

Empirische Befunde zur Wirkung von Assistenz – eine Literaturrecherche

Erik Pischel

25. Juni 2001

In dieser Arbeit stelle ich dreizehn Studien vor, in denen empirisch die Wirkung von Assistenzsystemen auf den Nutzer untersucht wird.

Inhaltsverzeichnis

1	Einleitung	2
2	AMANT und DULBERG: Navigation und Assistenz	2
3	SPROULL et al.: Ein Gesicht als Interface	4
4	DIETZ und LANG: Effektive Agenten	7
5	MOON und NASS: Adaptive Agenten und Persönlichkeit	9
6	KING und OHYA: Agentenrepräsentationen	10
7	VAN MULKEN et al.: Der Persona Effekt	12
8	KODA und MAES: Agenten mit Gesichtern	14
9	CASELL und THÓRISSON: Die Macht eines Nickens und eines Blickes	15
10	CASELL und VILHJÁLMSOON: Autonomes kommunikatives Verhalten	18
11	Untersuchungen zum Anthropomorphismus in der Sekundärliteratur	21
	Literaturverzeichnis	23

1 Einleitung

Assistenten und Agenten bilden seit einigen Jahren einen Schwerpunkt innerhalb der Softwareergonomie-Forschung. Welchen Effekt haben verschiedene Formen von Assistenz auf den Nutzer? Dazu gab es eine Reihe von Experimenten. Die Ergebnisse einiger dieser Studien möchte ich hier vorstellen. Dabei werde ich zunächst die die Autoren interessierende Fragen, dann das Experiment und schließlich die Ergebnisse anführen.

2 AMANT und DULBERG: Navigation und Assistenz

Navigation ist die Domäne des Experimentes von AMANT und DULBERG (Amant & Dulberg, 1998). Die Metapher der Navigation begegnet uns in vielen Bereichen: beim Durchsuchen von Informationsräumen (z.B. World Wide Web), Virtual reality Umgebungen und Entscheidungsräumen ebenso wie beim Wissens- und Prozessmanagement. Navigation löst das Problem nicht direkt, sondern setzt den nötigen Kontext für andere, effizientere Operationen.

2.1 Untersuchte Fragen

Amant und Dulberg entwickelten einen Assistenten, der Vorschläge für das Finden eines gesuchten Knotens in einem Baum gibt. Sie fragten sich, inwieweit die Aufgabenbewältigung von der *Qualität* des Assistenten abhängt.

Als Qualität eines Assistenten bezeichnen sie *Präzision* und *recall*. Um diese beiden Begriffe zu erklären, nutzen sie eine Analogie mit Informationssystemen. Präzision ist hier der relevante Teil der Ergebnismenge eine Abfrage, Recall das Verhältnis von relevanten und gefundenen Informationseinheiten zu der Gesamtmenge der relevanten Informationseinheiten. Der Unterschied scheint zunächst subtil, ein Beispiel soll ihn daher verdeutlichen: wenn bei einer Informationsabfrage 17 von 20 Ergebnissen relevant sind, ist das eine gute Präzision, bei drei von 20 eine schlechte. Wenn aber im ersten Fall insgesamt 100 relevante Datensätze in der Datenbank enthalten sind, ist die Zahl von 17 gelieferten ungenügend, der Recall ist gering. Auf die intelligente Assistenz angewendet, heißt dies in etwa: Präzision ist der Anteil der Ratschläge oder Aktionen des Assistenten, die zur Lösung der Aufgabe geeignet sind; Recall ist der Anteil der geeigneten Ratschläge oder Aktionen an der Gesamtheit der Alternativen.

Amant und Dulberg stellten nun folgende Fragen:

- Ist irgendeine Assistenz besser als keine?
- Ändert sich die Aufgaben-Performanz mit der Präzision der Assistenz?
- Ändert sich die Aufgaben-Performanz mit der Recall der Assistenz?
- Welcher der vorhandenen Effekte ist stärker? Sind sie unabhängig?

2.2 Das Experiment

Die 23 Probanden (Studenten eines HCI-Kurses) wurden gebeten, in einem Baum einen Knoten mit bestimmten Eigenschaften zu suchen. Jeder Knoten hatte eine Zahl, eine Form und

einen Buchstaben als Eigenschaft. Zu Beginn war nur die Wurzel des Baumes sichtbar. Durch Anklicken eines Knotens wurden seine Kinder aufgedeckt. Durch diese Art der Navigation mußte der gesuchte Knoten gefunden werden. Tiefe des Baumes war begrenzt, und die Eigenschaften der Kinder-Knoten ähnelten der des Eltern-Knotens. Der Assistent machte Vorschläge für den Pfad zum gesuchten Knoten, als textuelle Liste und durch Einfärbung der vorgeschlagenen Knoten.

Die Probanden wurden eingewiesen und konnte sich in einer Übungsphase (ohne Assistenten) mit dem System vertraut machen. Diese vereinfachte Situation beinhaltet nach Meinung von Amant und Dulberg wichtige Interaktionselemente eines Nutzers mit einem Assistenten:

- Nutzerwissen über die Domäne, das ausreicht, um die Aufgabe zu bewältigen. Dies ist jedoch einfacher mit Assistenz als ohne.
- Der Assistent ist nicht perfekt. Er gibt manchmal nützliche, manchmal irreführende Ratschläge.
- Der Nutzer muss entscheiden, ob er den Rat des Assistenten befolgt oder nicht.
- Die Performanz bei der Aufgabe hängt zum Teil vom Nutzer, zum Teil vom Assistenten ab.

Um die Qualität des Assistenten zu beeinflussen, machten die Autoren ihn zwar allwissend, aber fehlerhaft: aus der Menge der relevanten Knoten wurde ein Teil gegen irrelevante Knoten ausgetauscht. Der verbleibende Anteil an relevanten Knoten p bezeichnen die Autoren als *Präzision*. Der *Umfang* (im Original: scope) ist die Zahl der Ratschläge. Hier variierten die Autoren zwischen lokalem und globalem Umfang. Im lokalen Fall betreffenden die Ratschläge nur Knoten in näherer Umgebung der aktuellen Suche, im globalen Fall alle Knoten, die zum Zielknoten führen. Umfang ist lt. den Autoren ein schwacher Ersatz für Recall.

Präzision und Umfang sind die *unabhängigen Variablen*. Amant und Dulberg erstellten fünf Versuchsbedingungen nach der Varianzanalyse. Jeder Proband war diesen Bedingungen ausgesetzt, jedoch in unterschiedlicher Reihenfolge. Die Bedingungen sind in Tabelle 1 dargestellt.

Bedingung	Umfang	Präzision
Bedingung 0	—	—
Bedingung 1	global	0,25
Bedingung 2	lokal	0,25
Bedingung 3	global	0,75
Bedingung 4	lokal	0,75

Tabelle 1: Versuchsbedingungen bei Amant und Dulberg

Die *abhängigen Variablen* waren:

- Selektionen: Die Anzahl der expandierten Knoten,

- Dauer: Die Dauer des Versuchs.

und wurden durch Auswertung einer Aufzeichnung der Nutzer-Aktionen gewonnen. Sie sollten die Performanz der Aufgabe messen.

2.3 Ergebnisse

Die erhobenen Daten belegen, dass die Bedingung 0 (ohne Assistenz) wesentlich schlechtere Performanzergebnisse bedingte als die Bedingungen mit Assistenz: im Mittel wurden 47 mehr Selektionen und $7\frac{1}{2}$ Minuten mehr Zeit benötigt. Bei nur einem Probanden war die Leistung ohne Assistenz besser als mit Assistenz.

Bedingung	Selektion	Dauer	Std.abweichung
ohne Assistenz	65,0	503,3	7,7
global / 0,25	21,4	104,8	4,9
lokal / 0,25	23,5	122,0	5,2
global / 0,75	13,7	55,9	4,1
lokal / 0,75	11,3	44,3	3,9

Tabelle 2: Ergebnisse d. Experiments von Amant und Dulberg

Die statistische Auswertung der in Tabelle 2 gezeigten Ergebnisse ergab, daß sowohl Selektion als auch Dauer ausschließlich von der Präzision abhängen. Der Umfang (als Ersatz für Recall) hat keinen Effekt. Es gibt keine gegenseitige Abhängigkeiten. Die anderen kontrollierten Faktoren (verschiedene Probanden, Abfolge der Versuchsanordnungen, Problemgröße und -komplexität) zeigten keinen Effekt.

Wenn die Präzision gering ist, erhöht ein lokaler Umfang jedoch die Performanz.

3 SPROULL et al.: Ein Gesicht als Interface

Sproull et al. haben die Verhaltensunterschiede des Nutzers bzgl. eines textuellen Interfaces vs. eines Interface mit einem realistischen Gesicht in einer größeren Studie (Sproull et al., 1996) untersucht. Ihr Ziel war es, diese Unterschiede aufzuzeigen. Dabei ging es sowohl um Unterschiede in der Art, wie das System wahrgenommen und bewertet wird, als auch wie sich der Benutzer unter Eindruck dieses Systems verhält.

3.1 Untersuchte Fragen

Ausgehend von Erkenntnissen über Reaktionen auf Gesichter stellten Sproull et al. folgenden Hypthesen auf:

1. Die Probanden werden Persönlichkeitsmerkmale des Gesichtsinterfaces, die mit dem Aussehen verknüpft werden, anders attributieren, als die des Textinterfaces. Der Unterschied

bei Persönlichkeitsmerkmalen, die nicht mit dem Aussehen verknüpft werden, wird geringer ausfallen.

2. Die Probanden werden erregter und aufmerksamer sein, wenn sie mit dem Gesichtsin-terface interagieren, als wenn sie dies mit dem Textinterface tun.
3. Die Probanden werden sich selbst positiver beschreiben und sich weniger eröffnen, wenn sie mit dem Gesichtsin-terface interagieren, als wenn sie dies mit dem Textinterface tun.
4. Männer und Frauen werden sich in ihren Antworten auf ein Gesichtsin-terface stärker unterscheiden, als in ihren Antworten auf das Textdisplay.

3.2 Das Experiment

Das System, an welchem die Probanden sich versuchten, wurde als virtueller Karriere-Berater ausgegeben. Das erlaubte den Autoren, eine ganze Reihe von psychologischen Testitems darzu-bieten und diese zur Auswertung heranzuziehen.

Die Studie umfaßte 130 Studenten der Bostoner Universität mit einem Durchschnittsalter von 20,7 Jahren. Die Zahl der Männer und Frauen war ungefähr gleich.

Es wurden drei Versuchsbedingungen geschaffen. Einmal war es ein textuelles Interface, mit dem die Probanden interagierten. In den anderen beiden Bedingungen war es ein realistisch aussehendes, weibliches Gesicht. Dazu kam eine synthetische Stimme, wobei das Gesicht die Lippen synchron bewegte. Die beiden Bedingungen unterschieden sich in dem Gesichtsausdruck in den Gesprächspausen: der eine war ein ernster (die Mundwinkel und Augenbrauen leicht nach unten gezogen), der andere ein freundlicher (Mundwinkel und Augenbrauen nach oben gezogen).

Im Versuch konnten sich die Probanden nach eine Einführung zunächst an die Bedienung des Systems gewöhnen. Danach wurden vom System einige Standard-Tests durchgeführt: bei den Gesichtsin-terfaces wurden die Items gesprochen, und nur die Antworten textuell dargestellt. Im zweiten Teil interagierten die Probanden mit dem Berater (dazu wurde WEIZENBAUMS ELIZA-Programm modifiziert). Hier sprach das Gesicht die Worte des Berates, die jedoch zusätzlich nach einer Sekunde als Text dargestellt wurden. Im Anschluß füllten die Probanden einen post-experimentellen Fragebogen aus.

3.3 Ergebnisse

Soziale Wahrnehmung des Fragenden Wie Tabelle 3 (nach Sproull *et al.*, 1996) zeigt, be-werteten die Probanden die Persönlicheitsmerkmale des Fragenden (d.h. des Agenten) je nach Versuchsbedingungen unterschiedlich. Die Tabelle gibt an, in welcher Richtung die Merkmale der Gesichtsin-terfaces signifikant anders eingeschätzt wurden als die des Textinterfaces. Die Unterteilung in Merkmale, die mit dem Aussehen verbunden sind, und solcher, die nicht damit verbunden sind, übernahmen Sproull *et al.* von WARNER und SUGARMAN¹.

¹im englischen Original heißen die Merkmale *social evaluation, sociability, intellectual evaluation, potency, acti-vity* und *emotionality*.

Wahrnehmung	freundliches Gesicht	ernstes Gesicht
Mit Erscheinung verbundene Merkmale		
Soziale Bewertung	negativer	negativer
Geselligkeit	—	weniger wahr
Intellektuelle Bewertung	—	—
Nicht mit Erscheinung verbundene Merkmale		
Potenz	—	—
Aktivität	weniger wahr	weniger wahr
Emotionalität	—	—

Tabelle 3: Bewertung der Persönlichkeitattribute bei SPROULL et al.

Die Autoren sehen in den Ergebnissen die Bestätigung ihrer ersten Hypothese. Die paradox scheinende negativere Bewertung der Aktivität der animierten Interfaces mag in dem vorliegenden Gesicht oder in der langsamen, unbetont sprechenden und synthetischen Stimme resultieren.

Erregung und Aufmerksamkeit Die Antworten auf post-experimentelle Fragen nach der Erfahrung der Probanden mit dem System beinhaltete laut Faktorenanalyse drei Faktoren. Einer davon, den die Autoren mit *Happiness* bezeichnen, schließt auch die Zufriedenheit des Benutzers mit dem System ein. Hier konnten keine signifikanten Unterschiede zwischen den Versuchsbedingungen gemessen werden.

Ein anderer Faktor, *Erregung* genannt, zeigt Unterschiede auf: bei den Gesichtsisnterfaces beschrieben sich die Probanden weniger entspannt und weniger selbstsicher.

Probanden, die mit den Gesichtsisnterfaces arbeiteten, brauchten mehr Zeit zum Beantworten der Fragen der psychologischen Tests, als die Probanden mit dem Textinterfaces. Obwohl sie alternative Erklärungen nicht ausschließen können, schließen die Autoren daraus, daß die Probanden die zusätzliche Zeit zum sorgfältigerem Nachdenken über die Fragen nutzten und daher dem Experiment eine höhere Aufmerksamkeit schenkten. Zusätzlich zeigte sich, daß die Versuchspersonen in den Bedingungen mit Gesichtsisnterface signifikant mehr Fragen übersprangen als die Versuchspersonen in den anderen Bedingungen, aber nur in den Tests, die bekanntermaßen sozial beeinflussbarer sind (soziale Erwünschtheit und Altruismus). Dies kann bedeuten, daß die Personen gewisse persönliche Fragen vermieden.

Die Autoren sehen sich aufgrund dieser Ergebnisse in ihrer Hypothese 2 bestätigt. Die Antworten sind in Tabelle 4 (nach Sproull et al., 1996) zusammengefaßt.

Selbst-Präsentation Die Probanden, die die Gesichtsisnterfaces benutzten, stellten sich selbst positiver bezüglich Altruismus und sozialer Erwünschtheit dar. Diese beiden Probandengruppen schrieben auch weniger in ihren Antworten auf die Fragen des Beraters. Durch diese Ergebnisse sehen die Autoren ihre Hypothese 3 gestützt.

Antwort	freundliches Gesicht	ernstes Gesicht
Erregung	größer	größer
Performanz bei Selbst-Präsentation		
Dauer	größer	größer
Übersprungene Items gesamt	—	—
Altruismus	mehr	—
Soziale Erwünschtheit	mehr	—
Selbst-Wert	—	—
Selbst-Präsentation		
Altruismus	größer	größer
Soziale Erwünschtheit	—	—
Selbstwert	—	—
Interaktion mit dem Berater		
Dauer	—	—
Anzahl Wörter	—	kleiner
Anzahl der Interaktionsrunden	—	—

Tabelle 4: Verhaltensantworten bei SPROULL et al.

Unterschiede zwischen Männer und Frauen Die Autoren interessierten sich auch für die Interaktion zwischen Geschlecht und Versuchsbedingung (siehe Hypothese 4). Während sich sowohl Männer als auch Frauen altruistischer zeigten, wenn sie mit einem Gesichtsinterface interagierten, waren aber die Unterschiede zwischen den Geschlechtern nur bei dem freundlichen Gesicht groß. Frauen schrieben mehr Worte, wenn sie ein Textinterface benutzten. Bei den Männern verhielt es sich genau umgekehrt. Im allgemeinen antworteten die Männer positiver auf die Gesichtsdisplays, wogegen die Frauen positiver auf das Textdisplay antworteten. Auch mochten die Männer die Gesichtsdisplays mehr, die Frauen dagegen das Textdisplay. Dieses Ergebnis könnte mit dem noch zu unnatürlich wirkenden Gesicht und der synthetischen Sprache der Gesichtsinterfaces zusammenhängen. Frauen könnten dem abneigend gegenüberstehen, während Männer davon eher technisch fasziniert sind, so die Autoren.

4 DIETZ und LANG: *Äffektive Agenten*

Der in diesem Experiment (Dietz & Lang, 1999) genutzte Agent *b’Nex* gehört zum Polara Projekt. „The Polara Project is an effort to create an interactive multimedia story-world that takes the user on an interactive journey to the stars.“ (Dietz & Lang, 1999) *b’Nex* ist eine Comic-Figur und führt den Benutzer durch die Polara-Welt. Die Autoren beschloßen, den Agenten mit „selbst“-generierten Emotionen auszustatten. Mit dem Experiment versuchten die sie herauszufinden, welchen Einfluß dies auf die Benutzer hat.

DIETZ und LANG orientierten sich beim Begriff *Emotion* an den dimensional Theorien. Nach diesen gibt es drei Dimensionen, mit denen sich die jede Emotion beschreiben läßt: Valenz

(traurig – glücklich), Erregung (ruhig – aufgeregt) und Dominanz (kontrolliert – unkontrolliert).

Für ihren Agenten entwarfen die Autoren nun das sogenannte AVC Modell, welches drei Komponenten beinhaltet: Temperament, Stimmung und Emotion. Diese drei Komponenten haben Parameter in den genannten drei Dimensionen. Das Temperament ist statisch, Stimmung und Emotionen des Agenten können sich nach einem bestimmten Modell ändern.

Der Körperausdruck des Agenten soll dem emotionalen Zustand des Agenten entsprechen. Dazu wird jedes „Körperteil“ einer oder zwei emotionalen Dimensionen zugeordnet, z.B. die Form der Augenbraue der Valenz.

4.1 Untersuchte Fragen

DIETZ und LANG stellten drei Hypothesen auf:

1. Die Probanden werden berichten, mehr Emotionen gehabt zu haben, wenn sie mit dem emotionalen Agent interagieren, als wenn sie mit dem Agenten ohne Emotionen interagieren.
2. Die Probanden werden den emotionalen Agenten als menschlicher, freundlicher und „better to work with“ (Dietz & Lang, 1999) bewerten, als den nicht-emotionalen Agenten.
3. Die Probanden werden bei einer Gedächtnis-Aufgabe besser abschneiden, wenn sie mit dem emotionalen Agenten arbeiten, als wenn sie mit dem nicht-emotionalen Agenten arbeiten würden.

4.2 Das Experiment

Abhängig von den Aktionen des Benutzers und dem momentanen Zustand ändert der emotionale Agent seinen Körperausdruck und sein Aktivitätsniveau. Das Experiment besaß zwei unabhängige Variablen:

- Aufgabe: es gab eine Aufgabe, bei der die Probanden ein Interview mit emotionalen Items durchführten, und eine Lern-Aufgabe;
- Emotion: der Agent besaß einmal Emotionen, ein anderes Mal nicht.

Zunächst wurden die Probanden vom Agenten interviewt. In der Bedingung mit Emotion, näherte der Agent seine Stimmung der an, die er vom Probanden mitgeteilt bekam. In der nicht-emotionalen Bedingung ändert der Agent seinen emotionalen Zustand nicht. Danach bewerteten die Probanden die Persönlichkeit, Freundlichkeit und Funktionalität des Agenten.

Im zweiten Teil des Experiments führten die Probanden eine Gedächtnisaufgabe mit Hilfe des Agenten durch. Der Agent lehrte ihnen eine Liste von Fakten und fragte diese später ab. Der emotionale Agent erzeugte eine emotionale Rückmeldung zusätzlich zur Information, ob das Item korrekt war. Danach bewerteten die Probanden wieder Funktionalität und Persönlichkeit des Agenten.

Über den Umfang Stichprobe sagen die Autoren nichts. Aus den Daten entnehme ich, daß es vermutlich 80 Probanden waren.

4.3 Ergebnisse

Nach der erste Hypothese müßten die Probanden von mehr Gefühlen in der emotionalen Bedingung berichten. Dies wurde durch die eigenen Bewertung der Probanden nach einer Phase des Experimentes gemessen. Die Daten bestätigen diese Hypothese für beide Aufgabenbedingungen.

Nach der zweiten Hypothese müßten die Probanden den emotionalen Agenten günstiger bewerten. Dies geschah in keiner der beiden Aufgabenbedingungen statistisch signifikant. Nach der Lernaufgabe bewerteten die Probanden den emotionalen Agenten in allen untersuchten Dimensionen günstiger als den nicht-emotionalen Agenten, bei der Interview-Aufgabe jedoch bewerteten sie den nicht-emotionalen Agenten in der Mehrzahl der Dimension günstiger. Die Hypothese wurde nicht bestätigt.

Nach der dritten Hypothese müßten die Probanden bei der Lernaufgabe mit dem emotionalen Agenten besser abschneiden als mit dem nicht-emotionalen Agenten. Die absoluten Werte der Resultate zeigen in diese Richtung, jedoch ist der Effekt nicht statistisch signifikant.

Dietz und Lang finden die Resultate vorsichtig ermutigend. Obwohl sie in der Mehrzahl nicht statistisch signifikant waren, zeigten sie im wesentlichen in die erwartete Richtung. Aufgrund der Beschränkungen solle man aus ihnen nicht schließen, daß Emotionen bei Agenten kaum Effekte auf den Nutzer haben.

5 MOON und NASS: Adaptive Agenten und Persönlichkeit

Agenten sollen adaptiv sein, ist eine häufige Forderung. Doch was das genau heißt, bleibt oft unklar. Was bedeutet Adaptivität für die Persönlichkeit? Moon und Nass haben dies in einem Experiment (Moon & Nass, 1996) untersucht.

Die Autoren gehen von der sozialpsychologischen Regel des „Gesetzes der Anziehung“ aus. Einer Studie² von NASS et al. zufolge soll dies auch für die Menschen und Computer zutreffen: „[...] dominant people prefer to interact with dominant computers; submissive people prefer to interact with submissive computers.“ (Moon & Nass, 1996). Nun fragten sich Sproull et al., welche Effekte eine Persönlichkeitsänderung des Agenten auf den Nutzer hat.

5.1 Untersuchte Fragen

Ausgehend von der Gain-Loss-Theorie³ sagen die Autoren voraus,

- daß Menschen sicher eher zu Individuen hingezogen fühlen, die sich von einem entgegengesetzten Persönlichkeitsverhalten zum gleichen Persönlichkeitsverhalten ändern, verglichen mit Individuen, die die ganze Zeit über ein gleiches Persönlichkeitsverhalten aufweisen,
- daß Menschen eine Veränderung vom entgegengesetzten Persönlichkeitsverhalten zum gleichen Verhalten vorziehen gegenüber der entgegengesetzten Richtung.

²in Nass, Moon, Reeves und Dryer (1995)

³in Aronson und Linder (1965)

Die Frage für MOON und NASS ist, ob diese Voraussagen auch für Mensch-Computer-Beziehungen gelten.

5.2 Das Experiment

44 als dominant und 44 als unterwürfig⁴ Probanden wurden zufällig auf vier Versuchsbedingungen aufgeteilt. Die Probanden mußten in zwei Runden eine Aufgabe erfüllen. In jeder Runde war der Computer entweder dominant oder unterwürfig. Daraus ergeben sich die vier Versuchsbedingungen: (für die beiden Interaktionsrunden) dominant–dominant, unterwürfig–dominant, dominant–unterwürfig und unterwürfig–unterwürfig.

Die Unterschiede zwischen dominantem und unterwürfigem Verhalten des Computers bestanden nur in sprachindizierten Hinweisen und Vertrauensstufen.

Nach den beiden Interaktionsrunden wurden die Probanden mittels eines Fragebogens nach einer Bewertung des Computers und der Interaktion befragt. Die abhängigen Variablen waren dabei die Indizes soziale Anziehung, intellektuelle Anziehung, Nützlichkeit und emotionale Zufriedenheit.

5.3 Ergebnisse

Die Ergebnisse bestätigen die Vorhersagen. In der Bedingung ungleich–gleich wurden verglichen mit der Bedingung gleich–gleich signifikant höhere Werte bei allen vier abhängigen Variablen erzielt. Im Vergleich zur Bedingung gleich–ungleich wurden in allen Variablen außer „soziale Anziehung“ signifikant höhere Werte erzielt.

Moon und Nass sprechen von drei Ergebnisse:

1. Nutzer mögen Computer mit ihrem Persönlichkeitstyp.
2. Nutzer ziehen adaptive Computer denen vor, die über die Zeit konstant gleich bleiben.
3. Die bevorzugte Richtung der adaptive Änderung ist die des Persönlichkeitstyps des Nutzer.

6 KING und OHYA: Agentenrepräsentationen

In ihrem Experiment verglichen King und Ohya (1996) Agentenrepräsentationen, ohne jedoch dem Nutzer Funktionalität anzubieten. Dies gilt es, bei der folgender Darstellung zu beachten.

6.1 Untersuchte Fragen

Die Autoren fragten sich, wie die Repräsentation von Agenten die von Nutzern wahrgenommene Eigenschaften beeinflussen. Insbesondere sei die Wahrnehmung von „Vertrauen“ für den Nutzer wichtig.

⁴engl. *submissive*

6.2 Das Experiment

An dem Experiment nahmen 18 erwachsene Probanden teil. Sie wurden so ausgewählt, daß eine gewisse Bandbreite an Erfahrung und Verständnis von Computerbenutzung vorhanden war. Weiterhin waren die eine Hälfte der Versuchspersonen männlich, die andere weiblich. Eine Hälfte entstammte dem westlichen, die andere dem östlichen Kulturkreis.

King und Ohya wählten zwanzig Stimuli aus. Diese reichten von einfachen geometrischen Formen bis zu dreidimensionalen, voll ausgeprägten menschlichen Formen. Auch sog. Chernoff Gesichter (mit positiven und negativen Gesichtsausdrücken) und Karikaturen waren darunter. Einige Formen waren zwei-, andere dreidimensional. Einige waren auch mit zufälligem Lidschlag ausgestattet. Zusätzlich wählten die Autoren auch geometrische Formen aus, die zwar nicht einem Kopf ähnelten, jedoch in etwa dessen geometrische Komplexität besaßen.

Der Proband nahm vor einem wandgroßen Display Platz. Durch eine reflektierende Markierung auf der Haut des Probanden konnten die Experimentatoren dessen Kopfbewegungen verfolgen. Das Display präsentierte die Stimuli monochrom. Das System wurde so kalibriert, daß der Proband, indem er seinen Kopf und/oder seinen Oberkörper bewegte, den Stimulus aus verschiedenen Perspektiven wahrnehmen konnte. Die Stimuli wurden bei abgedunkeltem Licht gezeigt und erschienen für 15 Sekunden.

Insgesamt sah jeder Proband 40 Stimuli in zufälliger Reihenfolge (jede vorhandene Form unter zwei Bedingungen. Leider teilten die King und Ohya nicht mit, was diese Bedingungen waren und worin sie sich unterschieden). Zunächst mußte der Proband die wahrgenommenen Eigenschaften des jeweiligen Stimulus einer mehr-dimensionalen Skala zuordnen: dies war ein Dreieck mit den Seiten „Objekt“, „Agent“ und „Ereignis“. Der Proband traf die Zuordnung mittels eines Kreuzes innerhalb des Dreiecks.

Dann wurde der Proband gebeten, sich den Stimulus erneut zu betrachten und diesmal genau einem der obigen Begriffen zuzuordnen.

Schließlich sollte er die „Intelligenz“ oder „potenzielle Intelligenz“ des Stimulus auf einer 10-stufigen Skala bewerten.

6.3 Ergebnisse

Die folgenden Ergebnisse bezeichnen KING und OHYA in ihrer Arbeit als vorläufig. Die Probanden bewerteten die voll ausgeprägte, menschliche Form signifikant häufiger als „Agenten“ als andere anthropomorphe und andere Formen. Sie bewerteten die Intelligenz von menschlichen Formen signifikant höher als die von Karikaturen und Chernoff Gesichtern.

Die Probanden stufen unter den menschlichen Formen die drei-dimensionalen Formen signifikant häufiger als Agenten ein als andere Objekte. Dagegen waren die Bewertungen unter den eher symbolischen Karikaturen und Chernoff Gesichtern ähnlich. Die Versuchspersonen schätzten die Gruppe der anthropomorphen Formen als intelligenter und „agentenhafter“ ein als den Rest der Stimuli. Am intelligentesten und „agentenhaftesten“ schätzten sie die menschlichen Formen mit zufälligem Lidschlag ein.

Bzgl. der Anwendung dieser Ergebnisse verweisen die Autoren verweisen darauf, dass die mittels der gewählten Repräsentation des Agenten beim Nutzer erzeugten Stufen von Intelligenz und Vertrauen auch mit der tatsächlichen „Intelligenz“ und der Vertrauenswürdigkeit des

Agenten übereinstimmen sollten. Eine Diskrepanz würde den Nutzer verwirren.

7 VAN MULKEN *et al.*: *Der Persona Effekt*

van Mulken, André und Müller (1998) führten ein Experiment mit ihrem PPP („Personalized, Plan-based Presenter“) genannten System durch. (van Mulken *et al.* (1998) beschreiben vorwiegend das Experiment, André, Rist und Müller (1998) gehen vorwiegend auf das System ein.) PPP ist eine comic-artige, menschliche Figur, die Führungen bzw. Präsentationen im Netz gibt. Ihr Verhalten wird durch Anweisungen in einem Skript festgelegt. Insbesondere sind dies Zeigebewegungen in Kombination mit Sprachäußerungen mittels Sprachsynthese. Außerdem besitzt PPP eigenes Verhalten, um es lebhafter erscheinen zu lassen. Tabelle 5 gibt eine Klassifikation dieses eigenen Verhaltens (nach André *et al.*, 1998). Das entgeltliche Verhalten ergibt sich aus den Skript-Anweisungen und dem eigenen Verhalten.

Aktionstyp	Zweck	Beispiel
Pausen	einen lebendigeren Eindruck schaffen, ohne abzulenken	umhersehen, -schlendern
Navigation	die Person zur nächsten Position zu bewegen	umdrehen und laufen
sofortige Reaktion	auf externe Ereignisse, die direkte Rückmeldung verlangen, reagieren	springen

Tabelle 5: Eigenes Verhalten des PPP

7.1 Untersuchte Fragen

Die Autoren fragten sich,

- ob es einen Persona-Effekt (also einen positiven Effekt eines animierten Agenten auf den Nutzer) bzgl. der subjektiven Bewertung der Präsentation gibt,
- ob dieser Effekt das Verstehen der Präsentation erleichtert.

Für den ersten Teil erwarteten van Mulken *et al.* eine positivere Bewertung für den PPP.

7.2 Das Experiment

15 weibliche und 13 männliche Personen nahmen an dem Experiment teil, allerdings wurden die Daten zweier Versuchsteilnehmer aus technischen Gründen nicht berücksichtigt. Das Durchschnittsalter betrug 28 Jahre.

Es gab zwei unabhängige Variablen:

1. die Darstellung des präsentierenden Agenten: zum einen ein animierter Agent (Persona vorhanden), zum anderen ein sprechender Zeigestock (Persona abwesend).
2. der Informationstyp: zum einen technisch, zum anderen nicht-technisch.

Die erste Variable wurde interindividuell, die zweite intraindividuell variiert: jede Versuchsperson sah beide Typen von Informationen, aber jeweils entweder nur mit oder ohne Agent.

Der animierte Agent nutzte einen Zeigestock, um auf Dinge zu zeigen. Um wirklich nur die Anwesenheit des Agenten auszuschalten, behielten die Autoren den Zeigestock in der Versuchsbedingung ohne animierten Agenten bei.

Für die technische Versuchsbedingung benutzten van Mulken *et al.* die Informationen über vier verschiedene Flaschenzugsysteme. Die Probanden mußten später Fragen zur Konfiguration und der Kinematik der Systeme beantworten. Bei der nicht-technische Versuchsbedingung wurden zehn fiktive Institutsmitarbeiter mit Photo, Name, Beruf und Zimmernummer vorgestellt. Hier sollten sich die Versuchspersonen an diese Angaben nach der Präsentation erinnern.

Die Probanden sahen zunächst vier Präsentationen über Flaschenzüge und dann die Präsentation über die fiktiven Mitarbeiter. Nach jeder Präsentation mußten sie die Schwierigkeit der Präsentation einschätzen und einen Test durchführen. Zum Schluß füllten sie einen Fragebogen aus.

7.3 Ergebnisse

Im Falle des technischen Systems (Flaschenzüge) fanden die Probanden die Erklärungen leichter verständlich, wenn sie von dem animierten Agenten gegeben wurden. Dagegen gab es keinen Unterschied in der Präsentation der fiktiven Mitarbeiter. Auch wurde bei der Flaschenzug-Bedingung der animierte Agent als hilfreicher und unterhaltender angesehen als der Zeigestock. Wieder gab es diesen Unterschied nicht bei der Mitarbeiter-Repräsentation.

Angesichts dieser Domänen-Spezifität stellten van Mulken *et al.* die Hypothese auf, daß dieser Effekt am Aussehen des animierten Agenten liege: der Agent trug einen „Blaumann“ und hätte somit sehr gut zur technischen Flaschenzug-Präsentation gepaßt.

van Mulken *et al.* fragten die Probanden, ob sie sich von dem animierten Agenten abgelenkt fühlten, und ob er ihnen geholfen hätte, sich zu konzentrieren. Sie verglichen die Werte der beiden Informationstypen und stellten keinen Unterschied bzgl. der Ablenkung fest. Bei dem Flaschenzugsystem fühlten die Probanden sich jedoch signifikant konzentrierter.

Die Zahl der Fehler in den Antworten auf die Fragen zu den Präsentationen unterschied sich nicht signifikant zwischen der Bedingung mit dem animierten Agenten und dem Zeigestock. Hier gab es keine Unterschiede zwischen den Informationstypen.

Im Falle des Flaschenzug-Systems wurden korrekte Antworten auf die Problemlöse-Fähigkeit des Probanden in Zusammenwirkung mit einem adäquaten mentalen Modell gewertet. Im Falle der Mitarbeiter-Bedingung baten van Mulken *et al.* die Versuchspersonen, sich an die Namen und Berufe der vorgestellten Personen zu erinnern. Eine korrekte Antwort führten sie auf die Gedächtnisleistung des Probanden zurück. So kamen sie zu dem Schluß, daß ein animierter Agent weder zu erhöhten Problemlöse-Fähigkeiten noch zu einer höheren Gedächtnisleistung anregt.

50% der Probanden würde eine Präsentation mit einem animierten Agenten bevorzugen, 43% würden dies von dem zu präsentierenden Material abhängig machen und 7% wünschten sich Präsentationen ohne animierten Agenten.

8 KODA und MAES: Agenten mit Gesichtern

Koda und Maes (1996) wollten herausfinden, ob es eher vorteilhaft ist, einen Agenten mit einem Gesicht zu repräsentieren. Sie nutzen ein Poker-Spiel als Testumgebung für ihre Studie.

8.1 Untersuchte Fragen

Koda und Maes fragten sich:

- Mögen Leute den Anblick eines Gesichtes oder von Gesichtsausdrücken bei einem Agenten?
- Nutzen Leute das Gesicht, um das Verhalten des Agenten zu interpretieren?
- Erhöht ein Gesicht als Interface das Engagement einer Person für eine Aufgabe?
- Unter der Voraussetzung, daß obige Fragen positiv beantwortet werden: welche Charakteristiken eines Gesichtes lassen einen Agenten intelligent, freundlich und komfortabel erscheinen?

8.2 Das Experiment

Zehn Studenten des MIT Media Lab nahmen als Versuchspersonen an dem Experiment teil. Das Durchschnittsalter betrug 27 Jahre.

Die Probanden spielten jeder 18 Runden Poker gegen vier Agenten. Es gab fünf verschiedene Agenten, die sich in ihrer Repräsentation unterschieden: kein Gesicht, ein weibliches, realistisches Gesicht, ein weibliches Karikatur-Gesicht, ein Smiley-Gesicht und die Karikatur eines Hundegesichtes. Es gab demnach Unterschiede in der Menschlichkeit und dem Grad der Abstraktion. Jeder Agent hatte zehn verschiedene Gesichtsausdrücke, die er während des Spiels gemäß seines emotionalen Zustandes änderte. Der Algorithmus, der sie des Pokerspiels befähigte, war bei allen Agenten derselbe. Dies wußten die Probanden jedoch nicht.

Nach den 18 Poker-Runden fühlten die Versuchspersonen einen Fragebogen aus.

8.3 Ergebnisse

Die Probanden empfanden den Agenten mit dem realistischen Gesicht am intelligentesten, gefolgt von den beiden Karikaturen.

Koda und Maes fragten zudem, wie sympathisch die Gegenspieler den Probanden waren. Hier zeigte sich, daß sie die Gegenspieler mit Gesichtern sympathischer fanden als den unsichtbaren Gegner. Dies steht in einem gewissen Gegensatz zu den Resultaten von Sproull et al., die fanden, daß die soziale Bewertung eines Agenten mit Sprache und Gesicht geringer ausfällt

als bei einem Agenten mit einem Textinterface. Unter den Gesichtsagenten war die Hunde-Karikatur am beliebtesten. Hier besteht ein Gegensatz zwischen wahrgenommener Intelligenz und Sympathie.

Die Autoren faßten ihre Beobachtungen wie folgt zusammen:

- Die Probanden versuchten, die Gesichter und Gesichtsausdrücke zu deuten, was sie von ihrer Arbeit abhielt.
- Gesichter werden als sympathisch und einnehmend in einem Pokerspiel angesehen.
- Je realistischer ein menschliches Gesicht, desto eher wird es als intelligent, sympathisch und komfortabel angesehen.
- Eine Tierfigur ist sympathischer als ein menschliches Gesicht.

Koda und Maes empfehlen, Agenten mit Gesichtern im Entertainmentbereich einzusetzen. Auch für Aufgaben, bei denen der Nutzer Engagement zeigen muß (z.B. Bildung und Training), sei ein solcher Agent von Vorteil.

9 CASSELL und THÓRISSON: Die Macht eines Nickens und eines Blickes

Cassell und Thórisson (1998) präsentieren in ihrer Arbeit einen personifizierten, animierten Agent mit der Fähigkeit, multimodale Konversation mit einem Nutzer in Echtzeit durchzuführen. Mit diesem Agenten untersuchen sie zwei menschliche Eigenschaften, die als besonders nützlich für konversationale Systeme herausgestellt werden: *emotional feedback* und *envelope feedback*.

Emotional feedback bezeichnet die Technik, eine bestimmte Emotion durch einen bestimmten Gesichtsausdruck darzustellen. *Envelope feedback* meint non-verbales Verhalten während eines Gesprächs von Angesichts zu Angesicht, die der animierte Agent als Antwort auf die kommunikative Aktionen des Nutzers generiert. Dies können zum Beispiel Handgesten, Blicke zum Nutzer oder weg von ihm und Kopfnicken sein. Hier ist das Timing besonders wichtig.

In ihrer Studie verglichen Cassell und Thórisson diese beiden Arten von non-verbalem Feedback.

9.1 Untersuchte Fragen

Cassell und Thórisson fragten sich:

- Hilft non-verbales Feedback überhaupt bei animierten Agenten?
- Wenn ja, welche Art von non-verbalem Feedback?

Sie vermuteten, daß dieser Vergleich zugunsten von *emotional feedback* ausgehen würde.

Die Probanden interagierten mit einem Agenten, der ihnen einiges über das Sonnensystem erklären konnte. Sie erstellten drei Versuchsbedingungen: (CONT) der Agent gibt kein non-verbales Feedback, (EMO) der Agent gibt emotionales Feedback und (ENV) der Agent gibt

envelope feedback. Näheres dazu im Abschnitt „Das Experiment“. Die Autorinnen stellten acht Hypothesen auf:

1. Zwischen den Bedingungen CONT und EMO wird es keine Unterschiede geben bzgl. der relativen Beteiligungen der Nutzer.
2. Die relative Zahl der Beteiligungen der Probanden wird in der Bedingung ENV kleiner sein als in den anderen beiden Bedingungen.
3. Die Probanden in der Bedingung CONT werden genauso oft zögern wie in der Bedingung EMO.
4. In der Bedingung ENV werden die Probanden relativ gesehen weniger häufig zögern als in den Bedingungen EMO und CONT.
5. Die Probanden in der Bedingung CONT werden genauso oft gleichzeitig wie der Agent sprechen wie die Probanden in der Bedingung EMO.
6. In der Bedingung ENV werden die Probanden relativ gesehen weniger oft gleichzeitig wie der Agent als in den Bedingungen EMO und CONT.
7. Es wird keinen signifikanten Unterschied in den Ratings des Agenten zwischen den Bedingungen CONT und EMO geben.
8. Die Probanden in der Bedingung ENV werden den Agenten höher bewerten als die Probanden in den Bedingungen CONT und EMO.

Die Daten für die Hypothesen 1–6 wurden durch Videoband-Analysen gewonnen. Dies geschah unabhängig durch zwei Analysten in einem Doppelblind-Design. Die Reliabilität dieser Analysen war sehr gut ($r = 0.95$). Die Daten für die Hypothesen 7 und 8 wurden durch Fragebogen erhoben.

9.2 Das Experiment

Für ihr Experiment nutzten Cassell und Thórisson ein „Ymir“ genanntes System zur Generation von voll automatisierten Charakteren mit Echtzeit- und multi-modalen Fähigkeiten, die mit dem Nutzer von Angesicht zu Angesicht interagieren können. Die Figur des Agenten erschien auf einem Monitor und ein Model des Sonnensystems auf einem anderen. Das System konnte die Augenbewegungen des Probanden verfolgen. Der Nutzer trug einen „tracking suit“; dies erlaubte es dem System, sich eine visuelle Repräsentation seines Oberkörpers zu errechnen.

Der Nutzer konnte dem System Fragen über das Sonnensystem und seine Planeten stellen.

Es wurden drei animierte, personifizierte Agenten erstellt. Sie unterschieden sich nur in ihrem Gesicht und ihrer Stimme. Zusätzlich, abhängig von der Versuchsbedingung, unterschieden sie sich in ihren Antworten auf die Konversation: in der Kontrollbedingung (CONT) gab der Agent nur inhaltliche Rückmeldungen, in der zweiten Bedingung (ENV) zusätzlich *envelope feedback* und in der dritten Bedingung (EMO) zusätzlich zur Kontrollbedingung *emotional feedback*.

Das Verhalten des Agenten faßten die Autorinnen wie folgt zusammen:

1. CONT

- Ausführen von Kommandos („Will you show me what Mars looks like?“) und Beantwortung von Fragen.
- Verbale Bestätigung beim Ausführen von Kommandos (z. B. „okey-dokey, let’s go to Mars!“)

2. EMO

- Verwirrter Ausdruck, wenn er eine Äußerung nicht verstanden hat.
- Lächelt, wenn er vom Nutzer angesprochen wird und er einer Aufforderung nachkommt.

3. ENV

- Kopf und Augen dem Nutzer beim Zuhören und der Aufgabe beim Ausführen eines Kommandos zuwenden.
- Blick abwenden und Augenbrauen heben, wenn Gesprächsführung vom Nutzer an den Agenten geht.
- Blick zurück zum Nutzer, wenn die Gesprächsführung vom Agenten an den Nutzer geht.
- Mit den Fingern klopfen um zu zeigen, daß er „lebt“.
- Handgeste begleitend zum verbalen Inhalt.

Zu den abhängigen Variablen: Die Erfahrung der Nutzer mit der Lebensähnlichkeit des Agenten und der Einfachheit der Interaktion wurden durch einen Fragebogen erhoben. Die Effizienz der Interaktion wurde gemessen durch die relative Zahl von Äußerungen (Verhältnis Beteiligungen des Nutzer zu Beteiligungen des Agenten), der relativen Zahl des Zögerns des Probanden (bezogen auf die absolute Zahl seiner Beiträge) und relative Zahl der Frustrationen (bezogen auf die absolute Zahl seiner Beiträge).

In einer Vorstudie testeten Cassell und Thórisson mit 24 Probanden, dass die emotionalen Gesichtsausdrücke bei allen genutzten Charakteren korrekt erkannt wurden.

Die unabhängigen Variablen waren:

- Agent (namens „Gandalf“, „Roland“ und „Bilbo“),
- feedback-Bedingung (ENV, CONT und EMO),
- Reihenfolge der Gesichter und
- Reihenfolge der Bedingungen.

Die ersten beiden Variablen waren intraindividuell und die beiden letzten interindividuell.

Die zwölf Probanden waren beiderlei Geschlechts und zwischen 22 und 37 Jahre alt. Sie waren Anfänger im Umgang mit Computern und wurden instruiert, daß der Agent Kenntnisse über das Sonnensystem besitzt und sie mit ihm so normal wie möglich interagieren sollen.

Sie konnten sich in 4–8 Gesprächsrunden mit den Agenten vertraut machen, bevor das Experiment begann. Die Probanden interagierten im Durchschnitt sieben Minuten mit jedem der drei Agenten, dazwischen gab es jeweils eine fünfminütige Pause.

9.3 Ergebnisse

Sechs der acht Hypothesen wurden bestätigt. Insgesamt unterstützen die Ergebnisse die Signifikanz von *envelope feedback* über *emotional feedback* und rein inhaltliche Rückmeldungen. In keinem Fall wurde emotionales Feedback anders eingeschätzt als in der Kontrollbedingung. Die Autorinnen fanden keinen Effekt der Reihenfolgen von Charakteren und Bedingungen.

In der ENV Bedingung machten die Probanden weniger Gesprächsbeiträge pro Agenten-Gesprächsbeitrag als in den anderen Bedingungen. Cassell und Thórisson sehen darin eine ungefähre Schätzung für die Effizienz der Unterhaltung.

In dem Fragebogen mußten die Probanden die Sprache des Agenten, den Interaktionsstil und die Lebensähnlichkeit einschätzen. Für die Forscherinnen überraschend erreichten die Agenten in der ENV Bedingung eine hohe Bewertung bzgl. der Sprachfähigkeiten, und zwar absolut und verglichen mit den anderen beiden Bedingungen. Auch empfanden die Versuchspersonen in der ENV Bedingungen das Gespräch reibungsloser als in den anderen beiden Bedingungen.

Die Hypothesen vier und sechs wurden nicht bestätigt. In der Bedingung ENV zögerten die Probanden häufiger und sprachen auch häufiger zur gleichen Zeit wie der Agent. Cassell und Thórisson erklären dies zum einen mit Problemen der vorliegenden Implementationen mit zu lebenshaften Nutzern. Das Ins-Wort-Fallen sei typisch für Mensch-zu-Mensch Kommunikation. Außerdem habe der Agent Probleme mit dem Ins-Wort-Fallen, was wiederum zum Zögern beim Nutzer führen könnte.

Zusammenfassend, so die Autorinnen, hätten sie demonstriert, daß Systeme zur Konversation, die eine anthropomorphe Figur nutzen, besser funktionieren und besser akzeptiert werden können, wenn sie Elemente des *envelope feedback* anwenden würden.

10 CASSELL und VILHJÁLMSOON: *Autonomes kommunikatives Verhalten*

Cassell und Vilhjálmssoon (1999) beschäftigten sich damit, wie man das kommunikative Verhalten von Avataren in graphischen Chats verbessern kann. In aktuellen Chatsystemen, beklagen sie, kann der Nutzer zur gleichen Zeit entweder das Verhalten des Avatars steuern oder mit anderen Nutzern Nachrichten austauschen. Ist der Nutzer damit beschäftigt, Nachrichten einzutippen, stände die Avatare nur bewegungslos herum. Dies liefe der natürlichen Kommunikation zuwider, denn hier ist ein beachtlicher Teil non-verbale Kommunikation im Spiel.

Zwar könnten die Nutzer in den neuesten Systemen verschiedene Animationen oder emotionale Zustände aus einem Menü wählen, aber non-verbale Kommunikation sei oft spontan und der Nutzer zu beschäftigt, um das Verhalten seines Avatars zu kontrollieren.

Wäre es nicht wünschenswert, wenn Avatare autonom kommunikatives Verhalten zeigen würden, fragten sich Cassell und Vilhjálmssoon. Hier stellt sich die Frage, was der Nutzer bevorzugt: direkte Manipulation oder Autonomie des Agenten.

Der Ansatz von Cassell und Vilhjálmsón betrachtet den Avatar als autonomen Agenten und wendet die Technik des *envelope feedback* (siehe Abschnitt 9) an. Allerdings ist die Autonomie auf eine Reihe von kommunikativen Ausdrücken des Gesichts und des Kopfes begrenzt. Der Nutzer hat weiterhin die Kontrolle über die Navigation und den Inhalt des Gespráches. Er behált gewissermaßen die Kontrolle auf einer höheren Ebene über das Verhalten des Avatars. Damit werden laut Cassell und Vilhjálmsón folgende Probleme angesprochen:

Kontrollkomplexität Der Nutzer manipuliert einige wenige Parameter, die seine aktuellen kommunikativen Absichten repräsentieren. Er muß sich nicht mit der Animation einer menschlichen Figur auseinandersetzen.

Spontane Reaktion Der Avatar zeigt spontane Reaktionen auf andere Avatare, ohne daß der Nutzer diese explizit angeben muß.

Diskrete Nutzereingaben Weil der Avatar autonom ist, kann er auch ohne Nutzereingaben angemessenes Verhalten zeigen.

Übertragung von der Nutzerwelt in die Cyperwelt Da der Nutzer und der Avatar sich in verschiedenen Umgebungen aufhalten, würde es wenig Sinn machen, die eine Welt in die andere zu übertragen, also zum Beispiel live das Bild des Nutzers in den Chat zu übertragen. Kontrolle auf der Ebene der Intention und Autonomie auf der Ebene des unfreiwilligen kommunikativen Verhaltens könnten es jedoch dem Avatar erlauben, in der Situation geeignete kommunikative Hinweise zu geben.

10.1 Untersuchte Fragen

Für ihr Experiment benutzten Cassell und Vilhjálmsón ein 3D-Chat Programm namens *BodyChat*. In diesem Chat werden die Avatare als 3D-Modelle von Oberkörpern von cartoon-ähnlichen, menschlichen Figuren repräsentiert. Nutzer können ihre Avatare mittels Tasten navigieren, mit der Maus Kommando-Parameter eingeben und mit anderen Nutzern mittels Text kommunizieren. Dabei erscheinen die Worte des Sprechers oberhalb des Kopfes des Avatars. Es können sich nur jeweils zwei Personen miteinander unterhalten.

Das kommunikative Verhalten des Avatars hängt von den Intentionen des Nutzers ab. Diese kann der Nutzer über Kommando-Parameter beschreiben: er kann mitteilen, mit wem er sprechen möchte, ob er überhaupt gespráchsbereit ist oder ob er eine Gespräch beenden möchte. Aus diesen Intentionen generiert der Avatar dann autonomes Verhalten. Bei der Wiedergabe von Wörtern versucht der Avatar einen angemessenen Gesichtsausdruck zu machen, z.B. kann er das Wort „very“ durch ein leichtes Kopfnicken betonen.

Im wesentlichen wollten Cassell und Vilhjálmsón überprüfen, welche Effekte ein solcher autonomer Avatar auf den Nutzer eines graphischen Chats hat. Obwohl sie bereits die Vorzüge eines autonomen Avatars schilderten, haben sie doch auch Bedenken:

- Non-verbale Kommunikation ist noch nicht vollständig verstanden. Würde das System auch nur annähernd einer natürlichen menschlichen Kommunikation ähneln?
- Wünscht der Nutzer einen solch autonomen Avatar? Oder will er die Kontrolle über ihn?

- Lenkt der Avatar nicht einfach nur von der eigentlichen Kommunikation ab?

10.2 Das Experiment

An dem Experiment nahmen 24 Probanden teil. Sie hatten Erfahrung mit textbasierten Chatsystemen, jedoch keine mit graphischen Chatsystemen. Ihnen wurde die Aufgabe gestellt, andere Nutzer im Chat zu treffen und so viel wie möglich übereinander herauszufinden.

In der Chatumgebung befanden sich vier verschiedene Nutzer (die jedoch alle von einem Experimentator gesteuert wurden). Es gab drei verschiedene Versionen von BodyChat, die sich bis auf folgendes nicht unterschieden:

- Die *autonome* Version besaß alle bisher beschriebenen Fähigkeiten.
- Die *manuelle* Version besaß die gleichen Verhaltensmöglichkeiten, jedoch mußte der Nutzer sie von Hand aus einem Menü auswählen, um sie zu aktivieren.
- In der *sowohl-als-auch* Version besaßen die Avatare autonomes Verhalten, und die Nutzer konnten zusätzlich über Menüs weiteres Verhalten generieren.

Es gab auch noch eine vierte Version, die jedoch gesondert betrachtet wurde:

- Die *nichts* Version besaß kein kommunikatives Verhalten, der Nutzer konnte den Avatar lediglich durch den Raum bewegen.

Die Probanden konnten das System für 45 Minuten benutzen, dann mußten sie einen Fragebogen ausfüllen. Außerdem sollten sie alle Fakten, die sie über die anderen scheinbaren Nutzer erfahren hatten, auflisten. Das galt dann als Maß für die Aufgabenperformanz.

10.3 Ergebnisse

Die Gespräche in der autonomen Bedingung dauerten signifikant länger als in den anderen beiden Bedingungen. Laut Cassell und Vilhjálmsón könnte dies darauf hindeuten, daß die Probanden des autonomen Systems interessierter an der Kommunikation seien. Außerdem erinnerten sich in die Versuchspersonen der autonomen Version an mehr Fakten als in der sowohl-als-auch und in der manuellen Version (in dieser Reihenfolge). Dies könnte, so die Autoren, an erhöhtem Engagement der Probanden liegen, weil sie ihre Aufmerksamkeit nicht zwischen Kommunikation und Kontrolle des Avatars teilen mußten.

Cassell und Vilhjálmsón fragten die Probanden, wie natürlich sie das Verhalten des Avatars und die Interaktion fanden. Diese beiden zusammen zeigten, daß die Nutzer des autonomen Systems dieses als natürlicher beurteilten als die beiden anderen Systeme.

Die Probanden fanden außerdem, daß das autonome System expressiver war verglichen mit dem manuellen System, aber nicht signifikant expressiver als das sowohl-als-auch System. Dies weist darauf hin, so die Autoren, daß sich die Expressivität erhöht, wenn autonomes Verhalten generiert wird. Allerdings fanden die Nutzer des sowohl-als-auch Systems ihre Erfahrung langweiliger als die Nutzer des autonomen Systems.

Schließlich fanden die Probanden in der autonomen Bedingung, daß sie mehr Kontrolle über den Konversation hatten als die Probanden des sowohl-als-auch und des manuellen Systems (in

dieser Reihenfolge). Mit diesem Ergebnis scheinen sich zwei Gegensätze (Kontrolle vs. Autonomie) zu vereinen.

Schließlich verglichen Cassell und Vilhjálmsson die autonome Bedingung mit der nichts Bedingung. Die autonome Bedingung wurde als signifikant natürlicher eingestuft als die nichts Bedingung. Aber es gab keinen Unterschied hinsichtlich der Expressivität und Kontrolle. Außerdem wurde die durchschnittliche Gesprächsdauer in der nichts Bedingung nur von der in der autonomen Bedingung übertroffen. Die Autoren sehen sich in ihrer Schlußfolgerung unterstützt, daß die Nutzer Kontrolle und Spaß primär dem Gespräch selber entnehmen und sich durch die Kontrolle des Verhaltens ihres Avatars abgelenkt fühlen.

11 Untersuchungen zum Anthropomorphismus in der Sekundärliteratur

Die folgenden Untersuchung habe ich der Arbeit von Dehn und van Mulken (1999) entnommen, da ich die entsprechenden Originalarbeiten leider nicht finden konnte. Daher sind die Informationen dieser Arbeiten sehr spärlich und betreffen größtenteils nur das Resultat. Da sie jedoch wichtige Aussagen über anthropomorphe vs. nicht-anthropomorphe Agenten liefern, habe ich sie hier aufgeführt.

11.1 LESTER et al.: Der Persona Effekt

In einer Studie stellten LESTER et al.⁵ zwei pädagogische Agenten gegenüber: ein voll animierter, ausdrucksstarker Agent, der aufgaben- und domänenspezifische Ratschläge gibt, und ein ruhender Agent, der keine Ratschläge gibt.

Die Autoren fanden keine signifikanten Unterschiede, wenn es um die Unterhaltsamkeit der beiden Agententypen ging.

Die Nützlichkeit des ruhenden Agenten wurde höher bewertet als die des animierten Agenten. Andererseits stuften die Probanden den Rat des animierten Agenten als nützlicher ein. Dies verwundert jedoch nicht, da der ruhende Agent gar keinen Ratschläge gab.

Der animierte Agent wurde als glaubwürdiger empfunden.

11.2 LESTER et al.: Pädagogische Agent und das Problemlösen

In einer zweiten Studie beschäftigten sich LESTER et al.⁶ mit der Auswirkung von verschiedenen Agenten auf die Effektivität von Problemlösen und das Lernen.

Dazu wurden fünf Versionen eines animierten Agenten eingesetzt. Diese Versionen unterschieden sich in der Art des Ratschlages (aufgabenspezifisch vs. prinzipienbasiert) und in der Modalität, in der sie gegeben wurden (ausschließlich verbal vs. verbal und animiert).

⁵LESTER, J. C., CONVERSE, S. A., KAHLER, S. E., BAHLOW, S. T., STONE, B. A. & BHOGAL, R. S. (1997). The persona effect: affective pedagogical agents. In S. PEMBERTON, Ed. *Human Factors in Computing Systems: CHI'97 Conference Proceedings*, 359–366

⁶LESTER, J. C., STONE, B. A., CONVERSE, S. A., KAHLER, S. E. & BAHLOW, S. T. (1997). Animated pedagogical agents and problem-solving effectiveness: a large-scale empirical investigation. In S. DU BOULAY & MIZOGUCHI, Eds. *Proceedings of the 8th World Conference on Artificial Intelligence in Education*, 23–30

Kinder einer Mittelschule interagierten im Rahmen eines Lernprogrammes der botanischen Anatomie und Physiologie mit diesen Agenten. Sie erhielten die Aufgabe, eine Pflanze zu erschaffen, die an eine gegebene Umwelt möglichst gut angepaßt ist. Als Maß der Fähigkeit zum Problemlösen wählten LESTER et al. die Zahl der Fehler, die gemacht wurden. Tatsächlich fanden sie einen Zusammenhang zwischen dem Typ des Agenten und der Zahl der Fehler: für einfache Probleme gab es keine Unterschiede, wohl aber für komplexe Probleme. Hier wurden bei einem Agenten, der keine Ratschläge gab, mehr Fehler gemacht, als bei allen Ratschläge erteilenden Agenten. Allerdings gab es innerhalb der Gruppe der ratschlagenden Agenten keine Unterschiede.

LESTER et al. schließen nun daraus, daß animierte Agenten Problemlösen verbessern. Allerdings halten Dehn und van Mulken diesen Schluß nicht für überzeugend. Sie kritisieren zu Recht, daß die Ergebnisse nur besagen, daß Ratschläge das Problemlösen verbessern können. Daß animierte Agenten und verschiedene Arten von Ratschlägen dazu nötig sind, besagen sie nicht.

Weiterhin führten LESTER et al. Pre- und Post-Tests über das Biologie-Wissen der Probanden durch. Der Unterschied im Rating sollte den Umfang des Erlernten infolge des Lernprogrammes erfassen. Für alle Versuchsbedingungen gab es signifikanten Unterschiede zwischen Pre- und Posttests. Jedoch ließen sich keine statisch signifikanten Unterschiede zwischen den Agententypen nachweisen.

11.3 TAKEUCHI und NAITO: Soziale Interaktion und Agenten mit Gesicht

TAKEUCHI und NAITO ließen in ihrer Untersuchung⁷ zwei menschliche Gegner ein *Memory*-ähnliches Kartenspiel am Computer spielen. Auf dem Display war einmal ein Gesicht, ein anderesmal ein dreidimensionaler Pfeil zu sehen. Die Figuren unterstrichen den Stand des Spieles, z.B. schauten oder zeigten sie auf den Spieler, der an der Reihe war. Außerdem zeigte das Gesicht emotionale Ausdrücke, z.B. ein fröhliches, wenn einer der beiden Gegner zwei passende Karten fanden.

Lediglich sieben Probanden nahmen an dem Experiment teil. Dies begrenzt die Aussagekraft dieser Studie.

Das Display mit dem Gesicht erzeugte mehr Augenkontakt als das Display mit dem Pfeil, fanden die Autoren. Dies läßt darauf schließen, daß das Gesicht mehr Aufmerksamkeit der Probanden auf sich zog. Anscheinend zog das jedoch gleichzeitig eine Ablenkung von der Hauptaufgabe nach sich: die Versuchspersonen in der Bedingung mit dem Gesicht benötigten eine größere Reaktionszeit als die Probanden in der Bedingung mit dem Pfeil. Den Schluß von TAKEUCHI und NAITO, die Probanden konzentrierten sich nicht auf das Spiel, hinterfragen jedoch Dehn und van Mulken. Ihrer Meinung nach mag auch eine größere Komplexität der Situation für die Erhöhung der Reaktionszeit eine Rolle gespielt haben.

Die Probanden empfanden das Display mit dem animierten Gesicht als unterhaltsamer als das Display mit Pfeil.

⁷TAKEUCHI, AKIKAZU & NAGAGO, KATASHI (1995). Situated facial displays: towards social interaction. In I. KATZ, R. MACKL, L. MARKS, M. B. ROSSON & J. NIELSON, Eds. *Human Factors in Computing Systems: CHI'95 Conference Proceedings*, 450–455. New York: ACM Press

LITERATURVERZEICHNIS

Andererseits empfanden die Versuchspersonen das Display mit dem Pfeil als nützlicher verglichen mit dem Display mit dem Gesicht.

Literaturverzeichnis

- Amant, R. S. & Dulberg, M. S. (1998). An experiment with navigation and intelligent assistance. *Intelligent User Interfaces*, 171–178.
- André, E., Rist, T. & Müller, J. (1998). Guiding the user through dynamically generated hypermedia presentations with a life-like character. In *Proceedings of international conference on intelligent user interfaces (iui-98)*. San Francisco.
- Aronson, E. & Linder, D. (1965). Gain and loss of esteem as determinants of interpersonal attractiveness. *Journal of Experimental Social Psychology*, 1, 156-171.
- Cassell, J. & Thórisson, K. R. (1998). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*.
- Cassell, J. & Vilhjálmsson, H. (1999). Fully embodied conversational agents: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, 2, 45–64.
- Dehn, D. M. & van Mulken, S. (1999). The impact of animated interface agents: a review of empirical research. *Int. J. Human-Computer Studies*, 51.
- Dietz, R. B. & Lang, A. (1999). Effective agents: Effects of agent affect on arousal, attention, liking and learning. In *Proceedings of international cognitive '99*. San Francisco, CA.
- King, W. J. & Ohya, J. (1996). The representation of agents: Anthropomorphism, agency, and intelligence. In R. Bilger, S. Guest & M. J. Tauber (Eds.), *Human factors in computing systems: Chi'96 electronic conference proceedings*. (Online verfügbar via <http://info.acm.org/sigchi/chi96/proceedings/shortpap.htm>)
- Koda, T. & Maes, P. (1996). Agents with faces: the effect of personification. In *Proceedings of the 5th iee international workshop on robot and human communication (ro-man'96)* (pp. 189–194).
- Moon, Y. & Nass, C. I. (1996). Adaptive agents and personality change: Complementarity versus similarity as forms of adaptation. In R. Bilger, S. Guest & M. J. Tauber (Eds.), *Human factors in computing systems: Chi'96 electronic conference proceedings*. (Online verfügbar via <http://info.acm.org/sigchi/chi96/proceedings/shortpap.htm>)
- Nass, C., Moon, Y., Reeves, B. & Dryer, D. C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43, 223–239.
- Sproull, L., Subramani, M., Kiesler, S., Walker, J. H. & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction*, 11, 97–124.

LITERATURVERZEICHNIS

van Mulken, S., André, E. & Müller, J. (1998). The persona effect: how substantial is it? In H. Johnson, L. Nigay & C. Roast (Eds.), *People and computers xiii: Proceedings of hcl'98* (pp. 53–66). Berlin: Springer.