zu				
<b>Leinwand_Text</b>	wie gehts dir							

Wie das Transkript dokumentiert, erblickt Didi zunächst MAX und den vor ihm befindlichen Stehtisch (Spalten 1 und 2) und bewegt sich dann auf diesen zu. Noch in der Bewegung entbietet Didi dann mit den Worten „hallo MAX“ einen verbalen Gruß (4) und vollzieht auf diese Weise (s)einen Teil einer Eröffnungssequenz einer Interaktion. Indem er ihn *anspricht*, markiert Didi zugleich, dass er annimmt, MAX könne ihn hören und verstehen. Der abschließende Blick auf MAX (8) lässt sich vor diesem Hintergrund als Signal der Turnübergabe interpretieren, das zugleich anzeigt, dass nun eine Reaktion seitens des Agenten erwartet werden kann. Wie der folgende Ausschnitt zeigt, bleibt diese jedoch aus.

Tab. 2: Ereignisabfolge der Interaktionen, Sequenzen 9–14.

	9	10	11	12	13	14
<b>Didi_v</b>				GU:T gehts.		
<b>Didi_blick</b>	MAX	Textfeld	MAX		Textfeld	nach links (Bildschirm auf anderem Stehtisch)
<b>Didi_move</b>		steht einen Schritt vor dem Stehtisch mit der Tastatur				
<b>Leinwand_Text</b>	wie gehts dir					

Vonseiten des Agenten erfolgt keinerlei sichtbare Reaktion. Didi, der inzwischen einen Schritt vor dem Stehtisch zum Halten gekommen ist (9), richtet seinen Blick auf das Textfeld (10) und äußert dann die Worte „GU:T geht’s“ (12), die man klar als Antwort auf die im Textfeld verbliebene Frage „wie geht’s dir“ verstehen kann. Offensichtlich interpretiert Didi den Text im Feld als Beitrag von MAX und eventuell auch als Reaktion auf seinen vorangegangenen Gruß. Kontrafaktisch, aber (wie in 3.1 ausgeführt) durchaus erwartbar bringt Didi damit die Annahme zum Ausdruck, MAX würde schriftsprachlich kommunizieren, könne aber hören. Dies bestätigt sich auch in der Blickbewegung, die sich zunächst vom Textfeld weg und hin zum verbal adressierten verkörperten Agenten bewegt (11), um sich noch während des Sprechens wieder dem Textfeld zuzuwenden (13), auf dem offenbar eine Antwort erwartet wird. Das Textfeld bleibt jedoch unverändert, so dass Didi beginnt, die weitere Umgebung des Artefakts in Augenschein zu nehmen (14). Dies kann als Suchbewegung gedeutet werden, mit der er nach Hinweisen Ausschau hält, die er als Ressourcen nutzen kann, um zu verstehen, was falsch läuft und warum eine Reaktion von MAX ausbleibt. Wie der folgende Transkriptausschnitt zeigt, ist dieser Versuch erfolgreich.

Tab. 3: Ereignisabfolge der Interaktionen, Sequenzen 15–18.

	15	16	17	18
<b>Didi_blick</b>	MAX	Hinweisschild	Tastatur	
<b>Didi_move</b>				tritt an den Stehtisch und bewegt die Hände Richtung Tastatur

Im Zuge der Suchbewegung landet Didis Blick auf dem Hinweisschild (16) und anschließend auf der Tastatur (17), was dazu führt, dass er noch näher an den Stehtisch herantritt und seine Hände aus den Hosentaschen nimmt (18). Auf diese Weise bereitet Didi einen ‚Medienwechsel‘ vor, mit dem er vom Medium des Mündlichen ins Medium des Schriftlichen wechselt. Denn nachdem er die Hände aus den Hosentaschen genommen hat, legt er sie auf die Tastatur und beginnt eine Eingabe (19).

Tab. 4: Ereignisabfolge der Interaktionen, Sequenzen 19–23.

	19	20	21	22	23
<b>Didi_blick</b>		<i>Textfeld</i>		<i>Tastatur</i>	<i>Textfeld</i>
<b>Didi_tipp</b>	g g u t		(3,3'')		

An dieser Stelle kann man nicht nur einen Medienwechsel beobachten, sondern damit verbunden auch einen Orientierungswechsel: Während sich Didi zuvor noch multimodal an einem gemeinsamen Wahrnehmungsraum mit MAX orientiert hat (Zuwenden, Anschauen, Ansprechen), orientiert er sich nun an einer (zumindest seinerseits) rein textbasierten Kommunikation. Sein Blick wechselt für ca. eine halbe Minute ausschließlich zwischen Textfeld und Tastatur hin und her und ignoriert den Agenten. Nachdem er die zusätzlichen Informationen über die Funktionsweise des Agenten zur Kenntnis genommen hat, passt er sein Verhalten also an die nun erkannten technischen Restriktionen der ‚Interaktion‘ mit MAX an.

Zunächst tippt Didi „g g u t“ (19), was als (fehlerhafte) eliptische Form einer Wiederholung des zuvor Gesagten gedeutet werden kann. Das heißt, Didi nimmt eine Korrektur seines bisherigen Verhaltens vor, indem er seine Antwort in eine für MAX wahrnehmbare mediale Form überführt. Allerdings schickt er die Antwort nicht ab, sondern blickt zunächst auf (20) und lässt seinen Blick dann über Tastatur und Textfeld wandern (22 & 23). Damit stoppt Didi zunächst sein aktuelles Vorhaben und es stellt sich die Frage, wie es weitergeht. Denkbar ist, dass er beim Blick aufs Textfeld (20) den Rechtschreibfehler erkennt und nun eine entsprechende Reparatur vorbereitet.

Tab. 5: Ereignisabfolge der Interaktionen, Sequenzen 24–30.

	24	25	26	27	28	29	30
<b>Didi_blick</b>	<i>Textfeld</i>		<i>Tastatur</i>	<i>Textfeld</i>			<i>Tastatur</i>
<b>Didi_tipp</b>	<<löscht> t u g>	(4,0'')			<<löscht> g>	(2,5'')	

Genau dies scheint zunächst auch der Fall zu sein. Denn Didi löscht die drei Buchstaben t, u und g (24), wodurch allein der Buchstabe g im Textfeld stehen bleibt. Dies spricht dafür, dass er eine Rechtschreibkorrektur plant und vorhat, das Wort „gut“ noch einmal korrekt einzugeben. Nach einer langen Pause von 4 Sekunden (25) löscht er jedoch auch den verbleibenden Buchstaben (28), woraufhin eine erneute Pause entsteht (29). Anzunehmen ist, dass er nun noch einmal neu ansetzt und nicht nur eine Rechtschreibkorrektur vornimmt, sondern auch seine Worte anders wählt.

Tab. 6: Ereignisabfolge der Interaktionen, Sequenzen 31–35.

	31	32	33	34	35
<b>Didi_blick</b>		Textfeld		Tastatur	
<b>Didi_tipp</b>	s e h r		(3,8'')		s c h l e c h t <Enter>

Dies geschieht tatsächlich. Denn Didi gibt nun die Worte „sehr schlecht“ ein (31 & 35), wobei zwischen der Eingabe des ersten und des zweiten Wortes wieder fast vier Sekunden Pause liegen (33), in denen er seinen Blick nacheinander auf Textfeld und Tastatur richtet (32 & 34) und zu überlegen scheint, wie er seinen mit dem Wort „sehr“ begonnenen Zug fortsetzen soll. In der wiederholten Unterbrechung zeigt Didi somit eine gewisse Unschlüssigkeit, möglicherweise auch Unsicherheit an, was er eingeben soll. Gleichzeitig weisen Didis Zögern und die langen Pausen darauf hin, dass er keinen praktischen Handlungsdruck, wie er in Face-to-Face-Situationen herrscht,<sup>43</sup> mehr verspürt, nachdem er erkannt hat, dass MAX ausschließlich über die Texteingabe adressierbar ist. Hatte Didi im Medium der Mündlichkeit noch spontan mit einer ganz üblichen Floskel („GU:T geht’s“ (12)) geantwortet, fallen ihm nun (ohne Zeitdruck) andere Fortsetzungsmöglichkeiten der Interaktion ein und er entscheidet sich schließlich dafür mit den ‚Normalerwartungen‘, wie sie in zwischenmenschlichen Begrüßungen gemeinhin gelten,<sup>44</sup> zu brechen. Eine Äußerung schlechten Befindens, zumal in der gesteigerten Variante „*sehr* schlecht“, ist unüblich und wäre eigentlich nur in Interaktionen zwischen miteinander vertrauten Personen erwartbar – und hier auch nur, wenn es der sich entsprechend äßernden Person auch wirklich schlecht ginge. Angesichts dessen lässt sich Didis korrigierte Antwort plausibel nur als explizite Distanzierung von entsprechenden Normalerwartungen deuten. Einerseits zeigt er hierdurch möglichen Zaungästen an, dass er sich nicht *ernsthaft* auf einen Austausch mit dem artifiziellen Gegenüber einlässt. Andererseits leitet er auf diese Weise aber auch einen Test der kommunikativen Fähigkeiten des Artefaktes ein. Denn es stellt sich nun die Frage, ob MAX den Normverstoß als solchen erkennt und wie er hierauf reagiert.

Tab. 7: Ereignisabfolge der Interaktionen, Sequenzen 36–37.

	36	37
<b>MAX_v</b>	(3,0'')	ich spüre negative schwingungen

Tatsächlich dauert es wiederum einen Moment, bis etwas passiert (36). Nach drei Sekunden antwortet MAX dann aber mit den Worten „ich spüre negative Schwin-

<sup>43</sup> Vgl. Garfinkel 1967, 12.

<sup>44</sup> Vgl. Sidnell 2010, 208.

gungen“ (37). Damit schließt er zwar inhaltlich an Didis Eingabe an, jedoch in hochgradig spezifischer Weise. Denn während (aus Didis Perspektive) die Worte „sehr schlecht“ in unernster Weise auf seinen allgemeinen Zustand referieren, bezieht sich MAX' Äußerung auf das Verhältnis zwischen Beiden, das atmosphärisch gestört sei. Dies könnte nun wiederum als humorige Reaktion auf Didis Beitrag interpretiert werden, mit der MAX anzeigt, dass er diesen Unernst verstanden hat und somit seine kommunikative Kompetenz beweist. Genauso ist es aber auch möglich, den Anschluss als unpassend zu deuten und entsprechend als Hinweis auf Limitationen der kommunikativen Fähigkeiten von MAX.

Menschliche User\*innen können nur spekulieren, welche der Lesarten angemessen ist und entsprechend testweise reagieren. Dagegen haben wir die Möglichkeit, die Black Box MAX zu öffnen und einen analytischen Blick auf die interne Interaktionsarchitektur des Agenten zu werfen. Dies werden wir nun mit dem ersten tatsächlich vom Agenten produzierten Beitrag auch tun, um zu zeigen, wie dieser funktioniert und welche Konsequenzen dies im Fortgang der Interaktion zeitigt.

## Analyse der ‚internen‘ Funktionsweise von MAX

Ergänzend zur Analyse des Interfaces von MAX widmet sich der zweite Teil der Analyse nun also der ‚Innenwelt‘ des Agenten und zeigt anhand der Untersuchung von Programmierskripten, die Aufschluss über konkrete Interpretations-, Klassifikations- und Assoziationsprozesse geben, wie diese analog zur Interfacegestaltung Möglichkeiten der Interaktion eröffnet und beschränkt.

Nutzer\*innen selbst können aus dieser Perspektive als ‚soziale Black Boxes‘ der technischen Black Box Max angenommen werden, die in Prozessen der „Verkomplizierung“<sup>45</sup> Stück für Stück parametrisiert und mit Mitteln der Motivations- und Entscheidungsstruktur des Agenten eingeordnet und durchleuchtet werden. Angesichts dieser technischen Strukturierung von sozialer Umwelt lassen sich dann auch Erkenntnisse darüber gewinnen, wie ein auf Grundlage von programmierlogisch angelegten Prozessen des Metrisierens, Kategorisierens und Klassifizierens ‚denkendes‘ und ‚handelndes‘ Artefakt Anschlusshorizonte der Interaktion mitbestimmt.

## Die interne Systemarchitektur von MAX

Im Wesentlichen folgt die Systemarchitektur von MAX dem sogenannten BDI-Konzept.<sup>46</sup> Das heißt, er verfügt über Annahmen (Beliefs) über die Welt (und sein Gegenüber) und besitzt Wünsche (desires) sowie Intentionen. Dabei bestimmt ein

<sup>45</sup> Bischof/Heidt 2018 im Anschluss an Latour 1996.

<sup>46</sup> Vgl. Huber 1999.



sogenannter ‚BDI interpretier‘ „anhand der vorliegenden Überzeugungen (*beliefs*), den aktuellen Wünschen und Zielen sowie den Handlungsoptionen eine aktuelle *Intention*“.<sup>47</sup> Wichtig ist hierbei, dass MAX „mehrere Wünsche (*desires*) haben [kann], von denen mit einer *utility function* das höchstbewertete ausgewählt wird, um zur aktuellen Intention zu werden“ und dass die angesprochenen Handlungsoptionen „in Form abstrakter Pläne vor[liegen], die durch Vorbedingungen, Kontextbedingungen, erreichbare Konsequenzen und eine Prioritätsfunktion beschrieben sind“.<sup>48</sup> Zu unterscheiden ist bei den Handlungsplänen zudem zwischen globalen (top level) Metaplänen und Plänen, die sich auf lokale (low level) Dialogziele beziehen. Der Agent kann also durchaus das globale Ziel verfolgen, den Namen seines Gegenübers herauszufinden und gleichzeitig im konkreten Dialog (low level) feststellen, dass der Nutzer im Moment kein Interesse daran hat seinen Namen zu nennen und entsprechend sein übergeordnetes Dialogziel zwischenzeitlich ‚hinten an stellen‘,<sup>49</sup> um auf einen geeigneten Moment zu warten, den Metaplan doch noch umzusetzen.<sup>50</sup>

Zusätzlich zu diesem ‚kognitiven‘ Apparat verfügt MAX über eine Emotionskomponente, die seinen ‚Gemütszustand‘ festlegt und ebenfalls Einfluss auf die Intentionbildung hat (Abb. 5). So kann der Agent auf Erwartungsenttäuschungen, die sich daraus ergeben, dass bestimmte (zunächst kognitiv aufgeworfene) Dialogziele nicht erreicht werden, mit Veränderungen des Emotionssystems reagieren, die bspw. sein Interesse abschwächen, die eigene Initiative aufrecht zu erhalten. Veränderungen der Emotionen werden zudem durch bestimmte Mimiken und Gestiken angezeigt, um so für MAX‘ Gegenüber sichtbar zu werden.

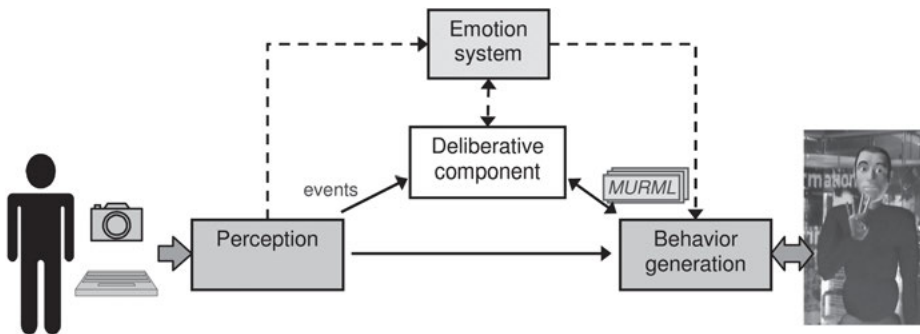


Abb. 5: Die Systemarchitektur von MAX.

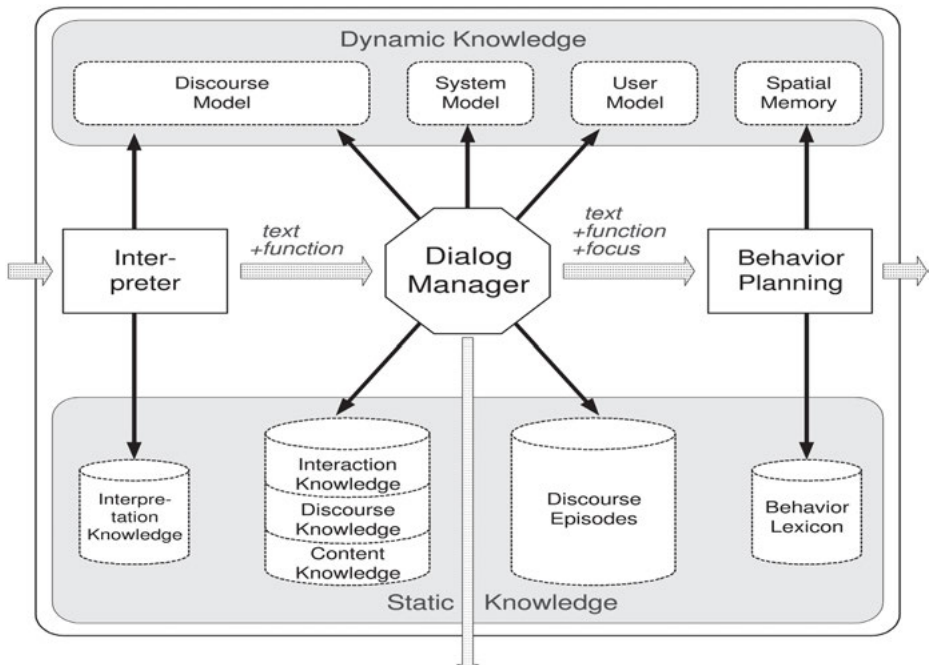
<sup>47</sup> Wachsmuth 2010, 149.

<sup>48</sup> Wachsmuth 2010, 149.

<sup>49</sup> Lotze 2016, 51.

<sup>50</sup> Kopp et al. 2005, 336f.

Wichtig ist zudem, dass MAX über statische und dynamische Wissensbestände verfügt, also einerseits Weltwissen bereithält, das ihm ‚einprogrammiert‘ wurde und andererseits dynamisches Wissen direkt aus Interaktionsverläufen ‚erwirtschaftet‘, das – zumindest wenn alles funktioniert – zwischengespeichert und mit neuen Gesprächen wieder gelöscht wird (Abb. 6). Im dynamischen Wissensbestand sind vor allem das ‚User Model‘ und das ‚Discourse Model‘ dafür verantwortlich, wie MAX die in der Konversation aufgeworfenen Informationen auf den Nutzer und die mit dem Nutzer verbundene Interaktionsgeschichte bezieht.



**Abb. 6:** Die interpretativen Komponenten von MAX' deliberativer Komponente.

Die Art und Weise, in der menschliche Kommunikationsteilnehmer\*innen in der Systemarchitektur ‚kognitiv repräsentiert‘ werden, ist dabei ‚listenförmig‘. MAX antizipiert sein Gegenüber als Person mit bestimmten Eigenschaften (z. B. in den Dimensionen Alter, Name, Geschlecht, Beruf, etc.) und versucht im Kommunikationsverlauf Dialogziele zu erreichen, die die einzelnen Dimensionen mit bestimmten Werten überschreiben. Gesprächspartner\*innen werden also als Ressourcen zur Beschreibung von leeren ‚Slots‘ im Inneren des Agenten mobilisiert.

Konkret arbeitet die Systemarchitektur dann so, dass sie einerseits entsprechend der eigenen Pläne proaktiv Äußerungen generiert und andererseits Äußerungen der Gesprächspartner\*innen verarbeitet, um diese 1) zu registrieren 2) zur Veränderung

der internen Zustände zu verwenden und 3) auf diese angemessen zu reagieren. Hierzu werden Äußerungen nach einem bestimmten Schema auf ihre kommunikativen Funktionen hin abgesucht, die auf drei Ebenen angesiedelt werden. Auf der ersten Ebene wird jeder Äußerung eine performative Funktion zugeschrieben, die entweder darin besteht Informationen anzubieten (provide) oder zu erbitten (askFor). Auf der zweiten Ebene wird spezifiziert, auf welchen Aspekt des laufenden Gespräches sich eine Äußerung bezieht. Differenziert wird hier zwischen dem Inhalt des Gespräches (content level), dem „interaction level“, auf dem es um die Etablierung, Fortsetzung und Beendigung des Gespräches geht und dem „discourse level“ auf dem das Themenmanagement in der laufenden Interaktion geklärt wird. Auf der dritten Ebene geht es bei der Bestimmung der kommunikativen Funktion einer Äußerung dann um den konkreten Inhalt einer Äußerung.<sup>51</sup>

Kommunikative Funktionen werden im System in der Form <performative>. <reference level>. <content>[arguments] abgelegt und gegebenenfalls mit zusätzlichen Argumenten in Klammern ergänzt. Eine Äußerung wie „Hallo“ kann dann zum Beispiel als „provide.interaction.greeting“ interpretiert werden, während eine Äußerung wie „Lass uns doch über Fussball reden“ als „askFor.discourse.topic.sports“ verarbeitet wird.<sup>52</sup> Je nach aktuellem Systemzustand, führt ein entsprechend interpretierter Input zur Anwendung des situativ passendsten Plans, der dann ausgeführt wird und einen entsprechenden Output generiert, dessen kommunikative Funktion gespeichert wird, um diese mit dem nächsten Input abzugleichen usw. So dynamisch die Outputs für MAX' Gegenüber damit auch sein können, so berechenbar bleibt aber die Logik der im Inneren des Systems ablaufenden ‚Testchecks‘ und Handlungskonsequenzen. Analog zu unseren Ausführungen weiter oben, legt also auch und gerade die Systemarchitektur der Innenwelt des Agenten bestimmte ‚Korridore‘ der Kommunikation nahe, über die überhaupt erst mit MAX gesprochen werden kann. Interaktionsmöglichkeiten werden also auch hier durch die Systemarchitektur gleichermaßen eröffnet und eingeschränkt – und zwar *gleichzeitig* zu den Affordanzen und Begrenzungen, die auf der Oberfläche und in der äußeren Umgebung des Artefaktes kommunikative Prozesse rahmen. Wie dies geschieht, zeigen wir im Fortgang der Interaktionsanalyse, in der wir die Betrachtung des interaktiven Geschehens zwischen MAX und Didi mit der Betrachtung der intern – gemäß der Systemarchitektur – ablaufenden Prozesse verschränken. Dabei starten wir mit der ersten für MAX registrierbaren Äußerung „sehr schlecht“.

---

51 Kopp et al. 2005, 335.

52 Kopp et al. 2005, 335.

## Fortsetzung der Interaktionsanalyse

user<sub>t</sub>

### 1 sehr schlecht

Mit der Eingabe „sehr schlecht“ (1) leistet Didi wie oben beschrieben einen Anschluss an eine dem Agenten zugeschriebene Frage, die dieser jedoch nie gestellt hat. Die Frage ist nun wie MAX mit dieser Eingabe innenweltlich umgeht und wie es zu seiner Äußerung „ich spüre negative Schwingungen“ kommt. Hierauf geben nun die von MAX aufgerufenen internen Prozesse Auskunft.

max<sub>c</sub>

```
2 p0129: 7.333333333333334 - curInput::!askFor(disliking) - smalltalk.
  disliking
```

Wie in dem Ausschnitt zu sehen ist, schreibt MAX dem aktuellen Input („curInput“) die kommunikative Funktion „askFor(disliking).smalltalk.disliking“ zu (2). Das heißt, der Äußerung „sehr schlecht“ wird intern zugeschrieben, Informationen anzufordern (verbunden mit dem optionalen Argument „disliking“, welches den Charakter der Anfrage spezifiziert), die sich auf das „content level“ beziehen<sup>53</sup> und inhaltlich Missfallen ausdrücken.

Kontrastiert man dies mit der kommunikativen Funktion, welche die Äußerung als Reaktion auf „wie gehts“ erfüllt, wird bereits deutlich, dass Max' Interpretation dieser zuwiderläuft, was auch Konsequenzen für das weitere Prozessieren besitzt. Denn entsprechend der zugeschriebenen Funktion wählt MAX (unter Berücksichtigung seines aktuellen Systemzustands) einen passend erscheinenden Handlungsplan aus. Dieser trägt den Namen „smalltalk.disliking“ und legt sowohl bestimmte Veränderungen des internen Zustands des Systems als auch einen Output mit bestimmter kommunikativer Funktion fest.

max<sub>c</sub>

```
3 <rule name="smalltalk.disliking" utility="-2">
4   <match>
5     <convfunction type="!askFor" modifier="disliking"/>
6   </match>
7   <action>
8     <command function="trigger-emotions" arguments="SPONTANEOUS
9     -20"/>
9     <random>
```

<sup>53</sup> Smalltalk wird auf der Ebene des Inhaltes der ‚Interaktionen‘ verortet (vgl. Kopp et al 2005, 337).

```

10         <act function="provide.discourse.disagree">Ich spuere
           negative Schwingungen.</act>
11         <act>Lass uns lieber ueber was sprechen was du magst.</act>
12     </random>
13     </action>
14 </rule>

```

So sinkt der Nützlichkeitswert des Handlungsplans um zwei Punkte, wenn er aktiviert wird (2), wodurch MAX' interne Prioritäten und Handlungsziele verändert werden. Zudem wird über die Kommandofunktion „trigger-emotions“ (8) ein bestimmter Gesichtsausdruck festgelegt, den Max annimmt, wenn der Plan exekutiert wird.<sup>54</sup> Auf der Handlungsebene wird darüber hinaus festgelegt, dass MAX mit einer Äußerung reagiert, welche die Funktion „provide.discourse.disagree“ (10) erfüllt. Die konkrete Äußerung wird dabei zufällig aus zwei möglichen Formulierungen ausgewählt (9–12) und dann ausgespielt.

$max_t$

15 Ich spuere negative Schwingungen.

$max_c$

```

16 <act function="provide.discourse.disagree" emphasis="none">Ich spuere
   negative Schwingungen.</act>

```

Im vorliegenden Fall ist dies die Antwort „Ich spuere negative Schwingungen“ (15) deren kommunikative Funktion anschließend gespeichert wird (16). Wie beschrieben, macht diese Antwort auf Didi's Eingabe nur begrenzt Sinn. Der Agent reagiert als gäbe es auf Ebene des Smalltalks sachliche Differenzen, obwohl Didi ja augenscheinlich eine wenn auch unernste Antwort auf die Frage nach seinem Gemütszustand gegeben hat. Bereits hier sieht man also wie sich ‚Innenwelt‘ des Agenten und kommunikative ‚Außenwelt‘ voneinander entkoppeln, wodurch in Echtzeit kommunikative Probleme erzeugt werden. Dies setzt sich im weiteren Verlauf fort. Denn nachdem MAX seinen Output als kurzfristige Reaktion auf Didis Eingabe produziert hat, aktiviert er ein übergeordnetes Ziel, das kurzzeitig in den Hintergrund gerückt ist.

$max_c$

```

17 g0016: get-name - hold-initiative - goals.user.getName
18 <rule name="goals.user.getName" utility="10">
19     <goal name="get-name" context="emptyslot name" />
20     <action function="take-initiative">

```

---

54 Vgl. hierzu Becker et al. 2004.

```

21     [...]
22 </action>
23 <action function="hold-initiative">
24     <switch var="$cycles">
25         <cond value="1">
26             [...]
27         </cond>
28         <cond value="2,3">
29             <random>
30                 <act>Jetzt aber weiter mit deinem Namen.</act>
31             </random>
32             <random>
33                 <block>
34                     <act>Du kannst Dir ja auch einen Namen ausdenken
35                     wenn Du Deinen nicht sagen willst.</act>
36                     <act function="askFor.content.name">Also?</act>
37                 </block>
38                 <act function="askFor.content.confirmation"> Willst
39                 Du ihn wirklich nicht sagen? Ich verrat es auch
40                 nicht weiter. Ok?</act>
41                 <act function="askFor.content.name">Wie hat Dich
42                 Deine Mutter frueher immer genannt?</act>
43                 <act function="askFor.content.confirmation">Hast Du
44                 vielleicht einen Spitznamen?</act>
45             </random>
46         </cond>
47     <else>
48         [...]
49     </else>
50 </switch>
51 </action>
52 <action function="resume-initiative">
53     [...]
54 </action>
55 </rule>

```

Max verfolgt während er seine Regeln abspult das globale Dialogziel, den Namen seines Gegenübers herausfinden zu wollen (17). Die „hold-initiative“ Information signalisiert, dass der Agent dieses Ziel beibehält, es also bereits vorher erreichen wollte aber zwischenzeitlich darauf umgestellt hat auf Eingaben des Nutzers zu reagieren, weil dieser ihm ein Interesse zur Initiativenübernahme signalisierte. Der Nutzer ist für ihn also nicht ‚neu‘, was bedeutet, dass MAX Informationen des Discourse-Models

aus der letzten Interaktion noch nicht gelöscht hat, in der er bereits mehrfach nach dem Namen des (anderen) Nutzers gefragt hat. MAX nimmt also an, dass er sich noch in Interaktion mit einer anderen Person befindet, was einerseits einen Hinweis darauf gibt, dass er die Kamera nicht nutzt, um zu registrieren, wann Personen den Raum vor dem Stehtisch betreten und verlassen und andererseits auch verstehbar macht, warum er zuvor Didis Eingangsäußerung nicht als Teil einer (unernsten) Begrüßungshandlung interpretiert hat, sondern als inhaltlichen Beitrag zu einem laufenden Smalltalk.

In der Konsequenz befindet sich MAX damit aber gleichsam ‚im falschen Film‘ und produziert weiter in der Form nicht erwartbare Anschlüsse. Denn infolge der Feststellung, dass er bereits mehrfach erfolglos versucht hat, den Namen des Users herauszufinden (28)<sup>55</sup> unternimmt er einen neuen Anlauf und produziert einen Output, der auf der Außenseite der Kommunikation offensichtlich keinen Sinn macht, da der Museumsbesucher gerade erst vor die Tastatur getreten ist. Intern speichert MAX zugleich ab, dass sein Beitrag die kommunikative Funktion „askFor.content.confirmation“ (37, 39) erfüllt, wodurch er in seiner internen Logik nicht mehr wirklich eine Frage stellt, die darauf abzielt eine Eingabe zu bekommen, die das Nutzermodell in der Dimension Name mit einem Wert überschreibt, sondern eine Ja/Nein-Frage, die konfirmiert oder abgelehnt werden kann.<sup>56</sup> Wie weiter unten zu sehen ist, hat dies Konsequenzen für den weiteren Verlauf des Geschehens.

max<sub>t</sub>

**51 Jetzt aber weiter mit deinem Namen.**

**52 Hast Du vielleicht einen Spitznamen?**

Lässt sich der erste Teil der Antwort des Agenten „Jetzt aber weiter mit deinem Namen [...]“ (51) gerade noch als unhöflicher erster Versuch interpretieren, den Namen des Gegenübers zu erfahren, fällt der zweite Teil „[...] Hast Du vielleicht einen Spitznamen?“ (52) als zu spezifisch für den bisherigen Konversationsverlauf auf. Während das Insistieren also in der internen Logik von MAX durchaus Sinn macht, da er sich in einer bereits länger anhaltenden Interaktion wähnt, gilt dies für Didi nicht, dem die internen Prozesse des Agenten nicht transparent sind. Dieser muss sich also fragen, wie es zu dieser Äußerung kommt und wie er darauf reagieren soll.

---

<sup>55</sup> Dies wird anhand der nachfolgend ausgespielten Reaktion deutlich. Denn diese wird aus möglichen Reaktionen generiert, die erfolgen, wenn MAX bereits zwei- oder dreimal versucht hat, den Namen seines Gegenübers zu erfahren (28).

<sup>56</sup> Nur eine der vier Antwortmöglichkeiten („Wie hat Dich Deine Mutter frueher immer genannt?“ (45)) korrespondiert mit dem Dialogue Act „askFor.content.name“, setzt also den für den weiteren Konversationsverlauf relevanten Kontext auf ‚Eingangstext als Namen des Gegenübers speichern‘. Wäre zufällig diese Möglichkeit ausgegeben und deren Funktion entsprechend gespeichert worden, wäre die weitere ‚Interaktion‘ womöglich anders verlaufen.

user<sub>t</sub>  
53 Didi

Der Nutzer sieht nun über die unerwartete Form der Frage nach seinem Namen hinweg und antwortet, indem er „didi“ eingibt (53). Dabei affirmiert er die Frage danach *ob* er einen Spitznamen hat, indirekt durch Mitteilung eben dieses Spitznamens und verzichtet darauf, die Frage explizit mit einem ‚ja‘ zu beantworten. Indem er ein als Spitznamen verstehbares Wort eingibt, beantwortet er also sowohl die Frage und gibt zugleich seinen (Spitz-)Namen preis.

Im Folgenden beginnt MAX mit der Interpretation von „didi“ (53), indem er den Input gegen unterschiedliche Merkmale testet, die der Agent mit seiner letzten Mitteilung im Kurzzeitgedächtnis hinterlegt hat. Was im zwischenmenschlichen Verkehr unproblematisch erscheint, stellt den Agenten MAX jedoch vor große Probleme, da er offensichtlich mit einer ja oder nein-Antwort rechnet und mit Abweichungen von diesem Schema nicht umgehen kann.

max<sub>c</sub>  
54 p0106: smalltalk.fallback.repeatAskFor

So registriert MAX offensichtlich, dass keine ‚confirmation‘ (positiv wie negativ) auf seine Entscheidungsfrage erfolgt ist, da ein eine ‚Fallback-Regel‘ ausgelöst wird (54), die darauf zielt, die eigene Anfrage zu wiederholen und hierfür erneut verschiedene Äußerungen vorhält, von denen eine zufällig ausgewählt wird (59–64).

```
15.02.2017 11:08, maxc
55 <rule name="smalltalk.fallback.repeatAskFor" utility="-15">
56 [...]
57   <action>
58     <command function="add-context" arguments="repeated-yn-
      question ultrashort"/>
59     <random>
60       <act function="$lastFunc">Was soll das denn heissen - Ja
      oder Nein?</act>
61       <act function="$lastFunc">Heisst das jetzt Ja oder
      Nein?</act>
62       <act function="$lastFunc">Ist das ein Ja oder ein
      Nein?</act>
63       <act function="$lastFunc">Was willst du damit sagen - Ja
      oder Nein?</act>
64     </random>
65   </action>
66 </rule>
```



Parallel zu diesem Ziel, das sich aufgrund von ‚low-level‘-Abwägungen anhand der konkreten, laufenden Interaktionsgeschichte etabliert, läuft aber auch das globale Metaziel – nämlich den Namen des Nutzers herausfinden zu wollen – weiter.

max<sub>c</sub>

67 g0016: get-name - hold-initiative - goals.user.getName

In der Konsequenz exekutiert MAX seine kurzfristigen und langfristigen Pläne direkt hintereinander und produziert gemäß der Vorgaben der Zufallsauswahl (vgl. 59–64 und 29–40) Output, der sich aus Äußerungen der „smalltalk.fallback.repeatAskFor“ Regel *und* des „user.getName“ Ziels zusammensetzt.

max<sub>t</sub>

**68 Was willst du damit sagen - Ja oder Nein? Jetzt aber weiter mit deinem Namen. Willst Du ihn wirklich nicht sagen? Ich verrat es auch nicht weiter. Ok?**

Diese Äußerungen behandeln Didis vorangegangene Eingabe nun – aus Perspektive der Kommunikation – kontrafaktisch weder als (eindeutige) Antwort auf die Frage, ob er einen Spitznamen hat noch als Preisgabe seines Namens. In der Operationslogik des Agenten macht dies zwar Sinn, dennoch können für ein solches Verhalten in der Konversation keinerlei legitime Gründe identifiziert werden, da Didi die eingeforderten Antworten aus seiner Perspektive ja bereits mit der Eingabe des Wortes „didi“ erbracht hat.

Kaum verwunderlich bricht Didi daraufhin die Interaktion ab und verlässt den Stehtisch, während MAX noch eine Weile fortfährt, ihn zu weiteren Eingaben zu überreden, bis er schließlich zu einem Zeitpunkt, an dem Didi sich schon länger nicht mehr im Raum befindet, eine Verabschiedungsfloskel von sich gibt. Deutlich wird damit, dass nicht nur die Interfacegestaltung kommunikative Probleme evoziert, sondern auch die Systemarchitektur des Agenten MAX. So fällt auf, dass der Agent nicht nur Merkmale vorangegangener Interaktionen in die Laufende hinein ‚verschleppt‘, sondern auch, dass zwischen (top level) Metaplänen und Unterzielen Synchronisationsprobleme auftreten, die auf den Konversationsverlauf ‚durchschlagen‘. MAX möchte den Namen seines Gegenübers herausfinden, geht aber (durch fehlende Selektivität des Gedächtnisses) davon aus, dass eine Selbstauskunft des Nutzers nicht mehr zu erwarten ist und versucht dann stattdessen eine Konfirmation (Ja/Nein-Antwort) auf die Frage nach der Existenz eines Spitznamens zu bekommen. Was er im Zuge dieser low level ‚Verengung‘ nicht mitbekommt ist, dass der User tatsächlich (s)einen Namen „didi“ offenbart. Dies verweist schließlich darauf, dass dem seine

Pläne abarbeitenden MAX die Fähigkeit fehlt mit der Vagheit und Indexikalität<sup>57</sup> sozialer Prozesse angemessen umzugehen. In der Konsequenz verhärtet sich die Konversationsstruktur derart, dass weder der Nutzer noch MAX weitere Wissensreservoirs mobilisieren können, die eine Fortführung des Gespräches begründen könnten.

## Schluss

Zieht man nun die Analyse von Interfacegestaltung, Systemarchitektur und sich vor diesem Hintergrund vollziehender ‚Interaktion‘ zusammen, wird deutlich, dass und wie die verschiedenen Ebenen der Interaktionsarchitektur je für sich aber auch in ihrem Zusammenspiel dazu beitragen, Möglichkeiten der Interaktion zu eröffnen und – im vorliegenden Fall vor allem auch zu beschränken und zu erschweren. Dies beginnt bereits damit, dass MAX’ Interface ‚auf den ersten Blick‘ mündliche Interaktion auf beiden Seiten zu ermöglichen scheint, was jedoch nicht der Fall ist und erst herausgefunden werden muss. Die Schwierigkeiten setzen sich damit fort, dass Text, der sich im Textfeld befindet plausibel aber dennoch falsch als Output von MAX gedeutet werden kann und im vorliegenden Fall auch gedeutet wird, wodurch sich ein Anschluss Didis ergibt, der für ihn Sinn macht, aber von MAX nicht angemessen verarbeitet werden kann, was Auslöser einer ganzen Kaskade an als missverständlich zu lesenden Äußerungen und Interpretationen ist. Rahmungsleistungen der Interaktionsarchitektur des ‚Interfaces‘ (genauer: des Textfeldes) und solche der Texterkennung in der Systemarchitektur treten entsprechend auseinander, da sie nicht hinreichend aufeinander abgestimmt sind, was zu den rekonstruierten kommunikativen Problemen führt.

Verallgemeinert man diese Einsichten, bestätigt sich zunächst, dass Räume, Interfaces und programmierte Dialogsysteme die Art und Weise (mit-)bestimmen, in der mit einem technischen Artefakt umgegangen werden kann. Eine solche Erkenntnis hat das Programm der Interaktionsarchitekturanalyse jedoch nicht exklusiv. Vielmehr ist diese Einsicht auch Ausgangspunkt und Ergebnis von Analysen, die Architekturen und Interfaces als Dispositive begreifen und in der Tradition der STS versuchen, hinter intuitiven Interfaces die machtvollen Entscheidungen und ‚Übersetzungsprozesse‘ zu beschreiben, welche erst zu funktionierender Technik führen. Die Besonderheit der Interaktionsarchitekturanalyse und mit ihr der Analyse medialer Affordanzen besteht aber darin, dass sie von vornherein darauf eingestellt ist, auch Widersprüchlichkeiten und Probleme in den Blick zu nehmen. Damit stellt sie in unseren Augen eine wichtige Analyseperspektive bereit, um Artefaktarrangements ‚in Aktion‘ zu untersuchen.

---

57 Vgl. Abels 2010, 131ff.

## Literaturverzeichnis

- Abels, Heinz (2010), *Interaktion, Identität, Präsentation. Kleine Einführung in interpretative Theorien der Soziologie*, Wiesbaden.
- Arminen, Ilkka/Licoppe, Christian/Spagnolli, Anna (2016), „Respecifying mediated interaction“, in: *Research on Language and Social Interaction* 49 (4), 290–309.
- Bearne, Eve/Kress, Gunther (2001), „Editorial“, *Reading, literacy and language* 35 (3), 89–93.
- Becker, Christian/Kopp, Stefan/Wachsmuth, Ipke (2004), „Simulating the emotion dynamics of a multimodal conversational agent“, in: *Workshop on Affective Dialogue Systems*, LNAI 3068, Berlin/Heidelberg, 154–165.
- Bischof, Andreas/Heidt, Michael (2018), „Die Verkomplizierung des Komplexen: Die Technisierung von Kommunikation in der Genese von Medientechnik“, in: Christian Katzenbach, Christian Pentzold, Sigrif Kannengießer, Marian Adolf u. Monika Taddicken (Hgg.), *Neue Komplexitäten für Kommunikationsforschung und Medienanalyse. Analytische Zugänge und empirische Studien* (Digital Communication Research 4), Berlin, 51–71.
- Garfinkel, Harold (1967), *Studies in Ethnomethodology*, Englewood Cliffs (NJ).
- Garfinkel, Harold/Wieder, Lawrence (1992), „Two incommensurable, asymmetrically alternate technologies of social analysis“, in: Graham Watson, Robert M. Seiler (Hgg.), *Text in context: Studies in ethnomethodology*, Newbury Park (CA), 175–206.
- Gesellensetter, Lars (2004), *Ein planbasiertes Dialogsystem für einen multimodalen Agenten mit Präsentationsfähigkeit* (Diplomarbeit, eingereicht an der Technischen Fakultät der Universität Bielefeld), Bielefeld.
- Gibson, James J. (1966), *The senses considered as perceptual systems*, Boston.
- Gnosa, Tanja (2018), „ANT goes dispositif: Überlegungen zu einer methodischen Verschränkung von Akteur-Netzwerk-Theorie und Dispositivanalyse“, in: *Le foucauldien*, 1 (4), 1–31.
- Goffman, Erving (1980 [1974]), *Rahmen-Analyse – Ein Versuch über die Organisation von Alltagserfahrungen* (suhrkamp taschenbuch wissenschaft 329), Frankfurt a. M. – ders. (1974), *Frame-Analysis. An Essay on the Organization of Experience*, New York.
- Goffman, Erving (2005), *Rede-Weisen. Formen der Kommunikation in sozialen Situationen*, Konstanz.
- Hausendorf, Heiko/Schmitt, Reinhold/Kesselheim, Wolfgang (2016), *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum* (Studien zur deutschen Sprache 72), Tübingen.
- Hausendorf, Heiko/Schmitt, Reinhold (2016a), „Interaktionsarchitektur und Sozialtopographie: Basiskonzepte einer interaktionistischen Raumanalyse“, in: Heiko Hausendorf, Reinhold Schmitt u. Wolfgang Kesselheim (Hgg.), *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum*, Tübingen, 27–54.
- Hausendorf, Heiko/Kesselheim, Wolfgang (2016b), „Die Lesbarkeit des Textes und die Benutzbarkeit der Architektur. Text- und Interaktionslinguistische Überlegungen zur Raumanalyse“, in: Heiko Hausendorf, Reinhold Schmitt u. Wolfgang Kesselheim (Hgg.), *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum*, Tübingen, 55–88.
- Hausendorf, Heiko/Schmitt, Reinhold (2016c), „Standbildanalyse als Interaktionsanalyse: Implikationen und Perspektiven“, in: Heiko Hausendorf, Reinhold Schmitt u. Wolfgang Kesselheim (Hgg.), *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum*, Tübingen, 161–188.
- Heath, Christian/vom Lehn, Dirk (2008), „Configuring ‚Interactivity‘: Enhancing Engagement in Science Centres and Museums“, in: *Social Studies of Science* 38 (1), 63–91.
- Huber, Marcus J. (1999), „JAM: A BDI-theoretic mobile agent architecture“, in: *Proceedings of the third annual conference on Autonomous Agents*, New York.
- Hutchby, Ian (2001), „Technologies, Texts and Affordances“, in: *Sociology* 35 (2), 441–456.
- Krummheuer, Antonia (2010), *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*, Stuttgart.

- Kopp, Stefan/Gesellensetter, Lars/Krämer, Nicole/Wachsmuth, Ipke (2005), „A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application“, in: Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier u. Thomas Rist (Hgg.), *Intelligent Virtual Agents*, Berlin/Heidelberg, 329–343.
- Latour, Bruno (1999), *Pandora's hope. Essays on the reality of science studies*, Cambridge (MA).
- Latour, Bruno (1996), „On Interobjectivity“, in: *Mind, Culture, and Activity* 3 (4), 228–245.
- Law, John (2013 [2009]), „Akteur-Netzwerk-Theorie und materiale Semiotik“, in: Tobias Conradi, Heike Derwanz u. Florian Muhle (Hg.), *Strukturentstehung durch Verflechtung. Akteur-Netzwerk-Theorie(n) und Automatismen* (Schriftenreihe des Graduiertenkollegs ‚Automatismen‘ 4), München, 21–48. – ders. (2009), „Actor Network Theory and Material Semiotics“, in: Bryan S. Turner (Hg.), *The New Blackwell Companion to Social Theory*, Malden (MA), 141–158.
- Lehn, Dirk vom (2006), „Die Kunst der Kunstbetrachtung: Aspekte einer pragmatischen Ästhetik in Kunstaustellungen“, in: *Soziale Welt* 57 (1), 83–99.
- Lotze, Netaya (2016), „Chatbots. Eine linguistische Analyse“, in: Jens Runkehl, Peter Schlobinski u. Torsten Siever (Hgg.), *Sprache. Medien. Innovationen* 9, Bern.
- Matthewman, Steve (2013), „Michel Foucault, Technology, and Actor-Network Theory“, in: *Techné. Research in Philosophy and Technology* 17 (2), 274–292.
- Pinch, Trevor J. (1992), „Opening Black Boxes: Science, Technology and Society“, in: *Social Studies of Science* 22 (3), 487–510.
- Pinch, Trevor J./Bijker, Wiebe E. (1984), „The Social Construction of Facts and Artefacts: or How the Sociology of Science and the Sociology of Technology might Benefit Each Other“, in: *Social Studies of Science* 14, 399–441.
- Seier, Andrea (2013), „Von der Intermedialität zur Intermaterialität, Akteur-Netzwerk-Theorie als ‚Übersetzung‘ post-essentialistischer Medienwissenschaft“, in: *Zeitschrift für Medien- und Kulturforschung* 2, 149–165.
- Sidnell, Jack (2010), *Conversation Analysis: An Introduction*, Oxford, Wiley-Blackwell.
- Star, Susan-Leigh (1990), „Power, Technology and the Phenomenology of Conventions: On being Allergic to Onions“, in: *The Sociological Review* 38 (1), 26–56.
- Wachsmuth, Ipke (2010), „Ich, Max‘ – Kommunikation mit Künstlicher Intelligenz“, in: Tilmann Sutter, Alexander Mehler (Hgg.), *Medienwandel als Wandel von Interaktionsformen*, Wiesbaden.
- Zillien, Nicole (2009), „Die (Wieder)Entdeckung der Medien. Das Affordanzkonzept in der Mediensoziologie“, *Sociologia Internationalis* 46 (2), 161–181.

## Bildnachweise

Abb. 1–4: Bilder aus eigenem Datenbestand.

Abb. 5–6: Kopp et al. 2005, 332f.