

Virtuelle Nothilfe

Ein Experiment zum Effekt von
virtueller Hilfe, Gewalt und Nothilfe
auf Hilfe- und Gewaltverhalten

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

des Fachbereichs Humanwissenschaften

der Universität Osnabrück

vorgelegt

von M. Rohangis Mohseni

aus Köln

Osnabrück, 2013



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

1	INHALTSVERZEICHNIS	
1	INHALTSVERZEICHNIS	2
2	ABSTRACT	8
3	EINLEITUNG	9
4	THEORETISCHER HINTERGRUND	12
4.1	Unterhaltungsmedien	12
4.2	Computerspiele	13
4.2.1	Abgrenzung von anderen Unterhaltungsmedien	13
4.2.2	Genres	14
4.2.3	Verbreitung	15
4.2.4	Forschungsbereiche	16
4.2.5	Unterhaltung und Präsenzgefühl	16
4.3	Gewaltverhalten	20
4.3.1	General Aggression Model	21
4.3.2	Catalyst Model	23
4.3.3	Social Interactionist Model	25
4.3.4	Moral Management Model	29
4.3.5	Catharsis Model	35
4.3.6	Bewertung der Theorien	38
4.3.7	Meta-Analysen	39
4.4	Operationalisierungen des Gewaltverhaltens	43
4.4.1	Verfahren	43
4.4.2	Validitätsprobleme	47
4.4.3	Bewertung der Verfahren	52
4.4.4	Auswahl	62
4.5	Hilfeverhalten	63
4.5.1	Prosoziale Hilfeleistung	64
4.5.2	Altruistische Hilfeleistung	65

Virtuelle Nothilfe

4.5.3	Hilfeleistung in Notsituationen	68
4.5.4	Hilfeleistung in Zivilcoursesituationen	69
4.5.5	Bewertung der Theorien	70
4.5.6	Meta-Analysen	71
4.6	Operationalisierungen des Hilfeverhaltens.....	73
4.6.1	Verfahren.....	73
4.6.2	Validitätsprobleme	75
4.6.3	Bewertung der Verfahren und Auswahl.....	75
5	THEORETISCH-INHALTLICHE HYPOTHESEN	76
5.1	Gewaltverhalten.....	76
5.2	Hilfeverhalten.....	78
6	DESIGN.....	80
7	OPERATIONALISIERUNG	82
7.1	Unabhängige Variablen.....	82
7.1.1	Wahl des Computerspiels.....	82
7.1.2	Spielaufgaben	83
7.1.3	Optimierungen	86
7.1.4	Blockade von Schummelmöglichkeiten	89
7.2	Abhängige Variablen.....	90
7.2.1	Gewaltverhalten.....	90
7.2.2	Hilfeverhalten.....	93
7.3	Störvariablen.....	94
7.3.1	Biologisches Geschlecht	95
7.3.2	Trait-Gewaltbereitschaft	95
7.3.3	Trait-Hilfsbereitschaft	95
7.3.4	Gewaltspielkonsum/Prosozialspielkonsum	97
7.3.5	Eigenschaften der Spielaufgabe	98
7.3.6	Subjektive Erregung	99

Virtuelle Nothilfe

7.3.7	Schwierigkeiten mit der Steuerung.....	99
7.3.8	Spieldauer.....	100
7.3.9	Vorkenntnisse mit Computerspielen	101
7.3.10	Verdachtsbildung	103
7.3.11	Nichtbefolgung.....	111
7.3.12	Zeitdruck.....	111
7.3.13	Schulform	112
7.3.14	Alter und Fachsemester	112
7.3.15	Erhebungswelle	112
7.3.16	Emotionen	112
8	EMPIRISCH-INHALTLICHE HYPOTHESEN	113
8.1	Gewaltverhalten.....	113
8.2	Hilfeverhalten.....	114
9	DURCHFÜHRUNG.....	115
9.1	Vortests.....	115
9.2	Hauptuntersuchung.....	115
9.2.1	Stichprobe	115
9.2.2	Zuordnung zu den Versuchsbedingungen.....	116
9.2.3	Ort.....	117
9.2.4	Aufklärung, Einverständnis und Datenschutz	117
9.2.5	Probandenschutz.....	118
9.2.6	Teilnahmevergütung	118
9.2.7	Versuchsablauf	118
10	STATISTISCHE VORHERSAGEN	120
10.1	Gewaltverhalten.....	120
10.2	Hilfeverhalten.....	121
11	AUSWERTUNG	122
11.1	Stichprobenbeschreibung	122

11.2	Datenprüfung	122
11.2.1	Fehlende Werte.....	122
11.2.2	Ausreißer	123
11.3	Validität der UV	124
11.3.1	Anzahl getöteter Banditen	124
11.3.2	Tutorialdauer.....	124
11.3.3	Aufgabendauer.....	125
11.3.4	Gewalt- und Hilfegehalt der Aufgabe.....	126
11.4	Validität der AV	129
11.4.1	Motive für Gewaltverhalten.....	129
11.4.2	Verdachtsbildung	133
11.5	Interne Validität	134
11.6	Prüfung der Anwendungsvoraussetzungen	138
11.6.1	Auswertungsverfahren.....	138
11.6.2	Varianzhomogenität.....	138
11.6.3	Normalverteilung	139
11.6.4	Homogenität der Regressionen	141
12	DISKUSSION DER STATISTISCHEN VORHERSAGEN.....	142
12.1	Gewaltverhalten.....	142
12.1.1	Einzelvergleiche.....	142
12.1.2	ANOVA.....	143
12.1.3	ANCOVA.....	144
12.2	Hilfeverhalten.....	148
12.2.1	Einzelvergleiche.....	148
12.2.2	ANOVA.....	149
12.2.3	ANCOVA.....	151
13	DISKUSSION DER EMPIRISCH-INHALTLICHE HYPOTHESEN.....	154
13.1	Gewaltverhalten.....	154

13.2	Hilfeverhalten.....	155
14	DISKUSSION DER VALIDITÄT	156
14.1	Gewaltverhalten.....	156
14.1.1	Validität der UV	156
14.1.2	Validität der AV	156
14.1.3	Interne Validität	158
14.2	Hilfeverhalten.....	159
14.2.1	Validität der UV	159
14.2.2	Validität der AV	159
14.2.3	Interne Validität	159
15	DISKUSSION DER THEORETISCH-INHALTLICHEN HYPOTHESEN	161
15.1	Gewaltverhalten.....	161
15.2	Hilfeverhalten.....	162
15.3	Generalisierbarkeit	163
15.3.1	Populationen	163
15.3.2	Situationen	163
16	GENERELLE DISKUSSION.....	166
17	STÄRKEN, SCHWÄCHEN UND OPTIMIERUNGSVORSCHLÄGE	168
17.1	Stärken.....	168
17.2	Schwächen und Optimierungsvorschläge	168
18	LITERATUR	172
19	ANHANG	183
A.	Verwendete Hardware	183
B.	Installierte Modifikationen	184
C.	Optimierung der Datei „Oblivion.ini“	185
D.	Videoeinstellungen.....	186
E.	Eigenschaften der Spielaufgabe	189
F.	Fragebogen der Anamnese	190

Virtuelle Nothilfe

G.	Notizbogen der Spielergebnisse	192
H.	Fragebogen der Hauptuntersuchung	193
I.	Fragebogen der Nachbefragung	201
J.	Einstellungen des CRTT	202
K.	Instruktionen zum CRTT (Vortest)	203
L.	Instruktionen zum CRTT (Hauptuntersuchung)	205
M.	Eckenzählaufgabe	207
N.	Datenverteilung beim Gewaltverhalten	216
O.	Datenverteilung beim Hilfeverhalten	218
P.	Homogenität der Regressionen	220
Q.	Formeln	221
20	ENDNOTEN	222

VIRTUELLE NOTHILFE

Eine aktuelle Meta-Analyse von Anderson et al. (2010) zeigt, dass Gewaltverhalten in Computerspielen das Gewaltverhalten im realen Leben erhöht, während es prosoziales Verhalten verringert. Daneben führen einige Studien von Greitemeyer und Osswald (2010) sowie von Gentile et al. (2009) zu der Schlussfolgerung, dass das Hilfeverhalten in Computerspielen das Hilfeverhalten im realen Leben erhöht. Bisher gibt es aber keine Studien, die die kombinierte Wirkung von Gewalt und Hilfe in Computerspielen untersuchen, obwohl diese Kombination typisch für Gewaltspiele ist (Anderson et al., 2010). In gewalthaltigen Rollenspielen bestehen viele Aufgaben darin, einer virtuellen Person durch die Anwendung von Gewalt zu helfen. Die vorliegende Studie soll diesen empirischen Mangel beheben, indem untersucht wird, ob gewaltsame Nothilfe zu gesteigertem Gewalt- und/oder Hilfeverhalten im realen Leben führt. Zu diesem Zweck wurde das Rollenspiel „Oblivion“ modifiziert, um vier verschiedene Experimentalbedingungen zu realisieren: (1) Nothilfe, (2) Töten, (3) Hilfe und (4) Schatzjagd. Werden diese Bedingungen miteinander verglichen, dann scheint Nothilfe zu einer Verringerung des Hilfeverhaltens im realen Leben zu führen und gleichzeitig das Gewaltverhalten zu erhöhen. Die Ergebnisse stehen im Einklang mit dem Moral Management Model (Hartmann & Vorderer, 2010; Hartmann, im Druck), welches auf Banduras Theorie des Moral Disengagement basiert (Bandura, 2002).

VIOLENT EMERGENCY ASSISTANCE IN COMPUTER GAMES

A recent meta-analysis of Anderson and colleagues (2010) shows that violent behavior in computer games promotes violent behavior in real-life and inhibits prosocial behavior. A couple of studies conducted by Greitemeyer and Osswald (2010) lead to the conclusion that helping behavior in computer games furthers helping behavior in real-life. There exist no studies examining the combined effect of violence and helping in computer games, although this combination is typical for violent video games (Anderson et al., 2010). In violent RPGs, a lot of tasks consist of helping someone by using violence. The present study addresses this issue and bridges the current empirical gap by investigating if violent emergency assistance furthers helping behavior and/or violent behavior in real-life. To accomplish that, the role-playing game “Oblivion” was modified to create four different experimental conditions: (1) violent emergency assistance, (2) killing, (3) helping, and (4) treasure hunting. Comparing these conditions, violent emergency assistance seemingly reduces helping behavior in real-life and at the same time furthers violent behavior. The results are in unison with the moral management model (Hartmann & Vorderer, 2010; Hartmann, in press), which is based on Banduras Theory of Moral Disengagement (Bandura, 2002).

3 EINLEITUNG

In der Wissenschaft existiert eine anhaltende Debatte über die Wirkung von Gewalt in Computerspielen. Dabei befürchten einige der Diskutierenden (z. B. Anderson et al., 2010), dass der dauerhafte Konsum von gewalthaltigen Spielen zu einer Erhöhung des Gewaltverhaltens führt. Diese Debatte findet sich auch in der Gesellschaft wieder. Nach einer Studie von Lawrence und Birkland (2004, zit. n. Ferguson, 2007a) waren Gewaltmedien (wie z. B. das Computerspiel „Doom“) und andere Bestandteile der Popkultur (wie z. B. die „Gothic“ Subkultur) die am zweithäufigsten in der New York Times und der Los Angeles Times diskutierten Ursachen für den Amoklauf in Columbine. Tatsächlich zeigt eine Studie von Feltes und Goldberg (2006), dass Gewalt an Schulen inzwischen zum Alltag von Schüler/innen gehört: In ihrer Untersuchung an über viertausend Achtklässlern des Schuljahrs 2003/2004 aus Bochum und Herne berichten 13.5% der Schüler/innen, schon einmal jemanden so geschlagen oder getreten zu haben, dass dieser zum Arzt musste, 9.5% jemanden genötigt zu haben und 8.5% jemanden beraubt zu haben.

Es gibt aber nur eine einzige Untersuchung (Ferguson, 2008), in der geprüft wurde, ob der Konsum von Gewaltspielen tatsächlich im Zusammenhang mit *Amokläufen* steht. Auch fanden sich bis vor Kurzem nur wenige Studien, die einen Zusammenhang zwischen Gewaltspielkonsum und *Aggression* untersucht haben. Beispielsweise sind in die Meta-Analyse von Anderson (2004) nur 44 Artikel mit 97 Studien aus dem Zeitraum von 1983 bis 2004 eingegangen, in die von Ferguson (2007a) nur 24 Artikel mit 25 Studien aus dem Zeitraum 1995 bis 2005. Erst die aktuellste Meta-Analyse von Anderson et al. (2010) enthält 136 Artikel mit 381 Studien aus dem Zeitraum von 1983 bis 2008. Die meisten der durchgeführten Studien basieren auf sogenannten (Ego-)Shootern, in denen der/die Spieler/in mit einer virtuellen Schusswaffe auf gegnerische Spielfiguren schießt.

Im Gegensatz zu dieser eher kulturpessimistischen Forschung gibt es inzwischen auch einige wenige kulturoptimistische Studien zum Zusammenhang zwischen prosozialen Spielen und Hilfeverhalten. Beispielsweise konnten Experimente von Greitemeyer und Oswald (2010) zeigen, dass prosoziale Spiele wie „Lemmings“ (DMA Design, 1991) das Hilfeverhalten fördern können.

Neben rein prosozialen Spielen und reinen Gewaltspielen gibt es auch Computerspiele mit Mischformen von Gewalt und Hilfsbereitschaft. In der weitverbreiteten Gattung des computerbasierten „Rollenspiels“ nimmt der Spieler bzw. die Spielerin die Rolle einer anderen Person ein. Diese andere Person ist in der Regel ein/e Held/in, die/der hilfebedürftigen Personen helfen muss. Dabei kommt es häufiger vor, dass eine Person gerettet werden muss, und dies notfalls mit Gewalt. Die ausgeübte Gewalt stellt eine Art Notwehr dar, bei der aber nicht die physische Unversehrtheit der eigenen, sondern die einer dritten Person geschützt wird.

Virtuelle Nothilfe

Im Folgenden wird diese Art von gewalthaltigem Hilfeverhalten daher als *Nothilfe* bezeichnet.

Anderson et al. (2010) vermuten, dass Nothilfe-Aufgaben beim Spieler eine Art „Helden“-Skript aktivieren könnten, das bestimmte Formen von Hilfeverhalten fördert. Bis zum Jahr 2010 hat sich allerdings keine Studie mit der Wirkung solcher Kombinationen von Gewalt- und Hilfeverhalten beschäftigt, wobei sich inzwischen einige in der Durchführung befinden (Anderson et al., 2010). Dies ist ein empirischer Missstand, da nach einer Studie der ASC Games (ohne Jahresangabe, zit. n. Vorderer, Bryant, Pieper & Weber, 2006) Rollenspiele in der Beliebtheit bei Computerspieler/innen knapp hinter Shootern auf Platz 2 rangieren.

Um diesen Missstand zu beheben, soll in dieser Studie der Frage nachgegangen werden, wie sich virtuelle Nothilfe auf das Hilfe- und Gewaltverhalten der Spieler/innen auswirkt. Begonnen wird mit einer Darlegung des theoretischen Hintergrunds für Computerspiele (Genre, Verbreitung, Forschungsbereiche, Unterhaltungswirkung, Präsenzgefühl). Im Anschluss werden Theorien des Gewaltverhaltens (mit dem Social Interactionist Model als Rahmentheorie) und übliche Möglichkeiten der Operationalisierung behandelt. Der *Competitive Reaction Time Task* wird als das am besten geeignete Verfahren ausgewählt. Strukturgleich werden in der Folge Theorien des Hilfeverhaltens (mit dem Motivationsmodell für Hilfeleistung als Rahmentheorie) und alle im Rahmen von prosozialen Spielen verwendeten Operationalisierungen für Hilfeverhalten diskutiert, wonach das Paradigma *Willingness to assist in another experiment* als praktikabelstes Messverfahren ausgewählt wird.

Zu Gewalt im Computerspiel wird eine Hypothese zum Gewaltverhalten abgeleitet, zu Hilfe im Computerspiel eine zum Hilfeverhalten. Diese Hypothesen stellen Replikationen der bisherigen Forschung dar. Darüber hinaus werden zu Nothilfe im Computerspiel auf Basis der diskutierten Theorien zwei Hypothesen zum Gewaltverhalten und zwei weitere zum Hilfeverhalten generiert. Um diese sechs Hypothesen prüfen zu können, wird das weit verbreitete Computerrollenspiel „The Elder Scrolls IV: Oblivion“ mit Hilfe eines sogenannten Editors modifiziert, um vier experimentelle Bedingungen zu realisieren („Helfen durch Töten“, „Helfen ohne Töten“, „Töten ohne Helfen“ sowie „Ohne Helfen und ohne Töten“). Gleichzeitig werden verschiedene Störvariablen erhoben, um ihren Einfluss kontrollieren zu können (Geschlecht, Spielkonsum, subjektive Erregung, diverse Eigenschaften der Spielaufgabe, Schwierigkeiten mit der Steuerung, Spieldauer, Vorkenntnisse mit Computerspielen, Verdachtsbildung, Nichtbefolgung, Zeitdruck, Schulform, Alter, Fachsemester, Erhebungswelle, Emotionen).

Nach drei Vorstudien wird das Experiment mit 186 Osnabrücker Studierenden durchgeführt, die als Voraussetzung Erfahrung in der Spielsteuerung mitbringen müssen und für ihre Teilnahme mit Versuchspersonenstunden bzw. zehn Euro vergütet werden. Das Experiment beginnt mit einem Täuschungsmanöver, gefolgt von der Anamnese und einem Training des *Competitive Reaction Time Task*. Danach spielen die Probanden das Computerspiel und ab-

Virtuelle Nothilfe

solvieren direkt im Anschluss den *Competitive Reaction Time Task*. Nach diesem füllen sie einen Fragebogen aus, in welchem eine Reihe der Störvariablen erhoben wird. Danach wird mittels *Willingness to assist in another experiment* das Hilfeverhalten gemessen. Das Experiment endet mit der Probandenaufklärung.

Die Auswertung der so erhobenen Daten beginnt mit einer Überprüfung der Datenkorrektheit und der Validität der unabhängigen und abhängigen Variablen. Hierbei fällt auf, dass die Stärke des Treatments „Hilfe im Spiel“ suboptimal ausfällt. Außerdem wird festgestellt, dass ca. ein Drittel der Probanden im Verlauf des *Competitive Reaction Time Task* diesen durchschauen, wobei das Durchschauen nicht mit den Versuchsbedingungen konfundiert ist. Hiernach erfolgt die Prüfung der internen Validität an, wobei ein unerwünschter Mediationseffekt identifiziert wird. Danach werden die Anwendungsvoraussetzungen für die geplanten statistischen Tests geprüft, wo nur unkritische Verletzungen gefunden werden.

Die statistische Prüfung der sechs Hypothesen führt teilweise zu unerwarteten Ergebnissen. Hinzu kommt, dass die Varianzaufklärung insgesamt nicht sehr groß ist und dass die Störvariablen zusammengekommen mehr Varianz aufklären als die unabhängigen Variablen. Die Ergebnisse werden zuerst im Rahmen der Ergebnisse der Validitätsprüfungen und im Anschluss im Rahmen der Theorien diskutiert. Danach wird auf die Generalisierbarkeit der Ergebnisse eingegangen. Die Arbeit schließt mit einer generellen Diskussion sowie einer Analyse der Stärken, Schwächen und Optimierungspotenziale der Studie. Im Anhang findet sich eine möglichst exakte Reproduktion der verwendeten Untersuchungsmaterialien, die eine Replikation der Studie erleichtern sollen.

4 THEORETISCHER HINTERGRUND

Es stellt sich die Frage, ob das spielerische Absolvieren von Aufgaben gewalttätigen Rettens (im Folgenden kurz als „Nothilfe“ bezeichnet) dazu führt, dass die Spieler/innen im realen Leben mehr Gewaltverhalten zeigen, oder dass sie mehr Hilfeverhalten zeigen, oder vielleicht auch beides zusammen. Diese Frage lässt sich zu der generelleren Frage umformulieren „Wie wirken Hilfe- und Gewaltanteile im Computerspiel auf aggressives bzw. helfendes Verhalten“. Die Computerspiele selbst fallen in den Forschungsbereich der Medienpsychologie, genauer in die Kategorien der „Unterhaltungsmedien“ und der „interaktiven Medien“. Daher soll der erste Abschnitt des Theorieteils Unterhaltungsmedien im Allgemeinen und der zweite Computerspiele im Speziellen behandeln. Im Anschluss befasst sich der dritte Abschnitt mit Theorien zur Aggression, der vierte mit Theorien zu Hilfeverhalten.

Wann immer möglich werden hierbei aus mehreren Gründen handlungstheoretische Theorien bevorzugt (zu den Unterschieden zwischen Handeln, Tun und Verhalten siehe Groeben, 1986; Erb, 1997; für eine stark verkürzte Zusammenfassung Mohseni, 2007, S. 5-7). Erstens sollte man aus ethischen Gründen¹ immer mit einer handlungstheoretischen Modellierung beginnen und nur dann, wenn für diese Modellierung keine empirische Geltung nachgewiesen werden kann, auf tuns- oder verhaltenstheoretische Modellierungen wechseln (Erb, 1997, S. 216f.). Zweitens bestehen bei den verhaltenstheoretischen Modellierungen von aggressivem bzw. helfendem Verhalten oft Probleme auf der definitorischen Ebene (siehe Abschnitt 4.3 bzw. 4.5). Drittens kann es leicht zu Inkonsistenzen führen, wenn handlungs- und verhaltenstheoretische Modellierungen zugleich verwendet werden, da erstere den Probanden Reflexivität und potenzielle Rationalität zusprechen und für unverzichtbar zur Erklärung des Handelns halten, während letztere den Probanden diese Eigenschaften absprechen bzw. für irrelevant halten. Vor der Beschäftigung mit diesem Problem im Rahmen der Theorien zu Aggression und Hilfe soll aber zuerst auf den Bereich der Medien und Computerspiele eingegangen werden.

4.1 UNTERHALTUNGSMEDIEN

Es erscheint sinnvoll, sich allgemein mit der psychologischen Wirkung von Unterhaltungsmedien zu beschäftigen, weil Unterhaltung inzwischen einen hohen Stellenwert im Alltag eingenommen hat. Bereits jetzt stehen den Menschen in den Industrieländern 30% der Wachzeit zur freien Verfügung (Zillmann, 2000, S. 17), wobei die meiste Zeit zum Fernsehen verwendet wird (Robinson, 1981, zit. n. Zillmann, 2000). Es besteht aber Uneinigkeit darüber, ob Unterhaltung eher positive oder eher negative Wirkungen hat. Diese Debatte geht auf die französischen Intellektuellen Pascal und Montaigne während des 16. und 17. Jahrhunderts zurück (Zillmann, 2000, S.15). Dabei vertrat Pascal als Vertreter christlicher Werte den Standpunkt, Unterhaltung sei eine Lust des Fleisches und daher abzulehnen, während Montaigne sie als Möglichkeit sah, der unbefriedigenden Realität für einen Moment zu entfliehen

(dies nennt sich heute „Eskapismus“). Interessanterweise gleichen sich die Standpunkte darin, dass sie voraussetzen, dass das Leben ein inadäquater Zustand ist, der überwunden werden muss – bei Pascal hin zu einem perfekten Leben nach dem Tode, bei Montagne hin zu einem Leben ohne Aversionen. Zillmann (2000, S. 15) weist daher zu Recht darauf hin, dass es noch eine dritte Position gibt – Unterhaltung als Bereicherung („enrichment“) eines guten Lebens.

Außerdem warnt Zillmann (2000, S. 15) davor, dass die Auseinandersetzung, ob Unterhaltung positiv oder negativ wirkt, häufig auf persönlichen Erfahrungen statt auf empirischem Wissen beruht. Es ist aber schwierig, die Positivwirkungen zu beurteilen, da es hier viel weniger Studien gibt als zu Negativwirkungen (Happ, Melzer & Steffgen, 2010b, S. 29). Aber auch bei der Beurteilung der empirischen Ergebnisse zu Negativwirkungen verbleibt das Problem, dass diese oft schwer zu deuten sind. Tedeschi und Felson (1994, S. 122) ziehen beispielsweise aus Langzeitstudien zur Wirkung von gewalthaltigen Fernsehprogrammen auf das Gewaltverhalten den Schluss, dass die bisherigen Studien keinen Zusammenhang nachweisen konnten. Trotzdem gehen sie davon aus, dass es wahrscheinlich einen kleinen Effekt gibt, wobei der genaue Wirkmechanismus unklar ist. Anderson, Gentile und Buckley (2007, S. 25) ziehen aus bisheriger Forschung zu Gewaltfernsehen hingegen den Schluss, dass es ein signifikanter Risikofaktor für Gewalt und Aggression sei.

Aus dieser generellen Debatte lassen sich folgende Forderungen ableiten: Es sollten nicht nur negative, sondern auch positive Wirkungen ins Auge gefasst werden. Diese sollten empirisch geprüft werden. Dabei sollte auf eine ausreichend große Stichprobe geachtet werden, da der Effekt möglicherweise nicht sehr groß ist. Außerdem sind Theorien zu bevorzugen, die nicht nur Zusammenhangstrukturen, sondern auch Wirkprozesse modellieren. Welche Theorien das sein könnten, soll im Folgenden diskutiert werden.

4.2 COMPUTERSPIELE

In Form einer Gegenstandsbestimmung soll als Erstes geklärt werden, was Computerspiele² von anderen Unterhaltungsmedien unterscheidet, und welche Varianten es gibt. Um ihre Alltagsrelevanz zu belegen, soll als Nächstes die Verbreitung von Computerspielen aufgezeigt werden. Danach soll dargestellt werden, welche Wirkungen von Computerspielen bereits untersucht wurden und welche nicht. Als Letztes soll analysiert werden, inwiefern die Wirkung von Computerspielen von ihrer Unterhaltungswirkung bzw. ihrer Absorptionskraft abhängt.

4.2.1 ABGRENZUNG VON ANDEREN UNTERHALTUNGS MEDIEN

Computerspiele sind ein vergleichsweise junges Medium. Aus diesem Grund werden zur Theoriebildung häufig (etwas) traditionellere Medien herangezogen. Grodal (2000, zit. n. Klimmt, 2001, S. 23) beispielsweise zieht eine Parallele zwischen Filmen und Computerspie-

len, mit dem Unterschied, dass in Computerspielen aktiv eingegriffen werden kann. Klimmt (2001) sieht die Besonderheit bei Computerspielen darin, dass hier Handlungsmöglichkeiten in Kombination mit Handlungsnotwendigkeiten geboten werden. Ein Film dagegen erzeuge zwar Handlungsnotwendigkeiten, biete aber keine Möglichkeit des Eingreifens, während ein Spielzeug Handlungsmöglichkeiten biete, aber keinen Handlungsdruck erzeuge. Die Kombination von Handlungsmöglichkeit und Handlungsnotwendigkeit solle eine besonders hohe Nutzungsmotivation erzeugen. Sie sei vergleichbar mit der Kombination von Fähigkeit und Tätigkeit im Flow-Konzept von Csikszentmihalyi (1999, zit. n. Klimmt, 2001, S. 30ff.), weswegen beim Computerspielen eine hohe Flow-Wahrscheinlichkeit bestehe.

Klimmt (2006a, S. 71) konzipiert den Prozess des Computerspielens als eine Kette von Input-Output-Loops, die auf einer höheren Ebene Spielepisoden bilden. In diesen Episoden haben Spieler bestimmte Ziele im Sinn, z. B. die Bearbeitung von Problemen und Herausforderungen. Eine Episode besteht dabei aus einer Ausgangslage, einer Handlung sowie einem Ergebnis. Die Ausgangslage ist durch Elemente geprägt, die autonom funktionieren und dabei Handlungsdruck auf die Spieler ausüben. Beispielsweise wird den Spielern in Kampfepisoden über die Narration meistens die Rolle einer Heldenfigur und den autonomen Elementen die Rolle des Feindes bzw. Bösewichts zugewiesen (Klimmt, 2006a, S. 73).

Computerspiele können sich hinsichtlich ihrer Eigenschaften stark voneinander unterscheiden. Eine Einteilung der Spiele nach ihren Eigenschaften erfolgt üblicherweise über Genres. Nach Klimmt (2006a, S. 21) sei diese Art der Einteilung inzwischen wenig sinnvoll, da die neuerdings auftretenden zahlreichen Mischformen keine exakte Abgrenzung mehr ermöglichen würden. Klimmt (2001b, zit. n. Klimmt, 2006a, S. 21) hält es daher für besser, Spiele mithilfe ihrer drei konstitutiven Merkmale zu beschreiben: (1) dem narrativen Kontext (d. h. der Rahmengeschichte), (2) der Spielaufgabe und (3) der Darstellungsform (d. h. der Präsentation des Spielraums sowie der Spielzeit). In diesem speziellen Fall scheint es aber trotz der fließenden Übergänge leicht möglich zu sein, zur anschaulichen Beschreibung der besonderen Eigenschaften des hier verwendeten Computerspiels auf Genres zurückzugreifen.

4.2.2 GENRES

Die Aufteilung in bestimmte Genres erfolgte bisher eher unsystematisch. Smith (2006) fasst die Aufteilungen verschiedener Autoren zusammen, indem er neun Genres herausstellt, die sich in jeder Aufteilung wiedergefunden haben: (1) Sportspiele, (2) Fahr- bzw. Rennspiele, (3) Simulationen, (4) Strategiespiele, (5) Rollenspiele, (6) Shooter, (7) Kampfspiele (im Deutschen oft als „Prügelspiele“ bekannt), (8) Action-Adventures³ und (9) Puzzlespiele. Da dieses Experiment mit einem Action-Rollenspiel durchgeführt wurde, sollen die betroffenen Genres „Action-Adventures“ und „Rollenspiele“ im Folgenden kurz beschrieben werden.

Nach Smith (2006, S. 51f.) sind die Entwicklung der Spielfigur und die Interaktion mit der Spielwelt die beiden konstitutiven Eigenschaften eines Rollenspiels. Der Autor dieses Textes

ist der Ansicht, dass hier eine dritte Eigenschaft ergänzt werden sollte, und zwar das Miterleben einer Geschichte (Narration), da Rollenspiele im Vergleich zu anderen Spielformen einen besonderen Schwerpunkt auf die Narration legen. Die ersten computerbasierten Rollenspiele beispielsweise waren Umsetzungen von klassischen Pen&Paper-Rollenspielen. In solchen Rollenspielen versetzen sich die Spieler in die Rolle einer fiktiven Figur. Im Extremfall wird diese Figur in schauspielerischer Art und Weise ausagiert (wie es z. B. auch beim sogenannten „Live-Rollenspiel“ der Fall ist). Ein solches Ausagieren ist in einer computerbasierten Umsetzung aber nicht möglich. Daher wurde in den klassischen computerbasierten Rollenspielen stattdessen eine umfangreiche Geschichte erzählt, in deren Verlauf dem Spieler immer wieder Handlungsalternativen ermöglicht wurden, welche die Geschichte (entscheidend) beeinflussten. Es gibt inzwischen aber auch Rollenspiele, bei denen diese Eigenschaft weniger bzw. so gut wie gar nicht vorhanden ist.

Rollenspiele sind ein verbreitetes Genre. Eine explorative Studie der Annenberg Studies on Computer Games Group zeigt beispielsweise, dass bei den Befragten $N = 297$ Spieler/innen die Shooter mit 57,5% das am meisten favorisierte Genre darstellen, direkt gefolgt von 54,8% für Rollenspiele (ASC Games, ohne Jahresangabe, zit. n. Vorderer, Bryant, Pieper & Weber, 2006).

Da das in diesem Experiment verwendete Action-Rollenspiel eine Mischung aus Action-Adventure und Rollenspiel darstellt, soll auch letzteres Genre kurz beschrieben werden. Bei Action-Adventures geht es nach Smith (2006, S. 52) hauptsächlich darum, dass der Spieler eine Reihe von Hindernissen überwindet bzw. Herausforderungen bewältigt, um ein bestimmtes Ziel im Spiel zu erreichen. Der Unterschied zum Rollenspiel würde darin bestehen, dass die Spielfigur im Action-Adventure statisch ist, während sie im Rollenspiel vom Spieler ausgestaltet würde. Nach Smith (2006, S. 52) können Ego-Shooter als ein Subgenre des Action-Adventures angesehen werden.

4.2.3 VERBREITUNG

Es erscheint sinnvoll, sich mit der psychologischen Wirkung von Computerspielen zu beschäftigen, weil diese unter den Unterhaltungsmedien anscheinend einen hohen Stellenwert eingenommen haben. Für Vorderer und Bryant (2006, S. XI) beispielsweise ist es offensichtlich, dass Computerspiele (zumindest für die jüngere Generation) zu einem der wichtigsten Unterhaltungsmedien geworden sind. Nach Klimmt (2006a, S. 22f.) ist es aber schwierig, zu Nutzungshäufigkeit bzw. Nutzungsintensität valide Daten zu bekommen, da es keine standardisierten Messsysteme mit repräsentativen Stichproben gibt. Die Zahlen würden meistens aus Befragungsstudien stammen sowie aus Marktdaten der Branchenverbände (in denen raubkodierte Spiele fehlen). Trotzdem würden die verfügbaren Zahlen eine näherungsweise Abschätzung der realen Nutzung ermöglichen, weswegen hier einige Befunde berichtet werden. Bereits im Jahr 2000 spielte jeder siebte Bundesbürger ab 14 Jahren min-

destens gelegentlich Computerspiele (Burda Advertising Center, 2000, zit. n. Klimmt, 2001, S. 22), während es im Jahr 2003 schon jeder vierte war (Burda-Medienforschung, 2004, zit. n. Klimmt, 2006a, S. 24). Der Anteil der 12- bis 19-Jährigen, die täglich oder mehrmals pro Woche Computerspiele spielen, ist von 37.6% in 1998 auf 40.9% in 2002 gestiegen (Feierabend & Klingler, 2003b, zit. n. Klimmt, 2006a, S. 23). Werden nur Jungen berücksichtigt, fallen die Werte mit 51.5% in 1998 bzw. 62.0% in 2002 sogar noch höher aus.

Diese starke Verbreitung findet sich aber nicht nur in Deutschland. Nach einer Studie der Entertainment Software Association spielen 50% aller Amerikaner Computerspiele, wobei das Durchschnittsalter der Spieler 29 Jahre beträgt und sich die Gruppe zu 39% aus Frauen zusammensetzt (ESA 2004b, zit. n. Vorderer, Bryant, Pieper & Weber, 2006). Ähnliche Zahlen finden sich schon bei Kindern. Eine Studie der Kaiser Family Foundation zeigt, dass 49% der Kinder im Alter von 0 bis 6 bereits ein Videospielgerät zu Hause stehen haben (Kaiser Family Foundation, 2003, zit. n. Vorderer, Bryant, Pieper & Weber, 2006). Jungen spielen im Durchschnitt 13, Mädchen 6 Stunden pro Woche Videospiele (Gentile, Walsh, Ellison, Fox & Cameron, 2004). Dabei scheinen Kinder immer mehr Zeit mit Videospielen zu verbringen (Anderson et al., 2007, S.7).

4.2.4 FORSCHUNGSBEREICHE

Bei der Erforschung von Computerspielen wurden nach Klimmt (2006a, S. 26) bisher hauptsächlich drei Bereiche untersucht: (1) geschlechtsspezifische Unterschiede im Interesse an und Erfolg bei Computerspielen, (2) mögliche aggressionsfördernde Wirkungen und (3) pädagogische Nutzungsmöglichkeiten. Lee und Peng (2006) benennen ebenfalls drei Bereiche, allerdings fehlt in ihrer Aufzählung die Untersuchung geschlechtsspezifischer Unterschiede; stattdessen führen sie die Untersuchung genereller Wirkungen von Computerspielen an. Der Erforschung prosozialer Wirkungen von prosozialen Spielen wurde dagegen bisher wenig Beachtung geschenkt (Gentile et al., 2009; vgl. auch Abschnitt 4.5.6).

In beiden Aufzählungen zeigt sich, dass für den Bereich der Emotionen lediglich die Aggression (in Form aggressiver Gefühle) genauer betrachtet wurde, obwohl diese im Alltag nicht unbedingt zu den wichtigsten Emotionen zählt. In einer Untersuchung von Mees und Schmitt (2003, S. 234f.), bei der $N = 202$ Probanden typische Beispiele für Gefühle nennen sollten, erhielt die Aggression zum Beispiel nur 19 Nennungen und rangierte damit auf Platz 22. Angesichts dessen scheint es möglich, dass andere Emotionen auch beim Spielen eine größere Rolle als die Aggression spielen könnten.

4.2.5 UNTERHALTUNG UND PRÄSENZGEFÜHL

Bei der Erforschung von Computerspielen macht es aus zwei Gründen Sinn, die Unterhaltungswirkung des verwendeten Spiels zu maximieren: Zum einen sollten Probanden umso mehr Teilnahmemotivation aufbringen, je unterhaltender das zu spielende Computerspiel

Virtuelle Nothilfe

auf sie wirkt. Dies sollte die Bereitschaft erhöhen, das Spiel bis zum Ende der zu lösenden Aufgabe zu spielen, ohne zwischendurch aus Langeweile Pausen einzulegen oder sogar ab-zubrechen. Zum anderen ist für Klimmt und Vorderer (2003, zit. n. Tamborini & Skalski, 2006, S. 235) die Unterhaltungswirkung nichts anderes als eine angenehme Form der Absorption in das Spiel (engl. „absorption“) bzw. des Präsenzgefühls im Spiel (engl. „presence“). Klimmt (2006a, S. 146-156) konnte beispielsweise zeigen, dass sowohl der inhaltliche als auch der darstellungsbezogene Realismus des Spiels (die das Präsenzgefühl erhöhen sollten) den Unterhaltungswert erhöhen. Tamborini (2000, zit. n. Tamborini & Skalski, 2006, 235f.) nimmt außerdem an, dass das im Spiel simulierte Verhalten umso stärker auf das Verhalten im Alltag einwirkt, je mehr die Spieler das Gefühl haben, im Spiel zu sein. Zur Optimierung der Effektstärke scheint es daher sinnvoll, durch die Operationalisierung ein möglichst hohes Präsenzgefühl zu erzeugen. Dazu ist es notwendig, sich die Forschungsergebnisse zu Unterhaltungs- und Präsenzwirkung genauer anzusehen.

Zur Unterhaltungswirkung von Computerspielen existieren nur wenige Forschungsergebnisse, da sie erst in neuerer Zeit von einer Forschungsgruppe untersucht wurde, die sich um Klimmt herum gebildet hat (z. B. in Klimmt, 2006a). Im Rahmen dieser Forschung wurden verschiedene Emotionen als relevant für die Unterhaltungswirkung befunden (z. B. Stolz, Freude, Spaß, Frustration etc.). Hierzu liegen allerdings vergleichsweise wenige Studien vor. Für diese Untersuchung ist insbesondere der Befund von Klimmt (2006a, S. 153-159; 172-178) relevant, dass die Wirkung eines Spiels gleichzeitig von den Eigenschaften der Spielaufgabe als auch von den Eigenschaften der Spieler abhängt. Beispielsweise errechnet sich die subjektive Wahrscheinlichkeit P , eine Aufgabe im Spiel zu bewältigen, nach folgender Formel (Klimmt 2006a, S. 173):

$$P_{(\text{Bewältigung, subjektiv})} = \text{Handlungsmöglichkeiten} - \text{Handlungsnotwendigkeit} + \text{Spielkompetenz}$$

Ein Computerspiel erzeugt eine Handlungsnotwendigkeit in der Regel dadurch, dass ein/e Spieler/in das Spiel verliert, wenn er/sie bestimmte Handlungen unterlässt. Die objektive Spielaufgabenschwierigkeit ergibt sich aus der Differenz von Handlungsmöglichkeiten (die z. B. beim Schach in der Anzahl der eigenen Figuren und der sich daraus ergebenden Zugmöglichkeiten bestehen würden) und Handlungsnotwendigkeit (beim Schach die Anzahl und die daraus resultierenden Zugmöglichkeiten der gegnerischen Figuren). Je höher die Fähigkeit des Spielers im Vergleich zu dieser objektiven Aufgabenschwierigkeit ist, desto eher wird die Spielaufgabe bewältigt. Die sich so ergebende subjektive Schwierigkeit steht im Zusammenhang mit der Unterhaltungswirkung. Nach Csikszentmihalyi (1999, zit. n. Klimmt, 2001, S. 30) führt eine Überforderung zu Angst, während eine Unterforderung zu Langeweile führt. Klimmt (2001) folgert daraus, dass eine zu schwierige Spielaufgabe beängstigend (bzw. frustrierend) sein kann. Nach Anderson et al. (2010) führen zu schwierige bzw. zu schnelle Spiele zu Frustration und Ärger, was wiederum aggressive Kognitionen aktivieren kann. Zur Vermeidung von Frustration werden Computerspiele daher so gestaltet, dass die Schwierigkeit

Virtuelle Nothilfe

sich nach und nach steigert, damit Über- und Unterforderung ausbleiben (Klimmt, 2001). Werden sie dennoch als zu leicht oder zu schwer wahrgenommen, dann sind sie nicht mehr unterhaltsam (Klimmt, Hartmann & Frey, 2007, zit. n. Hartmann & Vorderer, 2010). Klimmt (2006a, S. 146-156) konnte beispielsweise zeigen, dass eine schlechte Leistung negative Emotionen fördert. Dies gilt insbesondere bei leichten Aufgaben, wahrscheinlich, weil hier das „Versagen“ auf die eigene Fähigkeit attribuiert wird. Gleichzeitig konnte Klimmt nachweisen, dass positive Emotionen nur bei ausreichend hoher Schwierigkeit der Aufgabe auftreten. Dabei gibt es nach Klimmt (2006a) Erfahrungsunterschiede, die diesen Effekt moderieren. Geübte Spieler bewältigen auch schwierige Spielabschnitte, weil sie „mit hohem Tempo alle denkbaren Handlungsmöglichkeiten überblicken, die bestmögliche Wahl zwischen diesen Handlungsmöglichkeiten treffen und die ausgewählte Handlungsoption mit dem notwendigen Geschick umsetzen“ (S. 72). Das macht es umso schwieriger, eine Spiel-aufgabe so zu gestalten, dass sie sowohl für ungeübte als auch für geübte Spieler nicht zu leicht und nicht zu schwer ist.

Neben der Spielschwierigkeit spielen auch negative Gefühle und unter Umständen sogar die Hintergrundmusik eine Rolle für den Unterhaltungswert. Klimmt (2006a, S. 146-156) konnte beispielsweise zeigen, dass negative Gefühle den Unterhaltungswert senken. Zehnder und Lipscomb (2006) versuchten, einen direkten Effekt von Musik im Spiel auf die Immersion nachzuweisen (Immersion beschreibt nach Tamborini und Skalski (2006) das Gefühl, von einer virtuellen Umgebung umgeben zu sein und mit dieser zu interagieren). Ein direkter Effekt konnte nicht nachgewiesen werden, es ergaben sich aber signifikante Interaktionen (die allerdings nicht berichtet werden), wenn man Geschlecht, Alter und Schulabschluss mitein-bezog.

Neben dem Unterhaltungswert spielt auch das Präsenzgefühl eine Rolle dabei, wie stark ein Computerspiel auf den Spieler einwirkt. Wie bereits erwähnt, wird von Klimmt und Vorderer (2003, zit. n. Tamborini & Skalski, 2006, S. 235) die Unterhaltungswirkung mit einer angenehmen Form von Absorption bzw. Präsenzgefühl gleichgesetzt. Beim Präsenzgefühl handelt es sich nach Tamborini und Skalski (2006) um den Eindruck, „in der Spielwelt“ zu sein, während es sich nach Lombard und Ditton (1997, zit. n. Tamborini & Skalski, 2006, S. 226) um den Eindruck handelt, dass die Wahrnehmungen nicht durch ein Medium vermittelt werden (engl. „non-mediation“). Das Konstrukt des Präsenzgefühls lässt sich nach Tamborini und Skalski (2006) in drei Dimensionen einteilen: (1) räumliches Präsenzgefühl (engl. „spatial presence“), (2) soziales Präsenzgefühl (engl. „social presence“) und (3) Selbst-Präsenzgefühl (engl. „self-presence“). Beim räumlichen Präsenzgefühl handelt es sich um den Eindruck, sich physisch in einer virtuellen Umwelt aufzuhalten bzw. virtuelle physische Objekte als real wahrzunehmen. Das soziale Präsenzgefühl beschreibt den Eindruck, dass virtuelle soziale Akteure real sind. Das Selbst-Präsenzgefühl besteht in der Wahrnehmung, dass das virtuelle Selbst dem realen entspricht.

Virtuelle Nothilfe

Nach Steuer (1992, zit. n. Tamborini & Skalski, 2006, S. 228) ergibt sich das *räumliche* Präsenzgefühl durch die Interaktion von Wahrnehmungsbreite (d. h. die Anzahl der gleichzeitig aktivierten Wahrnehmungskanäle) und –tiefe (d. h. die Intensität der transportierten Information, die unter anderem auch von der realistischen Darstellung abhängt), wobei es aber auch bei geringer Breite und Tiefe vorhanden sein kann, wenn die Wahrnehmungen auf den Kanälen aufeinander abgestimmt sind (z. B. wenn man die Auswirkungen einer Spielhandlung gleichzeitig sehen, hören und durch Vibrationen des Spielcontrollers fühlen kann). Das räumliche Präsenzgefühl wird nach Tamborini und Skalski (2006) durch Involviertheit (engl. „involvement“) und Immersion bedingt. Die Involviertheit resultiere aus dem Bedeutungsgehalt, der durch Form und Inhalt einer Umgebung hergestellt werde. Das könne zum einen dadurch erreicht werden, dass die Objekte und Szenerien im Spiel so arrangiert werden, wie es von den Spielern erwartet wird, und zum anderen dadurch, dass der dramatische Inhalt und die bedeutungsvolle Handlung des Spiels mit den Erfahrungen der Spieler in Einklang gebracht werden. Die Immersion hingegen ergäbe sich durch die von der Technik hergestellte Kontrolle der Umgebungsreize. Daher würde Immersion in dem Maße ansteigen, in dem das Medium (1) den Rezipienten von seiner physischen Umgebung abschirme, in dem es (2) in ihm den Eindruck erzeuge, in der virtuellen Umgebung zu existieren, statt nur ein Beobachter zu sein, sowie in dem es (3) das Gefühl hervorrufe, sich in natürlicher Weise in der virtuellen Umgebung bewegen zu können sowie mit dieser natürlich interagieren zu können.

Das *soziale* Präsenzgefühl lässt sich nach Tamborini und Skalski (2006) in die drei Dimensionen „Kopräsenzgefühl“ (engl. „copresence“), psychologische Involviertheit (engl. „psychological involvement“) und die verhaltensmäßige Beschäftigung (engl. „behavioral engagement“) unterteilen. Kopräsenz beschreibt nach Goffmann (1959, zit. n. Tamborini & Skalski, 2006, S. 231) die Wahrnehmung, dass eine andere Person anwesend ist, es kann sich nach Tamborini und Skalski (2006) aber auch um das räumliche Präsenzgefühl handeln, dass eine andere Person anwesend ist und/oder dass man sich gegenseitig wahrnimmt (engl. „mutual awareness“). In Videospielen seien meistens virtuelle Figuren sichtbar, was ein Kopräsenzgefühl verursachen würde. Um das Gefühl gegenseitiger Wahrnehmung zu erzeugen, reicht nach Biocca et al. (2003, zit. n. Tamborini & Skalski, 2006, S. 231) die Sichtbarkeit aber nicht aus, solange nicht auch der Eindruck vorhanden ist, dass das Gegenüber über Intelligenz verfügt. Dies wird durch den Begriff der psychologischen Involviertheit abgebildet, die nach Tamborini und Skalski (2006) so etwas wie einen Sinn für Intelligenz darstellt. Damit diese Involviertheit eintritt, ist es nach Dennett (1987, zit. n. Tamborini & Skalski, 2006, S. 232) nötig, dass die virtuelle Figuren Hinweisreize auf ihre „intentionalen“ Zustände geben (z. B. indem die Figuren mit dem Spieler kommunizieren). In eine ähnliche Richtung geht der Begriff der verhaltensmäßigen Beschäftigung, mit dem nach Tamborini und Skalski (2006) das sozial erscheinende Verhalten der Spielfiguren gemeint ist. Die Autoren nennen als Beispiele „reden, plaudern, sich abwechseln (engl. „turn taking“), Augenkontakt und nonverbales Spiegeln“. Zu Letzterem gibt es eine Studie von Bracken et al. (2003, zit. n. Tamborini & Skalski, 2006, S.

232), die belegt, dass Spieler die nonverbalen Hinweisreize der Spielfiguren für eine Reihe von Zwecken nutzen, unter anderem zur Vorhersage des gegnerischen Verhaltens.

Welche Auswirkungen hat nun das Präsenzgefühl? Nach Tamborini (2000, zit. n. Tamborini & Skalski, 2006, S. 235f.) bauen Spieler durch ihre Handlungen im Spiel (d. h. das Vorbereiten von Plänen, das Treffen von Entscheidungen sowie das motorische Ausführen ihrer Pläne) mentale Modelle von diesen simulierten Handlungen auf. Daher nimmt er an, dass das im Spiel simulierte Verhalten umso stärker auf das Verhalten im Alltag einwirkt, je mehr die Spieler das Gefühl haben, im Spiel zu sein. In eine ähnliche Richtung geht die Annahme von Klimmt (2001), dass Computerspiele wegen der aktiven Eingriffsmöglichkeit emotional besonders stark wirken, da die Spieler hier nicht die (Miss-)Erfolge der Protagonisten lediglich miterleben, sondern über ihre Handlungen im Spiel eigene (Miss-)Erfolge erleben können. Das gilt auch für aggressive Emotionen. Klimmt, Schmid, Nosper, Hartmann und Vorderer (2006) weisen z. B. darauf hin, dass die Verbindung zwischen medialer Darstellung und Gewaltverhalten bei Computerspielen stärker ausfallen müsste, da hier der Konsument nicht passiv zuschaut, sondern aktiv virtuelle Gewalt ausübt. Einen Beleg dafür stellt die Studie von Ploman, Orobio de Castro und van Aken (2008, zit. n. Happ, Melzer & Steffgen, 2010b, S. 25) dar, in deren Rahmen gezeigt wurde, dass das Spielen eines Gewaltspiels eine stärkere Wirkung hatte als das bloße Zuschauen bei demselben Spiel. Zusätzlich führt die wahrgenommene Gewalt im Spiel zu einer Erhöhung des Präsenzgefühls (Nowak, Krcmar & Farrar, 2008, zit. n. Happ, Melzer & Steffgen, 2010b, S. 26).

4.3 GEWALTVERHALTEN

Nachdem im vorigen Abschnitt geklärt wurde, was Computerspiele sind und welche Prozesse ihre Wirkung beeinflussen, sollen als Nächstes Theorien dargestellt werden, die erklären wollen, wieso Computerspiele eine Auswirkung auf das Gewaltverhalten haben (können). Die im Bereich der Gewaltspielforschung dafür am meisten genutzte Theorie ist das General Aggression Model (abgekürzt „GAM“), welches daher zuerst beschrieben werden soll. Im Anschluss sollen die inhärenten Schwächen dieser Theorie aufgezeigt werden, die es sinnvoll erscheinen lassen, bei der Hypothesenbildung auch andere Theorien zugrunde zu legen. Das Social Interactionist Model bietet beispielsweise eine gute Grundlage, um eine stringente Definition der Kernbegriffe zu leisten, und spannt gleichzeitig einen aus ethischen Gründen zu bevorzugenden handlungstheoretischen Rahmen auf. Das Moral Management Model hingegen scheint gut dazu geeignet zu sein, die in Computerspielen relevanten moralischen Prozesse abzubilden. Daher sollen diese Theorien im Anschluss an das GAM erläutert werden. Das Purifikationskonzept im Rahmen des Catharsis Model soll ebenfalls aufgeführt werden, da es Ansätze zur Überwindung der negativen Auswirkungen von Gewaltdarstellungen bietet. Die letzte erwähnenswerte Theorie stellt das Catalyst Model dar, da es die einzige Theorie mit einem Schwerpunkt auf genetische Prädispositionen ist.

4.3.1 GENERAL AGGRESSION MODEL

Das General Aggression Model wurde entwickelt, um Kernideen aus den folgenden früheren Modellen zu integrieren: der sozialen Lerntheorie und verwandten sozialkognitiven Theorien, dem kognitiven neoassoziationistischen Modell, dem sozialen Informationsverarbeitungsmodell, dem affektiven Aggressionsmodell, der Skripttheorie und dem Erregungstransfermodell⁴ (Anderson et al., 2007, S. 40f.). Das GAM basiert auf der Annahme, dass menschliche Erinnerungen, Gedanken und Entscheidungsprozesse als komplexe assoziative Netzwerke dargestellt werden können (siehe Abb. 1), in denen die Knoten kognitive Konzepte und Emotionen repräsentieren (Anderson et al., 2007, S. 40f.). Diese Netzwerke beschreiben, wie sich eine aggressive Persönlichkeit durch Lernprozesse herانبildet (Anderson et al., 2007, S. 43). Das Netzwerkmodell ist aber ein generelles Lernmodell, welches sich auf alle möglichen Lerninhalte beziehen lässt. Daher wird die Wirkung von Gewaltspielen in der ursprünglichen Fassung des GAM zusätzlich mit den folgenden spezifischeren Ursachen erklärt (Anderson & Bushman, 2001, S. 355):

1. Beobachtungslernen aggressiver Verhaltensweisen,
2. Priming von Aggressionskognitionen,
3. Erzeugen eines aggressiven emotionalen Zustands oder
4. Erhöhung der physiologischen Erregung.

Mit diesen Ursachen will das GAM erklären, warum Gewaltspiele die Aggressionsbereitschaft steigern und die Bereitschaft zu prosozialem Verhalten mindern. Die von den Spieler/innen aus ihrer Innensicht genannten Gründe für ihre Spielaktivitäten sind dagegen für die Erklärung irrelevant. Es handelt sich somit um ein verhaltenstheoretisches Modell, was prinzipiell kein Nachteil sein muss. Es besteht aber ein Problem hinsichtlich der Ursache „physiologische Erregung“, weil diese nicht spezifisch für aggressives Verhalten ist, sondern jedes gerade ablaufende Verhalten unspezifisch aktiviert (Zillmann, 1983, zit. n. Tedeschi & Felson, 1994, S. 66). Somit führt eine erhöhte Erregung nicht zwingend zu einer höheren Aggressionsbereitschaft. Bei den aggressiven Kognitionen lässt sich ebenfalls darüber streiten, ob sie tatsächlich erklärungskräftig sind. Während nach Anderson und Dill (2000) aggressive Kognitionen der entscheidende Mediator für Gewaltverhalten sein könnten, halten Ferguson et al. (2008a) aggressive Kognitionen als Erklärung des Gewaltverhaltens für unbrauchbar, da es nach dem Konsum gewalthaltiger Spiele nicht verwunderlich sei, dass die Spieler anschließend an die Gewaltinhalte des Spiels denken würden. Für eine Operationalisierung von Aggression hat dies zur Folge, dass man besser nicht (ausschließlich) aggressive Kognitionen bzw. die physiologische Erregung erheben sollte, sondern auf die Messung aggressiver Emotionen bzw. aggressiven Verhalten zurückgreifen sollte.

In der aktuellsten Fassung unterscheidet das GAM zwischen proximalen Ursachen und Prozessen, die in der aktuellen Situation wirken, und distalen, die erst über einen längeren Zeitraum ihre Wirkung entfalten (siehe Abb. 2). Die distalen Ursachen bestehen dabei aus biolo-

Virtuelle Nothilfe

gischen und umweltbedingten Faktoren. Zu den biologischen Faktoren zählen ein niedriger Serotoninspiegel, eine geringe physiologische Erregbarkeit, Defizite in der kognitiven Kontrolle, ADHS sowie andere genetische Faktoren; zu den umweltbedingten Faktoren zählen familiäre Sitten, kommunale Gewalt, eine Kultur der Gewalt, Mediengewalt sowie extreme soziale Verhältnisse (Anderson et al., 2007, S. 47).

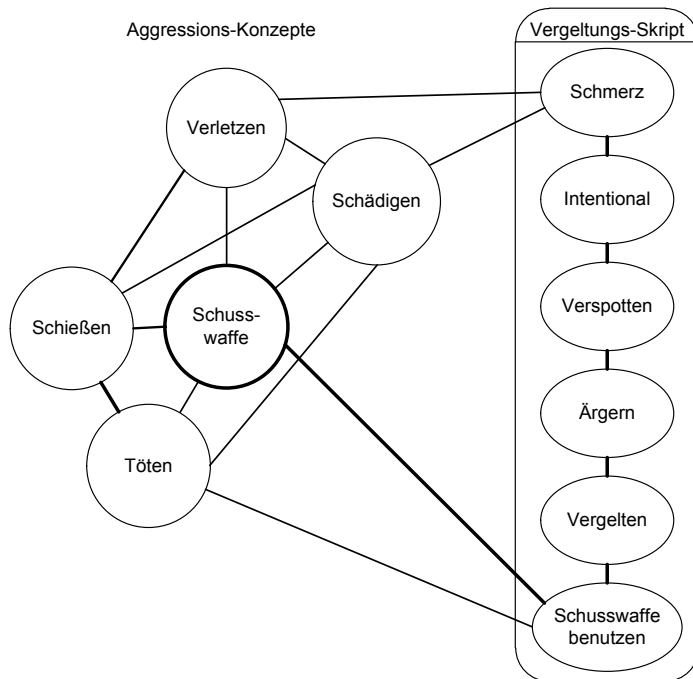


Abb. 1: Vereinfachtes assoziatives Netzwerk mit Aggressions-Konzepten und einem Vergeltungs-Skript (Anderson, Benjamin & Bartholow, 1998, entnommen aus Anderson et al., 2007, S. 43)

Die Faktoren fördern aggressive Überzeugungen und Einstellungen, aggressive Wahrnehmungsschemata, aggressive Erwartungsschemata und aggressive Verhaltensskripte sowie die Aggressionsdesensibilisierung. Dies wiederum führt zur Ausbildung einer aggressiven Persönlichkeit. Nach Anderson et al. (2010) führt am ehesten die kognitive Route zu einer permanenten Persönlichkeitsänderung, da die Wirkungen des negativen Affekts und der physiologischen Erregung sich relativ schnell verflüchtigen würden. Die aggressive Persönlichkeit beeinflusst zum einen, mit welchen aggressionsförderlichen Eigenschaften eine Person in eine potenziell aggressionsauslösende Situation hineingeht, und zum anderen, wie situative Auslöser von der Person wahrgenommen werden. Zu den aggressionsförderlichen Eigenschaften der Person zählen ein erhöhter/instabiler Selbstwert, proaggressive Überzeugungen, proaggressive Einstellungen, aggressive Verhaltensskripte, mit Aggression verbundene Ziele sowie der Glaube, in einer feindseligen Welt zu leben (engl.: „hostile world schema“); zu den situativen Auslösern zählen Provokation, Frustration, Schmerz und Unbehagen, schlechte Laune, sozialer Stress sowie aggressive Primes, zu denen auch Gewaltmedien gezählt werden (Anderson et al., 2007, S. 47).

Virtuelle Nothilfe

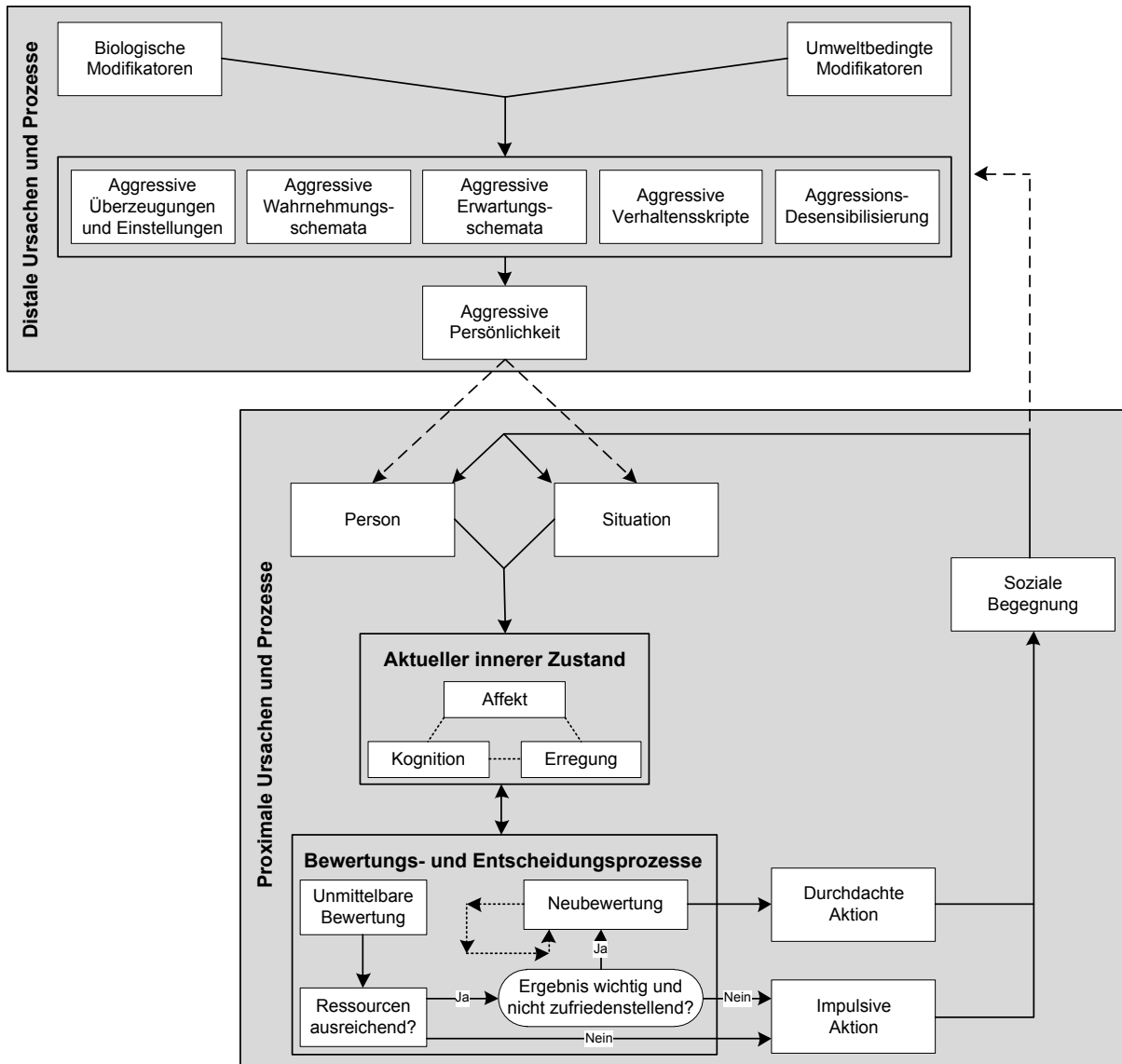


Abb. 2: Das „General Aggression Model“ der Gewalt (zusammengestellt aus drei Abbildungen in Anderson et al., 2007)

Situative und personale Faktoren wirken sowohl additiv als auch in Interaktion miteinander und bilden auf diese Weise den aktuellen inneren Zustand. Der innere Zustand wiederum beeinflusst zahlreiche Bewertungs- und Entscheidungsprozesse. Diese münden schlussendlich entweder in einer durchdachten oder einer impulsiven Aktion, die auf die soziale Begegnung einwirkt und somit die nächste Runde im Kreislauf des proximalen Prozesses begründet. Die soziale Begegnung hat aber auch einen Einfluss auf die distalen Prozesse bzw. Ursachen und somit auf das Verhalten in zukünftigen Begegnungen.

4.3.2 CATALYST MODEL

Das GAM lässt sich den sozialen Lerntheorien zuordnen, welche die Ursache der Aggression (hauptsächlich) in der Umwelt verorten. Dieser Theorie stellen Ferguson et al. (2008a) eine

Virtuelle Nothilfe

biologische Persönlichkeitstheorie namens „Catalyst Model“ gegenüber, die sie aus den Ausführungen von Pinker (2002) entwickelt haben (siehe Abb. 3). Im Catalyst Model wird die Ursache der Aggression (hauptsächlich) in der genetischen Veranlagung in Form einer genetischen Prädisposition (hauptsächlich bei Männern) verortet, die im Kindesalter zu einem aggressiven Temperament und nach dem Erwachsenwerden zu einer aggressiven Persönlichkeit führt. Umweltfaktoren, insbesondere der Einfluss familiärer Gewalt, moderieren die biologischen Ursachen. Die Theorie besagt weiterhin, dass aggressive Menschen dann mehr Gewaltverhalten zeigen, wenn sie durch die Umwelt stärker belastet werden. Je aggressiver ein Mensch ist, desto schwächer darf der Stressor sein, um Gewaltverhalten auszulösen. Nur unter extremen Umständen (wie z. B. kriegsähnlichen Zuständen) besteht die Möglichkeit, dass eine größere Anzahl an Menschen aggressives Verhalten zeigt. Umweltfaktoren sind somit nicht die Ursache des Gewaltverhaltens, aber sie können es verstärken.

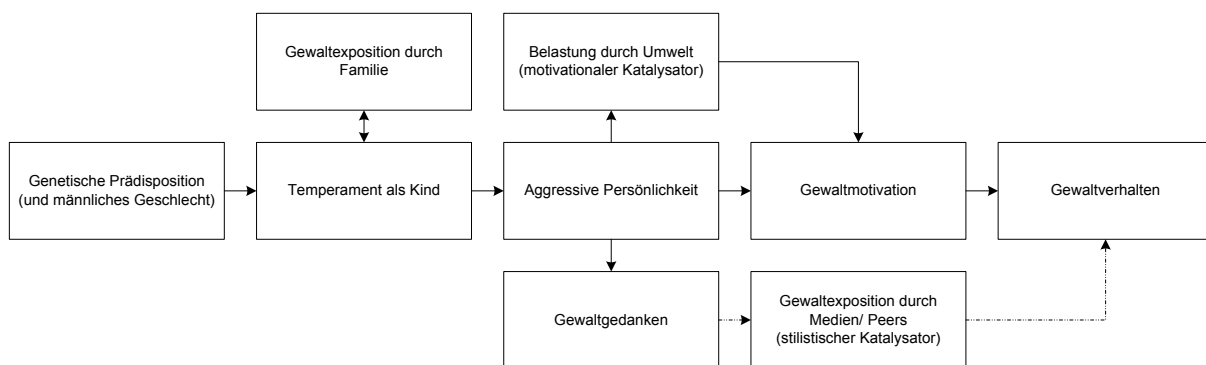


Abb. 3: Das „Catalyst Model“ der Gewalt (nach Ferguson et al., 2008a)

In einer solchen Theorie ist auch die Gewaltexposition durch die Medien (einschließlich gewalthaltiger Computerspiele) keine Ursache für Gewaltverhalten. Stattdessen haben Gewaltspiele die Rolle stilistischer Katalysatoren, d. h., eine aggressive Person zeigt Gewaltverhalten in einer Form, wie sie sie bereits in einem Computerspiel gesehen hat. Hätte sie das Spiel nicht konsumiert, würde sie dennoch gewalttätig werden, lediglich die Form wäre eine andere. Es wird angenommen, dass aggressive Personen aktiv nach Modellen für ihr Gewaltverhalten suchen, während nicht-aggressive Personen nach nicht-aggressiven Modellen suchen. Wenn man ihnen die Wahl lässt, würde eine aggressive Person somit Gewaltspiele bevorzugen, während nicht-aggressive Personen gewaltfreie Spiele bevorzugen würden. Diese aktive Selbstselektion ist in der Theorie der Grund, warum der Konsum von Gewaltspielen mit Gewaltverhalten korreliert. Das Catalyst Model ist eine relativ neue Theorie. Sie wurde im Rahmen der Gewaltspielforschung bisher nur bei Ferguson et al. (2008a) auf ihre Gültigkeit geprüft. Dabei kontrastieren die Autoren eine randomisierte Zuordnung zu einem gewalthaltigen bzw. einem gewaltfreien Computerspiel mit einer von den Probanden selbst gewählten Zuordnung. Sollte sich zwischen den randomisierten Bedingungen ein Unterschied zeigen, würde dies laut den Autoren für die Gültigkeit des GAM sprechen. Sollte sich hier kein Unterschied zeigen, dafür aber bei selbst gewählter Zuordnung, dann würde dies stattdessen

für die Gültigkeit des Catalyst Models sprechen, da hier die aggressiven Personen die Gewaltspielbedingung bevorzugen müssten. Wie von den Autoren angenommen zeigte sich ein Unterschied nur bei selbst gewählter Zuordnung, was für das Catalyst Model und gegen das GAM spricht. Leider enthält die Studie gravierende methodische Schwächen. So erkennen die Autoren zwar, dass die Nullhypothese der Forschungshypothese entspricht, aber sie führen weder eine Betafehlerkontrolle noch eine Poweranalyse durch. Zusätzlich sind die Gruppen so klein, dass nur große Effekte gefunden werden können. Hinzu kommt, dass bei einer Bonferroni-Korrektur statt des Betafehlers der Alphafehler korrigiert wurde. Nimmt man dies alles zusammen, so wird das GAM beim Theorienvergleich ziemlich benachteiligt. Außerdem ist die Selbstselektion mit dem Geschlecht konfundiert, da bei den Bedingungen mit selbst gewählter Zuordnung das Geschlechterverhältnis stark unterschiedlich ist (2 Männer zu 1 Frau in der gewalthaltigen Bedingung vs. 1 Mann zu 5 Frauen in der gewaltfreien Bedingung). Die Autoren verwenden das Geschlecht als Kovariate in einer ANCOVA, dies ist aber aufgrund der Konfundierung nicht sinnvoll. Kovariaten müssen nämlich unabhängig von den unabhängigen Variablen sein (vgl. Field, 2009, S. 397f.; Tabachnick & Fidell, 2007, S. 200; Bortz & Döring, 1995, S. 509), weil sich der Zusammenhang zwischen der abhängigen und der unabhängigen Variable verringert, wenn eine Kovariate dieselbe Varianz in der abhängigen Variable erklärt wie eine der unabhängigen Variablen. Somit sollten aus den Ergebnissen der Studie keine voreiligen Schlüsse hinsichtlich der Überlegenheit des Catalyst Models gezogen werden.

4.3.3 SOCIAL INTERACTIONIST MODEL

Tedeschi und Felson (1994) argumentieren, dass eine stringente Definition von aggressivem Verhalten nicht möglich ist, ohne sich dabei auf die Intentionen des Verhaltenden zu beziehen. Sie plädieren dafür, die bisherigen verhaltenstheoretischen Definitionen (wie z. B. die im Rahmen des GAM) aufzugeben und eine handlungstheoretische Sichtweise zu nutzen, die sie selbst als „social interactionist view“ bezeichnen. Zum Beleg führen sie die bisher erfolgten Definitionen von Aggression an. Die erste stammt von Buss (1961, zit. n. Tedeschi & Felson, 1994, S. 160), der aggressives Verhalten definiert als „jegliches Antwortverhalten, das anderen Organismen einen schädigenden Stimulus verabreicht“ (engl. „any response that delivers noxious stimuli to another organism“). Tedeschi und Felson (1994) kritisieren drei Dinge an dieser Definition: erstens umfasst sie auch unbeabsichtigte Schädigungen (z. B. Unfälle), zweitens lässt sie Schädigungen durch absichtliche Unterlassung unberücksichtigt (z. B. wenn man jemanden verbluten lässt) und drittens sind misslungene absichtliche Schädigungen ebenfalls nicht beinhaltet (z. B. wenn man jemanden erschießen will, aber daneben trifft). Zumindest Letzteres bessert Buss (1971, , zit. n. Tedeschi & Felson, 1994, S. 161) durch eine Änderung der Definition nach, bei der aggressives Verhalten „den Versuch [darstellt], einen schädigenden Stimulus zu verabreichen, auch wenn dieser Versuch nicht erfolgreich ist“ (engl. „the attempt to deliver noxious stimuli regardless of whether it is successful“). Ab-

Virtuelle Nothilfe

gesehen davon, dass bei dieser Definition Unterlassungshandlungen weiterhin unberücksichtigt bleiben, wird von Tedeschi und Felson (1994, S. 161f.) außerdem kritisiert, dass der Begriff des Versuchs ohne Rückgriff auf Intentionen nicht sinnvoll definierbar ist. Aus dem beobachtbaren Verhalten allein lässt sich nicht erkennen, ob jemand etwas „versucht“. Dazu ist es notwendig, sich in dessen Lage zu versetzen und so auf die hinter dem Verhalten liegenden Absichten zu schließen.

Baron und Richardson (1994) definieren Aggression daher als ein Verhalten, bei dem einer anderen Person absichtlich eine Schädigung zugefügt wird, obwohl diese das nicht möchte. Die Schädigung gegen den Willen des Anderen soll Aggression abgrenzen gegenüber rationalen Schädigungen (z. B. ein chirurgischer Eingriff, bei dem ein Teil des Körpers beschädigt wird, um einen Tumor zu entfernen) sowie gegenüber masochistischen Praktiken. Allerdings wird der Begriff der Absicht nicht weiter expliziert. Diese begriffliche Vagheit würde es laut Tedeschi und Felson (1994, S. 349) Frustrations-Aggressions-Theoretikern ermöglichen, reaktive Aggression gleichzeitig als intentional und unabsichtlich zu begreifen. Außerdem würden einige Autoren Motive mit Intentionen verwechseln (Tedeschi & Quigley, 2000). Nach Schutz (1967, zit. n. Tedeschi & Felson, 1994, S. 165) bezieht sich eine Intention auf ein erwünschtes Handlungsergebnis, während das Motiv den Grund dafür darstellt, warum das Ergebnis erwünscht ist. Wenn beispielsweise ein Kind gegenüber einem Geschwister eine Drohung ausspricht, um an die Fernbedienung zu gelangen, dann besteht der intentionale Teil in der Drohung, während der motivationale Teil in dem Wunsch besteht, ein bestimmtes Fernsehprogramm sehen zu können. Auch für diese Arbeit ist die Unterscheidung relevant, weil die Probanden nach den Intentionen für ihr aggressives Verhalten gefragt werden (siehe Abschnitt 11.4.1) und von diesen Intentionen auf die dahinterliegenden Motive geschlossen wird.

Einen empirischen Beleg für die Relevanz von Intentionen sehen Tedeschi und Felson (1994) in einer Studie von Epstein und Taylor (1967, zit. n. Tedeschi & Felson, 1994, S. 224) sowie in einer Studie von Ohbuchi und Kambara (1985, zit. n. Tedeschi & Felson, 1994, S. 224). In der erstgenannten Studie wurde der Competitive Reaction Time Task (CRTT) eingesetzt, wobei die Probanden die Einstellungen des „Gegners“ (der in Wirklichkeit ein Computerprogramm ist, siehe dazu Abschnitt 4.4.1.2) sehen konnten. Im CRTT hört der Proband nur dann einen unangenehmen Ton, wenn er eine Runde verloren hat. Die Einstellungen des Gegners haben daher in den Runden keinen Effekt, in denen der Proband gewinnt. Dennoch bemaßen die Probanden ihre Einstellungen auch danach, was der Gegner in diesen Runden für sie vorgesehen hatte. Es kam also nicht auf die tatsächliche Schädigung durch den Gegner an, sondern auf die beabsichtigte Schädigung. In der Studie von Ohbuchi und Kambara wurde die Manipulation so gestaltet, dass die tatsächliche und die beabsichtigte Schädigung gegenläufig waren. Hier zeigte sich, dass die Probanden ihre Einstellungen nach der beabsichtigten und nicht nach der tatsächlichen Schädigung bemaßen. Daneben benennen die Autoren zahlreiche empirische Belege gegen die Geltung der Frustrations-Aggressions-Theorien bzw.

der Theorie reaktiver Aggression. Da diese hier nicht relevant sind, sei der interessierte Leser an Tedeschi und Felson (1994) verwiesen.

Tedeschi und Felson (1994) betrachten im Rahmen ihrer Theorie aggressives Handeln als eine von drei Varianten des nötigenden⁵ Handelns (engl. „coercive behavior“). Die Erweiterung des Aggressionskonzepts um Nötigungen hat den Vorteil, dass schädliche Handlungen miteingefasst werden, bei denen die Schädigung Bestandteil der Nötigung ist, aber vom Täter nicht beabsichtigt wird. So besteht z. B. die Intention bei einem Raubüberfall in der Erlangung von Wertgegenständen und nicht in der Schädigung des Opfers, obwohl das Opfer eindeutig geschädigt wird; somit ist ein erfolgreicher Raubüberfall kein aggressiver Akt, aber eindeutig eine Nötigungshandlung (Tedeschi & Felson, 1994). Die Autoren unterscheiden bei Nötigungshandlungen die drei Varianten „Drohung“ (engl. „threat“), „Bestrafung“ (engl. „punishment“) und „körperliche Gewalt“ (engl. „bodily force“). Unter einer „Drohung“ verstehen sie eine Botschaft, die einer Person die Absicht verkündet, ihr einen Schaden zuzufügen. Drohungen können entweder zur Erzwingung von Gefügigkeit oder zur Schädigung dienen (z. B. indem sie Angst erzeugen). Unterschieden werden Drohungen, die an eine Bedingung geknüpft sind, von solchen, bei denen das nicht der Fall ist. Eine bedingte Drohung (engl. „contingent threat“) hat die Form „tust du nicht das, was ich will, dann tue ich X“, wobei X für eine schädigende Handlung steht, während eine unbedingte Drohung (engl. „non-contingent threat“) die Form „Ich werde X tun“ aufweist. Bei bedingten Drohungen besteht keine Schädigungsabsicht, bei unbedingten Drohungen schon. „Bestrafungen“ unterscheiden sich von Drohungen darin, dass sie keine Botschaften, sondern Handlungen sind, bei denen eine Schädigungsabsicht besteht. Bestrafungen sind damit eine Umschreibung für dasselbe Phänomen, welches bei anderen Autoren als „aggressives Verhalten“ bezeichnet wird. Die Autoren verwenden einen neuen Terminus, weil der Begriff möglichst wertneutral sein soll, was beim Begriff des aggressiven Verhaltens eindeutig nicht der Fall ist. Der Einfachheit halber soll in diesem Rahmen trotzdem der übliche Terminus beibehalten werden. Das Konzept der „Drohung“ wird jedoch übernommen, da die Drohung ohne Schädigungsabsicht nicht vom Aggressionsbegriff abgedeckt wird und bei dem zu wählenden Messverfahren durchaus eine Rolle spielen kann. Unter der letzten Variante des Nötigungsverhaltens, der „körperlichen Gewalt“, verstehen die Autoren die Verwendung physischen Kontakts, um eine andere Person zu einem bestimmten Handeln zu zwingen oder sie auf ein bestimmtes Handeln einzuschränken. Im Gegensatz zur Bestrafung besteht hier das Ziel nicht in der Schädigung des Gegenübers, sondern in der Erzwingung von Gefügigkeit. Unter die Definition fällt sowohl das Ausnutzen körperlicher Überlegenheit als auch die Verwendung von Waffen sowie die Nutzung des Überraschungsmoments. Auch dieses Konzept wird von den Autoren übernommen, da es bei dem zu wählenden Messverfahren relevant werden kann.

In der Theorie von Tedeschi und Felson (1994, S. 349) ist die Schädigung immer instrumenteller Natur, d. h., sie dient der Erreichung distaler Ziele (z. B. der Erzielung ausgleichender Gerechtigkeit). Die Autoren unterscheiden dabei drei Formen der Schädigung: (1) physische

Virtuelle Nothilfe

Bestrafung, (2) Deprivation und (3) soziale Schädigung. Die Entscheidung für oder gegen Nötigungshandlungen betrachten die Autoren im Rahmen einer rationalen Entscheidungstheorie (engl. „rational choice theory“). Nötigungshandlungen sind dann umso wahrscheinlicher, wenn ihr subjektiver Nutzen die subjektiven Kosten übersteigt. Kosten können durch das Ziel der Nötigung auferlegt werden (z. B. wenn Vergeltung zu befürchten ist), durch unbeteiligte Dritte (z. B. wenn diese die Nötigung sanktionieren) sowie durch die Umstände (z. B. wenn sich das Opfer der Nötigung in bestimmten Situationen besser oder schlechter wehren kann). Es ist trotzdem möglich, dass Nötigungshandlungen trotz großer damit verbundener Kosten erfolgen, und zwar dann, wenn die Nötigung als das geringste mehrerer Übel angesehen wird. Bei der Kosten-Nutzen-Abwägung ist außerdem zu beachten, dass diese nicht nur die mit der Nötigung verfolgten Ziele (terminale Werte; engl. „terminal values“) betrifft, sondern auch die Handlungsprozedur selbst (prozedurale Werte; engl. „procedural values“). Zum Beispiel gibt es Personen, die eine Nötigung an sich als etwas so Negatives ansehen, dass sie unter keinen Umständen bereit sind, eine solche durchzuführen.

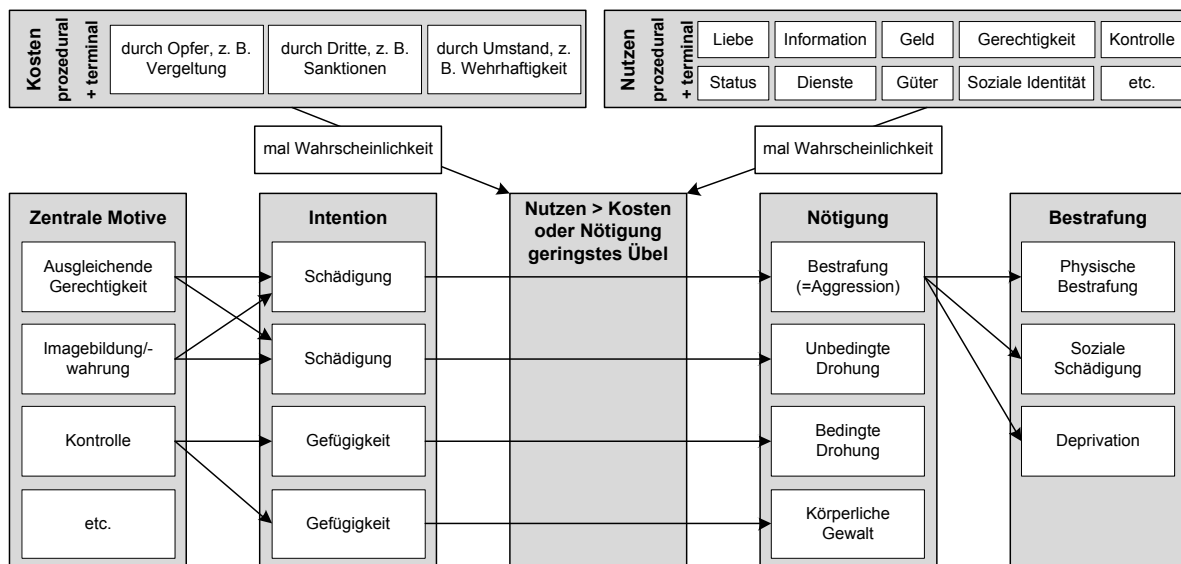


Abb. 4: Vereinfachtes „Social Interactionist Model“ der Gewalt (in Anlehnung an Tedeschi & Felson, 1994)

Beim Nutzen benennt Foa (1971, zit. n. Tedeschi & Felson, S. 183) als typische Motive z. B. Liebe, Status, Information, Dienste, Güter, Geld, Gerechtigkeit, soziale Identität etc. (engl.: love, status, information, money, goods and services, justice, social identity). Davon halten Tedeschi und Felson (1994) drei bestimmte Motive für besonders relevant, um Nötigungshandlungen zu erklären, wobei sich die Motive nicht gegenseitig ausschließen: (1) Ausübung von Kontrolle über das Handeln anderer Personen, (2) Ausgleichung von Gerechtigkeit und (3) Etablierung/Bewahrung des Image. Bei der Motivlage der ausgleichenden Gerechtigkeit führt die Handlung einer anderen Person zu einem Beschwer, das durch eine Bestrafung ausgeglichen werden soll, um so die Gerechtigkeit wiederherzustellen. Die Stärke der Bestrafung richtet sich dabei direkt nach der Höhe des Beschwer, kann aber auch davon beeinflusst werden, dass man den Übeltäter durch die Bestrafung umerziehen möchte und/oder

Virtuelle Nothilfe

dass man durch die Bestrafung vor weiteren Attacks abschrecken möchte. Bei der Motivlage der Etablierung/Bewahrung des Image kann es darum gehen, bei anderen den Eindruck zu erwecken, man sei mächtig, mutig und kompetent. Es kann aber auch darum gehen, das Gesicht zu wahren oder als „harter Kerl“ zu gelten.

4.3.4 MORAL MANAGEMENT MODEL

Eine handlungstheoretische Theorie, die im Gegensatz zum Social Interactionist Model direkt aus dem Bereich der Gewaltspielforschung stammt, findet sich bei Klimmt, Schmid, Nosper, Hartmann und Vorderer (2006; 2008). Die Autoren kritisieren an der bisherigen Gewaltspielforschung sowie an den zugehörigen Theorien (z. B. dem GAM), dass zwar die Auswirkung des Konsums untersucht wird, nicht aber die dahinterliegenden psychologischen Mechanismen. Die Theorien würden höhere kognitive Prozesse außen vor lassen. Außerdem bliebe unberücksichtigt, dass der Gewaltkonsum bei Computerspielen in einen Unterhaltungskontext eingebettet ist. Beides sei aber relevant, da Gewaltmedien verstärkt konsumiert werden, wenn sie unterhaltsam sind, und die Unterhaltungswirkung wiederum hänge von einer komplexen höheren kognitiven Verarbeitung ab. Dazu passt auch der Befund von Potter und Tomasello (2003, zit. n. Klimmt et al., 2006), nach welchem die subjektive Verarbeitung medialer Gewalt für aggressives Verhalten relevanter ist als die objektive Intensität der Gewaltdarstellung im Medium.

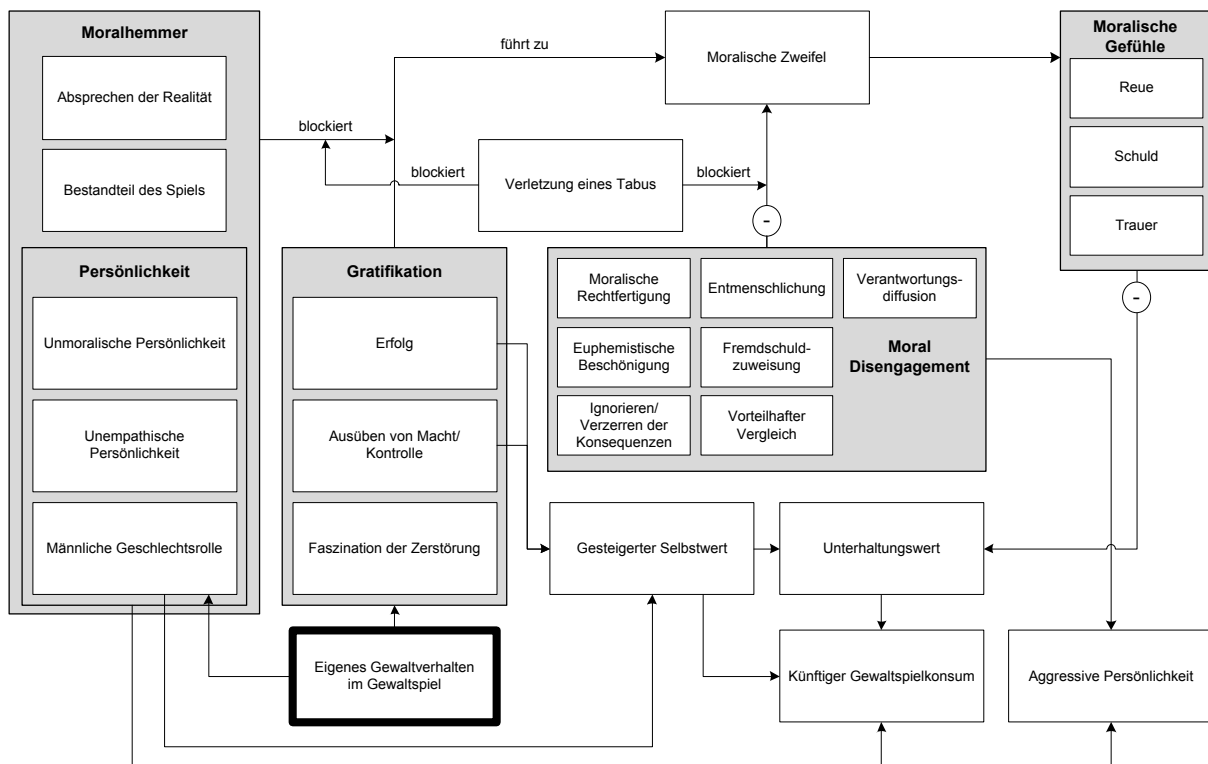


Abb. 5: Das „Moral Management Model“ der Gewalt (in Anlehnung an Klimmt et al., 2006; Klimmt et al. 2008; Hartmann, im Druck)

Virtuelle Nothilfe

Es ist verwunderlich, dass Gewaltspiele überhaupt unterhaltend wirken, da Gewalttaten normalerweise unangenehme Gefühle der Reue, Schuld und Trauer beim Täter verursachen (Düwell, Hübenthal & Werner, 2002, zit. n. Klimmt et al., 2008). Klimmt et al. (2006; 2008) konstruieren daher eine Theorie des moralischen Managements, die die Phänomene in einen logisch konsistenten Zusammenhang bringen soll. Im Folgenden wird diese Theorie als „Moral Management Model“ bezeichnet.

Nach dem Moral Management Model (siehe Abb. 5) können der ästhetische Genuss (bzw. die Faszination) beim Zerstören von Objekten, das Ausüben von Macht/Kontrolle, das Erleben von Erfolgen beim Lösen von Aufgaben/Herausforderungen sowie das Heranbilden/Ausleben einer männlichen Geschlechtsrollenidentität Gründe für virtuelle Gewalt sein. Macht- und Erfolgserlebnisse sowie das erfolgreiche Bilden/Ausleben einer männlichen Identität führen zu einem gesteigerten Selbstwert. Die Erfolgserlebnisse sind größer, wenn das Spiel Gewalt erfordert, da es sich bei bewaffneten Konflikten um Wettbewerbssituationen mit hohem Einsatz handelt. Der gesteigerte Selbstwert wiederum erhöht zum einen den Unterhaltungswert des Spiels und führt zum anderen zu einem (super)männlichen, positiven und stolzen Selbstbild. Ein hoher Unterhaltungswert sowie eine Verstärkung eines positiven Selbstbildes erhöhen die Wahrscheinlichkeit, dass auch in Zukunft Gewaltspiele konsumiert werden.

Wenn virtuelle Gewalt ausgeübt wird, kann sie zu moralischen Zweifeln führen. Dies liegt laut Hartmann, Toz und Brandon (2010) darin begründet, dass Nutzer sowohl den Computer als auch computerbasierte Agenten wie Menschen behandeln. Hartmann und Vorderer (2010) definieren virtuelle Gewalt daher als „Schädigung quasi-sozialer Charaktere“ (engl. „harm to quasi-social characters“). Das Auftreten von moralischen Zweifeln wird dadurch wahrscheinlicher, dass es sich in der Regel um extreme Gewalt handelt, die nicht nur gegen Objekte, sondern auch gegen Personen ausgeübt wird. Nur wenn Gewalt im Hinblick auf ein männliches Selbstbild als etwas absolut Positives aufgefasst wird, sollten überhaupt keine moralischen Zweifel auftreten. Moralische Zweifel können aversive Gefühle wie Reue, Schuld und Trauer auslösen, welche den Unterhaltungswert reduzieren und somit die Wahrscheinlichkeit senken, auch in Zukunft Gewaltspiele zu konsumieren.

Die moralischen Zweifel können aber durch verschiedene Strategien abgewehrt werden, um den Unterhaltungswert des Spiels aufrechtzuerhalten. Eine Strategie, die nur im Computerspiel möglich ist, ist die Unterscheidung zwischen der virtuellen Realität und der Wirklichkeit (Realitäts-Spiel-Unterscheidung). In Computerspielen werden keine lebendigen Wesen verletzt und keine realen Objekte zerstört. Somit sind die virtuellen Gewalttaten nicht real („Es handelt sich nur um ein Spiel“). Diese Strategie funktioniert aber nicht immer, da im Falle eines hohen Involvements bzw. einer hohen Immersion die Unterscheidung zwischen Spiel und Realität schwierig wird (Klimmt et al., 2006). Ein hohes Involvement wiederum ist für Spieler erstrebenswert, da es den Unterhaltungswert steigert (Sherry, 2004; Skalski et al.,

Virtuelle Nothilfe

2006; beide zit. n. Hartmann et al., 2010). Hartmann et al. (2010) finden es daher unwahrscheinlich, dass Spieler kontinuierlich den Realismus des Spiels anzweifeln. Zusätzlich zu dieser Abwehrstrategie können Spieler ihre Gewalt aber auch auf verschiedene Arten rechtfertigen. Klimmt et al. (2006; 2008) integrieren dazu die Theorie des „Moral Disengagement“ (dt.: „Freimachung von der Moral“) von Bandura (2002) in ihr Modell, in welchem die folgenden Möglichkeiten aufgeführt werden:

- Moralische Rechtfertigung (engl.: „moral justification“): Rechtfertigen der Gewalt durch Rückgriff auf höhere soziale Normen. Zum Beispiel ist die Gewalt nötig, um die Demokratie zu schützen.
- Euphemistische Beschönigung (engl.: „euphemistic labeling“): Verwenden von Begriffen, die die Gewalt weniger schlimm erscheinen lassen. Beispielsweise wird statt von „töten“ von „neutralisieren“ gesprochen.
- Ignorieren/Verzerren der Konsequenzen (engl.: „disregard/distortion of consequences“): Herabspielen der Gewaltauswirkungen, was die Gewalt weniger schlimm erscheinen lässt.
- Entmenschlichung (engl. „dehumanization“): Absprechen menschlicher Würde bzw. menschlicher Qualitäten und/oder Zuschreibung tierischer Qualitäten, um moralische Bedenken abzuschwächen.
- Fremdschuldzuweisung (engl.: „attribution of blame“): Rechtfertigen der Gewalt durch das Argument, das Opfer hätte es nicht anders verdient.
- Vorteilhafter Vergleich (engl.: „advantageous comparison“): Vergleichen der eigenen Gewalttaten mit schlimmeren Taten anderer Personen, die die eigene Gewalt weniger schlimm erscheinen lassen.
- Verantwortungsdiffusion (engl. „displacement/diffusion of responsibility“): Zuschreiben der Verantwortung zu Anderen (z. B. Kommandierenden). Gewalt ist die notwendige Konsequenz eines Befehls „von oben“ oder ein Ergebnis abstrakter sozialer Prozesse (z. B. des „Systems“), wodurch das Gewalt ausübende Individuum sich nicht die Schuld an der Gewalt gibt.

Bei der Anwendung der Abwehrstrategien sind verschiedene Varianten denkbar (Klimmt et al., 2008): Beispielsweise könnte die Rahmenhandlung des Spiels („Retter der Demokratie“ etc.) genutzt werden, um alle darauffolgenden Gewalttaten zu rechtfertigen (vgl. Klimmt et al., 2006). Alternativ könnten bestimmte Fälle von Gewalt bestimmte Kombination von Abwehrstrategien aktivieren. Beispielsweise könnte der Spieler eines Rollenspiels zu der Ansicht gelangen, Monstertyp A verdiene es einfach, getötet zu werden (=Fremdschuldzuweisung), während Monstertyp B getötet werden müsse, weil der König es verlangt (=Verantwortungsdiffusion). Nach Hartmann und Vorderer (2010) erfolgt die Abwehr automatisch, weswegen sie einer Person nicht unbedingt bewusst ist. Eine regelmäßige Anwendung der Abwehrstrategien im Gewaltspiel sollte nach Klimmt et al. (2008) theoretisch (über Lernprozesse und/oder Priming) dazu führen, dass diese Strategien auch im realen Leben schneller

Virtuelle Nothilfe

und leichter Anwendung finden, was langfristig zur Ausbildung einer aggressiven Persönlichkeit führt.

Gewaltspiele enthalten außerdem Hinweisreize, die Spielern die Abwehr moralischer Bedenken erleichtern sollen (Klimmt et al., 2008). Diese Hinweisreize sollen es ermöglichen, dass der Spieler Gewalt anwenden kann, ohne sich dadurch schuldig, reuig oder traurig zu fühlen. Im Moral Management Model sind solche Hinweisreize die Ursache für die virtuelle Gewalt (Hartmann & Vorderer, 2010) und nicht etwa dysfunktionale Persönlichkeitsmerkmale wie im Catalyst Model (siehe Abschnitt 4.3.2). Klimmt et al. (2008) ordnen den Abwehrstrategien der Moral Disengagement Theorie typische Hinweisreize in Spielen zu:

- **Moralische Rechtfertigung:** Die Rahmenhandlung der meisten Gewaltspiele gibt eine moralische Rechtfertigung für die Gewalt, typischerweise in Form einer Gut-gegen-Böse-Geschichte (z. B. „Verteidiger der Freiheit“, „Beschützer der Schwachen“).
- **Euphemistische Beschönigung:** Die meisten Gewaltspiele nutzen militärische Kontexte, in denen sich die Strategien wiederfinden, die Kommandanten im realen Leben zur Beschönigung von Gewalt verwenden (z. B. „Eliminieren des Feindes“, „Neutralisieren der Bedrohung“).
- **Ignorieren/Verzerren der Konsequenzen:** Solche Hinweisreize finden sich in fast allen Gewaltspielen, sowohl auf der narrativen Ebene (es tauchen keine Witwen oder Waisen der Getöteten im Spiel auf), als auch auf der visuellen (Leichen verschwinden nach kurzer Zeit).
- **Entmenschlichung:** Viele Spiele enthalten Monster und Fantasiegestalten, die sich in bestimmter Hinsicht von normalen Menschen unterscheiden (z. B. „Zombies“). Es gibt aber auch Formen impliziter Entmenschlichung durch Rückgriff auf Stereotype und Fremdgruppenabwertung (z. B. „islamistischer Terrorist“).
- **Fremdschulduzuweisung:** Diese Strategie wird in Spielen hauptsächlich dadurch gefördert, dass sich der Spieler in einer Notwehrsituation wiederfindet. Gegnerische Figuren, die den „unschuldigen“ Spieler angreifen, haben eine gewalttätige Antwort verdient, da sie selbst zu ungerechtfertigter Gewalt gegriffen haben.
- **Vorteilhafter Vergleich:** In Gewaltspielen tauchen oft Kreaturen auf, die extrem rohe Formen von Gewalt verwenden. Im Vergleich zu den Gewalttaten dieser Kreaturen erscheinen die des Spielers moderat.
- **Verantwortungsdiffusion:** Gewaltspiele mit militärischem, polizeilichem oder geheimdienstlichem Kontext fördern diese Strategie, da sie Autoritäten (z. B. „Regierungsmitglieder“) enthalten, die die Gewalt anordnen.

Zur Prüfung des noch relativ jungen Moral Management Model wurden bereits einige Studien durchgeführt. Beispielsweise haben Klimmt et al. (2006) Tiefeninterviews an zehn Computerspielern mit unterschiedlicher Spielerfahrung vorgenommen. Die Befragten rechtfertigten ihren Gewaltspielkonsum theoriekonform mit den Abwehrstrategien „Realitäts-Spiel-

Unterscheidung“, „Fremdschuldzuweisung (Notwehr)“, „Moralische Rechtfertigung (das Böse bekämpfen)“, „Entmenschlichung“ und „Euphemistische Beschönigung“. Etwas abweichend von der ursprünglichen Theorie fand sich aber auch das Argument, die Gewalt sei ein notwendiger Bestandteil von spielerischen Wettbewerben. Zwar ähnelt die Strategie dem „Ignorieren der Konsequenzen“, weil durch die Betonung des Wettbewerbsaspekts die negativen Konsequenzen in den Hintergrund gerückt werden. Auf der anderen Seite wird die Schädigung (z. B. Tötung des Gegners) des Gegenübers aber nicht beabsichtigt, sondern lediglich in Kauf genommen. Ein vergleichbares Phänomen findet sich im Bereich des Sports, wo der sportliche Kontext bei gewalthaltigen Sportarten (z. B. Boxen) die Moralstandards der beteiligten Personen verändert (Bredemeier & Shields, 1986, zit. n. Hartmann, im Druck). Zum Beispiel ist es beim Boxen völlig in Ordnung, dem Gegenüber direkt ins Gesicht zu schlagen. Im Computerspielbereich scheint diese Art der Rechtfertigung insbesondere bei Multiplayer-Spielen Anwendung zu finden, wo es vermehrt auf den Wettbewerb ankommt. Moralische Bedenken spielen hier anscheinend gar keine Rolle mehr (Klimmt et al., 2006). Ein wieder besser zur ursprünglichen Theorie passendes Ergebnis bei Klimmt et al. (2006) war, dass Gewalt gegen Personen in einigen Aussagen durchaus als etwas Positives bewertet wird, während sich in anderen Aussagen zeigt, dass sich dies ins Gegenteil verkehren kann, wenn starke moralische Bedenken (z. B. Gewalt gegen Kinder/Unschuldige, explizite Darstellung von Schmerzen etc.) induziert werden. Hierdurch zeigt sich, dass moralische Konflikte nicht vollständig durch einen Fokus auf den Wettbewerb beseitigt werden können. Hartmann et al. (2010) haben zwei Experimente durchgeführt, um zu untersuchen, inwiefern das Schuldgefühl von der Empathie (als Persönlichkeitsmerkmal) und von der durch das Spiel gegebenen Gewaltrechtfertigung abhängt. In beiden Experimenten fand sich ein Interaktionseffekt in der Art, dass Personen sich nur bei der Anwendung von ungerechtfertigter Gewalt schuldig fühlen, nicht aber bei der Anwendung gerechtfertigter Gewalt. Das Schuldgefühl ist dann umso stärker, je mehr Empathie die Personen besitzen. Dieser Interaktionseffekt war aber knapp nicht signifikant, was bei der ersten Studie an der zu kleinen Stichprobe ($N = 49$) und bei der zweiten am zu schwachen Treatment ($\Delta M = 0.49$, $.90 \leq SD \leq .94$) gelegen haben könnte. Trotzdem belegen die Experimente, dass Gewaltspiele moralische Bedenken verursachen können. Eine unveröffentlichte Studie von Hartmann (2011, zit. n. Hartmann, im Druck) zeigt, dass bei Personen mit hohen moralischen Standards der Gewaltspielkonsum sowie die Freude an virtueller Gewalt geringer sind. Hartmann und Vorderer (2010) konnten in einer Studie zeigen, dass Spieler weniger Schuld, Scham, Nervosität und Irritation verspüren, wenn ihr Gegner sich vorher unmoralisch verhalten hat (d. h. der Gegner hat Zivilisten erschossen). In der Studie war dieses Treatment trotz Randomisierung mit der Störvariable „nur ein Experiment“ konfundiert (d. h. wie sehr die Probanden ihr Verhalten mit dem Argument „das ist nur ein Experiment“ rechtfertigten), weswegen das Ergebnis laut den Autoren eher explorativen Wert hat. Daher führten sie ein weiteres Experiment durch, bei dem das Problem nicht auftrat. Hier resultierte das Ergebnis, dass Spieler weniger Schuld und weniger negativen Affekt verspürten, wenn sie Spielfiguren töteten, deren sozialen Hintergrund

sie nicht kannten. Außerdem ergab sich ein Interaktionseffekt in der Art, dass der Unterhaltungswert größer war, wenn entweder die Gewalt gerechtfertigt war und ihre Konsequenzen sichtbar, oder wenn die Gewalt ungerechtfertigt war und die Konsequenzen nicht sichtbar.

Neben den Studien, die direkt zur Prüfung des Moral Management Model durchgeführt wurden, finden sich auch Studien von Autoren, die gänzlich andere Ziele verfolgten, deren Ergebnisse aber das Moral Management Model stützen. Zwei Studien von Konijn et al. zeigen, dass aggressives Verhalten reduziert wird (Konijn, Bijvank & Bushman, 2007) bzw. die Empathie steigt (Konijn et al., 2008, zit. n. Happ, Melzer & Steffgen, 2010a), wenn man Empathie für das Opfer induziert. Eine Studie von Happ, Melzer und Steffgen (2010a) wiederum belegt, dass Spieler mit hoher Empathie in einem Gewaltspiel weniger Spaß und weniger Erfolg haben, und dass männliche Spieler generell weniger Empathie und Mitleid zeigen. Außerdem konnten die Autoren zeigen, dass durch die experimentelle Induktion von Empathie das Mitleid mit dem Gegner steigt, insbesondere wenn man die Rolle eines Bösewichts übernehmen muss.

Aufgrund der Ergebnisse bisheriger Studien modifizierte Hartmann (im Druck) das Moral Management Model in mehreren Punkten. Als Erstes ergänzte er drei weitere moralhemmende Faktoren: die Persönlichkeitseigenschaften „Empathiemangel“ (engl. „low on trait empathy“) und „Moralmangel“ (engl. „weak internalization of prosocial norms“) sowie die Strategie „Bestandteil des Spiels“ (engl. „part of the game“). Als Nächstes führte Hartmann (im Druck) das Konzept des Tabus ein. Es gibt Verstöße gegen Moralprinzipien, die so extrem sind, dass die Moralhemmer bzw. die Abwehrstrategien nicht ausreichen. Als Letztes präzierte er die Abwehrstrategie „Absprechen der Realität“ (er nennt sie „just a game“ bzw. „not really happening“), indem er ein duales Prozessmodell einführte. Im dualen Prozessmodell unterscheidet Hartmann (im Druck) zwischen einer erfahrungsbasierten (engl. „experiential“) und einer reflexiven (engl. „rational“) Informationsverarbeitung. Die erfahrungsbasierte Verarbeitung basiere eher auf assoziativen als auf logischen Verknüpfungen. Reize würden mit sensorischen Mustern im Gedächtnis verglichen und bei einer Übereinstimmung als real wahrgenommen. Beispielsweise fühle sich ein realistisch dargestellter computerbasierter Agent oder eine plastische virtuelle Szenerie intuitiv „real“ an (Schubert, 2009, zit. n. Hartmann, im Druck). Dies fördere ein Gefühl der Präsenz. Beide Verarbeitungsmodi liefen parallel ab. Dadurch könne sich eine mediale Darstellung „real“ anfühlen, obwohl der Mediennutzer wisse, dass sie es nicht ist. Dies löse den Widerspruch auf zwischen dem Gefühl von Spielern, dass ein Spiel real sei („Präsenz“) und der Aussage von Spielern, dass ein Spiel „nicht real“ sei. Die erfahrungsbasierte Verarbeitung sei der vorherrschende Verarbeitungsmodus. Dies liege erstens daran, dass sie schneller ablaufe. Zweitens würde sie weniger kognitive Verarbeitungskapazität benötigen. Dadurch würde bei Pseudorealitäten zeitlich vorangehend ein emotionaler Eindruck induziert, der erst nachträglich als unecht identifiziert würde (Zillmann, 2006, zit. n. Hartmann, im Druck). Drittens wäre dieses als unecht identifizieren nicht im Sinne des Spielers, da es zur Distanzierung gegenüber dem Spiel und somit zu

einem sinkenden Unterhaltungswert führe. Bei Hartmann und Vorderer (2010) finden sich weitere Studien, die das Primat der erfahrungsbasierten Verarbeitung belegen.

4.3.5 CATHARSIS MODEL

Bei der modernen Begriffsauslegung im Rahmen behavioristischer Forschung wird unter Katharsis die Abfuhr aufgebaute aggressiver Spannung durch aggressives Verhalten verstanden, was im Anschluss das Auftreten weiteren aggressiven Verhaltens reduzieren soll. Nach Tedeschi und Felson (1994, S. 365) würden alle Forscher nach einer Prüfung bisheriger Forschungsergebnisse zu dem Schluss kommen, dass Katharsis nicht zu einer nachfolgenden Reduktion aggressiven Verhaltens führen würde, sondern in einigen Fällen sogar zu einer Erhöhung (siehe dazu auch weiter unten die Ergebnisse zur sog. aktionspsychologischen Purgation). Somit könnte man zu dem Schluss kommen, dass die Katharsistheorie (engl. „Catharsis Model“) widerlegt ist. Scheele (1999) sowie Anderson et al. (2007, S. 144-149) weisen aber darauf hin, dass der Katharsisbegriff in der Psychologie nicht dem ursprünglichen aristotelischen Konzept entspricht. Nach Scheele (1999) herrsche zwar Uneinigkeit darüber, was exakt nach Aristoteles unter Katharsis verstanden werden solle. Das liege daran, dass die von Aristoteles verwendeten Begriffe „Phobos“ (dt. „Furcht“, „Schauer“) und „Eleos“ (dt. „Mitleid“, „Jammer“) mehrdeutig seien. Es herrsche aber Konsens, dass das Erleben von Phobos und Eleos bei den Zuschauern zu einer psychischen Reinigung führen solle. Bei Aristoteles handelt es sich also um ein rezeptionspsychologisches Konzept, d. h., die *Rezeption* eines Dramas hat eine bestimmte Wirkung. In der behavioristischen Forschung handelt es sich hingegen um ein aktionspsychologisches Konzept, d. h., die Wirkung entsteht durch das *Ausüben* aggressiven Verhaltens. Zwischen den verschiedenen Katharsis-Konzepten gibt es zahlreiche Unterschiede. Aus den bisher vorliegenden zahllosen Auslegungen des Katharsisbegriffs rekonstruierte die philosophisch-philologische Debatte laut Scheele (1999) zwei Extrempole, die sie als „Purgation“ und „Purifikation“ bezeichnet.

Bei der *Purgation* handelt es sich um einen homöostatischen Prozess, bei dem Spannungen abgeführt werden, die durch ein Zuviel an belastenden Emotionen entstanden sind. In der rezeptionspsychologischen Variante der Purgation erleben die Zuschauer stellvertretend Phobos und Eleos, was zu einer relativ automatischen Entladung der belastenden Emotionen führt. Der kathartische Effekt besteht in der Wiederherstellung des Wohlbefindens. Die auf diesem Konzept basierende kathartische Methode von Breuer und Freud (1895, zit. n. Scheele, 1999, S. 24ff.) sieht Kognitionen (d. h. affektkontrollierendes Bewusstsein) als Ursache für eine Behinderung der Entladung. Um die Entladung zu erreichen und dadurch Emotion und Verhalten zu reintegrieren, müssen kognitive Kontrollfunktionen somit vorübergehend außer Kraft gesetzt werden. Die behavioristische Forschung erweiterte dieses Katharsiskonzept um die Produktionsinstanz, d. h., es ging nicht mehr um die Rezeption von aggressivem Verhalten als Zuschauer, sondern um dessen Ausübung als Akteur. Somit handelt es sich hierbei um die aktionspsychologische Variante der Purgation.

Virtuelle Nothilfe

Bei der *Purifikation* hingegen handelt es sich um einen Prozess moralischer Entwicklung, in dessen Rahmen eine moralische Läuterung durch stellvertretendes (Mit-)Leiden erfolgt. Der kathartische Effekt besteht in einer moralischen Entwicklung hin zu einem sittlicheren Bewusstsein, wobei die moralische Entwicklung wiederum in einer zunehmenden Übereinstimmung von Kognition und Emotion besteht, d. h. „[d]as Individuum fühlt sich befreit von Kognitionen, Emotionen, Motivationen, die es nicht wollte und freier im Hinblick darauf zu ‚denken‘, zu ‚fühlen‘ und zu ‚handeln‘ wie es ‚eigentlich‘, eben im Sinne seiner von ihm subjektiv höher bewerteten moralischen Standards, denken, fühlen und handeln will“ (Scheele, 1999, S. 26f.). Als notwendige Voraussetzung für einen solchen Prozess sieht Scheele (1999) zum einen das Ziel, das „Gute zu wollen“ (S. 42), und zum anderen eine optimale Distanz zum moralischen Problem (S. 47). Ist die Distanz zu groß, wird der moralischen Emotion kaum bis keine Aufmerksamkeit geschenkt, ist sie zu klein, führt die starke Emotionalisierung zu leerlaufendem Denken, was eine kognitive Problembearbeitung unmöglich macht.

Tabelle 1: Purgations- vs. Purifikations-Merkmale (nach Scheele, 1999; Scheele & DuBois, 2006)

Purgation	Purifikation
Kathartischer Prozess	
Integration von Emotion und Verhalten durch Abreagieren/Ausagieren	Integration von Emotion und Kognition durch Umstrukturieren/Transformieren
Erlebensorientierung (absorbierende A-Reflexivität)	Verarbeitungsorientierung (involvierte Reflexivität)
unmittelbare Spannungsauflösung	aufgeschobene Spannungsauflösung
somatisch manifeste Intensität	verbal manifeste Extensität
Kathartischer Effekt	
unmittelbare Entspannung	mittelbare Entspannung
mittelbare Heilung, d. h. Wiederherstellung von emotionalem Wohlbefinden	unmittelbare Heilung, d. h. Stärkung von psychosozialer Motivation
(homöostatisches Modell)	(Entwicklungsmodell)

Einen Überblick über die Unterschiede zwischen Purgation und Purifikation (nach Scheele, 1999; Scheele & DuBois, 2006) gibt Tabelle 1. Nach Scheele (1999, S. 29) handelt es sich nicht um sich gegenseitig ausschließende Konzeptionen, sondern um Modellierungen von zwei unterschiedlichen psychischen Phänomenen, die unabhängig voneinander und sogar gleichzeitig auftreten können. Somit können prinzipiell beide Konzepte gleichzeitig in einer Untersuchung Anwendung finden. Im Bereich der Computerspielforschung wäre es sogar möglich, sowohl die rezeptions- als auch die aktionspsychologische Variante der Purgation zu studieren, da der Spieler zum einen die Rolle eines Zuschauers und zum anderen die eines

Virtuelle Nothilfe

Akteurs einnimmt. Ob eine Untersuchung der Konzepte aber Sinn macht, hängt teilweise davon ab, inwieweit sich die Konzepte bisher bewährt haben. Nützlich sind dazu die von Scheele (1999) berichteten empirischen Befunde zur rezeptions- und aktionspsychologischen Purgation sowie zur Purifikation.

Die Befunde⁶ zur *rezeptionspsychologischen Purgation* im Rahmen von Banduras Lernen am Modell fasst sie wie folgt zusammen: (1) aggressives Verhalten kann durch Beobachtung erlernt und ohne Überlernen lange behalten werden (Lern-Effekt), (2) stellvertretend erlebte Gewalt erhöht das Aggressionspotenzial, wenn (a) die Gewalt aus der Täter-Perspektive erlebt und als positiv dargestellt wird oder wenn (b) die Gewalt aus der Opfer-Perspektive erlebt und als ungerechtfertigt dargestellt wird (Stimulations-Effekt), und (3) stellvertretende Bestrafung führt zu Furcht vor Bestrafung und senkt so das Aggressionspotenzial (Hemmungs-Effekt). Scheele schließt allerdings nicht aus, dass eine vergeltungsmotivierte Aggressionsbereitschaft in Grenzen durch *Purifikations*-Prozesse überwunden werden kann. Solange dies nicht geschieht, muss allerdings befürchtet werden, dass die Rezeption von Aggression zum Erlernen aggressiven Verhaltens führt. Der Hemmungs-Effekt wirkt dem nur begrenzt entgegen, da er laut Kornadt (1982, zit. n. Scheele & DuBois, 2006) nur dazu führt, dass das aggressive Verhalten nicht ausgeübt wird, aber nicht dazu, dass die Aggressionsmotivation abgebaut wird.

Ebenso berichtet Scheele (1999, S. 37) eine Zusammenfassung der Befunde zur *aktionspsychologischen Purgation*: (1) die Aggressions-Motivation reduziert sich nur dann, wenn das Aggressions-Ziel subjektiv erreicht wurde, (2) nicht aber, wenn ein unspezifisches rein motorisches Ausagieren stattfindet. Nach Scheele besteht hier kein Widerspruch zum oben erwähnten Lern-Effekt, da nach Heckhausen (1989) die Aggressions-Handlungen durch die Erreichung des Aggressions-Ziels kurzfristig zu einer Desaktivierung der Aggressions-Volition führen, aber langfristig durch Lernen am Erfolg in einer Stärkung des Aggressions-Motivs münden. Diese Befunde stehen im Einklang mit dem Social Interactionist Model. Der Befund, dass unspezifisches Ausagieren wirkungslos ist, widerspricht aber der Alltagserfahrung, dass ein unspezifisches Ausagieren befreiend wirken kann. Scheele (1999, S. 38) nimmt daher an, dass für das Ausagieren in der Regel als positiv erlebte Tätigkeiten genutzt werden (z. B. Sport), die zu einer Distanzierung von Ärger-Emotion, -Motivation und -Handlung führen. Diese Annahme lässt sich ebenfalls mit dem Social Interactionist Model vereinbaren.

Zum *Purifikations*-Konzept existiert bisher nur eine einzige empirische Prüfung. Bönke (1989, zit. n. Scheele & DuBois, 2006) versuchte nachzuweisen, dass Aggression durch moralische Gefühle wie Mitleid, Schuld, Scham etc. gehemmt werden kann. Er konnte bei einem experimentellen Vergleich von drei Theaterstücken die theoretisch zu erwartende Art der Purifikations-Wirkung korrekt vorhersagen. Scheele (1999) ist außerdem der Ansicht, dass die Forschung zur Aggressions-Hemmung als Anhaltspunkt herhalten kann, solange man unterstellt, dass eine emotional-motivationale Zustandsähnlichkeit zwischen Hemmung und Läuterung

besteht. Als empirischen Beleg führt sie hier eine Studie von Rule und Ferguson (1986, zit. n. Scheele, 1999, S. 33) zur Wirkung von Gerechtigkeitsaspekten an. In dieser zeigte sich, dass die Probanden nach der Rezeption von gerechtfertigter im Vergleich zu ungerechtfertigter Gewalt feindseliger (d. h. enthemmter) reagierten. Dieser Befund steht auch im Einklang mit dem Moral Management Model. Als weiteren Beleg führt Scheele an, dass die Probanden aggressiver (d. h. enthemmter) in der Bedingung mit vergeltungsmotivierter Aggression im Vergleich zur Bedingung mit prosozial motivierter Aggression waren. Außerdem stammt von Anderson et al. (2008, zit. n. Hartmann et al., 2010) ein Beleg für die Geltung des *Purifikations*-Konzeptes, der bei Scheele noch nicht zu finden ist. Die Autoren führen Studien zur Fernsehgewalt an, die belegen würden, dass das Fokussieren auf den Schmerz und das Leiden der Opfer die negativen Auswirkungen des Gewaltkonsums reduzieren könne. Ein weiterer solcher Beleg ist die bereits in Abschnitt 4.3.4 zitierte Studie von Happ, Melzer und Steffgen (2010a), bei der durch die experimentelle Induktion von Empathie das Mitleid mit dem Gegner anstieg, insbesondere wenn man die Rolle eines Bösewichts übernahm.

4.3.6 BEWERTUNG DER THEORIEN

Die Entscheidung zwischen einem verhaltenstheoretischen Subjektmodell (wie es beim GAM vorliegt) oder einem handlungstheoretischen Subjektmodell (wie es beim Social Interactionist Model oder beim Moral Management Model vorliegt) hat direkte Folgen für die Auswahl der Aggressionstheorie und damit auch für die Hypothesenbildung und insbesondere die Operationalisierung. In der Handlungstheorie sind Intentionen definitorischer Bestandteil. Im Rahmen der Operationalisierung muss somit eine Schädigungsintention nachweisbar sein, da es sich sonst bei dem beobachteten Verhalten um etwas anderes als Aggression handeln kann. Daher erscheint es sinnvoll, vor der Auswahl der Operationalisierung eine Entscheidung für eines der beiden Subjektmodelle zu treffen. Hier fiel die Wahl auf die handlungstheoretische Sichtweise, da nur diese alle konzeptuellen Probleme bei der Definition des Aggressionsbegriffs zu lösen scheint. Damit scheiden das GAM und das Catalyst Model als Grundlage aus, weil beide durch das dahinterliegende verhaltenstheoretische Subjektmodell Probleme mit einer stringenten Definition des Konstrukts Aggression haben. Ein zweiter, weniger wissenschaftstheoretischer Grund, der gegen das GAM spricht, ist die ungenaue Modellierung von Wirkungsprozessen. Außerdem wird nicht spezifiziert, warum gerade interaktive Gewaltmedien eine starke aggressionsförderliche Wirkung besitzen sollen (Klimmt & Trepte, 2003; Klimmt, 2006a, S. 29). Drittens fehlt es an experimentellen Langzeitstudien, die nach Happ, Melzer und Steffgen (2010b, S. 25) entscheidend für die Prüfung des GAM sein könnten. Leider gibt es bisher nur korrelative Langzeitstudien (siehe dazu Abschnitt 4.3.7). Das Social Interactionist Model ist daher als allgemeines handlungstheoretisches Rahmenmodell zu bevorzugen. Da das Moral Management Model und das Scheelesche Catharsis Model ebenfalls auf einem handlungstheoretischen Subjektmodell basieren, sind sie in dieser Hinsicht kompatibel zum Social Interactionist Model. Es erscheint daher ratsam,

das generellere Social Interactionist Model um die spezielleren Eigenschaften der anderen beiden Theorien zu ergänzen. Aus dem Moral Management Model lassen sich Aussagen zu Gewaltwirkungsprozesse speziell in Computerspielen ableiten, während das Purifikationskonzept im Catharsis Model Ansatzpunkte zur Überwindung der Negativwirkung liefert.

4.3.7 META-ANALYSEN

Inzwischen existieren eine Reihe von Meta-Analysen zu der Frage, ob der Konsum von Gewaltmedien zu aggressivem bzw. kriminellern Verhalten führt (siehe Tabelle 2). Die Ergebnisse dieser Analysen sind nützlich für die Hypothesenbildung und sollen daher aufgeführt werden. Hinsichtlich *kriminellen* Verhaltens kommen Savage und Yancey (2008) in ihrer Meta-Analyse zu dem Schluss, dass eine Wirkung von Gewaltmedien (bei denen Computerspiele nicht enthalten sind) auf kriminelles aggressives Verhalten nicht nachgewiesen werden konnte. Ebenso kommt Ferguson (2008) nach einer Analyse verschiedener Behördenberichte zu dem Ergebnis, dass der Konsum von Gewaltspielen nicht im Zusammenhang mit Amokläufen steht. Dies könnte daran liegen, dass kriminelles aggressives Verhalten und insbesondere Amokläufe sehr extreme Formen von Gewalt darstellen. Sie kommen eher selten vor, weil Menschen gehemmt sind, Gewalt gegen andere Menschen auszuüben. Daher ist es möglich, dass es einen Zusammenhang gibt, der aber sehr klein ist. Dafür spricht, dass Meta-Analysen zu Gewaltspielen und *aggressivem* Verhalten kleine Effekte finden. Happ, Melzer und Steffgen (2010b) geben einen Literaturüberblick über bisher durchgeführte Meta-Analysen dieser Art. Die ersten stammen von Anderson und Bushman (2001) sowie von Sherry (2001). Anderson führte zu seiner Analyse zwei Aktualisierungen durch, und zwar Anderson (2004) sowie Anderson et al. (2010). Außerdem gibt es eine weitere Meta-Analyse von Bushman und Huesmann (2006). In den Meta-Analysen von Anderson kommen die Autoren zu der Ansicht, dass Gewaltspiele aggressives Verhalten (bzw. aggressive Gefühle, aggressive Gedanken und physiologische Erregung) erhöhen und gleichzeitig prosoziales Verhalten verringern. Dabei gibt es in den Effektstärken keine signifikanten Unterschiede zwischen experimentellen und korrelativen Studien, zwischen publizierten und unpublizierten Studien sowie zwischen jüngeren und älteren Probanden (Anderson & Bushman, 2001). In der ersten Meta-Analyse aus dem Jahr 2001 zeigen sich aber auch Begrenzungen der Allgemeinwirkung:

1. Der Effekt ist größer, wenn Aggression als „Gewalt gegen Dinge“ operationalisiert wird ($r+ = .41$); bei der Operationalisierung als „Gewalt gegen Menschen“ ist er kleiner ($r+ = .14$).
2. Zum Zeitpunkt der Analyse existierten keine Langzeitstudien, sodass nichts über die langfristigen Auswirkungen von Gewaltspielen bekannt war. Dennoch gingen die Autoren unter Verweis auf lerntheoretische Gesetzmäßigkeiten davon aus, dass regelmäßiger Konsum von Gewaltspielen die Neigung zu aggressivem Verhalten erhöht.

Virtuelle Nothilfe

Tabelle 2: Überblick über vorhandene Meta-Analysen zu Gewaltspielen

Autor	UV	Moderator	AV	N	K	r+	Fail-Safe-N
Sherry, 2001	Spiele		Aggression	1716	25	.15	-
Anderson & Bushman, 2001	Spiele	Probandenalter ¹ Studiendesign ² Publikationsstatus ³	aggressives Verhalten	3033	33	.19	-
			aggressive Kognition	1495	20	.27	-
			aggressiver Affekt	1151	17	.18	-
			physiologische Erregung	395	7	.22	-
			prosoziales Verhalten	676	8	-.16	-
Anderson, 2004	Spiele	Studiendesign ² Studienqualität ⁴	aggressives Verhalten	5240	32	.20	-
			aggressive Kognition	2567	19	.23	-
			aggressiver Affekt	2016	19	.17	-
			physiologische Erregung	508	9	.17	-
			Hilfeverhalten	683	7	-.20	-
Bushman & Huesmann, 2006	Medien (mit Spielen)	Studiendesign ² Probandenalter ¹	aggressives Verhalten	48430	262	.19	151984
			aggressive Kognition	22967	140	.18	34447
			Ärgergefühl	4838	50	.27	8654
			physiologische Erregung	1356	27	.26	927
			Hilfeverhalten	3243	59	-.08	157
Ferguson, 2007a	Spiele	Studienqualität ⁴ Probandenalter Publikationsjahr	aggressives Verhalten	483	5	.29	36
			aggressive Kognition	992	12	.25	167
			physiologische Erregung	363	4	.27	8
			prosoziales Verhalten	374	3	.30	9
Ferguson, 2007b	Spiele	keine	aggressives Verhalten	3602	17	.14	277
			visuospatiale Kognition	384	7	.49	330
Savage & Yancey, 2008	Medien (ohne Spiele)	keine	kriminelles aggressives Verhalten	[19; 10064]	[4;16]	[-.16;.28]	-
Anderson et al., 2010	Spiele	Kultur ⁵ Studiendesign ⁶ Probandenalter Probandengeschlecht Messverfahren ⁷ Gewaltspielkonsum ⁸ Spielerperspektive ⁹ Spielerrolle ⁹ Spielziel ⁹ Spielzeit ⁹	aggressives Verhalten	68313	140	.19	-
			aggressive Kognition	24534	95	.16	-
			aggressiver Affekt	17370	62	.14	-
			physiologische Erregung	9645	23	.18	-
			prosoziales Verhalten	9645	23	-.10	-
			Empathie/Desensibilisierung	8538	32	-.18	-

¹ Kind vs. Erwachsener. ² Experimentell vs. Nicht-Experimentell. ³ Publiziert vs. Nicht-Publiziert. ⁴ Best Practice vs. Nicht-Best-Practice.

⁵ Ost (Japan) vs. West (U. S. A.). ⁶ Experimentell vs. Querschnitt vs. Langzeit. ⁷ in Experimenten: CRTT vs. Andere; in Querschnittstudien: eine AV (physische Aggression) vs. mehrere AVen & Gewalt vs. Aggression. ⁸ in Querschnitt- und Langzeitstudien: Konsum gewalthaltiger Spiele vs. Konsum gewalthaltiger Spielszenen. ⁹ nur bei Experimenten.

Der zweite Kritikpunkt muss inzwischen relativiert werden, da es nun korrelative Langzeitstudien gibt, die eine Langzeitwirkung belegen (z. B. Kristen, 2005; Hopf, Huber & Weiß, 2008; amerikanische Studien finden sich bei Anderson et al., 2007, S. 5 und S. 95-119). In der deutschen Studie von Hopf et al. (2008) beispielsweise ist der Gewaltspielkonsum der größte

Risikofaktor für gewalttätiges *kriminelles* Verhalten, wobei dieser Befund in einem gewissen Widerspruch zu der oben aufgeführten Meta-Analyse von Savage und Yancey (2008) steht. Die amerikanischen Studien belegen einen Effekt der reinen Videospielzeit auf *aggressives* Verhalten (Anderson et al., 2007, S. 5). Laut den Autoren mangelt es aber weiterhin an experimentellen Langzeitstudien. Ebenso gibt es keine Langzeitstudien, bei denen physiologische Erregung oder Empathieverlust/Desensibilisierung die abhängigen Variablen sind (Anderson et al., 2010).

Weitere Kritik an der Meta-Analyse von Anderson und Bushman (2001) sowie an einer Meta-Analyse von Sherry (2001) wird von Ferguson (2007a, S. 473) geübt. Der Autor kritisiert an den beiden Meta-Analysen, dass nicht überprüft wurde, ob ein Publikations-Bias vorliegt. Seine eigene Meta-Analyse experimenteller Studien (2007a, S. 476f.) kommt zu dem Schluss, dass zumindest für aggressives Verhalten ein solcher Bias⁷ anzunehmen ist: Nach Orwin's Fail-Safe N würden nur sechs weitere nicht-publizierte Studien mit Nulleffekt reichen, um den Zusammenhang zwischen aggressivem Verhalten und Gewaltspielen nicht-signifikant werden zu lassen. Die Analyse nicht-experimenteller Studien kommt nach Ferguson zu einem vergleichbaren Ergebnis⁸; hier würden bereits drei Studien ausreichen. Aus zusätzlich durchgeführten Moderator-Analysen zieht Ferguson (2007a, S. 479f.) den Schluss, dass der unstandardisierte Einsatz von Messverfahren zu größeren (überschätzten) Effektstärken führt, wahrscheinlich weil er Möglichkeiten zur Ergebnismanipulation eröffnet. Ferguson findet dies problematisch, weil 62% der analysierten Studien keine standardisierten und reliablen Maße verwendet haben. An Fergusons Meta-Analyse selbst lässt sich wiederum kritisieren, dass er nur publizierte Studien aufgenommen hat, dass er nur neue Studien aufgenommen hat, und dass bei den neuen Studien viele fehlen (Anderson et al., 2010). In der aktuellsten Meta-Analyse von Anderson et al. (2010) finden sich laut den Autoren keine Hinweise auf einen solchen Bias. Über die Trim&Fill-Prozedur schätzen die Autoren ab, wie groß der Einfluss eines möglichen Bias auf die gefundenen Effektstärken wäre. Bei experimentellen Studien und bei Querschnittsstudien verändern sich die Effektstärken nur um 0.017, bei Langzeitstudien um 0.008. Sensitivitätsanalysen zeigen zudem, dass sich die Effektstärken durch methodisch schwache Studien nur marginal verändern. Daher sollen im Folgenden die Ergebnisse dieser aktuellen Meta-Analyse für die vollständige Stichprobe (statt der Ergebnisse für die „Best Practice“ Studien) aufgeführt werden.

Die Autoren berichten für die abhängige Variable „Gewaltverhalten“ in nicht-experimentellen Längsschnitt-Studien eine Effektstärke in Höhe von $r^+ = .20$ (aus $k = 14$ Studien mit $n = 5513$ Probanden), für nicht-experimentelle Querschnitt-Studien $r^+ = .19$ ($k = 81$; $n = 59336$) und für experimentelle Querschnitt-Studien $r^+ = .18$ ($k = 45$; $n = 3464$). Die Autoren berichten auch vergleichbar große Effektstärken für andere abhängige Variablen⁹: aggressive Kognition $r^+ = .21$ ($k = 48$; $n = 4289,5$); aggressiver Affekt $r^+ = .18$ ($k = 37$; $n = 3015$); physiologische Erregung $r^+ = .18$ ($k = 15$; $n = 969$). In den experimentellen Studien finden sich keine moderierenden Einflüsse durch die Spielerperspektive (erste vs. dritte Person), die

Spielerrolle (Held oder Verbrecher), die Spielzeit oder den Spielgegner (Mensch oder Nicht-Mensch). Die Autoren schließen daraus, dass die kurzfristige Wirkung von Gewaltspielen in Experimenten durch Primingprozesse zu Stande kommt.

Ferguson (2007a) übt aber auch grundsätzliche Kritik an den *Primärstudien*, auf denen die Meta-Analysen basieren. Diese Kritik hat zum Teil Folgen für die Gültigkeit der Meta-Analysen. Sie stützt sich auf vier Punkte (2007a, S. 480f.): Erstens würden in die Stichproben statt Gewalttättern immer (gesunde) Kinder und Studenten eingehen, obwohl bei nicht-klinischen Populationen das Gewaltpotenzial eher gering ist (Bodeneffekt). Dieser Kritikpunkt erscheint nur dann relevant, wenn die Primärstudien einen Zusammenhang zwischen Gewaltspielen und Amokläufen herstellen wollen. Sobald es um die Auswirkung des Gewaltspielkonsums in nicht-klinischen Populationen geht, ist es durchaus sinnvoll, Untersuchungen mit Kindern und Studenten durchzuführen. Zweitens wird kritisiert, dass unterschiedliche Aggressionsmaße verwendet würden, ohne deren Übereinstimmung zu berichten. In der Tat ist es sinnvoll, solche Übereinstimmungen zu berichten, um die Validität der Aggressionsmaße bewerten zu können. Dies würde sich von den Autoren der betreffenden Primärstudien aber leicht bewerkstelligen lassen. Im Rahmen von Meta-Analysen spielt dieser Kritikpunkt eine geringere Rolle, da sich die Übereinstimmungen auch zwischen den Effektstärken der Primärstudien berechnen lassen (siehe z. B. Carlson, Marcus-Newhall & Miller, 1989). Drittens wird moniert, dass mögliche Drittvariablen ignoriert würden. Das biologische Geschlecht würde z. B. einen großen Teil der Varianz erklären, der ansonsten dem Einfluss von Gewaltspielen zugeschrieben würde. Trait-Aggression würde zwar meist erhoben, allerdings wären die verwendeten Persönlichkeitstests leicht zu durchschauen, und gerade Gewalttäter würden zum Lügen neigen (Hare, 1993, zit. n. Ferguson, 2007a, S. 481). Dieser Kritikpunkt greift nur, wenn die Stichproben zu klein sind, um durch eine randomisierte Zuteilung zu einer Gleichverteilung der Drittvariablen zwischen den Gruppen zu kommen. In solchen Fällen bietet sich eine Erhebung der Drittvariablen an. Um das Durchschauen der Tests zu erschweren, erscheint es sinnvoll, diese losgelöst von der eigentlichen Untersuchung durchzuführen (z. B. in einer Sammelerhebung aller möglichen Persönlichkeitseigenschaften zu Beginn des Studiums). Werden Drittvariablen erhoben, dann ist es außerdem sinnvoll, sie in einer Meta-Analyse als potenzielle Moderatoren aufzunehmen, so wie es bei Anderson et al. (2010) mit dem biologischen Geschlecht der Fall ist. Viertens wird von Ferguson (2007a) kritisiert, es würde den Messinstrumenten (speziell dem CRTT) an Validität und Reliabilität mangeln. Dieser Kritikpunkt wiegt am schwersten, da er die Validität aller Primärstudien und damit auch aller Meta-Analysen bedroht. Daher soll im nächsten Abschnitt die von verschiedenen Autoren geübte Kritik an den Messinstrumenten ausführlich diskutiert und bewertet werden.

Wenn unterstellt wird, dass die Messinstrumente zu validen Ergebnissen geführt haben, dann legen die Meta-Analysen von Anderson und Bushman (2001), Anderson (2004) sowie Anderson et al. (2004; 2010) überzeugend dar, dass der Konsum von Gewaltspielen bei

nicht-klinischen Populationen kurzfristig zu höheren Werten in verschiedenen Aggressionsmaßen führt. Die Ergebnisse sind außerdem sehr nützlich, um die praktische Bedeutsamkeit eines Effekts zu bewerten, da es dafür keine festen Regeln gibt (siehe dazu Hussy & Möller, 1994). Insbesondere die Meta-Analyse von Anderson et al. (2010) ist dafür gut geeignet, weil dort nach Studientyp differenziert wird (Experiment, Querschnitt, Längsschnitt) und nur neuere Studien mit höherer methodischer Qualität eingeflossen sind, was die Vergleichbarkeit mit dieser Studie erhöht. Die Autoren berichten für experimentelle Querschnitt-Studien zur Wirkung von Gewaltspielen auf Gewaltverhalten eine Effektstärke von $r^+ = .181$. Wird die Korrelation in Hedges g umgerechnet, so ergibt sich $g = 0.364$. Dies mag auf den ersten Blick ein zu vernachlässigender Effekt sein. Anderson et al. (2007, S. 48-52) sowie Anderson et al. (2010) argumentieren aber, dass es sich bei Gewaltspielen um einen von mehreren Risikofaktoren handelt, die sich in ihrer Wirkung aufaddieren können (eine solche Kumulation wurde bisher jedoch nur in korrelativen Langzeitstudien nachgewiesen, nicht aber in experimentellen). Bei multikausalen Phänomenen sei ohnehin nicht zu erwarten, dass ein einzelner Risikofaktor einen größeren Teil der Varianz erklärt. Im Vergleich mit anderen Risikofaktoren zählt der Effekt sogar zu den größeren (Anderson et al., 2007, S. 143).

4.4 OPERATIONALISIERUNGEN DES GEWALTVERHALTENS

Im Folgenden werden verschiedene experimentelle Verfahren zur Messung von Gewaltverhalten beschrieben und die ihnen innewohnenden (Validitäts-)Probleme erläutert, um eine begründete Auswahl für diese Untersuchung treffen zu können. Aus der Theoriediskussion ergibt sich dabei als wichtige Anforderung, dass das Verfahren die Schädigungsabsicht sicherstellen muss.

4.4.1 VERFAHREN

Nach Tedeschi und Quigley (1996) wurden bisher hauptsächlich vier Verfahren zur Messung von aggressivem Verhalten in Experimenten eingesetzt: das „Teacher/Learner Paradigm“ (dt. „Milgram-Paradigma“, auch als „Buss Aggression Machine“ bekannt), das „Essay Evaluation Paradigm“ (dt. „Aufsatzbeurteilung“), das „Competitive Reaction Time Test Paradigm“ (abgekürzt „CRTT“; auch benannt als „Competitive Reaction Time Game“ oder „Competitive Reaction Task“; dt. „Reaktionszeit-Wettbewerb“) als direkter Nachfolger des „Taylor Aggression Paradigm“ (abgekürzt „TAP“) und das „Bobo Modeling Paradigm“ (dt. „Bobo-Puppenmodell“). Ritter und Eslea (2005) ergänzen diese Liste klassischer Verfahren um das „Point Subtraction Aggression Paradigm“ (abgekürzt „PSAP“; dt. „Punkt-Subtraktion“), weil es ihrer Ansicht nach ebenfalls oft verwendet wurde. Daneben benennen sie mit dem „Hot Sauce Paradigm“ (dt. „Scharfe Soße“), dem „Bungled Procedure Paradigm“ (dt. „Vermasselter Ablauf“) sowie dem „Experimental Graffiti Paradigm“ (dt. „Experimentelles Graffiti“) drei neue Erfolg versprechende Verfahren. Von allen genannten Verfahren zählen Anderson et al. (2010) das TAP und seine Variante CRTT sowie das Hot Sauce Paradigm zu den qualitativ

hochwertigen Operationalisierungen, wobei der CRTT die am häufigsten verwendete sei. Alle Verfahren sollen im Folgenden kurz vorgestellt und danach mit ihren Vor- und Nachteilen erläutert werden.

4.4.1.1 TEACHER/LEARNER UND ESSAY EVALUATION

Im „Teacher/Learner Paradigm“ haben die Probanden die Rolle eines „Lehrers“ inne, der einem „Schüler“ (meistens ein Konföderierter) bei einer Gedächtnisaufgabe beim Lernen helfen soll (Tedeschi & Quigley, 1996; Ritter & Eslea, 2005). Der „Lehrer“ hält sich in der Regel in einem anderen Raum auf und bekommt dort Informationen über die Korrektheit der Antworten des „Schülers“. Falsche Antworten soll der „Lehrer“ durch Erteilen eines Stromstoßes bestrafen. Der Stromstoß kann in seiner Intensität und Dauer variiert werden. Die Methode des „Essay Evaluation Paradigm“ unterscheidet sich vom Teacher/Learner Paradigm darin, dass der Proband die Stromstöße verabreicht, um einen Aufsatz des Konföderierten zu bewerten. Die beste Bewertung besteht in einem Elektroschock, die schlechteste in zehn Schocks.

4.4.1.2 TAP UND CRTT

Beim „Taylor Aggression Paradigm“ (TAP) von Taylor (1967) wird der Versuchsperson mitgeteilt, sie würde in einem Reaktionszeitwettbewerb gegen eine andere Person antreten, bei dem es darum geht, so schnell wie möglich eine Taste zu drücken. In Wahrheit „spielt“ die Versuchsperson gegen einen Computer, der Sieg und Niederlage schon vor Spielbeginn festgelegt hat. Der „Verlierer“ erleidet einen aversiven Reiz, dessen Stärke der „Gewinner“ zu Beginn einer Runde festlegt. In der ursprünglichen Variante bestanden die Reize aus Elektroschocks (Taylor, 1967). Das Verfahren wurde 1989 von Bushman modifiziert, der den Elektroschock gegen einen lauten Ton ersetzte und die Möglichkeiten hinzufügte, auch die Reizdauer einzustellen sowie den Wert Null für Lautstärke und Dauer einzustellen (Bushman & Saults, 2006). Diese Variante nennt Bushman „Competitive Reaction Time Task“ (CRTT). Als Maß für aggressives Verhalten gilt in der Regel die Dauer und/oder die Lautstärke, die die Versuchsperson für die Bestrafung des „Verlierers“ festlegt (vgl. Anderson & Bushman, 1997; Anderson & Dill, 2000; Anderson, Lindsay & Bushman, 1999; Anderson, Anderson, Dorr, DeNeve & Flanagan, 2000; Anderson et al., 2004, Studie 2; Bartholow & Anderson, 2002; Bushman & Anderson, 1998; Carnagey & Anderson, 2005, Studie 3; Giancola & Chermack, 1998; Lindsay & Anderson, 2000). Wie Dauer und Lautstärke zu einem Maß verrechnet werden, ist allerdings nicht standardisiert. Elson (2011) verglich in seiner Arbeit die folgenden bisher eingesetzten Auswertevarianten:

1. den Durchschnitt der Lautstärke über alle Trials bilden (z. B. Sestir & Bartholow, 2010),
2. den Durchschnitt der Dauer über alle Trials bilden,

Virtuelle Nothilfe

3. den Durchschnitt der Lautstärke über alle Trials bilden, den Durchschnitt der Dauer über alle Trials bilden, beides miteinander multiplizieren (z. B. Bartholow, Sestir & Davis, 2005),
4. die Quadratwurzel der Dauer über alle Trials bilden, mit der Lautstärke multiplizieren, über diese Werte den Durchschnitt bilden (z. B. Carnagey & Anderson, 2005),
5. den Durchschnitt der Lautstärke nach Siegrunden bilden (z. B. Anderson & Dill, 2000),
6. den Durchschnitt der Lautstärke nach Niederlagerunden bilden (z. B. Anderson & Dill, 2000),
7. den Durchschnitt der Dauer nach Siegrunden bilden (z. B. Anderson & Dill, 2000),
8. den Durchschnitt der Dauer nach Niederlagerunden bilden (z. B. Anderson & Dill, 2000),
9. die Häufigkeit über die extremsten Einstellungen (ab Wert 8 auf der 10er-Skala) bilden (z. B. Anderson & Carnagey, 2009),

Daneben führt Elson (2011) auch eine weitere Variante von Anderson et al. (2004) an, bei der in der zweiphasigen Version des CRTT (siehe oben unter „Konstruktvalidität“) zur Auswertung der Durchschnitt der Lautstärke in Runde 1, 2 bis 9, 10 bis 17 und 18 bis 25 gebildet wurde, was zu vier Gewaltvariablen führte. Keine Erwähnung findet bei Elson (2011) eine Variante von Bushman und Baumeister (1998; auch verwendet bei Konijn et al., 2007), bei der nur der Wert aus der ersten Runde verwendet wird, da die Einstellungen in den Folgerunden lediglich die reziproke Antwort des Probanden auf das darstellen würden, was der Gegenspieler ihrer Vermutung nach getan hat (am Ende dieses Abschnitts finden sich weitere Informationen zu dieser Variante). Bei Pawelczyk (2011) findet sich außerdem eine völlig neue Variante: Lautstärke und Dauer jeder Runde wurden miteinander multipliziert, um im Anschluss per Faktorenanalyse festzustellen, welche Runden sich ohne Nebenladungen zusammenfassen lassen (hier waren es Runden 1-8 und 18-25).

4.4.1.3 BOBO MODELLING

Beim „Bobo Modeling Paradigm“ von Bandura (1973) sieht der Proband (üblicherweise ein Kind), wie ein Erwachsener aggressiv mit einem aufblasbaren Spielzeug-Clown spielt. Es lässt sich experimentell variieren, ob das Verhalten des Erwachsenen belohnt oder bestraft wird. Gemessen wird das aggressive Verhalten des Probanden gegenüber dem Clown, nachdem ihm erlaubt wurde, mit dem Clown zu spielen.

4.4.1.4 POINT SUBTRACTION AGGRESSION

Beim „Point Subtraction Aggression Paradigm“ (kurz „PSAP“) von Cherek (1981) sollen Probanden am Computer eine Taste 100 Mal drücken, um dadurch einen Geldbetrag in Höhe von 10 Cent zu gewinnen. Alternativ können sie eine andere Taste 10 Mal drücken, um einem Gegner, der angeblich in einem anderen Raum sitzt, dieselbe Summe von dessen Kon-

Virtuelle Nothilfe

to abzuziehen. Durch 10-maliges Drücken einer dritten Taste können sie diesem Gegner einen Rauschton von einer Sekunde Länge verabreichen. Am Bildschirm können die Probanden sehen, wie viel Geld sie erspielt haben und ob ihr Gegner Geld von ihrem Konto abgezogen hat. In Wirklichkeit zieht der Computer automatisch Geld nach einem vorher festgelegten Schema ab. Als Aggressionsmaß gilt, wie oft der Proband die Subtraktions-Taste gedrückt hat. Das Drücken der Rauschton-Taste wird nicht als Aggression gewertet, da das Subtrahieren von Geld von den Probanden als extrem aversiv empfunden würde, während der Rauschton nur eine Belästigung darstelle (Cherek, 1981). Diese Einschätzung überrascht, wenn man bedenkt, dass im CRTT solche Rauschtöne als aversiver Reiz gelten. Der Grund für den Befund könnte darin liegen, dass der Rauschton im PSAP lediglich 60 dB laut ist, was im CRTT einer der niedrigsten Stufen entspricht.

4.4.1.5 HOT SAUCE

Die Methode „Hot Sauce Paradigm“ wurde von McGregor et al. (1998) bzw. Lieberman, Solomon, Greenberg und McGregor (1999) entwickelt. Probanden sollen die Speise eines Konföderierten mit scharfer Chilisoße würzen, wobei der Konföderierte scharfe Speisen angeblich nicht mag. Der Konföderierte hat den Probanden vorher provoziert, in der Regel durch Verabreichung eines mit Essig versetzten Saftes (Lieberman & Greenberg, 1998, zit. n. Lieberman et al., 1999; McGregor et al., 1998, Studie 4) oder durch Schreiben eines Aufsatzes, in dem die Weltanschauung des Probanden abgewertet wird (McGregor et al., 1998, Studie 1).

Aufgrund des Vorwurfs, dass es keine nicht-aggressive Verhaltensalternative bei dem Verfahren geben würde, wurden verschiedene Erweiterungen des Paradigmas vorgenommen. Böhm und Streicher (2010) erweiterten das Verfahren um eine aggressive Alternative für Frauen (Mayonnaise) und eine für beide Geschlechter neutrale Alternative (fettarmer Joghurt). Eine ähnliche Erweiterung findet sich bei Beier und Kutzner (2011a; 2011b). Die Autoren verwendeten Saft statt Soße und kreierten dabei eine „angenehme“ Variante (Orangensaft), eine „neutrale“ Variante (Früchtetee) und eine „scharfe“ Variante (Orangensaft mit Tabasco).

4.4.1.6 BUNGLED PROCEDURE

Im „Bungled Procedure Paradigm“ (übersetzt „vermasselter Ablauf“) von Russell et al. (1996; 2002; beide zit. n. Ritter & Eslea, 2005, S. 410f.) soll der Proband mit einer Pellet- oder Paintball-Waffe auf eine mit Schutzkleidung ausgestattete Frau schießen. Die Aufgabe wird als eine Form männlicher Unterhaltung präsentiert. Das aggressive Verhalten ist operationalisiert als Schusskraft der gewählten Waffe multipliziert mit der Anzahl der Kugeln. Bevor ein Proband tatsächlich auf die Frau schießen kann, wird die Prozedur abgebrochen. Es wird be-

hauptet, der Proband gehöre eigentlich in die Kontrollgruppe und sei versehentlich in der falschen Versuchsbedingung gelandet (daher der Name „vermasselter Ablauf“).

4.4.1.7 EXPERIMENTAL GRAFFITI

Bei der „Experimental Graffiti and Tearing Procedure“ von Norlander et al. (zit. n. Ritter & Eslea, 2005, S. 411) erhalten die Probanden eine Illustration von „Adam und Eva im Garten Eden“, auf die sie etwas zeichnen sollen. Beurteiler bestimmen im Anschluss die Menge an „Graffiti“, den Grad an Zerstörung sowie aggressive und sexuelle Inhalte, die dem Bild auf diese Weise hinzugefügt wurden. Alles dies gilt als Maß für aggressives Verhalten. Als Nächstes erhalten die Probanden eine Abbildung von „Samson und der Löwe“, deren aggressiver Charakter bei den Probanden aggressives Verhalten provozieren soll. Die Probanden werden instruiert, die Abbildung in kleine Stücke zu zerreißen und in einen Umschlag zu legen, der halb so groß ist wie die Abbildung. Die Anzahl der Stücke gilt als Maß für aggressives Verhalten.

4.4.2 VALIDITÄTSPROBLEME

Allen oben genannten Verfahren wird von einigen Autoren (z. B. Tedeschi & Quigley, 1996) vorgeworfen, dass es ihnen an Validität mangeln würde. Dieser Vorwurf basiert auf der Tatsache, dass keine Operationalisierung die Intentionen und/oder Motive der Probanden systematisch erfasst, obwohl die Schädigungs*absicht* definitorischer Bestandteil des Konstrukts ist (Tedeschi & Quigley, 1996; Ritter & Eslea, 2005). Die Intentionen bzw. Motive einer Person müssen also bekannt sein, um zu entscheiden, ob es sich um aggressives Verhalten oder um etwas anderes handelt. Durch die Unterschlagung der Intentionen bei der Messung von Aggression ist das Konstrukt nach Tedeschi und Quigley unterrepräsentiert (2000, S. 128ff.).

Dagegen wird von anderen Autoren (z. B. Giancola & Chermack, 1998) eingewendet, dass gängige Maße für Aggression (Dauer vs. Intensität vs. Häufigkeit; schriftliche vs. physische Ausdrucksform) trotzdem hoch miteinander korrelieren (Carlson et al., 1989). Einige Autoren sehen dies als Beweis für die Validität (z. B. Anderson et al., 2007, S. 23; Konijn et al., 2007). Nach Tedeschi und Quigley (2000) ist dieser Beweis nicht zwingend, da es immer noch möglich wäre, dass kein einziges Maß Aggression erfasst, sondern etwas anderes wie Konformität (engl. „conformity“, siehe Abschnitt 4.4.2.1) oder ausgleichende Gerechtigkeit (engl. „retributive justice“ oder kurz „retribution“, siehe Abschnitt 4.4.2.2). Als Beleg führen sie an, dass in der Studie von Carlson et al. (1989) die damals gängigen Maße verbaler und körperlicher Aggression zwar durch die Variablen „Frustration“ und „Ärger“, nicht aber durch die Variable „persönlicher Angriff“ vorhergesagt werden konnten. Gerade ein persönlicher Angriff sollte aber die stärkste Provokation darstellen und damit am ehesten zu einer aggressiven Reaktion führen.

Virtuelle Nothilfe

Einen weiteren Beleg für die Validität der Operationalisierungen sehen einige Autoren (z. B. Konijn et al., 2007; Giancola & Chermack, 1998) darin, dass „reale“ und laborbasierte Maße für Aggression von situationalen und personalen Variablen (z. B. Geschlecht, Provokation, Alkoholkonsum, Temperatur) in gleicher Weise beeinflusst werden (Anderson & Bushman, 1997). Bei Giancola und Chermack (1998) wird diese Form des Validitätsnachweises als „Gruppendiskrimination“ bezeichnet. Tedeschi und Quigley (2000) sehen hierin aber nur einen schwachen Validitätsnachweis, da es sich um einen logischen Fehlschluss handelt. Wenn aggressive Personen ein Verhalten A im Labor häufiger zeigen als nichtaggressive Personen, und wenn aggressive Personen im Feld ein Verhalten B häufiger zeigen als nichtaggressive Personen, dann gilt nicht zwingend A = B. Wenn z. B. Personen durch eine hohe Temperatur aggressiver werden und die hohe Temperatur sowohl zu einer negativeren Bewertung anderer Personen im Labor als auch zu vermehrter Flüssigkeitsaufnahme im Feld führt, dann lässt sich nicht logisch daraus schließen, dass eine erhöhte Flüssigkeitsaufnahme dasselbe Verhalten darstellt wie eine negativere Bewertung anderer Personen (Tedeschi & Quigley, 2000).

4.4.2.1 FEHLENDE SCHÄDIGUNGSABSICHT

Da die Intentionen der Probanden nicht erfasst werden, besteht Unsicherheit hinsichtlich der Frage, ob die Probanden eine Schädigung ihres Gegenübers beabsichtigen. Nach Gottfredson und Hirschi (1993, zit. n. Tedeschi & Quigley, 1996, S.165) würden Aggressionsexperimente lediglich zeigen, dass Probanden gefällige, fügsame und für soziale Normen empfängliche Menschen sind.

4.4.2.1.1 KONFORMITÄT UND IMPLIZITE AUFFORDERUNG

Tedeschi und Quigley (1996) schlussfolgern, dass es sich bei Aggressionsexperimenten um einfache Variationen des Milgram-Experiments zur Konformität handelt. Die Versuchspersonen seien gezwungen, aversive Reize zu verabreichen. Daher würden sie nach Hinweisen (engl. „cues“) in ihrer Umwelt suchen, die ihnen vermitteln, welches Verhalten von ihnen erwartet wird. Dabei müsse der Hinweis nicht einmal eine Information über die Richtung des Verhaltens beinhalten (sei „mehr“ oder sei „weniger“ aggressiv). Es reiche aus, dass die Versuchspersonen glauben, von ihnen würde das Verabreichen mehr oder weniger starker aversiver Reize erwartet. Die Handlungen würden dann mit der Intention durchgeführt, den Experimentator zufriedenzustellen. Die Versuchspersonen würden sich nicht als aggressiv erleben, da ihnen die Rolle des „Lehrers“, des „Evaluators“ oder des „Wettbewerbers“ vorher zugewiesen wurde (Tedeschi & Quigley, 2000, S. 131; Baron & Richardson, 1994). Dieses Problem wird dadurch begünstigt, dass zahlreiche Hinweisreize den Probanden nahelegen, dass das „aggressive“ Verhalten harmlos ist und gebilligt wird (Ritter & Eslea, 2005). Giancola und Chermack (1998) hingegen sind der Ansicht, ein Problem mit Konformität bzw. implizitem Aufforderungscharakter (engl. „demand characteristics“) läge nicht vor. Verschiedene

Virtuelle Nothilfe

Studien würden zeigen, dass immer nur sehr wenige Versuchspersonen die Messung von Aggression durchschauen. Nach Tedeschi und Quigley (2000) ist es aber für das Vorliegen einer impliziten Aufforderung überhaupt nicht notwendig, dass die Probanden das Ziel des Experiments erkennen. Es reicht aus, wenn die Probanden aus Hinweisreizen das von ihnen erwünschte Verhalten erschließen und sich dementsprechend verhalten, um die Versuchsleiter zufriedenzustellen.

4.4.2.1.2 KEINE VERHALTENSALTERNATIVEN

Ein solcher Hinweisreiz kann z. B. darin bestehen, dass es keine Möglichkeit gibt, sich nicht-aggressiv zu verhalten. Möglicherweise schlussfolgern die Probanden aus diesem Umstand, dass von ihnen das Verabreichen aversiver Reize erwartet wird. Wenn eine sinnvolle Verhaltensalternative eingeführt wird, dann kann sich die Wahl des Verhaltens verändern. Eine Studie von Hokanson (1970, zit. n. Tedeschi & Quigley, 1996, S. 171) zeigt beispielsweise, dass Frauen sich häufiger für Belohnungen entscheiden, wenn ihnen die Wahl gelassen wird, auf eine Provokation mit Belohnung oder Bestrafung zu reagieren. Dies könnte eine Form von Beschwichtigungsverhalten darstellen. Giancola und Chermack (1998) antworten auf den Vorwurf der fehlenden Verhaltensalternative, dass die Probanden sehr wohl nicht-aggressive Verhaltensmöglichkeiten geboten bekämen. Im CRTT lässt sich beispielsweise der Wert Null einstellen; im PSAP kann der Knopf für den Gewinn von Punkten gedrückt werden. Das Problem bei diesen Verhaltensoptionen ist aber, dass die Probanden nicht mehr mit ihrem „Gegenspieler“ interagieren. Die Alternative könnte von den Probanden daher als Teilnahmeverweigerung („dropping out of the experiment“) interpretiert werden (Ritter & Eslea, 2005, S. 414f.), wodurch wieder ein Problem durch Konformität bzw. implizite Aufforderung vorliegen würde.

4.4.2.1.3 NICHT-AGGRESSIVE MOTIVE DURCH COVER STORIES

Ein weiteres Problem entsteht durch die verwendeten Cover Stories. Wenn die Probanden fest davon überzeugt sind, dass sie durch die aversiven Reize beim Lernen helfen (z. B. Teacher/Learner Paradigm) oder ihre Siegchancen verbessern (z. B. CRTT), dann besteht ihrerseits keine Schädigungsabsicht (Ritter & Eslea, 2005). Da es sich beim Teacher/Learner Paradigm laut Cover Story um ein Experiment zum Lernen handelt, könnten die Probanden durchaus die Schlussfolgerung ziehen, dass die ihnen unbekannte Person besonders dumm sein muss, da sie ohne Grund angreift/provoziert (Tedeschi & Quigley, 1996). Sie erteilen dann möglichst viele bzw. starke Schocks, da eine dumme Person stärker von der lernförderlichen Behandlung profitieren sollte. In dem Fall würde es sich sogar um prosoziales Verhalten handeln. Tatsächlich fanden Baron und Eggleston (1972, zit. n. Tedeschi & Quigley, 1996, S. 167; vgl. auch Ritter & Eslea, 2005, S. 413) eine positive Korrelation zwischen selbst eingeschätztem Altruismus und Intensität der Schocks. In einer Studie von Rule und Nesdale (1974, zit. n. Tedeschi & Quigley, 1996, S. 167; s. auch Ritter & Eslea, 2005, S. 413) fand sich

ein weiterer Beleg für diese Interpretation. Dort gab es zwei experimentelle Manipulationen: Die Probanden wurden vom Konföderierten beleidigt (oder nicht) und ihnen wurden erklärt, dass Schocks das Lernen fördern (oder nicht). Wenn Schocks das Lernen fördern sollten, dann verringerte eine vorangegangene Beleidigung die Stärke des Schocks, d. h. die Beleidigung verringerte die prosoziale Motivation der Probanden. Wenn Schocks das Lernen nicht fördern sollten, dann war es für die Schockstärke egal, ob die Probanden vorher beleidigt wurden. Eine Untersuchung von Melburg und Tedeschi (1989, zit. n. Tedeschi & Quigley, 1996, S. 174) zeigte, dass auch Selbstpräsentation ein Motiv für die Verabreichung der Schocks sein kann. Es wurde experimentell manipuliert, wie die Aufgabenlösung eines Probanden beurteilt wurde. Schnitt der Proband schlechter ab als der Konföderierte, so verabreichte er mehr Schocks. Dazu passend korrelierte die Variable „Besorgnis um das eigene Ansehen“ (engl. „identity concerns“) mit der Anzahl der Schocks, nicht hingegen die Variable „Frustration“.

Giancola und Chermack (1998) stimmen dieser Kritik am Teacher/Learner Paradigm zu, sind aber der Meinung, dass dieses Problem TAP und PSAP nicht betreffen würde. Hier würde ein Hereinfallen auf die Cover Story eine aggressive Reaktion nicht ausschließen. Ritter und Eslea (2005) dagegen resümieren, dass das Verabreichen unangenehmer Reize (Schocks, scharfe Soße) zwar mehr oder weniger schädlich sei, dass daraus aber nicht automatisch folge, dass von außen gleich aussehende Verhaltensweisen von den gleichen Motiven ausgelöst würden (siehe dazu auch Abschnitt 11.4.1). Dazu passt auch das Ergebnis des oben beschriebenen Experiments von Melburg und Tedeschi (1989, zit. n. Tedeschi & Quigley, 1996, S. 174), bei dem die Schockeinstellungen im Teacher/Learner Paradigm nicht von der vorangegangenen Frustration abhingen, sondern von der Sorge um das eigene Image.

Aus den oben diskutierten Problemen folgt letztendlich, dass das bei den Probanden beobachtete Verhalten eine Schädigungsabsicht beinhalten kann, aber nicht muss. Im Rahmen rein beobachtender Forschung lässt sich prinzipiell keine Entscheidung darüber treffen, ob eine solche Intention vorliegt oder nicht, da Intentionen und Motive nicht direkt beobachtbar sind und die Validität von Probandenaussagen im Rahmen desselben Experiments nicht gesichert werden kann. Hierfür wäre ein zweistufiges Vorgehen wie das von Groeben (1986) vorgeschlagene nötig (eine Erläuterung des Verfahrens findet sich in Abschnitt 14.1.2). Eine solche zweistufige Validierung wurde aber bisher für kein einziges der genannten Messverfahren durchgeführt.

4.4.2.2 EINGESCHRÄNKTE VERALLGEMEINERBARKEIT

Neben dem Vorwurf, dass die Verfahren möglicherweise etwas anderes als aggressives Verhalten erheben, wird auch argumentiert, dass es sich selbst im Fall einer validen Messung um sehr spezielle Formen von Aggression handeln könnte, was die Verallgemeinerbarkeit einschränken würde.

4.4.2.2.1 RETRIBUTION

Beispielsweise wäre es möglich, dass die Verfahren nur Vergeltungsaggression erfassen, die erst auf eine Provokation hin erfolgt. Probanden verhalten sich hier nach dem Prinzip „Wie du mir, so ich dir“. Die meisten Verfahren beinhalten Provokationen, die aufgrund des Verstoßes gegen die Gegenseitigkeitsnorm als ungerecht wahrgenommen werden. Anders ausgedrückt handelt es sich um eine Negativvariante der Reziprozität (Tedeschi & Quigley, 1996), die eine wichtige Rolle beim aggressiven Verhalten spielt (Axelrod, 1984, zit. n. Konijn et al., 2007). Dies wäre ein Problem für die Konstruktvalidität, weil es sich bei der Retribution um eine sehr spezielle Form von aggressivem Verhalten handelt. Das Konstrukt wäre unterrepräsentiert, weil proaktive Formen von Aggression (d. h. ohne vorangegangene Provokation) nicht operationalisiert werden. Gerade bei typischen Verbrechen wie Vergewaltigung, Raub, Mord etc. dürfte es sich aber um proaktive Aggression handeln (Tedeschi & Quigley, 1996). Auf solche Verbrechen ließen sich die Ergebnisse der experimentellen Aggressionsstudien dann nicht verallgemeinern.

4.4.2.2.2 FEINDSELIGKEIT

Eine ähnliche potenzielle Einschränkung der Verallgemeinerbarkeit besteht darin, dass feindselige Probanden möglicherweise nur die ihnen gebotene Gelegenheit nutzen, um ihre feindselige Persönlichkeit auszuleben (Tedeschi & Quigley, 1996, S. 167). Dies wird dadurch ermöglicht, dass der Versuchsleiter aggressives Verhalten nicht sanktioniert, sondern toleriert. Tedeschi und Quigley beziehen sich dabei auf drei Untersuchungen (Follingstad, Kallichman, Cafferty & Vormbrock; 1992; Shemberg, Leventhal & Allman, 1968; Wolfe & Baron, 1971; alle drei zit. n. Tedeschi & Quigley, 1996), in denen Männer und Frauen mit hoher aggressiver Persönlichkeit höhere Schockintensitäten im Teacher/Learner Paradigm verwendet haben als solche mit geringer aggressiver Persönlichkeit. Alternativ könnte es sich nicht um feindselige Personen handeln, sondern um solche, die zu Verantwortungsverschiebung (im Sinne einer Persönlichkeitseigenschaft) neigen. Verantwortungsverschiebung (engl. „displacement of responsibility“) ist eine von mehreren Abwehrstrategie im Rahmen der Moral Disengagement Theory, mit denen Menschen ihr unmoralisches Verhalten rechtfertigen (vgl. Bandura, 2002; Klimmt et al., 2006; 2008; siehe Abschnitt 4.3.4). Giancola und Chermack (1998) sehen trotz der Kritik von Tedeschi und Quigley (1996) kein Problem darin, da aggressives Verhalten auch in der realen Welt oft legitimiert würde (insbesondere im Fall von Kriegen). Aber selbst wenn man dieses Argument akzeptiert, bleibt ein Problem für die Konstruktvalidität bestehen: möglicherweise erfassen die Messverfahren nicht allgemein Aggression, sondern speziell Feindseligkeit bzw. Verantwortungsverschiebung.

4.4.2.2.3 RÄUMLICHE DISTANZ

Eine weitere Einschränkung der Konstruktvalidität wird dadurch verursacht, dass in den meisten Paradigmen eine räumliche Distanz zwischen Proband und Konföderiertem besteht.

Virtuelle Nothilfe

Dadurch sind die Ergebnisse nicht auf die im Alltag wesentlich üblicheren Situationen verallgemeinerbar, in denen sich Aggressor und Opfer von Angesicht zu Angesicht gegenüberstehen und die Ausübung von physischer Gewalt echten Körperkontakt erfordert. In der Studie von Milgram (1963, zit. n. Ritter & Eslea, 2005, S. 415) zeigt sich beispielsweise, dass eine Erhöhung der Distanz zu einer Verringerung der Hemmschwelle für aggressives Verhalten führt. Dies könnte an dem höheren emotionalen Abstand (schließlich hört und sieht man sein Opfer nicht mehr), an dem geringeren Retaliationsrisiko und an der geringeren notwendigen Anstrengung liegen (Ritter & Eslea, 2005).

4.4.2.3 SOZIALE ERWÜNSCHTHEIT

Im Rahmen einer Meta-Analyse zur Wirkung einer sichtbaren Waffe auf aggressives Verhalten stellen Carlson, Marcus-Newhall und Miller (1990, S. 626) fest, dass Probanden sich weniger aggressiv verhalten, wenn sie erkannt haben, dass es um die Erfassung ihres aggressiven Verhaltens geht (engl. „hypothesis awareness“), bzw. wenn sie befürchtet haben, dass ihr Verhalten bewertet wird (engl. „evaluation apprehension“). Die Autoren führen das darauf zurück, dass aggressives Verhalten als negativ bewertet wird und die Probanden bemüht sind, einen guten Eindruck zu machen, indem sie kein unerwünschtes Verhalten zeigen (engl. „impression management“). Die soziale Unerwünschtheit aggressiven Verhaltens führt somit zu einer Dissimulation. Schlimmstenfalls sind die Versuchsbedingungen mit dem Erkennen der Täuschung konfundiert, was zu einer schweren Störung der internen Validität führen kann. Bestenfalls liegt keine Konfundierung vor, was „nur“ die Variabilität in der abhängigen Variable und somit die Präzision der Untersuchung reduziert. Es ist daher sehr wichtig, dass die Probanden die Täuschungsmanöver vor und während der Messung nicht durchschauen. Die Probanden dürfen außerdem nicht befürchten, dass die Messung zu einer Bewertung ihres Verhaltens führt.

4.4.3 BEWERTUNG DER VERFAHREN

Die oben aufgeführten Validitätsprobleme treffen in unterschiedlichem Ausmaß auf die verschiedenen Messverfahren zu. Daher soll dasjenige Verfahren eingesetzt werden, bei dem die Validitätseinschränkungen insgesamt am geringsten ausfallen. Zu diesem Zweck werden im Folgenden alle Probleme diskutiert, die beim jeweiligen Verfahren vorliegen. Die Diskussion einer Kriteriumsvalidierung erübrigt sich, da eine solche für kein einziges Verfahren durchgeführt wurde, was an den damit verbundenen ethischen Problemen liegen dürfte.

4.4.3.1 TEACHER/LEARNER UND ESSAY EVALUATION

4.4.3.1.1 KONSTRUKTVALIDITÄT

Tedeschi und Quigley (1996) weisen darauf hin, dass das Teacher/Learner Paradigm und das Essay Evaluation Paradigm mehrere Mängel gemeinsam haben, die die Konstruktvalidität

Virtuelle Nothilfe

bedrohen bzw. einschränken: (1) In beiden Verfahren gibt es eine Autorität (= Versuchsleiter), die das aggressive Verhalten toleriert (*Konformität bzw. implizite Aufforderung*); (2) in beiden gibt es nur eine Reaktionsmöglichkeit, nämlich sich aggressiv zu verhalten (*implizite Aufforderung*); (3) in beiden ist die Distanz zwischen Proband und Konföderiertem hoch (*Distanz*); (4) beide verwenden eine Cover Story, aufgrund derer die Probanden die Bestrafung aus altruistischen Gründen durchführen könnten (nämlich um dem „Schüler“ zu helfen) und sich dadurch nicht als aggressiv erleben würden (*nicht-aggressive Motive*). Solche Probanden würden den aggressiven Akt des Konföderierten als Beleg für dessen Dummheit wahrnehmen. Da die Schocks das Lernen fördern sollen, würden dümmere Menschen stärkere Schocks benötigen. Lieberman et al. (1999) weisen außerdem darauf hin, dass viele der üblichen Probanden die Milgram-Studien kennen, sodass sie wissen, dass in Wirklichkeit keine Stromstöße verabreicht werden. Dies kann im ungünstigen Fall einer Konfundierung zu einer Störung der Validität führen. Giancola und Chermack (1998) nennen als Beleg für die Validität des Verfahrens eine Studie, in denen per se aggressivere Menschen höhere Werte im Teacher/Learner Paradigm einstellen als weniger aggressive. Ein solcher Validitätsnachweis durch „Gruppendiskrimination“ ist aber schwach, da er auf einem logischen Fehlschluss beruht (s. Abschnitt 4.4.2).

4.4.3.1.2 PRAKTISCHE PROBLEME

Lieberman et al. (1999) benennen außerdem zwei praktische Probleme: Erstens ist die nötige technische Ausrüstung teuer, und zweitens verweigern Ethik-Kommissionen in der Regel die Durchführung.

4.4.3.2 TAP UND CRTT

4.4.3.2.1 KONSTRUKTVALIDITÄT

Die Kritik am Teacher/Learner Paradigm lässt sich ebenfalls an TAP und CRTT üben. Auch hier gibt es nach Tedeschi und Quigley (1996) einen Versuchsleiter, der das aggressive Verhalten toleriert (*Konformität bzw. implizite Aufforderung*); auch hier ist die Distanz zwischen Proband und Konföderiertem hoch (*Distanz*); auch hier gibt es keine nicht-aggressive Verhaltensalternative (*implizite Aufforderung*). Nach Tedeschi und Quigley (2000) kommt hinzu, dass die Versuchspersonen zwar die Intensität des Stromschlags variieren können, jedoch nicht die Wahl haben, gar keinen Stromschlag zu verabreichen (*implizite Aufforderung*). Giancola und Chermack (1998, S. 244) argumentieren hingegen, dass im TAP auch eine nicht-aggressive Option zur Verfügung steht, da die niedrigste Schockeeinstellung von den Probanden nach deren Aussage als nicht-aggressives Verhalten wahrgenommen würde. Im CRTT besteht dieses Problem nicht mehr, da (zumindest in aktuellen Versionen des Programms) der Proband den Wert Null für die Lautstärke einstellen kann. Außerdem kann er als weitere Verhaltensalternative seinen „Gegner“ durch eine Zahlung von Geld beschwichtigen. Neben

diesen beiden Möglichkeiten sind keine weiteren nicht-aggressiven Verhaltensweisen vorgesehen. Es lässt sich darüber streiten, ob diese Verhaltensalternativen eine sinnvolle Interaktion mit dem Gegenüber ermöglichen.

Tedeschi und Quigley (1996) sehen im Vergleich zum Teacher/Learner Paradigm einen Vorteil darin, dass eine Cover Story fehlt und die Schädigungsabsicht damit eher angenommen werden kann (*nicht-aggressive Motive*). Sie benennen aber gleich mehrere alternative Motive für das Einstellen der Schockstärken durch die Probanden: um den Gegner davon abzuhalten, dass dieser höhere Werte einstellt (soziale Kontrolle), um nicht als Verlierer dazustehen (Selbstdarstellung bzw. Imagebildung), um sich für die Provokation zu rächen (Retribution bzw. ausgleichende Gerechtigkeit) oder um sich am Verhalten des anderen zu orientieren (Konformität). Zum einen ist dadurch unklar, ob eine Schädigungsabsicht bei jedem Probanden angenommen werden kann. Zum anderen können sich die Schockstärken aufschaukeln, d. h., die Probanden verabreichen immer lautere und längere Töne, weil die Konföderierten damit angefangen haben. Bei sehr wettbewerbsorientierten Probanden sei sogar vorstellbar, dass sie nur deshalb laute Töne verabreichen, weil sie den Gegner behindern und so den Wettbewerb gewinnen wollen. Bartholow und Anderson (2002) modifizierten daher das Verfahren, indem sie das CRTT zweimal hintereinander durchführen, wobei der Gegner die Einstellungen nur im ersten Durchgang und der Proband sie nur im zweiten Durchgang vornehmen kann. Dies soll ein Aufschaukeln der Reaktionen vermeiden. Abgesehen davon, dass sich diese Modifikation nicht durchgesetzt hat, sind bei ihrer Anwendung alternative Verhaltensklärungen weiterhin möglich, wie z. B. den Gegner behindern zu wollen. Daher versuchen Anderson und Murphy (2003), auch die Intentionen der Versuchsperson zu erfassen, indem sie die Versuchspersonen nach dem CRTT in einem Fragebogen (der sogenannten „Aggressive Motives Scale“) nach ihren Intentionen fragen. Anderson et al. (2004, S. 218f.) berichten für die aus zwei Items bestehende Skala „instrumentelle aggressive Motivation“ eine Korrelation in Höhe von $r = .57$ ($p < .001$) und für die Skala „Rachemotivation“ eine interne Konsistenz in Höhe von $\alpha = .74$. Es ist allerdings fraglich, ob durch eine nachträgliche Befragung die wirklichen Intentionen der Versuchsperson erfasst werden, oder ob es sich um eine bloße Rechtfertigung ihres vorangegangenen Verhaltens handelt.

In mehreren Studien wurde ein Nachweis konvergenter Validität durch „Gruppendiskrimination“ durchgeführt. Personen mit hoher Trait-Aggression stellen höhere Werte im TAP ein (Taylor, 1967; Giancola & Zeichner, 1995b). Giancola und Chermack (1998) nennen drei weitere Studien, in denen per se aggressivere Menschen höhere Werte im TAP einstellen als weniger aggressive. Tedeschi und Quigley (2000) kritisieren zwar nicht die Studien an sich, behaupten aber, es könnte sich um eine selektive Auswahl handeln, da Studien mit Nullkorrelation wie die von Muntaner et al. (1990, zit. n. Tedeschi und Quigley, 2000, S. 133) unerwähnt bleiben. Giancola und Zeichner (1995c, zit. n. Giancola & Chermack, 1998, S. 247) konnten außerdem nachweisen, dass ihre Version des TAP bei Trunkenheit mit der BDHI Assault Scale korreliert. Insgesamt lässt sich festhalten, dass die Studien zwar die konvergente

Virtuelle Nothilfe

Validität des Verfahrens stützen, aber nur indirekt und schwach, da der Nachweis durch Gruppendiskrimination geführt wurde.

Daneben existieren (mindestens) zwei Studien zur diskriminanten Validität des Verfahrens. Nach Bernstein, Richardson und Hammock (1987) hängt der TAP nicht mit Hilfe, Schuld, Verdacht, indirekter Feindseligkeit und nach innen gerichtetem Ärger zusammen, er korreliert also nicht mit Konstrukten, die nichts mit Aggression zu tun haben. In der Studie wurden folgende Maße verwendet: zwei Formen von Hilfeverhalten, eine Variante des Gefangenendilemmas (Prisoner's Dilemma Game), das Teacher/Learner Paradigm sowie der TAP. Die Hilfe Maße korrelierten aber nicht miteinander, was ihre Validität infrage stellt. Aggression im Teacher/Learner Paradigm korrelierte zwar mit der Schockintensität im TAP, aber nach Bernstein et al. liegt signifikante Methodenvarianz vor, da in beiden Paradigmen die Schockereinstellungen die abhängige Variable darstellen. Tedeschi und Quigley (2000) kommen daher zu dem Schluss, dass diese Ergebnisse die diskriminante Validität des TAP nicht stützen. Daneben finden sie es überraschend, dass der TAP nicht mit Wettbewerb korreliert, obwohl die Versuchspersonen instruiert werden, dass sie sich in einer Wettbewerbssituation befinden. Neben der Studie von Bernstein et al. (1987) existiert noch eine weitere von Giancola und Zeichner (1995b), in der eine Variante des TAP nicht mit Subskalen von Instrumenten korreliert, die etwas anderes als Aggression erfassen.

Von Anderson (z. B. Anderson et al., 2004, S. 216) werden als Gegenbeleg zum Vorwurf mangelnder Validität häufig verschiedene Validierungsstudien angeführt. An diesen Validierungsstudien kritisieren Ferguson (2007a) sowie Ferguson, Smith, Miller-Stratton, Fritz und Heinrich (2008b) mehrere Punkte: (1) Die externe Validität sei trotz der gegenteiligen Behauptung von Anderson und Dill (2000) bisher nicht gesichert, denn keine der Validierungsstudien würde zeigen, dass eine höhere Anzahl an lauten Rauschtönen mit einem externalen Aggressionsindikator (wie beispielsweise krimineller Gewalt) zusammenhängt. Stattdessen würde die Studien nur eine sehr indirekte Methode zur Sicherung der externen *Validität* verwenden, indem sie aufzeigen, dass korrelative und experimentelle Studien zu denselben Ergebnissen gelangen. Außerdem würden Untersuchungen zur *Reliabilität* des CRTT fehlen.

4.4.3.2.2 UNSTANDARDISIERTE AUSWERTUNG

Ferguson et al. (2008a, 2008b) kritisieren, dass zwar die Erhebung der Daten standardisiert sei¹⁰, nicht aber die Auswertung der Ergebnisse. Es bleibe offen, ob Aggression als Lautstärke, als Dauer oder beides gemessen werden soll bzw. ob ein Summenwert oder ein Cut-Off verwendet werden soll (zu den Varianten siehe Abschnitt 4.4.1.2). In den Studien von Anderson würden die Ergebnisse immer auf eine andere Weise ausgewertet, was eine Form von „capitalization on chance“ darstellen würde. In einigen Studien werden außerdem mehrere Scores gleichzeitig für die Auswertung herangezogen (z. B. Anderson & Dill, 2000: Lautstärke und Dauer nach Sieg und nach Niederlage; aber auch wie bereits erwähnt Anderson et al.,

2004: Runde 1, 2 bis 9, 10 bis 17 und 18 bis 25). Ferguson et al. (2008a, 2008b) kritisieren an einer parallelen Berechnung mehrerer Scores, dass eine notwendige α -Fehler-Adjustierung nicht vorgenommen wird. Dieses Vorgehen stelle ebenfalls eine Form von „capitalization on chance“ dar. Eine solche Aufteilung in vier Maße findet sich auch in den Validierungsstudien zum CRTT wieder. Auch hier wurde keine Korrektur (z. B. durch Verwendung einer kombinierten Effektstärke) durchgeführt. Ferguson et al. (2008a, 2008b) zweifeln außerdem an der Notwendigkeit einer Differenzierung zwischen Sieg- und Niederlage-Durchgängen (vgl. Anderson & Dill, 2000), da in ihrer Studie (2008a) die Konsistenz des Lautstärkemaßes über alle Durchgänge hinweg $\alpha = .90$ bzw. die des Dauermaßes $\alpha = .98$ betrug, was dafür spricht, dass kein Unterschied zwischen Sieg- und Niederlage-Durchgängen bestehen kann, weil die Konsistenz in dem Fall niedriger ausfallen müsste. Außerdem stellen sie fest, dass die Korrelation zwischen Lautstärke und Dauer lediglich $r = .30$ beträgt. In einer weiteren Studie (2008b) versuchten die Autoren, Lautstärke (als Summenwert sowie als Cut-Off) und Dauer (als Summenwert) mit verschiedenen (indirekten) Maßen für Aggression (Trait, mangelnde Impulskontrolle, mangelnde verbale Intelligenz) in Beziehung zu setzen. Personen mit höherer Trait-Aggression stellten höhere Lautstärken ein, während Personen mit höherer Impulskontrolle bzw. mit höherer verbaler Intelligenz höhere Dauerwerte einstellten. Aufgrund dieser Ergebnisse zweifeln die Autoren an der Validität der Dauer als Maß für Aggression. Sie führen daher eine „Standardisierung“ durch, indem sie das Dauermaß entfernen. An der Studie lässt sich allerdings kritisieren, dass die Stichprobe nur aus Studierenden bestand, was die Varianz bei Impulskontrolle und Intelligenz einschränkt. Außerdem war sie mit $N = 53$ recht klein, sodass bei nicht-signifikanten Effekten offenbleibt, ob diese auch in der Population nicht vorhanden sind. Hinzu kommt, dass die Autoren bei der Auswertung nicht zwischen Männern und Frauen unterschieden, was die niedrige Korrelation zwischen Lautstärke und Dauer erklären könnte. Giancola und Zeichner (1995a) zeigen nämlich, dass differenzielle Validität vorliegt: Alkoholisierte (und damit aggressivere) weibliche Studenten erhöhen nur die Dauer, nicht aber die Stärke des Schocks, während alkoholisierte männliche Studenten sowohl Dauer als auch Stärke erhöhen. Dies könne laut den Autoren dafür sprechen, dass die Dauer ein Maß für indirekte Aggression darstellt (siehe auch Ritter & Eslea, 2005). Wenn das stimmt, würde eine Entfernung des Dauermaßes den CRTT für weibliche Probanden ungeeignet machen. Die Autoren nennen aber auch zwei Tatsachen, die gegen ihre Interpretation sprechen: zum einen, dass die Erhöhung der Dauer auf durch Alkohol verminderte Bewegungsfähigkeit zurückzuführen sein könnte; zum anderen, dass indirekte Aggression definiert ist als Aggression durch soziale Manipulation und/oder als Aggression, bei der dem Opfer die Identität des Täters nicht bekannt ist. Differenzielle Validität würde die bei Ferguson et al. (2008a) zu findende Korrelation zwischen Geschlecht und Lautstärke-Einstellung in Höhe von $r = .28$ erklären können. Differenzielle Validität findet sich auch in einer Studie von Hokanson (1970, zit. n. Tedeschi & Felson, 1994, S. 45), bei der Männer im CRTT anders reagierten als Frauen. In dieser Studie wurde eine Variante des CRTT eingesetzt, bei der neben Bestrafung auch Belohnung und Aussetzen möglich ist. Das Programm wurde so eingestellt, dass

der Computer dennoch ständig bestraft. Männer erteilten daraufhin Schocks zwecks Vergeltung, während Frauen Belohnungen zwecks Beschwichtigung bevorzugten. Bei beiden Geschlechtern führte diese spezifische Reaktion zu einem raschen Absinken ihrer physiologischen Erregung, während jede andere Reaktion die Erregung nur langsam schwinden ließ, was als kathartischer Effekt der spezifischen Reaktion interpretiert wurde.

Der Vorschlag von Ferguson et al. (2008a, 2008b), zwecks Standardisierung auf das Dauermaß zu verzichten, könnte daher den CRTT als Aggressionsmaß für Frauen ungeeignet machen. Auf der anderen Seite zeigt die Studie von Elson (2011, S. 62), dass eine Standardisierung dringend angestrebt werden sollte. Elson verglich verschiedene Varianten der Auswertung (siehe Abschnitt 4.4.1.2). Je nach Variante war Gruppe A aggressiver als Gruppe B, oder Gruppe B aggressiver als A, oder es gab keinen Unterschied zwischen den Gruppen. Bei einer Standardisierung wäre zu berücksichtigen, dass nach Bushman und Baumeister (1998) sowie nach Konijn et al. (2007) der Wert aus der ersten Runde des CRTT das beste Maß für Aggression darstellt, da die Einstellungen in den Folgerunden die reziproke Antwort des Probanden auf das darstellen, was der Gegenspieler ihrer Vermutung nach getan hat. Es handelt sich hier um sozialpsychologische Prozesse (z. B. Tit for Tat), die nicht mit Aggression zusammenhängen (müssen). Dabei reagiert der Proband anscheinend eher auf die dem Gegenüber unterstellte Schädigungsabsicht als auf die tatsächliche Wirkung der Einstellungen (Epstein & Taylor, 1967; Ohbuchi & Kambara, 1985; beide zit. n. Tedeschi & Felson, 1994, S. 224). Neben diesem Problem ist es außerdem denkbar, dass der Proband Zweifel an der Echtheit des Gegenspielers bildet, wenn dieser sich anders „verhält“, als der Proband von einem Menschen erwartet. Dies würde alle Messwerte in denjenigen Runden verfälschen, die auf das Durchschauen der Täuschung folgen.

4.4.3.2.3 PRAKTISCHE PROBLEME

Neben den genannten Validitätsmängeln sehen Lieberman et al. (1999) das praktische Problem, dass Ethikkommissionen das Verfahren oft nicht akzeptieren, weil die Versuchsperson hier eine Reihe von unangenehmen Tönen erleiden muss.

4.4.3.3 BOBO MODELLING

Nach Tedeschi und Quigley (1996) bzw. Tedeschi und Felson (1994) liegt der Verdacht nahe, dass es sich beim „Bobo Modeling Paradigm“ nicht um aggressives Verhalten handelt, sondern um eine Form von rauem Spielverhalten (engl. „rough and tumble play“) oder Imitation (engl. „follow the leader“). Schließlich sei die Bobo-Puppe so robust, dass selbst Erwachsene sie nicht beschädigen können. Deswegen gäbe es für Kinder keinen Grund, davon ausgehen, dass es ihnen möglich sei, der Puppe Schaden zuzufügen. Laut den Autoren gibt es keine Studie, die eine Schädigungsabsicht der Kinder nachweist. Außenstehende Beobachter hingegen würden das Verhalten der Kinder als Imitation interpretieren (Joseph, Kane, Nacci &

Tedeschi, 1997, zit. n. Tedeschi & Quigley, 1996, S. 170). Außerdem gibt es keine Studien, die belegen, dass das Verfahren auch bei Jugendlichen bzw. Erwachsenen eingesetzt werden kann. Da es somit für diese Untersuchung nicht verwendbar ist, wird hier nicht weiter auf die Konstruktvalidität des Verfahrens eingegangen.

4.4.3.4 POINT SUBTRACTION AGGRESSION

Im Gegensatz zum TAP bzw. CRTT ist der Proband beim PSAP nicht dazu gezwungen, sich aggressiv zu verhalten, um zu gewinnen. Den Vorwurf von Tedeschi und Quigley (1996), dass das Verfahren möglicherweise Konformität statt Aggression erhebt, entkräften Giancola und Chermack (1998, S. 244) daher mit dem Argument, dass im PSAP auch eine nicht-aggressive Option zur Verfügung steht. Bei dieser Option drückt der Proband eine Taste, um für sich selbst Punkte zu gewinnen. Da hier keine Interaktion mit dem Gegenüber mehr stattfindet, könnte diese „Alternative“ von den Probanden aber auch als Teilnahmeverweigerung („dropping out of the experiment“) interpretiert werden (vgl. Ritter & Eslea, 2005, S. 414f.), wodurch wieder ein Problem durch Konformität bzw. implizite Aufforderung entsteht. Außerdem finden sich hier die Probleme der Distanz zwischen Aggressor und Opfer (*Distanz*) sowie der Tolerierung der Aggression durch den Versuchsleiter (*Konformität bzw. implizite Aufforderung*) wieder.

4.4.3.4.1 KONSTRUKTVALIDITÄT

Zu dem Verfahren liegen mehrere Studien vor, die die konvergente Validität über Gruppendiskrimination geprüft haben. Giancola und Chermack (1998) nennen vier Studien, in denen per se aggressivere Menschen höhere Werte im PSAP einstellen als weniger aggressive. Cherek, Spiga und Steinberg (1989, zit. n. Giancola & Chermack, 1998) sowie Cherek, Steinberg, Kelly und Robinson (1987, zit. n. Giancola & Chermack, 1998) konnten außerdem zeigen, dass sich das Punktabzugsverhalten und das Punktgewinnverhalten der Probanden im PSAP nach Alkoholenuss unabhängig voneinander verändern. Giancola und Chermack (1998) interpretieren das als Nachweis dafür, dass im PSAP aggressives Verhalten von anderem Antwortverhalten (z. B. generelle Erregung oder Sedierung) abgegrenzt ist. Da in den Studien die Schädigungsabsicht der Probanden unerfasst bleibt, lässt sich aber nicht sicher feststellen, ob es sich überhaupt um aggressives Verhalten handelt. Außerdem ist ein Validitätsnachweis über Gruppendiskrimination indirekt und schwach. Es existiert aber auch eine Studie mit direkterem Nachweis der konvergenten Validität. Cherek, Moeller, Schnapp und Dougherty (1997, zit. n. Giancola & Chermack, 1998, S. 247) konnten nachweisen, dass das PSAP hoch mit selbst berichteter Aggression zusammenhängt. Hier bleibt aber offen, wie valide ein solcher Selbstbericht ist.

4.4.3.5 HOT SAUCE

4.4.3.5.1 KRITERIUMSVALIDITÄT

Eine Kriteriumsvalidierung des Hot Sauce Paradigm im engeren Sinne ist bisher nicht erfolgt. Lieberman et al. (1999) begründen aber die externe Validität des Verfahrens damit, dass scharfe Gewürze bereits für echte Angriffe sowie in Fällen von Kindesmissbrauch genutzt wurden (zu Beispielen siehe McGregor et al., 1998, Lieberman et al., 1999; Ritter & Eslea, 2005).

4.4.3.5.2 KONSTRUKTVALIDITÄT

Die konvergente Validität wird dadurch gestützt, dass das Verfahren mit Buss und Perrys Aggressionsfragebogen korreliert (Lieberman et al., 1999). Dies gilt allerdings nur, wenn die Versuchspersonen Informationen rational und nicht erfahrungsbezogen verarbeiten. Die Autoren erklären dieses Ergebnis dadurch, dass Menschen im erfahrungsbezogenen Modus weniger von ihren Persönlichkeitsdispositionen geleitet werden. Die Konstruktvalidität wird zusätzlich gestützt durch eine Validierung von McGregor et al. (1998) im Rahmen der „Terror Management Theory“. Die Autoren konnten zeigen, dass Probanden nach einer Provokation mehr scharfe Soße verabreichen. Die Provokation bestand in der Induktion von Mortalitäts-salienz und anschließender Bedrohung der Weltanschauung. Sobald die Probanden die Möglichkeit erhielten, vor der Verabreichung der Soße den Provokateur schriftlich (negativ) zu bewerten, sank die Menge verabreichter Soße. Die Autoren erklären das damit, dass die Bedrohung der Weltanschauung Dissonanz erzeugt und sowohl die Verabreichung der scharfen Soße als auch die Negativbewertung eine konkurrierende Möglichkeit zur Dissonanzreduktion darstellen. Ihre Ergebnisse würden somit die Konstruktvalidität des Hot Sauce Paradigms stützen. Ritter und Eslea (2005) weisen aber darauf hin, dass die Stimmungsinduktion fehlgeschlagen ist, weil es keinen signifikanten Unterschied zwischen der Experimental- und der Kontrollgruppe gab. Tatsächlich hatte die Experimentalgruppe sogar eine bessere Stimmung. Daher zweifeln Ritter und Eslea (2005) an einer Schädigungsabsicht in dieser Gruppe. Außerdem lässt sich einwenden, dass auch nicht-aggressive Verhaltensweisen Dissonanz reduzierend sein können und somit unklar bleibt, ob die Verabreichung scharfer Soße überhaupt einen aggressiven Akt darstellt. Einen Beleg für die Schädigungsabsicht sehen McGregor et al. darin, dass eine weitere Probandengruppe ($n = 10$) die im Durchschnitt verabreichte Soße als recht scharf (7,2 auf einer Skala von 1 bis 9) und als recht schmerzhaft (7,8 auf einer Skala von 1 bis 9) bewertete. Die Stichprobe ist mit $N = 10$ allerdings recht klein. Ritter und Eslea (2005) kritisieren außerdem, dass die Motive bisher nicht detailliert analysiert wurden, weswegen es noch keinen starken Beleg für eine Schädigungsabsicht geben würde.

Bei dem Verfahren finden sich auch die konstruktvaliditätseinschränkenden Probleme der Distanz zwischen Aggressor und Opfer und der Tolerierung der Aggression durch den Versuchsleiter wieder. Ein Vorteil des Verfahrens hingegen besteht darin, dass der Proband

Virtuelle Nothilfe

nicht dazu gezwungen ist, sich aggressiv zu verhalten. Die Instruktion impliziert jedoch, dass das Verabreichen scharfer Soße als Reaktion auf eine Provokation akzeptabel ist (Ritter & Eslea, 2005).

Außerdem gibt es bei der ursprünglichen Variante keine Möglichkeit, auf eine nicht-aggressive Weise mit dem Gegner zu interagieren (Ritter & Eslea, 2005). Dagegen finden sich in der Erweiterung von Böhm und Streicher (2010) neben der scharfen Soße die Alternativen „Joghurt“ und „Mayonnaise“. Dabei nahmen die Autoren an, dass es sich beim Joghurt um eine nichtaggressive Alternative handeln würde und bei der Mayonnaise um eine aggressive Alternative, die bevorzugt von Frauen verwendet werden sollte. Die Validierungsstudie der Autoren zeigt aber einen Geschlechtereffekt, der für eine differenzielle Validität spricht: Die Autoren führten ihre Validierungsstudie sowohl mit männlichen als auch mit weiblichen Probanden und Provokateuren durch. Die männlichen Probanden verwendeten wie erwartet nach der Provokation hauptsächlich scharfe Soße, die weiblichen verwendeten aber auch in größerem Umfang scharfe Soße statt wie erwartet hauptsächlich Mayonnaise. Bei der ursprünglichen Variante des Hot Sauce Paradigm findet sich ein ähnlicher moderierender Effekt des Geschlechts. Frauen verabreichen hier signifikant weniger scharfe Soße als Männer, aber nur dann, wenn sie befürchten, später auf ihr „Opfer“ zu treffen (Evers, Fischer, Rodriguez Mosquera & Manstead, 2005). Böhm und Streicher (2010) erklären sich die unerwartet starke Verwendung der scharfen Soße durch die Frauen folgerichtig dadurch, dass die Frauen in ihrer Studie keine Vergeltung oder Gefahr für sich selbst fürchten mussten, weil der Provokateur nicht anwesend war und auch kein späteres Treffen in Aussicht stand. Den Geschlechtereffekt hingegen erklären sie durch unterschiedliche Theorien über Lebensmittel, die sie mit Aussagen der Männer aus dem Abschlussgespräch stützen. Männer scheinen Mayonnaise im Gegensatz zu Frauen nicht als ungesund, sondern als „nahrhaft und wohltuend“ anzusehen, während sie den fettarmen Joghurt negativ bewerten, da dieser „dem Körper nicht viel geben“ würde. Alle genannten Erklärungen erfolgten aber erst nach der Sichtung der Ergebnisse. Ohne weitere Validierungsstudien bleibt es daher ungeklärt, ob bzw. unter welchen Bedingungen es sich beim Joghurt tatsächlich um eine neutrale Alternative bzw. bei der Mayonnaise um eine aggressiv-weibliche Alternative handelt.

Die Validierungsstudien von Beier und Kutzner (2011a; 2011b) zeigen ebenfalls ein gemischtes Bild. In der ersten Studie gab es vier Versuchsgruppen, wobei die Probanden der ersten Gruppe „angenehmen“ Orangensaft verabreichen mussten, die der zweiten „neutralen“ Früchtetee, die der dritten „scharfen“ Orangensaft mit Tabasco. Die Probanden der vierten Gruppe hatten die Wahl, welchen der drei Safttypen sie verabreichten. Die drei Safttypen waren vorab von $N = 19$ Versuchspersonen hinsichtlich ihres Geschmacks bewertet worden (S. Beier, persönliche Mitteilung, 20.01.2012): Der angenehme Saft wurde als angenehm bewertet und der scharfe Saft als unangenehm, während sich der neutrale Saft dazwischen befand¹¹. Ebenfalls erwartungskonform wurde nur der scharfe Saft als scharf bewertet. In der ersten Studie zeigte sich, dass Probanden nach einer Provokation in jeder der drei Be-

dingungen ohne Wahlmöglichkeit unabhängig vom Typ des Saftes immer mehr Saft verabreichen, d. h. nicht nur mehr scharfen Saft, sondern auch mehr neutralen und mehr angenehmen Saft. Dieses Ergebnis ist schwierig zu deuten. Möglicherweise hatten die Probanden in der ersten Studie ein anderes Geschmacksempfinden als die Probanden aus dem Vortest. Unter dieser Voraussetzung wäre es denkbar, dass sie jeden Saft als aversiv wahrnahmen. Dies erscheint jedoch wenig plausibel. Außerdem wählten auch unter Wahlmöglichkeit lediglich 3 von 18 Probanden den scharfen Saft. Die Autoren erklären sich das dadurch, dass ihre Provokation möglicherweise nicht stark genug war, um eine aggressivere Reaktion zu erzeugen. Um zu prüfen, ob die Probanden den Saft als schädlich bewerten, wurde eine weitere Studie durchgeführt. In dieser wurde erfasst, (1) wie sehr die Probanden glauben, dass ihr Gegenüber das ausgewählte Getränk (nicht) mag, (2) wie sehr ihnen gefallen würde, dass dem Gegenüber das Getränk nicht schmeckt und (3) wie sehr ihnen gefallen würde, dass dem Gegenüber das Getränk gut schmeckt (S. Beier, persönliche Mitteilung, 16.11.2011). In der zweiten Studie fehlten dafür die drei Bedingungen ohne Wahlmöglichkeit. Es zeigte sich wie erwartet, dass nach einer Provokation häufiger der scharfe Saft gewählt wurde. Es gab aber auch Probanden, die den scharfen Saft auswählten, ohne dass sie eine Schädigungsabsicht äußerten. Auch dieses Ergebnis kann unterschiedlich gedeutet werden. Entweder war für diese Probanden der scharfe Saft nicht unangenehm¹², oder sie waren aufgrund der sozialen Unerwünschtheit von aggressivem Verhalten nicht bereit, ihre aggressive Absicht zugeben. In einer dritten Studie wurden daher die Absichten hinter der Saftauswahl genauer untersucht. Um soziale Unerwünschtheit auszuschließen, handelte es sich um eine Szenariostudie. Die Probanden sollten sich vorstellen, in einem Experiment zu sein, bei dem sie (a) keine Wahlmöglichkeit bei der Saftauswahl haben, (b) eine implizite Wahlmöglichkeit haben und (c) eine explizite Wahlmöglichkeit haben. Die abhängige Variable bestand in der Wahrnehmung der hinter der Auswahl stehenden Absicht. Die Probanden waren bei allen drei Varianten der Ansicht, dass eine Schädigungsabsicht vorliegt, bei den Varianten mit Wahlmöglichkeit empfanden sie die Schädigungsabsicht aber als stärker. Wenn man die Ergebnisse aller drei Studien über das adaptierte Verfahren zusammennimmt, so bleiben noch einige Fragen zur Validität offen. Es dürfte aber lohnenswert sein, diese zu klären, da das Verfahren der klassischen Vorgehensweise überlegen zu sein scheint.

4.4.3.6 BUNGLED PROCEDURE

4.4.3.6.1 KONSTRUKTVALIDITÄT

Eine systematische Konstruktvalidierung des Verfahrens wird bisher nicht berichtet. Nach Ritter und Eslea (2005) besteht ein großer Vorteil des Verfahrens darin, dass es hohe Augenscheinvalidität besitzt, da viele Gewalttaten in der realen Welt unter Schusswaffeneinsatz begangen werden. Der größte Nachteil besteht allerdings darin, dass mit dem Paradigma durch den Abbruch des Schussvorgangs kein aggressives Verhalten erhoben wird, sondern nur die Intention zu aggressivem Verhalten. Außerdem bestehen Zweifel an der Schädi-

Virtuelle Nothilfe

gungsabsicht, da (a) die Probanden die Aufgabe als unterhaltsame Erfahrung wahrnehmen, wobei die Unterhaltungsratings der beste Prädiktor für das „Gewaltverhalten“ waren (Russel et al. 1996; 2002; beide zit. n. Ritter & Eslea, 2005, S. 412), (b) es keinen Zusammenhang zur aggressiven Persönlichkeit¹³ gibt (Russel et al., 2002, zit. n. Ritter & Eslea, 2005, S. 412), (c) das Schussziel Schutzkleidung trägt, was möglicherweise den Anschein erweckt, dass ein Treffer keine Folgen hat (Ritter & Eslea, 2005), (d) zahlreiche Hinweisreize vorhanden sind, die die Billigung des Verhaltens anzeigen (Russel et al., 1996, zit. n. Ritter & Eslea, 2005, S. 415) und (e) die Cover Story behauptet, dass es lediglich um die Bewertung einer neuen Form der Unterhaltung durch Männer geht. Zusätzlich fehlt es an einer Alternative zur „aggressiven“ Reaktion. Die Probanden könnten zwar beschließen, nicht auf ihr „Opfer“ zu schießen, dies könnte aber als Teilnahmeverweigerung empfunden werden: Durch die Cover Story könnten sie sich genötigt fühlen, einen Schuss abzugeben, damit sie über ihre Eindrücke berichten können (Ritter & Eslea, 2005).

4.4.3.7 EXPERIMENTAL GRAFFITI

4.4.3.7.1 KONSTRUKTVALIDITÄT

Auf der einen Seite wird die Konstruktvalidität des Verfahrens indirekt dadurch gestützt, dass junge Männer nach Alkoholkonsum mehr Graffiti produzierten, was als Verstärkung vandalistischen Verhaltens durch Alkohol interpretiert wurde (Kortynk & Perkins, 1983, zit. n. Ritter & Eslea, 2005, S. 411). Auf der anderen Seite bestehen Zweifel an der Validität, da Norlander et al. (1998, zit. n. Ritter & Eslea, 2005, S. 413) feststellten, dass Frauen beim Malen der Graffitis einen höheren Grad an Elaboration erreichten, was die Autoren vermuten lässt, dass es sich (zumindest bei den Frauen) um kreatives Verhalten handeln könnte. Gegen diese Interpretation spricht aber, dass Elaboration und Graffitiproduktion nicht korrelieren (Ritter & Eslea, 2005). Trotzdem bleibt offen, ob die Probanden überhaupt von einer Schädigung durch ihr Verhalten ausgehen. Ihnen wird der Wert der Illustrationen nicht bekannt gemacht, und es fehlt ein Treatment-Check, der eine Provokation durch die Illustrationen nachweist. Somit könnte es sich nach Ritter und Eslea (2005) auch hier um eine Form des „rough and tumble play“ handeln. Die Tatsache, dass Frauen höhere Werte beim Bemalen der Illustrationen erreichen, ohne dass es einen Geschlechterunterschied bei der Aggression oder dem Zerreißen der Illustrationen gibt, ist nach Ritter und Eslea (2005) ein Indiz dafür, dass das Verfahren möglicherweise Konformität erfasst. Schließlich werden die Probanden explizit instruiert, die Illustrationen zu bemalen bzw. zu zerreißen.

4.4.4 AUSWAHL

Die Auswahl eines geeigneten Verfahrens fällt schwer, da bei allen Zweifel an der Validität bestehen. Zum Teil liegt dies daran, dass aus ethischen Gründen nur relativ milde Provokationen bzw. Aggressionen möglich sind (Ritter & Eslea, 2005). Ein ideales Verfahren wäre

nach Tedeschi und Quigley (1996, S. 171) eines, in dem sich die Probanden unbeobachtet fühlen und die Gründe für ihr aggressives Verhalten selbst generieren. Mindestens aber sollten die Intentionen der Versuchspersonen untersucht werden und ihnen die Wahl zwischen verschiedenen aggressiven und nicht-aggressiven Maßnahmen zur Konfliktlösung gelassen werden (Tedeschi & Quigley, 1996; 2000). Dabei sollte den Probanden die aggressive Option nicht implizit nahegelegt werden (Ritter & Eslea, 2005). Daher wurde darauf geachtet, ein Verfahren auszuwählen, bei dem eine Erhebung der Intentionen möglich ist. Dadurch wurde das Bungled Procedure Paradigm ausgeschlossen, das aber ohnehin nicht eingesetzt werden sollte, da es nicht das Verhalten erfasst. Als Nächstes wurde das Bobo Modeling Paradigm ausgeschlossen, da unklar ist, ob es auch für Erwachsene geeignet ist. Beim Experimental Graffiti Paradigm gibt es empirische Indizien dafür, dass es statt aggressivem möglicherweise kreatives, konformes oder spielerisches Verhalten erhebt. Daher kam es ebenfalls nicht infrage. Das aggressive Verhalten im Point Subtraction Aggression Paradigm scheint im Vergleich zu den anderen Paradigmen keine allzu große Schädigung zu beinhalten, da dem Gegner hier lediglich Punkte abgezogen werden. Da außerdem keine Schädigungsabsicht nachgewiesen wurde, bleibt offen, ob das Verfahren überhaupt aggressives Verhalten erfasst. Daher wurde es aus der Auswahl ausgeschlossen. Das Teacher/Learner Paradigm hingegen wurde ausgeschlossen, weil es im Vergleich zum TAP/CRTT keine Vorteile hat, aber eine Cover Story beinhaltet, wegen der aggressives und prosoziales Verhalten konfundiert sind. Somit blieben CRTT und Hot Sauce Paradigm als mögliche Operationalisierungen übrig. Letztere hat den Vorteil, dass es in den modifizierten Varianten mehrere Verhaltensoptionen anbietet. Leider scheint die Modifikation von Böhm und Streicher (2010) zu einer differenziellen Validität bei Männern und Frauen zu führen. Da dieses Experiment auch mit weiblichen Probanden durchgeführt werden sollte, konnte diese Modifikation somit nicht eingesetzt werden. Ohne diese Modifikation hat das Verfahren aber keinen größeren Vorteil gegenüber dem CRTT. Aber auch in der unmodifizierten Variante sind die Werte von Männern nicht mit denen von Frauen vergleichbar, da das Geschlecht einen moderierenden Effekt auslöst. Die Modifikation von Beier und Kutzner (2011a; 2011b) scheint im Gegensatz dazu vielversprechend zu sein. Allerdings lag diese Modifikation noch nicht vor, als das Verfahren für dieses Experiment ausgewählt wurde, sodass sie nicht berücksichtigt werden konnte. Selbst wenn sie zu dem Zeitpunkt vorgelegen hätte, wäre sie genau wie das Originalverfahren und die Modifikation von Böhm und Streicher (2010) im Vergleich zum CRTT immer noch relativ neu und wenig erforscht, weswegen möglicherweise (weitere) noch nicht identifizierte Validitätsprobleme hätten enthalten sein können. Somit erschien das CRTT als dasjenige Verfahren, das im Vergleich mit den anderen die wenigsten Probleme enthält. Daher fiel die Wahl auf das CRTT.

4.5 HILFEVERHALTEN

Nach einer Diskussion der Theorien und Operationalisierungen zum Gewaltverhalten sollen als Nächstes in gleicher Weise Theorien zum Hilfeverhalten dargestellt werden. Beim Hilfe-

verhalten unterscheidet Bierhoff (2010, S. 13-16) zwischen hilfreichem Verhalten als Oberkategorie, prosozialem Verhalten als Unterform des hilfreichen Verhaltens und altruistischem Verhalten als Unterform des prosozialen Verhaltens. Die drei Formen unterscheiden sich wie folgt: Unter hilfreiches Verhalten fallen auch Aktivitäten, die durch berufliche Verpflichtungen bedingt sind (z. B. bei Feuerwehrleuten, Pflegekräften etc.). Um *prosoziales Verhalten* geht es hingegen nur dann, wenn die Aktivitäten nicht aufgrund von Dienstpflichten erfolgen. Prosoziales Verhalten kann das distale Ziel haben, das Wohlergehen einer anderen Person zu steigern, das eigene Wohlergehen zu steigern oder beides (siehe auch Abb. 8). Im Gewaltspielbereich ist das General Learning Model (GLM) eine Theorie, die zur Erklärung von prosozialem Verhalten eingesetzt wird. Diese Theorie soll zuerst beschrieben werden. Zur Erklärung von altruistischem Verhalten ist sie aber nicht geeignet, da sie nicht zwischen prosozial und altruistisch motiviertem Hilfeverhalten unterscheidet. Es wäre aber denkbar, dass sich virtuelles Hilfeverhalten nur durch eines der beiden Motive erklären lässt, weswegen eine Differenzierung grundsätzlich sinnvoll erscheint. Von *altruistischem Verhalten* spricht man dann, wenn das Wohl des Hilfebedürftigen das ausschlaggebende/entscheidende Ziel ist (Heckhausen, 1989, S. 298), wobei das eigene Wohlergehen durchaus ein Nebenziel sein darf (siehe dazu auch das Konstrukt „Zivilcourage“ in Abschnitt 4.5.4). Zur Beschreibung dieser Form von Hilfemotiv wird auf eine Theorie von Heckhausen zurückgegriffen. Für die Erklärung von Nothilfe (als Kombination von Hilfe- und Gewaltverhalten) scheinen aber Theorien zur Hilfe in Notsituationen bzw. in Zivilcourage-Situationen geeigneter zu sein, weswegen diese im Anschluss aufgeführt werden.

4.5.1 PROSOZIALE HILFELEISTUNG

Das General Aggression Model (siehe Abschnitt 4.3.1) wurde von Buckley und Anderson (2006) zum General Learning Model (GLM) verallgemeinert. Im Prinzip besteht der Unterschied zum GAM darin, dass das Modell nicht speziell das Erlernen von aggressivem Verhalten, sondern allgemein das Erlernen jeder beliebigen Verhaltensweise erklären will. Wie man in Abb. 6 sehen kann, beschreibt das vereinfachte Kurzzeitmodell des GLM, wie kurzfristige Lernprozesse generell ablaufen. Wenn z. B. eine Person nach einem Computerspiel die Gelegenheit erhält, einer anderen Person zu helfen oder zu schaden, dann hängt die Wahl der Verhaltensalternative davon ab, welche Verhaltensskripte durch das Spiel geprimed wurden und welches Verhalten durch das Spiel verstärkt wurde – erfordert das Spiel prosoziales Verhalten, dann erhöht sich das prosoziale Verhalten nach dem Spiel, erfordert es Gewalt, erhöht sich das aggressive Verhalten (Gentile et al., 2009). Das Langzeitmodell in Abb. 7 ergänzt das Kurzzeitmodell um den Aspekt, dass die Wiederholung eines Lernprozesses dauerhaft auf verschiedene kognitive und emotionale Eigenschaften einer Person und darüber auf ihre Persönlichkeit einwirkt.

Da den Verhaltensmotiven in dem Modell keine große Relevanz beigemessen wird, kann es nicht zwischen egoistisch und altruistisch motiviertem Hilfeverhalten differenzieren. Diese

Virtuelle Nothilfe

Unterscheidung ist aber wichtig, da altruistisches Hilfeverhalten in Kohlbergs Stufenmodell der Gerechtigkeitsmoral (vgl. Kohlberg, 2008; Heidbrink, 1996, S. 62-83; Mohseni, 2007, S. 13-15) als moralisch höher zu bewerten ist. Beispielsweise wäre egoistisches Hilfeverhalten eher auf den Stufen 1 bis 4 (Vermeidung von Strafe; eine Hand wäscht die andere; Konformität; Recht und Ordnung) zu verorten, während altruistisches Hilfeverhalten eher auf den Stufen 5 und 6 (Sozialvertragsordnung; kategorischer Imperativ) liegen würde. Außerdem ist Zivilcourage als eine Sonderform des Handelns in Notsituationen ebenfalls primär altruistisch motiviert. Eine Unterscheidung zwischen egoistischen und altruistischen Motiven kann aber nur unter Kenntnis der subjektiven Absichten erfolgen, weswegen diese Absichten mit in die Definition eingehen müssen. Daher reicht nach Heckhausen (1989, S. 279) eine behavioristische Verhaltensbeschreibung nicht aus. Dementsprechend konzipiert er ein handlungstheoretisches Motivationsmodell für Hilfeleistung, welches im nächsten Abschnitt beschrieben werden soll.

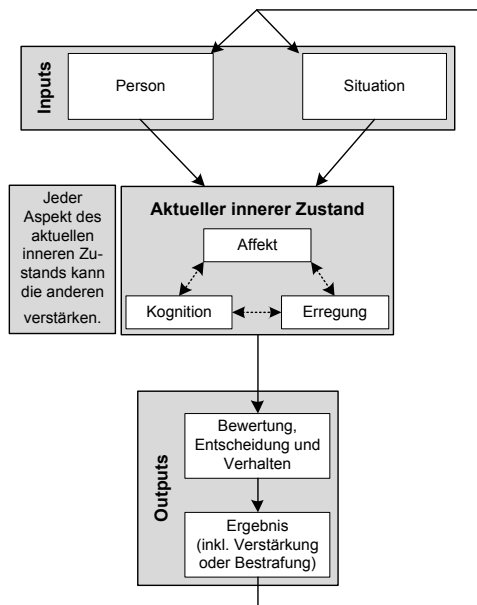


Abb. 6: Vereinfachte Darstellung der Kurzzeitprozesse im GLM, adaptiert nach Gentile et al., 2009, S. 754

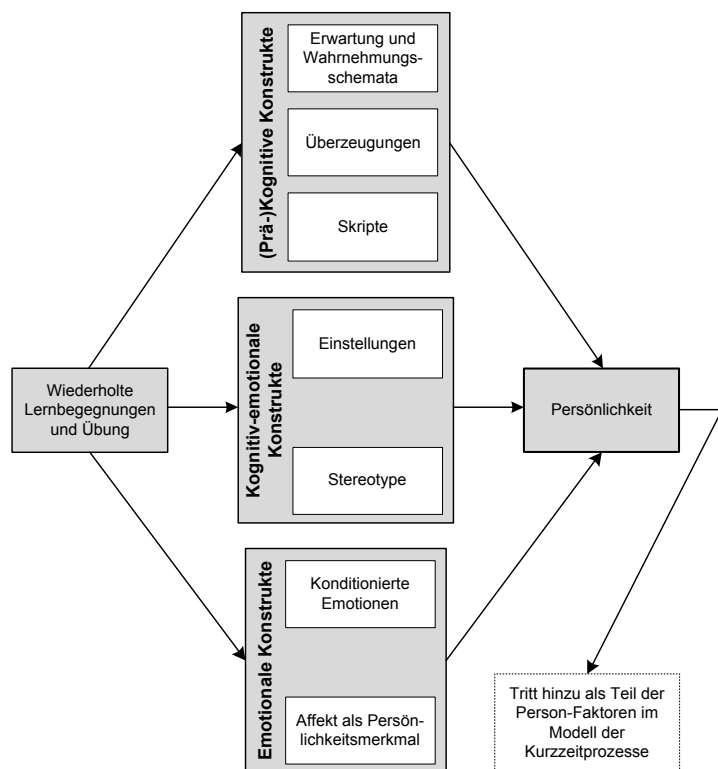


Abb. 7: Langzeitprozesse im GLM, adaptiert nach Gentile et al., 2009, S. 755

4.5.2 ALTRUISTISCHE HILFELEISTUNG

In Heckhausens Motivationsmodell expliziert er sowohl egoistisches als auch altruistisches Hilfehandeln (siehe Abb. 8). Hilfehandeln aus *egoistischen* Gründen lässt sich nach dem Mo-

Virtuelle Nothilfe

dell durch die Kosten, den Nutzen, die Fremd- sowie die Selbstbewertung erklären, die der Helfende als Folge seiner Handlung antizipiert. Zu Kosten und Nutzen zählt Heckhausen (1989, S. 301) materielle Güter, Zeitaufwand, Mühe, Verlust/Gewinn im Hinblick auf andere (selbstdienliche) Motive sowie die aversive Erregung (engl. „personal distress“). Übersteigen die erwarteten Kosten den erwarteten Nutzen, dann kommt es nach Schwartz und Howard (1981, zit. n. Bierhoff, 2010, S. 101-103) zu Abwehrprozessen (siehe Abschnitt 4.5.3). Neben Kosten und Nutzen ist auch die Fremdbewertung eine wichtige Folge. Darunter versteht Heckhausen (1989, S. 301) das Urteil anderer Personen bzw. der Öffentlichkeit über das eigene Handeln, insofern die Öffentlichkeit einem die Verantwortung für dieses Handeln zuschreibt. Hier spielen soziale Wertnormen eine Rolle, wie die Verantwortlichkeitsnorm und die Gegenseitigkeitsnorm. Im Prinzip geht es darum, durch die Einhaltung solcher Normen in der Öffentlichkeit eine positive Bewertung durch die Öffentlichkeit zu erfahren bzw. durch einen Verstoß gegen diese Normen das Gesicht zu verlieren. Damit ließe sich eine negative Fremdbewertung als eine spezielle Form von Kosten ansehen und eine positive als eine spezielle Form von Nutzen, auch wenn Heckhausens Modell dies so nicht (explizit) vorsieht.

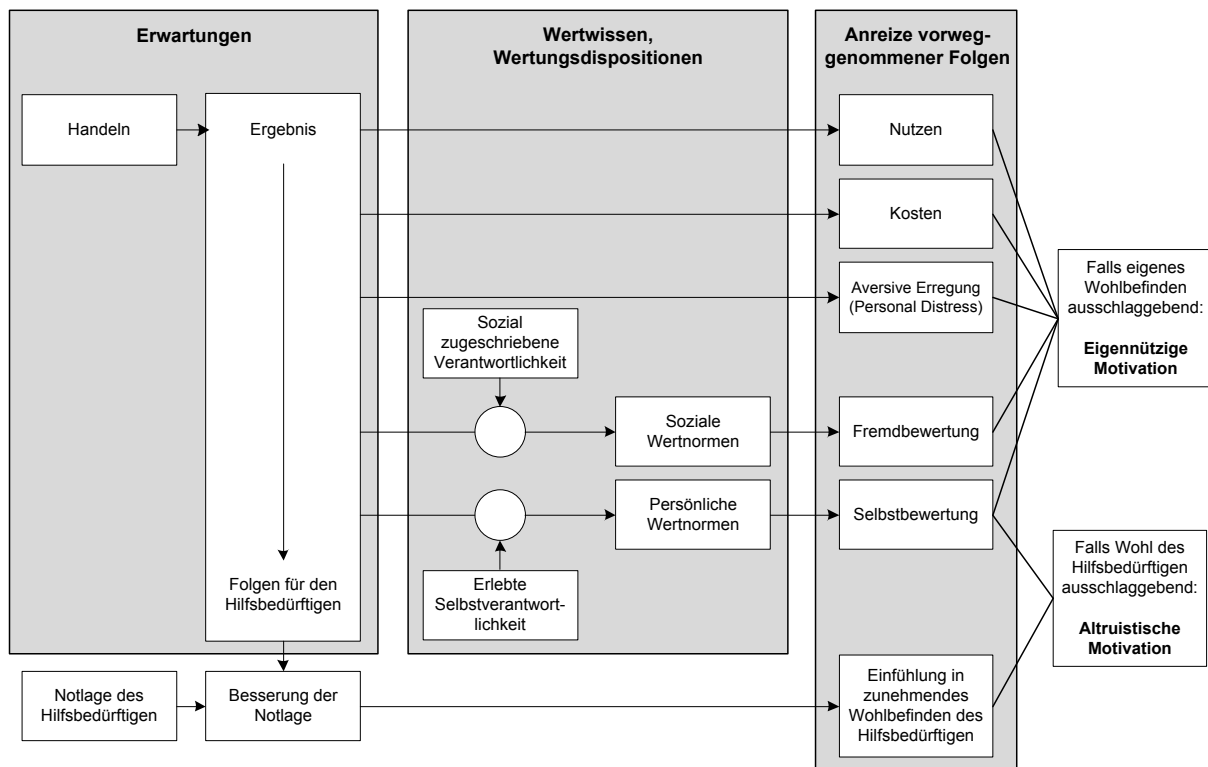


Abb. 8: Schema eines Motivationsmodells für Hilfeleistung (nach Heckhausen, 1989, S. 302)

Hilfhandeln aus *altruistischen* Gründen erklärt das Modell mit der antizipierten Selbstbewertung sowie der Einfühlung in die Notlage des Hilfsbedürftigen. Unter die Selbstbewertung fasst Heckhausen (1989, S. 301) positive und negative affektive Zustände, die durch persönliche Wertnormen verursacht werden, insofern man sich selbst (und nicht etwa äußeren Umständen oder anderen Personen) die Verantwortung für das eigene Handeln zu-

Virtuelle Nothilfe

schreibt. Zu den positiven affektiven Zuständen zählen die Selbstzufriedenheit darüber, den eigenen Wertnormen gerecht zu werden und die Genugtuung darüber, etwas Gutes getan zu haben; zu den negativen zählen Schuld und Scham, wenn man den eigenen Wertnormen nicht gerecht wurde. Hier findet sich eine Parallele zu den Aggressionstheorien „Moral Management Model“ und „Catharsis Model“, die ebenfalls solche moralischen Emotionen beinhalten. Wenn die Selbstbewertung allerdings lediglich zur Selbstwert-Erhöhung dient, dann wird sie nicht zu den altruistischen, sondern zu den egoistischen Zielen gezählt. Unter die emotionale Einfühlung in die Notlage des Hilfsbedürftigen fasst Heckhausen (1989, S. 301) die Antizipation, wie sich der Hilfebedürftige nach der Hilfehandlung fühlen wird. Dabei spielen Emotionen wie Erleichterung, Befreiung von Notgefühlen, Linderung von Schmerzen etc. eine Rolle. Die Einfühlung in den Hilfebedürftigen erfolgt über Sympathie, Mitleid und Empathie. Hierin liegt eine weitere Parallele zum Moral Management Model und zum Catharsis Model.

Egoistische und altruistische Gründe schließen sich in Heckhausens Modell nicht gegenseitig aus. Handlungen können gleichzeitig aus egoistischen als auch aus altruistischen Gründen erfolgen. Die Zuordnung erfolgt danach, welche Gründe die ausschlaggebenden sind. Vorschläge, wie sich das feststellen lassen könnte, finden sich bei Scheele und Kapp (2002, S. 27). Sind egoistische Gründe ausschlaggebend, so erfolgt eine Hilfehandlung immer dann, wenn der Gesamtnutzen die Gesamtkosten für den Helfenden übersteigen, während bei altruistischen Gründen eine Hilfehandlung am wahrscheinlichsten ist, wenn der Helfende sich in die Notlage des Hilfsbedürftigen einfühlt und sich von seinem Handeln eine Besserung der Notlage verspricht. Eine anti-altruistische Position, die altruistisches Helfen per se ausschließt, ist ohnehin nicht sinnvoll, da sie nach Scheele und Kapp (2002, S. 27) zu Erklärungslücken führt. Zum einen kann sie nicht erklären, warum es manchmal erheblich schwerer fallen kann, eine Hilfehandlung zu unterlassen, statt sie durchzuführen. Zum anderen wird ebenfalls nicht erklärt, warum das unangenehme Mitleiden selbst dann nicht zu einem Entziehen aus der Situation führt, wenn dies leicht zu bewerkstelligen wäre.

In Heckhausens Modell sind Persönlichkeitsdispositionen implizit mitgedacht, auch wenn diese nicht explizit abgebildet sind (1989, S. 303). Unter die Dispositionen fällt nach Heckhausen z. B. die Verantwortungsübernahme, während in Kapp und Scheeles (1996) Zivilcourage-Modell (siehe Abschnitt 4.5.4) die folgenden fünf Dispositionen genannt werden: (1) Verantwortungsbewusstsein, (2) Sensibilität/Bedürfnis, Leid zu vermindern, (3) Wissen über situationsspezifische Fähigkeiten, (4) Selbstsicherheit/Selbstbewusstsein und (5) Konfliktbereitschaft/Konfliktfähigkeit. Rein biologische Erklärungsansätze weist Heckhausen zurück, da die Sozialisation auf der einen Seite dazu führen könne, dass man nicht für seine engsten Angehörigen eintritt, und auf der anderen Seite, dass man allen Menschen (d. h. nicht nur Angehörigen) gegenüber altruistisch handle (1989, S. 281). Daher sind in dem Modell keine ausschließlich genetischen Erklärungsfaktoren enthalten.

4.5.3 HILFELEISTUNG IN NOTSITUATIONEN

Hilfesituationen unterteilt Heckhausen (1989, S. 282-285) in alltägliche Gefälligkeiten und in ernste Notlagen, wobei diese Unterscheidung im Modell nicht explizit abgebildet, aber implizit mitgedacht wird. Notsituationen zeichnen sich durch folgende Eigenschaften aus: (1) Sie kommen unvorhergesehen, sodass es keinen Handlungsplan gibt, (2) sie sind so selten, dass man keine Erfahrungen damit hat, (3) sie erfordern schnelles Reagieren, sodass längeres Nachdenken nicht möglich ist und (4) sie sind auch für den Helfer gefährlich. Nach Bierhoff (2010, S. 122) werden schwerwiegende Notsituationen im Vergleich zu leichteren eher als Bedrohung aufgefasst. In einem solchen Fall könne es zu Abwehrprozessen kommen, die eine Hilfeleistung unwahrscheinlich werden lassen. Schwartz und Howard (1981, zit. n. Bierhoff, 2010, S. 101-103) nennen die folgenden vier Abwehrmöglichkeiten: (1) Kleinreden der Notlage, (2) Verleugnung der Verpflichtung, (3) Verleugnung der Möglichkeit zu handeln und (4) Verleugnung der Möglichkeit, effektiv Hilfe zu leisten. Montada (2001, zit. n. Bierhoff, 2010, S. 66) nennt vier (zum Teil ähnliche) Möglichkeiten zur Abwehr von Verantwortung: (1) Kleinreden der Notlage, (2) Attribution der Schuld auf das Opfer, (3) Attribution der Verantwortung auf Andere und (4) Ansehen der eigenen Privilegien als gerechtfertigt. Bei Bandura (2002) finden sich im Rahmen des Moral Disengagement (siehe Abschnitt 4.3.4) acht Abwehrstrategien: (1) moralische Rechtfertigung, (2) euphemistische Beschönigung, (3) Ignorieren/Verzerren der Konsequenzen, (4) Entmenslichung, (5) Fremdschulduzuweisung, (6) vorteilhafter Vergleich und (7) Verantwortungsdiffusion. Alle genannten Abwehrstrategien haben gemeinsam, dass die Verantwortung für das eigene Handeln entweder wegrationalisiert oder wegattribuiert wird.

Bierhoff, Klein und Kramp (1990, zit. n. Bierhoff, 2010, S. 133-139) haben für Notsituationen (in ihrem Fall handelte es sich um Verkehrsunfälle) die folgenden Vorteile und Risiken herausgearbeitet: Bei den Vorteilen „Handeln nach Gewissen“ und „Kenntnisse zeigen“, gefolgt von „Mitgefühl ausdrücken“, „Schutz vor strafrechtlicher Verfolgung“ und „Selbstwertgefühl steigern“ sowie „Solidarität zeigen“ und „Anerkennung erhalten“; bei den Risiken „Fehler bei Erster Hilfe“, „Juristisches Nachspiel“, „Sich in Gefahr begeben“ und „Überfordert sein“, gefolgt von „Zeitverlust vermeiden“ und „Weitere Unfälle verursachen“. Bierhoff (2010, S. 136-138) fasst dies so zusammen, dass bei den Vorteilen das Verantwortungsgefühl dominiere und bei den Risiken die persönliche (In-)Kompetenz, weswegen beides zentral für Hilfe in Notsituationen sei. Dies zeige auch eine Studie von Schwartz und Ben David (1976, zit. n. Bierhoff, 2010, S. 137f.), bei der Verantwortung und Kompetenz systematisch variiert wurden. Bierhoff interpretiert die Ergebnisse dieser Studie so, dass nicht helfend eingegriffen wird, wenn die Verantwortung für die Notlage auf den Hilfsbedürftigen statt auf sich selbst attribuiert wird, oder wenn eine geringe Kompetenz als Entschuldigung vor sich und anderen genutzt wird. Eine Pfadanalyse der Daten von Bierhoff, Klein und Kramp (1990, zit. n. Bierhoff, 2010, S. 139-143) zeigt, dass sich die gefühlte Kompetenz nicht direkt auf die Hilfsbereitschaft auswirkt, sondern von Entschlussicherheit moderiert wird. Diese wiederum wirkt

Virtuelle Nothilfe

auf die Verantwortung und wird selbst von Instrumentalität mitbedingt. In einer der drei Unfallsituationen gab es außerdem eine Wirkung von Empathie auf die Verantwortlichkeit. Die Autoren erklären sich das so, dass in den anderen beiden Szenarien die Situation möglicherweise relativ unpersönlich war. Bleiben die Variablen „Training“ und „Alter“ unberücksichtigt, dann ergibt sich das Modell aus Abb. 9.

Das Modell lässt sich als Erweiterung der erlebten Selbstverantwortlichkeit in Heckhausens allgemeinem Motivationsmodell ansehen. In Notsituationen scheint diese insbesondere durch die Entschluss-sicherheit bedingt zu sein, welche wiederum von Instrumentalität und Kompetenz beeinflusst wird, aber auch Empathie kann eine Rolle spielen. Es gibt aber auch einige Parallelen zum Zivilcourage-Modell von Scheele und Kapp (2002), das im Folgenden erläutert werden soll.

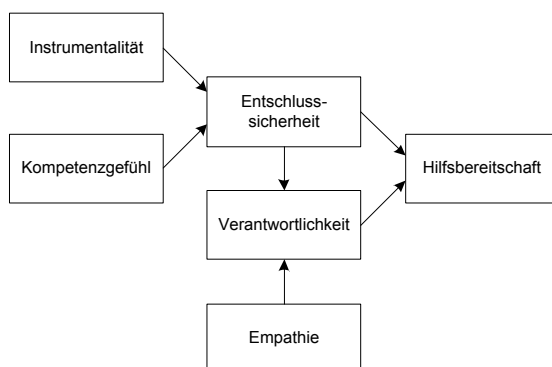


Abb. 9: Modell der Hilfsbereitschaft in Unfallsituationen (modifiziert nach Bierhoff, 2010, S. 142)

4.5.4 HILFELEISTUNG IN ZIVILCOURAGESITUATIONEN

Basierend auf subjektiven Theorien (siehe Abb. 10) definieren Scheele und Kapp (2002; siehe auch Scheele, 2004, S. 86) den Begriff der Zivilcourage wie folgt:

„Zivilcouragiertes Handeln (1) ist ein (verbales) Aufstehen/Handeln gegen kritikwürdige Zustände, Handlungen, Meinungen, etc., (2) wobei die (öffentliche) Mehrheit mit dieser Kritik nicht übereinstimmt; (3) enthält die Konfliktstruktur zwischen dem Bedürfnis, einerseits (soziale) Missstände zu bekämpfen, und andererseits dem Bedürfnis nach eigener physischer bzw. psychischer Unversehrtheit; (4) wobei die Bekämpfung der Missstände vor allem auf die altruistische Motivation der Verhinderung bzw. Verminderung von Leid anderer zurückgeht; (5) ist verbunden mit der Wahl zwischen der (subjektiv) belastenden Sanktion durch (über-)mächtige andere oder (bei Unterlassung von zivilcouragiertem Handeln) Belastung durch eigene Schuld- beziehungsweise Schamgefühle, (6) wobei sich diese Schuld- beziehungsweise Schamgefühle vor allem auf die moralische Integrität als zentralem Merkmal des eigenen Selbstkonzepts (Ideal-Selbst) beziehen; (7) es ist gekennzeichnet durch ein Veröffentlichungsobligat, wobei die moralische Empörung/Entrüstung das zentrale Moment für den Handlungsimpuls darstellt, (9) der durch Zielangemessenheit und RezipientInnenorien-

Virtuelle Nothilfe

tiertheit gekennzeichnet sein muss, (10) zugleich (aber) universelle Menschenrechte nicht verletzen darf.“ (S. 22)

Zivilcourage stellt somit eine Sonderform des altruistischen Hilfehandelns dar. Zivilcourage nach dieser Definition unterscheidet sich von dem hier verwendeten Begriff der Nothilfe hauptsächlich darin, dass bei Zivilcourage die mutige Hilfeleistung gegen den Konsens der Mehrheit stattfindet, während sie bei Nothilfe im Einklang mit der Mehrheitsmeinung steht. Zivilcourage unterscheidet sich von gewaltsamen Formen der Nothilfe außerdem darin, dass universelle Menschenrechte – zu denen auch die körperliche Unversehrtheit gehört – durch das zivilcouragierte Eingreifen nicht verletzt werden dürfen. Gewalt darf somit nur eingesetzt werden, wenn alle anderen Möglichkeiten ausgeschöpft sind und keine andere Wahl mehr bleibt. Nur wenn dies der Fall ist, kann gewaltsame Nothilfe als eine extreme Form zivilcouragierten Handelns betrachtet werden.

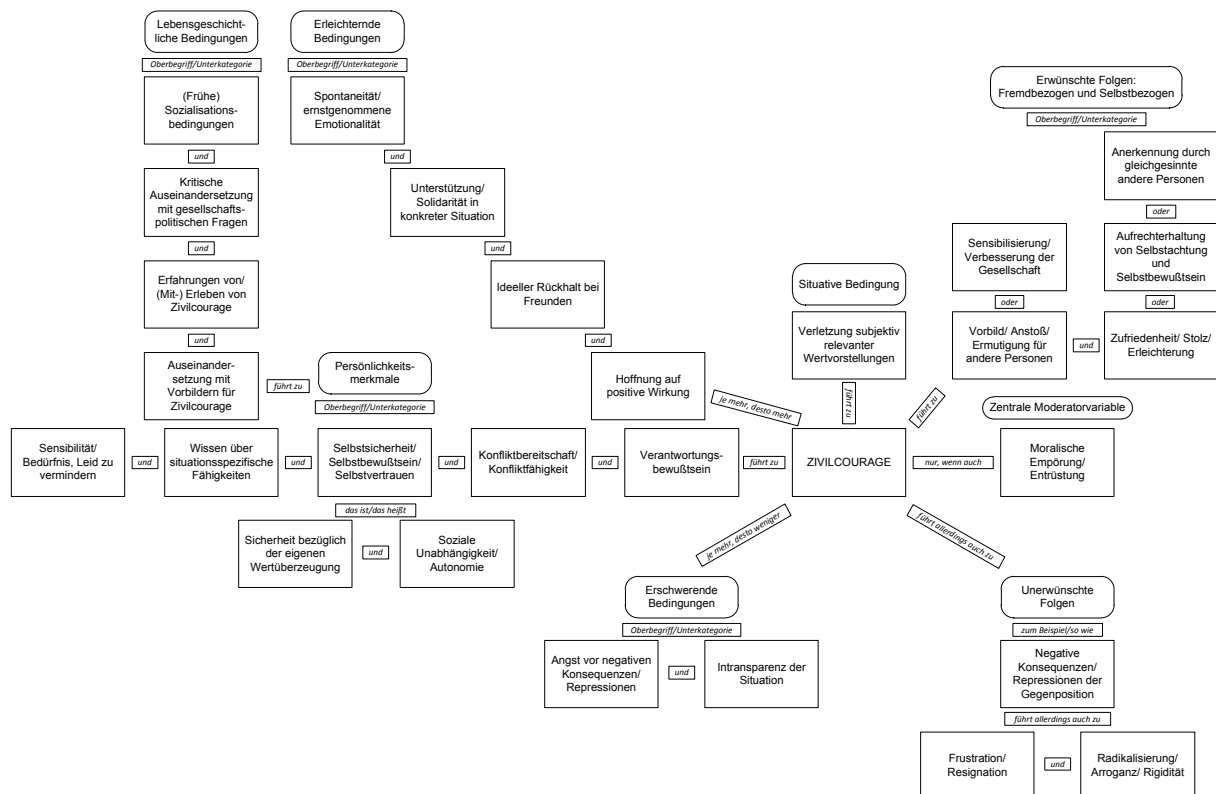


Abb. 10: Subjektive Theorien zu Zivilcourage: Empirischer Bereich (aus Kapp & Scheele, 1996, S. 137)

4.5.5 BEWERTUNG DER THEORIEN

Für die Ableitung der Hypothesen und für die Operationalisierung ist es wichtig, ob auf handlungs- oder verhaltenstheoretische Modelle zurückgegriffen wird, da die zugrunde gelegten Subjektmodelle nicht vollständig miteinander vereinbar sind (für eine wissenschaftstheoretische Begründung siehe Groeben, 1986). Dies zeigt sich z. B. daran, dass in handlungstheore-

tischen Modellen die Motive der Personen eine zentrale Rolle spielen, während sie bei verhaltenstheoretischen Theorien wie dem GLM unberücksichtigt bleiben. Ohne Kenntnis der Motive kann aber nicht ausgeschlossen werden, dass es sich bei dem beobachteten Verhalten um etwas anderes als Hilfeverhalten handelt. Hier besteht eine Parallelität zum aggressiven Verhalten, wo das gleiche Problem besteht. Beim Hilfeverhalten ist es außerdem nicht möglich, ohne Kenntnis der Motive zwischen egoistisch und altruistisch motiviertem Helfen zu unterscheiden. Daher sind handlungstheoretische Modelle zu bevorzugen. Das Modell von Heckhausen ist hierfür sehr gut geeignet, da es sowohl alltägliche Gefälligkeiten als auch Notsituationen erklären kann. Es erscheint aber ratsam, dieses generelle Modell um Aspekte von Scheeles Zivilcourage-Theorie zu ergänzen, da diese im Bereich der ernststen Notlagen wesentlich detaillierter gestaltet ist.

4.5.6 META-ANALYSEN

Zu den Auswirkungen von Gewaltspielen auf prosoziales Verhalten existieren nur wenige Studien (Happ, Melzer & Steffgen, 2010b, S. 30). Anderson et al. (2010) berichten in ihrer Meta-Analyse gerade einmal $k = 8$ experimentelle Querschnitt-Studien zur Wirkung von Gewaltspielen auf prosoziales Verhalten. Da unter prosozialem Verhalten auch Verhaltensweisen verstanden werden können, die nicht unbedingt etwas mit altruistischer Hilfe zu tun haben, wird die Anzahl der Studien mit Hilfeverhalten noch darunter liegen. Für die genannten $k = 8$ Studien (mit $n = 875$ Probanden) berichten Anderson et al. (2010) eine Effektstärke von $r^+ = -.16$. Auf den ersten Blick ist der Effekt damit vergleichbar mit der Wirkung von Gewaltspielen auf aggressives Verhalten. Dabei muss aber berücksichtigt werden, dass prosoziales Verhalten häufiger als alltägliche Gefälligkeit (z. B. heruntergefallene Bleistifte aufheben) denn als ernste Notlage operationalisiert worden ist. Da beim Eingreifen in ernststen Notlagen die Kosten in der Regel viel höher sind, dürfte die Effektstärke für solche Notlagen weit unter diesem Wert liegen. Beispielsweise zeigte sich in einer Studie von Chambers und Ascione (1987), dass die Wahl des Computerspiels sich auf das Spendeverhalten, aber nicht auf das Hilfeverhalten auswirkte. Happ, Melzer und Steffgen (2010b, S. 31) führen dieses Ergebnis darauf zurück, dass Spenden einfacher und weniger zeitraubend ist als Helfen.

Auch der Erforschung prosozialer Wirkungen von prosozialen Spielen wurde bisher wenig Beachtung geschenkt (Gentile et al., 2009). Von Greitemeyer und Oswald (2009; 2010), Greitemeyer, Osswald und Brauer (2010) sowie von Gentile et al. (2009) stammen die wenigen aktuellen Studien, die es zur prosozialen Wirkung von Computerspielen gibt.

Greitemeyer und Osswald (2009) führten zwei Experimente durch, in denen die Wirkung eines prosozialen Spiels (Lemmings) mit der eines neutralen Spiels (Tetris) verglichen wurde. Im ersten Experiment war die „Verfügbarkeit antisozialer Gedanken“ die abhängige Variable, im zweiten Experiment der „Hostile Attribution Bias“. Die Autoren konnten zeigen, dass das

Spielen eines prosozialen Spiels aggressive Kognitionen verringert. Greitemeyer und Osswald (2010) führten mit den Spielen Lemmings und Tetris vier weitere Experimente durch, wobei es im ersten Experiment zusätzlich eine Gewaltbedingung (Lamers) gab. Im ersten Experiment wurde die abhängige Variable „prosoziales Verhalten“ als Anzahl der Bleistifte operationalisiert, die der Proband innerhalb einer bestimmten Frist unaufgefordert aufhob, nachdem der Versuchsleiter sie „versehentlich“ vom Schreibtisch gestoßen hatte. Hier waren die Probanden in der prosozialen Bedingung hilfsbereiter als die in den anderen beiden Bedingungen ($\omega = 0.35, p < .05$). Im zweiten Experiment wurde prosoziales Verhalten als Zeitmenge operationalisiert, die ein Proband für weitere Experimente eines hilfebedürftigen Diplomanden opfern würde. In der prosozialen Bedingung waren mehr Probanden zu diesem Opfer bereit ($\omega = 0.46, p < .01$), und gleichzeitig fiel ihr Opfer größer aus ($d = 0.82, p < .05$). Im dritten Experiment galt es als prosoziales Verhalten, wenn die Probanden die Versuchsleiterin vor unerwünschten Zudringlichkeiten ihres „Ex-Freunds“ bewahrten, der von einem Konföderierten gespielt wurde. Der Sinn dahinter war, eine Operationalisierung zu nutzen, die mit höheren Kosten verbunden ist. Insgesamt weist diese Operationalisierung größere Übereinstimmung mit einer Zivilcourage-Situation auf (siehe Abschnitt 4.5.4). In der prosozialen Bedingung schritten wie erwartet mehr Probanden helfend ein ($\omega = 0.34, p < .05$). Allerdings können die Autoren nicht ausschließen, dass einige Probanden dabei eher die Absicht verfolgten, den Ex-Freund zu verletzen, als der Versuchsleiterin zu helfen. In dem Fall würde es sich um aggressives Verhalten oder eine Mischung aus aggressivem und prosozialem Verhalten handeln. Im vierten Experiment wurden neben prosozialem Verhalten (in Form der verschütteten Bleistifte) auch prosoziale Kognitionen erhoben, um deren Einfluss als Mediator überprüfen zu können. In der prosozialen Bedingung hatten die Probanden mehr prosoziale Kognitionen ($d = 1.44, p < .001$) und hoben mehr Bleistifte auf ($d = 0.89, p < .05$). Wie erwartet waren die Kognitionen ein signifikanter Mediator. Neben diesen Experimenten zum prosozialem Denken und Handeln führten Greitemeyer et al. (2010) auch zwei Studien zum prosozialem Affekt durch. Da der prosoziale Affekt in dieser Studie keine Rolle spielt, wird auf diese Studien mit Ausnahme der folgenden Ergebniszusammenfassung nicht weiter eingegangen: Prosoziale Spiele fördern Empathie und verringern Schadenfreude.

Gentile et al. (2009) führten eine korrelative, eine experimentelle sowie eine Längsschnittstudie durch. In der korrelativen Studie mit Schulkindern wurde Prosozialität über Fragebögen in vier Formen erhoben: Hilfeverhalten, kooperatives Verhalten, Empathie und emotionales Bewusstsein (engl. „emotional awareness“). Zusätzlich wurden aggressive Kognitionen in Form von „Akzeptanz von Aggression“ und „Hostile Attribution Bias“ gemessen. Die Ergebnisse lassen sich so zusammenfassen, dass der Konsum prosozialer Spiele die Prosozialität erhöht und aggressive Kognitionen verringert, während der Konsum von Gewaltspielen die Prosozialität senkt und aggressive Kognitionen erhöht. In der Längsschnittstudie sollten die Probanden die Häufigkeit der verschiedenen von ihnen in den letzten Monaten ausgeübten prosozialem Verhaltensweisen einschätzen. Dies galt als Maß für ihr prosoziales Verhal-

ten. In einem Strukturgleichungsmodell konnte dieses prosoziale Verhalten zu t_2 durch den Konsum prosozialer Spiele zu t_1 vorhergesagt werden ($\beta = .069, p < .001$). Ebenso konnte der Konsum prosozialer Spiele zu t_2 durch das prosoziale Verhalten zu t_1 vorhergesagt werden ($\beta = .074, p < .001$). Die Autoren bezeichnen dieses Muster als Engelskreis (engl. „upward spiral“). Die Effekte sind zwar klein, aber die Autoren führen das darauf zurück, dass zwischen den Messungen nur 3 bis 4 Monate lagen. In der Experimentalstudie spielten die Probanden ein prosoziales, ein neutrales oder ein gewalthaltiges Computerspiel. Im Anschluss sollten sie für einen anderen Probanden 11 aus 30 verschiedenen Tangram-Puzzles auswählen. Es gab 10 leichte, 10 mittelschwere und 10 schwere Puzzles. Ihnen wurde gesagt, dass der andere Proband 10 Dollar gewinnen würde, wenn er es schaffen würde, 10 Puzzles innerhalb von 10 Minuten zu lösen. Die Anzahl ausgewählter leichter Puzzles galt als Maß für Hilfeverhalten, die Anzahl ausgewählter schwerer Puzzles als Maß für aggressives Verhalten. Über alle Gruppen hinweg wurden mehr leichte als schwere Puzzles ausgewählt ($d = 0.49, p < .005$). Die Probanden in der Bedingung mit dem prosozialem Spiel wählten mehr leichte Puzzles als die Probanden in den anderen beiden Bedingungen ($d = 0.48, p < .005$), während die Probanden in der Gewaltspielbedingung mehr schwere Puzzles als die Probanden in den anderen Bedingungen wählten ($d = 0.46, p < .005$).

Wenn man die Ergebnisse der oben genannten Studien zusammennimmt, dann lässt sich schlussfolgern, dass das Spielen von Gewaltspielen die Wahrscheinlichkeit des Auftretens von prosozialem Verhalten verringert und die von aggressiven Kognitionen erhöht, während es beim Spielen prosozialer Spiele genau umgekehrt ist. Hierbei gibt es aber die Einschränkung, dass das für die Untersuchungen gewählte prosoziale Verhalten so gut wie nie mit hohen Kosten verbunden ist. Für die hier gewählte Operationalisierung „Eckenzählen“ werden dagegen mittelhohe Kosten vermutet, da das Eckenzählen im Vergleich zum Bleistifteaufheben mit einem größeren Zeitbedarf sowie einer leichten kognitiven Anstrengung einhergeht. Wird eine der aufgeführten Effektstärken für diese Studie zur Bewertung der praktischen Bedeutsamkeit herangezogen, dann dürfte diese den Zusammenhang zwischen Gewaltspielen und „mittelteurem“ Hilfeverhalten um einen nicht bekannten Betrag überschätzen. Leider bieten sich keine besseren Alternativen an. Aus diesem Grund wird der Wert aus der Meta-Analyse von Anderson et al. (2010) trotzdem für die Bewertung der praktischen Bedeutsamkeit herangezogen. Bei experimentellen Studien zur Wirkung von Gewaltspielen auf prosoziales Verhalten fanden die Autoren eine Effektgröße von $r = -.161$. Wird der Betrag der Effektgröße in Hedges g umgerechnet, so ergibt sich der Wert $g = 0.323$.

4.6 OPERATIONALISIERUNGEN DES HILFEVERHALTENS

4.6.1 VERFAHREN

Die Variable Hilfeverhalten lässt sich in Experimenten auf verschiedene Arten operationalisieren (eine Auflistung findet sich bei Eagly & Crowley, 1986, sowie Anderson & Bushman,

Virtuelle Nothilfe

2001). Übliche Varianten sind: Jemandem (Geld oder Bonbons) spenden, jemandem bei einer Aufgabenlösung helfen, und einem Gewaltopfer helfen (Anderson et al., 2010). Die Messung des Spendeverhaltens (z. B. Chambers & Ascione, 1987) stellt aber keine optimale Erhebung von Hilfeverhalten dar, da die Motive beim Spendeverhalten stark von der Situationseinschätzung abhängen. Die Versuchsperson könnte z. B. der Ansicht sein, dass der Spendenempfänger eine Bedürftigkeit nur vortäuscht, dass die Spende nicht zum gewünschten Ziel führt, dass die Ziele des Spendenempfängers nicht lohnenswert sind etc. Außerdem scheinen Spendeverhalten und Hilfeverhalten zwei voneinander unabhängige Verhaltensweisen zu sein (siehe Abschnitt 4.5.6).

In einigen Fällen wird das sogenannte „Gefangenendilemma“ genutzt, bei dem die Versuchsperson vor die Wahl gestellt wird, einen anderen Gefangenen zu verraten oder zu schweigen. Mit dem Gefangenendilemma wird aber eigentlich Kooperations- bzw. Wettbewerbsverhalten erhoben (Bernstein et al., 1987). Bei Gentile et al. (2009) findet sich ein Messverfahren, dass eine Art Mischung zwischen Kooperations- und Spendeverhalten erhebt. Ein Proband kann etwas Gutes tun, indem er für den „anderen“ Probanden möglichst leichte Puzzles auswählt, damit dieser eine höhere Chance auf einen Geldgewinn hat (siehe Abschnitt 4.5.6). Der Proband erhält aber keine Hinweise darauf, ob der „andere“ Proband in Not ist und auf den Geldgewinn in irgendeiner Form angewiesen ist. Daher könnten hinter diesem Verhalten auch andere Motive als Hilfsbereitschaft stecken. Angesichts der mit dem Hilfeverhalten verbundenen Kosten handelt es sich hier außerdem eher um eine alltägliche Gefälligkeit als um Hilfe in einer Notsituation. Eine weitere Variante einer solchen Gefälligkeit findet sich bei Greitemeyer und Osswald (2010, siehe Abschnitt 4.5.6): In Studie 1 wurde erhoben, wie viele Probanden vom Versuchsleiter „versehentlich“ verschüttete Bleistifte aufheben (engl. „spilled pencils“), wobei auch die Menge der Bleistifte gemessen wurde.

Bei Studien zu Hilfeverhalten in Notsituationen wird in der Regel eine solche vorgetäuscht, z. B. ein verstauchter Fuß, ein liegengebliebenes Auto, ein verlorener Brief etc. Oft wird nur erhoben, ob die Versuchsperson überhaupt helfend eingreift. Dies führt allerdings zu einer Varianzeinschränkung. Daher wird manchmal auch die Zeit gemessen, die zwischen einem Hilferuf und der Reaktion der Versuchsperson vergeht. Weitere Varianten metrischer Messungen finden sich bei Greitemeyer und Osswald (2010): In Studie 2 wurden die Probanden gefragt, ob bzw. wie viele Minuten sie bereit wären, direkt im Anschluss an die Untersuchung für einen hilfebedürftigen Diplomanden an einer weiteren Untersuchung teilzunehmen (engl. „willingness to assist“). Hier ist eine leichte Notlage des Diplomanden deutlich erkennbar. Die eindeutigste Notlage, deren Beseitigung mit potenziell hohen Kosten verbunden ist, findet sich aber in Studie 3, wo der Proband die Versuchsleiterin vor Zudringlichkeiten durch ihren „Ex-Freund“ bewahren musste.

4.6.2 VALIDITÄTSPROBLEME

Alle genannten Operationalisierungen haben gemeinsam, dass die Motive der Probanden nicht erfasst werden. Dadurch entstehen zwei Probleme. Zum einen kann bei dem beobachteten Verhalten nicht in jedem Fall zweifelsfrei auf ein Hilfemotiv geschlossen werden. Dass dieses Problem nicht nur rein theoretischer Natur ist, zeigte sich bei Greitemeyer und Osswald (2010) in Studie 3. Hier verfolgten einige Probanden weniger das Ziel, der Versuchsleiterin zu helfen. Stattdessen ging es ihnen hauptsächlich darum, den Angreifer verletzen. Somit handelt es sich eher um eine Form gewaltsamer Retribution bzw. ausgleichender Gerechtigkeit. Zum anderen kann nicht zwischen egoistischer und altruistischer Hilfe unterschieden werden. Damit lässt sich nicht sagen, wie stark das Verhalten von moralischen Kognitionen und Emotionen motiviert ist.

4.6.3 BEWERTUNG DER VERFAHREN UND AUSWAHL

Zur Untersuchung von Nothilfe wäre eine Operationalisierung ideal, die eine ernste Notlage simuliert. Hier bietet sich am ehesten die „Rettung der Versuchsleiterin“ an. Allerdings kommt diese Operationalisierung nicht infrage, da sie eher aggressives Verhalten statt helfendes Verhalten misst. Außerdem ist sie ethisch nur schwer zu vertreten. Die zweitbeste Operationalisierung ist „willingness to assist“. Hier ist die Notlage einer Person erkennbar, auch wenn es sich nicht um eine ernste Notlage handelt. Ein weiterer Vorteil¹⁴ ist, dass das Verfahren wahrscheinlich eher altruistische als egoistische Hilfe erfasst: Der Nutzen für den Probanden ist gering, da das von ihm erwartete Verhalten nur wenig Unterhaltungswert haben dürfte und wegen einer fehlenden Leistungsrückmeldung auch keine Selbsterkenntnismöglichkeiten bietet. Neben der fehlenden Schwere der Notlage besteht der größte Nachteil des Verfahrens darin, dass die Motive nicht erhoben werden. Diese lassen sich auch nicht sinnvoll erheben. Zum einen würde eine Erhebung das notwendige Täuschungsmanöver sabotieren, da kein sinnvoller Grund für die Erhebung der Motive angegeben werden kann. Zum anderen besteht die Gefahr, dass die Probanden in Rechtfertigungsdruck geraten und daher falsche Gründe nennen. Beides ließe sich nur durch ein zweistufiges Vorgehen wie bei Groeben (1986) lösen (siehe Abschnitt 4.4.2.1), aber dafür wären eigene Validierungsstudien notwendig, die hier nicht geleistet werden können. Ein weiterer Nachteil des Verfahrens liegt in den eher geringen Kosten, da der Proband lediglich Zeit und Konzentration aufwenden muss, aber nicht mit einer Schädigung seiner Person rechnen muss. Die genannten Nachteile treffen aber auf alle oben aufgeführten Operationalisierungen zu. Da „willingness to assist“ somit das beste Vorteil-Nachteil-Verhältnis aufweist, wurde es zur Messung von Hilfeverhalten ausgewählt.

5 THEORETISCH-INHALTLICHE HYPOTHESEN

Aus der Fragestellung, ob Nothilfesituationen im Computerspiel eher die Gewaltbereitschaft, eher die Hilfsbereitschaft oder beides gleichzeitig fördern, wurden drei theoretisch-inhaltliche Hypothesen (TIH) zum Gewaltverhalten und drei weitere zum Hilfeverhalten abgeleitet.

5.1 GEWALTVERHALTEN

Wie bei der Theoriendiskussion erwähnt, soll dem Social Interactionist Model als allgemeinem Rahmenmodell der Vorzug gegeben werden. Aus diesem Modell Vorhersagen zu beobachtbarem Verhalten abzuleiten ist aber schwierig, da es primär auf nicht beobachtbare Motive und Intentionen als Verhaltensmotivatoren abhebt. Dafür spielt es eine wichtige Rolle bei der Operationalisierung des Gewaltverhaltens, da es klarstellt, dass es sich ohne Schädigungsabsicht per Definition nicht um Gewaltverhalten handeln kann und die Schädigungsabsicht bei der Verhaltensbeobachtung sichergestellt werden muss.

Da hier insbesondere die Effekte der virtuellen Ausübung und weniger der Beobachtung von Gewalt interessieren, wird vom Catharsis Model nur die aktionspsychologische Variante (und nicht die rezeptionspsychologischen) berücksichtigt. Aus dieser sowie aus dem General Aggression Model und dem Moral Management Model lässt sich ableiten, dass die (virtuelle) Ausübung von Gewalt im Spiel zu einem Gewaltlernen führt, welches die Auftretenswahrscheinlichkeit von Gewalt im realen Leben erhöht. Die empirischen Befunde zur Purgation sind aber nicht ganz einschlägig, da sie im Gegensatz zu den Befunden zum General Aggression Model und zum Moral Management Model nicht auf Studien zu Gewaltspielen basieren. Allerdings existieren zum Moral Management Model noch nicht sehr viele Untersuchungen. Dagegen wird das General Aggression Model insbesondere von der Meta-Analyse von Anderson et al. gestützt, in welche 140 Studien zum aggressiven Verhalten im Kontext von Gewaltspielen eingeflossen sind. Die meisten dieser Studien kommen zu dem Ergebnis, dass der Konsum von Gewaltspielen zu einem erhöhten Gewaltverhalten führt. Die erste Hypothese stellt eine Replikation dieser Befunde dar:

TIH₁: Reines Gewaltverhalten (Töten) im Computerspiel erhöht das Gewaltverhalten im Vergleich zu gewaltfreiem Verhalten im Computerspiel.

Die Auswirkung von gewalttätigem Hilfeverhalten in Computerspielen wurde bisher nicht erforscht. Da es Gewalt beinhaltet, sollte es aber ebenfalls das Gewaltverhalten erhöhen:

TIH₂: Gewalttätiges Hilfeverhalten (Nothilfe) im Computerspiel erhöht das Gewaltverhalten im Vergleich zu gewaltfreiem Verhalten im Computerspiel.

Virtuelle Nothilfe

Eine weitere Frage ergibt sich, wenn Gewalt mit Hilfe und Gewalt ohne Hilfe miteinander verglichen werden. Erhöht gewalttätiges Hilfeverhalten im Vergleich zu reinem Gewaltverhalten die im realen Leben ausgeübte Gewalt stärker, oder ist es genau umgekehrt? Aus dem General Aggression Model lässt sich dazu keine Vorhersage ableiten. Im Rahmen des Moral Management Model sollte es sich beim gewalttätigen Hilfeverhalten um eine moralisch gerechtfertigte Form von Gewalt handeln, wonach das Gewaltverhalten hier am stärksten ausfallen sollte. Die Forschung zu Zivilcourage hingegen zeigt, dass Gewalt beim helfenden Eingreifen möglichst vermieden wird. Die Intention des Eingreifenden liegt hier in der Ausübung von Hilfe; Gewalt wird als letztes Mittel zur Erreichung dieses Ziels betrachtet. Daher sollte es in gewalttätigen Hilfesituationen zu weniger Gewalt kommen als in reinen Gewaltsituationen. Da sich die Theorien in ihrer Vorhersage widersprechen, muss eine Entscheidung getroffen werden, welcher Theorie bei der Hypothesenableitung der Vorzug zu geben ist. Am sinnvollsten wäre es, die besser bewährte Theorie heranzuziehen, aber beide Theorien sind empirisch noch wenig geprüft. Ein größerer Unterschied zwischen den Theorien besteht darin, dass die hier verwendete Zivilcourage-Theorie von einem Subjektmodell ausgeht, bei dem sich die subjektive Intention der Handelnden mit der objektiven Motivation deckt. Anders gesagt unterstellt die Theorie, dass die von den Handelnden benannten Gründe den tatsächlichen Ursachen entsprechen. Das Moral Management Model hingegen unterstellt, dass die von den Handelnden benannten Gründe nicht den tatsächlichen Ursachen entsprechen, sondern Abwehrstrategien für ihr unmoralisches Verhalten darstellen. Nach Erb (1997) sollte man aus ethischen Gründen „zunächst mit einer handlungstheoretischen Modellierung beginnen, bei der die Deckungsgleichheit von subjektiver Intention (aus der Ersten-Person-Perspektive) und ‚objektiver‘ Motivation (aus der Dritten-Person-Perspektive) impliziert ist, und nur, wenn für diese Modellierung keine empirische Geltung nachgewiesen werden kann, auf tuns- oder verhaltenstheoretische Ansätze zurückgreifen“ (S. 216-220; siehe dazu auch Mohseni, 2007, S. 5-7; Groeben, 1999, S. 367-388). Da die empirische Geltung für beide Theorien wie erwähnt bisher nur wenig überprüft wurde, wurde der Forschung zu Zivilcourage Vorrang bei der Hypothesenableitung eingeräumt.

Im Computerspiel wird dem Spieler aber meistens keine Wahl gelassen, ob er Gewalt ausübt oder nicht, da sich die Hilfeleistung in der Regel nur durch Gewaltanwendung realisieren lässt. Es lässt sich aber vermuten, dass selbst im Anschluss an ein (quasi vom Spiel erzwungenes) gewaltsames Eingreifen die Intention weiterhin in der Hilfeleistung besteht. Somit sollte es im Anschluss eher zu Hilfe- als zu Gewaltverhalten kommen. Daraus folgt, dass es nach gewalttätigen Hilfesituationen im Vergleich zu reinen Gewaltsituationen seltener zu Gewaltverhalten kommen sollte. Auf der Ebene von Gruppenvergleichen lässt sich dies auch wie folgt formulieren:

TIH₃: Reines Gewaltverhalten (Töten) im Computerspiel erhöht das Gewaltverhalten stärker als gewalttätiges Hilfeverhalten (Nothilfe) im Computerspiel.

5.2 HILFEVERHALTEN

Bei der Ableitung der Hypothesen zum Hilfeverhalten besteht dasselbe Problem wie bei der Ableitung der Hypothesen zum Gewaltverhalten. Eigentlich sollte dem handlungstheoretischen Motivationsmodell für Hilfeleistung als allgemeinem Rahmenmodell der Vorzug gegeben werden, aber auch hier ist es schwierig, aus dem Modell konkrete Vorhersagen zu beobachtbarem Verhalten abzuleiten, da es primär auf nicht beobachtbare Motive und Intentionen als Verhaltensmotivatoren abhebt. Dafür stellt es klar, dass ohne Kenntnis der Intentionen des Handelnden nicht entschieden werden kann, ob es sich um Hilfeverhalten handelt bzw. ob das Hilfeverhalten eher altruistischer oder eher egoistischer Natur ist.

Hinzu kommt nun, dass sich die allgemeineren Theorien zum Hilfeverhalten auf eine Vielzahl empirischer Studien stützen können, während es zum computerspielspezifischeren General Learning Model nur wenige Studien gibt. Aus letzterem lässt sich ableiten, dass reines Hilfeverhalten im Computerspiel das Hilfeverhalten im realen Leben erhöht. Dieser Aussage stehen die empirisch besser geprüften Theorien zum Hilfeverhalten aber nicht entgegen. Die vierte Hypothese stellt daher eine Replikation der Befunde zum General Learning Model dar:

TIH₄: Reines Hilfeverhalten (Hilfe) im Computerspiel erhöht das Hilfeverhalten im Vergleich zu hilfefreiem Verhalten im Computerspiel.

Die Auswirkung von gewalttätigem Hilfeverhalten im Computerspiel wurde bisher nicht erforscht. Da gewalttätiges Hilfeverhalten aber Hilfe beinhaltet, sollte es ebenfalls das Hilfeverhalten im Leben erhöhen:

TIH₅: Gewalttätiges Hilfeverhalten (Nothilfe) im Computerspiel erhöht das Hilfeverhalten im Vergleich zu hilfefreiem Verhalten im Computerspiel.

An dieser Stelle lässt sich wieder die Frage stellen, wie es im Vergleich der Bedingungen aussieht: Erhöht gewalttätiges Hilfeverhalten im Vergleich zu reinem Hilfeverhalten die im realen Leben ausgeübte Hilfe stärker, oder ist es genau umgekehrt? Aus dem General Learning Model lässt sich keine Vorhersage ableiten, möglicherweise aber aus der Zivilcourage-theorie von Scheele und Kapp (2002). Nach dieser Theorie wird Gewalt beim helfenden Eingreifen soweit wie irgend möglich vermieden. Passend dazu zeigt sich auch in der Meta-Analyse von Anderson et al. (2001), dass Menschen eher dazu bereit sind, Gewalt gegen unbelebte Dinge auszuüben ($r^+ = .41$), als gegen andere Menschen ($r^+ = .14$, $\chi^2[1] = 4.80$, $p < .05$). Anscheinend sind Menschen stark gehemmt, Anderen Gewalt anzutun. Wenn Hilfe nur noch mit Gewalt möglich ist, könnte dies dazu führen, dass ein helfendes Eingreifen vollständig ausbleibt. Im Computerspiel wird dem Spieler aber in der Regel keine Wahl gelassen, ob er Hilfe ausübt oder nicht, da sich das Spiel nur dann weiterspielen lässt, wenn die vorgegebenen

Virtuelle Nothilfe

Aufgaben erledigt werden. Es lässt sich vermuten, dass im Anschluss an ein (quasi vom Spiel erzwungenes) gewaltsames Eingreifen beim Helfen weiterhin eine Intention zur Vermeidung von Gewalt besteht. Daher sollte es auch im Anschluss an ein gewaltsames Eingreifen eher zu Hilfe- als zu Gewaltverhalten kommen. Daraus folgt, dass es nach gewalttätigen Hilfesituationen im Vergleich zu reinen Gewaltsituationen seltener zu Gewaltverhalten kommen sollte. Auf der Ebene von Gruppenvergleichen lässt sich das auch wie folgt formulieren:

TIH₆: Reines Hilfeverhalten (Hilfe) im Computerspiel erhöht das Hilfeverhalten stärker als gewalttätiges Hilfeverhalten (Nothilfe) im Computerspiel.

Alle Hypothesen sollten im Rahmen eines Experiments geprüft werden. Die erste unabhängige Variable (UV) bestand im Gewaltverhalten und die zweite UV im Hilfeverhalten, das der Spieler aufbringen muss, um im Spiel weiterzukommen („Töten“ vs. „Kein Töten“ im Fall des Gewaltverhaltens bzw. „Helfen“ vs. „Kein Helfen“ im Fall des Hilfeverhaltens). Die erste abhängige Variable (AV) war das aggressive Verhalten nach dem Spiel, die zweite das Hilfeverhalten nach dem Spiel. Unabhängig davon gab es mehrere Störvariablen, die der Kontrolle bedurften (siehe Abschnitt 7.3).

Um die Hypothesen experimentell zu prüfen, wurden in einem Computerspiel vier Bedingungen hergestellt (siehe Abb. 11). In der „Nothilfe“-Bedingung musste eine hilfebedürftige Person gerettet werden, indem alle Gegner getötet wurden. In der „Hilfe“-Bedingung musste ebenfalls eine hilfebedürftige Person gerettet werden, allerdings ohne Anwendung von Gewalt. In der „Töten“-Bedingung mussten alle Gegner getötet werden, ohne dass in irgendeiner Weise Hilfe ausgeübt wurde. In der „Schatzjagd“-Bedingung musste ein Schatz gestohlen werden, ohne dass in irgendeiner Weise Hilfe oder Gewalt ausgeübt wurde. Somit ergab sich folgender vollständig gekreuzter Versuchsplan:

		Faktor A	
		Gewaltverhalten im Spiel	
		A ₁ Töten	A ₂ Kein Töten
Faktor B	B ₁ Helfen	Nothilfe (Helfen durch Töten)	Hilfe (Helfen ohne Töten)
	B ₂ Kein Helfen	Töten (Töten ohne Helfen)	Schatzjagd (Ohne Helfen und ohne Töten)

Abb. 11: Versuchsplananlage

Im Rahmen dieses Designs lassen sich die sechs Hypothesen wie folgt veranschaulichen (siehe Abb. 12 bis Abb. 17). Die „größer als“-Zeichen sollen symbolisieren, dass bei Gültigkeit der Hypothese der Gruppenmittelwert in dieser Zelle größer sein muss als in der anderen.

Um eine Konfundierung mit unbekannten Störvariablen zu vermeiden, wurden die vier Spielaufgaben so ähnlich wie möglich gehalten. Aufgrund dieser Ähnlichkeit erschien es wenig sinnvoll, eine Versuchsperson alle vier Varianten absolvieren zu lassen. Daher wurden die Versuchsbedingungen nicht intra-, sondern interindividuell variiert.

Virtuelle Nothilfe

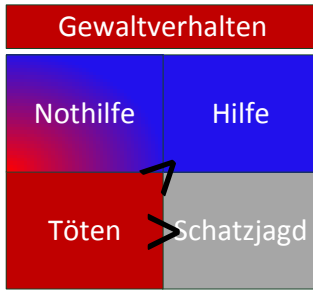


Abb. 12: TIH₁

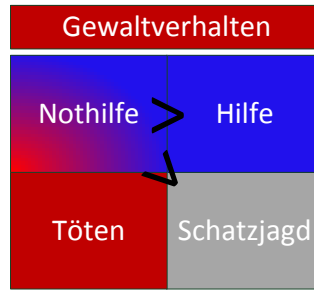


Abb. 13: TIH₂

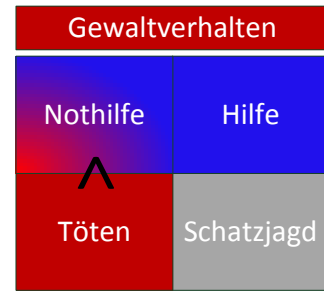


Abb. 14: TIH₃

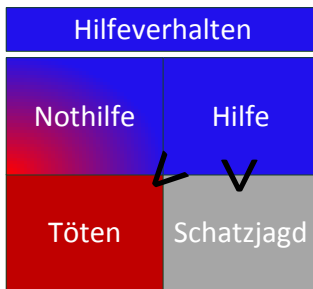


Abb. 15: TIH₄

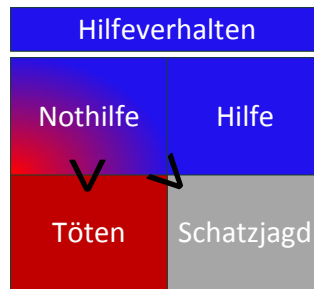


Abb. 16: TIH₅

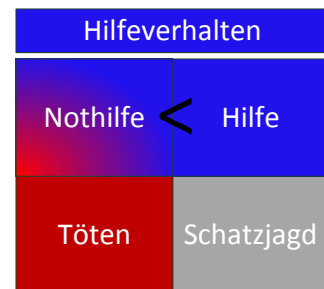


Abb. 17: TIH₆

Um die Orthogonalität der Effekte zu gewährleisten, wurde in allen vier Bedingungen dieselbe Anzahl Versuchspersonen untersucht. Mithilfe des Programms G*Power 3.0.1 von Faul, Erdfelder, Lang und Buchner (2007) wurde a priori die optimale Stichprobengröße berechnet. Für die Auswertung mittels Varianzanalyse beträgt sie $N = 180$ (F -Test, $f = .25$; $\alpha = .05$; $\beta = .20$, $df = 3$, $k = 4$). Für die Berechnung wurde mit $f = .25$ eine mittlere Effektstärke angenommen, da der Effekt für das Gewaltverhalten aus der Meta-Analyse von Anderson et al. (2010) mit $r = .19$ einen kleinen bis mittleren Effekt vermuten ließ. Da das biologische Geschlecht eine relevante Störvariable darstellt, sollte in jeder Bedingung das gleiche Verhältnis von Männern zu Frauen vorliegen. Daher sollten in jeder der 4 Gruppen jeweils 23 Männer und 23 Frauen untersucht werden, was die Stichprobengröße geringfügig auf $N = ((23 + 23) \times 4) = 184$ erhöhte.

7 OPERATIONALISIERUNG

7.1 UNABHÄNGIGE VARIABLEN

7.1.1 WAHL DES COMPUTERSPIELS

Die UVen „Hilfe im Spiel“ und „Gewalt im Spiel“ sollten als Aufgaben operationalisiert werden, die in einem PC-Spiel zu lösen sind. Nach Hartig, Frey und Ketzel (2003) sind die meisten üblichen Computerspiele für solche Zwecke ungeeignet, da sie keine Möglichkeit der Datenaufzeichnung bieten. Außerdem ist die experimentelle Variation der Inhalte selten möglich. Die übliche Vorgehensweise, für jede Versuchsbedingung ein eigenes Spiel zu verwenden, würde aber zur Konfundierung einer Bedingung mit den übrigen Merkmalen des verwendeten Spiels führen. Bei einem solchen Vorgehen befürchten Klimmt und Trepte (2003), dass die gewaltfreien Spiele im Vergleich zu den gewalthaltigen weniger aufregend gestaltet sein könnten und daher weniger unspezifische Erregung auslösen würden, was bei Gültigkeit von Zillmanns Erregungstransfertheorie zu einem anschließenden Unterschied im Gewaltverhalten führen würde. Besser ist es, alle Versuchsbedingungen in einem einzigen Spiel zu realisieren, weil so alle Spielmerkmale in den Bedingungen identisch sind (Hartig et al., 2003). Es musste daher auf eines der wenigen Spiele mit integriertem Editor zurückgegriffen werden, der die Möglichkeit der experimentellen Variation bietet (zum Editor siehe Bethesda Softworks, 2006b). Um die emotionale Beteiligung der Spieler zu erhöhen, sollte es sich außerdem um ein Action-Rollenspiel in der Ego-Perspektive handeln. Action-Rollenspiele sind Filmen ähnlich, nur dass der Spieler hier die Hauptrolle aktiv einnimmt und es mehr Gewaltszenen pro Zeiteinheit gibt. Die Ego-Perspektive sollte die Identifikation mit der Spielfigur verstärken. Eine größere Identifikation führt dazu, dass man sich weniger mit den Protagonisten des Spiels über deren Leistungen freut (Sozio-Emotionen), sondern mehr über die eigene Leistung (Ego-Emotionen), was vermutlich zu einer besonders intensiven emotionalen Wirkung führt (Klimmt, 2001).

Die Auswahlkriterien trafen nur auf das Spiel „The Elder Scrolls: Morrowind“ und dessen Nachfolger „The Elder Scrolls: Oblivion“ (Bethesda Softworks, 2006a) zu. Der Hersteller ist nach Klimmt (2006b) bekannt für audiovisuelle Darstellungsqualität, welche wichtig für ein möglichst starkes Präsenzgefühl ist (zum Begriff „Präsenzgefühl“ siehe Abschnitt 4.2.5). Im neueren Spiel „Oblivion“ ist die Grafik besser, sodass das Präsenzgefühl hier im Vergleich höher ausfallen dürfte. Dazu passt die Ansicht von Klimmt (2006b), Oblivion biete „außergewöhnlich intensives Unterhaltungserleben“ und sei damit „eine Fundgrube und ein ideales Testumfeld für multidimensionale Theorien interaktiver Unterhaltung“. Daher wurde Oblivion für die Durchführung des Versuchs ausgewählt.

In Oblivion befindet sich der Spieler in einer mittelalterlichen Fantasywelt, welche er aus der Ego-Perspektive betrachtet. Die Spielfigur wird mit der Tastatur bewegt. Mit der Maus wird

Virtuelle Nothilfe

die Blickrichtung der Spielfigur gesteuert. Diese Art der Darstellung ist für Anfänger allerdings nicht ganz unproblematisch, weil das Auseinanderfallen von Bewegungsrichtung und Blickrichtung zu Orientierungslosigkeit und sogar Übelkeit (sog. „cybersickness“) führen kann (siehe z. B. Hartig et al., 2003). Die Maustasten werden für Kämpfe mit anderen Spielfiguren benötigt (linke Taste: Angriff mit der Waffe; rechte Taste: Blocken). Mit anderen Spielfiguren kann aber nicht nur gekämpft, sondern auch kommuniziert werden. Dabei wählt der Spieler eine von mehreren Äußerungen aus einer vorgegebenen Liste aus. Je nach Wahl der Äußerung „antwortet“ die andere Spielfigur. Die Antworten sind dabei vertont, während die Sätze des Spielers stumm bleiben.

7.1.2 SPIELAUFGABEN

Zu Beginn des Spiels mussten alle Probanden – auch die, die das Spiel bereits kannten – über ein Tutorial im Spiel die Steuerung einüben. Das Tutorial war in allen Versuchsbedingungen identisch, mit Ausnahme des Kampftrainings, welches in den Bedingungen ohne Gewalt fehlte. Im Kampftraining mussten die Probanden erst gegen eine Skelettpuppe und danach gegen eine Riesenkrabbe antreten (siehe Abb. 18 und Abb. 19).



Abb. 18: Kampftraining gegen Skelettpuppe



Abb. 19: Kampftraining gegen Riesenkrabbe

Im Anschluss an das Tutorial erhielten die Probanden im Spiel von der Spielfigur „Auftraggeberin“ eine von vier verschiedenen Aufgaben (siehe Abb. 20 und Abb. 21). Es wurde mit Absicht eine weibliche Figur gewählt, da Männer aufgrund der Helden- und der Ritterlichkeitsnorm¹⁵ eher Frauen als Männern in Not helfen (Eagly & Crowley, 1986; Alfermann, 1996, S. 129-133). Über die Aufgaben wurden die vier Versuchsbedingungen realisiert. Dabei wurde darauf geachtet, dass die hilfehaltigen Aufgaben einer ritterlichen Heldentat möglichst nahe kommen.

Bei der „Nothilfe“-Aufgabe wurde den Probanden von der Auftraggeberin mitgeteilt, sie sei an einer tödlichen Grippe erkrankt (siehe Abb. 20). Banditen hätten den heilenden Trank gestohlen. Der Trank befände sich im Zentrum des Banditenlagers. Der Weg dorthin sei durch Türen versperrt. Die Banditen würden die Schlüssel zur nächsten Tür bei sich tragen. Um die jeweils nächste Tür zu öffnen, müsse der Proband daher den jeweiligen Banditen besiegen

Virtuelle Nothilfe

(zu den Banditen siehe Abb. 22 bis Abb. 27). Es sei zwecklos, den Banditen die Schlüssel zu stehlen. Die „Hilfe“-Aufgabe unterschied sich dadurch, dass die Auftraggeberin hier den Probanden mitteilte, die Schlüssel würden sich in Truhen befinden. Der Proband solle sich an den Banditen vorbeischieben. Es sei zwecklos, die Banditen anzugreifen, da diese nicht besiegt werden können. Um dies sicherzustellen, wurde die Spielfigur des Probanden „ohnmächtig“, sobald entweder ein Bandit die Spielfigur oder die Spielfigur einen Banditen mit der Waffe schlug.



Abb. 20: Spielfigur Auftraggeberin (in den Bedingungen Hilfe und Nothilfe)



Abb. 21: Spielfigur Auftraggeberin (in den Bedingungen Töten und Schatzjagd)

Die „Töten“-Aufgabe unterschied sich von der „Nothilfe“-Aufgabe darin, dass die Auftraggeberin nicht tödlich erkrankt war (siehe Abb. 21). Stattdessen teilte sie den Probanden mit, dass Banditen einen Trank geraubt hätten, der viel Geld wert sei. Für hundert Goldstücke würde sie dem Probanden verraten, wo der Trank zu finden sei. In der „Schatzjagd“-Aufgabe teilte die Auftraggeberin den Probanden ebenfalls mit, dass Banditen einen wertvollen Trank geraubt hätten. Allerdings solle der Proband sich an den Banditen vorbeischieben, da es zwecklos sei, die Banditen anzugreifen. Die Aufgaben wurden in jeder anderen Hinsicht so ähnlich wie möglich gehalten, um die Konfundierung einer Versuchsbedingung mit Störvariablen zu vermeiden. Dazu gehörte auch, dass alle Aufgaben in der gleichen Zeit lösbar sein sollten. Dabei wurde eine Zeit von ungefähr 20 Minuten (siehe Abschnitt 7.3.8) anvisiert. Da die Lösungszeit aber nicht nur von der Aufgabenschwierigkeit, sondern auch in stärkerem Maß von den Fähigkeiten eines Probanden abhing, war es nur schwer möglich, eine in allen Bedingungen absolut identische Lösungszeit sicherzustellen (siehe hierzu auch Abschnitt 11.3.2).

Im zweiten Vortest wurden die Spielaufgaben einer ersten Prüfung unterzogen (siehe Abschnitt 9.1). Dabei zeigte sich, dass einige Aspekte der Spielaufgabe suboptimal ausfielen. Die Spielfigur in der Hilfebedingung wurde häufiger ohnmächtig als in der Gewalt- und der Nothilfebedingung, weswegen die Schwierigkeit der gewaltfreien Bedingung gesenkt und die der gewalthaltigen erhöht wurde. Die Auftraggeberin wirkte in der Hilfebedingung nicht sehr hilfsbedürftig, weswegen der Eindruck der Hilfsbedürftigkeit durch eine kränklich wirkende

Virtuelle Nothilfe

Erscheinung inkl. Hustengeräuschen gesteigert wurde. Im Tutorial gab es (immer noch) zu viele Erklärungen, die weggeklickt werden müssen, deren Anzahl daher reduziert wurde.



Abb. 22: Spielfigur „Ben Bandit“



Abb. 23: Spielfigur „Bud Bandit“



Abb. 24: Spielfigur „Bill Bandit“



Abb. 25: Spielfigur „Brian Bandit“



Abb. 26: Spielfigur „Borg Bandit“



Abb. 27: Spielfigur „Boss Bandit“

Im dritten Vortest wurden die Spielaufgaben einer weiteren Prüfung unterzogen. Dabei wurde zum einen kontrolliert, ob die Probanden bei der jeweiligen Aufgabe zustimmen, dass für deren Lösung Hilfe- und/oder Gewaltverhalten erforderlich war bzw. dass der Inhalt des Spiels hilfe- und/oder gewalthaltig war (Manipulation Check). Zum anderen wurde überprüft, ob die Aufgaben sich hinsichtlich folgender relevanter (Stör-)Eigenschaften unterscheiden: Interessantheit, Spannung, Unterhaltung, Frustration, Schwierigkeit, subjektive Aktivierung, Präsenzgefühl, Realitätsnähe der Grafik und Gewalthaltigkeit der Grafik. Mit Aus-

Virtuelle Nothilfe

nahme der subjektiven Aktivierung, die mit den zwei Items „Munterkeit“ und „Lustlosigkeit“ erhoben wurde, wurde jede Störeeigenschaft in Form eines einzelnen Items erhoben. Die Bewertung erfolgte auf einer Skala von 1 (völlig falsch) bis 5 (völlig richtig).

Es wurde angestrebt, dass die Probanden die gewalthaltigen Bedingungen hinsichtlich der Inhalte und der Aufgaben im Vergleich zu den gewaltlosen als gewalthaltiger bewerten und ebenso die hilfehaltigen Bedingungen als hilfehaltiger im Vergleich zu den hilflosen. Dieses Ziel wurde leider nicht vollständig erreicht (siehe Anhang E). Zwar wurde wie erwünscht in den hilfehaltigen Bedingungen die Hilfehaltigkeit der *Spielaufgabe* und *Spielinhalte* höher eingeschätzt als in den hilflosen. Auch wurde in den gewalthaltigen Bedingungen die Gewalthaltigkeit der *Spielaufgabe* höher eingeschätzt als in den gewaltlosen. Ein unerwünschtes Ergebnis war aber, dass die Einschätzung der Gewalthaltigkeit der *Spielinhalte* in der reinen Gewaltbedingung niedriger ausfiel als in den beiden gewaltlosen Bedingungen.

Außerdem sollten sich die Bedingungen hinsichtlich der übrigen Variablen möglichst nicht unterscheiden. Auch dies war nicht der Fall, da das Präsenzgefühl und die Interessantheit in der reinen Gewaltbedingung im Vergleich zu den übrigen drei Bedingungen niedriger eingeschätzt wurden. Dies ist verwunderlich, da Gewaltinhalte das Präsenzgefühl erhöhen sollten (Nowak, Krcmar & Farrar, 2008, zit. n. Happ, Melzer & Steffgen, 2010b, S. 26). Da die Stichprobe klein war, könnte es sich möglicherweise um einen Zufallsbefund handeln.

Im dritten Vortest zeigte sich außerdem, dass die Spielfigur in den gewaltlosen Bedingungen viel häufiger ohnmächtig wurde als in den gewalthaltigen Bedingungen. Damit einher ging die Einschätzung der Probanden, dass die gewaltlosen Bedingungen schwieriger zu spielen wären als die gewalthaltigen. Als Ursache wurde vermutet, dass die Probanden in den gewaltlosen Bedingungen die Laufwege der Banditen erlernen müssen, weil sie sich sonst nur mit Glück vorbeischieben können, während ein Sieg durch Kämpfen fast keine Lern- und Übungsprozesse erfordert. Es war zu befürchten, dass die höhere Schwierigkeit zu mehr Frustration führen könnte, was wiederum die Aggression erhöhen würde. Im dritten Vortest zeigte sich dies zwar nur tendenziell, da die gewaltlosen Bedingungen nur als etwas frustrierender bewertet wurden als die gewalthaltigen. Um sicherzugehen, wurden die gewaltlosen Bedingungen trotzdem in der Art modifiziert, dass die Spielfigur beim Schleichen nicht mehr durch einen Banditen entdeckt werden konnte, sobald der betreffende Bandit die Spielfigur bereits 2 Mal ohnmächtig geschlagen hatte. Die Spielfigur hätte dadurch pro Bandit maximal 2 Mal und damit im gesamten Spiel maximal 12 Mal ohnmächtig werden können, solange der Proband sie wie instruiert schleichen ließ.

7.1.3 OPTIMIERUNGEN

Um das Präsenzgefühl (siehe Abschnitt 4.2.5) und damit die (emotionale) Wirkung des Computerspiels zu steigern, wurden folgende Optimierungsmaßnahmen durchgeführt. Zur Erhöhung der Immersion wurde das Computerspiel auf einem ausreichend schnellen PC ausge-

Virtuelle Nothilfe

führt, damit die Bildwiedergabe ohne Ruckeln erfolgen konnte. Des Weiteren wurde das Spiel in hoher Auflösung (1680x1050) auf einem größeren Flachbildschirm (22 Zoll) präsentiert, der für die schlierenfreie Wiedergabe schneller Bildfolgen geeignet war (siehe Anhang A). Zusätzlich wurde auf verschiedene Modifikationen zur Erhöhung der Auflösung der spielinternen Texturen zurückgegriffen (siehe z. B. dev/akm, 2009), um die grafische Darstellung realistischer erscheinen zu lassen (siehe Anhang B). Die eher eintönige Hintergrundmusik wurde gegen eine stimmungsvollere Musik aus dem Film „Conan der Barbar“ (Pouledouris, 1982) ersetzt. Die letzte immersionssteigernde Maßnahme bestand in der wahrnehmungsmäßigen Abschirmung der Spieler. Dies wurde erreicht, indem (1) die Tonausgabe des Spiels (und auch des CRTT; siehe Abschnitt 7.2.1) über geschlossene Kopfhörer mit Schalldämmungseigenschaften erfolgte, um störende Außengeräusche abzuschirmen, indem (2) die Erhebung in Einzelkabinen stattfand und indem (3) beim Spielen das Licht abgeschaltet wurde. Damit Letzteres nicht zu Problemen beim Steuern der Spielfigur führt, wurde eine beleuchtete Tastatur verwendet.

Die Involviertheit (als Bestandteil des räumlichen Präsenzgefühls) sollte durch eine möglichst glaubwürdige Rahmenhandlung sowie durch die Verwendung dekorativer Objekte in der Spielwelt erhöht werden, während das „behavioral engagement“ (als Bestandteil des sozialen Präsenzgefühls) durch eine Vertonung der Äußerungen der Spielfiguren (Auftraggeberin und Banditen) inklusive passender Lippenbewegungen gesteigert werden sollte. Dabei wurde für jede Spielfigur ein/e andere/r Synchronsprecher/in eingesetzt. Da die Auftraggeberin mit dem Spieler von Angesicht zu Angesicht kommunizierte, wurde zusätzlich die in ihrem Gesicht zu erkennende Emotion an ihre verbalen Äußerungen angepasst.

Es ist sinnvoll, auch das Aufkommen von Langeweile im Spiel zu verhindern. Langeweile ist nach Csikszentmihalyi (1990, zit. n. Klimmt, 2006a, S. 51) ein negativ bewerteter Zustand kognitiver Unterforderung, oftmals begleitet von aversiven Zuständen wie Reizbarkeit und Frustration. Frustration wiederum kann aggressives Verhalten begünstigen. Um Langeweile im Spiel zu minimieren, wurden verschiedene den Unterhaltungswert erhöhende Funktionen realisiert. Jeder einzelne Bandit wurde optisch individuell gestaltet (siehe Abb. 22 bis Abb. 27). Außerdem standen die Banditen nicht einfach in der Gegend herum, sondern gingen einer Tätigkeit nach: der erste Bandit aß etwas, der zweite lief Patrouille, der dritte arbeitete in einem Labor, der vierte kämpfte gegen Krabben, der fünfte bediente Hebel und der sechste lief wieder Patrouille. Diese Maßnahme sollte auch das Kopräsenzgefühl (insbesondere die „mutual awareness“) erhöhen. Die sich bewegenden Banditen sollten außerdem den Probanden in den gewaltlosen Bedingungen eine Herausforderung bieten. Der fünfte Bandit stach dabei besonders hervor, da er die Truhe mit dem Schlüssel darin bei jeder Hebelaktivierung an eine andere Stelle teleportierte. In den gewalthaltigen Bedingungen hingegen mussten die Probanden gegen die Banditen kämpfen, wobei der dritte Bandit im Kampf durchsichtig wurde, während der sechste Bandit beim Kampf durch Flammenwolken die Sicht des Spielers behinderte. Zusätzlich zu den Banditen wurden an verschiedenen Stellen

Virtuelle Nothilfe

Fallen versteckt, die die Probanden überraschen/erschrecken sollten (siehe Abb. 28 und Abb. 29).



Abb. 28: Pfeilfalle



Abb. 29: Morgensternfalle

Damit keine unnötige Frustration aufkommen sollte, wurden die Spielabschnitte so gestaltet, dass sich die Spieler fast nicht verlaufen konnten. So gab es nur einen einzigen recht kurzen Pfad mit wenigen Abzweigungen. Das Verlies (engl. „dungeon“) bestand außerdem aus Teilabschnitten, innerhalb derer man zu einem bereits abgeschlossenen Abschnitt nicht zurückgehen konnte.



Abb. 30: Spieler während Ohnmacht



Abb. 31: Spieler nach Ohnmacht

Ein technisches Problem bestand darin, dass das Spiel automatisch mit dem Tod der Spielfigur endet. Dies hätte dazu geführt, dass für die meisten Probanden das Lösen der Spiel Aufgabe unmöglich gewesen wäre, und dies insbesondere in der gewaltfreien Bedingung, wo ein einziger Schlag eines Banditen tödlich ist (zu den Gründen dafür siehe Abschnitt 7.1.4). Daher wurde ein Mechanismus im Spiel implementiert, der bei einem Absinken der Lebensenergie unter einen kritischen Wert die Spielfigur „ohnmächtig“ werden lässt. Die Spielfigur fällt dabei auf den Boden und lässt sich in diesem Zustand nicht mehr steuern (siehe Abb. 30). Kurze Zeit später wird die Spielfigur an einen sogenannten „Checkpoint“ zurückversetzt, der sich in einem Abstand zum zuletzt bekämpften Banditen befindet. Oben im Bild wird dabei der Hinweis „Eine unbekannte Macht hat Euch in Sicherheit gebracht“ eingeblendet (siehe Abb. 31), um den Probanden einen plausiblen Grund für die Abläufe zu bieten. Zur Si-

cherheit wurde außerdem bei Erreichen eines Checkpoints automatisch der Spielstand gespeichert. Würde die Spielfigur trotz der ergriffenen Gegenmaßnahmen sterben, statt ohnmächtig zu werden, so ließe sich das Spiel am letzten Checkpoint wieder aufnehmen.

7.1.4 BLOCKADE VON SCHUMMELMÖGLICHKEITEN

Es wäre möglich gewesen, dass die Probanden in den gewaltlosen Versuchsbedingungen trotz anderslautender Instruktion gegen die Banditen kämpfen. Um dies zu unterbinden, führte ein einziger Treffer eines Banditen zur sofortigen Ohnmacht der Spielfigur des Probanden. Das Gleiche passierte, falls der Proband mit seiner Waffe einen der Banditen schlug. Durch diese Lösung bestand allerdings die Gefahr, dass die Probanden die gewaltlosen Bedingungen als schwieriger und/oder frustrierender bewerten könnten. Dies wurde mit entsprechenden Fragen nach dem Spiel geprüft (siehe Abschnitt 7.3.5).

Ein weitaus größeres Problem bestand darin, dass sich erfahrene Spieler innerhalb des Spiels einen Vorteil hätten verschaffen können, indem sie spezielle Funktionen benutzen. Beispielsweise kann man an kritischen Stellen durch Druck auf die F5-Taste den Spielstand speichern (sog. „Quick Save“) und bei Eintritt unerwünschter Folgen diesen Spielstand per F9-Taste laden („Quick Load“). Noch problematischer ist aber die Möglichkeit, im Spiel durch Drücken der ^-Taste die sogenannte „Konsole“ zu öffnen. Die Konsole dient Entwicklern zum Testen von Funktionen, indem spielverändernde Befehle eingegeben werden können. Mit ihrer Hilfe kann z. B. durch Eingabe des Befehls „Kill“ ein anvisierter Gegner sofort eliminiert werden. Das Verwenden der Konsole bzw. der Quick-Save-Funktion lässt sich leider nicht durch das Entfernen der entsprechenden Tasten lösen, da die Probanden die Funktionen im Options-Menü des Spiels jederzeit auf eine andere Taste legen können. Daher wurde Oblivion um den Oblivion Scripting Extender (kurz: OBSE) erweitert, mit dessen Befehl „Disable Control“ alle störenden Funktionen deaktiviert wurden. Dies waren neben „Quick Save“ und „Quick Load“ auch die Funktionen „Auto Move“ (der Spieler läuft automatisch vorwärts), „Toggle Point Of View“ (der Spieler wechselt von der Ego-Perspektive in eine Ansicht hinter die Spielfigur), „Block“ (der Spieler blockt einen Angriff des Gegners ab), „Cast“ (der Spieler zaubert). Die Konsole lässt sich über Disable Control leider nicht abschalten. Hierfür wurde das Plugin „Change Console Key“ eingesetzt, mit dessen Hilfe die Konsolentaste auf eine Taste gelegt wurde, die auf der Tastatur nicht vorhanden ist. Da Change Console Key über eine außerhalb des Spiels liegende Datei (sog. „ini-Datei“) konfiguriert wird, war es den Probanden nicht möglich, die Einstellung im Spiel zu ändern.

Den Probanden blieb zum Schummeln also nur noch die Möglichkeit, über das Spielmenü einen Spielstand anzulegen und diesen im Bedarfsfall zu laden. Ein so angelegter Spielstand lässt sich innerhalb des Spiels aber nicht mehr löschen und wäre damit dem Versuchsleiter beim Notieren der Spielergebnisse sowie beim Sichern der Spieldateien aufgefallen.

7.2 ABHÄNGIGE VARIABLEN

Analog zu den zwei UVen „Gewalt im Spiel“ und „Hilfe im Spiel“ gibt es zwei AVen: Gewaltverhalten nach dem Spiel und Hilfeverhalten nach dem Spiel. Um Schlussfolgerungen zum tatsächlichen Verhalten ziehen zu können, wurden die Variablen nicht in Form von Befragungen bzw. projektiven Verfahren, sondern in Form von Beobachtungen echten Verhaltens erhoben.

7.2.1 GEWALTVERHALTEN

Eine ausführliche Diskussion der verschiedenen Verfahren zur Erhebung von Gewaltverhalten findet sich in Abschnitt 4.4. An dieser Stelle wird daher nicht erneut auf die Vor- und Nachteile des gewählten Verfahrens eingegangen, sondern beschrieben, wie die Operationalisierung konkret erfolgte.

7.2.1.1 COMPETITIVE REACTION TIME TASK

Für die Erhebung des Gewaltverhaltens wurde das CRTT eingesetzt. Zur Durchführung des CRTT bietet Bushman ein spezielles Programm namens „Competitive Reaction Task Reward & Punish“ (CRT R&P)¹⁶ an (Bushman & Baumeister, 1998). Zum Einsatz kam Version 3.2.5 vom 08.03.2010. In dieser Version kann dem Gegenspieler zusätzlich zum Rauschton ein Geldbetrag „verabreicht“ werden (Bushman & Saults, 2006). Diese Funktion wurde aber nicht genutzt, weil sie anscheinend differenziell valide ist: Die Belohnung des Gegners kann bei Frauen eine Form von Beschwichtigungsverhalten sein (Hokanson, 1970, zit. n. Tedeschi & Quigley, 1996, S. 171), während bei Männern unklar bleibt, was sie mit der Belohnung des Gegners beabsichtigen.

In dem Programm werden insgesamt 25 Runden gespielt. In jeder Runde stellt die Versuchsperson vorab die Lautstärke und Dauer eines Rauschtons ein, mit dem der „Verlierer“ der Runde bestraft werden soll. Lautstärke und Dauer lassen sich in Stufen von 0 bis 10 variieren, wobei bei der Lautstärke Stufe 1 einem Ton von 50 dB(A) und Stufe 10 von 105 dB(A) entspricht, während bei der Dauer Stufe 1 einem Ton von 0.5 Sekunden und Stufe 10 von 10.0 Sekunden Länge entspricht. Um die Versuchsperson zu provozieren, gewinnt der Computer immer die erste Runde. Die anderen Runden sind so eingestellt, dass die Versuchsperson die Hälfte der Durchgänge (also 12 Runden) gewinnen kann. Die Betonung liegt auf „kann“, da der Proband nach zwei Sekunden automatisch verliert („Valid Reaction Time Range (before Automatic Loss)“, siehe Abb. 55 in Anhang J) und nach drei Sekunden der Durchgang automatisch beendet wird („Time Before Automatic Trial Termination“). Diese Maßnahmen sollen die Täuschung aufrechterhalten, dass der Gegenspieler real ist. Lautstärke und Dauer der vom Computer verabreichten Töne wurden so eingestellt, dass die Intensitäten mit moderaten Einstiegswerten beginnen, auf minimale Werte zurückfallen und dann in einem mög-

Virtuelle Nothilfe

lichst zufällig anmutenden Muster kontinuierlich bis zum Maximalwert ansteigen. Das Ansteigen der Lautstärke hat den Sinn, die Auswirkungen einer wachsender Provokation untersuchen zu können (Giancola & Chermack, 1998). Die exakten hier verwendeten Einstellwerte finden sich im Anhang J.

Um eine Schädigung des Hörorgans auszuschließen, wurde die Lautstärke der Kopfhörer mithilfe eines Schallpegel-Messgerätes kalibriert¹⁷. Bushman und Sauls (2006) empfehlen als Lautstärke auf Intensitätsstufe 10 einen Wert von 105 dB(A). Sicherheitshalber wurde die Lautstärke in diesem Experiment in Anlehnung an Ferguson et al. (2008b) auf 95 dB(A) begrenzt, was der Lautstärke in einer Diskothek entspricht. Von dieser Stufe ausgehend wurde jede Intensität um fünf Dezibel abgesenkt. Es resultierten die Kalibrierwerte aus Abb. 55, dort zu finden in der Tabelle oben links. An einem zweiten baugleichen Computer mit identischer Ausstattung wurden im CRT R&P dieselben Kalibrierwerte eingetragen. Danach wurde als Probe die Lautstärkemessung an diesem Computer wiederholt. Es resultierten identische Messwerte, sodass die gefundenen Kalibrierwerte als hinreichend genau angesehen werden können.

Da die Auswertung des CRTT nicht standardisiert ist, musste eine Entscheidung getroffen werden, welche Reaktionen im CRTT als Aggressionsmaß herangezogen werden sollen (siehe hierzu Abschnitt 4.4.1.2). In Anlehnung an Bushman und Baumeister (1998) wurde lediglich die Einstellung in der ersten Runde des CRTT ausgewertet. Aufgrund der Feststellung von Ferguson et al. (2008a), dass Sieg-Durchgänge sich nicht von Niederlage-Durchgängen unterscheiden, wurde nicht zwischen diesen differenziert. Die Autoren empfehlen außerdem, nur die Lautstärke als Maß für Aggression heranzuziehen. Dies erschien aber wenig sinnvoll, weil sich in der Studie von Giancola und Zeichner (1995a) gezeigt hatte, dass bei weiblichen Probanden nur die Dauer, nicht aber die Schockstärke als Maß für Aggression herangezogen werden kann. Daher wurde entschieden, bei *weiblichen* Probanden nur die Dauer als Maß der Aggression zu verwenden, auch wenn die genauen Gründe für den Geschlechtsunterschied noch nicht abschließend geklärt sind. Bei *männlichen* Probanden hingegen wurde nur die Lautstärke verwendet. Dadurch ergab sich das Folgeproblem, dass die Einstellung eines Tons von 0.0 Sekunden Dauer auch bei maximaler Lautstärke (bzw. eines Ton von 0 dB(A) auch bei maximaler Dauer) als nicht aggressiv bewertet werden muss, weil schlicht nichts zu hören ist. Um diesem Umstand Rechnung zu tragen, wurde bei Einstellung einer Dauer von 0.0 Sekunden die Lautstärke auf 0 dB(A) festgelegt (bzw. bei Einstellung eines Tons von 0 dB(A) die Dauer auf 0.0 Sekunden). Zu dem Computerprogramm von Bushman und Baumeister (1998) existiert eine englischsprachige Probandeninstruktion von Bushman und Sauls (2006). Diese wurde auf Deutsch übersetzt. Im dritten Vortest (siehe Abschnitt 9.1) zeigte sich, dass die Probanden Schwierigkeiten hatten, die Instruktion zu verstehen. Daher wurde die Instruktion komplett überarbeitet (siehe Anhang K).

Virtuelle Nothilfe

Leider ist die Nullwerteinstellung des Programms „CRT R&P“ von Bushman und Sauls (2006) verunglückt. Der Wert Null lässt sich nämlich nur dann einstellen, wenn der Proband zuerst einen Wert über Null einstellt und danach wieder auf Null zurückstellt. Dadurch kann der Proband leicht übersehen, dass die Einstellung von Null überhaupt möglich ist. Somit besteht die Gefahr, dass der Proband zu der Ansicht gelangt, die Einstellung von Null wäre eine Form von Teilnahmeverweigerung (Ritter & Eslea, 2005, S. 414f.; siehe auch Abschnitt 4.4.2.1). Daher wurde in der Instruktion darauf hingewiesen, wie sich der Nullwert einstellen lässt. In der Neufassung wurde außerdem versucht, die im Vortest identifizierten Probleme zu lösen: Erstens kam regelmäßig die Nachfrage, warum die Reaktionszeiten in einem Wettbewerb erhoben werden und nicht als individuelle Messung. Dies wurde nun damit begründet, dass jeder Proband in einer anderen Versuchsbedingung sei, und geprüft werden solle, nach welcher Versuchsbedingung man kürzere Reaktionszeiten habe. Zweitens hatten einige Probanden versucht, zu gewinnen, indem sie schon bei Gelb repetitiv geklickt hatten. Die Probanden wurden daher instruiert, dies zu unterlassen, weil man so keine Reaktionszeiten messen könne und weil das Programm so nicht überlistet werden könne. Drittens fanden es einige Probanden unplausibel, dass die Farbe der visuellen Anzeige in allen Runden anzeigt, dass der Gegenspieler noch mit Einstellen beschäftigt sei. Sie zogen daraus die Schlussfolgerung, dass es keinen Gegenspieler gibt. Daher wurde in der neuen Instruktion erläutert, dass die Anzeige immer gleich bleibt, auch wenn der Gegenspieler mit dem Einstellen fertig ist. Dies wurde damit gerechtfertigt, dass man sich sonst nicht immer rechtzeitig auf den Start des Reaktionsteils einstellen kann.

7.2.1.2 AGGRESSIVE MOTIVES SCALE

Aus dem Social Interactionist Model (siehe Abschnitt 4.3.3) lässt sich ableiten, dass es sich nur dann um Gewaltverhalten handelt, wenn die Schädigung absichtlich und zur Erreichung eines distalen Ziels erfolgt. Um dies überprüfen zu können, füllten die Probanden im Anschluss an das CRTT die „Aggressive Motives Scale“ (Anderson und Murphy, 2003) aus. Die AMS besteht aus sechs Items zur Motivation, und zwar „Ich wollte meinen Gegenspieler verletzen“ („CRTT Gegner verletzen“), „Ich wollte meinen Gegenspieler verrückt machen“ („CRTT Gegner verrückt machen“), „Ich wollte mich an meinem Gegenspieler rächen für die Lautstärke, die er/sie eingestellt hat“ („CRTT an Gegner rächen“), „Ich wollte meinen Gegenspieler behindern, um zu gewinnen“ („CRTT Gegner behindern“), „Ich wollte beeinflussen, welche Lautstärke mein Gegenspieler einstellt“ („CRTT Gegnerton beeinflussen“) , „Ich wollte ihm/ihr lautere Töne verpassen als er/sie mir verpasst hat“ („CRTT lauter als Gegner“). Mit den ersten drei Items der Skala soll überprüft werden, ob die Versuchspersonen während des CRTT Intentionen hat, die eine Schädigung des Gegners beinhalten, während die übrigen drei Items überprüfen sollen, ob auch nichtaggressive Intentionen vorliegen. Daneben enthält die AMS als siebtes Item die Frage, wie sehr sich der Proband während des CRTT geärgert hat („CRTT Ärger Proband“). Da die Einstellungen des Computerprogramms die Proban-

Virtuelle Nothilfe

den frustrieren sollen, soll so geprüft werden, ob diese Manipulation im CRTT erfolgreich ist. Der Bogen wurde um einige Fragen ergänzt, die das Vorhandensein eines Gegenspielers glaubwürdiger machen sollten. Dadurch sollte die Wirkung des Täuschungsmanövers überprüfbar gemacht und auch über den CRTT hinaus aufrecht erhalten werden. Für die Formulierung der Items wurde auf ein Experiment von Giancola und Zeichner (1995b) zurückgegriffen. Giancola und Zeichner baten ihre Probanden „to describe their subjective impressions of their opponent, how well he performed on the task, and the effectiveness of the task as a measure of RT“ (S. 202). Daraus wurden die drei Items „Wie gut eignet sich das Reaktionszeit-Spiel Ihrer Ansicht nach zum Messen von Reaktionszeiten?“ („CRTT als Reaktionszeittest geeignet“), „Reagierte Ihr Gegenspieler Ihrer Ansicht nach schneller als Sie?“ („Gegner ist schneller“) und „Bitte beschreiben Sie in wenigen Worten, welchen Eindruck Sie von Ihrem Gegenspieler gewonnen haben“ abgeleitet. Zusätzlich wurde die Frage hinzugefügt „Wie sehr hat sich Ihr Gegenspieler Ihrer Ansicht nach geärgert?“ („Ärger Gegner“).

7.2.2 HILFEVERHALTEN

Zur Erhebung des Hilfeverhaltens sollte eine Form von „willingness to assist“ verwendet werden (siehe Abschnitt 4.6.3). Um die Ergebnisse später mit der Studie von Greitemeyer und Oswald (2009) vergleichen zu können, wurde deren Operationalisierung als Grundlage verwendet und wie folgt modifiziert: Den Versuchspersonen wurde vorgetäuscht, dass die Experimentatoren von einem Diplomanden um einen Gefallen gebeten worden seien. Der Diplomand wolle untersuchen, wie gut man sich nach längeren Experimenten konzentrieren könne. Die Experimentatoren würden den Probanden gern kurz einen Probedurchgang machen lassen, damit er wisse, worum es gehe. Dann könne er entscheiden, ob er dem Diplomanden helfen möchte oder nicht. Da die Untersuchung des Diplomanden mit der der Experimentatoren nichts zu tun habe, bekäme der Proband keine zusätzliche Vergütung. Willigte der Proband in die Teilnahme ein, dann wurde ihm über ein Computerprogramm¹⁸ das Anliegen des Diplomanden aus dessen Perspektive erläutert (siehe Anhang M). Der Proband bearbeitete im Anschluss sechs Konzentrationsaufgaben, deren Schwierigkeit sich steigerte (siehe Abb. 32 bis Abb. 37).

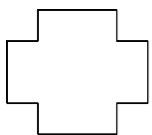


Abb. 32:
Aufgabe 1
(leicht)

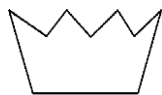


Abb. 33:
Aufgabe 2
(leicht)

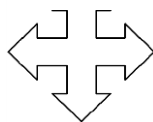


Abb. 34:
Aufgabe 3
(mittel)

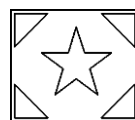


Abb. 35:
Aufgabe 4
(mittel)

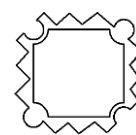


Abb. 36:
Aufgabe 5
(schwer)

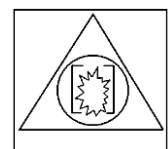


Abb. 37:
Aufgabe 6
(schwer)

Virtuelle Nothilfe

Er musste die Ecken eines Polygons zählen und im Anschluss angeben, ob die Anzahl gerade oder ungerade ist. Nach den sechs Probeaufgaben sollte er angeben, wie viele weitere Aufgaben dieser Art (maximal 200) er für den Diplomanden bearbeiten möchte. Damit die Probanden im Anschluss nicht sinnlos Aufgaben bearbeiteten, stürzte das Programm ab, sobald der Proband die Aufgabenanzahl eintrug. Der Versuchsleiter prüfte anhand der Reaktion der Probanden auf den Programmabsturz, ob die Probanden erkannten, dass es sich hierbei um eine Täuschung handelt. Falls dies der Fall war, wurde es notiert. Als Operationalisierung der AV Hilfeverhalten diente die Anzahl der Aufgaben, die der Proband bearbeiten wollte. Um auszuschließen, dass Zeitdruck keine Störwirkung auf die Hilfsbereitschaft ausübt, wurde den Probanden schon vor dem Experiment mitgeteilt, dass das Experiment zwei Stunden dauern würde, obwohl die Zeit für das Experiment so bemessen war, dass die Probanden noch zwischen 15 und 45 Minuten Zeit für diese Aufgabe gehabt hätten. Der Zeitdruck wurde zur Sicherheit als Störvariable operationalisiert, indem die dem Probanden für diese Aufgabe verbleibende Zeit berechnet wurde (siehe Abschnitt 7.3.12).

7.3 STÖRVARIABLEN

Folgende Störvariablen wurden über Fragebögen erfasst: das biologische Geschlecht, die Persönlichkeitseigenschaften „Gewaltbereitschaft“ und „Hilfsbereitschaft“, der Gewaltspielkonsum, die Eigenschaften der Spielaufgabe, die subjektive Erregung, die Schwierigkeiten mit der Steuerung, die Spieldauer sowie die Vorkenntnisse mit Computerspielen. Außerdem wurden folgende allgemeine Störvariablen erhoben: die Verdachtsbildung (engl. „suspicion“, d. h., die Versuchsperson bildete einen Verdacht über das wahre Ziel der Untersuchung und passte ihr Verhalten daran an), sowie die Nichtbefolgung von Anweisungen (engl. „non-compliance“, d. h., die Versuchsperson hielt sich absichtlich nicht an die Instruktionen). Aus dem Untersuchungsablauf ergaben sich weitere Störvariablen: der Zeitdruck bei der Hilfsbereitschaftsaufgabe, die Schulform der Probanden, ihr Alter und ihr Fachsemester sowie die Welle der Datenerhebung. Die Meta-Analyse von Anderson und Bushman (2001) zeigt, dass Gewaltspiele aggressive Emotionen verursachen können. Es bleibt aber ungeklärt, ob, und falls ja, welche anderen Emotionen ebenfalls durch Gewaltspielkonsum beeinflusst werden. Als Grundlage für eine spätere explorative Analyse wurden daher auch Emotionen erhoben, die im Rahmen dieser Studie aber nicht weiter analysiert wurden. Die Messung der Emotionen wird hier nur aufgeführt, weil sie Auswirkungen auf den Versuchsablauf sowie die Erhebung der subjektiven Erregung hatte.

Wenn die Störvariablen über Fragebögen erfasst wurden, dann wurden soweit wie möglich einheitliche verbale Marker verwendet. Die verwendeten Formulierungen stammen von Rohrmann (1978, zit. n. Bortz und Döring, 1995, S. 164), der empirisch belegen konnte, dass die Formulierungen „zumindest annähernd äquidistante Ausprägungen des Merkmalskontinuums markieren“. Das bedeutet, dass den Versuchspersonen der Abstand z. B. zwischen

„nie“ und „selten“ genauso groß vorkommt wie der Abstand zwischen „selten“ und „gelegentlich“.

7.3.1 BIOLOGISCHES GESCHLECHT

Nach Klimmt (2006a, S. 122) hat sich das biologische Geschlecht in vielen Studien zu Computerspielen als bedeutsame Moderatorvariable erwiesen. Männer spielen mehr Gewaltspiele und sind physisch aggressiver (Anderson et al., 2010). Wahrscheinlich ist das soziale Geschlecht relevanter als das biologische; für den deutschen Sprachraum stehen hier aber keine zuverlässigen Messinstrumente zur Verfügung (vgl. Mohseni, 2007). Daher wurde zumindest das biologische Geschlecht erhoben (siehe Anhang F). Außerdem wurde versucht, bei der Zuordnung der Personen zu den Versuchsbedingungen dasselbe Geschlechterverhältnis in jeder Bedingung zu realisieren (siehe Abschnitt 9.2.2).

7.3.2 TRAIT-GEWALTBEREITSCHAFT

Um vorab vorhandene Persönlichkeitsunterschiede hinsichtlich Gewalt- und Hilfsbereitschaft statistisch kontrollieren zu können, wurden Gewalt- und Hilfsbereitschaft als Persönlichkeitseigenschaft (engl. „trait“) mittels Fragebögen erhoben. In angloamerikanischen Gewaltspielstudien wird zur Messung von Gewaltbereitschaft in der Regel das aus 29 Items bestehende „Aggression Questionnaire“ von Buss und Perry (1992) eingesetzt. Um maximale Vergleichbarkeit zu erreichen, wurde die deutsche Version dieses Verfahrens „Aggressionsfragebogen von Buss und Perry“ (Amelang und Bartussek, 1997) verwendet. Diese Variante ist für den deutschen Sprachraum validiert (Herzberg, 2003; Collani & Werner, 2005). In der Validierungsstudie von Herzberg (2003) zeigte sich aber, dass zwei Items im deutschen Sprachraum problematisch sind. Nach Herzberg gab es diesen Befund auch bei anderen Autoren (Bryant & Smith, 2003; Nakano, 2001; beide zit. n. Herzberg, 2003). Die beiden Items wurden daher in dieser Untersuchung nicht verwendet, sodass eine Gesamtskala mit 27 Items resultierte. Für diese Gesamtskala ermittelte Herzberg (2003) eine Retest-Reliabilität von $r = .80$ und ein Cronbachs Alpha von $\alpha = .88$.

7.3.3 TRAIT-HILFSBEREITSCHAFT

Da Hilfsbereitschaft bisher nicht als Störvariable in Computerspielstudien erhoben wurde, fiel die Wahl eines geeigneten Verfahrens hier schwerer. Zur Auswahl standen die Kurzfassung BAS-II-K der Braunschweiger Adjunktions-Skalen (Bottenberg, Bilsky, Schindler & Wehrmeyer, 1979a, 1979b; Bottenberg, Langer & Hansen, 1979), die Prosozialitäts-Skalen von Bilsky (1981), die Helping Dispositions Scales (HDS) von Severy (1975) sowie die Subskala „Hilfsbereitschaft“ der Prosocial Personality Battery (PSB) von Penner, Craiger, Fritzsche und Freifeld (1995) in der deutschen Fassung von Staufienbiel (2000). Gegen den Einsatz der HDS sprach, dass es bisher keine deutsche Variante gibt. Auch die Prosozialitäts-Skalen von Bilsky

Virtuelle Nothilfe

erschieden problematisch, da die Formulierung bei einigen Items nicht mehr zeitgemäß wirkte (z. B. Item 6 „*Notzuchtverbrecher* und andere Sexualstraftäter verdienen keine Nachsicht“), der Inhalt einiger Items überholt wirkte (z. B. Item 7 „Ich begrüße Bestrebungen, die es Gefangenen erlauben, daß ihre Ehefrauen über Nacht bei ihnen bleiben“) und einige Items etwas anderes als Prosozialität zu messen schienen (z. B. Item 17 „Eine betrunkene Frau ekelt mich an“ oder Item 46 „Den Kommunismus würden mehr Menschen für gut halten, wenn sie nur etwas über ihn wüßten“). Dagegen schienen die BAS-II-K sowie die PSB beide besser geeignet zu sein. Hinsichtlich der Gütekriterien gab es zwischen diesen beiden Instrumenten nur geringe Unterschiede. Die Retest-Reliabilität der BAS-II-K beträgt $r_{tt} = .75$ (Cronbachs Alpha wird nicht berichtet). Die Gütekriterien der PSB werden weder für die Einzelskalen noch für die Gesamtskala berichtet, sondern für zwei Faktoren („Hilfsbereitschaft“ und „auf Andere gerichtete Empathie“), auf die sich die Items bei einer Faktorenanalyse verteilen. Bei beiden Faktoren liegt Cronbachs Alpha über $\alpha = .80$; die Retest-Reliabilität für den Faktor „Hilfsbereitschaft“ liegt bei $r_{tt} = .77$ und die für den Faktor „Empathie“ bei $r_{tt} = .85$ (Penner et al., 1995). Die Wahl fiel schließlich auf die PSB in der deutschen Fassung von Staufenberg (2000), da sie für eine zuverlässige Messung von Hilfsbereitschaft nur die Hälfte der Items (19 statt 40) benötigt.

Im Pretest zeigte sich, dass die Items der PSB von den Probanden kritisch bewertet wurden. Moniert wurde, dass einige Formulierungen unklar sind. Z. B. ist im Item „Wenn ich jemanden sehe, der in einer Notlage unbedingt Hilfe braucht, drehe ich durch“ nicht klar, ob mit Durchdrehen eher eine Art moralische Empörung oder eine Form von psychischer Überforderung gemeint ist. Außerdem fragen die meisten Items bereits gezeigtes Hilfeverhalten ab. Dabei sind sie aber sehr spezifisch (z. B. Weihnachtskarten gekauft; Person über Straße geholfen etc.), ohne dass sichergestellt wurde, dass diese Verhaltensweisen paradigmatisch für hilfsbereite Persönlichkeiten sind. Es ist daher denkbar, dass auch sehr hilfsbereite Personen (z. B. mangels Gelegenheit) viele der Verhaltensweisen noch nie gezeigt haben.

Daher wurde die Entscheidung für den Einsatz der PSB wieder verworfen und stattdessen auf die Subskala „Adjutons-Affirmierung“ der BAS-II-K zurückgegriffen. Von den 20 Items wurden 5 eliminiert, weil sie sich nicht auf die Hilfsbereitschaft der befragten Person beziehen („Mein Vater hat vielen Menschen geholfen.“, „Meine Mutter hat vielen Menschen geholfen.“, „Was Hilfsbereitschaft betrifft, war mir mein Vater ein Vorbild.“, „Was Hilfsbereitschaft betrifft, war mir meine Mutter ein Vorbild.“ und „Ich habe die schwerste Krise in meinem Leben nur meistern können, weil andere mir geholfen haben.“). Um die Reliabilität der stark gekürzten Skala nicht zu gefährden, wurde sie um die aus 13 Items bestehende Subskala „Motivation to Help“ aus der HDS¹⁹ ergänzt. Das Cronbachs Alpha für diese Subskala beträgt $\alpha = .88$, das für die Gesamtskala $\alpha = .94$. Es resultierte eine Gesamtskala mit 28 Items.

Beide Trait-Fragebögen wurden im ersten Vortest überprüft. Dabei wurden die Items der Trait-Gewaltbereitschafts-Skala per Zufallsprinzip mit denen der Hilfsbereitschafts-Skala

Virtuelle Nothilfe

vermischt. Dadurch sollte verschleiert werden, was die Fragebögen messen. Im Vortest gaben alle Probanden an, beim Ausfüllen die Vermutung gebildet zu haben, dass es im Experiment um Gewalt- und Hilfeverhalten (bzw. prosoziales Verhalten) geht. Nach Ferguson (2007a) ist dies kein unerwartetes Problem, da Fragebögen dieser Art leicht zu durchschauen sind.

Um die die Messung der AVen nicht zu verfälschen, hätte die die Messung der Gewalt- und Hilfsbereitschaft erst am Ende des Experimentes durchgeführt werden können. Dadurch wäre die Messung aber durch die Versuchsbedingungen kontaminiert worden. Eine statistische Kontrolle des Einflusses von Gewalt- und Hilfsbereitschaft auf die AVen wäre dann höchstwahrscheinlich nicht mehr möglich gewesen. Für die Probanden wäre es außerdem unplausibel gewesen, erst das Ende des Experimentes zu verkünden, sie dann um die Teilnahme an dem Experiment eines Diplomanden zu bitten (was ihre Hilfsbereitschaft messen sollte), und sie danach weitere Fragebögen ausfüllen zu lassen. Daher wurde auf den Einsatz der Fragebögen vollständig verzichtet.

7.3.4 GEWALTSPIELKONSUM/PROSOZIALSPIELKONSUM

Um zu prüfen, ob die Probanden möglicherweise bereits abgestumpft gegenüber Gewalt in Computerspielen sind, sollte ihr regelmäßiger Konsum gewalthaltiger Spiele erhoben werden. Zu diesem Zweck kann z. B. das „Video Game Questionnaire“ (VGQ) von Anderson und Dill (2000) verwendet werden. Dieses Instrument erfasst, welche Gewaltspiele die Probanden in der Vergangenheit wie oft gespielt haben (Anderson et al., 2007, S. 65): Die Versuchsperson soll die fünf am häufigsten gespielten Computerspiele auflisten und dann vier Fragen zur Spielhäufigkeit (in den letzten Monaten, in Klasse 11/12, in Klasse 9/10, in Klasse 7/8), zwei zum Gewaltgehalt (Inhalt, Grafik) und eine zum Genre beantworten. Anderson und Dill (2000, Studie 1) berichten für das VGQ ein Cronbachs α in Höhe von $\alpha = .86$ für das Gewalt-rating sowie $\alpha = .84$ für die Spielhäufigkeit. Ferguson et al. (2008a) dagegen differenzierten beim Einsatz des VGQ zwischen Gewaltgehalt der Grafik und Gewaltgehalt des Inhalts („storyline“) und berichten für beides jeweils $\alpha = .88$. Die beiden Maße haben eine Übereinstimmung von $\alpha = .96$ und messen laut den Autoren somit dasselbe. Die Autoren berichten außerdem eine Übereinstimmung der Spielhäufigkeit zwischen Vergangenheit und Gegenwart in Höhe von $\alpha = .83$.

Aufbauend auf dem VGQ haben Gentile, Walsh, Ellison, Fox und Cameron (2004) bzw. Anderson et al. (2007, S. 64f., S. 98f.) die „Media Violence Exposure Scale“ (MVE) erstellt, die in kürzerer Form die Exposition von Gewalt in den drei Bereichen „Computerspiele“, „TV“ sowie „Video-/Kinofilme“ erfasst. Zu jedem Bereich sollen die drei am häufigsten konsumierten Exemplare (Spiel, Fernsehprogramm bzw. Film) aufgelistet werden. Danach soll angegeben werden, wie oft das Exemplar konsumiert wurde und wie gewalthaltig es ist. Gentile et al. haben für den Summenwert dieser Skala eine befriedigende Konsistenz von $\alpha = .80$ gefun-

den, während die Konsistenz bei Anderson et al. (2007, S. 65) sogar bei $\alpha = .90$ lag. Gentile et al. (2009) weisen darauf hin, dass die interne Konsistenz der Skala zwangsweise eingeschränkt ist, da es nicht nur die Gruppe der Gewaltmedienkonsumenten und die der Gewaltmedienabstinenten gibt, sondern auch Mischformen (z. B. wenn nur Gewaltspiele, nicht aber Gewaltfilme konsumiert werden). Um die Validität der selbst eingeschätzten Gewalthaltigkeit zu belegen, führten Bushman et al. (2008, zit. n. Gentile et al., 2009, S. 756) außerdem einen Vergleich mit Expertenratings über 100 häufig genannte Spiele durch. Die Korrelation betrug $r = .75$ ($p < .001$).

In dieser Untersuchung wurde aus ökonomischen Gründen nur der Teil der MVE eingesetzt, der sich auf Computerspiele bezieht (siehe Anhang G). Da Greitemeyer und Osswald (2010) ihre Probanden nach der Prosozialität des Computerspiels fragten, wurde die MVE für dieses Experiment um eine solche Frage ergänzt. Um Fluktuationen im Spielverhalten abzumildern, wurde außerdem die Frage nach der Nutzungshäufigkeit in Anlehnung an das VGQ in zwei Fragen aufgeteilt: eine zur Nutzung im letzten Jahr und eine zur Nutzung in den Jahren davor. Über die drei gespielten Computerspiele und die beiden Zeiträume wurde im Anschluss ein mit der Häufigkeit gewichteter Summenwert für den Gewaltspielkonsum und ein ebensolcher Wert für den Prosozialspielkonsum errechnet.

7.3.5 EIGENSCHAFTEN DER SPIELAUFGABE

Es gibt verschiedene Eigenschaften der Spielaufgabe, die einen Einfluss auf die AV Gewaltverhalten haben können. Eine durch das Spiel erhöhte „subjektive Erregung/Aktivierung“ kann aggressives Verhalten direkt beeinflussen (siehe Abschnitt 7.3.6). Andere Eigenschaften können indirekt die Aggression beeinflussen. Hierzu zählen insbesondere die „Schwierigkeit“ des Spiels und die „Frustration“ durch das Spiel. Eine zu hohe Schwierigkeit kann beängstigend und frustrierend sein, während eine zu geringe Schwierigkeit langweilen kann (siehe Abschnitt 4.2.5). Langeweile kann entweder zu Frustration oder zu geringerer subjektiver Erregung führen. Ein starkes Präsenzgefühl fördert den Unterhaltungswert (siehe Abschnitt 4.2.5), der wiederum negative Gefühle reduziert. Angst und Frustration könnten die Aggression erhöhen, während fehlende Erregung und das Ausbleiben negativer Gefühle sie senken könnten.

Die vier Versuchsbedingungen sollten daher hinsichtlich ihrer Eigenschaften „subjektive Erregung/Aktivierung“, „Schwierigkeit“, „Frustration“, „Interessantheit“, „Spannungswert“, „Unterhaltungswert“ und „Präsenzgefühl“ vergleichbar sein, weil die experimentelle Manipulation sonst mit einer dieser Variablen konfundiert sein könnte. Diese subjektiven Spieleigenschaften wurden von den Probanden per Fragebogen bewertet (siehe Anhang G). Dazu wurde das „Video Game Rating Sheet“ (VGRS) von Anderson und Ford (1986) verwendet. Dieses besteht aus insgesamt sieben Items, die auf einer siebenstufigen Skala die folgenden Eigenschaften eines Spiels erfassen sollen: Schwierigkeit („How difficult was X“), Unterhal-

tungswert („How enjoyable was X“), Frustration („How frustrating was X“), Spannungswert („How exciting was X“), Geschwindigkeit („How fast was the action of X“), Gewaltgehalt des Spielinhalts („How violent was the content of X“) sowie Gewaltgehalt der grafischen Darstellung („How violent were the graphics of X“). Die Güte des Gesamtverfahrens ist unbekannt, da Gütekriterien nicht berichtet werden. Alle Items wurden für diese Untersuchung ins Deutsche übersetzt (siehe Anhang G). Um den Fragebogen an die anderen Bögen anzupassen, wurde statt einer sieben- eine fünfstufige Skala verwendet. Außerdem wurden die folgenden Items ergänzt: Ich finde, dass „die Spielaufgabe interessant war“ (Interessantheit), „die Spielaufgabe mich munter gemacht hat“ (Subjektive Erregung, siehe Abschnitt 7.3.6), „die Spielaufgabe mich lustlos gemacht hat“ (Subjektive Erregung), „ich in dem Spiel versunken bin“ (Präsenzgefühl) und „die Grafik im Spiel realistisch war“ (Darstellungsrealismus als Voraussetzung für Präsenzgefühl). Außerdem wurden drei Items ergänzt, mit denen der Erfolg der experimentellen Manipulation geprüft werden sollte (siehe Abschnitt 1.1.1).

7.3.6 SUBJEKTIVE ERREGUNG

Um eine Konfundierung zu vermeiden, sollten die vier Versuchsbedingungen auch vergleichbar sein hinsichtlich der subjektiven Erregung/Aktivierung (engl. „perceived arousal“), die sie beim Spieler auslösen. Im Bereich der Gewaltspiele gibt es mit der „Perceived Arousal Scale“ (Anderson, Deuser & DeNeve, 1995) ein Verfahren zur systematischen Erhebung der subjektiven Erregung. Die Skala besteht aus 24 Adjektiven, die den aktuellen Zustand der Erregung beschreiben (z. B. „energiegeladen“, „munter“, „ausgelaugt“, „müde“). Auf einer 5-stufigen Skala gibt der Proband an, inwieweit der jeweilige Zustand vorliegt.

Im zweiten Vortest gab es beim Ausfüllen der Perceived Arousal Scale (PAS) das Problem, dass diese von der Struktur her dem EMO16 sehr ähnlich ist und die Inhalte sich teilweise überschneiden, was die Probanden irritierte. Die beiden Verfahren ließen sich aber nicht zusammenfassen, da der EMO16 Substantive verwendet, zu denen es nicht immer ein korrespondierendes Adjektiv gibt, während die PAS Adjektive verwendet, zu denen es nicht immer ein korrespondierendes Substantiv gibt. Daher wurde auf den Einsatz der vollständigen PAS verzichtet; nur die Items „munter“ und „lustlos“ wurden in die Fragen zur Spielaufgabe integriert (siehe Anhang G).

7.3.7 SCHWIERIGKEITEN MIT DER STEUERUNG

Um Unterschiede in den Fähigkeiten bei der Benutzung von Maus und Tastatur zu reduzieren, wurde zu Beginn des Computerspiels ein Training der Spiel-Steuerung, ein sogenanntes „Tutorial“, durchlaufen. In einem separaten Vortest wurde die Brauchbarkeit des Tutorials geprüft. Außerdem wurden unter dem Monitor zwei Aufsteller positioniert, die das Tastaturlayout sowie die Mausbelegung visualisierten (siehe Abb. 38 und Abb. 39). Zusätzlich wurde eine beleuchtete Tastatur verwendet, um den Probanden die Tastensuche im Dunkeln zu er-

Virtuelle Nothilfe

leichtern. Nach dem Spiel sollten die Versuchspersonen angeben, ob ihnen die Steuerung bereits vor dem Spiel bekannt war (siehe Anhang G). Danach sollten sie bewerten, ob sie trotz Tutorial Schwierigkeiten mit der Steuerung hatten. Außerdem wurde die Zeit erfasst (siehe Anhang G), die für das Absolvieren des Tutorials bzw. der Spielaufgabe benötigt wurde, da eine geringe Lösungszeit ein Indikator für eine hohe Spielfähigkeit des Probanden sein kann (unter der Annahme, dass ein Proband sich als Ziel primär die Lösung der Aufgabe gesetzt hat und nicht etwa die Exploration des Spiels). Schließlich wurde vom Spiel automatisch gespeichert, wie oft die Spielfigur des Probanden ohnmächtig geworden ist. Diese fünf Variablen (Steuerungsvorkenntnisse, subjektive Schwierigkeiten, Tutorialdauer, Aufgabendauer, Ohnmachten der Spielfigur) gingen als separate Störvariablen in die Auswertung ein.

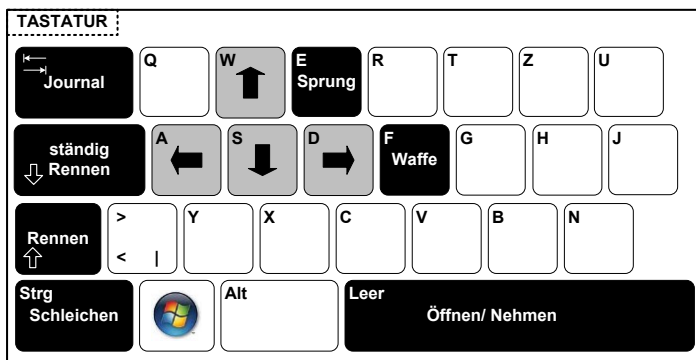


Abb. 38: Info Tastaturlayout

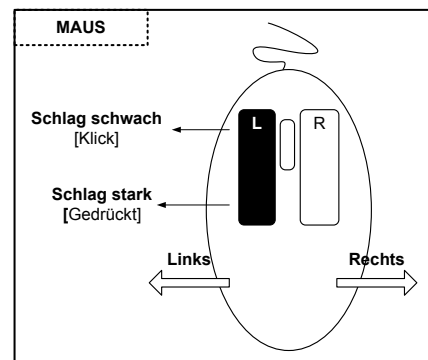


Abb. 39: Info Mausbelegung

Während des ersten Vortests wurde das Problem identifiziert, dass die Menge der Hinweise im Tutorial für die „Amateure“ (hier die vier Spieler ohne Oblivion-Erfahrung) genau richtig war, für die „Profis“ (hier die zwei Spieler mit Oblivion-Erfahrung) aber viel zu hoch. Außerdem fühlten sich die Amateure selbst nach dem Absolvieren des Tutorials im Umgang mit der Steuerung weniger sicher als die Profis. Als Letztes bestand ein wesentlicher Unterschied zwischen Profis und Amateuren darin, dass die Profis für das Durchlaufen des Tutorials nur ein Drittel der Zeit benötigten. Es wurde angenommen, dass diese Unterschiede sich in der Hauptuntersuchung noch verschärfen würden. Daher wurde entschieden, in der Hauptuntersuchung nur erfahrenere Spieler/innen als Probanden zuzulassen.

7.3.8 SPIELDAUER

Im Gegensatz zu dieser Studie spielen Proband/innen im Rahmen der Gewaltspielforschung das Computerspiel nicht solange, wie sie zur Lösung einer vorgegebenen Aufgabe benötigen, sondern für eine vorher festgelegte Zeit. Dabei hat sich eine Dauer von 20 Minuten etabliert (z. B. Anderson et al., 2007, Studie 1; Fischer, Kubitzki, Guter & Frey, 2007, Studie 2; Carnagey & Anderson, 2005; Irwin & Gross, 1995, zit. n. Anderson et al., 2007, S. 33); es gibt aber auch Autoren, die mit noch kürzeren Treatments arbeiten (z. B. 15 Minuten bei Anderson & Dill, 2000, Studie 2; 12 Minuten bei Elson, 2010; 10 Minuten bei Greitemeyer & Osswald,

2009, Studie 1 und 2; 2010, Studie 2; Hartmann & Vorderer, 2010; 8 Minuten bei Greitemeyer & Osswald, 2010, Studien 1 und 3). Nach Anderson et al. (2010) ist die kurzfristige Wirkung von Gewaltspielen auf das aggressive Verhalten hauptsächlich auf Priming-Prozesse zurückzuführen. In einem solchen Fall sei es egal, ob die Spieldauer 15 oder 30 Minuten betragen würde. Nur im Fall von Beobachtungslernprozessen hätte eine Spieldauer über 15 Minuten eine stärkere Wirkung. Dies sei aber nur dann der Fall, wenn die Probanden die Aggressionsskripte nicht bereits verinnerlicht hätten (z. B. Kinder), wenn das modellierte Aggressionsverhalten neuartig sei und wenn die Situation während der Aggressionsmessung der Situation im Computerspiel hinsichtlich der Charaktere, der Provokation und der verfügbaren Gewaltreaktion sehr ähnlich sei. Alle diese Punkte treffen in diesem Experiment nicht zu. Daher scheint eine Spieldauer von 15 bis 20 Minuten angemessen.

Im Gegensatz zu den üblichen Experimenten sollten die Probanden in dieser Studie im Spiel eine Aufgabe lösen, ohne dass vorher klar war, wie viel Zeit sie für die Lösung benötigen werden. Die Lösungszeit ist aber zentral, da sie der Treatment-Dauer entspricht. Aus diesem Grund wurden für jeden Probanden zwei Zeiten per Stoppuhr erfasst, und zwar zum einen die für das Durchlaufen des Tutorials benötigte Zeit, und zum anderen die für das Lösen der Spielaufgabe benötigte Zeit (siehe Anhang G). Diese Zeitmessungen sind unter bestimmten Bedingungen auch ein Indikator für die spielerischen Fähigkeiten des Probanden (siehe Abschnitt 7.3.7). Die Spielaufgabe wurde so optimiert, dass die Lösungszeit für die meisten Probanden ca. 20 Minuten betragen sollte.

7.3.9 VORKENNTNISSE MIT COMPUTERSPIELEN

Die Aufgabenschwierigkeit spielt in Wechselwirkung mit den spielerischen Fähigkeiten eine Rolle für den Unterhaltungswert, der wiederum mit negativen Gefühlen zusammenhängt (siehe Abschnitt 4.2.5). Diese negativen Gefühle können sich auf das aggressive Verhalten auswirken. Daher ist es wichtig, dass die Probanden vergleichbare Fähigkeiten zur Lösung der Spielaufgabe mitbringen. Es wurde aber vermutet, dass Probanden ohne Erfahrung mit Oblivion oder ähnlichen Spielen mehr Schwierigkeiten beim Lösen der Spielaufgabe haben könnten als solche mit Erfahrung, was sich auch im ersten Vortest zeigte (siehe Abschnitt 7.3.7). Da die Hypothesen eher für Hobbyspieler als für Nichtspieler Geltung beanspruchen, wurde daher beschlossen, nur Hobbyspieler als Probanden zuzulassen. Bei der Werbung für die Teilnahme wurde potenziellen Probanden mitgeteilt, dass sie nur dann teilnehmen können, sofern sie die Figur in Oblivion oder ähnlichen Spielen (z. B. Gothic, Counter-Strike, World of Warcraft, GTA, Second Life etc.) steuern können. In der Voruntersuchung wurden die Vorkenntnisse mit solchen Spielen und mit Oblivion erhoben (siehe Anhang F). Um bei den Hobbyspielern die Erfahrung mit Oblivion feiner differenzieren zu können, wurde auch danach gefragt, ob sie schon einmal eine Modifikation für Oblivion benutzt bzw. erstellt haben (siehe Anhang F), ob sie das Spiel schon vor dem Tutorial steuern konnten (siehe Anhang G), wie nützlich das Tutorial war, wie sicher ihr Umgang mit dem Spiel nach dem Tutorial

war, wie oft sie auf das Poster mit der Tastenbelegung geschaut haben und wie oft sie in den letzten 90 Tagen Oblivion gespielt haben (vgl. Anhang F).

Von Anfang an wurde befürchtet, dass die letzte Frage schwierig zu beantworten sein könnte, da Computerspielen oft eine regelmäßig auftretende Verhaltensweise ist. In einem solchen Fall ist es nach Schwarz und Oyserman (2001, S. 136f.) unwahrscheinlich, dass jede Episode einzeln im Gedächtnis gespeichert wird; stattdessen verschmelzen die Episoden zu einer globalen Repräsentation ohne Zeit- oder Ortsinformation. Schwarz und Oyserman empfehlen, als Gegenmaßnahme eine kurze und noch nicht lange zurückliegende Referenzperiode zu wählen. Allerdings weisen sie darauf hin, dass kurze Perioden bei Probanden, die das Verhalten selten ausüben (in unserem Fall bei den Amateuren), zu vielen „Null“-Antworten führen können. Außerdem würden geringe Häufigkeiten eher überschätzt, während hohe Häufigkeiten eher unterschätzt werden. Zur Förderung der Erinnerung empfehlen die Autoren, erinnerungsdienliche Hinweisreize einzusetzen. Sie schlagen vier verschiedene Strategien vor, die sich auch kombinieren lassen: (1) Dekomposition der Frage, (2) Verringerung des Zeitdrucks/ Erhöhung der Motivation, (3) Optimierung der zeitlichen Suchrichtung sowie (4) Verankerung an salienten Ereignissen.

1. Die Strategie „Dekomposition der Frage“ besteht darin, die Frage in Teilfragen zu zerlegen (z. B. eine Frage nach der allgemeinen Computerspieldauer in Fragen zur Spieldauer bestimmter Spiele zu zerlegen). Diese Strategie macht in unserem Fall wenig Sinn, da die Frage schon auf ein bestimmtes Spiel abzielt, und die Aktivitäten innerhalb des Spiels nicht weiter sinnvoll zerlegt werden können. Außerdem ist Dekomposition nach Schwarz und Oyserman nur dann erinnerungsförderlich, wenn die Teilfragen auf seltene und gut erinnerbare Ereignisse abzielen.
2. Die Strategie „Verringerung des Zeitdrucks/ Erhöhung der Motivation“ besteht darin, durch explizite Instruktion darauf hinzuweisen, dass die Frage sehr wichtig ist und sich für die Beantwortung Zeit genommen werden soll. Allerdings funktioniert sie nur, wenn sie für wenige Fragen sparsam eingesetzt wird. Im Vortest- Fragebogen wurde eine solche Instruktion realisiert.
3. Die Strategie „Optimierung der zeitlichen Suchrichtung“ besteht darin, eine chronologische Folge von der Gegenwart in die Vergangenheit zu wählen, weil es leichter ist, sich an kürzlich erfolgte Instanzen zu erinnern, die wiederum als Hinweisreiz für zeitlich davorliegende Instanzen dienen können. Dies gilt nicht, wenn vorangegangene Instanzen eine logische Voraussetzung für spätere Instanzen darstellen, was hier aber nicht der Fall war. Daher wurde diese Strategie bei der Formulierung des Items berücksichtigt.
4. Die Strategie „Verankerung an salienten Ereignissen“ besteht darin, ein gut erinnerbares Ereignis (z. B. Weihnachten) als Verankerung einzusetzen. Zum einen hilft dies, bei Zeiträumen wie z. B. „12 Monate“ den Interpretationsspielraum zu minimieren (ist das Kalenderjahr gemeint, wird der aktuelle Monat mitgerechnet etc.). Zum an-

deren ist ein solcher Anker erinnerungsförderlicher als eine abstrakte Datumsangabe (wie z. B. „seit dem ersten März“). Ein Anker dieser Art ließ sich in unserem Bogen jedoch nicht verwenden, da die Untersuchungsteilnehmer zu unterschiedlichen Zeitpunkten befragt wurden. Als Notlösung wurde der Zeitraum in Form von Tagen statt Monaten abgefragt.

Trotz der weitestmöglichen Berücksichtigung der genannten Empfehlungen von Schwarz und Oyserman sind die Antworten auf das Item „Spielhäufigkeit“ wahrscheinlich ungenau. Sie könnten dennoch genügen, um Amateure von Profis unterscheiden zu können. Die Autoren empfehlen, zur Sicherheit die Schwierigkeit und die Genauigkeit der Beantwortung von den Befragten einschätzen zu lassen und explizit auf die Möglichkeit einer späteren Korrektur hinzuweisen (S. 142). Diesen Empfehlungen wurde ebenfalls gefolgt. Die resultierenden Formulierungen finden sich in Abschnitt 9.1.

Im ersten Vortest zeigte sich, dass die Probanden die Frage trotzdem nicht beantworten können. Daher wurden in Anlehnung an die Media Violence Exposure Scale (siehe Abschnitt 7.3.4) gröbere Antwortkategorien gewählt ([fast] nie – ein paar Mal im Monat – ca. 1 Mal pro Woche – ca. 2 bis 3 Mal pro Woche – [fast] täglich). Zusätzlich wurde das Item-Format an die anderen Items angeglichen, sodass das Item lautete „Ich spiele Oblivion am PC bzw. habe früher Oblivion am PC gespielt“. Die Störvariable „Vorkenntnisse mit Computerspielen“ wurde aus der Summe der Items „Vorkenntnisse mit Oblivion“, „Vorkenntnisse mit anderen Rollenspielen“ und „Vorkenntnisse mit ähnlich gesteuerten Spielen“ erstellt. Die Vorkenntnisse mit Oblivion gingen dreifach, die mit anderen Rollenspielen zweifach und die mit ähnlich gesteuerten Spielen einfach gewichtet in die Summe ein. Dadurch sollte der Umstand berücksichtigt werden, dass die Erfahrung mit dem Spiel Oblivion sich stärker auf die Steuerfähigkeiten auswirkt als die Erfahrung mit anderen Rollenspielen, und dass die Erfahrung mit anderen Rollenspielen sich stärker auswirkt als die Erfahrung mit lediglich ähnlich zu steuernden Spielen.

7.3.10 VERDACHTSBILDUNG

Bei der Verdachtbildung geht es um das Problem, dass die Probanden ihr Verhalten verändern, wenn sie einen Verdacht gebildet haben, worum es in der Untersuchung gehen könnte. Aus diesem Grund wurde am Ende des Experiments direkt danach gefragt. Während die Erfassung des Hilfeverhaltens während der Vortests nicht zu einem Verdacht geführt hat, wurde beim Gewaltverhalten im zweiten Vortest festgestellt, dass viele Probanden aus unterschiedlichen Gründen erkannt haben, dass sie gegen einen Computer angetreten sind, dass es in der Untersuchung um Aggression geht und dass der CRTT ihre Aggression misst. Jene Probanden, die eine Erfassung ihrer Aggression vermuten, könnten wegen der sozialen Unerwünschtheit aggressiven Verhaltens zur Dissimulation neigen (siehe Abschnitt 4.4.2.3). Wenn sie dies vor bzw. während der Durchführung des CRTT erkannt haben, könnte bei die-

Virtuelle Nothilfe

sen Probanden der Zusammenhang zwischen Computerspiel und aggressivem Verhalten niedriger ausfallen. Bei jenen Probanden, die erkannt haben, dass sie nicht gegen einen menschlichen Gegenspieler angetreten sind, ist die *Richtung* der Verhaltensänderung schwieriger vorherzusagen. Auf der einen Seite könnten sie extrem hohe Werte im CRTT eingestellt haben, um auszutesten, ob die Reaktion darauf eher für einen Menschen oder eher für einen Computer spricht. Auf der anderen Seite könnten sie die Werte beliebig eingestellt haben, da ein Computer die eingestellten Töne ohnehin nicht „hören“ kann.

Um die Verdachtsbildung zu verhindern, wurde daher ein Täuschungsmanöver mit Konföderiertem eingeführt. Dieser traf immer erst zum Zeitpunkt des Untersuchungsbeginns ein und gab sich dann beim Versuchsleiter als Proband aus. War der eigentliche Proband bereits vor Untersuchungsbeginn da, so wurde er gebeten, auf den zweiten Teilnehmer zu warten. Kam der Proband hingegen zu spät, so wartete der Konföderierte vor dem Labor. Sobald der eigentliche Proband eintraf, teilte der Versuchsleiter dem Konföderierten mit, dass der zweite Versuchsleiter noch nicht eingetroffen sei. Er begleitete den Konföderierten in den „Versuchsraum“ und bat ihn, sich bis zum Eintreffen des zweiten Versuchsleiters schon einmal die Instruktionen zum Versuch durchzulesen. Der Raum war identisch in der Ausstattung mit dem echten Versuchsraum. Um die Täuschung nicht zu gefährden, lagen beide Räume so weit auseinander, dass man von dem echten Versuchsraum keine Sicht- und Hörverbindung zu diesem Raum hatte. Der Konföderierte wurde gebeten, auf keinen Fall zu verschwinden, falls der zweite Versuchsleiter nicht innerhalb der nächsten fünf Minuten einträte. Stattdessen solle er in dem Fall zum „ersten“ Versuchsleiter in dessen Versuchsraum gehen. Dem Konföderierten wurde kurz erklärt, wie er diesen Raum finden kann. Im Anschluss daran ging der Versuchsleiter mit dem Probanden in den Versuchsraum und bat ihn, ebenfalls die Instruktionen zu lesen. Dann wurde mit dem Probanden die Voruntersuchung durchgeführt. Danach wurde die schriftliche Einverständniserklärung eingeholt. Während der Proband die Erklärung ausfüllte, verließ der Versuchsleiter den Raum und gab vor, zu prüfen, ob der zweite Versuchsleiter auch wirklich eingetroffen sei. Kurz vor Beginn des CRTT verließ der Versuchsleiter ein weiteres Mal den Raum, um zu prüfen, ob der „andere Proband“ auch schon mit dem Spielen fertig wäre. Diese Manöver sollten den Probanden glauben machen, dass er im CRTT gegen einen menschlichen Gegner antritt. Bei einigen Probanden wurde an Hand ihres Namens und ihres Studienfachs schon vor Beginn der Untersuchung angenommen, dass diese den Konföderierten persönlich (er-)kennen könnten. Es wurde befürchtet, dass sie ihn dadurch entweder entlarven oder aber ihr Verhalten ihm gegenüber verändern würden (z. B. indem sie ihm im CRTT weniger laute Töne verabreichten). Bei diesen Probanden übernahm der Versuchsleiter die Rolle des Konföderierten.

Die Wirksamkeit der Täuschung wurde durch zwei Maßnahmen überprüft: Zum einen mussten die Probanden im Anschluss an das CRTT einen Online-Fragebogen ausfüllen, in welchem sie das CRTT und ihren Gegenspieler bewerten sollten (siehe Abschnitt 7.2.1). Zum anderen wurden die Probanden während der Aufklärung gebeten, mögliche Täuschungsmanöver zu

Virtuelle Nothilfe

benennen. Der Versuchsleiter paraphrasierte die Antworten der Probanden und hielt diese schriftlich fest. Er gab außerdem eine subjektive Bewertung ab, ob der Proband das Täuschungsmanöver nicht, teilweise oder gänzlich durchschaut hatte. „Teilweise durchschaut“ wurde kodiert, wenn der Proband entweder erkannte, dass es keinen menschlichen Gegner gab, oder dass das CRTT Aggression messen sollte. Erkannte der Proband beides, wurde „gänzlich durchschaut“ kodiert. Diese Kodierung erwies sich jedoch als unzureichend, da dem Versuchsleiter für eine genauere Beurteilung die Informationen fehlten, die der Proband im Fragebogen angegeben hatte.

Daher wurde eine Inhaltsanalyse vorgenommen, die eine exaktere Kodierung ermöglichen sollte. Als Analyseeinheit (vgl. Rustemeyer, 1992; Merten, 1995) wurden alle Aussagen gewählt, die sich auf das Täuschungsmanöver beziehen, und zwar sowohl die Angaben im Fragebogen als auch die Angaben gegenüber dem Versuchsleiter. Es wurden auch geschlossene Fragen berücksichtigt, wenn aus den Antworten Rückschlüsse gezogen werden konnten. Folgende geschlossenen Fragen gingen in die Analyse ein:

1. Wie gut eignet sich das Reaktionszeit-Spiel Ihrer Ansicht nach zum Messen von Reaktionszeiten?
2. Wie sehr haben Sie sich während des Reaktionszeit-Spiels geärgert?
3. Wie sehr hat sich Ihr Gegenspieler Ihrer Ansicht nach geärgert?
4. Reagierte Ihr Gegenspieler Ihrer Ansicht nach schneller als Sie?
5. Ich wollte meinen Gegenspieler behindern, um zu gewinnen.
6. Ich wollte beeinflussen, welche Lautstärke mein Gegenspieler einstellt.
7. Ich wollte meinen Gegenspieler verrückt machen.
8. Ich wollte meinen Gegenspieler verletzen.
9. Ich wollte mich an meinem Gegenspieler rächen für die Lautstärke, die er/sie eingestellt hat.
10. Ich wollte ihm/ihr lautere Töne verpassen als er/sie mir verpasst hat.
11. Haben Sie eine Vermutung, warum es in der Untersuchung geht?

Aus dem Fragebogen wurden außerdem folgende offenen Fragen analysiert:

12. Bitte beschreiben Sie in wenigen Worten, welchen Eindruck Sie von Ihrem Gegenspieler gewonnen haben.
13. [Falls Sie eine Vermutung haben, worin es in der Untersuchung geht.] Was vermuten Sie, worum es geht?
14. [Falls Sie eine Vermutung haben, worin es in der Untersuchung geht.] Ab welchem Zeitpunkt hatten Sie diese Vermutung?
15. Gibt es sonst etwas, das Sie mitteilen möchten? Nutzen Sie die Gelegenheit für Kommentare und Anmerkungen!

Zusätzlich wurden folgende offenen Fragen aus der Probandenaufklärung hinzugenommen:

16. Welche Täuschungsmanöver gab es?
17. Was war das Ziel der Untersuchung?
18. Sonstige Anmerkungen [des Versuchsleiters]

7.3.10.1 ERSTELLUNG DER KATEGORIENSYSTEME

Nach der Durchführung des Experiments, aber noch vor der Sichtung der Daten, wurden deduktiv die zwei Kategoriensysteme „Aggression durchschaut“ und „Computergegner durchschaut“ erstellt (siehe Tabelle 3 und Tabelle 4), um mit den gebildeten Variablen Moderatoranalysen durchzuführen.

Tabelle 3: Kategoriensystem „Aggression durchschaut“

Oberkategorie	Kode	Unterkategorie
Keine Aussagen zum Experiment/ Keine Ahnung, was das Experiment untersuchen soll	11	Keine Aussagen/ Keine Ahnung
Das Experiment soll etwas untersuchen (außer Aggression und Gefühle)	21	Vor CRTT
	22	Während CRTT
	23	Nach CRTT
	24	Unklar
Das Experiment soll Gefühle untersuchen	31	Vor CRTT
	32	Während CRTT
	33	Nach CRTT
	34	Unklar
Das Experiment soll Aggression untersuchen	41	Vor CRTT
	42	Während CRTT
	43	Nach CRTT
	44	Unklar
Das Reaktionszeit-Spiel soll Aggression untersuchen	51	Vor CRTT
	52	Während CRTT
	53	Nach CRTT
	54	Unklar
Widersprüchliche Aussagen zum Experiment	61	Widersprüche
Sonstiges	71	Vor CRTT
	72	Während CRTT
	73	Nach CRTT
	74	Unklar

Hinsichtlich des Zeitpunkts wurde immer dann „Vor CRTT“ kodiert, wenn die Aussagen des Probanden sich auf Ereignisse beziehen, die vor dem CRTT stattfanden (z. B. vor dem Experiment, zu Experimentbeginn, während der Erklärung des CRTT, während des Spiels Oblivion etc.). „Während CRTT“ hingegen wurde kodiert, wenn die Aussagen sich auf Ereignisse beziehen, die nach dem Einstellen der Regler durch den Probanden in der ersten Runde des CRTT stattfanden. Bei der Kategorie „Nach CRTT“ wiederum beziehen sich die Aussagen auf

Virtuelle Nothilfe

Ereignisse nach Abschluss des CRTT, während bei der Kategorie „Unklar“ die Aussagen unklar lassen, wann die benannten Ereignisse stattfanden.

Tabelle 4: Kategoriensystem „Computergegner durchschaut“

Oberkategorie	Kode	Unterkategorie
Keine Aussagen	11	Keine Aussagen
Proband ist der Gegner	21	Proband ist Gegner
	22	Proband ist nicht Gegner (nach Aufklärung)
Jemand anderes ist (wahrscheinlich) Gegner	31	Vor CRTT
	32	Während CRTT
	33	Nach CRTT
	34	Unklar
Computer ist (wahrscheinlich) Gegner	41	Vor CRTT
	42	Während CRTT
	43	Nach CRTT
	44	Unklar
Widersprüchliche Aussagen zum Gegner	51	Gegner existiert nicht, kein Mensch etc.
	52	Gegner ist Konföderierter
Sonstiges	61	Vor CRTT
	62	Während CRTT
	63	Nach CRTT
	64	Unklar

An dieser Stelle zeigte sich bereits, dass es besser gewesen wäre, spezifische Fragen zum Zeitpunkt des Durchschauens zu stellen. Dies unterblieb leider, da nach dem dritten Vortest nicht erwartet wurde, dass eine so große Anzahl an Probanden die Täuschungsmanöver durchschaut. Dadurch fehlten nun Informationen, die für eine präzisere Kodierung nützlich gewesen wären. Bei der Kodierung des Zeitpunktes mussten somit einige nicht überprüfbare Annahmen gemacht werden: Wenn ein Proband noch kurz nach dem Reaktionsspiel der Ansicht war, gegen den anderen Probanden gespielt zu haben, später (d. h. vor oder während der Aufklärung) aber der Ansicht war, schon davor die Täuschung erkannt zu haben, dann wurde davon ausgegangen, dass diese Erkenntnis erst nach dem Reaktionszeit-Spiel erfolgt ist. Wenn ein Proband kurz nach dem Reaktionsspiel der Ansicht war, nicht gegen den anderen Probanden gespielt zu haben, aber keine Angaben dazu machte, wann er diese Überzeugung gewonnen hat, dann wurde davon ausgegangen, dass er dies während des Reaktionszeit-Spiels und nicht schon davor erkannt hat.

Die Kategoriensysteme wurden mit Beispielen aus 13 Datensätzen expliziert, die aus dem Pool der 41 Datensätze stammten, welche nicht in die spätere statistische Auswertung eingingen (d. h. die Datensätze der Probanden aus dem dritten Vortest sowie die der Probandinnen). Die explizierten Kategoriensysteme wurden durch drei Kodierer an weiteren 14 Datensätzen aus diesem Pool geprüft. Dabei wurde die Kodierer-Übereinstimmung über Fleiss' κ bestimmt (zur Berechnung siehe Bortz, Lienert & Boehnke, 2000; Rustemeyer, 1992; Landis & Koch, 1977), welches mit dem webbasierten Programm „ReCal3“ (vgl. Freelon, 2010) berechnet wurde. Für das erste Kategoriensystem „Aggression durchschaut“ resultier-

te $\kappa = .43$. Dieser Wert kann als mittelmäßig bezeichnet werden (vgl. Grouven, Bender, Ziegler & Lange, 2007, S. 66; Landis & Koch, 1977, S. 165). Für das zweite Kategoriensystem „Computergegner durchschaut“ resultierte $\kappa = .56$, das ebenfalls als mittelmäßig bezeichnet werden kann. Aufgrund dieses suboptimalen Ergebnisses wurde eine Kodierer-Schulung durchgeführt, bei der die Nichtübereinstimmungen geprüft und diskutiert wurden. Im Anschluss wurde das Kategoriensystem an den verbliebenen 14 Datensätzen aus dem Pool geprüft. Nun resultierte für das erste Kategoriensystem $\kappa = .56$, das als mittelmäßig bezeichnet werden kann, und für das zweite Kategoriensystem ein ebenfalls als mittelmäßig zu bezeichnendes $\kappa = .58$. Dieses Ergebnis wurde als wenig zufriedenstellend bewertet, da für die geplante Kovarianzanalyse (siehe Abschnitte 0 und 12.2.3) eine möglichst exakte Kodierung der Störvariablen nötig war. Daher wurden die Kategoriensysteme vor der eigentlichen Kodierung noch einmal präzisiert²⁰.

7.3.10.2 KODIERUNG DER PROBANDENAUSSAGEN

Im Anschluss wurden die Aussagen der ersten 93 Probanden mithilfe der Kategoriensysteme von denselben drei Kodierern verkodet, wobei sich für das erste Kategoriensystem $\kappa = 0.63$ und für das zweite $\kappa = 0.56$ ergab. Dieses Ergebnis war erneut nicht zufriedenstellend. Daher wurden beide Kategoriensysteme entgegen den Prinzipien der Inhaltsanalyse vor der Kodierung der übrigen 93 Datensätze erneut überarbeitet. Durch die Überarbeitung der Kategoriensysteme wurde die Objektivität der Kodierung der bereits kategorisierten Datensätze verletzt, da sich nachträglich nicht überprüfen lässt, wie stark die Kodierer übereingestimmt hätten, wenn sie von vornherein die neuen Kategoriensysteme verwendet hätten. Diese Verletzung erschien aber notwendig, um präzisere Kategoriensysteme zu erhalten, die eine Erstellung zuverlässiger Störvariablen ermöglichen.

Beim Kategoriensystem „Aggression durchschaut“ wurden lediglich die Zuordnungsregeln präzisiert²¹, während beim Kategoriensystem „Computer als Gegner durchschaut“ neben ähnlichen Präzisierungen²² zusätzlich eine weitere Kategorie eingefügt wurde. In diese Kategorie wurden die Probanden subsumiert, bei denen die Angaben zum Gegner inkonsistent waren. Durch die Einführung der neuen Oberkategorie „Widersprüchliche Aussagen zum Gegner“ konnten in den bereits kodierten Daten 22 Abweichungen aufgelöst werden. Außerdem wurden weitere Fälle gefunden, bei denen diese Kategorie anzuwenden war.

Nach ihrer Erstellung wurden die neuen Kategoriensysteme auf alle Datensätze angewendet. Die Kodierung der 93 letzten Datensätze ergab $\kappa = 0.73$ für das Kategoriensystem „Aggression durchschaut“, das als „gut“ bewertet werden kann, bzw. $\kappa = 0.84$ für das Kategoriensystem „Computergegner durchschaut“, welches als „sehr gut“ bewertet werden kann. Es resultierte für jeden Probanden ein Kode für die Störvariable „Aggression durchschaut“ sowie für die Störvariable „Computergegner durchschaut“.

Virtuelle Nothilfe

Tabelle 5: Häufigkeitsverteilung im Kategoriensystem „Aggression durchschaut“

Kode	Aggression durchschaut	Häufigkeit
11	Keine Aussagen/ Keine Ahnung	21
21	Unspezifisch (Vor CRTT)	10
22	Unspezifisch (Während CRTT)	4
23	Unspezifisch (Nach CRTT)	1
24	Unspezifisch (Zeitpunkt unklar)	1
31	Emotionen (Vor CRTT)	17
32	Emotionen (Während CRTT)	4
33	Emotionen (Nach CRTT)	7
34	Emotionen (Zeitpunkt unklar)	3
41	Aggression (Vor CRTT)	17
42	Aggression (Während CRTT)	0
43	Aggression (Nach CRTT)	8
44	Aggression (Zeitpunkt unklar)	2
51	CRTT Aggressionsmaß (Vor CRTT)	59
52	CRTT Aggressionsmaß (Während CRTT)	17
53	CRTT Aggressionsmaß (Nach CRTT)	15
54	CRTT Aggressionsmaß (Zeitpunkt unklar)	0
61	Widersprüche	0
71	Sonstiges (Vor CRTT)	0
72	Sonstiges (Während CRTT)	0
73	Sonstiges (Nach CRTT)	0
74	Sonstiges (Zeitpunkt unklar)	0
	Summe	186

Tabelle 6: Häufigkeitsverteilung im Kategoriensystem „Computergegner durchschaut“

Kode	Computergegner durchschaut	Häufigkeit
11	Keine Aussagen	2
21	Proband ist Gegner	93
22	Proband ist nicht Gegner (Nach Aufklärung)	28
31	Jemand anderes ist Gegner (Vor CRTT)	0
32	Jemand anderes ist Gegner (Während CRTT)	0
33	Jemand anderes ist Gegner (Nach CRTT)	0
34	Jemand anderes ist Gegner (Zeitpunkt unklar)	0
41	Computer ist Gegner (Vor CRTT)	0
42	Computer ist Gegner (Während CRTT)	18
43	Computer ist Gegner (Nach CRTT)	3
44	Computer ist Gegner (Zeitpunkt unklar)	0
51	Widersprüche (Gegner existiert nicht)	38
52	Widersprüche (Gegner ist Konföderierter)	4
61	Sonstiges (Vor CRTT)	0
62	Sonstiges (Während CRTT)	0
63	Sonstiges (Nach CRTT)	0
64	Sonstiges (Zeitpunkt unklar)	0
	Summe	186

Auffällig ist, dass die Kategorie „Computer ist Gegner (Vor CRTT)“ gar nicht besetzt ist. Dies liegt wahrscheinlich daran, dass es keine direkte Frage danach gab, wann die Probanden diese Täuschung durchschaut haben. Solange ein Proband nicht von sich aus angab, dies von Anfang an vermutet zu haben, wurde angenommen, dass er dies erst während des CRTT erkannt hatte. Diese Annahme lässt sich mangels Daten aber nicht überprüfen.

7.3.10.3 BILDUNG DER VERDACHTSBILDUNGS-VARIABLEN

Um den Einfluss des Durchschauens auf die AVen prüfen zu können, wurden aus den Codes Störvariablen generiert. Dabei wurde zuerst berücksichtigt, dass ein Proband die Täuschung vor dem CRTT erkannt haben muss, damit sie sich auf sein Verhalten im CRTT auswirken kann. Erfolgte das Erkennen der Täuschung erst nach der ersten Runde des CRTT, so war dies für die Messung unerheblich, da das aggressive Verhalten nicht als die Summe der Einstellungen im CRTT, sondern lediglich als die Einstellung in der ersten Runde des CRTT operationalisiert wurde. Leider war es nicht möglich, sicher festzustellen, ab wann die Täuschung durchschaut wurde. Zum einen wurde nicht direkt nach dem Zeitpunkt gefragt, sodass bei einer fehlenden Zeitangabe von einem Durchschauen nach dem CRTT ausgegangen wurde. Aber selbst wenn der Proband eine Angabe zum Zeitpunkt gemacht hatte, so bestand die Möglichkeit einer verzerrten Erinnerung. Daneben konnten alle Aussagen von sozialer Erwünschtheit, Konformität (engl. „compliance“) etc. verfälscht sein. Außerdem wurde den Probanden das Erkennen des Untersuchungsziels bzw. der Aggressionsmessung dadurch erleichtert, dass am Ende des Experiments nach aggressiv machenden Spielen gefragt wurde. Die Aussagen der Probanden sollten daher mit Vorsicht interpretiert werden.

Trotz der genannten Probleme wurde entschieden, folgende Codes als Durchschauen der Täuschung anzusehen: für die Variable „Aggression durchschaut“ der Kode 51 (das Reaktionszeit-Spiel soll Aggression messen), für die Variable „Computergegner durchschaut“ die Codes 31 (Jemand anderes ist Gegner) und 41 (Computer ist Gegner). Nach dieser Kodierung hatten $n = 59$ Probanden bereits vor der eigentlichen Durchführung des CRTT erkannt, dass es Aggression messen soll, während kein einziger Proband schon vor dem CRTT erkannt hatte, dass der Gegenspieler nicht echt war. Nun lässt sich einwenden, dass es für die Beeinflussung des Verhaltens im CRTT schon ausreicht, wenn ein Proband erkannt hat, dass es im Experiment um Aggression geht, ohne dies auf das CRTT zu beziehen. Daher wurde eine weitere Version der Störvariable erstellt, indem für „Aggression durchschaut“ der Kode 41 (das Experiment soll Aggression untersuchen) ebenfalls als Durchschauen gewertet wurde. In dieser Version hatten $n = 76$ Probanden schon vor dem CRTT erkannt, dass es um Aggression geht. Auch beim Kategoriensystem „Computergegner durchschaut“ lässt sich ein vergleichbarer Einwand führen. In der Kategorie „Widersprüchliche Aussagen zum Gegner“ fehlt nämlich die Information, ab wann die Probanden zu der Ansicht gelangt sind, dass es sich beim Gegner nicht um einen Menschen bzw. um einen Konföderierten gehandelt hat. Möglicherweise haben die Probanden diese Einsicht schon vor dem CRTT gehabt und den Gegenspieler

nur deswegen im Anschluss an das CRTT wie einen Menschen beschrieben, (a) weil sie dazu aufgefordert wurden (engl. „compliance“), (b) weil sie meinten, dass dies von ihnen erwartet wird (engl. „demand characteristics²³“), (c) weil sie das Experiment nicht stören wollten (Sozialpsychologie des Experiments), (d) weil sie sich nicht sicher genug waren etc. Diese Vermutungen erscheinen zwar plausibel, sie lassen sich aber nachträglich nicht überprüfen. Um zumindest feststellen zu können, ob sich bei diesen Probanden eine Auswirkung auf ihr Gewaltverhalten ergeben hat, wurde auch in diesem Fall eine weitere Version der Störvariable erstellt, indem zusätzlich die Kodes 51 und 52 (Widersprüchliche Aussagen zum Gegner) als Durchschauen gewertet wurden. In dieser Version hatten $n = 42$ Probanden „schon vor dem CRTT“ erkannt, dass sie nicht gegen einen echten Gegner angetreten sind.

7.3.11 NICHTBEFOLGUNG

Nichtbefolgung kann daraus resultieren, dass die Versuchsperson das Experiment sabotieren will (die sog. „negativistische Versuchsperson“ in der Sozialpsychologie des Experiments), weil der Versuchsperson langweilig ist und sie deswegen interessanteren Tätigkeiten nachgeht, oder weil der Versuchsperson die gestellten Aufgaben/Fragen unangenehm sind. Bei einer Nichtbefolgung zu Sabotagezwecken liegt der besondere Fall vor, dass die Versuchsperson höchstwahrscheinlich keine ehrliche Antwort darauf geben wird, ob sie sich an die Instruktionen gehalten hat. Aus dem Grund wurde zur Erhebung der Nichtbefolgung lediglich gegen Ende der Untersuchung danach gefragt, ob die Versuchsperson die Lust zur Teilnahme verloren hat und ob sie einige Fragen zu unangenehm zu beantworten fand (siehe Anhang H).

7.3.12 ZEITDRUCK

Unter Zeitdruck verringert sich die Hilfsbereitschaft (Heckhausen, 1989, S. 295). Eine Erklärung dafür liefert eine Studie von Berkowitz (1970, zit. n. Heckhausen, 1989, S. 304), nach der Beschäftigungsdruck und Selbstbefangenheit beim Erledigen persönlicher Anliegen vorübergehend zu einem starken Nachlassen empathischer Einfühlung führen. Daher wurde erfasst, wie viel Zeit einem Probanden für das Bearbeiten derjenigen Aufgabe blieb, mit welcher seine Hilfsbereitschaft erfasst werden sollte. Dazu wurde vom Versuchsleiter schriftlich die Uhrzeit des Untersuchungsbeginns festgehalten. Am Ende des elektronischen Fragebogens wurde automatisch ein Zeitstempel erstellt, der den Beginn der „Konzentrationsaufgabe“ markiert. Der Zeitraum zwischen diesen beiden Zeitpunkten wurde von der gesamten Untersuchungszeit (zwei Stunden) abgezogen, um so die Zeit zu ermitteln, die dem Probanden für die Durchführung der „Konzentrationsaufgabe“ verblieb. Diese Zeit bildet die Störvariable „Restzeit für Hilfeverhalten“.

7.3.13 SCHULFORM

Da aus praktischen Gründen nicht nur Studierende der Universität Osnabrück, sondern auch der Fachhochschule Osnabrück teilgenommen haben, wurde diese Variation als Störvariable aufgenommen (siehe Anhang F).

7.3.14 ALTER UND FACHSEMESTER

Um sicherzugehen, dass sich die Studierenden nicht zu stark hinsichtlich ihres Alters und Fachsemesters unterscheiden, wurden diese beiden Merkmale ebenfalls erfasst (siehe Anhang F).

7.3.15 ERHEBUNGSWELLE

Da die Datenerhebung zwei Semester lang stattfand, kann nicht ausgeschlossen werden, dass das zwischenzeitliche Geschehen zwischen den Semestern einen Unterschied zwischen den Erhebungswellen verursacht hat, z. B. weil sich Informationen zum Experiment herumgesprochen haben. Daher wurde die Erhebungswelle als Störvariable aufgenommen.

7.3.16 EMOTIONEN

Zur Operationalisierung eines breiten Spektrums an Emotionen gibt es nur wenige etablierte Verfahren. Der bei Computerspieluntersuchungen häufiger eingesetzte PANAS (Watson, Clark & Tellegen, 1998; deutsche Version: Krohne, Egloff, Kohlmann & Tausch, 1996; Egloff, Schmukle, Burns & Schwerdtfeger, 2006) kam hier nicht infrage, da er zwar verschiedene Emotionen in Form von Adjektiven abfragt, diese aber lediglich als Grundlage für die Skalen „positiver Affekt“ (PA) und „negativer Affekt“ (NA) dienen. Daher fiel die Entscheidung zugunsten des EMO16 (Schmidt-Atzert & Hüppe, 1996), der sechzehn voneinander abgrenzbare Emotionen erfasst. Um möglichst alle spielrelevanten Emotionen zu erfassen, wurde der EMO16 um vier weitere Emotionen aus dem Struktur-Modell von Mees (1991) ergänzt, nämlich „Leid“, „Mitfreude“, „Schadenfreude“ und „Hoffnung“. Außerdem wurde die Emotion „Sexuelle Erregung“ gestrichen, um die Versuchspersonen nicht unnötig zu irritieren.

8 EMPIRISCH-INHALTLICHE HYPOTHESEN

Da eine Entscheidung über alle zu verwendenden Operationalisierungen gefallen ist, lassen sich nun die theoretisch-inhaltlichen Hypothesen weiter auf die Ebene der empirisch-inhaltlichen Hypothesen ableiten:

8.1 GEWALTVERHALTEN

TIH₁: Reines Gewaltverhalten (Töten) im Computerspiel erhöht das Gewaltverhalten.

EIH₁: Wenn Personen im Computerspiel Oblivion Gegner töten müssen, ohne dabei in irgendeiner Weise Hilfe auszuüben, dann werden sie im Anschluss im Competitive Reaction Time Task intensivere Rauschtöne einstellen als Personen, die in Oblivion keine Gegner töten müssen.

TIH₂: Gewalttätiges Hilfeverhalten (Nothilfe) erhöht das Gewaltverhalten.

EIH₂: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, indem sie Gegner töten, dann werden sie im Anschluss im Competitive Reaction Time Task intensivere Rauschtöne einstellen als Personen, die in Oblivion keine Gegner töten müssen.

TIH₃: Reines Gewaltverhalten (Töten) im Computerspiel erhöht das Gewaltverhalten stärker als gewalttätiges Hilfeverhalten (Nothilfe).

EIH₃: Wenn Personen im Computerspiel Oblivion Gegner töten müssen, ohne dabei in irgendeiner Weise Hilfe auszuüben, dann werden sie im Anschluss im Competitive Reaction Time Task intensivere Rauschtöne einstellen als Personen, die Gegner töten müssen, um eine hilfebedürftige Spielfigur zu retten.

8.2 HILFEVERHALTEN

TIH₄: Reines Hilfeverhalten (Hilfe) im Computerspiel erhöht das Hilfeverhalten.

EIH₄: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, ohne dabei in irgendeiner Weise Gewalt anzuwenden, dann werden sie im Anschluss freiwillig mehr Eckenzahl-Aufgaben für einen hilfebedürftigen Diplomanden lösen wollen als Personen, die in Oblivion keine hilfebedürftige Spielfigur retten müssen.

TIH₅: Gewalttätiges Hilfeverhalten (Nothilfe) erhöht das Hilfeverhalten.

EIH₅: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, indem sie alle Gegner töten, dann werden sie im Anschluss freiwillig mehr Eckenzahl-Aufgaben für einen hilfebedürftigen Diplomanden lösen wollen als Personen, die in Oblivion keine hilfebedürftige Spielfigur retten müssen.

TIH₆: Reines Hilfeverhalten (Hilfe) im Computerspiel erhöht das Hilfeverhalten stärker als gewalttätiges Hilfeverhalten (Nothilfe).

EIH₆: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, ohne dabei in irgendeiner Weise Gewalt anzuwenden, dann werden sie im Anschluss freiwillig mehr Eckenzahl-Aufgaben für einen hilfebedürftigen Diplomanden lösen wollen als Personen, die Gegner töten müssen, um eine hilfebedürftige Spielfigur zu retten.

9 DURCHFÜHRUNG

Nach der Entscheidung über die Operationalisierungen und der Ableitung der Hypothesen auf die Ebene der empirisch-inhaltlichen Hypothesen wurden mehrere Vortests vor der eigentlichen Untersuchung durchgeführt, mit denen Material, Versuchsablauf etc. überprüft und im Anschluss optimiert werden sollten. Die Ergebnisse dieser Vortests wurden bereits überall dort erwähnt, wo sie sich auf Material, Ablauf etc. ausgewirkt haben. Im Folgenden soll daher nur das Vorgehen bei den Vortests und darauf folgend die Hauptuntersuchung beschrieben werden.

9.1 VORTESTS

Im ersten Vortest wurde das Tutorial zum Erlernen der Steuerung einem Vortest mit $N = 6$ Versuchspersonen unterzogen, um zu prüfen, ob es verständlich und in allen vier Versuchsbedingungen vergleichbar schwierig, spannend etc. war. Dazu wurde ein Fragebogen entwickelt, mit dem die Probanden das Tutorial einerseits hinsichtlich der Nützlichkeit zum Erlernen der Steuerung und andererseits hinsichtlich potenzieller Störvariablen (Schwierigkeit, Frustration, Unterhaltungswert, Spannungswert, Gewalthaltigkeit) bewerten sollten. Der Fragebogen sollte später in der Hauptuntersuchung für die Bewertung der Spielaufgabe verwendet werden. Der Störvariablenteil des Bogens basierte auf dem „Video Game Rating Sheet“²⁴ (siehe Abschnitt 7.3.5) von Anderson und Ford (1986).

Der zweite Vortest diente der Optimierung von Materialien und Abläufen sowie der Fehler-suche. Insgesamt $N = 8$ Probanden (sieben männliche und ein weiblicher) absolvierten daher den vollständigen Versuchsablauf in den Räumen und mit der Ausrüstung der Hauptuntersuchung. Der Hälfte der Proband/innen war die Forschungsfrage bekannt.

Der dritte Vortest sollte die Qualität der UV-Manipulation sichern. Alle Probanden absolvierten wieder den vollständigen Versuchsablauf in den Räumen und mit der Ausrüstung der Hauptuntersuchung. Von den insgesamt $N = 23$ Probanden waren $n = 14$ männlich und $n = 9$ weiblich. Keinem Probanden war die Forschungsfrage bekannt.

9.2 HAUPTUNTERSUCHUNG

9.2.1 STICHPROBE

Die Hauptuntersuchung sollte mit jungen Erwachsenen durchgeführt werden, weil einerseits bei Kindern und Jugendlichen die Gefahren einer schädlichen Wirkung besonders hoch erscheinen, andererseits aber eine stärker ausgeprägte Fähigkeit zum moralischen Urteilen vorhanden sein muss, die bei Kindern zum Teil fehlt. Um die Probespielzeit gering zu halten und das Problem der Cybersickness zu umgehen (siehe Abschnitt 7.1.1), sollten die Probanden möglichst bereits über Erfahrung mit Computerspielen in der Ego-Perspektive verfügen.

Virtuelle Nothilfe

Eine echte Zufallsstichprobe war praktisch nicht herstellbar, da es keine vollständige Liste aller jungen Erwachsenen gibt, die im Osnabrücker Raum leben und die Steuerung des verwendeten Computerspiels beherrschen. Daher wurden im Rahmen einer Gelegenheitsstichprobe Studierende aller Fachbereiche der Universität sowie der Fachhochschule Osnabrück untersucht. Idealerweise sollten genauso viele Männer wie Frauen teilnehmen, um auf diese Weise die Störvariable „biologisches Geschlecht“ zu kontrollieren. Es war aber zu befürchten, dass sich weniger Frauen als Männer für die Teilnahme anmelden, da das gewählte Computerspiel eher von Männern gespielt wird.

Die erste Erhebungswelle des Experiments fand während des Sommersemesters 2010 vom 18. Mai 2010 bis einschließlich 14. Juli 2010 statt, die zweite im Wintersemester 2010/2011 vom 27. September 2010 bis einschließlich 13. November 2010. Während der ersten Erhebungswelle musste bei 2 weiblichen Probanden das Experiment abgebrochen werden, weil diese die Spielaufgabe nicht lösen konnten. Es wurden insgesamt 126 vollständige Datensätze erhoben. Von diesen mussten 2 eliminiert werden, da in einem Fall Verständnisprobleme seitens eines kaum Deutsch sprechenden Probanden vermutet wurden und in einem anderen Fall ein Proband durch einen Programmfehler die Aufgabe in der Nothilfebedingung lösen konnte, ohne den sechsten Banditen zu bekämpfen. Es verblieben 124 verwertbare Datensätze (100 Probanden und 24 Probandinnen). Da nach der ersten Welle nur 24 Probandinnen teilgenommen hatten, erschien es unrealistisch, in der zweiten Welle auf eine Gesamtanzahl von 97 Probandinnen zu kommen. Daher wurde beschlossen, die Untersuchung nur mit männlichen Probanden durchzuführen und die bereits mit Probandinnen erhobenen Daten in der zweiten Welle noch einmal mit männlichen Probanden nachzuerheben. In Abweichung zur ersten Welle musste bei dieser Erhebung eine andere Person als Versuchsleiter/Konföderierter (siehe Abschnitt 9.2.7) eingesetzt werden. Außerdem wurde die Instruktion zum Reaktionszeitenspiel auf der Basis der ersten Erhebung geringfügig verändert, damit weniger Verständnisschwierigkeiten auftreten sollten.

Während der zweiten Welle musste bei 2 männlichen Probanden das Experiment abgebrochen werden; in dem einen Fall, weil der Proband aufgrund von Sprachschwierigkeiten nicht verstanden hatte, dass nach dem Tutorial die Zeit gemessen werden soll, und in dem anderen Fall, weil der Proband noch vor der Durchführung des CRTT bereits den anschließenden Fragebogen über das CRTT geöffnet hatte. Am Ende der zweiten Welle resultierte ein Gesamtdatensatz mit 28 weiblichen und 186 männlichen Probanden, wobei nur letztere in die statistischen Analysen einfließen.

9.2.2 ZUORDNUNG ZU DEN VERSUCHSBEDINGUNGEN

Idealerweise sollte die Zuordnung zu den Versuchsbedingungen randomisiert erfolgen, um auch unbekannte Störvariablen zu kontrollieren. Dabei sollten sich die Probanden gleichmäßig über die Versuchsbedingungen verteilen. Da die Variablen Gewalt- und Hilfsbereitschaft

Virtuelle Nothilfe

vom Geschlecht beeinflusst werden und gleichzeitig vorhergesehen wurde, dass weniger Frauen als Männer teilnehmen würden, wurde befürchtet, dass sich bei einfacher Randomisierung keine gleiche Verteilung von Männern und Frauen über die Versuchsbedingungen ergeben könnte. Aus dem Grund wurde eine Mischform aus Parallelisieren und Randomisieren verwendet. Für jedes Geschlecht getrennt wurden Blöcke aus jeweils vier Probanden gebildet. Innerhalb eines Blocks erfolgte die Zuordnung zu den vier Versuchsbedingungen randomisiert. Nach Abschluss eines Blocks befand sich so in jeder Bedingung die gleiche Anzahl an Männern bzw. Frauen. Dadurch sollte ein annähernd gleich großes Geschlechterverhältnis zwischen den Versuchsbedingungen erreicht werden.

Um auszuschließen, dass der Versuchsleiter einen Probanden vor dem Spiel hypothesenkonform beeinflussen kann, wurde die Zuordnung zu einer Bedingung uneinsehbar im Spiel vorgenommen. Erst am Ende des Spiels wurden über ein Skript im Spiel die relevanten Statistiken (Versuchsbedingung, Probandennummer, Anzahl der besiegten Banditen sowie Anzahl der Bewusstlosigkeiten) sichtbar gemacht. Die Versuchsbedingung wurde dabei mit Alpha, Beta, Gamma oder Delta bezeichnet, damit die Probanden nicht erkennen konnten, worum es in der Bedingung geht.

Am Ende der Datenerhebung verteilten sich die männlichen Probanden wie folgt auf die Bedingungen: Nothilfe $n = 45$, Hilfe $n = 48$, Töten $n = 45$ und Schatzjagd $n = 48$. Die ungleiche Zellenbesetzung ergab sich aus zwei Gründen. Zum einen wurden nicht verwertbare Datensätze versehentlich überkompensiert, sodass je zwei Probanden zu viel in die Hilfe- und Schatzjagdbedingung gelangt sind. Zum anderen ist die Probandenanzahl ($N = 186$) nicht glatt durch vier teilbar, weswegen in zwei Bedingungen je ein Proband mehr vorkommt, wobei es sich ungünstigerweise wieder um die Hilfe- und Schatzjagdbedingung handelt.

9.2.3 ORT

Die Datenerhebung wurde im Versuchssteuerungslabor der Universität Osnabrück durchgeführt. Um eine gegenseitige Beeinflussung der Probanden untereinander zu vermeiden, wurden ausschließlich Einzelerhebungen durchgeführt, bei denen der Proband allein in einer Kabine saß, die von der Umgebung abgeschottet war. In seltenen Fällen wurden zwei Erhebungen parallel durchgeführt (d. h. in zwei Kabinen, mit zwei Probanden, zwei Versuchsleitern und zwei Konföderierten). Um auch hier auszuschließen, dass sich die Probanden gegenseitig beeinflussen, lag jede Kabine außerhalb der Sicht- und Hörweite der anderen Kabine.

9.2.4 AUFKLÄRUNG, EINVERSTÄNDNIS UND DATENSCHUTZ

Die Teilnahme am Experiment erfolgte freiwillig und nach Aufklärung der Probanden über mögliche Risiken sowie Einholen einer Einverständniserklärung. Den Probanden wurde der Schutz ihrer Daten zugesichert. Die Namen der Versuchspersonen wurden aufgezeichnet,

um im Falle einer Beschwerde die Messdaten der Person zuordnen zu können. Die Liste mit den Namen wurde separat von den Messdaten geführt und in einer verschlüsselten und passwortgeschützten Containerdatei gespeichert, auf die nur der Versuchsleiter Zugriff hatte.

9.2.5 PROBANDENSCHUTZ

Der Schutz der Probanden wurde sichergestellt durch die dem Experiment vorangehende Anamnese, die Überwachung des Versuchs und die Möglichkeit, den Versuch jederzeit abbrechen. Folgende Kriterien galten als Ausschlusskriterien für die Versuchsteilnahme: Tinnitus, Schwerhörigkeit, Geräuschphobie (Knalltrauma), Kreislaufprobleme und kardiovaskuläre Erkrankungen.

9.2.6 TEILNAHMEVERGÜTUNG

Die Probanden erhielten für die Teilnahme entweder zwei Versuchspersonenstunden oder zehn Euro. Zusätzlich konnten sie an einer Verlosung von ca. einhundert verschiedenen Preisen (hauptsächlich gebrauchte Videospiele, CDs und DVDs) teilnehmen.

9.2.7 VERSUCHSABLAUF

Bei Eingang der Anmeldung eines Probanden zum Versuch wurde überprüft, ob der Proband den Konföderierten möglicherweise kennen könnte. In den Fällen, in denen dies befürchtet wurde, tauschte der Versuchsleiter mit dem Konföderierten die Rolle. Dadurch sollte vermieden werden, dass die Probanden sich bei Kenntnis des Gegners im CRTT anders verhalten, z. B. indem sie „Milde“ walten lassen.

Jeder Proband wurde zu Beginn des Experiments in einer schriftlichen Instruktion über seine Rechte und Pflichten sowie die Risiken der Teilnahme aufgeklärt und über den Ablauf des Versuchs informiert. In der Instruktion wurde als Ziel des Experiments angegeben, dass die Frage untersucht würde, wie sich die kognitive und motorische Leistung in Computerspielen auf andere kognitive und motorische Aufgaben sowie die aktuelle Stimmung auswirken würde (ähnlich: Anderson & Dill, 2000, S. 784; Ferguson et al., 2008a, S. 319).

In einer kurzen Voruntersuchung wurde geprüft, ob keine gesundheitlichen Ausschlusskriterien vorliegen und ob das Teilnahmekriterium „Erfahrung mit der Spielsteuerung“ erfüllt war. War das der Fall, wurde die schriftliche Einverständniserklärung eingeholt. Während der Voruntersuchung wurden außerdem die potenziellen Störvariablen „Lärm“, „Versuchsleiter“, „Konföderierter“, „Alter des Probanden“, „Hochschule des Probanden“, „Lehreinheit des Probanden“ und „Studienfach des Probanden“ erfasst (siehe Anhang F).

Nach der Voruntersuchung begann die eigentliche Versuchsdurchführung, die aus mehreren Schritten bestand. Während jedes Schritts verließ der Versuchsleiter den Raum, um die Re-

Virtuelle Nothilfe

aktionen des Probanden nicht zu beeinflussen. Die Probanden wurden instruiert, nach Beendigung eines Schritts den Versuchsleiter über eine Funkklingel davon in Kenntnis zu setzen. Im ersten Schritt wurden die Probanden in die Benutzung des CRTT-Programms eingeführt, wie es bei Durchführung des CRTT üblich ist (z. B. Anderson et al., 2007, S. 65). Danach gaben die Probanden in einem Online-Fragebogen Auskunft über ihren aktuellen emotionalen Zustand (EMO16). Im Anschluss durchliefen sie in Oblivion zuerst ein Training der Steuerung (das sogenannte Tutorial) und mussten direkt danach eine von vier Spielaufgaben lösen. Die Zuordnung eines Probanden zu einer Aufgabe wurde automatisch durch das Spiel anhand einer vorher festgelegten Zufallsverteilung vorgenommen (siehe Abschnitt 9.2.2). Nach dem Spiel absolvierten die Probanden den CRTT. Im Anschluss daran beantworteten sie mehrere Online-Fragebögen (Aggressive Motives Scale, Fragen zum CRTT, EMO16, Fragen zur Spielsteuerung und zur Spielaufgabe inkl. subjektiver Erregung, Media Violence Exposure Scale, Verdachtbildung, Nichtbefolgung). Nach dem Ausfüllen des Fragebogens wurde den Probanden mitgeteilt, dass ein Diplomand sie um einen Gefallen bitten möchte. Sie wurden gebeten, an einem Konzentrationstest teilzunehmen, der in Wirklichkeit das Hilfeverhalten der Probanden erheben sollte. Zum Schluss wurden die Probanden vom Versuchsleiter befragt, ob sie Täuschungsmanöver identifizieren konnten. Die Antworten wurden stichwortartig notiert. Erst danach wurden die Probanden darüber aufgeklärt, dass ihr Gegner im CRTT computerbasiert war. Sie wurden gebeten, diese Information nicht an die Öffentlichkeit zu tragen, bis die Studie beendet ist. Die Probanden wurden darüber informiert, dass eine vollständige Aufklärung erst am Ende der Studie erfolgt, indem die Ziele der Studie sowie die Täuschungsmanöver auf der Homepage veröffentlicht werden. Als Letztes erhielten die Probanden ihre Vergütung.

10 STATISTISCHE VORHERSAGEN

Nach der Erhebung der Daten ist es vor der Durchführung der statistischen Tests sinnvoll, zuerst aus den empirisch-inhaltlichen Hypothesen die folgenden statistischen Vorhersagen abzuleiten:

10.1 GEWALTVERHALTEN

ElH₁: Wenn Personen im Computerspiel Oblivion Gegner töten müssen, ohne dabei in irgendeiner Weise Hilfe auszuüben, dann werden sie im Anschluss im Competitive Reaction Time Task intensivere Rauschtöne einstellen als Personen, die in Oblivion keine Gegner töten müssen.

SV₁: Der Mittelwert der Intensität im CRTT aus Runde 1 ist in der Gruppe „Töten“ größer als in der Gruppe „Hilfe“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

ElH₂: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, indem sie Gegner töten, dann werden sie im Anschluss im Competitive Reaction Time Task intensivere Rauschtöne einstellen als Personen, die in Oblivion keine Gegner töten müssen.

SV₂: Der Mittelwert der Intensität im CRTT aus Runde 1 ist in der Gruppe „Nothilfe“ größer als in der Gruppe „Hilfe“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

ElH₃: Wenn Personen im Computerspiel Oblivion Gegner töten müssen, ohne dabei in irgendeiner Weise Hilfe auszuüben, dann werden sie im Anschluss im Competitive Reaction Time Task intensivere Rauschtöne einstellen als Personen, die Gegner töten müssen, um eine hilfebedürftige Spielfigur zu retten.

SV₃: Der Mittelwert der Intensität im CRTT aus Runde 1 ist in der Gruppe „Töten“ größer als in der Gruppe „Nothilfe“.

10.2 HILFEVERHALTEN

ElH₄: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, ohne dabei in irgendeiner Weise Gewalt anzuwenden, dann werden sie im Anschluss freiwillig mehr Eckenzahl-Aufgaben für einen hilfebedürftigen Diplomanden lösen wollen als Personen, die in Oblivion keine hilfebedürftige Spielfigur retten müssen.

SV₄: Der Mittelwert der angegebenen Aufgabenanzahl bei der Eckenzahlaufgabe ist in der Gruppe „Hilfe“ größer als in der Gruppe „Töten“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

ElH₅: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, indem sie alle Gegner töten, dann werden sie im Anschluss freiwillig mehr Eckenzahl-Aufgaben für einen hilfebedürftigen Diplomanden lösen wollen als Personen, die in Oblivion keine hilfebedürftige Spielfigur retten müssen.

SV₅: Der Mittelwert der angegebenen Aufgabenanzahl bei der Eckenzahlaufgabe ist in der Gruppe „Nothilfe“ größer als in der Gruppe „Töten“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

ElH₆: Wenn Personen im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten müssen, ohne dabei in irgendeiner Weise Gewalt anzuwenden, dann werden sie im Anschluss freiwillig mehr Eckenzahl-Aufgaben für einen hilfebedürftigen Diplomanden lösen wollen als Personen, die Gegner töten müssen, um eine hilfebedürftige Spielfigur zu retten.

SV₆: Der Mittelwert der angegebenen Aufgabenanzahl bei der Eckenzahlaufgabe ist in der Gruppe „Hilfe“ größer als in der Gruppe „Nothilfe“.

11 AUSWERTUNG

Um die statistischen Vorhersagen zu prüfen, muss als Nächstes die Art der Auswertung bestimmt werden. Um eine möglichst hohe Teststärke zu erhalten, wurden zur Testung der Hypothesen mehrere geplante Einzelvergleiche durchgeführt. Bei allen gerichteten Hypothesen wurde dabei einseitig getestet. Daran schloss sich eine Datenexploration in Form von ANOVAs und ANCOVAs an. Vor der eigentlichen Testung wurde die obligatorische Prüfung der Datenqualität durchgeführt. Im Folgenden finden sich Informationen zur realisierten Stichprobe, zur Datenqualitätsprüfung, zum Erfolg der experimentellen Manipulation, zur Auswahl der statistischen Tests sowie zur eigentlichen Hypothesenprüfung.

11.1 STICHPROBENBESCHREIBUNG

Die Stichprobe bestand aus $N = 186$ männlichen Studierenden, von denen $n = 102$ im Sommersemester 2010 und die übrigen $n = 84$ im Wintersemester 2010/2011 teilnahmen. An der Universität Osnabrück studierten $n = 139$ Probanden, an der Fachhochschule Osnabrück $n = 47$. Zu dem Zeitpunkt betrug ihr Fachsemester im Durchschnitt $M = 4.36$ Semester ($SD = 3.47$, $Mdn = 3$, Min. = 1, Max. = 19) und ihr Alter $M = 23.65$ Jahre ($SD = 3.31$, $Mdn = 23$, Min. = 18, Max. = 41).

Die Probanden verteilten sich wie folgt auf die Versuchsbedingungen: Nothilfe $n = 45$, Hilfe $n = 48$, Töten $n = 45$, Schatzjagd $n = 48$. Das Durchlaufen des Tutorials dauerte im Durchschnitt $M = 3.63$ Minuten ($SD = 1.27$, Min. = 1.73, Max. = 10.32). Nach dem Spiel sollten die Probanden auf einer Skala von „1 = völlig falsch“ bis „5 = völlig richtig“ angeben, ob sie nach dem Tutorial sicher im Umgang mit der Steuerung waren. Der Skalenwert lag im Mittel bei $M = 4.00$ ($SD = 0.88$). Das Durchlaufen der eigentlichen Spielaufgabe dauerte durchschnittlich $M = 22.39$ Minuten ($SD = 7.02$, Min. = 8.47, Max. = 54.13). Dabei wurde die Spielfigur im Schnitt $M = 5.49$ Mal ($SD = 4.29$, Min. = 1, Max. = 26) ohnmächtig. Diese Werte weisen darauf hin, dass trotz Rückgriff auf erfahrene Computerspieler und Vorschaltung eines Tutorials zwischen den Probanden immer noch sichtbare Fähigkeitsunterschiede vorhanden waren.

11.2 DATENPRÜFUNG

Vor der Datenauswertung wurden die Daten überprüft (vgl. Tabachnick & Fidell, 2007, Kap. 4). Der erste Schritt bestand darin, herauszufinden, ob alle Werte innerhalb eines plausiblen Bereichs lagen. Gefundene Fehler wurden korrigiert.

11.2.1 FEHLENDE WERTE

Der nächste von den genannten Autoren empfohlene Schritt ist die Analyse fehlender Daten, die hier für die Prüfung der Hypothesen keinen Sinn macht, da für alle Probanden zwangsweise vollständige Werte für alle (un-)abhängigen Variablen vorliegen. Sogar bei den Störva-

Virtuelle Nothilfe

riablen sind die meisten Datensätze vollständig; in vereinzelten Fällen fehlen bei einer Variable bis zu 4%. Die fehlenden Werte ergeben sich dadurch, dass eine Versuchsperson im Fragebogen eine Frage mit „nicht sinnvoll beantwortbar“ beantworten konnte. Aus praktischen Gründen wurden diese fehlenden Werte für die weiteren Analysen durch EM-Schätzer ersetzt.

11.2.2 AUSREIßER

Für jede der beiden AVen wurde eine univariate Ausreißeranalyse durchgeführt, indem geprüft wurde, ob der z -standardisierte Wert $z = 3.29$ übersteigt (siehe Tabachnick & Fidell, 2007, S. 72 ff.). Bei beiden AVen wurden jeweils drei Ausreißer diagnostiziert (siehe Tabelle 7).

Tabelle 7: Anzahl der Ausreißer in den Versuchsbedingungen

AV	Versuchsbedingung			
	Nothilfe ($n=45$)	Hilfe ($n=48$)	Töten ($n=45$)	Schatzjagd ($n=48$)
Hilfeverhalten	1	1	1	0
Gewaltverhalten	2	0	1	0

Beim Gewaltverhalten hatten drei Probanden als Lautstärke den Maximalwert 10 eingestellt (siehe Abb. 40), während beim Hilfeverhalten drei Probanden angegeben hatten, 100 Aufgaben lösen zu wollen (siehe Abb. 41).

Es wurde überprüft, ob diese Ausreißer auf eine fehlerhafte Datenerfassung oder eine fehlerhafte Spezifizierung fehlender Werte zurückzuführen waren, was nicht der Fall war. Tabachnick und Fidell (2007, S.73) empfehlen in solchen Fällen entweder die Entfernung der Ausreißer aus dem Datensatz oder die Verminderung des Messwerts. Ersteres sei dann anzuwenden, wenn anzunehmen ist, dass der Ausreißer nicht aus der Grundgesamtheit stammt; Letzteres sei anzuwenden, wenn anzunehmen ist, dass der Ausreißer aus einer Grundgesamtheit stammt, in der die Werte extremer verteilt sind als in einer Normalverteilung. Es gab keinen Grund für die Annahme, dass die Ausreißer nicht aus der Grundgesamtheit stammen. Eine Elimination hätte außerdem zu neuen Problemen führen können, da die Ausreißer beim Gewaltverhalten immer in den gewalthaltigen Bedingungen lagen und die Ausreißer beim Hilfeverhalten in zwei von drei Fällen in den hilfehaltigen Bedingungen, wodurch die Elimination den Zusammenhang zwischen UVen und AVen stärker beeinflusst hätte. Daher wurden die Messwerte vermindert, indem derjenige Messwert zugewiesen wurde, der rechnerisch bei einem Wert von $z = 3.29$ zustande kommen würde. Beim Gewaltverhalten wurde der Wert somit von 10 auf 8 reduziert, während er beim Hilfeverhalten von 100 auf 79 gemindert wurde.

Virtuelle Nothilfe

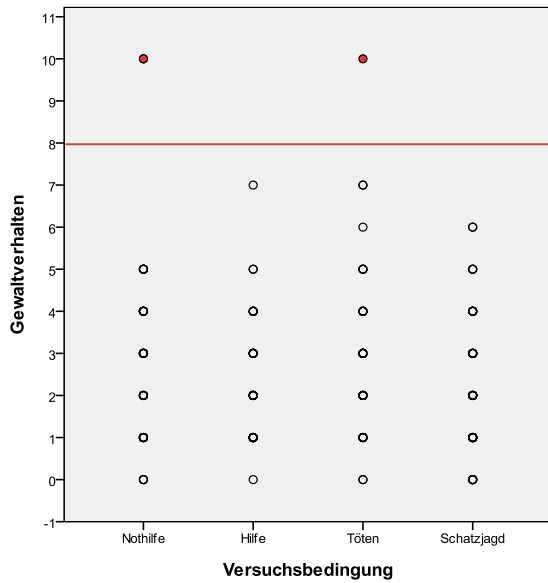


Abb. 40: Ausreißer beim Gewaltverhalten

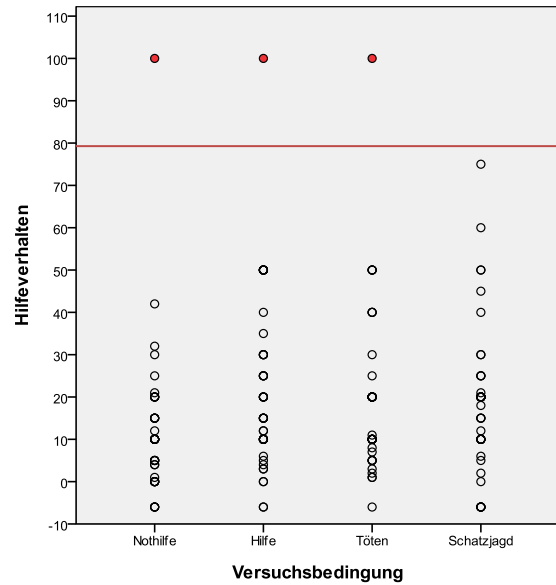


Abb. 41: Ausreißer beim Hilfeverhalten

11.3 VALIDITÄT DER UV

Während des Experiments wurden verschiedene Maßnahmen realisiert, mit denen überprüft werden sollte, ob die experimentelle Manipulation erfolgreich war. Hierzu gehörte die Zählung der getöteten Banditen, die Messung der Tutorial- und Aufgabendauer sowie die subjektive Bewertung der Gewalt- und Hilfehaltigkeit der Aufgabe.

11.3.1 ANZAHL GETÖTETER BANDITEN

Im Spiel wurde automatisch im Hintergrund die Anzahl der Banditen erfasst, die vom Spieler getötet wurden. Die Anzahl konnte am Ende des Spiels vom Versuchsleiter eingesehen werden und wurde notiert. Eine Prüfung der Daten zeigte das zu erwartende Ergebnis, dass die Spieler in den gewalthaltigen Bedingungen immer sechs Banditen getötet haben, während sie in den gewaltfreien Bedingungen keinen einzigen getötet haben²⁵.

11.3.2 TUTORIALDAUER

Es lässt sich darüber streiten, ob auch schon das Tutorial zum Treatment dazugezählt werden sollte oder nicht. Sicherheitshalber wurde daher auch die Dauer des Tutorials erfasst (siehe Tabelle 8). Um die Probanden in den gewaltfreien Versuchsbedingungen im Tutorial nicht mit Gewalt zu konfrontieren, enthielt das Tutorial nur in den gewalthaltigen Bedingungen eine Kampfübung. Dadurch war das Tutorial in diesen Bedingungen prinzipbedingt länger. Die gewalthaltige Nothilfebedingung ($M = 4.26$) unterscheidet sich aber nicht signifikant von der gewalthaltigen Tötenbedingung ($M = 4.57$), $t(88) = 1.128$, $p = .262$, und die gewalt-

Virtuelle Nothilfe

freie Hilfebedingung ($M = 2.94$) unterscheidet sich nicht signifikant von der gewaltfreien Schatzjagdbedingung ($M = 2.86$), $t(94) = 0.624$, $p = .534$.

Tabelle 8: Für die Lösung des Tutorials benötigte Zeit in Minuten (Hauptuntersuchung)

Spielzeit	Versuchsbedingung			
	Nothilfe ($n=45$)	Hilfe ($n=48$)	Töten ($n=45$)	Schatzjagd ($n=48$)
<i>M</i>	4.26	2.94	4.57	2.86
<i>SD</i>	1.10	0.60	1.53	0.61
<i>Min</i>	2.45	1.73	2.82	1.85
<i>Max</i>	8.57	4.40	10.32	4.43

11.3.3 AUFGABENDAUER

Die Dauer für die Lösung der Spielaufgabe stellt in diesem Experiment die Dauer des Treatments dar. In den üblichen Experimenten spielen die Probanden das Computerspiel für eine konstante Zeit. Dies war hier nicht möglich, da die Probanden die Spielaufgabe lösen mussten und die Spielzeit somit mit der Lösung der Aufgabe endete. Dadurch konnte die Spielzeit nicht konstant gehalten werden. Aus dem Grund wurde sie erfasst, um ihre Auswirkung prüfen und ggf. statistisch kontrollieren zu können. Im Idealfall sollte die Lösungszeit in allen Versuchsbedingungen gleich lang sein. Dies war leider nicht ganz der Fall, wie sich in Tabelle 9 ablesen lässt. Während sich Mittelwert, Minimum, Maximum und Schiefe zwischen der Nothilfe-, Hilfe- und Gewaltbedingung nur geringfügig unterscheiden, fällt die Schatzjagdbedingung aus dem Rahmen. Ein Post-Hoc-Mittelwertvergleich (LSD) lässt erkennen, dass sich die Schatzjagdbedingung tatsächlich signifikant von allen anderen Bedingungen unterscheidet ($p < .01$).

Tabelle 9: Für die Lösung der Spielaufgabe benötigte Zeit in Minuten (Hauptuntersuchung, mit Ausreißern)

Spielzeit	Versuchsbedingung			
	Nothilfe ($n=45$)	Hilfe ($n=48$)	Töten ($n=45$)	Schatzjagd ($n=48$)
<i>M</i>	23.47	22.88	23.96	19.42
<i>SD</i>	6.72	7.85	7.71	4.66
<i>Min</i>	13.75	11.33	14.13	8.47
<i>Max</i>	51.88	54.13	53.68	29.50
<i>Schiefe</i>	1.98	1.79	2.05	0.56

Dieser Unterschied wird durch Ausreißer verursacht. Wie sich in Abb. 42 sehen lässt, liegt die Spielzeit in den meisten Fällen zwischen 10 und 35 Minuten. In den ersten drei Gruppen finden sich aber auch jeweils zwei bis drei Probanden mit einer Spielzeit von über 40 Minuten (siehe Abb. 42), während in der Schatzjagdbedingung ein Proband mit einer besonders kurzen Spielzeit von unter 10 Minuten zu finden ist.

Virtuelle Nothilfe

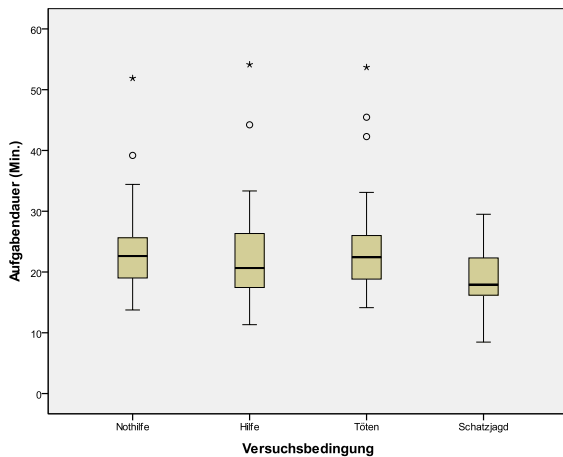


Abb. 42: Box-Plot zu Ausreißern bei der Lösungszeit

Hinsichtlich der Spielzeit können diese Probanden als Ausreißer angesehen werden. Durch eine Elimination würden sich Maximum, Minimum und Schiefe der Versuchsbedingungen einander angleichen (siehe Tabelle 10)

Tabelle 10). Der Unterschied zwischen der Schatzjagdbedingung und den anderen Bedingungen wäre aber weiterhin signifikant ($p < .05$). Eine Elimination würde außerdem bei den statistischen Analysen zu einem spürbaren Verlust an Freiheitsgraden führen. Daher wurden die Fälle nicht eliminiert und die Spielzeit stattdessen in den ANCOVAs als Störvariable hinzugefügt.

Tabelle 10: Für die Lösung der Spielaufgabe benötigte Zeit (Hauptuntersuchung, ohne Ausreißer)

Spielzeit	Versuchsbedingung			
	Nothilfe (<i>n</i> =43)	Hilfe (<i>n</i> =46)	Töten (<i>n</i> =42)	Schatzjagd (<i>n</i> =47)
<i>M</i>	22.45	21.73	22.31	19.66
<i>SD</i>	4.60	5.58	4.47	4.42
<i>Min</i>	13.75	11.33	14.13	13.82
<i>Max</i>	34.42	33.33	33.10	29.50
<i>Schiefe</i>	0.32	0.48	0.39	0.85

11.3.4 GEWALT- UND HILFEGEHALT DER AUFGABE

Für die Hauptuntersuchung sollte wie schon für die Voruntersuchungen überprüft werden, ob die Manipulation der UVen erfolgreich war. Bei erfolgreicher Manipulation müssten die Probanden die Bedingungen mit Hilfe- bzw. Gewaltanforderungen als solche erkannt haben. Für die Prüfung wurden vier Items herangezogen: das Item zum Gewaltgehalt des Inhalts „Ich finde, dass die Inhalte des Spiels voll von Gewaltverhalten waren“ (Gewalt im Inhalt) aus dem Video Game Rating Sheet (siehe Abschnitt 7.3.5) sowie die dem Sheet hinzugefügten Items zum Gewaltgehalt der Aufgabe „Ich finde, dass die Inhalte des Spiels voll von Gewaltverhalten waren“ (Gewalt in Aufgabe), zum Hilfegehalt des Inhalts „Ich finde, dass die Inhal-

Virtuelle Nothilfe

te des Spiels voll von Hilfeverhalten waren“ (Hilfe im Inhalt) und zum Hilfegehalt der Aufgabe „Ich finde, dass die Spielaufgabe Hilfeverhalten erforderte“ (Hilfe in Aufgabe).

Tabelle 11: Subjektive Bewertung von Hilfe- und Gewalthaltigkeit der Spielaufgabe (Hauptuntersuchung)

Eigenschaft d. Aufgabe	Versuchsbedingung											
	Nothilfe (n=45)			Hilfe (n=48)			Töten (n=45)			Schatzjagd (n=48)		
Hilfe	ja			ja			nein			nein		
Gewalt	ja			nein			ja			nein		
	<i>M</i>	<i>SD</i>	<i>Schiefe</i>	<i>M</i>	<i>SD</i>	<i>Schiefe</i>	<i>M</i>	<i>SD</i>	<i>Schiefe</i>	<i>M</i>	<i>SD</i>	<i>Schiefe</i>
Gewalt in Aufgabe¹	4.02	1.23	-1.18	1.21	0.58	3.34	4.38	0.94	-1.71	1.21	0.58	3.34
Gewalt im Inhalt¹	3.27	1.07	-0.22	2.42	0.94	0.17	3.51	1.12	-0.74	2.33	1.04	0.11
Gewalthaltigkeit	3.64	0.84	-0.42	1.81	0.56	0.33	3.94	0.87	-1.11	1.77	0.58	0.23
Hilfe in Aufgabe¹	3.10	1.32	-0.07	3.54	1.32	-0.70	2.62	1.28	0.44	2.60	1.28	0.16
Hilfe im Inhalt¹	2.64	0.75	-0.01	2.85	0.99	0.03	1.92	0.89	0.48	2.29	1.10	0.24
Hilfeshaltigkeit	2.87	0.92	-0.12	3.20	0.99	-0.77	2.27	0.98	0.33	2.45	1.00	0.16

¹ 5-stufige Skala von 1 (völlig falsch) bis 5 (völlig richtig).

Auffällig sind die hohen Schiefe-Werte der Variable „Gewalt in Aufgabe“, die über die Gruppen die Schiefe wechseln (siehe Tabelle 11). In den gewalthaltigen Bedingungen ist sie links-schief (*Schiefe* = -1.18 bzw. *Schiefe* = -1.71), während sie in den gewaltfreien Bedingungen rechtsschief ist (jeweils *Schiefe* = 3.34). Eine mögliche Ursache hierfür könnte ein Decken- bzw. Bodeneffekt sein. Möglicherweise hätten die Probanden in den gewalthaltigen Bedingungen noch höhere Werte vergeben bzw. die Probanden in den gewaltfreien Bedingungen noch niedrigere Werte, wenn die Antwort-Skala mehr als fünf Stufen gehabt hätte.

Für die weiteren Analysen wurde geprüft, ob die Items zu einer Skala zusammengefasst werden können. Dafür spricht zum einen eine explorative Hauptkomponentenanalyse mit Varimax-Rotation, die zwei Faktoren findet (sowohl nach Scree-Plot als auch nach Kaiser-Kriterium), wobei die Hilfeshaltigkeits-Items auf dem ersten Faktor laden (.88 und .87), die Gewalthaltigkeits-Items auf dem zweiten (.85 und .85), und die Nebenladungen alle vom Betrag geringer als .16 ausfallen. Zum anderen korrelieren die Gewalthaltigkeits-Items ($r = .44$, $p < .01$) signifikant miteinander, ebenso die Hilfeshaltigkeits-Items ($r = .53$, $p < .01$). Daher wurden die Gewalthaltigkeits-Items per arithmetischer Mittelung zu einer Gewalthaltigkeits-skala (Cronbachs $\alpha = .58$) und die Hilfeshaltigkeits-Items zu einer Hilfeshaltigkeitsskala ($\alpha = .68$) zusammengefasst.

Virtuelle Nothilfe

Bei einer erfolgreichen Manipulation sollte sich folgendes Muster ergeben: In den hilfehaltigen Versuchsbedingungen „Nothilfe“ und „Hilfe“ sollte die Hilfehaltigkeit höher eingeschätzt werden als in den hilfefreien Bedingungen „Töten“ und „Schatzjagd“, während die Gewalthaltigkeit in den gewalthaltigen Versuchsbedingungen „Nothilfe“ und „Töten“ höher eingeschätzt werden sollte als in den gewaltfreien „Hilfe“ und „Schatzjagd“. Um zu prüfen, ob dies zutrifft, wurden t -Tests für die Unterschiede hinsichtlich der Aufgaben durchgeführt. Dabei wurde auch die Effektgröße über Hedges g bestimmt (Hedges, 1981). Dieser Parameter ist eine Effektgröße für Mittelwertunterschiede, wobei im Gegensatz zum eng verwandten d von Cohen zur präziseren Schätzung statt der einfachen Streuung die gepoolte Streuung verwendet wird. Nach einer Daumenregel von Cohen (1988, S. 24-27) handelt es sich bei $d = 0.20$ um einen kleinen, bei $d = 0.50$ um einen mittleren und bei $d = 0.80$ um einen großen Effekt. Aus der Berechnung resultierten folgende Ergebnisse:

Die wahrgenommene Gewalthaltigkeit korreliert signifikant mit der UV „Gewalt im Spiel“ ($r = .81, p < .001$). Außerdem ergibt sich ein signifikanter Unterschied zwischen den gewalthaltigen ($M = 3.79$) und den gewaltfreien Bedingungen ($M = 1.79$), $t(153) = 18.58, p < .001, g = 2.76$. Dies entspricht einem großen Effekt. Im Vergleich dazu erreichten Greitemeyer und Osswald (2010) in ihrer Pilotstudie einen noch höheren Wert von $g = 5.41$, während Greitemeyer, Osswald und Brauer (2010, Studie 2) nur auf einen wesentlich geringeren Wert in Höhe von $g = 0.83$ kommen. Hieran erkennt man, wie nützlich es für die Generalisierbarkeit gewesen wäre, wenn die Autoren für jedes ihrer Folgeexperimente einen Wert berichtet hätten.

Die Hilfehaltigkeit korreliert signifikant mit der UV „Hilfe im Spiel“ ($r = .33, p < .001$). Außerdem findet sich ein signifikanter Unterschied zwischen den hilfehaltigen ($M = 3.04$) und den hilfefreien Bedingungen ($M = 2.36$), $t(184) = 4.75, p < .001, g = 0.70$. Hierbei handelt es sich um einen mittleren Effekt. Im Vergleich dazu erreichten fünf andere Studien beim Vergleich zwischen der prosozialen und der neutralen Bedingung (Greitemeyer & Osswald, 2009; 2010; Greitemeyer, Osswald & Brauer, 2010) im Durchschnitt einen größeren Wert in Höhe von $g = 1.18$ (Min. = 0.63; Max. = 1.86). Allerdings wird von den Autoren nicht für jedes Experiment ein eigener Wert berichtet, was erneut die Generalisierbarkeit reduziert.

Nimmt man alle Ergebnisse zusammen, so lässt sich die Manipulation der Gewalthaltigkeit als erfolgreich bewerten, auch wenn mit einer veränderten Antwortskala möglicherweise noch größere Messunterschiede hätten erzielt werden können. Die Manipulation der Hilfehaltigkeit dagegen führt zwar ebenfalls zu signifikanten, aber vergleichsweise kleinen Unterschieden. Eine Ursache für die Schwäche des Treatments könnte darin liegen, dass der Gewaltgehalt des Spiels durch die regelmäßigen Kämpfe salient bleibt, während das Sammeln der Schlüssel aus den Truhen den Hilfegehalt dieser Tätigkeit nicht ständig ins Bewusstsein rückt, sodass der Hilfegehalt nur am Anfang und am Ende der Spielaufgabe präsent ist, aber nicht mitten drin. Der Manipulation-Check (siehe Abschnitt 1.1.1) zeigt aber auch, dass die

Virtuelle Nothilfe

wahrgenommene Hilfehaltigkeit mit $M = 2.6$ in den hilfereien Bedingungen recht hoch ausfällt. Im Vergleich dazu fiel die wahrgenommene Gewalthaltigkeit in den gewaltfreien Bedingungen mit $M = 1.2$ wesentlich niedriger aus. Der geringe Unterschied in der wahrgenommenen Hilfehaltigkeit könnte dazu führen, dass die Wirkung der Hilfehaltigkeit auf das nach dem Spiel gemessene Hilfeverhalten kleiner ausfällt als die Wirkung der Gewalthaltigkeit auf das nach dem Spiel gemessene Gewaltverhalten.

11.4 VALIDITÄT DER AV

Um die Validität der AV Gewaltverhalten sicherzustellen, wurde geprüft, ob beim Aggressionsmaß die definitorisch erforderliche Schädigungsabsicht vorlag. Da während der Untersuchung der Eindruck entstand, dass die Probanden die Täuschungsmanöver im CRTT durchschaut hatten, wurde außerdem überprüft, inwieweit dieser Eindruck zutrifft.

11.4.1 MOTIVE FÜR GEWALTVERHALTEN

Mit der erweiterten Aggressive Motives Scale (AMS; siehe Abschnitt 7.2.1.2) sollte die Schädigungsabsicht im CRTT festgestellt und die Wirkung des Täuschungsmanövers überprüft werden. Der in den Probanden erzeugte Ärger, der im Mittel zwischen „kaum“ und „mittelmäßig“ liegt (siehe Tabelle 12), kann als erfolgreiche Provokation durch den Computergegner gedeutet werden.

Tabelle 12: Mittelwerte der erweiterten Aggressive Motives Scale (Hauptuntersuchung)

Item	Versuchsbedingung							
	Nothilfe		Hilfe		Töten		Schatzjagd	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
CRTT als Reaktionszeittest geeignet	3.38	0.68	3.58	0.85	3.64	0.64	3.42	0.94
Gegner ist schneller	3.39	0.87	3.47	0.68	3.00	0.93	3.43	0.65
Ärger Gegner	2.71	0.84	2.44	1.00	2.76	0.86	2.47	0.85
Ärger Proband	2.40	0.72	2.50	1.01	2.33	0.83	2.65	0.89
An Gegner rächen	2.22	1.22	2.35	1.36	2.38	1.42	2.58	1.43
Gegner verrückt machen	1.76	0.93	1.65	1.02	1.89	1.21	2.02	1.14
Gegner verletzen	1.24	0.74	1.08	0.28	1.32	0.79	1.17	0.63
Gegnerton beeinflussen	3.07	1.27	3.33	1.31	3.18	1.40	3.63	1.36
Gegner behindern	1.64	0.96	1.58	0.99	1.70	1.08	1.67	0.93
Lauter als Gegner	1.44	0.87	1.56	0.97	1.84	1.13	1.89	1.23

5-stufige Skala mit 1 (gar nicht), 2 (kaum), 3 (mittelmäßig), 4 (ziemlich) und 5 (außerordentlich).

Der Wert wäre eventuell höher ausgefallen, wenn der Computer in jeder Runde die maximale Lautstärke appliziert hätte (was aber wahrscheinlich dazu geführt hätte, dass die Probanden schlussfolgern, dass sie nicht gegen einen anderen Probanden spielen). Außerdem scheinen die Probanden die Cover Story akzeptiert zu haben, da sie das CRTT im Durchschnitt für mittelmäßig bis ziemlich geeignet zur Erfassung von Reaktionszeiten halten. Der ausgelöste Ärger und die Bewertung des CRTT als geeignetes Reaktionszeitspiel würden im

Virtuelle Nothilfe

Prinzip für den Erfolg der Täuschungen sprechen, gäbe es nicht weitere schriftliche und mündliche Hinweise der Probanden, die dagegen sprechen (siehe Abschnitt 11.4.2).

Ein weiteres Problem liegt darin, dass die Probanden sich im Durchschnitt „kaum“ am Gegner rächen wollten, ihn „kaum“ verrückt machen wollten und ihn so gut wie „gar nicht“ verletzen wollten, was eher gegen eine starke Schädigungsabsicht der Probanden spricht. Außerdem fallen die Werte der nicht-aggressiven Motive „den Gegner behindern“ und „lauter als der Gegner sein“ im Vergleich dazu höher aus und die des Motivs „die Lautstärke des Gegners beeinflussen“ mit einem Wert zwischen „mittelmäßig“ und „ziemlich“ sogar deutlich höher. Dies spricht dafür, dass die Probanden mehr nicht-aggressive als aggressive Intentionen im CRTT besaßen. Leider ist die AMS nicht geeignet, dies systematisch zu überprüfen, da nur ein Teil der theoretisch möglichen nicht-aggressiven Intentionen erfasst wird.

Um ein Muster in den Intentionen zu identifizieren, wurde eine explorative Hauptkomponentenanalyse mit Varimax-Rotation durchgeführt. Nach dem Kaiser-Kriterium (Eigenwerte größer als 1) ergab sich eine Lösung mit vier Faktoren, während der Scree-Plot für eine Lösung mit einem oder zwei Faktoren (siehe Abb. 43) und die Parallelanalyse nach Horn (durchgeführt mittels Makro von O'Connor, 2000) für eine Lösung mit zwei Faktoren sprach (siehe Abb. 44).

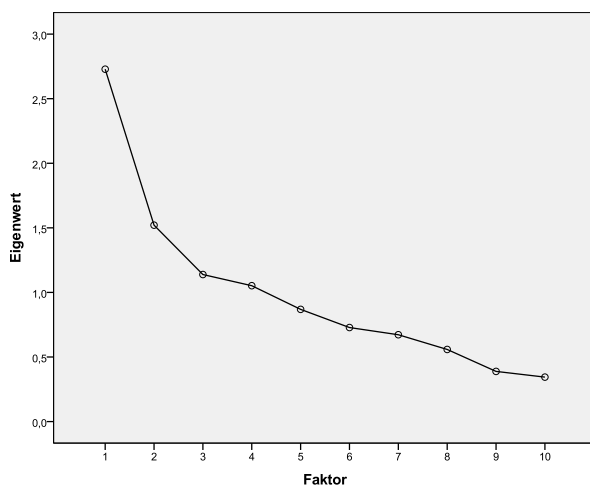


Abb. 43: Scree-Plot zur Faktoranalyse der Intentionen im CRTT

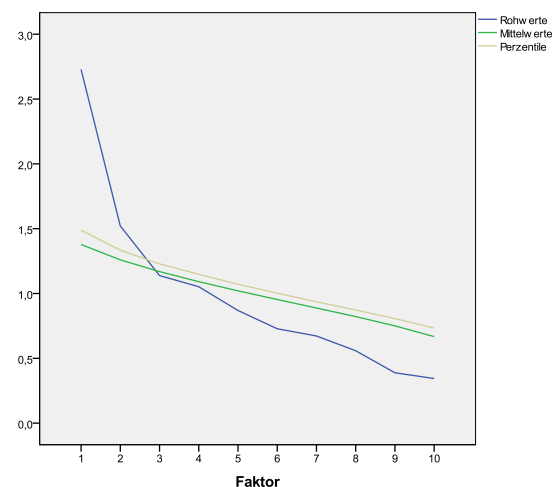


Abb. 44: Parallelanalyse zur Faktorenanalyse der Intentionen im CRTT

Die Analyse des Ladungsmusters spricht hingegen eher für eine Lösung mit drei Faktoren, da der dritte Faktor gut interpretierbar ist und noch 12% der Varianz aufklärt. Auf dem ersten Faktor laden diejenigen drei Items am höchsten, die eine Schädigungsabsicht klar erkennen lassen (siehe Tabelle 13). Hier wollen die Probanden den Gegner verrückt machen, verletzen und behindern. Das dahinterliegende Motiv bleibt aber vage, da nicht erkennbar ist, ob die Probanden ihr Image als harter Kerl aufpolieren wollen, ob es ihnen auch um Rache geht

Virtuelle Nothilfe

oder ob sie gänzlich andere Beweggründe haben. Daher wurde dieser Faktor nicht nach dem Motiv, sondern nach dem gezeigten aggressiven Verhalten benannt. In Anlehnung an Tedeschi und Felson (1994) wurde dafür statt des Begriffs „Aggression“ der neutralere Begriff „Bestrafung“ verwendet (siehe dazu Abschnitt 4.3.3). Auf dem zweiten Faktor laden die Items am höchsten, die etwas mit Ärger und Rache zu tun haben. Hier haben sich die Probanden geärgert und dafür gerächt, indem sie lautere Einstellungen als ihr Gegner gewählt haben. Möglicherweise ist der dem Gegner unterstellte Ärger darauf zurückzuführen, dass die Probanden davon ausgehen, dass ihre Racheaktionen den Gegner geärgert haben. Dieser Faktor lässt sich in Anlehnung an Tedeschi und Felson (1994) als „Ausgleichende Gerechtigkeit“ oder als „Retribution“ bezeichnen. Auf dem dritten Faktor laden die Items am höchsten, die etwas mit der Beeinflussung des Gegners zu tun haben. Hier halten die Probanden den CRTT dazu geeignet, Reaktionszeiten zu erheben. Gleichzeitig glauben sie, dass der Gegner schneller sei als sie selbst. Möglicherweise ist das der Grund dafür, die Lautstärkeeinstellungen des Gegners beeinflussen zu wollen. Zu diesem Zweck stellen die Probanden besonders niedrige Werte ein (was man in Tabelle 13 an dem negativen Vorzeichen des Zusammenhangs zwischen dem Faktor und den CRTT-Werten ablesen kann), in der Hoffnung, dass ihr Gegner auch niedrige Werte einstellt.

Tabelle 13: Faktorenanalyse der erweiterten Aggressive Motives Scale zwecks Prüfung von Motiven im CRTT

Faktor	Dreifaktorenlösung			Zweifaktorenlösung	
	Bestrafung	Retribution	Deeskalation	Bestrafung	Retribution
Varianzaufklärung	21%	21%	12%	20%	22%
Gegner verrückt machen	.77	.31	-.09	.75	.36
Gegner behindern	.74	-.05	-.08	.74	.01
Gegner verletzen	.70	.03	.10	.66	.12
Lauter als Gegner	.59	.47	.23	.50	.57
Gegnerton beeinflussen	-.28	.47	.62	-.43	.56
An Gegner rächen	.23	.81	.15	.14	.84
Ärger Proband	.05	.69	-.20	.04	.63
Ärger Gegner	.08	.63	-.23	.09	.58
CRTT als RZT geeignet	.02	-.08	.69	-.12	.07
Gegner ist schneller	.07	-.16	.36	.00	-.07

Ladungen >.30 auf mehr als einem Faktor sind fett dargestellt.

Dieses Verhalten ähnelt einer umgekehrten Variante der bedingten Drohung (siehe Abschnitt 4.3.3), einer umgekehrten Variante der Retribution bzw. einer Positivvariante der Reziprozität (siehe Abschnitt 4.4.2.2): Im Vergleich zu der Drohung „wenn du hohe Lautstärken einstellst, dann mache ich das auch“ hat sie die Form „stelle ich niedrige Lautstärken ein, dann tue du das auch“. Im Gegensatz zu den anderen beiden Faktoren besteht die Beson-

Virtuelle Nothilfe

derheit darin, dass bei diesem Faktor klar erkennbar ist, dass keine Schädigungsabsicht besteht, sondern im Gegenteil die Schädigung der eigenen Person abgewendet werden soll. Um dieses Ziel zu erreichen, setzen die Probanden nicht auf Nötigung, sondern auf Deeskalation bzw. Beschwichtigung. Daher wurde dieser Faktor als „Deeskalation“ bezeichnet.

Im Anschluss wurde eine Einschluss-Regression mit Gewaltverhalten als AV und den Motiven als Prädiktoren durchgeführt (siehe Tabelle 14). Bei der Interpretation ist zu berücksichtigen, dass die AV Gewaltverhalten nur aus der ersten Runde des CRTT besteht, während sich die Motive auf alle Runden des CRTT beziehen.

Tabelle 14: Multiple Regression der abhängigen Variable „Gewaltverhalten“ auf die Motive im CRTT

Variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	VIF
Bestrafung	.62	.12	.36	.000	1.00
Retribution	.16	.12	.10	.160	1.00
Deeskalation	-.19	.12	-.11	.102	1.00

$R^2 = .15$ ($p = .000$), korrigiertes $R^2 = .14$.

Das Gesamtmodell erklärt 14-15% der Varianz der AV ($p < .01$). Die Motive im CRTT sind somit gut zur Vorhersage der Einstellungen in der ersten Runde des CRTT geeignet. Verwunderlich ist aber, dass es auch einen Zusammenhang zum Retributionsmotiv gibt. Die AV Gewaltverhalten besteht aus den Einstellungen in der ersten Runde des CRTT, aber ein Retributionsmotiv kann sich eigentlich erst nach der ersten Runde heranbilden. Möglicherweise ist die Aussage, sich am Gegner rächen zu wollen, nur eine nachgeschobene Rechtfertigung für das eigene aggressive Verhalten. Anhand der auf dem Faktor ladenden Items lässt sich nämlich ablesen, dass es sich um ein Verhalten mit Schädigungsabsicht handelt. Der Zusammenhang zwischen Gewaltverhalten und den Gewaltmotiven „Bestrafung“ bzw. „Retribution“ scheint damit auf den ersten Blick die Validität des CRTT zu stützen.

Für die Validität des CRTT als Maß für State-Aggression würde sprechen, wenn die reine Gewaltbedingung zu aggressiven Motiven führt und die reine Hilfebedingung zu prosozialen Motiven. Zur Nothilfebedingung ist eine Vorhersage zu diesem Zeitpunkt nicht sinnvoll, da die Forschungsfrage gerade darin besteht, ob Nothilfe eher zu Gewalt oder eher zu Hilfeverhalten führt. Wenn man sich also die Verteilung der Motive auf die Versuchsbedingungen ansieht, dann wäre ein Muster zu erwarten, bei dem die *aggressiven* Motive „Bestrafung“ bzw. „Retribution“ in der Tötenbedingung höher ausfallen als in der Hilfe- und der Schatzjagdbedingung. Das *prosoziale* Motive „Deeskalation“ hingegen sollte in der Hilfebedingung höher ausfallen als in der Töten- und der Schatzjagdbedingung.

Wie sich in Tabelle 15 ablesen lässt, ist dies nicht immer der Fall. Während das Bestrafungsmotiv noch dem vorhergesagten Muster entspricht, fällt das Retributionsmotiv aus dem Rahmen, da es in der Schatzjagdbedingung am höchsten ausfällt. Dies gilt auch für das eher prosoziale Deeskalationsmotiv, das zwar in der Hilfebedingung wie erwartet ausfällt, aber

Virtuelle Nothilfe

auch in der Schatzjagdbedingung einen vergleichbar hohen Wert erzielt. Wenn man die Höhe der Mittelwerte mit Maxima und Minima vergleicht, so sieht man außerdem, dass es sich eher um Trends als große Effekte handelt. Zu diesem Ergebnis kommen auch a priori Kontraste. Wird beim Bestrafungsmotiv die Tötenbedingung mit der Hilfe- und Schatzjagdbedingung kontrastiert, so ist das Ergebnis mit $p = .141$ nicht signifikant. Das gleiche ergibt sich beim Retributionsmotiv mit $p = .405$. Wird beim Deeskalationsmotiv die Hilfebedingung mit der Töten- und der Schatzjagdbedingung kontrastiert, ergibt sich ebenfalls mit $p = .302$ ein nicht signifikantes Ergebnis. Daher ist zu befürchten, dass die von den Probanden verfolgten Motive weniger von der Versuchsbedingung, sondern stärker von ihrer Persönlichkeit abhängen. Dies wäre eine Einschränkung des Einsatzbereichs, da der CRTT somit Traits (z. B. aggressive bzw. prosozialen Persönlichkeit, siehe Abschnitt 4.4.2.2) statt States erheben würde. Eine Überprüfung dieser Vermutung ist in diesem Experiment aber nicht möglich, da aufgrund der Ergebnisse des zweiten Vortests (siehe Abschnitt 9.1) auf den Einsatz entsprechender Persönlichkeitsfragebögen verzichtet wurde.

Tabelle 15: Mittelwerte der Motive in den Versuchsbedingungen

Variable	Min	Max	Nothilfe <i>M (SE)</i>	Hilfe <i>M (SE)</i>	Töten <i>M (SE)</i>	Schatzjagd <i>M (SE)</i>
Bestrafung	-1.29	5.70	-.02 (0.86)	-.15 (0.86)	.14 (1.27)	.04 (0.97)
Retribution	-2.30	2.87	-.12 (0.80)	-.07 (0.97)	.01 (1.04)	.18 (1.16)
Deeskalation	-3.06	2.47	-.21 (0.93)	.13 (1.03)	-.04 (0.93)	.11 (1.08)

11.4.2 VERDACHTSBILDUNG

Täuschungsmanöver, bei denen ein Verdacht gebildet werden konnte, gab es bei beiden AVen. Bei der AV Hilfeverhalten mutmaßte in der Hauptuntersuchung lediglich ein einziger Proband am Ende des Experiments, dass es den vorgegebenen hilfebedürftigen Diplomanden nicht geben könnte. Allerdings konnte der Proband sich keinen Reim darauf machen, was für einen Zweck diese Täuschung gehabt haben könnte. Daher scheint bei dieser AV die Verdachtbildung keine große Rolle gespielt zu haben. Hinsichtlich der AV Gewaltverhalten wurde bereits über eine Inhaltsanalyse (siehe Abschnitt 7.3.10) festgestellt, dass die Probanden die Täuschung in verschiedener Hinsicht durchschaut haben. Hier wurden die Variablen „Aggression durchschaut“ und „Computergegner durchschaut“ gebildet. Um zu prüfen, ob es zu einer Konfundierung dieser Variablen mit den Versuchsbedingungen gekommen sein könnte, wurde als erstes geprüft, wie sich die täuschungsdurchschauenden Probanden auf die Versuchsbedingungen verteilen (siehe Tabelle 16).

Während in jeder Bedingung ungefähr gleich viele Probanden erkannt haben, dass sie nicht gegen einen anderen Probanden antreten, so haben in der Hilfebedingung weniger Probanden erkannt, dass das CRTT ihr Gewaltverhalten messen sollte. Da Gewalt im Allgemeinen ein unerwünschtes Verhalten darstellt, kann angenommen werden, dass Probanden geringere Werte einstellen, wenn sie wissen, dass ihr Gewaltverhalten erfasst werden soll.

Virtuelle Nothilfe

Tabelle 16: Mittelwerte der abhängigen Variablen „Gewaltverhalten“ und „Hilfeverhalten“ in den Versuchsbedingungen im Vergleich zur prozentualen Verteilung der Variablen „Aggression durchschaut“ und „Computergegner durchschaut“

Parameter der Variablen	Versuchsbedingung			
	Nothilfe	Hilfe	Töten	Schatzjagd
<i>M</i> Gewaltverhalten	2.82	2.21	2.96	2.27
<i>M</i> Hilfeverhalten	11.69	22.13	18.91	18.29
% Aggression durchschaut (sicher)	33.3%	22.9%	37.8%	33.3%
% Aggression durchschaut (sicher+möglich)	44.4%	29.2%	48.9%	41.7%
% Computergegner durchschaut (möglich) ¹	22.2%	20.8%	22.2%	25.0%

¹ Die Kategorie „Computergegner durchschaut (sicher)“ fehlt hier, weil nach Kodierung der Inhaltsanalyse kein einziger Proband vor Durchführung des CRTT vermutet hatte, dass der Gegner nicht echt ist.

Es ist daher denkbar, dass der Wert für das Gewaltverhalten in der Hilfebedingung noch niedriger ausgefallen wäre, wenn hier mehr Probanden die Täuschung erkannt hätten. Um diese Vermutung auf Plausibilität zu prüfen, wurden die drei generierten Störvariablen mit beiden AVen korreliert (siehe Tabelle 17). Die Störvariablen wurden dabei so kodiert, dass im Falle des Durchschauens der Wert $i = +1$ kodiert wurde, andernfalls $i = -1$.

Tabelle 17: Korrelationen zwischen den Variablen „Aggression durchschaut“ bzw. „Computergegner durchschaut“ und den abhängigen Variablen „Gewaltverhalten“ bzw. „Hilfeverhalten“

Variable	Korrelationen			
	<i>r</i> Gewalt	<i>p</i> Gewalt	<i>r</i> Hilfe	<i>p</i> Hilfe
% Aggression durchschaut (sicher)	-.11	.122	.03	.682
% Aggression durchschaut (sicher+möglich)	-.06	.427	.04	.624
% Computergegner durchschaut (möglich) ¹	-.14	.059	.12	.104

¹ Die Kategorie „Computergegner durchschaut (sicher)“ fehlt hier, weil nach Kodierung der Inhaltsanalyse kein einziger Proband vor Durchführung des CRTT vermutet hatte, dass der Gegner nicht echt ist.

Es zeigt sich, dass die Korrelationen alle nicht signifikant sind ($p > .05$) und maximal 2% der Varianz einer AV aufklären. Negative Werte der Störvariablen (=Täuschung wurde nicht durchschaut) kovariieren mit Gewaltverhalten und positive Werte mit Hilfeverhalten. Anders ausgedrückt führt das Durchschauen der Täuschungen zu vermindertem Gewaltverhalten bzw. erhöhtem Hilfeverhalten. Daher wurde geprüft, ob es sich hier möglicherweise um Moderatoren handelt. Dies ist nicht der Fall, da beide Variablen keinen signifikanten Zusammenhang zu den UVen aufweisen. Somit liegt auch keine Konfundierung vor. Da die Störvariablen aber einen (wenn auch geringen) Einfluss auf die AVen zu haben scheinen, wurden sie bei der Auswahl der Störvariablen für die ANCOVAS berücksichtigt. Dabei wurde von den beiden Varianten der Variable „Aggression durchschaut“ nur die „sichere“ verwendet, da die Korrelation hier größer ausfiel.

11.5 INTERNE VALIDITÄT

Hinsichtlich der UVen fand sich keine offensichtliche Störungen der internen Validität, wie beispielsweise ein Programmierfehler, durch den in einigen Fällen weniger Banditen zu be-

Virtuelle Nothilfe

seitigen gewesen wären, oder ein Umsetzungsfehler, bei dem das Spiel durch Schummeln hätte verkürzt werden können. Der Einfluss unbekannter Störvariablen sollte eigentlich durch die randomisierte Zuordnung und die dafür ausreichend große Stichprobe kontrolliert werden. Somit könnte man im Prinzip davon ausgehen, dass keine Störung der internen Validität aufgetreten sein kann. Um sicherzugehen, soll trotzdem im Folgenden überprüft werden, ob eine der erfassten Störvariablen konfundiert ist, da jede Konfundierung eine Bedrohung der internen Validität darstellt.

Tabelle 18: Bivariate Korrelationen zwischen den unabhängigen Variablen „Gewalt im Spiel“ bzw. „Hilfe im Spiel“ und den Störvariablen zwecks Prüfung auf Konfundierungen

Störvariablengruppe	Störvariable	UV Gewalt im Spiel	UV Hilfe im Spiel
		<i>r</i>	<i>r</i>
Gewalt-/ Prosozial- spielkonsum	Gewaltspielkonsum	-.02	-.04
	Prosozialspielkonsum	.08	.07
Eigenschaften der Spelaufgabe	Spelaufgabe Interessantheit	-.19**	-.05
	Spelaufgabe Spannung	-.21**	-.01
	Spelaufgabe Unterhaltung	-.02	-.04
	Spelaufgabe Frustration	-.08	.04
	Spelaufgabe Schwierigkeit	-.07	.06
	Spelaufgabe Präsenzgefühl	-.01	.03
	Spelaufgabe Realitätsnähe Grafik	-.06	-.14
	Spelaufgabe Aktivierung (munter)	-.03	-.03
Subjektive Erregung	Spelaufgabe Aktivierung (lustlos)	-.04	.13
	Oblivion Steuerung bekannt	.07	.07
Schwierigkeiten mit Steuerung	Oblivion Steuerung sicher	-.22**	.00
	Aufgabendauer (Minuten)	.18*	.11
Vorkenntnisse mit Computerspielen	Vorkenntnisse mit Computerspielen	-.05	-.08
Verdachtsbildung	Aggressionsmaß durchschaut	.08	-.08
	Computergegner durchschaut	-.01	-.03
Nichtbefolgung	Lust verloren	-.01	.00
Zeitdruck	Restzeit für Hilfeverhalten (Minuten)	-.19**	.02
Schulform	Schulform	.01	-.04
Alter	Alter	.12	-.13
Fachsemester	Fachsemester	.10	-.04
Erhebungswelle	Erhebungswelle	-.10	-.02

* $p < .05$. ** $p < .01$. Da die gewaltfreien (bzw. hilfefreien) Bedingungen mit dem Wert $i = -1$ kodiert wurden, die gewalthaltigen (bzw. hilfehaltigen) hingegen mit $i = +1$, bedeutet ein positiver Zusammenhang, dass die Werte der Variable in den gewalthaltigen (bzw. hilfehaltigen) Bedingungen größer sind als in den gewaltfreien (bzw. hilfefreien).

Von einer Konfundierung spricht man, wenn „die unabhängige Variable mit weiteren Variablen, die die abhängige Variable ebenfalls beeinflussen, *konfundiert* bzw. überlagert ist, so daß letztlich nicht entschieden werden kann, welche Variablen für die Unterschiede in der abhängigen Variablen verantwortlich sind“ (Bortz & Döring, 1995, S. 491). Eine Konfundierung liegt also dann vor, wenn eine Störvariable sowohl mit einer UV und gleichzeitig mit einer AV korreliert ist. Um mögliche Konfundierungen aufzuspüren, wurden die UVen im ers-

Virtuelle Nothilfe

ten Schritt mit allen Störvariablen bivariat korreliert (nur die Störvariable „Fragen unangenehm“ blieb unberücksichtigt, da niemand die Fragen unangenehm fand). Insgesamt gingen 23 Störvariablen in die Analyse ein (siehe Tabelle 18).

Alle signifikanten Korrelationen wurden als potenziell problematisch betrachtet, wobei der α -Fehler für den Signifikanztest auf 5% festgesetzt wurde. Hierbei ist zu berücksichtigen, dass bei diesem Fehlerniveau bei 23 Korrelationen, die auf Signifikanz getestet werden, 1 Test auch dann „zufällig“ signifikant werden kann, wenn es in der Grundgesamtheit keine Unterschiede gibt. Bei der UV „Gewalt im Spiel“ ergaben sich sieben signifikante Korrelationen, bei der UV „Hilfe im Spiel“ keine (Tabelle 18). Eine Konfundierung liegt allerdings nur dann vor, wenn die Störvariable gleichzeitig mit einer UV und einer AV korreliert. Daher wurde im nächsten Schritt geprüft, ob die sieben ermittelten Störvariablen signifikant mit einer der beiden AVen korrelieren (siehe Tabelle 19).

Tabelle 19: Bivariate Korrelationen zwischen der unabhängigen Variable „Gewalt im Spiel“ und den abhängigen Variablen zwecks Prüfung auf Konfundierungen

Störvariable	UV Gewalt im Spiel	AV Gewaltverhalten	AV Hilfeverhalten
	<i>r</i>	<i>r</i>	<i>r</i>
Spielaufgabe Interessantheit	-.19**	-.08	.16*
Spielaufgabe Spannung	-.21**	-.03	.22**
Oblivion Steuerung sicher	-.22**	-.03	.10
Aufgabendauer (Minuten)	.18*	.06	-.10
Restzeit für Hilfeverhalten (Minuten)	-.19**	-.08	-.03

Lediglich zwei der sieben Variablen, die signifikant mit der UV „Gewalt im Spiel“ korrelierten, wiesen auch einen signifikanten Zusammenhang zu einer AV auf, und zwar die Störvariablen „Spielaufgabe Spannung“ und „Spielaufgabe Interessantheit“. Die Probanden in den gewaltfreien Bedingungen fanden die Spielaufgabe interessanter ($r = -.19, p < .01$) und spannender ($r = -.21, p < .01$). Gleichzeitig war das Hilfeverhalten ausgeprägter, wenn die Spielaufgabe als interessanter ($r = .16, p < .05$) bzw. spannender ($r = .22, p < .01$) wahrgenommen wurde. Diese Zusammenhänge könnten eine ernst zu nehmende Störung der internen Validität darstellen. Warum aber finden sich diese Zusammenhänge? Eine interessantere und spannendere Aufgabe könnte möglicherweise zu einer stärkeren Aktivierung des Spielers führen. Tatsächlich korreliert sowohl die Interessantheit mit der Aktivierung ($r = .32, p < .001$ für das Item „munter“ bzw. $r = -.31, p < .001$ für das Item „lustlos“) als auch die Spannung ($r = .51, p < .001$ für das Item „munter“ bzw. $r = -.30, p < .001$ für das Item „lustlos“). Die erhöhte Aktivierung korreliert außerdem signifikant mit dem Hilfeverhalten ($r = .17, p < .05$ für das Item „munter“ bzw. $r = -.18, p < .05$ für das Item „lustlos“), wenn auch die Korrelationen dem Betrag nach klein ausfallen. Somit erhärtet sich der Verdacht, dass der Zusammenhang zwischen Hilfeverhalten und Spannung/Interessantheit von der Munterkeit/Lustlosigkeit mediiert werden könnte. Um dies zu prüfen, wurden zwei Mediatoranaly-

Virtuelle Nothilfe

sen mit Hilfeverhalten als AV durchgeführt. Als UV wurde in dem einen Fall „Spannung“ und im anderen „Interessantheit“ verwendet. Die Variablen „Aktivierung (munter)“ und „Aktivierung (lustlos)“ wurden per arithmetischer Mittelung zu einem einzigen Mediator „Aktivierung“ zusammengefasst. Dies schien gangbar zu sein, denn obwohl die Variablen nur zu $r = .36$, $p < .001$ korreliert sind, so ergibt sich nach der Spearman-Brown-Formel, dass das ursprüngliche Instrument „Perceived Arousal Scale“ mit 24 Items eine Reliabilität in Höhe von $r = .87$ gehabt hätte. Die Analysen wurden mit dem Makro von Preacher und Hayes (2008a, 2008b) durchgeführt. Bei der UV „Interessantheit“ (siehe Abb. 45) waren alle drei Pfade signifikant ($p_a < .001$, $p_b < .05$, $p_c < .05$). Nach Hinzunahme des Mediators war der Pfad von der UV zur AV nicht mehr signifikant ($p_c = .257$). Damit sind alle Bedingungen für das Vorliegen einer vollständigen Mediation erfüllt. Das 95%-Konfidenzintervall $CI = [.226; 2.235]$ eines Bootstrapping-Tests mit 1.000 Proben zeigt, dass der Mediationseffekt signifikant ist ($p < .05$).

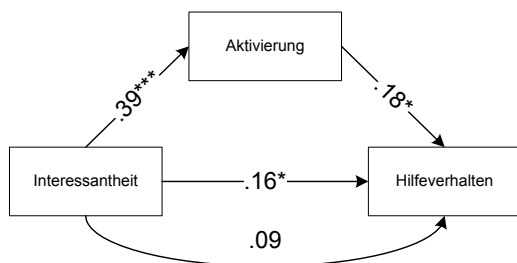


Abb. 45: Mediation von Interessantheit über Aktivierung auf Hilfeverhalten

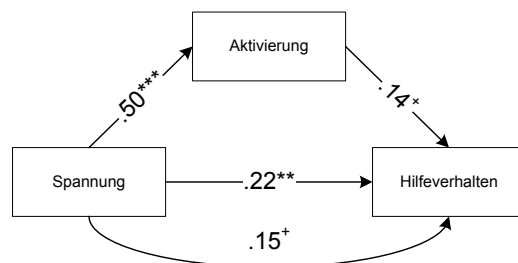


Abb. 46: Mediation von Spannung über Aktivierung auf Hilfeverhalten

Bei der UV „Spannung“ (siehe Abb. 46) hingegen waren nur zwei der drei Pfade signifikant auf dem 5%-Niveau und der dritte auf dem 10%-Niveau ($p_a < .001$, $p_b < .10$, $p_c < .01$). Nach Hinzunahme des Mediators wurde der Zusammenhang zwischen UV und AV zwar geringer, war aber noch auf dem 10%-Niveau signifikant ($p_c < .10$). Das 95%-Konfidenzintervall $CI = [-.008; 2.317]$ des Bootstrapping-Tests zeigte außerdem, dass der Mediationseffekt knapp die Signifikanz verfehlt ($p > .05$). Der Sobel-Test war ebenfalls nur auf dem 10%-Niveau signifikant. Nimmt man diese Ergebnisse zusammen, so scheint ein eindeutiger Mediationseffekt für die UV „Interessantheit“ vorzuliegen, während ein Mediationseffekt für die UV „Spannung“ zwar erkennbar ist, aber höchstens auf dem 10%-Niveau Signifikanz erreicht. Insgesamt hat sich die Vermutung bestätigt, dass die gewaltfreien Spielaufgaben interessanter (spannender) waren, was anscheinend dazu geführt hat, dass die Probanden in diesen Bedingungen aktiver waren. Die erhöhte Aktivierung wiederum hat zwar nicht das Gewaltverhalten, dafür aber das Hilfeverhalten gesteigert. Somit könnte eine bedeutsame Einschränkung der internen Validität vorliegen, da die UV „Gewalt im Spiel“ mit der Störvariable „Spannung“ (und schwächer mit „Interessantheit“) teilweise vermischt ist. Eine abschließende Bewertung dieses Problems wird in Abschnitt 14.2.3 vorgenommen.

11.6 PRÜFUNG DER ANWENDUNGSVORAUSSETZUNGEN

Nach der Analyse der Manipulation Checks und der Konfundierungen bestand der nächste Schritt darin, das Auswertungsverfahren für die eigentliche Hypothesenprüfung auszuwählen. Da das gewählte Verfahren bestimmte Anwendungsvoraussetzungen mitbringt, musste außerdem geprüft werden, ob diese erfüllt sind. Im Anschluss an die Prüfung sollte außerdem festgestellt werden, welchen Einfluss die Störvariablen auf die gefundenen Zusammenhänge haben.

11.6.1 AUSWERTUNGSVERFAHREN

Das zur Hypothesenprüfung notwendige Design beinhaltet zwei UVen und zwei AVen. Damit bietet sich im Prinzip die multivariate Varianzanalyse (MANOVA) als geeignetes Auswertungsverfahren an. Ein Vorteil der MANOVA ist, dass beide AVen zugleich analysiert werden, wohingegen eine Auswertung mit mehreren ANOVAs zu einer Kumulierung des α -Fehlers führen würde und eine mögliche Interaktion zwischen den AVen unberücksichtigt ließe. Allerdings ist die MANOVA weniger gut geeignet, wenn die AVen sehr hoch oder gar nicht korrelieren (Woodward et al., 1990, zit. n. Tabachnick & Fidell, 2007, S. 268). Letzterer Fall ist hier eingetreten, da die AVen lediglich zu $r = 0.08$, $p = .26$ korrelieren. Nach Tabachnick und Fidell (2007, S. 268) hat die MANOVA in solch einem Fall weniger Teststärke als einzeln durchgeführte ANOVAs. Daher empfehlen sie, einzelne ANOVAs zu berechnen und der α -Fehler-Kumulierung durch eine Bonferroni-Korrektur entgegenzuwirken. Da die ANOVA aber nur die ungerichtete Hypothese prüft, dass sich mindestens eine Zelle von mindestens einer anderen unterscheidet, während die zu prüfenden Hypothesen gerichtete Unterschiede zwischen bestimmten Zellen vorhersagen, wäre im Anschluss an die ANOVA ein multipler Mittelwertvergleich durchzuführen. Die dafür zur Verfügung stehenden a posteriori Tests besitzen aber weniger Teststärke als sogenannte „a priori Einzelvergleiche“ (auch „a priori Kontraste“ genannt). Daher wurden die Hypothesen direkt anhand solcher Einzelvergleiche geprüft. Normalerweise wäre auch hier eine Bonferroni-Korrektur notwendig gewesen. Eine solche Korrektur ist nach Bortz (1999, S. 262) aber dann nicht nötig, wenn es sich um „a priori formulierte Einzelvergleichshypothesen [handelt], die theoretisch gut begründet sind, oder die aufgrund von Vorversuchen aufgestellt werden konnten“. Da dies gegeben ist, wurde auf eine Korrektur des α -Fehler-Niveaus verzichtet, um die Power für die Testungen hoch zu halten.

11.6.2 VARIANZHOMOGENITÄT

Da alle vier Versuchsbedingungen in die Prüfung der Hypothesen eingehen, muss die Homogenität der Varianzen aller vier Bedingungen sichergestellt werden. Die Prüfung der Varianzhomogenität ist bei a priori Kontrasten, ANOVA und ANCOVA identisch. Es können verschiedene Tests herangezogen werden. Hussy und Jain (2002, S. 197) empfehlen den Levene-Test,

da dieser nicht so empfindlich auf die Verletzung der Normalverteilungsannahme reagiert. Die Autoren weisen außerdem darauf hin, dass bei Homogenitätstests der β -Fehler besonders relevant ist, da die Nullhypothese hier die „Wunschhypothese“ darstellt. Um den β -Fehler für diese Tests gering zu halten, wurde er auf $\beta = .05$ und der α -Fehler somit auf $\alpha = .20$ festgesetzt. Bei der ANOVA zur AV Gewaltverhalten ist der Levene-Test mit $F(3, 182) = 1.242, p = .27$ nicht signifikant. Für die AV Hilfeverhalten ergibt der Levene-Test mit $F(3, 182) = 1.560, p = .20$ ebenfalls ein nicht signifikantes Ergebnis, wenn auch nur knapp. Die Annahme der Varianzhomogenität kann somit bei beiden AVen beibehalten werden.

Für die ANCOVA ergeben sich etwas andere Prüfgrößen. Bei der AV Gewaltverhalten ist der Levene-Test mit $F(3, 182) = 2.071, p < .20$ signifikant auf dem 20%-Niveau. Bei der ANCOVA zur AV Hilfeverhalten ist das Ergebnis des Levene-Tests mit $F(3, 182) = 2.089, p < .20$ ebenfalls signifikant auf dem 20%-Niveau. Daher stellt sich die Frage, welche Auswirkungen die Verletzung der Varianzhomogenitätsannahme auf die Ergebnisse der ANCOVA haben könnte. Nach Glass, Peckham und Sanders (zit. n. Bortz, 1999, S. 276) wird der F -Test selbst bei Varianzheterogenität nur unerheblich beeinflusst, solange die Stichproben gleich groß sind. Dies ist hier der Fall.

11.6.3 NORMALVERTEILUNG

Die verwendeten Auswertungsverfahren setzen voraus, dass alle Stichproben (Faktorstufen) aus normalverteilten Populationen stammen (Bühner & Ziegler, 2009, S. 368). Diese Voraussetzung lässt sich aber nicht direkt überprüfen. Stattdessen kann nur überprüft werden, ob die gemessenen Daten der AVen in den Versuchsbedingungen normalverteilt sind. Ist dies der Fall, kann auf das Vorliegen einer Normalverteilung in der Grundgesamtheit geschlossen werden. Die Normalverteilung lässt sich mit dem Shapiro-Wilk-Test prüfen. Bei beiden AVen ist dieser in allen Bedingungen signifikant (siehe Tabelle 20), was gegen eine Normalverteilung spricht.

Tabelle 20: Test der abhängigen Variablen auf Normalverteiltheit nach Shapiro-Wilk

AV	Versuchsbedingung							
	Nothilfe		Hilfe		Töten		Schatzjagd	
	<i>W</i>	<i>p</i>	<i>W</i>	<i>p</i>	<i>W</i>	<i>p</i>	<i>W</i>	<i>p</i>
Gewaltverhalten	.920	.004	.862	.000	.930	.009	.927	.005
Hilfeverhalten	.799	.000	.915	.002	.847	.000	.892	.000

Der Test kann allerdings in großen Stichproben schon bei kleinen Abweichungen von einer Normalverteilung signifikant werden. Daher wird empfohlen, sich zusätzlich P-P- oder Q-Q-Plots, Histogramme, Schiefe und Kurtosis anzusehen (Field, 1999, S. 148). In einer Normalverteilung haben Schiefe und Kurtosis den Wert null²⁶. Bei der AV Gewaltverhalten sind nur in der Hilfebedingung sowohl Schiefe als auch Kurtosis signifikant von Null verschieden, während bei der AV Hilfeverhalten in allen Bedingungen Schiefe und Kurtosis signifikant von Null

Virtuelle Nothilfe

verschieden sind, mit Ausnahme der Kurtosis in der Hilfebedingung (siehe Tabelle 21). Tabachnick und Fidell (2007, S. 79ff.) empfehlen bei großen Stichproben, sich zusätzlich die Verteilung anzusehen, da auch diese Tests bei großen Stichproben schon bei geringen Abweichungen zu signifikanten Ergebnissen führen würden, weswegen in solchen Fällen der Betrag relevanter sei als das Ergebnis des Signifikanztests. Die Autorinnen empfehlen sogenannte „expected normal probability plots“, bei denen die Abweichungen der Messwerte von den in einer Normalverteilung zu erwartenden Werten dargestellt werden. Wenn die Messwerte normalverteilt sind, dann liegen die Abweichungen bei den „expected normal probability plots“ auf einer Diagonalen.

Tabelle 21: Schiefe und Kurtosis der abhängigen Variablen in den Versuchsbedingungen

AV	Parameter	Versuchsbedingung			
		Nothilfe	Hilfe	Töten	Schatzjagd
Gewaltverhalten	Schiefe	0.791	1.220	0.809	0.520
	SE Schiefe	0.354	0.343	0.354	0.343
	z Schiefe	2.234	3.557*	2.285	1.516
	Kurtosis	0.941	1.841	0.527	-0.478
	SE Kurtosis	0.695	0.674	0.695	0.674
	z Kurtosis	1.354	2.731*	0.758	0.709
Hilfeverhalten	Schiefe	2.417	0.892	1.489	1.111
	SE Schiefe	0.354	0.343	0.354	0.343
	z Schiefe	6.828*	2.601*	4.206*	3.239*
	Kurtosis	9.437	0.511	2.479	2.124
	SE Kurtosis	0.695	0.674	0.695	0.674
	z Kurtosis	13.578*	0.758	3.567*	3.151*

* z-Werte größer oder gleich 2,33 sind signifikant bei $p < .01$.

Bei der AV Gewaltverhalten sehen die Daten in den Histogrammen/Plots einigermaßen normalverteilt aus (siehe Anhang N). Dabei zeigen insbesondere die Histogramme in allen vier Bedingungen eine sichtbare Rechtsschiefe, was im Einklang mit den positiven Schiefe-Werten steht. Bei der AV Hilfeverhalten sehen die Werte in der Nothilfebedingung bis auf die Ausreißer und die sichtbare Rechtsschiefe ebenfalls noch einigermaßen normalverteilt aus, während sich in der Hilfebedingung so etwas wie ein zweiter Gipfel beim Wert 50 bildet. Die Werte in der Töten- und der Schatzjagdbedingung scheinen sich zweigipflig zu verteilen, wobei zwischen den Abstufungen größere Lücken zu erkennen sind.

Da die Daten insbesondere bei der AV Hilfeverhalten offenbar nicht normalverteilt sind, stellt sich die Frage, welche Auswirkungen dies auf die statistischen Tests haben könnte. Nach Bortz (1999, S. 138ff.) sind Mittelwerte bei hinreichend großem Stichprobenumfang ($n \geq 30$) nach dem zentralen Grenzwerttheorem angenähert normalverteilt. Daher müssen für den t -Test die Grundgesamtheiten nur bei kleineren Stichproben normalverteilt sein. Da die Zellengrößen in diesem Experiment $n \geq 45$ betragen, wäre nach Bortz eine Prüfung der Normalverteilung im Prinzip nicht erforderlich. Außerdem reagiere der t -Test laut Bortz robust auf Verletzungen seiner Voraussetzungen (insbesondere dann, wenn die Populationen ein-

gipflig-symmetrisch verteilt sind, was hier aber nicht vorausgesetzt werden kann). Selbst bei ungleichen Stichprobengrößen würde die Präzision nicht beeinflusst, solange Varianzhomogenität gegeben sei. Bei der ANOVA sieht es ähnlich aus. Nach Glass, Peckham und Sanders (zit. n. Bortz, 1999, S. 276) können Abweichungen von der Normalität vernachlässigt werden, wenn die Verteilungen in der Population schief sind. Die Verteilungen in der Population sind zwar nicht bekannt, aber die Verteilungen in der Stichprobe sind eindeutig schief. Die Teststärke würde laut den Autoren durch schmalgipflige Verteilungen vergrößert, durch breitgipflige hingegen verringert. In diesem Fall sind drei der vier Verteilungen eher schmalgipflig und eine eher breitgipflig, was die Teststärke leicht vergrößern müsste. Selbst bei Varianzheterogenität würde der F -Test nach den Autoren nur unerheblich beeinflusst, solange die Stichproben gleich groß sind. Erst Varianzheterogenität in Kombination mit ungleich großen Stichproben würde zu falschen Ergebnissen führen. Beides trifft hier nicht zu. Zusammengefasst müsste somit auch die ANOVA zu zuverlässigen Ergebnissen führen.

11.6.4 HOMOGENITÄT DER REGRESSIONEN

Bei der ANCOVA gibt es im Vergleich zur ANOVA zusätzlich die Voraussetzung der Homogenität der Regressionen (Field, 1999, S. 399; Tabachnick & Fidell, 2007, S. 202f.). Inhaltlich bedeutet dies, dass die Beziehung zwischen einer AV und einer Kovariate über alle Versuchsbedingungen gleich sein muss, was sich an den entsprechenden Regressionsgeraden ablesen lässt, die eine homogene Steigung aufweisen müssen (z. B. Field, 1999, S. 413). Die Voraussetzung lässt sich prüfen, indem die Interaktion einer Kovariate mit der UV geprüft wird. Wird diese Interaktion signifikant, ist die Voraussetzung verletzt. Da hier zwei UVen vorliegen, müssen für jede Kovariate mehrere Interaktionsterme gebildet werden, und zwar $UV_1 \times Kovariate_x$, $UV_2 \times Kovariate_x$ und $UV_1 \times UV_2 \times Kovariate_x$. Der Test lässt sich mit dem ANCOVA-Verfahren durchführen, indem zu dem eigentlichen Modell die genannten Interaktionsterme hinzugefügt werden (siehe z. B. Tabachnick & Fidell, 2007, S. 230-237). Wird einer dieser Interaktionsterme signifikant, ist die Voraussetzung homogener Regressionen verletzt. Dies ist nicht der Fall (siehe Anhang P). Daher müsste die ANCOVA ebenfalls zu zuverlässigen Ergebnissen führen.

12 DISKUSSION DER STATISTISCHEN VORHERSAGEN

12.1 GEWALTVERHALTEN

12.1.1 EINZELVERGLEICHE

Einen Überblick über die Mittelwerte der AV Gewaltverhalten in den Versuchsbedingungen gibt Abb. 47. Augenscheinlich sind die Mittelwerte in der Nothilfe- und der Tötenbedingung größer als die in der Hilfe- und der Schatzjagdbedingung. Auf Basis dieser Werte wurden Einzelvergleiche durchgeführt, mit denen die ersten drei statistischen Vorhersagen geprüft wurden. Da es sich um gerichtete Hypothesen handelt, wurden wie eingangs erwähnt einseitige t -Tests durchgeführt.

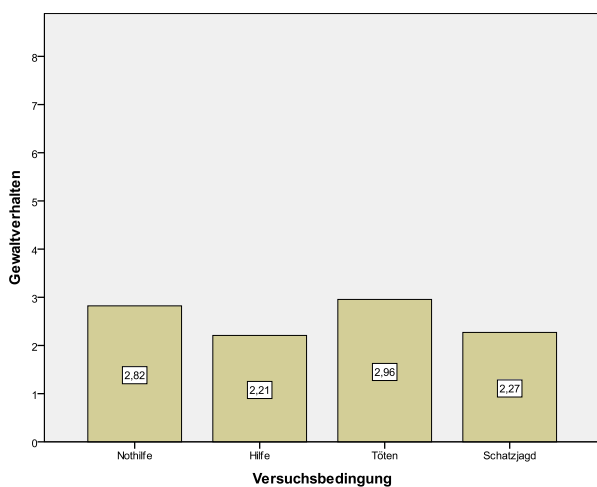


Abb. 47: Mittelwerte der abhängigen Variable „Gewaltverhalten“ in den Versuchsbedingungen

SV₁: Der Mittelwert der Intensität im CRTT aus Runde 1 ist in der Gruppe „Töten“ größer als in der Gruppe „Hilfe“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

Der Mittelwert in der Tötenbedingung ($M = 2.956$) ist tatsächlich signifikant größer als der Mittelwert in der Hilfebedingung ($M = 2.208$), $t(182) = 2.143$, $p < .05$, $g = 0.463$ und auch als der Mittelwert in der Schatzjagdbedingung ($M = 2.271$), $t(182) = 1.964$, $p < .05$, $g = 0.395$. Die gesamte Testhypothese wird somit angenommen mit einer Irrtumswahrscheinlichkeit von $\alpha = .05$.

SV₂: Der Mittelwert der Intensität im CRTT aus Runde 1 ist in der Gruppe „Nothilfe“ größer als in der Gruppe „Hilfe“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

Der Mittelwert in der Nothilfebedingung ($M = 2.822$) ist zwar signifikant größer als der in der Hilfebedingung ($M = 2.208$), $t(182) = 1.761$, $p < .05$, $g = 0.373$, er ist aber nicht signifikant

Virtuelle Nothilfe

größer als der in der Schatzjagdbedingung mit $t(182) = 1.581, p = .07, g = 0.313$. Die Nullhypothese wird somit angenommen mit einer Irrtumswahrscheinlichkeit von $\beta = .20$ (zur Logik dieser Entscheidung siehe Hussy & Jain, 2002, S. 161f.).

SV₃: Der Mittelwert der Intensität im CRTT aus Runde 1 ist in der Gruppe „Töten“ größer als in der Gruppe „Nothilfe“.

Der Mittelwert in der Tötenbedingung ($M = 2.956$) ist nicht signifikant größer als der in der Nothilfebedingung ($M = 2.822$), $t(182) = 0.377, p = .35, g = 0.072$. Die Nullhypothese wird somit angenommen mit einer Irrtumswahrscheinlichkeit von $\beta = .20$. Der g -Wert zeigt außerdem, dass ein Unterschied zwischen den Gruppen quasi nicht vorhanden ist.

12.1.2 ANOVA

Um die Auswirkungen der Störvariablen zu prüfen, wurde a posteriori eine ANOVA durchgeführt, deren Ergebnisse mit einer ANCOVA verglichen wurden. Die ANOVA führt zu folgenden Resultaten (siehe Tabelle 22): Der Haupteffekt „Gewalt im Spiel“ ist signifikant auf dem 1%-Niveau. Alle anderen Effekte sind nicht signifikant.

Tabelle 22: Zweifaktorielle ANOVA mit „Gewaltverhalten“ als abhängige Variable

Quelle der Varianz	SS	df	MS	F	p	η^2	ω^2
Gewalt im Spiel	19.584	1	19.584	6.936	.009	0.037	0.031
Hilfe im Spiel	0.445	1	0.445	0.158	.692	0.001	0.000
Interaktion	0.058	1	0.058	0.021	.886	0.000	0.000
Fehler	513.885	182	2.824				
Gesamt	533.962	186					

Nach der Daumenregel von Cohen (1988, S. 273-289) lässt sich $f^2 = .10$ als kleiner Effekt bezeichnen und $f^2 = .25$ als mittlerer Effekt. Der Parameter f^2 lässt sich in η^2 umrechnen (siehe Anhang Q). Daraus folgt, dass es sich bei $\eta^2 = .01$ um einen kleinen und bei $\eta^2 = .06$ um einen mittleren Effekt handelt (vgl. Hussy & Jain, 2002, S. 157). Der Haupteffekt „Gewalt im Spiel“ lässt sich danach als kleiner Effekt bewerten. Nach Bortz (1999, S. 269f.) lässt sich der prozentuale Anteil bestimmen, den die Haupteffekte und die Interaktion an der Aufklärung der AV haben, indem η^2 mit 100 multipliziert wird. Der signifikante Haupteffekt „Gewalt im Spiel“ klärt also 3.7% der Varianz der AV Gewaltverhalten auf. Laut Bortz (1999, S. 269f.) bzw. Tabachnick und Fidell (2007, S. 55) gilt η^2 für die Verhältnisse in der Stichprobe, während ω^2 die Verhältnisse in der Population schätzt (die Formel zur Berechnung von findet sich in Anhang Q). In der Population klärt der Haupteffekt „Gewalt im Spiel“ mit $\omega^2 = 0.031$ noch 3.1% der Varianz der AV Gewaltverhalten auf.

12.1.3 ANCOVA

Um zu prüfen, ob die Ergebnisse der ANOVA sich durch Hinzunahme von Störvariablen möglicherweise verändern, wurde im Anschluss eine Kovarianzanalyse (ANCOVA) durchgeführt.

12.1.3.1 STÖRVARIABLENAUSWAHL

Im Rahmen dieses Experiments wurde eine Vielzahl von Störvariablen erhoben, von denen die Mehrzahl durch theoretische Überlegungen bzw. durch die Ergebnisse empirischer Studien begründet ist. Die nicht auf diese Weise abgeleiteten Störvariablen wie Alter, Hochschulzugehörigkeit, Erhebungswelle etc. wurden erhoben, weil die Treatments trotz Randomisierung möglicherweise nicht zu kovariierenden Bedingungen geführt haben und somit die Unterschiede zwischen den Gruppen zum Teil auch auf diese stichprobenspezifischen Störgrößen rückführbar sein könnten. In eine ANCOVA sollten aber nur wenige, theoretisch oder empirisch gut begründete Störvariablen einfließen, da die Hinzunahme einer weiteren Störvariable immer einen Freiheitsgrad kostet (Tabachnick & Fidell, 2007, S. 211f.). In so einem Fall bietet es sich an, über eine multiple lineare Regression diejenigen Störvariablen auszuwählen, die in der Stichprobe den stärksten Zusammenhang mit der AV aufweisen. Dabei stellt sich die Frage, welche Regressionsmethode hierfür geeigneter ist. Die Methode „Schrittweise“ (engl. „stepwise regression“), die von Tabachnick und Fidell (2007, S. 211f.) für diesen Zweck empfohlen wird, wählt in jedem Schritt den erklärungskräftigsten Prädiktor aus. Die Hinzufügung weiterer Prädiktoren endet, sobald ein aufzunehmender Prädiktor keinen signifikanten Aufklärungsgewinn bietet. Dieses Verfahren hat aber zwei Nachteile: Zum einen ist die Aufnahme eines Prädiktors nicht unabhängig von den bereits aufgenommenen Prädiktoren. Es kann sich sogar eine grundsätzlich andere Auswahl ergeben, je nachdem, mit welchem Prädiktor die Regression begonnen wird. Zum anderen wird für diese Form der Regression eine recht große Stichprobe von mindestens 40 Personen pro Prädiktor benötigt, wenn die Ergebnisse generalisierbar sein sollen (Tabachnick & Fidell, 2007, S. 123). Bei ca. 25 Störvariablen würden somit $N \geq 1000$ Personen benötigt, die hier nicht zur Verfügung stehen. Die alternative Methode „Einschluss“ nimmt alle Prädiktoren gleichzeitig in die Analyse auf. Für diese Methode werden nach einer Daumenregel von Green (1991, zit. n. Tabachnick & Fidell, 2007, S. 123) $N \geq 104 + m$ Personen benötigt (bei $\alpha=.05$, $\beta=.20$ und mittlerer Effektstärke), wenn es allein um die Testung der einzelnen Prädiktoren geht, wobei m für die Anzahl der Prädiktoren steht. Nach dieser Methode müssten bei ca. 25 Störvariablen also lediglich $N \geq 129$ Personen zur Verfügung stehen, was hier gegeben ist. Bei einer schief verteilten AV werden allerdings mehr Personen benötigt (Green, 1991, zit. n. Tabachnick & Fidell, 2007, S. 123). Die Methode hat aber ebenfalls zwei Nachteile: Zum einen existiert kein Entscheidungskriterium zur Auswahl der am besten geeigneten Prädiktoren, zum anderen kann sich der Zusammenhang zwischen den ausgewählten Prädiktoren und der AV stark verändern, sobald die unerwünschten Prädiktoren aus der Gleichung entfernt werden. Um die Nachteile der beiden Verfahren zu kompensieren, wurde hier ein zweistufiges Verfahren an-

gewandt. Im ersten Schritt wurde eine Regression nach der Methode „Einschluss“ vorgenommen. Hier wurden alle erklärungskräftigen Prädiktoren mit einem Betagewicht von $\beta \geq .10$ ausgewählt. Störvariablen, die nachgewiesen konfundiert waren (siehe Abschnitt 11.4.1), gingen nicht in die Regressionen ein. Kovariaten müssen nämlich unabhängig von den UVen sein (vgl. Field, 2009, S. 397f.; Tabachnick & Fidell, 2007, S. 200; Bortz & Döring, 1995, S. 509), weil sich der Zusammenhang zwischen der AV und der UV verringert, wenn eine Kovariate dieselbe Varianz in der AV erklärt, die bereits von einer UV aufgeklärt wird. Alle Prädiktoren, die nach der schrittweisen Regression übrig blieben, gingen als Kovariaten in die ANCOVA ein.

Da die lineare Regression nur lineare Zusammenhänge finden kann, wurde zur Prüfung des Zusammenhangs für jede Kombination aus Störvariable und AV ein bivariates Streudiagramm erstellt (aufgrund der großen Menge werden die Diagramme hier nicht dargestellt). In den Diagrammen waren keine Hinweise auf kurvilineare Zusammenhänge zu erkennen. Außerdem wurde bei den sieben dichotomen Störvariablen überprüft, ob eine der beiden Ausprägungen eine Häufigkeit von 90% oder mehr aufweist, da eine so verteilte Variable nach Rummel (1970, zit. n. Tabachnick & Fidell, 2007, S. 73) dazu führen würde, dass die Korrelationskoeffizienten beschnitten werden und die Werte in der weniger häufigen Kategorie einen größeren Einfluss hätten als die anderen. Dies war bei der Störvariablen „Fragen unangenehm“ der Fall, die sich als Konstante erwies (d. h. kein Proband fand eine Frage unangenehm), sowie bei „Lust verloren“, wo 93% der Werte in einer Kategorie lagen (d. h. die meisten Probanden verloren die Lust nicht). Diese beiden Variablen wurden vor Durchführung der Regressionen entfernt.

12.1.3.2 AUSWAHL-REGRESSION

Die Anwendungsvoraussetzungen der Regression sind Normalverteilung, Homoskedastizität und fehlende Multikollinearität. Da die Regression nur lineare Zusammenhänge finden kann, sollte außerdem Linearität gegeben sein. Schließlich sollten keine Ausreißer vorhanden sein, da diese das Ergebnis der Regression stark verzerren können. Bis auf die Multikollinearität können die Voraussetzungen anhand der Verteilung der Residualwerte (siehe Abb. 48 und Abb. 49) geprüft werden.

Aus den Residualplots lässt sich folgern (siehe Field, 2009, S. 248f.), dass in den Regressionen sowohl Normalität gegeben ist (daran erkennbar, dass die Punkte sich gleichmäßig ober- und unterhalb der Horizontalen verteilen) als auch Linearität (daran erkennbar, dass die Punkte sich nicht u-förmig verteilen). Es sind auch keine größeren Ausreißer zu finden (daran erkennbar, dass ein Punkt weit entfernt von den anderen Punkten liegt). Der zweite Residualplot ist allerdings schwieriger zu interpretieren, was daran liegt, dass die verbliebenen Prädiktoren beide dichotom sind. Hier könnte es sein, dass die Normalitätsannahme verletzt ist.

Virtuelle Nothilfe

Ein zusätzlich aus den Residuen erstelltes Histogramm (siehe Anhang N) lässt erkennen, dass dies nicht der Fall ist.

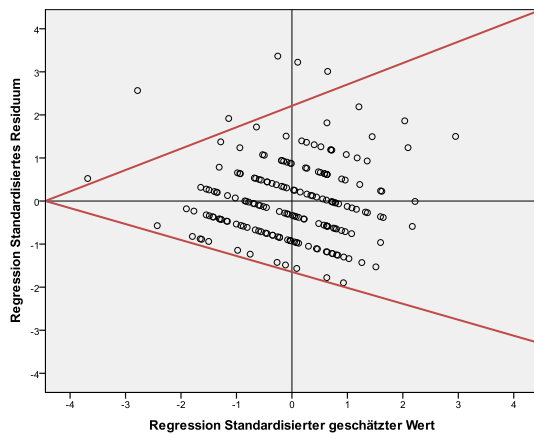


Abb. 48: Residualplot Gewaltverhalten (Einschluss-Regression)

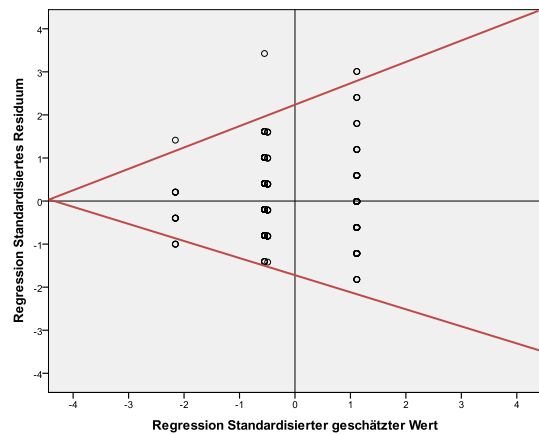


Abb. 49: Residualplot Gewaltverhalten (Schrittweise-Regression)

In beiden Fällen liegt Heteroskedastizität vor (daran erkennbar, dass die Punkte sich trichterförmig verteilen). Die Heteroskedastizität führt aber nicht zu einem invaliden Ergebnis, sondern zu abgeschwächten Zusammenhängen zwischen den Störvariablen und der AV (Tabachnick & Fidell, 2007, S. 127), was für diesen Anwendungsfall ein vernachlässigbares Problem ist. Die wahrscheinlichste Ursache für die Heteroskedastizität dürfte darin liegen, dass einige Störvariablen schief verteilt sind, andere aber nicht (Tabachnick & Fidell, 2007, S. 127).

Tabelle 23: Einschluss-Regression der abhängigen Variable „Gewaltverhalten“ auf alle Störvariablen (dargestellt werden nur Prädiktoren mit $|\beta| \geq .10$); Schrittweise-Regression auf die verbliebenen Störvariablen

Störvariable	Einschluss			Schrittweise		
	β	p	VIF	β	p	VIF
Erhebungswelle	-.19	.019	1.19	-.20	.006	1.02
Computergegner durchschaut	-.17	.024	1.08	-.17	.024	1.02
Spielaufgabe Interessantheit	-.16	.150	2.27			
Spielaufgabe Aktivierung (lustlos)	-.14	.118	1.64			
Aggressionsmaß durchschaut	-.11	.255	1.11			
Spielaufgabe Schwierigkeit	-.11	.255	1.64			
Spielaufgabe Präsenzgefühl	-.10	.264	1.51			

Einschluss-Regression: $R^2 = .15$ ($p = .171$), korrigiertes $R^2 = .04$. Schrittweise-Regression: $R^2 = .06$ ($p = .004$), korrigiertes $R^2 = .05$.

Ob Multikollinearität vorliegt, wurde anhand der Toleranzwerte der Prädiktoren festgestellt, die einen Wert von $Tol_j = 0.20$ nicht unterschreiten dürfen. Der kleinste Toleranz-Wert in der Einschluss-Regression betrug $Tol_j = 0.37$, der kleinste im letzten Schritt der Schrittweise-Regression betrug $Tol_j = 0.98$. Die Regressionen sollten daher valide Ergebnisse erbringen.

Virtuelle Nothilfe

Die Einschluss-Regression führte zu den in Tabelle 23 aufgeführten Störvariablen. Mit den verbliebenen Störvariablen wurde im Anschluss die Schrittweise-Regression durchgeführt. Für die ANCOVA verblieben die zwei Störvariablen „Erhebungswelle“ und „Computergegner durchschaut“. Diese hängen wie folgt mit der AV zusammen: Die Probanden der zweiten Erhebungswelle wiesen im Durchschnitt niedrigere Werte im CRTT auf als die der ersten Welle. Diejenigen Probanden, die wahrscheinlich durchschaut hatten, dass sie gegen den Computer angetreten sind, stellten ebenfalls niedrigere Werte im CRTT ein.

12.1.3.3 ERGEBNISSE

Die ANCOVA führt zu folgenden Resultaten (siehe Tabelle 24): Der Haupteffekt „Gewalt im Spiel“ ist knapp nicht mehr auf dem 1%-Niveau signifikant. Der Haupteffekt „Hilfe im Spiel“ sowie die Interaktion der beiden Haupteffekte bleiben nicht signifikant. Dafür sind alle Störvariablen (mindestens) signifikant auf dem 5%-Niveau, was aufgrund ihrer Auswahl über die Regressionen aber auch nicht anders zu erwarten war.

Tabelle 24: ANCOVA mit der unabhängigen Variable „Gewaltverhalten“ und den Störvariablen „Erhebungswelle“, „CRTT Gegner verrückt machen“ und „Computergegner durchschaut“ als Kovariaten

Quelle der Varianz	SS	df	MS	F	p	η^2	ω^2
ANOVA							
Gewalt im Spiel	19.584	1	19.584	6.936	.009	0.037	0.031
Hilfe im Spiel	0.445	1	0.445	0.158	.692	0.001	0.000
Interaktion	0.058	1	0.058	0.021	.886	0.000	0.000
Fehler	513.885	182	2.824				
Gesamt	533.962	186					
ANCOVA							
Gewalt im Spiel	15.535	1	15.535	5.756	.017	0.031	0.024
Hilfe im Spiel	0.722	1	0.722	0.268	.602	0.001	0.000
Interaktion	0.003	1	0.003	0.001	.990	0.000	0.000
Erhebungswelle	17.981	1	17.981	6.662	.009	0.036	0.028
Computergegner durchschaut	13.739	1	13.739	5.091	.037	0.028	0.021
Fehler	444.697	180	2.484				
Gesamt	533.962	186					

Nach der Daumenregel von Cohen (1988, S. 273-289) lässt sich $\eta^2 = .01$ als kleiner Effekt bezeichnen, $\eta^2 = .06$ als mittlerer Effekt und $\eta^2 = .14$ als großer Effekt. In diesem Sinne handelt es sich bei den Effekten der Kovariaten um kleine Effekte. Der signifikante Haupteffekt „Gewalt im Spiel“ klärt mit $\eta^2 = 0.031$ noch 3.1% der Varianz in der Stichprobe bzw. mit $\omega^2 = 0.024$ ca. 2.4% der Varianz in der Population auf. In der ANOVA war der Wert mit $\eta^2 = 0.037$ (bzw. $\omega^2 = 0.031$) nur geringfügig größer. Dies weist darauf hin, dass die von den Störvariablen aufgeklärte Varianz größtenteils unabhängig ist von der Varianz, die der Haupteffekt aufklärt. Die Kovariate „Erhebungswelle“ übertrifft den Haupteffekt in der Varianzaufklärung. Alle Kovariaten klären zusammengekommen 6.4% (4.9%) auf.

12.2 HILFEVERHALTEN

Nach der ausführlichen Testung der drei Hypothesen zum Gewaltverhalten sollen nun die übrigen drei Hypothesen zum Hilfeverhalten geprüft werden. Die Prüfung erfolgt dabei auf dieselbe Weise: Sie beginnt mit Einzelvergleichen, die von einer ANOVA und einer ANCOVA gefolgt werden.

12.2.1 EINZELVERGLEICHE

Einen Überblick über die Mittelwerte der AV Hilfeverhalten in den Versuchsbedingungen gibt Abb. 50. Augenscheinlich ist der Mittelwert in der Nothilfebedingung geringer als in allen anderen, während der Mittelwert in der Hilfebedingung etwas größer zu sein scheint als die in der Töten- und der Schatzjagdbedingung. Auf Basis dieser Werte wurden Einzelvergleiche durchgeführt, mit denen die letzten drei statistischen Vorhersagen geprüft wurden. Wie bei den Einzelvergleichen zum Gewaltverhalten wurden auch hier einseitige t -Tests durchgeführt.

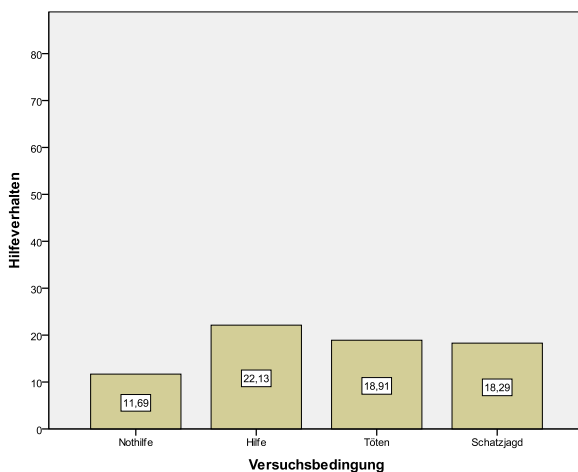


Abb. 50: Mittelwerte der abhängigen Variable „Hilfeverhalten“ in den Versuchsbedingungen

SV₄: Der Mittelwert der angegebenen Aufgabenanzahl bei der Eckenzählaufgabe ist in der Gruppe „Hilfe“ größer als in der Gruppe „Töten“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

Der Mittelwert in der Hilfebedingung ($M = 22.125$) ist weder signifikant größer als der in der Tötenbedingung ($M = 18.911$), $t(182) = 0.918$, $p = .18$, $g = 0.181$, noch als der in der Schatzjagdbedingung ($M = 18.292$), $t = 1.113$, $p = .13$, $g = 0.216$. Die Nullhypothese wird somit angenommen mit einer Irrtumswahrscheinlichkeit von $\beta = .20$ (zur Logik dieser Entscheidung siehe Hussy & Jain, 2002, S. 161f.).

Virtuelle Nothilfe

SV₅: Der Mittelwert der angegebenen Aufgabenanzahl bei der Eckenzählaufgabe ist in der Gruppe „Nothilfe“ größer als in der Gruppe „Töten“ und gleichzeitig größer als in der Gruppe „Schatzjagd“.

Der Mittelwert in der Nothilfebedingung ($M = 11.689$) ist nicht signifikant größer als in der Tötenbedingung ($M = 18.911$), $t(182) = 2.062$, $p < .05$, $g = -0.455$, und ebenfalls nicht signifikant größer als in der Schatzjagdbedingung ($M = 18.292$), $t(182) = 1.885$, $p < .05$, $g = -0.415$. Die Nullhypothese wird somit angenommen mit einer Irrtumswahrscheinlichkeit von $\beta = .20$.

SV₆: Der Mittelwert der angegebenen Aufgabenanzahl bei der Eckenzählaufgabe ist in der Gruppe „Hilfe“ größer als in der Gruppe „Nothilfe“.

Der Mittelwert in der Hilfebedingung ($M = 22.125$) ist signifikant größer als der in der Nothilfebedingung ($M = 11.689$), $t(182) = 2.980$, $p < .01$, $g = 0.624$. Die Testhypothese wird somit angenommen mit einer Irrtumswahrscheinlichkeit von $\alpha = .05$.

12.2.2 ANOVA

Nach den Einzelvergleichen zum Hilfeverhalten folgt nun zu Vergleichszwecken eine a posteriori durchgeführte ANOVA. Die ANOVA führt zu folgenden Resultaten (siehe Tabelle 25): Die Interaktion sowie der Haupteffekt „Gewalt im Spiel“ sind signifikant auf dem 5%-Niveau. Der Haupteffekt „Hilfe im Spiel“ ist nicht signifikant.

Tabelle 25: Zweifaktorielle ANOVA mit „Hilfeverhalten“ als abhängige Variable

Quelle der Varianz	SS	df	MS	F	p	η^2	ω^2
Gewalt im Spiel	1119.100	1	1119.100	3.928	.049	0.021	0.015
Hilfe im Spiel	133.369	1	133.369	0.468	.495	0.003	0.000
Interaktion	1419.391	1	1419.391	4.982	.027	0.027	0.021
Fehler	51854.456	182	284.915				
Gesamt	54499.833	186					

Nach der Daumenregel von Cohen (1988, S. 273-289) handelt es sich bei dem Haupteffekt „Gewalt im Spiel“ sowie bei der Interaktion zwischen „Gewalt im Spiel“ und „Hilfe im Spiel“ jeweils um kleine Effekte. Dabei klärt der signifikante Haupteffekt in der Stichprobe 2.1% der Varianz der AV Hilfeverhalten auf, während die Interaktion in der Stichprobe 2.7% aufklärt. Somit werden von den signifikanten Effekten in der Stichprobe zusammen genommen 4.8% der Gesamtvarianz aufgeklärt. Dies ist knapp 1% mehr als bei der AV Gewaltverhalten, wo 3.7% der Gesamtvarianz aufgeklärt wurden. In der Population klärt der Haupteffekt noch 1.5% der Varianz auf, während die Interaktion 2.1% aufklärt. Zusammengenommen sind dies 3.6% der Varianz, die in der Population aufgeklärt werden.

Virtuelle Nothilfe

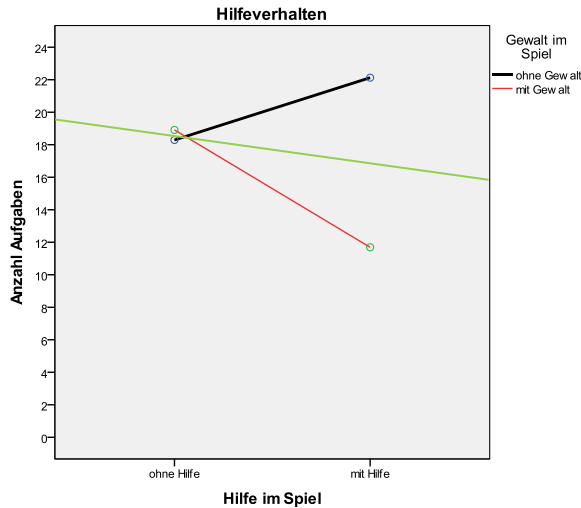


Abb. 51: Interaktion Hilfe x Gewalt

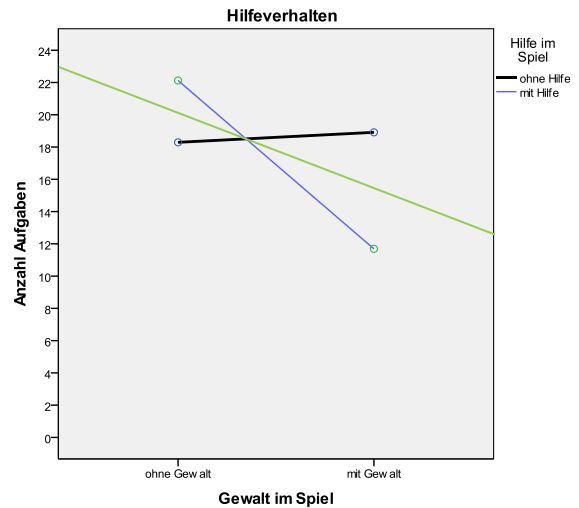


Abb. 52: Interaktion Gewalt x Hilfe

Es werden zwei Interaktionsdiagramme benötigt, um festzustellen, welche Form der Interaktion vorliegt. Da sich die Linien in beiden Diagrammen kreuzen, handelt es sich hier um eine disordinale Interaktion (die rote Linie in der Abbildung dient zur Veranschaulichung des jeweiligen Haupteffekts). Bei disordinalen Interaktionen können eventuell vorliegende Haupteffekte nicht sinnvoll interpretiert werden, sondern nur die einfachen Haupteffekte (Hussy & Jain, 2002, S. 189). Nach Bühner und Ziegler (2009, S. 419f.) ist nur das Diagramm des signifikanten Haupteffekts für die Interpretation relevant (in diesem Fall Abb. 52). Die Deutung lautet daher wie folgt: Lässt sich die Aufgabe im Spiel nur lösen, indem der Auftraggeberin geholfen wird, dann ist das Hilfeverhalten nach dem Spiel kleiner, wenn Gewalt angewendet werden musste. Zu derselben Deutung kommt man auch nach der Durchführung eines multiplen Mittelwertvergleichs, bei dem jeder Gruppenmittelwert gegen jeden anderen getestet wurde. Für solche Vergleiche sind laut Diehl und Arbinger (1992, S. 217ff.) drei Verfahren besonders gut geeignet, die sich hinsichtlich ihrer Anwendungsvoraussetzungen leicht unterscheiden. Bei varianzhomogenen Populationen und ungleicher Gruppengröße empfehlen die Autoren den hier eingesetzten Tukey-Kramer-Test (siehe Anhang P) mit $MS_{Fehler} = 284.915$, $N = 186$, $J = 4$, $df = 182$, $Q_k = 3.633$, $\alpha_F = .05$. Bei diesem Test resultiert für jedes Paar an Versuchsbedingungen ein Q^* -Wert, der auf Signifikanz geprüft wird, indem er mit einem kritischen Q_k -Wert verglichen wird (siehe Tabelle 26).

Tabelle 26: Multipler Mittelwertvergleich (Tukey-Kramer-Test) zum Hilfeverhalten

Bedingung	A: Nothilfe	B: Hilfe	C: Töten	D: Schatzjagd
<i>n</i>	45	48	45	48
<i>M</i>	11.689	22.125	18.911	18.292
A: Nothilfe	-	4.214*	2.870	2.666
B: Hilfe	-	-	1.298	1.573
C: Töten	-	-	-	0.250

Virtuelle Nothilfe

Das Ergebnis des Tukey-Kramer-Tests zeigt, dass sich mit $Q^*(182) = 4.214$, $p < .05$ lediglich die Nothilfe- von der Hilfebedingung signifikant hinsichtlich des Hilfeverhaltens unterscheidet. Alle anderen Gruppenunterschiede sind nicht signifikant. Zur Bewertung des Mittelwertunterschieds lässt sich Hedges g heranziehen, das in diesem Fall $g = 0.62$ beträgt. Da Hedges g und Cohens d eng verwandt sind, lässt sich für die Bewertung die Faustregel von Cohen (1988, S. 24-27) heranziehen, nach der es sich bei Werten ab $d = 0.50$ um einen mittleren und ab $d = 0.80$ um einen großen Effekt handelt. Der gefundene Mittelwertunterschied kann somit als „mittel“ bezeichnet werden.

12.2.3 ANCOVA

Die Ergebnisse der ANOVA sollen im Folgenden wieder mit den Ergebnissen einer ANCOVA verglichen werden. Zu diesem Zweck ist im ersten Schritt die Auswahl der Kovariaten durchzuführen.

12.2.3.1 STÖRVARIABLENAUSWAHL

Für die AV Hilfeverhalten wurden ebenfalls zur Auswahl der Störvariablen für die ANCOVA eine Einschluss-Regression sowie eine Schrittweise-Regression durchgeführt.

12.2.3.2 AUSWAHL-REGRESSION

Auch bei diesen Regressionen wurden die Anwendungsvoraussetzungen wieder anhand der Verteilung der Residualwerte (siehe Abb. 53 und Abb. 54) geprüft. Aus den Residualplots lässt sich folgern, dass in den Regressionen Linearität gegeben ist. Die Normalitätsannahme scheint hingegen leicht verletzt zu sein, da sich über der Horizontalen mehr Punkte befinden als darunter. Außerdem befindet sich im Plot der Schrittweise-Regression am linken Ende der Abszisse ein Ausreißer. Zusätzlich liegt wieder Heteroskedastizität vor. Die Verletzung von Homoskedastizität und Normalität führen aber beide nicht zu einem invaliden Ergebnis, sondern zu abgeschwächten Zusammenhängen zwischen Störvariablen und AV (Tabachnick & Fidell, 2007, S. 127), was für diesen Anwendungsfall ein vernachlässigbares Problem ist.

Ob Multikollinearität vorliegt, wurde wieder anhand der Toleranzwerte der Prädiktoren festgestellt, die einen Wert von $Tol_j = 0.20$ nicht unterschreiten dürfen. Der kleinste Toleranzwert in der Einschluss-Regression betrug $Tol_j = 0.41$, der kleinste am Ende der Schrittweise-Regression betrug $Tol_j = 0.89$. Die Regressionen sollten daher valide Ergebnisse erbringen. Die Einschluss-Regression führte zu den in Tabelle 27 aufgeführten Störvariablen.

Virtuelle Nothilfe

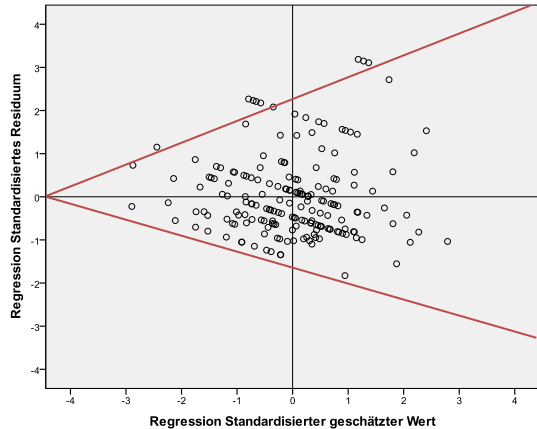


Abb. 53: Residualplot Hilfeverhalten (Einschluss-Regression)

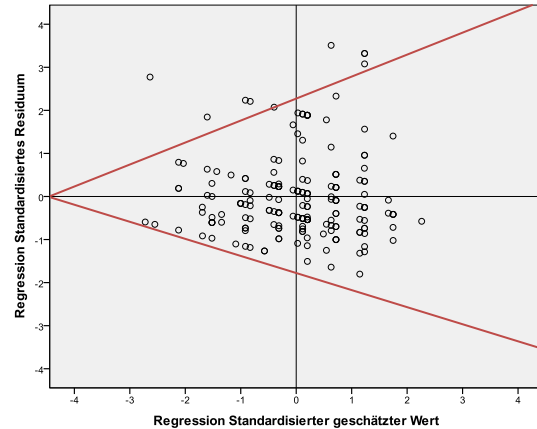


Abb. 54: Residualplot Hilfeverhalten (Schrittweise-Regression)

Mit den verbliebenen Störvariablen wurde im Anschluss die Schrittweise-Regression durchgeführt. Für die ANCOVA verblieben die Störvariablen „Spielaufgabe Realitätsnähe Grafik“, „Spielaufgabe Schwierigkeit“ sowie „Spielaufgabe Aktivierung (lustlos)“. Diese hängen wie folgt mit der AV zusammen: Die Probanden waren bereit, umso mehr Aufgaben für den Diplomanden zu bearbeiten, je mehr sie (a) die Grafik im Spiel realistisch fanden und (b) die Spielaufgabe schwierig fanden bzw. je weniger sie (c) die Spielaufgabe lustlos machte.

Tabelle 27: Einschluss-Regression der abhängigen Variable „Hilfeverhalten“ auf alle Störvariablen (dargestellt werden nur Prädiktoren mit $\beta \geq .10$); Prüfung der Prädiktoren auf Konfundierung; Schrittweise-Regression auf die verbliebenen Störvariablen

Störvariable	Einschluss			Schrittweise		
	β	p	VIF	β	p	VIF
Spielaufgabe Schwierigkeit	.18	.038	1.50	.17	.025	1.08
Aufgabendauer	-.17	.128	2.44			
Spielaufgabe Realitätsnähe Grafik	.16	.041	1.26	.17	.019	1.04
Computergegner durchschaut	.16	.034	1.08			
Spielaufgabe Aktivierung (lustlos)	-.16	.082	1.59	-.20	.009	1.12
Restzeit für Hilfeverhalten	-.14	.147	1.73			
Prosozialespielkonsum	.12	.138	1.21			

Einschluss-Regression: $R^2 = .16$ ($p = .054$), korrigiertes $R^2 = .06$. Schrittweise-Regression: $R^2 = .09$ ($p = .001$), korrigiertes $R^2 = .08$.

12.2.3.3 ERGEBNISSE

Die ANCOVA führt zu folgenden Resultaten (siehe Tabelle 28): Der Haupteffekt „Gewalt im Spiel“ ist zwar nur noch auf dem 10%-Niveau signifikant, klärt aber weiterhin nahezu dieselbe Menge an Varianz auf (1.9% vs. 2.1% in der Stichprobe bzw. 1.2% vs. 1.5% in der Population). Beim Haupteffekt „Hilfe im Spiel“ und bei der Interaktion hingegen ergeben sich keine Veränderungen. Dies weist darauf hin, dass die von den Störvariablen aufgeklärte Varianz zwar unabhängig ist von der Varianz, die vom Haupteffekt „Hilfe im Spiel“ und der Interaktion

Virtuelle Nothilfe

aufgeklärt wird, dass aber mindestens eine Störvariable einen Teil der Varianz in der AV aufklärt, die in der ANOVA vom Haupteffekt „Gewalt im Spiel“ aufgeklärt wurde. Eine bivariate Korrelation dieses Haupteffekts mit den Störvariablen hilft bei der Suche nach dieser Störvariable nicht weiter, da die Störvariablen „Spielaufgabe Schwierigkeit“ ($r = -.07, p = .32$), „Spielaufgabe Aktivierung (lustlos)“ ($r = -.04, p = .62$) und „Spielaufgabe Realitätsnähe Grafik“ ($r = -.06, p = .45$) alle ungefähr gleich niedrig mit dem Haupteffekt korrelieren. Möglicherweise liegt die Ursache auch darin, dass die gewaltlosen Spielaufgaben als interessanter/spannender wahrgenommen wurden, und dass Interessantheit/Spannungsgehalt der Spielaufgabe wiederum eine Auswirkung auf die AV Hilfeverhalten haben, die dabei von der Aktivierung moderiert wird (siehe Abschnitt 11.5).

Nach der Daumenregel von Cohen (1988, S. 273-289) lässt sich $\eta^2 = .01$ als kleiner Effekt bezeichnen und $\eta^2 = .06$ als mittlerer Effekt. In diesem Sinne handelt es sich bei den Effekten der Störvariablen um kleine Effekte. Alle Störvariablen zusammen genommen klären 8.9% der Stichprobenvarianz bzw. 6.4% der Populationsvarianz auf.

Tabelle 28: ANCOVA mit der unabhängigen Variable „Hilfeverhalten“ und den Störvariablen „Spielaufgabe Realitätsnähe Grafik“, „CRTT als RZT geeignet“, „Spielaufgabe Aktivierung (lustlos)“ und „Spielaufgabe Schwierigkeit“ als Kovariaten

Quelle der Varianz	SS	df	MS	F	p	η^2	ω^2
ANOVA							
Gewalt im Spiel	1119.100	1	1119.100	3.928	.049	.021	.015
Hilfe im Spiel	133.369	1	133.369	0.468	.495	.003	.000
Interaktion	1419.391	1	1419.391	4.982	.027	.027	.021
Fehler	51854.456	182	284.915				
Gesamt	54499.833	186					
ANCOVA							
Gewalt im Spiel	927.478		927.478	3.505	.063	.019	.012
Hilfe im Spiel	4.061		4.061	0.015	.902	.000	.000
Interaktion	1288.907		1288.907	4.870	.029	.026	.019
Spielaufgabe Realitätsnähe Grafik	1353.715		1353.715	5.115	.025	.028	.020
Spielaufgabe Schwierigkeit	996.383		996.383	3.765	.054	.021	.013
Spielaufgabe Aktivierung (lustlos)	1960.761		1960.761	7.409	.007	.040	.031
Fehler	47372.942	179	264.653				
Gesamt	54499.833	186					

13 DISKUSSION DER EMPIRISCH-INHALTLICHE HYPOTHESEN

13.1 GEWALTVERHALTEN

Nach den statistischen Prüfungen sollen im Folgenden die drei Hypothesen zum Gewaltverhalten von der Ebene der Testhypothese zurück auf die Ebene der empirisch-inhaltlichen Hypothese abgeleitet werden. Die Ableitung auf die Ebene der theoretisch-inhaltlichen Hypothesen erfordert eine umfassende Bewertung der Variablenvalidität und der internen Validität, weswegen sie erst im darauffolgenden Abschnitt erfolgt.

Die erste Testhypothese (TH_1) wurde angenommen. Für die statistische Vorhersage (SV_1) gilt somit: Der Mittelwert der AV Gewaltverhalten ist in der Gruppe „Töten“ größer als in der Gruppe „Hilfe“ und gleichzeitig größer als in der Gruppe „Schatzjagd“. Die Effektstärken der beiden Einzelvergleiche, der ANOVA und der ANCOVA sind dabei vergleichbar mit dem Wert aus der Meta-Analyse von Anderson et al. (2010) und können daher als praktisch bedeutsam angesehen werden. Wenn die präzisionsmindernden Störungen beim CRTT nicht aufgetreten wären, wäre der Effekt möglicherweise noch größer gewesen. Die erste empirisch-inhaltliche Hypothese (EIH_1) hat sich also bewährt: Personen, die im Computerspiel Oblivion Gegner töten müssen, ohne dabei in irgendeiner Weise Hilfe auszuüben, stellen im CRTT intensivere Rauschtöne ein als Personen, die in Oblivion keine Gegner töten müssen.

Die zweite Testhypothese (TH_2) wurde abgelehnt und die Nullhypothese angenommen. Personen, die im Computerspiel Oblivion eine hilfebedürftige Spielfigur retten mussten, indem sie Gegner töten, stellten also keine intensiveren Rauschtöne im CRTT ein als Personen, die in Oblivion keine Gegner töten mussten. Das Ergebnis resultiert aber aus dem Umstand, dass sich die Nothilfebedingung zwar signifikant von der Hilfebedingung unterschied, aber der Unterschied zur Schatzjagdbedingung knapp nicht signifikant wurde. Die Effektstärken hingegen sind beide vergleichbar mit der aus der Meta-Analyse von Anderson et al. (2010) und können als praktisch bedeutsam bewertet werden. Somit lässt sich festhalten, dass sich die zweite empirisch-inhaltliche Hypothese (EIH_2) nur bedingt nicht bewährt hat. „Bedingt“ nicht bewährt deshalb, weil auch der Unterschied zwischen Nothilfe- und Schatzjagdbedingung möglicherweise signifikant geworden wäre, wenn die Präzision der Untersuchung geringfügig höher gewesen wäre.

Die dritte Testhypothese (TH_3) wurde abgelehnt und die Nullhypothese angenommen. Dabei unterscheidet sich die Tötenbedingung auch von der Effektgröße her nur marginal von der Nothilfebedingung. Die dritte empirisch-inhaltliche Hypothese (EIH_3) hat sich daher nicht bewährt.

13.2 HILFEVERHALTEN

Im Folgenden sollen die drei Hypothesen zum Hilfeverhalten von der Ebene der Testhypothese zurück auf die Ebene der empirisch-inhaltlichen Hypothese abgeleitet werden.

Die vierte Testhypothese wurde abgelehnt und die Nullhypothese angenommen. Die Anzahl der Eckenzählaufgaben ist in der Hilfebedingung zwar größer als in den hilfefreien Bedingungen, aber die Unterschiede sind nicht groß genug, um signifikant zu werden. Im Vergleich zu der Effektgröße aus der Gewaltspiel-Meta-Analyse von Anderson et al. (2010) fallen die Effekte in dieser Studie ca. ein Drittel geringer aus, während sie im Vergleich zu den Prosozialspiel-Studien von Greitemeyer und Osswald (2009; Studie 2) bzw. von Gentile et al. um die Hälfte bis drei Viertel geringer sind. Die Ursache für die geringe Effektstärke dürfte zum größten Teil an dem schwachen Treatment „Hilfe im Spiel“ liegen. Mit einem stärkeren Treatment wären die Unterschiede zwischen den Gruppen möglicherweise signifikant geworden. Es könnte aber auch eine Rolle gespielt haben, dass es sich beim Eckenzählen um eine Gefälligkeit mit potenziell höheren Kosten handelt, aufgrund dessen ebenfalls kleinere Effekte resultieren dürften. Die vierte empirisch-inhaltliche Hypothese (EIH₄) hat sich daher nur bedingt nicht bewährt.

Die fünfte Testhypothese wurde abgelehnt und die Nullhypothese angenommen. Post-hoc durchgeführte Analysen zeigen, dass anscheinend der umgekehrte Zusammenhang gilt, d. h. in der Nothilfebedingung ist die Anzahl der Eckenzählaufgaben nicht größer, sondern kleiner. Dieser umgekehrte Zusammenhang weist außerdem eine praktisch bedeutsame Effektstärke auf, die sogar noch über der aus der Meta-Analyse von Anderson et al. (2010) liegt, vergleichbar mit der Effektstärke bei Gentile et al. (2009) ist, und immerhin halb so groß wie die bei Greitemeyer und Osswald (2009; Studie 2). Daher hat sich die fünfte empirisch-inhaltliche Hypothese (EIH₅) nicht bewährt.

Die sechste Testhypothese wurde angenommen. Der Effekt fällt doppelt so hoch aus wie der vergleichbare Wert aus der Meta-Analyse von Anderson et al. (2010), größer als der bei Gentile et al. (2009), aber immer noch nicht so hoch wie der bei Greitemeyer und Osswald (2009; Studie 2). Zusammengenommen ist er praktisch bedeutsam. Die sechste empirisch-inhaltliche Hypothese (EIH₆) hat sich somit bewährt.

14 DISKUSSION DER VALIDITÄT

14.1 GEWALTVERHALTEN

Aus der Hypothesentestung resultierte das Ergebnis, dass Personen, die im Computerspiel Oblivion Gegner töten müssen, ohne dabei in irgendeiner Weise Hilfe auszuüben, im CRTT intensivere Rauschtöne einstellen als Personen, die in Oblivion keine Gegner töten müssen. Damit daraus gefolgert werden kann, dass reines Gewaltverhalten im Computerspiel im Vergleich zu gewaltfreiem Verhalten das Gewaltverhalten erhöht, müssen Variablenvalidität und interne Validität gegeben sein, d. h.:

1. Die Versuchsbedingungen im Spiel müssen gewalthaltig vs. gewaltfrei bzw. hilfreich vs. hilfefrei gewesen sein (Validität der UV),
2. der CRTT muss das Gewaltverhalten valide erfasst haben (Validität der AV) und
3. der Effekt der Versuchsbedingungen auf das Gewaltverhalten darf nicht durch Drittvariablen gestört worden sein (interne Validität).

14.1.1 VALIDITÄT DER UV

Hinsichtlich der UV „Gewalt im Spiel“ ist die Variablenvalidität gegeben, da der Manipulation Check zeigt, dass bei der Gewalthaltigkeit des Spiels zwischen den gewalthaltigen und den gewaltfreien Versuchsbedingungen ein großer Unterschied besteht. Hingegen besteht hinsichtlich der UV „Hilfe im Spiel“ bei der Hilfehaltigkeit des Spiels nur ein kleiner Unterschied zwischen den hilfreichen und den hilfefreien Bedingungen. Dadurch wird die Variablenvalidität zwar nicht gestört, es ist aber mit kleineren Effekten zu rechnen.

14.1.2 VALIDITÄT DER AV

Die Variablenvalidität der AV „Gewaltverhalten“ ist nur bedingt gegeben. Zum einen lässt die Inhaltsanalyse der Probendenaussagen vermuten, dass beim CRTT eine größere Anzahl an Probanden bemerkt hat, dass der Test Aggression misst bzw. dass sie im Test nicht gegen einen menschlichen Mitspieler angetreten sind. Es ließ sich leider nicht für jeden einzelnen Fall klären, ob diese Erkenntnisse tatsächlich vor der Durchführung des CRTT gewonnen wurden. Dafür spricht aber, dass die Probanden tatsächlich niedrigere Werte im CRTT einstellten. Dieser Effekt war zwar nicht sehr groß, aber in Grenzen mit der Effektgröße der UV vergleichbar. Die ANCOVA zeigt aber, dass die von der Störvariable aufgeklärte Varianz größtenteils unabhängig von der Varianz ist, die der Haupteffekt aufklärt.

Ein weiteres potenzielles Validitätsproblem der AV „Gewaltverhalten“ zeigte sich darin, dass die Probanden des Sommersemesters 2010 im Durchschnitt höhere Werte im CRTT aufwiesen als die des Wintersemesters 2010/11. Die Effektgröße ist vergleichbar mit der der UV und kann somit als praktisch bedeutsam angesehen werden. Der Verdacht liegt nahe, dass

sich zwischenzeitlich Informationen über das Experiment herumgesprochen haben könnten. Es könnte aber auch sein, dass zwischen den Semestern in den Medien verstärkt über Gewaltspiele berichtet wurde, dass es Lehrveranstaltungen zu Gewaltspielen gab etc. Abschließend klären lässt sich dies leider nicht. Die ANCOVA zeigte aber auch hier, dass die von der Störvariable aufgeklärte Varianz größtenteils unabhängig von der Varianzaufklärung des Haupteffekts ist.

Zusammengenommen könnte das bedeuten, dass die Einstellungen im CRTT stärker von anderen Variablen als der Aggression der Probanden beeinflusst werden, dass die Aggression aber trotzdem noch einen messbaren und unabhängigen Einfluss auf die Einstellungen ausübt. Die geminderte Variablenvalidität scheint sich somit weniger auf die interne Validität, sondern vielmehr auf die Präzision der Untersuchung niederzuschlagen. Somit könnte man von einer ausreichend validen Messung ausgehen, gäbe es nicht ein grundsätzliches ungelöstes Problem, von dem sämtliche Verfahren zur Aggressionsmessung betroffen sind: Die mangelhafte Validierung.

Zum einen fehlt eine Kriteriumsvalidierung an einem Außenkriterium, was damit zusammenhängen dürfte, dass die Verfahren Verhalten erfassen sollen. Um ein Verfahren mit dem Kriterium zu vergleichen, müsste aggressives Verhalten induziert werden, was schnell zu ethischen Problemen führen dürfte. Ersatzweise ließe sich natürlich auftretendes aggressives Verhalten heranziehen (z. B. in Form von Lehrerbeobachtungen), dann ließe sich aber nicht klären, zu welchem Anteil das (vom Lehrer) beobachtete Verhalten von der Situation und zu welchem Anteil es von der Persönlichkeit verursacht wird. Zum anderen fehlt ebenfalls eine belastbare Konstruktvalidierung. Hier liegt das Problem darin, dass die Konvergenzvalidität ungeklärt bleibt, weil es eben kein Messverfahren gibt, welches zweifelsfrei aggressives Verhalten erfasst. Wie das Beispiel von Beier und Kutzner (2011a; 2011b) zeigt, ist es im Rahmen einer Konstruktvalidierung immerhin möglich, die Schädigungsabsicht sicherzustellen. Eine solche Validierung liegt bisher aber nur für das Hot Sauce Paradigm vor.

In dieser Studie wurden daher die Intentionen der Probanden direkt im Anschluss an den CRTT erhoben. Deren Analyse zeigte, dass aggressive Intentionen durchaus vorhanden waren. Die per Faktorenanalyse aus den Intentionen gebildeten Motive können außerdem das Gewaltverhalten vorhersagen (im Fall des Bestrafungsmotivs sogar signifikant). Die Analyse der Motive zeigte, dass die Probanden in den gewalthaltigen Bedingungen eher aggressive und in den gewaltfreien Bedingungen eher ein nicht-aggressives Motiv verfolgten, was ebenfalls für die Validität spricht.

Auf der anderen Seite waren die aggressiven Intentionen nicht sonderlich stark ausgeprägt, während die nicht-aggressiven Intentionen sogar stärker ausgeprägt waren. Dies könnte auf soziale Erwünschtheit rückführbar sein, die möglicherweise dadurch verstärkt wurde, dass viele Probanden das Täuschungsmanöver durchschaut haben. Wie bereits weiter oben beschrieben scheint das Durchschauen der Täuschung aber weniger die Validität, sondern eher

die Präzision der Untersuchung gemindert zu haben. Problematischer erscheint es, dass die Zusammenhänge zwischen den Motiven und den Versuchsbedingungen sehr klein und nicht signifikant waren, weswegen nicht auszuschließen ist, dass das Motiv stärker von der Persönlichkeit eines Probanden als von den Versuchsbedingungen abhängt. Eine Überprüfung dieser Vermutung ist in diesem Experiment aber nicht möglich, da aufgrund der Ergebnisse des zweiten Vortests auf den Einsatz entsprechender Persönlichkeitsfragebögen verzichtet wurde.

Es wäre aber auch möglich, dass die Aussagen der Probanden über ihre Intentionen nicht valide sind, da sie bei ihren Angaben nicht offen gewesen sein könnten (z. B. durch sozial erwünschte Antworten) und/oder da sie sich über ihre eigentlichen Motive (z. B. zum Schutz ihres Selbstwerts) getäuscht haben könnten. Im Rahmen rein beobachtender Forschung lässt sich prinzipiell keine sichere Entscheidung darüber treffen, ob ein Motiv vorliegt oder nicht, da Motive nicht direkt beobachtbar sind und die Validität der Probandenaussagen im Experiment nicht gesichert werden kann. Hierfür wäre ein zweistufiges Vorgehen wie das von Groeben (1986) vorgeschlagene nötig, bei dem in der ersten Stufe eine subjektive Theorie inklusive der Motive in einem Interview erhoben wird (mehr zu subjektiven Theorien findet sich bei Groeben, Wahl, Schlee & Scheele, 1988). Die Validität der Probandenaussagen wird dabei über Konsensvalidierung sichergestellt. Der Interviewer versucht, durch Annäherung an Habermas „ideale Sprechsituation“ eine Interviewsituation herzustellen, die von Gleichberechtigung, Partnerschaftlichkeit bzw. Hierarchiearmut geprägt ist. Bei Gelingen führt dies in der Regel zu offeneren Antworten des Befragten. Erst in einer zweiten Stufe wird die Relevanz der subjektiven Theorien für das Handeln der Probanden über Experimente geprüft (Handlungsvalidierung). Daher kann hier nicht mit Sicherheit von den Aussagen der Probanden auf ihre Motive geschlossen werden.

Die Ergebnisse zur Validität der AV Gewaltverhalten lassen sich wie folgt zusammenfassen: **Wenn man den Aussagen der Probanden traut und daneben keine Konfundierung der Versuchsbedingungen mit der Probandenpersönlichkeit unterstellt, dann sollte die Variablenvalidität gegeben sein, allerdings bei eingeschränkter Präzision.** Wenn man dagegen den Aussagen der Probanden misstraut oder die genannte Konfundierung unterstellt, bleibt die Variablenvalidität ungeklärt, da einschlägige Validierungsstudien fehlen, die das Gegenteil belegen könnten.

14.1.3 INTERNE VALIDITÄT

Interne Validität scheint gegeben zu sein, da hinsichtlich der AV Gewaltverhalten bei der Analyse der Konfundierungen keine problematischen Zusammenhänge gefunden wurden.

14.2 HILFEVERHALTEN

14.2.1 VALIDITÄT DER UV

Wie bereits erwähnt, besteht bei der Hilfefähigkeit des Spiels nur ein kleiner Unterschied zwischen den hilfefähigen und den hilfefreien Bedingungen, weswegen mit kleineren Effekten zu rechnen ist. Dieses Problem wird dadurch verschärft, dass es sich bei der AV um „mittelteures“ Hilfeverhalten handelt, aufgrund dessen ebenfalls kleinere Effekte resultieren dürften. Von diesem Problem sind alle drei Hypothesen zum Hilfeverhalten betroffen.

14.2.2 VALIDITÄT DER AV

Die Validität der AV Hilfeverhalten ist schon dadurch eingeschränkt, dass die Probanden nur die Anzahl an Aufgaben angeben mussten, diese im Anschluss aber nicht abarbeiten mussten, da das Programm vorher abbrach. Dadurch wurde nicht ihr Hilfeverhalten erhoben, sondern nur ihre Hilfsbereitschaft. Da die Probanden aber nicht wissen konnten, dass das Programm abbrechen wird, sollte die so erhobene Hilfsbereitschaft ziemlich verhaltensnah ausfallen. Dennoch wäre es denkbar, dass einige Probanden die Aufgabenbearbeitung nach einiger Zeit (z. B. aus Langeweile, Frustration etc.) abgebrochen hätten, wenn ihnen das möglich gewesen wäre.

Ein weiteres Problem ergibt sich aus verwendeten Cover Story. Die Aufgaben wurden als Konzentrationstest ausgegeben. Die Analyse der Störvariable „subjektive Aktivierung“ zeigte, dass diejenigen Probanden hilfsbereiter sind, die sich munterer bzw. weniger lustlos fühlen. Dieser Befund kann so gedeutet werden, dass das Verfahren durch seine Verkleidung als „Konzentrationsaufgabe“ möglicherweise nicht (nur) die Hilfsbereitschaft der Probanden erhebt, sondern (auch) ihre subjektive Wachheit. Dazu passt auch, dass in der ANCOVA neben der Aktivierung selbst zwei weitere Störvariablen eine Rolle spielten, die mit der Aktiviertheit der Probanden zusammenhängen könnten, nämlich die Aufgabenschwierigkeit (schwierige Aufgaben könnten stärker aktivieren) und die Realitätsnähe der Grafik (immersivere Szenarien könnten stärker aktivieren). Das Ergebnismuster der Hypothesentestung spricht zwar dagegen, dass das Verfahren (ausschließlich) die Wachheit erhebt, aber letztendlich lässt sich das nur durch eine Validierungsstudie überprüfen, die bisher nicht durchgeführt wurde.

14.2.3 INTERNE VALIDITÄT

Die Analyse der Konfundierungen identifizierte zwar eine Reihe von Störvariablen, die mit der UV einen Zusammenhang aufwiesen, zeigte aber auch, dass die gefundenen Zusammenhänge nahezu alle als unproblematisch angesehen werden können. Lediglich bei den Störvariablen „Interessantheit der Spielaufgabe“ bzw. „Spannungsgehalt der Spielaufgabe“ stellte sich die subjektive Aktivierung der Probanden als Mediator zwischen diesen Störvariablen

Virtuelle Nothilfe

und der AV Hilfeverhalten heraus. Dies wäre kein Problem, wären die Störvariablen nicht gleichzeitig mit der UV „Gewalt im Spiel“ teilweise vermischt: Die Spielaufgaben in den gewaltfreien Bedingungen (d. h. „Schleichen“ statt „Kämpfen“) wurden als interessanter bzw. spannender empfunden, was anscheinend zu einer erhöhten Aktivierung geführt hat, die wiederum das Hilfeverhalten gesteigert hat. Daher ist es schwierig zu differenzieren, zu welchem Anteil das Hilfeverhalten von der fehlenden Gewalt in den Spielaufgaben und zu welchem es von dem wach machenden Effekt ihres Spannungsgehalts bzw. ihrer Interessantheit verursacht wurde. Auf der anderen Seite zeigte sich bei der Hypothesentestung, dass die „interessanteren“ Bedingungen nicht zu mehr Hilfeverhalten führten als die „uninteressantere“ Tötenbedingung. Das gefundene Muster scheint vielmehr stärker von den spezifischen Eigenschaften der Nothilfebedingung verursacht zu werden als von der „Interessantheit“ der Spielaufgaben. Dazu passt auch das Ergebnis der AN(C)OVA, die einen signifikanten Interaktionseffekt findet, der von der Nothilfebedingung verursacht wird. Außerdem zeigt die ANCOVA, dass die von den Störvariablen aufgeklärte Varianz zum größten Teil unabhängig ist von der Varianz, die der Haupteffekt „Gewalt im Spiel“ aufklärt. Zu diesen Störvariablen zählt die „Spielaufgabe Aktivierung (lustlos)“ sowie die potenziell aktivierungsnahen Störvariablen „Realitätsnähe der Grafik“ und „Aufgabenschwierigkeit“. Daher wird davon ausgegangen, dass durch den Mediationseffekt zwar eine Einschränkung der internen Validität vorliegt, diese aber praktisch nicht bedeutsam ist.

15 DISKUSSION DER THEORETISCH-INHALTLICHEN HYPOTHESEN

Nach der Bewertung der empirisch-inhaltlichen Hypothesen und der aufgetretenen Störungen der Validität sollen im Folgenden alle sechs Hypothesen auf die Ebene der theoretisch-inhaltlichen Testhypothese abgeleitet werden.

15.1 GEWALTVERHALTEN

Wie oben dargelegt ist es strittig, ob der CRTT das Gewaltverhalten valide erfasst hat. Unter der Annahme, dass dies der Fall ist, lässt sich folgern, dass reines Gewaltverhalten im Computerspiel das Gewaltverhalten erhöht. Dieses Ergebnis ist wenig überraschend, steht es doch sowohl im Einklang mit bisherigen Gewaltspielstudien als auch mit den einschlägigen Theorien (d. h. dem General Aggression Model und dem Moral Management Model). Wie der spezifische Wirkmechanismus dabei aussieht, bleibt aber unklar. Das General Aggression Model postuliert hier Beobachtungslernen aggressiven Verhaltens, Priming aggressiver Kognitionen, Erzeugen aggressiver Emotionen sowie Erhöhung physiologischer Erregung. Das Moral Management Model dagegen postuliert einen Belohnungswert durch aggressives Verhalten (Lernen am Erfolg, Ausüben von Macht, Faszination durch Zerstörung), der mit moralischen Zweifeln konfligiert, die daher abgewehrt werden (müssen). Im Rahmen dieser Studie wurden keine Variablen gemessen, mit denen geprüft werden könnte, welche der genannten Wirkungsmechanismen zum Tragen kamen. Für eine Folgestudie wäre es daher beispielsweise interessant, zu erheben, ob und wann moralische Gefühle bzw. Zweifel auftreten, wie leicht sie abgewehrt werden etc., und wie sich dies auf das Gewaltverhalten nach dem Spiel auswirkt.

Dass Nothilfe im Computerspiel ebenfalls das Gewaltverhalten erhöht, ließ sich dagegen nicht zweifelsfrei bestätigen. Die fehlende Signifikanz des Effekts dürfte auf die mangelnde Präzision zurückzuführen sein. Hier könnten aufseiten der AV das Durchschauen des Täuschungsmanövers und das Überwiegen nicht-aggressiver Intentionen sowie aufseiten der UV die Schwäche des Treatments „Hilfe im Spiel“ eine Rolle gespielt haben. Möglicherweise ist das Ergebnis aber auch eine direkte Folge des gewählten Nothilfeszenarios. In der Nothilfebedingung stahlen die Täter eine lebenswichtige Medizin. Durch diese Tat wurde das Leben des Opfers nur indirekt bedroht, und es blieb offen, ob die Täter dabei eine Schädigung des Opfers beabsichtigten. Außerdem konnten die Probanden die Vertrauenswürdigkeit des Opfers nur anhand weniger Hinweisreize einschätzen. Es wäre möglich, dass die Probanden vermutet haben, dass sie vom scheinbaren Opfer getäuscht werden, damit es sie für seine Zwecke instrumentalisieren kann. Ein solcher dramaturgischer Kniff wäre in einem Rollenspiel nichts Ungewöhnliches. Die Befunde sprechen also weniger gegen die Gültigkeit der theoretischen Annahmen, sondern mehr gegen die gewählte Form der Operationalisierung. Es lässt sich daher weiterhin vermuten, dass Nothilfe im Computerspiel möglicherweise zu

erhöhtem Gewaltverhalten führt; dies müsste allerdings in einer Folgestudie geprüft werden.

Das dritte Ergebnis bestand darin, dass sich zwischen reinem Gewaltverhalten und Nothilfe kein Unterschied fand. Diese Vorhersage wurde aus der Zivilcouragetheorie abgeleitet, nach der Gewalt beim helfenden Eingreifen möglichst vermieden wird. Das scheint zumindest in dem hier verwendeten Nothilfeszenario nicht zuzutreffen. Das Moral Management Model scheint für das Ergebnis aber auch keine gute Erklärung zu liefern, da es vorhergesagt hätte, dass die Nothilfebedingung wegen der ethischen Rechtfertigung für die Gewalt das Gewaltverhalten stärker fördert als die reine Gewaltbedingung. Dieses Ergebnis ist ebenfalls nicht eingetreten. Am besten lässt es sich mit dem General Aggression Model in Einklang bringen, welches keinen Unterschied zwischen moralisch gerechtfertigter und ungerechtfertigter Gewalt macht. Allerdings entspricht das gefundene Ergebnis nicht der Vorhersage, weswegen die Irrtumswahrscheinlichkeit mit 20% vergleichsweise hoch ausfällt. Daher wäre auch hier eine Replikation dieses Ergebnisses notwendig, um eine besser abgesicherte Aussage treffen zu können.

15.2 HILFEVERHALTEN

Auch bei den Hypothesen zum Hilfeverhalten ist es strittig, ob das Hilfeverhalten valide erfasst wurde. Hier müssen die Annahmen gemacht werden, dass das gewählte Verfahren zur Messung des Hilfeverhaltens nicht nur die Bereitschaft zur Hilfe erhoben hat, sondern tatsächliches Hilfeverhalten, und dass es nicht die Wachheit der Probanden erhoben hat. Aber selbst mit diesen Annahmen ließ sich nicht zweifelsfrei bestätigen, dass reines Hilfeverhalten im Computerspiel das Hilfeverhalten erhöht. Die Richtung der Vorhersage stimmte zwar, aber der Effekt war zu klein, um signifikant zu werden. Auch hier dürfte wieder die Schwäche des Treatments „Hilfe im Spiel“ eine Rolle gespielt haben. Hinzu kommt, dass das Hilfeverhalten mit potenziell höheren Kosten verbunden war als die übliche Gefälligkeit. Das Ergebnis spricht damit weniger gegen die theoretischen Annahmen und das dahinterstehende General Learning Model, sondern mehr gegen die gewählte Operationalisierung, weswegen sich weiterhin vermuten lässt, dass reine Hilfe im Computerspiel möglicherweise zu erhöhtem Hilfeverhalten führt.

Anders dagegen sieht es bei der Wirkung von Nothilfe aus. Hier wurde aus dem General Learning Model abgeleitet, dass Nothilfe zu mehr Hilfeverhalten führt, aber das Gegenteil scheint der Fall zu sein. Damit spricht das Ergebnis nicht nur gegen das General Learning Model, sondern lässt sich auch schlecht mit der Zivilcouragetheorie vereinbaren, die Nothilfe als Extremform helfenden Verhaltens modellieren müsste. Am ehesten lässt es sich mit dem Moral Management Model vereinbaren, nach welchem Nothilfe eine ethisch gerechtfertigte Form von Gewalt darstellt, bei der moralische Zweifel leichter abgewehrt werden können oder gar nicht erst aufkommen. Da dieses Ergebnis aber wieder nicht der Vorhersage ent-

spricht und einer Irrtumswahrscheinlichkeit von 20% unterliegt, wäre eine Replikation notwendig.

Da Nothilfe das Hilfeverhalten anscheinend verringert, ist es nicht verwunderlich, dass sich die letzte Vorhersage bestätigen ließ, dass reine Hilfe zu mehr Hilfeverhalten führt als Nothilfe. Dieses Ergebnis steht im Einklang mit der für die Hypothesenbildung genutzten Zivilcourage-theorie, ließe sich aber auch mit dem Moral Management Model erklären. Wenn unterstellt wird, dass die Gewaltanteile in der Nothilfebedingung die Hilfeanteile übertroffen haben, ließe sich das Ergebnis sogar mit dem General Learning Model vereinbaren. Die Entscheidung für eine der Theorien ist wahrscheinlich deswegen so schwierig, weil im Rahmen dieser Studie keine Variablen gemessen wurden, mit denen der konkrete Wirkmechanismus geprüft werden könnte.

15.3 GENERALISIERBARKEIT

Nach der Diskussion der Ergebnisse stellt sich die Frage nach der Generalisierbarkeit auf andere Populationen und Situationen.

15.3.1 POPULATIONEN

In die Auswertung gingen nur Studenten ein, die bereits Erfahrung mit der Steuerung von 3D-Spielen haben. Daher können die Ergebnisse zum jetzigen Zeitpunkt weder auf Frauen noch auf Personen ohne Steuerungserfahrungen verallgemeinert werden. Die Generalisierbarkeit ist weiterhin dadurch eingeschränkt, dass es sich um eine Gelegenheitsstichprobe und nicht um eine Zufallsstichprobe handelt. Es ist denkbar, dass Mechanismen der Selbstselektion zum Tragen gekommen sind. Zum Beispiel könnten Probanden, die von ihren Fähigkeiten her zur Teilnahme geeignet gewesen wären, durch die Werbung den falschen Eindruck gewonnen haben, dass ihre Fähigkeiten zur Teilnahme nicht ausreichen. Solche Störeffekte lassen sich nur bei einer ausreichend großen Zufallsstichprobe sicher ausschließen. Auf der anderen Seite wird laut Hussy und Jain (2002, S. 193f.) in der Psychologie fast nie mit Zufallsstichproben gearbeitet, und trotzdem ist es möglich, zu neuen Erkenntnissen zu gelangen. Die Autoren vertreten daher die Ansicht, dass auch Gelegenheitsstichproben zur Hypothesenprüfung geeignet sind, solange die Unabhängigkeit der Messwerte gewährleistet ist (was hier der Fall ist). Da keine Mechanismen der Selbstselektion erkennbar wurden, wird (vorläufig) davon ausgegangen, dass die Gelegenheitsstichprobe zur Hypothesentestung geeignet war.

15.3.2 SITUATIONEN

Die Verallgemeinerbarkeit auf andere Situationen hängt eng mit der Operationalisierung zusammen. Auf der Seite der UV lässt sich die Frage stellen, ob sich die Ergebnisse auf andere Genres verallgemeinern lassen. Oblivion stellt einen typischen Vertreter der Gattung „narra-

tives Action-Rollenspiel in der Ego-Perspektive“ dar. Daher wird angenommen, dass die Geltung durchaus für andere Action-Rollenspiele dieser Art beansprucht werden kann. Darüber hinaus wird angenommen, dass die Geltung sogar für Action-Adventures in der Ego-Perspektive (einschließlich Ego-Shooter) beansprucht werden kann, solange diese vergleichbar dichte Geschichten erzählen, bei denen ein Mitfühlen mit den Charakteren ermöglicht wird²⁷.

Neben dem Genre stellt sich die Frage, ob die Ergebnisse aus dem Labor auf die reale Welt übertragbar sind. Grundsätzlich besteht die Gefahr, dass die Probanden im Labor ein anderes Verhalten zeigen, als sie dies unter natürlichen Bedingungen tun würden. Zumindest im Rahmen der Spielsituation fanden sich keine Hinweise hierfür. Die Generalisierbarkeit ist aus zwei Gründen dennoch eingeschränkt. Zum einen bewerteten die Probanden das Spiel nur als mittelmäßig interessant, unterhaltsam etc. (siehe Anhang E), obwohl es sich um ein populäres und aktuelles Spiel gehandelt hat. Dies könnte daran gelegen haben, dass die Probanden im Alltag die Wahl haben und sich so das Spiel aussuchen können, das ihnen am besten gefällt, während dies im Labor nicht möglich war. Wahrscheinlich reduziert dies die Intensität des Spielerlebnisses und damit die Wirkung des Spiels. Zum anderen wurde das Spiel nur für einige Minuten gespielt, während Rollenspiele im Alltag zum Teil über hunderte von Stunden konsumiert werden. Auch dies dürfte die Wirkung eingeschränkt haben. Anderson et al. (2007; 2010) nehmen zwar an, dass kleine kurzfristige Effekte einer einzelnen Spielepisode sich zu einem großen langfristigen Effekt aufaddieren (siehe Abschnitt 4.3.7). Eine solche Kumulation wurde bisher jedoch nur in korrelativen Langzeitstudien nachgewiesen, nicht aber in experimentellen. Damit bleibt die alternative Möglichkeit bestehen, dass die flüchtigen Priming-Effekte nur für kurze Zeit verhaltensrelevant werden (ähnlich subliminaler Werbebotschaften).

Auf der Seite der AV Gewaltverhalten lässt sich daran zweifeln, dass das Messverfahren eine Verallgemeinerung auf andere Formen aggressiven Verhaltens erlaubt. Erstens gibt es bei dem Verfahren einen Versuchsleiter, der das aggressive Verhalten toleriert. Zweitens besteht eine räumliche Distanz zwischen den Konfliktparteien. Drittens gibt es zur Gewalt keine Verhaltensalternative, mit der sich der Konflikt austragen oder klären ließe. Viertens werden möglicherweise sehr spezielle Formen von Aggression erhoben, die durch Retribution, Feindseligkeit oder Wegattribution der Verantwortung ausgelöst wurden. Damit sind die Ergebnisse lediglich auf Situationen übertragbar, in denen die Gewalt von einer Autorität erlaubt wird, in denen die Konfliktparteien sich nicht von Angesicht zu Angesicht gegenüberstehen, und in denen es keine Alternative zur Gewalt gibt. Möglicherweise ist die Generalisierung auch nur dann möglich, wenn die Gewalt durch Rache, Feindseligkeit oder Absprechen der Verantwortung ausgelöst wurde. Im Alltag sind aber Situationen wesentlich üblicher, in denen sich Aggressor und Opfer gegenüberstehen und die Ausübung von physischer Gewalt echten Körperkontakt erfordert. Dabei wird die Gewalt nicht von Autoritäten erlaubt, sondern bestraft. Außerdem wird vor dem Ausführen einer Gewalthandlung meistens versucht,

Virtuelle Nothilfe

den Konflikt gewaltfrei beizulegen. Somit scheint es nicht möglich zu sein, die hier gefundenen Ergebnisse auf typische Alltagssituationen zu übertragen.

Auf der Seite der AV Hilfeverhalten finden sich solche Einschränkungen bisher nicht. Das könnte aber auch daran liegen, dass die gewählte Operationalisierung „willingness to assist“ noch nicht validiert ist. Da sich die Operationalisierung an der im Alltag üblichen Gefälligkeit orientiert (wenn auch mit höheren Kosten), ist es aber naheliegend, zu vermuten, dass die Ergebnisse auf solche Gefälligkeitssituationen übertragen werden können. Eine Übertragung auf Hilfesituationen, in denen die Kosten sehr hoch sind (z. B. Notsituationen oder Zivilcoursituationen), scheint aber nicht angemessen zu sein.

Wenn man sich die Ergebnisse der Hypothesentestung anschaut, dann scheinen die meisten davon am ehesten mit der Theorie des Moral Management vereinbar zu sein. Allerdings sind die Ergebnisse (noch) nicht gut abgesichert und lassen sich (noch) nicht über die Gruppe der männlichen studierenden Hobbyspieler hinaus verallgemeinern. Außerdem bleibt offen, ob die Wirkung des Computerspiels lange genug anhält, um im Alltag relevant zu werden, und ob die im Labor gefundenen Ergebnisse auf Alltagssituationen übertragen werden dürfen, insbesondere beim Gewaltverhalten, wo zahlreiche Unterschiede zu normalen Konfliktsituationen bestehen (die Distanz zum Opponenten ist groß, eine Retribution durch den Opponenten ist unwahrscheinlich, die Gewalt wird von einer Autorität gebilligt und stellt die einzige vorhandene Option zur Konfliktlösung dar).

Andere zum Moral Management Model durchgeführte Studien lassen aber zumindest befürchten, dass der regelmäßige Konsum von Spielen mit gerechtfertigter Gewalt möglicherweise das moralische Denken (auch) dauerhaft reduzieren könnte. Interessanterweise findet sich sogar in einer Studie zum Hilfeverhalten ein Ergebnis, welches sich mit gerechtfertigter Gewalt erklären lässt: Bei Greitemeyer und Osswald (2010; Studie 3) stellte sich unerwarteter Weise heraus, dass einige Probanden anscheinend den „Ex-Freund“ der Versuchsleiterin unnötigerweise verletzen wollten, um diese aus einer Bedrängnis zu retten. Es wäre also wichtig, weitere Studien durchzuführen, die überprüfen, unter welchen Bedingungen eine moralische Rechtfertigung die Wahrscheinlichkeit für Gewalt erhöht.

Angesichts des möglichen Risikos könnte es empfehlenswert sein, auf den Konsum von Gewaltspielen zu verzichten. Aber das würde bedeuten, das Kind mit dem Bade auszuschütten, da diese Spiele als sehr unterhaltsam bewertet werden und durch die Unterhaltung positive Wirkungen wie z. B. Entspannung ermöglicht werden. Daher wäre es sinnvoller, zuerst nach Schutzfaktoren zu suchen, die die mögliche Negativwirkung verhindern können. Da das unerwünschte Verhalten vom Ausbleiben des moralischen Denkens moderiert wird, wäre es eventuell hilfreich, dieses Denken zu fördern. Das könnte zum Beispiel dadurch geschehen, dass die virtuellen Verstöße gegen die Moral nicht aus dem Bewusstsein verdrängt werden. Stattdessen sollten die moralischen Zweifel ganz im Sinne einer kathartischen Purifikation erlebt, durchlebt und ausgehalten werden. Dies ließe sich auf zwei Wegen bewerkstelligen.

- Die erste Methode könnte darin bestehen, beim Konsum von unmoralischen (Gewalt-)Spielen zwischendurch kognitiven Abstand zu gewinnen und sich die moralischen Implikationen der eigenen virtuellen Handlungen bewusst zu machen. Das dürfte angesichts der immersiven Sogwirkung und der zahlreichen Moral Disengagement Cues aber nicht gerade einfach werden und zu einer Reduktion des Unterhaltungswertes führen.

Virtuelle Nothilfe

- Eine andere Möglichkeit wäre es, Gewaltspiele mit moralischem Anspruch zu spielen. Ein illustratives Beispiel für ein solches Spiel stellt „Nier Gestalt“ vom Entwicklungsstudio Cavia (2010) dar. In diesem Spiel kämpft der Protagonist gegen Schattenwesen, die seine Tochter entführt haben. Am Ende des Spiels erfährt der Spieler, dass die Schattenwesen Menschen waren, die von einem außerirdischen Virus befallen wurden, der sie radikal verändert hat. Durch die Veränderung wurden die Menschen zu hüllenlose Seelen, welche nicht mehr mit Menschen kommunizieren können. Nach dem Ende lässt sich das Spiel ab der Mitte erneut starten, um ein weiteres Ende der Geschichte freizuschalten. Im Verlauf dieses zweiten Endes kann der Spieler die Sprache der Schattenwesen verstehen. Dadurch erfährt er auf dramatische Art und Weise, dass viele seiner Taten moralisch fragwürdig waren. Beispielsweise weiß er nun, dass ein kleiner Schatten, der anscheinend die Kontrolle über einen Kampfroboter übernommen hatte, in Wirklichkeit die lebendig gewordene „Seele“ eines Kindes war, dessen Mutter von Menschen gejagt und getötet wurde, sodass es einsam und allein verblieb. In seiner Einsamkeit befreundete es sich mit einem ebenso einsamen Roboter und erkundete mit diesem die Welt. Der Spieler wusste nicht um diese Umstände, als er in den Lebensraum des Schattenkindes und des Roboters eindrang. Er begriff das Duo als Bedrohung, die beseitigt werden musste. Durch diesen Kunstgriff verkehrt das Spiel die ursprüngliche moralische Rechtfertigung in ihr genaues Gegenteil eines moralischen Tabubruchs. Das dürfte bei den Spielern und Spielerinnen moralische Zweifel auslösen, die nicht so leicht abzuwehren sind. Es besteht aber auch die Gefahr, dass in so einem Fall der Unterhaltungswert leidet, wodurch wiederum die angenehme und entspannende Seite des Spiels reduziert wird. Auf Dauer könnte das Spiel ganz abgebrochen werden, um das Auftreten unangenehmer Gefühle zu vermeiden.

Es könnte sich für den Spieler aber lohnen, trotz eines gesenkten Unterhaltungswertes moralisierende Spiele zu spielen bzw. sich der moralischen Implikationen des virtuellen Handelns bewusst zu werden, da dies möglicherweise die moralische Entwicklung fördert. Scheele und DuBois (2006) sind überzeugt, dass „under certain circumstances, we consciously choose (morally) painful media experiences in order to achieve satisfaction of gaining in understanding“ (S. 417). Dies wäre allerdings für den Bereich der Computerspiele empirisch zu prüfen, was bisher nicht geschehen ist. Die Befunde zum Purifikationseffekt aus anderen medialen Bereichen lassen eine solche Prüfung lohnenswert erscheinen.

17 STÄRKEN, SCHWÄCHEN UND OPTIMIERUNGSVORSCHLÄGE

17.1 STÄRKEN

Die große Stärke der Studie liegen in der Maximierung der experimentellen Kontrolle.

- Erstens wurde die unabhängige Variable innerhalb desselben Computerspiels variiert, um identische Spielmerkmale in allen Versuchsbedingungen zu erhalten. Durch die experimentelle Variation war es überhaupt erst möglich, die kombinierte Wirkung von Hilfe und Gewalt im Spiel zu untersuchen.
- Zweitens wurde eine vergleichsweise große Stichprobe erhoben, um sicherzustellen, dass sich Störgrößen möglichst zufällig über die Bedingungen verteilen.
- Drittens wurde eine Vielzahl an Störvariablen erhoben, um überprüfen zu können, ob relevante Störgrößen sich tatsächlich zufällig über die Bedingungen verteilt haben.
- Viertens kann das Treatment in der vorliegenden Form als Grundlage für weitere Studien nützlich sein. Die Bewertung der Gewalthaltigkeit unterscheidet sich recht stark zwischen den gewaltlosen und den gewalthaltigen Bedingungen, sodass Untersuchungen zur Wirkung von virtueller Gewalt ohne weitere Optimierungen des Treatments machbar erscheinen.

17.2 SCHWÄCHEN UND OPTIMIERUNGSVORSCHLÄGE

Trotz der methodischen Sorgfalt weist die Studie aber auch mehrere Defizite auf.

- Das erste und größte Defizit besteht in der ungeklärten Validität der abhängigen Variablen. Beim Verfahren zur Messung des Gewaltverhaltens wurde bisher keine Validierung an einem Außenkriterium oder an einem anderen Verfahren, dass zweifelsfrei aggressives Verhalten erhebt, durchgeführt.
- Außerdem fehlt der Nachweis einer Schädigungsabsicht, weswegen das Verfahren möglicherweise etwas anderes als Aggression erhebt (z. B. soziale Kontrolle, Selbstdarstellung, Konformität, Wettbewerb etc.). Hier wäre eine Validierungsstudie notwendig, die sich an das Vorgehen von Beier und Kutzner (2011a; 2011b) anlehnen könnte. Für eine solche Studie sollte die Aggressive Motives Scale um Items ergänzt werden, mit denen auch Aussagen über weitere konkurrierende nicht-aggressive Motive gemacht werden können, wie z. B. Konformität.
- In dieser Studie hat anscheinend ca. ein Drittel der Probanden trotz des aufwendigen Täuschungsmanövers erkannt, dass das Verfahren Aggression messen soll. Dies wiederum hat offenbar zu sozial erwünschten Antworttendenzen geführt. Dieses Problem zu umgehen, stellt sich bei der Population der Computerspieler als schwierig dar. Personen, die regelmäßig Gewaltspiele konsumieren, werden in der Öffentlichkeit für dieses Freizeitverhalten häufig kritisiert. Dadurch bringen diese Hobbyspieler

Studien zu Computerspielen eine gewisse Skepsis entgegen, da sie vermuten, hier solle die Negativwirkung ihres Hobbys belegt und ein generelles Verbot von Gewaltspielen gerechtfertigt werden. Die anfängliche Skepsis wird dadurch weiter genährt, dass im CRTT unangenehme Rauschtöne verabreicht werden müssen, wofür keine Begründung geliefert wird. Nur auf den ersten Blick scheint es sinnvoll zu sein, den Probanden eine nicht-aggressive Alternativerklärung für die Verabreichung der Töne anzubieten. Auf den zweiten Blick kann denjenigen Probanden, die auf diese Alternativerklärung „hereinfallen“, keine Schädigungsabsicht mehr unterstellt werden. Hinzu kommt, dass das Taylor Aggression Paradigm, welches die Grundlage des CRTT ist, eine Verbesserung des Teacher/Learner Paradigm darstellt, welches in den Studien von Milgram Verwendung fand. Diese Studien sind aber inzwischen einer größeren Öffentlichkeit bekannt (z. B. aus den Medien oder dem Schulunterricht). Personen, die diese Studien kennen, entlarven das CRTT schnell als Verfahren zur Messung von Aggression. Dieser Sachverhalt spiegelt sich in den Kommentaren einiger Probanden wider, die bei der Erklärung des CRTT äußerten, dass sie das Verfahren an die „Studie mit den Stromstößen“ erinnern würde.

- In dieser Studie hat ca. ein Fünftel der Probanden nach der Aufklärung behauptet, erkannt zu haben, dass sie im CRTT nicht gegen einen anderen Probanden angetreten sind. Möglicherweise haben die Probanden eine Vorstellung davon, wie ihr Gegenspieler auf ihre Einstellungen zu reagieren hat. Werden ihre Annahmen enttäuscht, so führt das nicht selten dazu, dass die Probanden vermuten, der Gegner sei nicht echt. Auch dieser Sachverhalt findet sich in den Kommentaren einiger Probanden. Außerdem sprechen die Ergebnisse der Inhaltsanalyse ebenfalls dafür, da sich hier gezeigt hat, dass niemand die Existenz des Gegenspielers *vor* der Durchführung des CRTT angezweifelt hat, aber einer größeren Anzahl Probanden *während* des CRTT Zweifel kamen. Es scheint leider nicht möglich zu sein, die statischen Einstellungen im CRTT so zu wählen, dass möglichst wenige Probanden in ihren Annahmen enttäuscht werden, da jeder Proband eine andere Annahme haben kann. Beispielsweise stellen einige Probanden nur leise Töne ein und finden es unplausibel, dass ihr Gegner immer lauter wird, während andere direkt laute Töne einstellen und es unplausibel finden, dass ihr Gegner nicht mitzieht. Wieder andere finden es merkwürdig, dass sie in den Einstellungen des Gegners keine Systematik entdecken können. Um solche Probleme zu minimieren, müsste das CRTT wahrscheinlich dynamisch auf die Einstellungen der Probanden reagieren. Damit würde aber jeder Proband ein anderes Rauschton-„Treatment“ erhalten, wodurch die Messung des Gewaltverhaltens zwischen den Probanden nicht mehr vergleichbar wäre. Es erscheint daher vorteilhafter, nach der Datenerhebung nur die Daten aus der ersten Runde auszuwerten.
- Die Messpräzision des Verfahrens scheint nicht sonderlich hoch zu sein. Hier bleibt offen, ob dies an den oben genannten Problemen bei der Durchführung gelegen hat, oder ob die Messwerte möglicherweise statt vom aggressiven Zustand eher von der

aggressiven (bzw. antisozialen) Persönlichkeit des Probanden abhängen. Dieser Verdacht konnte mangels Erhebung der Persönlichkeit in diesem Experiment nicht abschließend geprüft werden, könnte aber in einer Validierungsstudie geprüft werden.

- Angesichts der genannten Probleme sind generelle Zweifel an der Validität des CRTT angebracht. Allerdings mangelt es an besseren Alternativen, was sich unter anderem auch darin manifestiert, dass in der Meta-Analyse von Anderson et al. (2010) die meisten Studien zum Gewaltverhalten auf dem CRTT basierten. Sollte der CRTT daher in weiteren Studien Verwendung finden, ohne dass vorher Validierungsstudien durchgeführt werden, dann scheint es in jedem Fall ratsam, die Probanden vor der Aufklärung nach etwaigen Täuschungsmanövern zu fragen und sich dabei auch nach dem Zeitpunkt zu erkundigen, an dem sie die jeweilige Täuschung durchschaut haben wollen. Dadurch ließe sich zumindest abschätzen, inwiefern das Durchschauen der Täuschungen sich auf die Messung des Gewaltverhaltens ausgewirkt haben kann. Alternativ könnte das Hot Sauce Paradigm weiterentwickelt werden. Die Ergebnisse der Validierungsstudien von Beier und Kutzner (2011a; 2011b) sind zumindest vielversprechend.
- Auch beim Verfahren zur Messung des Hilfeverhaltens wurde bisher keine Validierung durchgeführt. Hier bleibt außerdem ungeklärt, ob das Verfahren durch die Einkleidung als Konzentrationsaufgabe möglicherweise nicht nur von der Hilfsbereitschaft, sondern auch von der gefühlten Konzentrationsfähigkeit der Probanden abhängt. Auch dieses Verfahren wurden nur eingesetzt, weil es an besseren Alternativen mangelt.
- Daneben erhebt das Verfahren Hilfsbereitschaft statt Hilfeverhalten. Es ließe sich zwar leicht modifizieren, um echtes Verhalten zu erheben, allerdings wäre es dann unökonomisch. Daher wäre eine Pilotstudie sinnvoll, die ein so modifiziertes Verfahren mit dem hier verwendeten vergleicht, um zu überprüfen, ob sich tatsächlich größere Unterschiede ergeben. Im Rahmen dieser Pilotstudie sollte auch geprüft werden, ob durch die verwendete Cover Story die subjektive Wachheit der Probanden eine Rolle spielt. Dabei könnte es interessant sein, unterschiedliche Cover Stories in ihrer Auswirkung miteinander zu vergleichen.
- Weitere Defizite finden sich beim Treatment. Hier unterschied sich die Bewertung der Hilfefähigkeit nicht sehr stark zwischen den hilflosen und den hilfefähigen Bedingungen. Falls das Treatment in weiteren Studien zur Untersuchung der Wirkung von Hilfe im Spiel verwendet werden soll, wäre daher eine Optimierung notwendig. Der Hilfegehalt könnte in der Art operationalisiert werden, dass er nicht nur am Anfang und am Ende der Spielaufgabe ins Bewusstsein gerückt wird, sondern auch regelmäßig während der Aufgabe, z. B. indem sich die virtuelle Auftraggeberin für das Bewältigen von Teilschritten der Spielaufgabe regelmäßig bedankt. Außerdem sollte die Ursache dafür gefunden werden, warum in den eigentlich hilfefreien Bedingungen von den Probanden eine geringe Hilfefähigkeit wahrgenommen wird.

Virtuelle Nothilfe

- Ein weiteres Problem des Treatments liegt darin, dass sich die gewaltfreien und gewalthaltigen Bedingungen hinsichtlich Interessantheit und Spannung unterscheiden. Hier könnte im Fall einer Folgeuntersuchung ein Vortest helfen, dieses Problem zu vermeiden.
- Ein drittes Problem des Treatments liegt darin, dass Anfänger mit der Steuerung der Spielfigur bzw. mit dem Erlernen der Steuerung schnell überfordert sind. Es ist schwer zu sagen, wie die Steuerung vereinfacht bzw. besser erklärt werden kann, ohne dabei erfahrene Spieler mit zu geringer Komplexität bzw. unnötigen Instruktionen zu langweilen.
- Ein viertes aber kleineres Problem des Treatments liegt in der Qualität der grafischen Darstellung. Diese ist zwar akzeptabel, könnte aber gesteigert werden, um Präsenzgefühl und Immersion zu maximieren. Hierfür ließe sich ein Nachfolgespiel derselben Softwareschmiede verwenden, z. B. der Ego-Shooter „Fallout 3“ (Bethesda Softworks, 2008) oder das Ende 2011 erschienene Rollenspiel „Elder Scrolls V: Skyrim“ (Bethesda Softworks, 2011). Das Problem bei der Wahl eines möglichst aktuellen Spiels ist es aber, dass die Hardwareanforderungen exponentiell wachsen. Spiele werden inzwischen so entwickelt, dass sie erst auf der Hardware der Zukunft ihr volles Potenzial entfalten, was den maximalen Verkaufszeitraum der Spiele erhöhen soll. Um solche Spiele ab Erscheinungsdatum nutzen zu können, werden selbst unter Inkaufnahme diverser Einschränkungen für jeden PC Investitionen im vierstelligen Bereich notwendig. Die für eine Modifikation benötigte Entwicklungszeit kann aber dazu führen, dass das gewählte Spiel bei Fertigstellung der Modifikation trotzdem alt aussieht und damit bei den Probanden an Attraktivität verliert.

18 LITERATUR

- Alfermann, D. (1996). *Geschlechterrollen und geschlechtstypisches Verhalten*. Stuttgart u. A.: Kohlhammer.
- Amelang, M. & Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung* (4., überarbeitete und erweiterte Auflage). Stuttgart: Kohlhammer.
- Anderson, C. A. (2004). An update on the effects of playing violent video games. *Journal of adolescence*, 27, 113-122.
- Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M. & Flanagan, M. (2000). Tempera-
ture and aggression. *Advances in Experimental Social Psychology*, 32, 63-133.
- Anderson, C. A. & Bushman, B. J. (1997). External validity of "trivial" experiments: the case of
laboratory aggression. *Review of General Psychology*, 1, 19-41.
- Anderson, C. A. & Bushman, B. J. (2001). Effects of violent video games on aggressive behav-
ior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior:
A meta-analytic review of the scientific literature. *Psychological Science*, 12, 353-359.
- Anderson, C. A. & Carnagey, N. L. (2009). Causal effects of violent sports video games on ag-
gression: Is it competitiveness or violent content? *Journal of Experimental Social Psychol-
ogy*, 45, 731-739.
- Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin, A. J., Eubanks, J. & Valentine, J. C.
(2004). Violent video games: Specific effects of violent content on aggressive thoughts
and behavior. *Advances in Experimental Social Psychology*, 36, 199-249.
- Anderson, C. A., Deuser, W. E. & DeNeve, K. (1995). Hot temperatures, hostile affect, hostile
cognition, and arousal: Tests of a general model of affective aggression. *Personality and
Social Psychology Bulletin*, 21, 434-448.
- Anderson, C. A. & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and be-
havior in the laboratory and in life. *Journal of Personality and Social Psychology*, 78, 772-
790.
- Anderson, C. A. & Ford, C. M. (1986). Affect of the game player: Short-term effects of highly
and mildly aggressive video games. *Personality and Social Psychology Bulletin*, 12, 390-
402.
- Anderson, C. A., Gentile, D. A. & Buckley, K. E. (2007). *Violent video game effects on children
and adolescents. Theory, research and public policy*. Oxford u. A.: Oxford University Press.
- Anderson, C. A., Lindsay, J. J. & Bushman, B. J. (1999). Research in the psychological labora-
tory: Truth or triviality? *Current Directions in Psychological Science*, 8, 3-9.

- Anderson, C. A. & Murphy, C. R. (2003). Violent video games and aggressive behavior in young women. *Aggressive Behavior*, 29, 423-429.
- Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R. & Saleem, M. (2010). Violent video game effects on aggression, empathy, and pro-social behavior in Eastern and Western countries: A meta-analytic review. *Psychological Bulletin*, 136, 151-173.
- Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31, 101-119.
- Baron, R. A. & Richardson, D. R. (1994). *Human aggression* (2. Auflage). New York: Plenum Press.
- Bartholow, B. D. & Anderson, C. A. (2002). Effects of violent video games on aggressive behavior: Potential sex differences. *Journal of Experimental Social Psychology*, 38, 283-290.
- Bartholow, B. D., Sestir, M. A. & Davis, E. B. (2005). Correlates and consequences of exposure to video game violence: Hostile personality, empathy, and aggressive behavior. *Personality and Social Psychology Bulletin*, 31, 1573-1586.
- Beier, S. & Kutzner, F. (2011a, September). *Choose a juice! Der Effekt von Wahloptionen auf Aggression in einem modifizierten Hot-Sauce Paradigma*. Poster präsentiert auf der dreizehnten Tagung der Fachgruppe Sozialpsychologie (FGSP), Hamburg.
- Beier, S. & Kutzner, F. (2011b, November). *Choose a juice! The effect of choice options and intention on aggression in a modified hot-sauce paradigm* [Präsentation]. Präsentiert auf dem sechszehnten Workshop Aggression, Marburg.
- Bernstein, S., Richardson, D. & Hammock, G. (1987). Convergent and discriminant validity of the Taylor and Buss measures of physical aggression. *Aggressive Behavior*, 13, 15-24.
- Bethesda Softworks (2006a). *The Elder Scrolls IV: Oblivion* [Computerspiel]. Entwickelt von Bethesda Softworks. New York: 2K Games.
- Bethesda Softworks (2006b). *The Elder Scrolls Construction Set Wiki* [On-line Wiki]. Verfügbar unter: <http://cs.elderscrolls.com/constwiki/index.php> [18.05.2011].
- Bethesda Softworks (2008). *Fallout 3* [Computerspiel]. Entwickelt von Bethesda Softworks. New York: Bethesda Zenimax.
- Bethesda Softworks (2011). *The Elder Scrolls V: Skyrim* [Computerspiel]. Entwickelt von Bethesda Softworks. New York: 2K Games.
- Bierhoff, H. (2010). *Psychologie prosozialen Verhaltens. Warum wir anderen helfen* (2., vollständig überarbeitete Auflage). Stuttgart: Kohlhammer.

- Bilsky, W. (1981). Prosozialität. Zwei Skalen zur Beschreibung eines provisorischen Konzepts. *Diagnostica*, 17, 66-77.
- Böhm, F. & Streicher, B. (2010, September). *Fette Aggressionsmessung: eine geschlechtssensitive Erweiterung des Hot-Sauce-Paradigmas*. Poster präsentiert auf dem zweijährigen Treffen der Deutschen Gesellschaft für Psychologie (DGPS), Bremen.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5., vollständig überarbeitete und aktualisierte Auflage). Berlin u. A.: Springer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation*. Für Sozialwissenschaftler (2. Auflage). Berlin u. A.: Springer.
- Bortz, J., Lienert, G. A. & Boehnke, K. (2000). *Verteilungsfreie Methoden in der Biostatistik*. Berlin u. A.: Springer.
- Bottenberg, E. H., Bilsky, W., Schindler, U. & Wehrmeyer, H. (1979a). *Zur persönlichkeitspsychologischen Operationalisierung von Adjution: Konstruktion von Fragebogen-Skalen* (Braunschweiger Psychologische Arbeiten, 1979/2). Braunschweig: Technische Universität, Lehrstuhl für Psychologie.
- Bottenberg, E. H., Bilsky, W., Schindler, U. & Wehrmeyer, H. (1979b). *Zur empirischen Charakterisierung von Adjutions-Affirmierung und Adjutions-Disvalidierung: Wert-Orientierung, erzogenen-rekonstruierte elterliche Erziehung, Fremdbeurteilung* (Braunschweiger Psychologische Arbeiten, 1979/3). Braunschweig: Technische Universität, Lehrstuhl für Psychologie.
- Bottenberg, E. H., Langer, F. & Hansen, D. (1979). Zur persönlichkeitspsychologischen Exploration von Adjution: Hypothetisches Schema und Operationalisierung. *Psychologie und Praxis*, 23, 71-85.
- Buckley, K. E. & Anderson, C. A. (2006). A theoretical model of the effects and consequences of playing video games. In P. Vorderer & J. Bryant (Hrsg.), *Playing Video Games* (S. 363-378). Mahwah, NJ: Lawrence Erlbaum.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München u. A.: Pearson Studium.
- Bushman, B. J. & Anderson, C. A. (1998). Methodology in the study of aggression: Integrating experimental and nonexperimental findings. In R. Geen & E. Donnerstein (Hrsg.), *Human aggression: Theories, research and implications for policy* (S. 23-48). San Diego, CA: Academic Press.
- Bushman, B. J. & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, 75, 219-229.

- Bushman, B. J. & Huesmann, L. R. (2006). Short-term and long-term effects of violent media on aggression in children and adults. *Archives of Pediatrics Adolescent Medicine*, 160, 348-352.
- Bushman, B. J. & Saults, J. S. (2006). *The Competitive Reaction Time (CRT) measure of laboratory aggression* [Anleitung]. Verfügbar über [HTTP://uk.groups.yahoo.com/group/CRTRP](http://uk.groups.yahoo.com/group/CRTRP)
- Buss, A. H. & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63, 452-459.
- Carlson, M., Marcus-Newhall, A. & Miller, N. (1989). Evidence for a general construct of aggression. *Personality and Social Psychology Bulletin*, 15, 377-389.
- Carlson, M., Marcus-Newhall, A. & Miller, N. (1990). Effects of situational aggression cues: A quantitative review. *Journal of Personality and Social Psychology*, 58, 622-633.
- Carnagey, N. L. & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition and behavior. *Psychological Science*, 16, 882-889.
- Cavia (2010). *Nier Gestalt* [Computerspiel]. Entwickelt von Cavia. Japan, USA, EU: Square Enix.
- Chambers, J. H. & Ascione, F. R. (1987). The effects of prosocial and aggressive videogames on children's donating and helping. *Journal of Genetic Psychology*, 148, 499-505.
- Cherek, D. R. (1981). Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology*, 75, 339-345.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. Auflage). Hillsdale u. A.: Lawrence Erlbaum.
- Collani, G. von & Werner, R. (2005). Self-related and motivational constructs as determinants of aggression: An analysis and validation of a German version of the Buss-Perry Aggression Questionnaire. *Personality and Individual Differences*, 38, 1631-1643.
- dev/akm (2009). *dev/akm mods. Mods, Guides, Reviews for Elder Scrolls IV: Oblivion* [On-line Ressource]. Verfügbar unter: <http://devnull.sweetdanger.net> [18.05.2011].
- Diehl, J. M. & Arbinger, R. (1992). *Einführung in die Inferenzstatistik* (2. Auflage). Eschborn: Verlag Dietmar Klotz.
- DMA Design (1991). *Lemmings* [Computerspiel]. Entwickelt von DMA Design. Liverpool: Psygnosis.
- Eagly, A. H. & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, 100, 283-308.

- Egloff, B., Schmukle, S. C., Burns, L. R. & Schwerdtfeger, A. (2006). Spontaneous emotion regulation during evaluated speaking tasks: Association with negative affect, anxiety expression, memory and physiological responding. *Emotion*, 6, 356-366.
- Elson, M. (2011). *The effects of displayed violence and game speed in first-person shooters on physiological arousal and aggressive behavior*. Unveröffentlichte Diplomarbeit, Universität zu Köln.
- Erb, E. (1997). Gegenstands- und Problemkonstituierung: Subjekt-Modelle (in) der Psychologie. *Zur Programmatik einer sozialwissenschaftlichen Psychologie. Band I, 1. Halbband* (S. 139-239). Münster: Aschendorff.
- Evers, C., Fischer, A. H., Rodriguez Mosquera, P. M. & Manstead, A. S. R. (2005). Anger and social appraisal: A "spicy" sex difference? *Emotion*, 5, 258-266.
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Feltes, T. & Goldberg, B. (2006). Selbstberichtete Delinquenz, Viktimisierung und Verbrechensfurcht bei Schülern mit und ohne Migrationshintergrund - Ergebnisse einer Befragung von 4.000 Bochumer Schülerinnen und Schülern. In J. Obergfell-Fuchs & M. Brandenstein (Hrsg.), *Nationale und internationale Entwicklungen in der Kriminologie. Festschrift für Helmut Kury zum 65. Geburtstag* (S. 203-237). Frankfurt: Verlag für Polizeiwissenschaft.
- Ferguson, C. J. (2007a). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior*, 12, 470-482.
- Ferguson, C. J. (2007b). The Good, the Bad and the Ugly: A meta-analytic review of positive and negative effects of violent video games. *Psychiatric Quarterly*, 78, 309-316.
- Ferguson, C. J. (2008). The school shooting/violent video game link: Causal relationship or moral panic? *Journal of Investigative Psychology and Offender Profiling*, 5, 25-37.
- Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S. & Smith, S. M. (2008a). Violent video games and aggression: Causal relationship or byproduct of family violence and intrinsic violence motivation? *Criminal Justice and Behavior*, 35, 311-332.
- Ferguson, C. J., Smith, S., Miller-Stratton, H., Fritz, S. & Heinrich, E. (2008b). Aggression in the laboratory: Problems with the validity of the modified Taylor Competitive Reaction Time Test as a measure of aggression in media violence studies. *Journal of Aggression, Maltreatment & Trauma*, 17, 118-132.
- Field, A. (2009). *Discovering Statistics Using SPSS* (3. Auflage). London u. A.: Sage.

- Fischer, P., Kubitzki, J., Guter, S. & Frey, D. (2007). Virtual driving and risk taking: Do racing games increase risk-taking cognitions, affect, and behaviors? *Journal of Experimental Psychology: Applied*, 13, 22-31.
- Freelon, D. G. (2010). *ReCal: Intercoder reliability calculation as a web service*. *International Journal of Internet Science*, 5, 20-33.
- Gentile, D. A., Anderson, C. A., Yukawa, S., Ihori, N., Saleem, M., Ming, L. K., Shibuya, A., Liau, A. K., Khoo, A., Bushman, B. J., Huesmann, L. R. & Sakamoto, A. (2009). The effects of pro-social video games on prosocial behaviors: International evidence from correlational, longitudinal, and experimental studies. *Personality and Social Psychology Bulletin*, 35, 752-763.
- Gentile, D. A., Walsh, D. A., Ellison, P. R., Fox, M. & Cameron, J. (2004, Mai). *Media violence as a risk factor for children: A longitudinal study*. Paper presented at the American Psychological Society 16th Annual Convention, Chicago, Illinois.
- Giancola, P. R. & Chermack, S. T. (1998). Construct validity of laboratory aggression paradigms: A response to Tedeschi and Quigley (1996). *Aggression and Violent Behavior*, 3, 237-253.
- Giancola, P. R. & Zeichner, A. (1995a). An investigation of gender differences in alcohol-related aggression. *Journal of Studies on Alcohol*, 56, 573-579.
- Giancola, P. R. & Zeichner, A. (1995b). Construct validity of a competitive reaction-time aggression paradigm. *Aggressive Behavior*, 21, 199-204.
- Greitemeyer, T. & Osswald, S. (2009). Prosocial video games reduce aggressive cognitions. *Journal of experimental social psychology*, 45, 896-900.
- Greitemeyer, T. & Osswald, S. (2010). Effects of prosocial video games on prosocial behavior. *Journal of Personality and Social Psychology*, 98, 211-221.
- Greitemeyer, T., Osswald, S. & Brauer, M. (2010). Playing prosocial video games increases empathy and decreases schadenfreude. *Emotion*, 10, 796-802.
- Groeben, N. (1986). *Handeln, Tun, Verhalten als Einheiten einer verstehend-erklärenden Psychologie*. Tübingen: Francke.
- Groeben, N. (1999). Fazit: Die metatheoretischen Merkmale einer sozialwissenschaftlichen Psychologie. In N. Groeben (Hrsg.), *Zur Programmatik einer sozialwissenschaftlichen Psychologie*. Band I, 2. Halbband (S. 311-404). Münster: Aschendorff.
- Groeben, N., Wahl, D., Schlee, J. & Scheele, B. (1988). *Forschungsprogramm Subjektive Theorien: Eine Einführung in die Psychologie des reflexiven Subjekts*. Tübingen: Francke.
- Grouven, U., Bender, R., Ziegler, A. & Lange, S. (2007). Der Kappa-Koeffizient. *Deutsche Medizinische Wochenschrift*, 132, 65-68.

Virtuelle Nothilfe

- Happ, C., Melzer, A. & Steffgen, G. (2010a, September). *Mitleid mit dem Gegner: Fällt das Spielen schwer? Wirkung einer Empathie-Induktion auf Erleben und Verhalten in einem gewalthaltigen Videospiel*. Präsentiert auf dem DGPS-Kongress in Bremen vom 26.-30. September.
- Happ, C., Melzer, A. & Steffgen, G. (2010b). *Media, violence and prevention: A literature survey*. Luxembourg: Université du Luxembourg, INSIDE.
- Hartig, J., Frey, A. & Ketzel, A. (2003). Modifikation des Computerspiels Quake III Arena zur Durchführung psychologischer Experimente in einer virtuellen 3D-Umgebung. *Zeitschrift für Medienpsychologie*, 15, 149-154.
- Hartmann, T. (im Druck). Moral disengagement during exposure to media violence. In R. Tamborini (Hrsg.), *Media and Morality. Proceedings of the 2011 BEA Research Symposium*. New York: Taylor & Francis.
- Hartmann, T., Toz, E. & Brandon, M. (2010). Just a game? Unjustified virtual violence produces guilt in empathetic players. *Media Psychology*, 13, 339-363.
- Hartmann, T. & Vorderer, P. (2010). It's okay to shoot a character: moral disengagement in violent video games. *Journal of Communication*, 60, 94-119.
- Heckhausen, H. (1989). *Motivation und Handeln* (2., völlig überarbeitete und ergänzte Auflage). Berlin u. A.: Springer.
- Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107-128.
- Heidbrink, H. (1996). *Einführung in die Moralpsychologie* (2. Auflage). Weinheim: PVU.
- Herzberg, P. Y. (2003). Faktorstruktur, Gütekriterien und Konstruktvalidität der deutschen Übersetzung des Aggressionsfragebogens von Buss und Perry. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24, 311-323.
- Hopf, W. H., Huber, G. L. & Weiß, R. H. (2008). Media violence and youth violence. A 2-year longitudinal study. *Zeitschrift für Medienpsychologie*, 20, 79-96.
- Hussy, W. & Jain, A. (2002). *Experimentelle Hypothesenprüfung in der Psychologie*. Göttingen u. A.: Hogrefe.
- Hussy, W. & Möller, H. (1994). Hypothesen. *Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden. Serie I Forschungsmethoden der Psychologie, Band 1 Methodologische Grundlagen der Psychologie* (S. 475-507). Göttingen u. A.: Hogrefe.
- Kapp, F. & Scheele, B. (1996). Was verstehen Sie unter Zivilcourage? Annäherung an eine Psychologie des ‚aufrechten Gangs‘ mit Hilfe Subjektiver Theorien. *Gruppendynamik*, 27, 125-143.

- Klimmt, C. (2001). Computer-Spiel: Interaktive Unterhaltungsangebote als Synthese aus Medium und Spielzeug. *Zeitschrift für Medienpsychologie*, 13, 22-32.
- Klimmt, C. (2006a). *Computerspielen als Handlung. Dimensionen und Determinanten des Erlebens interaktiver Unterhaltungsangebote*. Köln: Herbert von Halem Verlag.
- Klimmt, C. (2006b). „Ein echtes Highlight ist der Wald“: Professionalisierung der Computerspiel-Vermarktung am Beispiel des Verkaufsstarts von „Oblivion“. *Zeitschrift für Medienpsychologie*, 18, 139-140.
- Klimmt, C., Schmid, H., Nosper, A., Hartmann, T. & Vorderer, P. (2006). How players manage moral concerns to make video game violence enjoyable. *Communications*, 31, 309-328.
- Klimmt, C., Schmid, H., Nosper, A., Hartmann, T. & Vorderer, P. (2008). 'Moral management': Dealing with moral concerns to maintain enjoyment of violent video games. In A. Sudmann-Jahn & R. Stockmann (Hrsg.), *Computer games as a sociocultural phenomenon: Games without frontiers - wars without tears* (S. 108-118). Hampshire, UK: Palgrave.
- Klimmt, C. & Trepte, S. (2003). Theoretisch-methodische Desiderata der medienpsychologischen Forschung über die aggressionsfördernde Wirkung gewalthaltiger Computer- und Videospiele. *Zeitschrift für Medienpsychologie*, 15, 114-121.
- Kohlberg, L. (2008). The development of children's orientations toward a moral order I. Sequence in the development of moral thought. *Human Development*, 51, 8-20.
- Konijn, E. A., Bijvank, M. N. & Bushman, B. J. (2007). I wish I were a warrior: The role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Developmental Psychology*, 43, 1038-1044.
- Kristen, A. (2005). *Aggressive Jungen und gewalthaltige Computerspiele*. Unveröffentlichte Dissertation, Fachbereich Erziehungswissenschaften und Psychologie der Freien Universität Berlin.
- Krohne, H. W., Egloff, B., Kohlmann, C.-W. & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der „Positive and Negative Affect Schedule“ (PANAS). *Diagnostica*, 42, 139-156.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Lieberman, J. D., Solomon, S., Greenberg, J. & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, 25, 331-348.
- Lindsay, J. J. & Anderson, C. A. (2000). From antecedent conditions to violent actions: A general affective aggression model. *Personality and Social Psychology Bulletin*, 26, 533-547.

- Lee, K. M. & Peng, W. (2006). What do we know about social and psychological effects of computer games? A comprehensive review of the current literature. In P. Vorderer & J. Bryant (Hrsg.), *Playing Video Games. Motives, Responses, and Consequences* (S. 325-346). Mahwah, NJ: Lawrence Erlbaum.
- McGregor, H. A., Lieberman, J. D., Greenberg, J., Solomon, S., Arndt, J., Simon, L. & Pyszczynski, T. (1998). Terror management and aggression: Evidence that mortality salience motivates aggression against worldview-threatening others. *Journal of personality and social psychology*, 74, 590-605.
- Mees, U. (1991). *Die Struktur der Emotionen*. Göttingen u. A.: Hogrefe.
- Mees, U. & Schmitt, A. (2003). Die Typikalität von Emotionsbegriffen: ihr Zusammenhang mit Erlebensdimensionen von Emotionen im semantischen Gedächtnis. In U. Mees & A. Schmitt (Hrsg.), *Emotionspsychologie: Theoretische Analysen und empirische Untersuchungen* (S. 221-262). Oldenburg: BIS.
- Merten, K. (1995). *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis* (2., verbesserte Auflage). Opladen: Westdeutscher Verlag.
- Mohseni, M. R. (2007). *Zivilcourage bei Jugendlichen* (überarbeitete Fassung von „Auswirkung von moralthematischem Kontext, sozialem Geschlecht und Angstbewältigungsstrategien auf die Bereitschaft zu zivilcouragiertem Handeln“ [On-line]. Verfügbar unter <http://psydok.sulb.uni-saarland.de/volltexte/2005/508> [20.07.2011]). Saarbrücken: VDM.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32, 396-402.
- Pawelczyk, A. (2011). *Gewaltanwendung und Hilfeleistung in Videospielen und ihre Wirkung auf aggressives Verhalten*. Unveröffentlichte Diplomarbeit, Universität Osnabrück.
- Penner, L. A., Craiger, J. P., Fritsche, B. A. & Freifeld, T. A. (1995). Measuring the prosocial personality. In J. N. Butcher & C. D. Spielberger (Hrsg.), *Advances in personality assessment* (Band 12, S. 147-163). Hillsdale, NJ: Erlbaum.
- Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York u. A.: Penguin Books.
- Pouledouris, B. (1982). *Conan the Barbarian* [CD]. Paris: Milan.
- Preacher, K. J. & Hayes, A. F. (2008a). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40, 879-891.
- Preacher, K. J. & Hayes, A. F. (2008b). INDIRECT [SPSS Makro]. Verfügbar unter: <http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html> [27.05.2011].

Virtuelle Nothilfe

- Ritter, D. & Eslea, M. (2005). Hot sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, 31, 407-419.
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Hrsg.), *The Handbook of Research Synthesis* (S. 231-244). New York: Russell Sage Foundation.
- Rustemeyer, R. (1992). *Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten*. Münster: Aschendorff.
- Savage, J. & Yancey, C. (2008). The effects of media violence exposure on criminal aggression: A meta-analysis. *Criminal Justice and Behavior*, 35, 772-791.
- Scheele, B. (1999). Theoriehistorische Kontinuität: Lernen von Aggression oder Möglichkeiten zur Katharsis. In N. Groeben (Hrsg.), *Zur Programmatik einer sozialwissenschaftlichen Psychologie*. Band I, 2. Halbband (S. 1-83). Münster: Aschendorff.
- Scheele, B. (2004). 'Weibliches' Mit-Leiden und 'männliches' Aktiv-Werden. Das Entwicklungsziel 'Zivilcourage'. In G. Meyer, U. Dovermann, S. Frech & G. Gugel (Hrsg.), *Zivilcourage lernen. Analysen - Modelle - Arbeitshilfen* (S. 86-94). Tübingen: Institut für Friedenspädagogik Tübingen.
- Scheele, B. & DuBois, F. (2006). Catharsis as a moral form of entertainment. In J. Bryant & P. Vorderer (Hrsg.), *Psychology of Entertainment* (S. 405-422). Mahwah, NJ: Lawrence Erlbaum.
- Scheele, B. & Kapp, F. (2002). Utopie Zivilcourage: zur Integration von Fürsorgemoral und öffentlichem Handeln. In Psychologisches Institut (Hrsg.), *Kölner Psychologische Studien*, VII, 1-59. Köln: Psychologisches Institut.
- Schmidt-Atzert, L. & Hüppe, M. (1996). Emotionsskalen EMO 16. Ein Fragebogen zur Selbstbeschreibung des aktuellen emotionalen Gefühlszustandes. *Diagnostica*, 42, 242-267.
- Schwarz, N. & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22, 127-160.
- Sestir, M. A. & Bartholow, B. D. (2010). Violent and nonviolent video games produce opposing effects on aggressive and prosocial outcomes. *Journal of Abnormal Psychology*, 119, 934-942.
- Severy, L. J. (1975). Individual differences in helping dispositions. *Journal of personality assessment*, 39, 282-292.
- Sherry, J. L. (2001). The effects of violent video games on aggression. *Human Communication Research*, 27, 409-431.
- Smith, B. P. (2006). The (computer) games people play. In P. Vorderer & J. Bryant (Hrsg.), *Playing Video Games. Motives, Responses, and Consequences* (S. 43-56). Mahwah, NJ: Lawrence Erlbaum.

- Staufenbiel, T. (2000). *Organizational citizenship behavior*. Unveröffentlichte Habilitationsschrift, Philipps-Universität Marburg.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5. Auflage). Boston u. A.: Pearson International Edition.
- Tamborini, R. & Skalski, P. (2006). The role of presence in the experience of electronic games. In P. Vorderer & J. Bryant (Hrsg.), *Playing Video Games. Motives, Responses, and Consequences* (S. 225-240). Mahwah, NJ: Lawrence Erlbaum.
- Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, 35, 297-310.
- Tedeschi, J. T. & Felson, R. B. (1994). *Violence, aggression, & coercive Actions*. Washington, DC: American Psychological Association.
- Tedeschi, J. T. & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior*, 1, 163-177.
- Tedeschi, J. T. & Quigley, B. M. (2000). A further comment on the construct validity of laboratory aggression paradigms: A response to Giancola and Chermack. *Aggression and Violent Behavior*, 5, 127-136.
- Vorderer, P. & Bryant, J. (Hrsg.). (2006). *Playing video games. Motives, responses, and consequences*. Mahwah, NJ: Lawrence Erlbaum.
- Vorderer, P., Bryant, J., Pieper, K. M. & Weber, R. (2006). Playing video games as entertainment. In P. Vorderer & J. Bryant (Hrsg.), *Playing Video Games. Motives, Responses, and Consequences* (S. 1-8). Mahwah, NJ: Lawrence Erlbaum.
- Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.
- Zehnder, S. M. & Lipscomb, S. D. (2006). The role of music in video games. In P. Vorderer & J. Bryant (Hrsg.), *Playing Video Games. Motives, Responses, and Consequences* (S. 241-258). Mahwah, NJ: Lawrence Erlbaum.
- Zillmann, D. (2000). The coming of media entertainment. *Media Entertainment* (S. 1-20). Mahwah, NJ: Lawrence Erlbaum.

19 ANHANG

A. VERWENDETE HARDWARE

Tabelle A: Verwendete Hardware

Komponente	Hersteller	Typ	Eigenschaften
Grafikkarte	XFX	Nvidia GeForce 8800 GTX	Schnelle Grafikkarte für flüssige Bildwiedergaberraten. Treiberstand: 197.13.
Prozessor	Intel	Core 2 Duo E8400 (E0)	Schneller Doppelkernprozessor für flüssige Bildwiedergaberraten
Prozessorkühler	Scythe	Mugen 2	Leise Kühlung
Netzteil	Enermax	Modu 525W	Stabile und leise Stromversorgung
Gehäuse	Silverstone	SST-KL02B	Offene Bauweise für eine gute Belüftung
Gehäuselüfter	Enermax	Magma 120	Leise Kühlung
Speicher	Corsair	2 x 2 GB (TWIN2X4096-6400C5)	Schneller RAM-Speicher
Festplatte	Samsung	Spinpoint F1 320 GB	Schnelle und gleichzeitig leise Festplatte für kurze Ladezeiten und damit ein ruckelfreies Spielerlebnis.
Entkoppler	Scythe	HDS 2	Festplattenentkopplung reduziert die Übertragung von Vibrationen und mindert so Störgeräusche.
Maus	Razer	Diamondback 3G	Hochauflösende Maus ermöglicht blitzschnelle Reaktionen. Die Empfindlichkeit lässt sich auch während des Spielens verstellen.
Tastatur	Microsoft	Microsoft Sidewinder X4	Tastatur mit beleuchteten Tasten.
Monitor	LG	227 WT-PF	22"-Breitbild-Monitor mit kurzen Reaktionszeiten für eine störeffektfreie Wiedergabe schneller Bildfolgen. Durch den erweiterten Farbraum wirken die Farben besonders plastisch.
Kopfhörer	T.Bone	HD990D	Dynamischer Kopfhörer, Schalldruck 105 db, Frequenzgang 10 Hz bis 20 kHz, Dämpfung von Außengeräuschen ca. 22 dB

B. INSTALLIERTE MODIFIKATIONEN

Tabelle B: Installierte Modifikationen (Mods)

Option	Wert	Begründung
Offizieller Patch	1.2.0416	Aktuellste Fehlerbehebungen
Oblivion Improved (NewRaven und Andere)	1.41	Enthält weitere wichtige Fehlerbehebungen sowie Verbesserungen (z. B. übersichtlicheres Inventar). Installiert wurde es ohne Shivering Isles, mit OBSE, mit Nvidia-HDR-Fix, mit Archive Invalidated, mit Übersetzungen, mit Karten, mit HUD und mit neuen Ladebildschirmen (weil es mit den alten abstürzte).
Oblivion Mod Manager (Timeslip)	1.1.12	Ermöglicht eine saubere und konfliktfreie Installation der Modifikationen. Eingesetzt wird BSA Redirection (plus „Pack Face Textures into Redirection“), da es sonst nicht mit "Archive Invalidated" von Oblivion Improved harmoniert. Die Plugins wurden in folgende Reihe gebracht: (1) Oblivion, (2) Global Settings, (3) Nvidia-HDR-Fix, (4) (ESM-Dateien), (5) Obl-IM.esp, (6) Immersive Health.esp, (7) Promo.esp.
Qarl's Texture Pack (Qarl)	3R	Viele Texturen (Wände, Steine, Wege) etc. sind höher aufgelöst und dadurch schärfer. Die „R“-Variante ist marginal unschärfer, reduziert dafür den Speicherbedarf der Grafikkarte.
Improved Facial Textures (Shady Tradesman)	2.1	Gesichter sind höher aufgelöst.
New Face Textures (Enayla)	2	Die Auftraggeberin hat feinere Hautstrukturen, Grübchen, Nasolabial- und Augenwinkelfalten.
MEAT: Mangler's Equipment & Ammo Textures (Mangler)	0.1	Waffen/Rüstungen sind plastischer.
Kalikuts Green Glass Retex (kalikut)	1.0	Die grünen Glaswaffen vom Borg Bandit sind weniger surreal und sehen mehr wie echtes Glas aus.
Mikes Clothes Replacer (mike-gemini)	1.0	Die Kleidung der Auftraggeberin ist höher aufgelöst.
Quality Female Body Textures (outlive)	1.0	Das Dekolleté der Auftraggeberin ist höher aufgelöst.
Hit Shader Replacement (Timeslip)	1.0	Wird der Spieler getroffen, so erhält er durch Anzeige eines dunkelroten Bildschirmrandes ein visuelles Feedback.
Unique Features (Von Djangos)	1.5	Verpasst Gesichtern individuelle Merkmale. Wurde eingesetzt, um der Auftraggeberin eine Pockennarbe zu verpassen.
Change Console Key (scanti)	2.0	Deaktiviert die Konsolen-Taste (benötigt OBSE >=1.3)

Die Angaben zum Autor des Mods finden sich in Klammern hinter dem Namen der Option. Eine Übersicht über typische Mods findet sich bei dev/akm (2009).

C. OPTIMIERUNG DER DATEI „OBLIVION.INI“

Tabelle C: Optimierung der Datei „Oblivion.ini“

Option	Wert	Begründung
[General]		
bUseHardDriveCache	1	Beschleunigt Ladevorgänge durch Nutzung des Festplatten-Cache.
SMainMenuMovieIntro	leer	Entfernt die Startfilme und beschleunigt so den Spielstart.
SIntroSequence	leer	Entfernt die Intros und beschleunigt so den Spielstart.
SMainMenuMovie	leer	Entfernt die animierte Karte im Spielmenü und beschleunigt so den Spielstart.
[HAVOK]		
iNumHavokThreads	5	
[Display]		
iShadowMapResolution	2048	Je höher, desto mehr Pixel werden für die Schattendarstellung verwendet. Ist insbesondere bei Dialogen mit anderen Figuren zu bemerken.
iShadowFilter	10	Höhere Werte führen zu einer weicheren Schattendarstellung.
[Interface]		
bAllowConsole	0	Soll die Debugging-Konsole deaktivieren, was aber nicht immer klappt. In dem Fall kann man OBSE installieren und mit dem OBSE-Plugin "ChangeConsoleKey" den Konsolenaufruf auf eine Taste verlegen, die über die Spieler-Tastatur nicht erreichbar ist.
[ControlsPlus]	182	Die Option wird vom OBSE-Plugin "ChangeConsoleKey" hinzugefügt. Der Wert 182 legt den Konsolenaufruf auf die Taste "LaunchApp1Key", die oft fehlt oder einen Treiber erfordert, um zu funktionieren. http://www.kellyethridge.com/vbcorlib/doc/ConsoleKeyInfo.ConsoleKey.html
[BlurShaderHDRInterior]		
fBlurRadius	6.0000	Reduziert die HDR-Beleuchtung, weil diese in Innenräumen übertrieben wirkt.
fBrightScale	1.7500	Reduziert die HDR-Beleuchtung.
[OpenMP]		
iThreads	20	Erhöht die maximale Anzahl der Threads, was eine Performanceverbesserung bewirken kann.
iOpenMPLevel	10	Erhöht die maximale Anzahl der Threads.

D. VIDEOEINSTELLUNGEN

Tabelle D1: Videoeinstellungen im Grafikkartentreiber

Option	Wert	Begründung
Anisotrope Filterung	16x	Verbessert die Symmetrie der perspektivisch entfernten Objekte.
Antialiasing – Gamma-Korrektur	Globale Einstellung (Ein) verwenden	
Antialiasing – Modus	Beliebige Anwendungseinstellung außer Kraft setzen	Aktiviert die Kantenglättung, was Treppenbildung in Kanten reduziert.
Antialiasing – Transparenz	Aus	„Ein“ kostet zu viel Leistung.
Antialiasing-Einstellung	8x	Stellt die Kantenglättung auf einen mittelhohen Wert. Hohe Werte sind besser, kosten aber Leistung.
Dreifach-Puffer	Globale Einstellung (Aus) verwenden	
Erweiterungsbeschränkung	Globale Einstellung (Aus) verwenden	
Fehlermeldung	Globale Einstellung (Aus) verwenden	
Maximale Anzahl der vorgeordneten Einzelbilder	Globale Einstellung (3) verwenden	
Multi-Display/gemischte GPU-Beschleunigung	Einzel-Display-Leistungs-Modus	Es wird nur eine Grafikkarte eingesetzt.
Texturfilterung - Anisotrope Abtastoptimierung	Aus	Optimierung verschlechtert die Bildqualität.
Texturfilterung - Negativer LOD-Bias	Clamp	Weil anisotrope Filterung aktiviert ist.
Texturfilterung - Trilineare Optimierung	Aus	Optimierung verschlechtert die Bildqualität.
Texturfilterungsqualität	Hohe Qualität	
Threaded-Optimierung	Ein	Unterstützung für Prozessoren mit mehreren Kernen.
Vertikale Synchronisierung	Globale Einstellung (Einstellung für 3D-Anwendung verwenden) verwenden	
Übereinstimmende Texture-Clamp	Globale Einstellung (Hardware verwenden) verwenden	

Virtuelle Nothilfe

Tabelle D2: Videoeinstellungen im Startmenü

Option	Wert	Begründung
Auflösung	1680 x 1050	Maximale Bildschirmauflösung für scharfe Darstellung
Anti-Aliasing	Keine (Bestleistung)	Als das Spiel erschien, konnten Nvidia-Grafikkarten HDR und AA noch nicht kombinieren. Daher wird bei Nvidia-Karten AA nicht über das Spiel, sondern indirekt im Treiber aktiviert.
Modus	Maximieren	Randlose Darstellung
Modus	VSyn ein	Verhindert sich gegeneinander verschiebende Bildschirmhälften (engl. „tearing“)
Bildschirmeffekte	HDR ein	Sorgt für dynamischere Beleuchtung
Entferntes Rendering	Alle ein	Erhöht den Realismus, weil auch weiter entfernte Objekte dargestellt werden
Qualität	Ultrahohe	Stellt automatisch die meisten der oben angegebenen Werte ein

Tabelle D3: Videoeinstellungen im Spielmenü

Option	Wert	Begründung
Auflösung	1680 x 1050	Siehe Tabelle A1
Grafikqualität	Hoch	Maximale Grafikqualität
Detail Bäume	Max	Maximale Grafikqualität
Detail NPCs	Max	Maximale Grafikqualität
Detail Gegenstände	Max	Maximale Grafikqualität
Detail Objekte	Max	Maximale Grafikqualität
Gras	Max	Maximale Grafikqualität
Distanz sehen	Max	Maximale Grafikqualität
Detail Entfernung	An	Maximale Grafikqualität
Entfernte Gebäude	An	Maximale Grafikqualität
Entfernte Bäume	An	Maximale Grafikqualität
Innere Schatten	Max	Maximale Grafikqualität
Äußere Schatten	Max	Maximale Grafikqualität
Körperschatten	An	Maximale Grafikqualität
Grasschatten	An	Maximale Grafikqualität
Laubwerksschatten	An	Maximale Grafikqualität
Schattenqualität	Hoch	Maximale Grafikqualität
Spiegelndes Licht	Max	Maximale Grafikqualität
HDR-Beleuchtung	An	Siehe Tabelle A2
Bloom	Aus	Bloom ist ein minderwertiger Ersatz für HDR.
Wasserdetail	Hoch	Maximale Grafikqualität
Wasserspiegelungen	An	Maximale Grafikqualität
Wasserverdrängung	An	Maximale Grafikqualität
Fensterspiegelungen	An	Maximale Grafikqualität
Blutspritzer	Hoch	Diese Einstellung ist realistischer.
Anti-Aliasing	Aus	Siehe Tabelle A2

Virtuelle Nothilfe

Tabelle D4: Audioeinstellungen im Spielmenü

Option	Wert	Begründung
Gesamtlautstärke	20	Die Kopfhörer sind maximal laut, um für das CRTT eine ausreichende Lautstärke zu erreichen. Für Oblivion ist dies viel zu laut.
Sprachlautstärke	Max	Die Stimmufnahmen sind etwas zu leise. Durch die Absenkung der anderen Lautstärken (s. unten) werden auf die Stimmufnahmen lauter.
Effektlautstärke	75	
Schrittlautstärke	75	
Musiklautstärke	75	

Virtuelle Nothilfe

E. EIGENSCHAFTEN DER SPIELAUFGABE

Tabelle E1: Subjektive Bewertung und objektive Schwierigkeit der Spielaufgaben (dritter Vortest)

Eigenschaft der Spielaufgabe	Versuchsbedingung							
	Nothilfe (n=6)		Hilfe (n=5)		Töten (n=6)		Schatzjagd (n=6)	
Hilfeshaltig	ja		ja		nein		nein	
Gewalthaltig	ja		nein		ja		nein	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hilfe in Aufgabe ¹	3.33	0.82	4.50	0.58	2.50	1.52	2.17	1.60
Hilfe im Inhalt ¹	3.00	1.26	3.50	0.58	1.83	0.41	1.83	0.75
Gewalt in Aufgabe ¹	4.00	0.63	1.20	0.45	4.50	0.55	1.00	0.00
Gewalt im Inhalt ¹	3.17	1.17	2.80	0.84	2.33	1.03	2.67	1.03
Gewalt in Grafik ¹	2.17	0.98	2.00	1.41	2.00	0.63	2.00	0.89
Realitätsnähe Grafik ¹	2.50	0.55	3.00	1.00	2.67	1.21	3.17	0.75
Präsenzgefühl ¹	3.67	0.82	3.40	1.82	2.17	1.33	3.33	1.51
Interessantheit ¹	3.00	0.63	3.40	0.55	2.17	0.41	3.67	0.82
Spannung ¹	3.50	0.55	3.40	0.89	2.83	3.33	3.33	1.37
Unterhaltung ¹	3.67	1.03	3.00	1.00	3.33	0.52	3.17	0.98
Aktivierung (Munter) ¹	3.33	0.82	3.20	1.10	3.50	1.05	3.50	0.84
Aktivierung (Lustlos) ¹	1.50	0.84	2.00	0.71	2.17	1.47	1.67	1.03
Frustration ¹	2.33	1.21	3.60	1.14	3.00	0.89	3.33	1.21
Schwierigkeit ¹	2.50	0.84	3.60	1.14	2.83	1.47	3.67	1.21
Spielerohnmacht ¹	2.83	0.75	5.80	1.79	2.17	0.41	13.33	9.65

¹ 5-stufige Skala von 1 (völlig falsch) bis 5 (völlig richtig).

Tabelle E2: Subjektive Bewertung der Spielaufgaben (Hauptuntersuchung)

Eigenschaft der Spielaufgabe ¹	Gesamt	Nothilfe	Hilfe	Töten	Schatzjagd
Interessantheit	3.09 (1.06)	2.78 (1.00)	3.27 (1.12)	2.98 (0.99)	3.29 (1.07)
Spannung	3.20 (1.11)	2.96 (1.00)	3.42 (1.16)	2.96 (1.00)	3.44 (1.18)
Unterhaltung	3.30 (1.11)	3.20 (1.18)	3.29 (1.25)	3.36 (0.96)	3.33 (1.06)
Frustration	2.92 (1.12)	2.76 (1.09)	3.17 (1.04)	2.89 (1.09)	2.85 (1.24)
Schwierigkeit	2.89 (1.08)	2.76 (1.05)	3.15 (1.09)	2.87 (1.24)	2.79 (0.90)
Präsenzgefühl	2.98 (1.09)	2.96 (1.28)	3.07 (0.96)	2.98 (1.14)	2.92 (1.01)
Realitätsnähe der Grafik	2.76 (1.09)	2.56 (1.14)	2.67 (1.17)	2.84 (1.02)	2.98 (1.02)
Aktivierung (munter)	3.08 (1.18)	2.96 (1.15)	3.13 (1.20)	3.13 (1.15)	3.10 (1.26)
Aktivierung (lustlos)	1.90 (1.09)	1.96 (1.15)	2.12 (1.17)	1.76 (0.96)	1.75 (1.04)

¹ 5-stufige Skala mit 1 (völlig falsch), 2 (ziemlich falsch), 3 (unentschieden), 4 (ziemlich richtig) und 5 (völlig richtig).

Markieren Sie so: ☐ ☒ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.

Korrektur: ☐ ☒ ☐ ☒ Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

1.1 Probanden-Nr.

Diagram illustrating the bit fields for the `uint8_t` type:

Field Name	Bit Range
100er	x0 - x9
10er	x0 - x9
1er	x0 - x9

- ☐ RM
- ☐ Ja

- ☐ RK ☐ NN
☐ Nein

10er □ □ □ □ □ □ □ □ □ □
1er □ □ □ □ □ □ □ □ □ □
 x_0 x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9

- ### 1.5 Datum: Monat

$\begin{array}{cccccccccc} \text{10er} & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ \text{1er} & \square & \square & \square & \square & \square & \square & \square & \square & \square \\ & x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 \end{array}$

- ### 1.6 Datum: Jahr

10er ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1er ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

x_0 x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9

- ## 1.7 VK

- ☐ RM
- ☐ NN

- RK □ TS

2.1 Alter

10er □ □ □ □ □ □ □ □ □
1er □ □ □ □ □ □ □ □ □
 x_0 x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9

- 2.2 Geschlecht
- 2.3 Hochschule
- 2.4 Lehreinheit (s. Liste)

- ☐ männlich
☐ FH

- ☐ weiblich
☐ Uni

- ## 2.5 Studienfach

--

- ## 2.6 Fachsemester

10er □ □ □ □ □ □ □ □ □
1er □ □ □ □ □ □ □ □ □
x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

Bitte wenden



3. Vorerfahrung mit Computerspielen

Ich möchte Ihnen nun einige Fragen zu Ihrer bisherigen Erfahrung mit Oblivion oder ähnlichen Spielen stellen. Dabei ist es nicht wichtig, ob Ihre Erfahrungen aktuell oder bereits einige Zeit her sind.

(Mindestens 1x Mittelkategorie)

- | | ca. 2 bis 3 Mal pro Woche
ein paar Mal im Monat
(fast) nie | (fast) täglich | weiß nicht |
|---|--|--------------------------|--------------------------|
| 3.1 Ich spiele PC-Spiele bzw. habe PC-Spiele gespielt, bei denen die Steuerung so ähnlich wie in Oblivion funktioniert (z. B. Egoshooter wie Counter Strike oder Call of Duty, aber auch World of Warcraft, GTA, Second Life etc.). | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3.2 Ich spiele PC-Spiele bzw. habe PC-Spiele gespielt, die so ähnlich sind wie Oblivion (z. B. Gothic). | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3.3 Ich spiele Oblivion am PC bzw. habe früher Oblivion am PC gespielt. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3.4 Ich benutze Modifikationen (sogenannte „Mods“) für Oblivion bzw. habe Mods benutzt. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3.5 Ich erstelle Mods für Oblivion bzw. habe Mods erstellt. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

4. Mögliche Ausschlusskriterien

- | | | |
|--|-----------------------------|-------------------------------|
| 4.1 Haben Sie Tinnitus (ständiges Rauschen und Pfeifen im Ohr)? | <input type="checkbox"/> Ja | <input type="checkbox"/> Nein |
| 4.2 Leiden Sie unter Schwerhörigkeit? | <input type="checkbox"/> Ja | <input type="checkbox"/> Nein |
| 4.3 Hatten Sie schon einmal eine unangenehme Erfahrung mit sehr lauten Geräuschen („Knalltrauma“)? | <input type="checkbox"/> Ja | <input type="checkbox"/> Nein |

Ausschluss, wenn die Erkrankung das Hörvermögen beeinträchtigt (z. B. Rhinitis, Sinusitis) oder das Herz-Kreislauf-System belastet (z. B. Kreislaufprobleme, kardiovaskuläre Erkrankungen).

- | | | |
|--|-----------------------------|-------------------------------|
| 4.4 Liegt bei Ihnen eine akute Erkrankung vor? | <input type="checkbox"/> Ja | <input type="checkbox"/> Nein |
| 4.5 Wenn ja, welche? | | |

- | | | |
|---|-----------------------------|-------------------------------|
| 4.6 Liegt bei Ihnen eine chronische Erkrankung vor? | <input type="checkbox"/> Ja | <input type="checkbox"/> Nein |
| 4.7 Wenn ja, welche? | | |



G. NOTIZBOGEN DER SPIELERGESBISSE

EvaSys	Spielergebnisse	Electric Paper
Universität Osnabrück 03 Diplomarbeiten/Promotionen		Medienstudie Spielergebnisse

Markieren Sie so: ☐ ☒ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
 Korrektur: ☐ ☒ ☐ ☐ Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

1. Vom Versuchsleiter auszufüllen

1.1 Probanden-Nr.
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

1.2 Geschlecht ☐ männlich ☐ weiblich

2. Spielergebnisse

2.1 Zeit Tutorial (Minuten)
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.2 Zeit Tutorial (Sekunden)
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.3 Spiel erfolgreich beendet ☐ ja ☐ nein

2.4 Zeit Spielaufgabe (Minuten)
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.5 Zeit Spielaufgabe (Sekunden)
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.6 Spielzeit laut Savegame (Minuten)
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.7 Spielzeit laut Savegame (Sekunden)
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.8 Versuchsbedingung ☐ Alpha ☐ Beta ☐ Gamma
☐ Delta

2.9 Anzahl Bewusstlosigkeitspokale
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

2.10 Anzahl Siegespokale
 100er ☐☐☐☐☐☐☐☐☐☐
 10er ☐☐☐☐☐☐☐☐☐☐
 1er ☐☐☐☐☐☐☐☐☐☐
 x0 x1 x2 x3 x4 x5 x6 x7 x8 x9

F954U8889P1PL0V1
20.05.2011, Seite 1/1

H. FRAGEBOGEN DER HAUPTUNTERSUCHUNG

1

2

3

4

5

6

7

8

9

Vom Versuchsleiter auszufüllen

Probanden-Nr.

902

Geschlecht
☐ männlich ☒ weiblich

1

2

3

4

5

6

7

8

9

Aktueller Gefühlszustand vor dem Spiel

Wir möchten herausfinden, wie sich Computerspiele auf die Stimmung auswirken. Daher möchten wir Ihren Gefühlszustand jetzt vor dem Spiel und später noch ein Mal nach dem Spiel erfragen.

Um zu verhindern, dass Sie eine Frage übersehen, wird geprüft, ob Sie bei jeder Frage ein Kreuz gemacht haben. Wenn Sie eine Frage unbeantwortet lassen möchten, kreuzen Sie bitte "weiß nicht" an.

Wie fühlen Sie sich jetzt? Beschreiben Sie bitte Ihren aktuellen Gefühlszustand mit Hilfe der vorgegebenen Wörter. Jedes Wort steht für einen Bereich von Gefühlen. Es schließt also ähnliche Gefühle ein, für die man auch ein anderes Wort verwenden könnte. Kreuzen Sie nun an, wie intensiv Sie gerade jedes Gefühl erleben! Je stärker das Gefühl ist, desto weiter rechts machen Sie das Kreuz.

		nicht vorhanden	eher schwach	mittel	eher stark	sehr stark	weiß nicht
Abneigung	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ärger	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neid	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Langeweile	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Angst	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unruhe	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Traurigkeit	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sehnsucht	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Virtuelle Nothilfe

Scham	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Schuldgefühl	<input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Freude	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Stolz	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Mitgefühl	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Zuneigung	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Überraschung	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Leid	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Schadenfreude	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Mitfreude	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Hoffnung	<input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Sonstige Gefühle	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>

Wenn Sie unter **Sonstige Gefühle** etwas angekreuzt haben, geben Sie bitte hier an, welche Gefühle Sie meinen.

Aufregung

1

2

3

4

5

6

7

8

9

Wichtige Hinweise zu Oblivion

Sie werden gleich das Computerspiel „Oblivion“ spielen. Das Spiel beginnt mit einem Tutorial, in dem Sie erklärt bekommen, wie die Steuerung mit Maus und Tastatur funktioniert. Auf dem Tisch stehen außerdem zwei Mini-Poster, die zeigen, was welche Taste bewirkt. Möglicherweise können Sie Oblivion bereits steuern. Auch dann müssen Sie das Tutorial durchlaufen, weil wir für alle Teilnehmer des Experiments die Bedingungen gleich halten müssen.

Sobald Sie das Tutorial durchlaufen haben, erhalten Sie direkt im Spiel von einer Figur eine Aufgabe. Diese Aufgabe sollen Sie so gut wie möglich lösen. Dazu gehört, dass Sie es tunlichst vermeiden sollten, dass Ihre eigene Spielfigur ohnmächtig wird oder sogar stirbt.

Außerdem wird gemessen, wie lange Sie für das Tutorial und das Lösen der Aufgabe benötigt haben. Die erste Zeitmessung beginnt am Anfang des Spiels, die zweite direkt nach der Einführung. Da der Versuchsleiter zu diesem Zeitpunkt nicht bei Ihnen im Raum ist, müssen Sie ihr/ihm bitte einen Hinweis geben, sobald Sie mit der Einführung fertig sind. Der Zeitpunkt wird Ihnen im Spiel angezeigt. Der Versuchsleiter wird dann Ihre Zeit stoppen. Das Spiel teilt Ihnen auch mit, wenn Sie die Aufgabe gelöst haben. Geben Sie dann bitte wieder so schnell wie möglich dem Versuchsleiter einen Hinweis. Er/Sie wird erneut die Zeit stoppen.

Bitte setzen Sie erst am Ende des Spiels die Kopfhörer ab. Lassen Sie danach das Spiel geöffnet, damit Ihre Spielergebnisse notiert werden können!

Bitte wenden Sie sich nun an den Versuchsleiter, damit er/sie das Spiel starten kann. Sie können diesen Fragebogen während des Spiels geöffnet lassen.

Bitte füllen Sie die folgenden Fragebögen erst aus, wenn Sie mit dem Spielen von Oblivion fertig sind!

Ich bin mit dem Spielen fertig.

☒ Ja

Virtuelle Nothilfe

1 2 3 4 5 6 7 8 9

Reaktionszeit-Spiel

Bitte beantworten Sie im Folgenden einige Fragen zum Reaktionszeit-Spiel.

	gar nicht	mittelmäÙig	äuÙerordentlich	ziemlich	kaum	weiß nicht
Wie gut eignet sich das Reaktionszeit-Spiel Ihrer Ansicht nach zum Messen von Reaktionszeiten?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wie sehr haben Sie sich während des Reaktionszeit-Spiels geärgert?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wie sehr hat sich Ihr Gegenspieler Ihrer Ansicht nach geärgert?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Reagierte Ihr Gegenspieler Ihrer Ansicht nach schneller als Sie?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bitte beschreiben Sie in wenigen Worten, welchen Eindruck Sie von Ihrem Gegenspieler gewonnen haben.						
...						

Bitte geben Sie an, welche der folgenden Gründe am besten beschreiben, warum Sie sich für eine bestimmte Lautstärke-Einstellung entschieden haben.

	gar nicht	mittelmäÙig	äuÙerordentlich	ziemlich	kaum	weiß nicht
Ich wollte meinen Gegenspieler behindern, um zu gewinnen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wollte beeinflussen, welche Lautstärke mein Gegenspieler einstellt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wollte meinen Gegenspieler verrückt machen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wollte meinen Gegenspieler verletzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wollte mich an meinem Gegenspieler rächen für die Lautstärke, die er/sie eingestellt hat.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich wollte ihm/ihr lautere Töne verpassen als er/sie mir verpasst hat.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Virtuelle Nothilfe

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

Aktueller Gefühlszustand nach dem Spiel

Wir möchten Ihren Gefühlszustand jetzt nach dem Spiel ein weiteres Mal erfragen.

Wie fühlen Sie sich jetzt? Beschreiben Sie bitte Ihren aktuellen Gefühlszustand mit Hilfe der vorgegebenen Wörter. Jedes Wort steht für einen Bereich von Gefühlen. Es schließt also ähnliche Gefühle ein, für die man auch ein anderes Wort verwenden könnte. Kreuzen Sie nun an, wie intensiv Sie gerade jedes Gefühl erleben! Je stärker das Gefühl ist, desto weiter rechts machen Sie das Kreuz.

	nicht vorhanden	sehr schwach	eher schwach	mittel	eher stark	sehr stark	weiß nicht
Abneigung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ärger	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Neid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Langeweile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Angst	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unruhe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Traurigkeit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sehnsucht	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scham	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Schuldgefühl	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Freude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Stolz	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitgefühl	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Zuneigung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Überraschung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Leid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Schadenfreude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mitfreude	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hoffnung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sonstige Gefühle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wenn Sie unter **Sonstige Gefühle** etwas angekreuzt haben, geben Sie bitte hier an, welche Gefühle Sie meinen.

Virtuelle Nothilfe

1 2 3 4 5 6 7 8 9

Steuerung von Oblivion

Bitte beantworten Sie im Folgenden einige Fragen zu der Steuerung von Oblivion.

Ich konnte bereits vor dem Durchspielen des Tutorials die Figur in Oblivion steuern.

☐ Ja ☐ Nein

	völlig falsch	ziemlich falsch	unentschieden	ziemlich richtig	völlig richtig	weiß nicht
Ich fand das Tutorial nützlich zum Erlernen der Steuerung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich fühle mich jetzt sicher im Umgang mit der Steuerung von Oblivion.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	nie	gelegentlich	seltener	immer	oft	weiß nicht
Ich musste auf eines der Mini-Poster schauen, auf denen die Tastenbelegung abgebildet ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Virtuelle Nothilfe

1 2 3 4 5 6 7 8 9

Eigenschaften der Spielaufgabe

Bitte beantworten Sie im Folgenden einige Fragen zu der Aufgabe, die Sie vorhin in Oblivion gespielt haben.

Was war Ihre Aufgabe im Spiel?

weiß nicht
 ziemlich richtig
 ziemlich falsch
 völlig richtig
 völlig falsch
 unentschieden

Ich finde, dass ...

... die Spielaufgabe interessant war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe spannend war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe unterhaltsam war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe frustrierend war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe schwierig war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe mich munter gemacht hat.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe mich lustlos gemacht hat.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... ich in dem Spiel versunken bin.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Grafik im Spiel realistisch war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Grafik im Spiel voll von Gewaltdarstellungen war.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Inhalte des Spiels voll von Gewaltverhalten waren.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Inhalte des Spiels voll von Hilfeverhalten waren.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe Gewaltverhalten erforderte.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
... die Spielaufgabe Hilfeverhalten erforderte.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>

Möchten Sie sonst noch etwas über das Spiel sagen? Nutzen Sie die Gelegenheit für Kommentare und Anmerkungen!

Virtuelle Nothilfe

1
2
3
4
5
6
7
8
9

Erfahrung mit Videospielen

Bitte beantworten Sie im Folgenden einige Fragen zu Ihren Erfahrungen mit Videospielen. Bitte erinnern Sie sich an die drei Videospiele, die Sie seit der Zeit nach der Grundschule am meisten gespielt haben. Gemeint sind Spiele am PC, an der Konsole und in der Spielhalle.

Titel meines am **häufigsten** gespielten Videospieles (= **Spiel 1**):

Titel meines am **zweithäufigsten** gespielten Videospieles (= **Spiel 2**):

Titel meines am **drithäufigsten** gespielten Videospieles (= **Spiel 3**):

	ca. 2 bis 3 Mal pro Woche (fast) täglich	ca. 1 Mal pro Woche (fast) nie	ein paar Mal im Monat (fast) nie	weiß nicht
Wie oft haben Sie Spiel 1 im letzten Jahr gespielt?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<input type="radio"/>
Wie oft haben Sie Spiel 1 in den Jahren davor gespielt?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<input type="radio"/>
Wie oft haben Sie Spiel 2 im letzten Jahr gespielt?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<input type="radio"/>
Wie oft haben Sie Spiel 2 in den Jahren davor gespielt?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<input type="radio"/>
Wie oft haben Sie Spiel 3 im letzten Jahr gespielt?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<input type="radio"/>
Wie oft haben Sie Spiel 3 in den Jahren davor gespielt?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>			<input type="radio"/>

Geben Sie nun bitte an, wie prosozial bzw. gewalthaltig Sie die Spiele finden. Zum Vergleich können Sie sich folgende Spiele vorstellen: *Lemmings* als prosoziales Spiel, *Counter-Strike* als gewalthaltiges Spiel und *Tetris* als ein Spiel, das weder prosozial noch gewalthaltig ist.

	außerordentlich ziemlich mittelmäßig kaum gar nicht	weiß nicht
Wie prosozial finden Sie Spiel 1 ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Wie prosozial finden Sie Spiel 2 ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Wie prosozial finden Sie Spiel 3 ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Wie gewalthaltig finden Sie Spiel 1 ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Wie gewalthaltig finden Sie Spiel 2 ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>
Wie gewalthaltig finden Sie Spiel 3 ?	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/>

Virtuelle Nothilfe

1 2 3 4 5 6 7 8 9

Abschließende Fragen

Ich möchte Sie nun abschließend bitten, einige Fragen zum Versuch an sich zu beantworten.

Haben Sie eine Vermutung, worum es in der Untersuchung geht?

☐ Ja ☐ Nein

Falls ja, was vermuten Sie, worum es geht?

Falls ja, ab welchem Zeitpunkt hatten Sie diese Vermutung?

Fanden Sie einige Fragen zu unangenehm oder/und zu privat, um diese zu beantworten?

☐ Ja ☐ Nein

Falls ja, welche?

Haben Sie im Verlauf des Versuchs die Lust verloren, weiter am Versuch teilzunehmen (z. B. weil Ihnen zu viele Fragen gestellt wurden)?

☐ Ja ☐ Nein

Falls ja, was war der Grund?

Gibt es sonst etwas, das Sie mitteilen möchten? Nutzen Sie die Gelegenheit für Kommentare und Anmerkungen!

Um den Fragebogen abzuschließen, klicken Sie unten rechts auf den Button "Absenden". Danach wenden Sie sich bitte an den Versuchsleiter, der Ihnen erklärt, wie es weitergeht. Vielen Dank!

I. FRAGEBOGEN DER NACHBEFRAGUNG

EvaSys	Untersuchungsziel	Electric Paper
Universität Osnabrück 03 Diplomarbeiten/Promotionen	Rohangis Mohseni Ziele	

Markieren Sie so: ☐ ☒ ☐ ☐ Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift. Dieser Fragebogen wird maschinell erfasst.
 Korrektur: ☐ ☒ ☐ ☐ Bitte beachten Sie im Interesse einer optimalen Datenerfassung die links gegebenen Hinweise beim Ausfüllen.

1. Vom Versuchsleiter auszufüllen

1.1 Probanden-Nr.

100er	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
10er	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
1er	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	

1.2 Geschlecht

☐ männlich

☐ weiblich

2. Untersuchungsziel

2.1 Welche Täuschungsmanöver gab es?

2.2 Reaktionszeit-Test durchschaut

☐ Ja

☐ Teils-Teils

☐ Nein

2.3 Konzentrations-Test durchschaut

☐ Ja

☐ Teils-Teils

☐ Nein

2.4 Was war das Ziel der Untersuchung?

2.5 Ziel der Untersuchung erkannt

☐ Ja

☐ Teils-Teils

☐ Nein

2.6 Sonstige Anmerkungen



Virtuelle Nothilfe

J. EINSTELLUNGEN DES CRTT

Tabelle J: Einstellungen des Competitive Reaction Time Tasks

Runde	Ergebnis Proband	Lautstärke Stufe	Lautstärke db (A)	Dauer Stufe	Dauer Sekunden
1	Verloren	3	60	3	1.5
2	Verloren	3	60	3	1.5
3	Gewonnen	4	65	3	1.5
4	Verloren	1	50	1	0.5
5	Verloren	2	55	1	1.5
6	Gewonnen	4	65	2	1.0
7	Gewonnen	1	50	4	2.0
8	Verloren	3	60	4	2.0
9	Gewonnen	2	55	2	1.0
10	Verloren	5	70	5	2.5
11	Gewonnen	5	70	5	2.5
12	Verloren	4	65	4	2.0
13	Gewonnen	6	75	6	3.0
14	Gewonnen	6	75	7	3.5
15	Verloren	7	80	6	3.0
16	Verloren	7	80	4	2.0
17	Gewonnen	4	65	7	3.5
18	Verloren	7	80	8	4.0
19	Gewonnen	8	85	7	3.5
20	Verloren	10	95	8	4.0
21	Gewonnen	8	85	9	4.5
22	Verloren	7	80	9	4.5
23	Gewonnen	9	90	10	5.0
24	Verloren	10	95	10	5.0
25	Gewonnen	9	90	7	3.5

Global Settings

File Help

Participant's Language: German

New Participant Entry & Selection Options:

☐ Trial Blocks
 ☐ Date of Birth
 ☐ Condition

Select Sound

Name: KipNoiseBlastStereo.wav

Sound Feedback

Intensity Actual dB Sound System Durations

Level Intensity Volume Volume (ms)

10 95 90 100 5000
 9 90 60 100 4500
 8 85 32 100 4000
 7 80 15 100 3500
 6 75 10 90 3000
 5 70 10 60 2500
 4 65 10 30 2000
 3 60 10 15 1500
 2 55 10 9 1000
 1 50 10 5 500

Feedback Type(s)

☒ Reward & Punish
☒ Punish Only
☒ Reward Only

Reward Function

\$ 0.05

Feedback Duration

3000

☐ Disable RT Practice

Reaction Time Modality

☒ Auditory
 ☐ Visual

Reaction Time Delays (ms)

Valid Reaction Time Range (before Automatic Loss)

Maximum RT 2000

Minimum RT 0

Random Delay Range from Warning (Yellow) to RT Stimulus (Red / Tone)

Maximum Time 2000

Minimum Time 1000

Time Before Automatic Trial Termination

Termination Time 3000

SIMULATED DELAYS

Delay at Start of Experiment (ms)

Simulated Wait for Opponent

Minimum Start Delay 1000

Maximum Start Delay 2000

Probability of No Delay 0

Delay at Start of Each Trial (ms)

Simulated Opponent Setting Levels

Minimum Trial Delay 1000

Maximum Trial Delay 2000

Probability of No Delay 0

Simulated Wait for Connection

Connection Delay 2000

Delay Between RT and Feedback

Feedback Delay 3000

Abb. 55: Einstellungen des CRT R&P

K. INSTRUKTIONEN ZUM CRTT (VORTEST)

- Damit wir ohne Zeitverzögerung nach dem Computerspiel das Reaktionszeit-Spiel durchführen können, soll es Ihnen schon jetzt erklärt werden.
- Sie werden nach dem Computerspiel gegen den anderen Probanden in einem Reaktionszeit-Spiel antreten.
- Jeder von Ihnen ist in einer anderen Versuchsbedingung, und wir möchten so sehen, nach welcher Versuchsbedingung man schneller auf eine Schaltfläche klicken kann.
- Sie bekommen das Spiel jetzt erklärt und können es nach der Erklärung ausprobieren.
- Das Spiel besteht aus 25 Durchgängen.
- Der **Verlierer** eines Durchgangs hört einen Rauschton.
- Sie stellen Lautstärke und Dauer des Tons für Ihren Gegenspieler ein, während Ihr Gegenspieler dasselbe für Ihren Ton macht.
- Hier (**->auf den linken Slider mit der Beschriftung „Lautstaerke“ zeigen**) stellen Sie die Lautstärke ein, die Ihr Mitspieler hören soll, wenn er langsamer als Sie reagiert hat.
- Nachdem Sie die Lautstärke eingestellt haben, stellen Sie hier (**->auf den mittleren Slider mit der Beschriftung „Dauer“ zeigen**) die Dauer ein. Die maximale Dauer beträgt 5 Sekunden (**-> auf Einstellung 5 des Sliders zeigen**).
- Diese Schaltfläche zeigt Ihnen später durch Ampelfarben die Spielphase an. Sie wechselt von grün über gelb zu rot.
- Wenn die Schaltfläche rot wird, drücken Sie die linke Maustaste so schnell sie können.
- Einige Probanden versuchen, sich einen Vorteil zu verschaffen, indem sie bei gelb dauerhaft klicken. Dadurch hat man aber keinen Vorteil, weil das Spiel erst weiterläuft, wenn man mit dem Klicken aufhört. Würden Sie und Ihr Gegner gleichzeitig Dauerklicken, würde das Spiel nie enden.
- Wenn Sie wie vorgesehen bei Rot geklickt haben, dann zeigt der Computer Ihnen an, ob Sie gewonnen oder verloren haben.
- Sie können hier jederzeit sehen, welche Einstellungen Ihr Mitspieler für Sie vorgesehen hatte (**-> auf die linken Slider unter der Überschrift „Rueckmeldung Ihres Gegenspielers an Sie“ zeigen**).
- Sie hören aber nur dann den dort eingestellten Rauschton, wenn Ihr Mitspieler schneller war.
- Auch Ihr Mitspieler bzw. Ihre Mitspielerin kann jederzeit sehen, welche Einstellungen Sie vorgesehen haben. Er bzw. sie hört den Rauschton aber nur, wenn Sie schneller waren.

Nun möchte ich Ihnen einige Beispiele für Lautstärken geben, die im Reaktionszeit-Spiel verwendet werden, damit Sie wissen, wie laut diese sind.

- (**-> Auf den Bildschirm zeigen**). Lautstärke 0 entspricht keinem Rauschton; Lautstärke 1 ist ein Ton mit 50 Dezibel; Lautstärke 2 ein Ton mit 55 Dezibel; Lautstärke 3 ein Ton mit

Virtuelle Nothilfe

60 Dezibel und so weiter. Wenn Ihr Mitspieler keinen Rauschton hören soll, dann müssen Sie wegen eines Programmierfehlers den Kasten mindestens auf Lautstärke 1 und von dort wieder zurück auf 0 stellen.

- Bitte setzen Sie nun die Kopfhörer auf.
- Ich möchte Ihnen nun ein Beispiel für einen Ton auf Stufe 1 geben, die 50 Dezibel entspricht. (-> **Stufe 1 für 2 Sekunden**).
- Konnten Sie den Ton auf beiden Ohren hören? (-> **Wenn der Proband „Nein“ sagt, ggf. das Kopfhörerkabel fest einstecken, weil es sich gelöst haben könnte, und dann solange probieren, bis der Proband auf beiden Ohren einen Ton hören kann**).
- Nun möchte ich Ihnen ein Beispiel für einen Rauschton auf Stufe 5 geben, was 70 Dezibel entspricht. (-> **Stufe 5 für 2 Sek.**).
- Als nächstes möchte ich Ihnen ein Beispiel für einen längeren Rauschton auf Stufe 5 geben. (-> **Stufe 5 für 5 Sek.**).
- Auch Sie können auf diese Weise die Dauer des Tons einstellen, den Ihr Mitspieler hört.
- Zum Schluss möchte ich Ihnen ein Beispiel für einen Ton auf Stufe 10 geben, was 95 Dezibel entspricht. Dies entspricht der Lautstärke in einer Diskothek. Sind Sie bereit für einen Ton auf Stufe 10? (**Stufe 10 für 1 Sekunde**).
- Bitte legen Sie nun die Kopfhörer ab.
- Ich möchte Sie nun das Einstellen der Regler und das Klicken auf die Schaltfläche üben lassen.
- (-> **Auf „Weiter“ klicken. Den Probanden das Einstellen und das Klicken üben lassen**).

Nach dem Computerspiel

- Zuerst muss ich prüfen, ob die andere Versuchsperson ebenfalls bereit ist. (-> **Nachsehen und fortfahren, falls alles OK ist**).

L. INSTRUKTIONEN ZUM CRTT (HAUPTUNTERSUCHUNG)

- Sie werden gleich Oblivion spielen. Der andere Proband auch.
- Jeder von Ihnen ist in einer anderen Versuchsbedingung, und wir möchten sehen, nach welcher Versuchsbedingung man kürzere Reaktionszeiten hat.
- Sie werden daher nach Oblivion gegen den anderen Probanden in einem Reaktionszeit-Spiel antreten.
- Damit wir mit möglichst wenig Zeitverzögerung nach Oblivion das Reaktionszeit-Spiel durchführen können, erklären wir es Ihnen schon jetzt.
- Das Spiel hat insgesamt 25 Runden. In jeder Runde sollen Sie so schnell wie möglich auf diese Schaltfläche (-> **zeigen**) klicken.
- Zu Beginn jeder Runde wird die Fläche erst grün, dann gelb, dann rot.
- Wenn die Schaltfläche rot wird, drücken Sie die linke Maustaste so schnell sie können.
- Einige Probanden haben versucht, zu gewinnen, indem sie bei gelb andauernd klicken (-> **demonstrieren**). Bitte tun Sie dies nicht. Zum einen können wir so keine Reaktionszeiten messen. Zum anderen hat das Spiel einen Schutz: Es läuft erst weiter, wenn Sie mit dem Klicken aufhören. Falls Sie und Ihr Gegner gleichzeitig Dauerklicken, würde das Spiel nie enden.
- Wenn Sie wie vorgesehen bei Rot geklickt haben, dann zeigt der Computer Ihnen hier an (-> **auf Fläche zeigen**), ob Sie gewonnen oder verloren haben.
- Jetzt kommt die Besonderheit: Der **Verlierer** eines Durchgangs hört über die Kopfhörer ein Geräusch.
- Am Anfang jeder Runde stellen Sie die Lautstärke und Dauer des Tons ein, den Ihr Gegenspieler hört, falls er verliert.
- Ihr Gegenspieler macht dasselbe für den Ton, den Sie hören, falls Sie verlieren.
- Hier (-> **auf „Lautstaerke“-Slider zeigen**) stellen Sie die Lautstärke ein.
- Lautstärke 0 entspricht keinem Rauschton; Lautstärke 1 ist ein Ton mit 50 Dezibel; dann wird es jeweils 5 Dezibel lauter. Lautstärke 5 sind 70 Dezibel und Lautstärke 10 sind 95 Dezibel. Ich gebe Ihnen nachher ein paar Probetöne.
- Hier (-> **auf „Dauer“-Slider zeigen**) stellen Sie die Dauer ein. Bei Null hört man keinen Ton; dann kommt eine halbe Sekunde, eine ganze Sekunde bis maximal 5 Sekunden (-> **auf Einstellung 5,0 des Sliders zeigen**).
- Am Ende jeder Runde können Sie sehen, welche Einstellungen Ihr Gegenspieler für Sie vorgesehen hatte (-> **auf die Slider unter der Überschrift „Rueckmeldung Ihres Gegenspielers an Sie“ zeigen**).
- Das dient dazu, dass Sie nicht glauben, dass der Ton lauter oder leiser war, als er in Wirklichkeit war.
- Ihr Gegenspieler hat den gleichen Bildschirm wie Sie vor sich. Er kann auch sehen, welche Einstellungen Sie gemacht haben.

Virtuelle Nothilfe

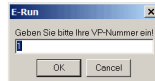
Nun möchte ich Ihnen ein paar Beispiele für die Lautstärken geben:

- Bitte setzen Sie die Kopfhörer auf.
- Wir beginnen mit einem Ton auf Stufe 1, was 50 Dezibel entspricht. (-> **Stufe 1 für 2 Sekunden**).
- Konnten Sie den Ton auf beiden Ohren hören? (-> **bei „Nein“ ggf. Kabel neu einstecken, Mute-Taste prüfen**).
- Als nächstes ein Ton auf Stufe 5. Das sind 70 Dezibel. (-> **Stufe 5 für 2 Sek.**).
- Jetzt ein Ton für 5 Sekunden, damit Sie sehen, wie lang das ist. (-> **Stufe 5 für 5 Sek.**).
- Zum Schluss ein Ton auf Stufe 10, was 95 Dezibel entspricht. Das ist Disco-Lautstärke. Sind Sie bereit? (**Stufe 10 für 1 Sekunde**).
- Sie können die Kopfhörer jetzt ablegen.
- Ich möchte Ihnen nun zeigen, wie das mit dem Klicken funktioniert. (-> **Auf „Weiter“ klicken.**)
- Zuerst stellen Sie die Regler ein. Dann klicken Sie auf die Schaltfläche. Diese wird grün – gelb – rot.
- Die Fläche wird übrigens immer kurz grün, auch wenn der Gegenspieler mit dem Einstellen fertig ist. Der Grund ist, weil zwischen gelb und rot manchmal nur eine Sekunde vergeht und Sie sonst keine Zeit haben, sich auf das Klicken einzustellen.
- Damit Sie nicht versehentlich die Runde starten, können Sie erst auf Bereit klicken, wenn Sie die Regler verstellt haben.
- Wenn Sie die Null einstellen wollen, müssen Sie daher erst beide Regler verschieben, und dann mindestens einen wieder auf Null zurückstellen.
- Bitte probieren Sie nun einmal aus, wie das Einstellen und Klicken funktioniert. (**Den Probanden das Einstellen und das Klicken üben lassen**).

Nach dem Computerspiel

- Zuerst muss ich prüfen, ob die andere Versuchsperson ebenfalls bereit ist. (-> **Nachsehen und fortfahren, falls alles OK ist**).
- Wir können nun das Reaktionszeit-Spiel beginnen. Wenn Sie möchten, können Sie vor dem Start noch einmal die Funktionen des Spiels ausprobieren.
- Haben Sie noch eine Frage? Wenn nicht, klicken Sie bitte auf Weiter (-> **Auf die Schaltfläche „Weiter“ zeigen**).
- Ich werde gleich den Raum verlassen. Benachrichtigen Sie mich bitte, wenn Sie mit dem Spiel durch sind.
- Setzen Sie bitte die Kopfhörer auf. Setzen Sie sie erst wieder ab, wenn das Reaktionszeit-Spiel beendet ist.
- Klicken Sie nun auf die Schaltfläche „Experiment beginnen“.

M. ECKENZÄHLAUFGABE



ZIEL DER STUDIE

Im Rahmen meiner Diplomarbeit möchte ich die Konzentrationsfähigkeit am Arbeitsplatz untersuchen.

Mich interessiert, wie gut sich Mitarbeiter konzentrieren können, wenn sie längere Zeit kognitiv anstrengende Tätigkeiten durchführen mussten.

(weiter mit der Leertaste)

Virtuelle Nothilfe

ZIEL DER STUDIE

Ich habe dazu einen Test entwickelt, der am Computer durchgeführt werden kann. Für den Test habe ich 200 Aufgaben erstellt. Da es aber zu lange dauern würde, immer alle 200 Aufgaben zu lösen, sollen in dieser VORSTUDIE die besten Aufgaben ausgewählt werden.

(weiter mit der Leertaste)



ZIEL DER STUDIE

Um die besten Aufgaben auswählen zu können, muss ich wissen, wie schwierig diese sind. Je mehr Aufgaben Sie bearbeiten, desto schneller kann ich die Vorstudie abschließen und mit meiner Hauptuntersuchung beginnen.

(weiter mit der Leertaste)



Virtuelle Nothilfe

ZIEL DER STUDIE

Ich erwarte aber nicht, dass Sie alle 200 Aufgaben bearbeiten. Sie machen zuerst einen Probedurchgang. Nach diesem geben Sie bitte an, wieviele Aufgaben sie noch lösen möchten.

Das Programm wählt dann aus den 200 Aufgaben diejenigen aus, die bisher von den wenigsten Personen bearbeitet wurden.

(weiter mit der Leertaste)



ABLAUF DER STUDIE

Zuerst erkläre ich, worin eine Aufgabe besteht.

Dann sollen Sie in einem Probedurchgang sechs Aufgaben lösen.

Danach folgt der eigentliche Hauptdurchgang.

(weiter mit der Leertaste)



Virtuelle Nothilfe

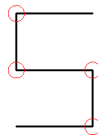
AUFGABENBESCHREIBUNG

Die Aufgabe besteht darin, die ECKEN einer Figur zu zählen.
Zur besseren Vorstellung folgt jetzt ein Beispiel, in
welchem die zu zählenden Ecken rot eingekreist sind.

(weiter mit der Leertaste)



BEISPIEL



(weiter mit Leertaste)



Virtuelle Nothilfe

ANLEITUNG

Die gezeigte Figur beinhaltet 4 (rot eingekreiste) Ecken.

Sie sollen diese Ecken zählen und danach angeben, ob die Zahl der Ecken, die Sie gezählt haben, GERADE oder UNGERADE war.

Bitte drücken Sie danach die folgende Taste:

GERADE --> Taste „G“

UNGERADE --> Taste „U“

(weiter mit der Leertaste)

↳

WICHTIGE HINWEISE

Die Aufgaben sollen Konzentrationsfähigkeit erfassen. Wichtiger als Schnelligkeit ist, dass Sie versuchen, korrekt zu antworten. Bearbeiten Sie die einzelnen Aufgaben nach Ihrem eigenen Tempo.

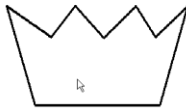
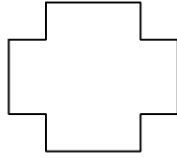
Damit Sie ein Gefühl für die unterschiedlichen Schwierigkeitsgrade bekommen, sind die Aufgaben im Probedurchgang unterschiedlich schwierig: Die ersten beiden Aufgaben sind leicht, die nächsten beiden mittel und die letzten beiden schwer.

Es folgt nun der PROBEDURCHGANG.

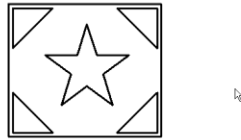
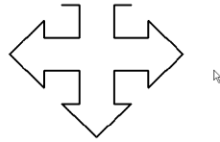
(zum Probedurchgang mit der Leertaste)

↳

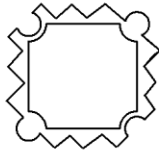
Virtuelle Nothilfe



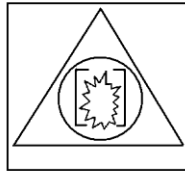
Virtuelle Nothilfe



Virtuelle Nothilfe



42



42

Virtuelle Nothilfe

HAUPTDURCHGANG

Der PROBEDURCHGANG ist beendet. Bitte tippen Sie nun ein, wieviele Aufgaben sie noch lösen möchten. Das Programm wählt dann aus den 200 Aufgaben diejenigen aus, die bisher von den wenigsten Personen bearbeitet wurden.

Danach startet der eigentliche HAUPTDURCHGANG. Dort müssen Sie so viele Aufgaben lösen, wie Sie an dieser Stelle angeben.

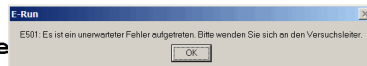
■

(Bitte die Aufgabenanzahl eintippen und danach die Eingabetaste drücken)

HAUPTDURCHGANG

Der PROBEDURCHGANG ist beendet. Bitte tippen Sie nun ein, wieviele Aufgaben sie noch lösen möchten. Das Programm wählt dann aus den 200 Aufgaben diejenigen aus, die bisher von den wenigsten Personen bearbeitet wurden.

Danach startet der eigentliche HAUPTDURCHGANG. Dort müssen Sie so viele Aufgaben lösen, wie Sie an dieser Stelle angeben.



100



(Bitte die Aufgabenanzahl eintippen und danach die Eingabetaste drücken)

N. DATENVERTEILUNG BEIM GEWALTVERHALTEN

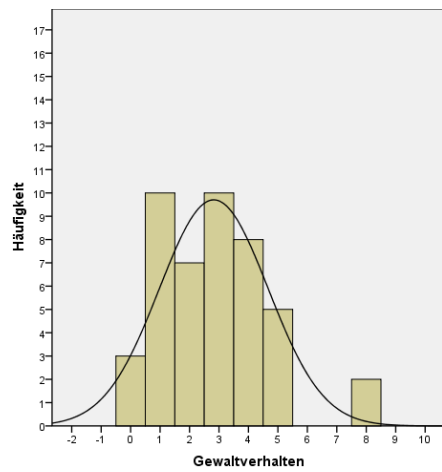


Abb. 56: Histogramm Nothilfe

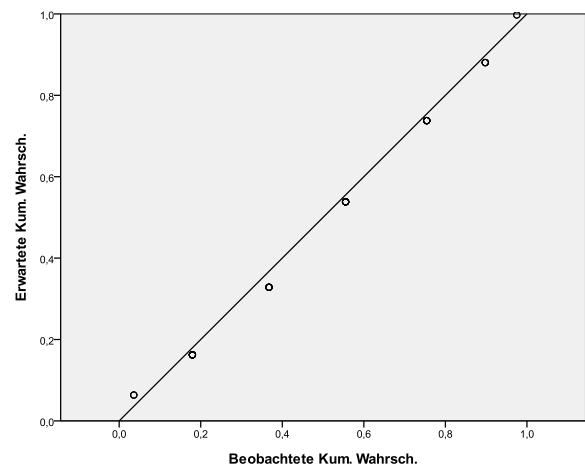


Abb. 57: Probability Plot Nothilfe

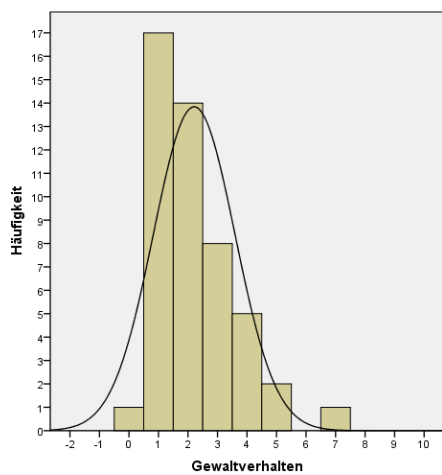


Abb. 58: Histogramm Hilfe

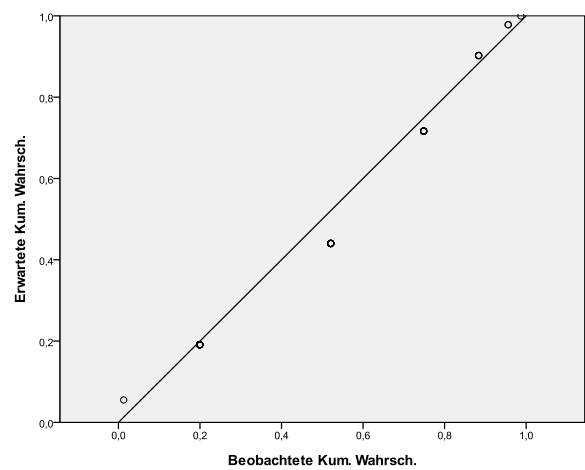


Abb. 59: Probability Plot Hilfe

Virtuelle Nothilfe

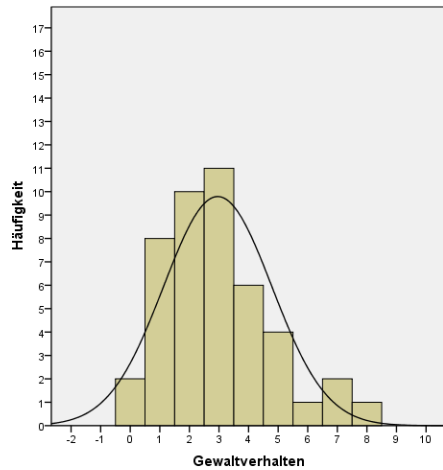


Abb. 60: Histogramm Töten

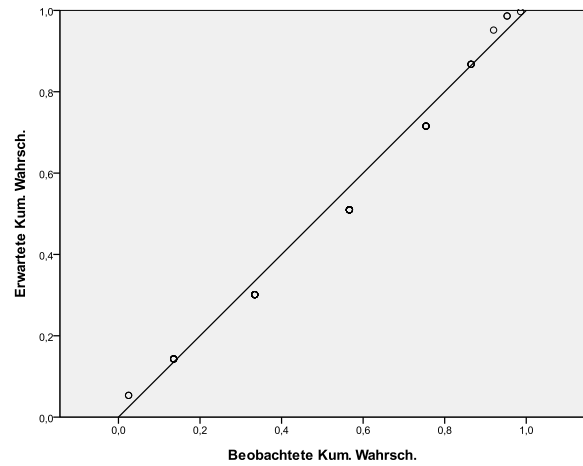


Abb. 61: Probability Plot Töten

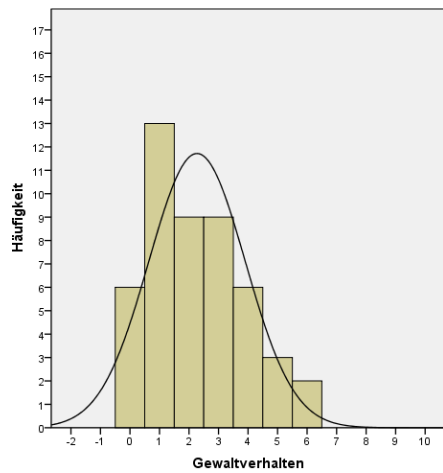


Abb. 62: Histogramm Schatzjagd

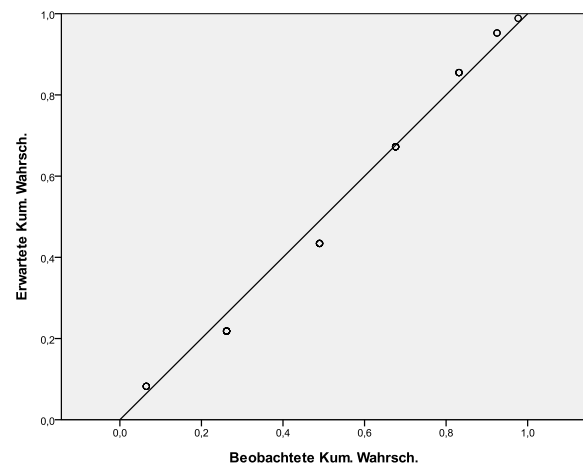


Abb. 63: Probability Plot Schatzjagd

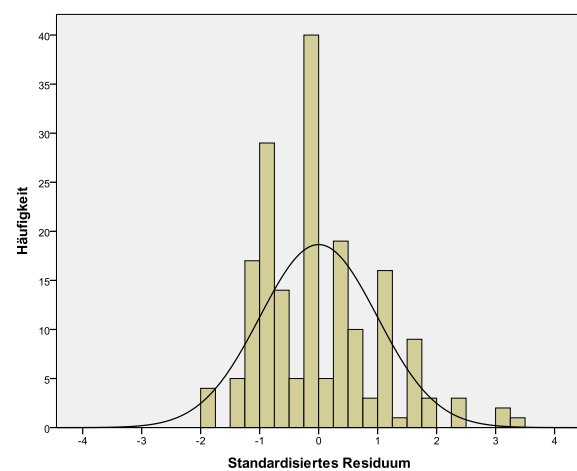


Abb. 64: Histogramm Schrittweise-Regression

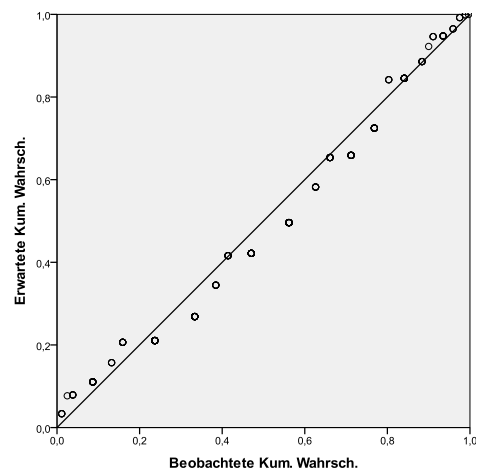


Abb. 65: Probability Plot Schrittweise-Regression

O. DATENVERTEILUNG BEIM HILFEVERHALTEN

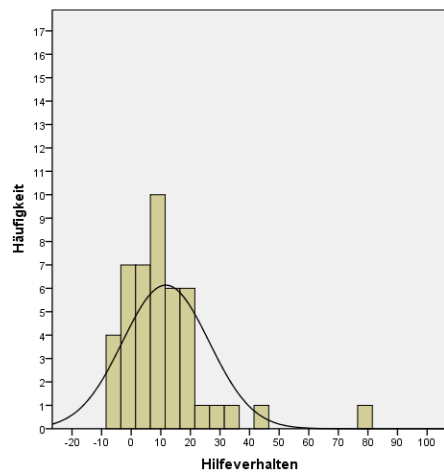


Abb. 66: Histogramm Nothilfe

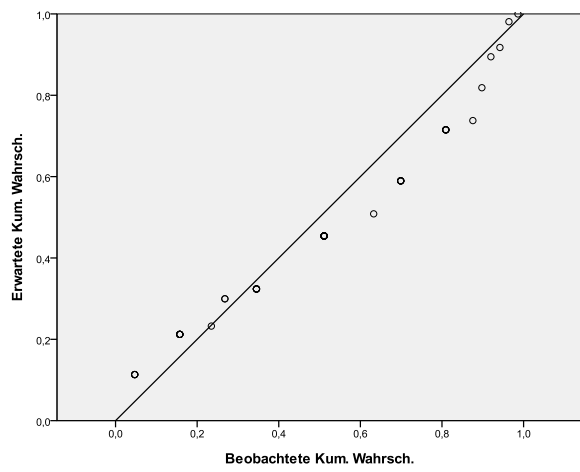


Abb. 67: Probability Plot Nothilfe

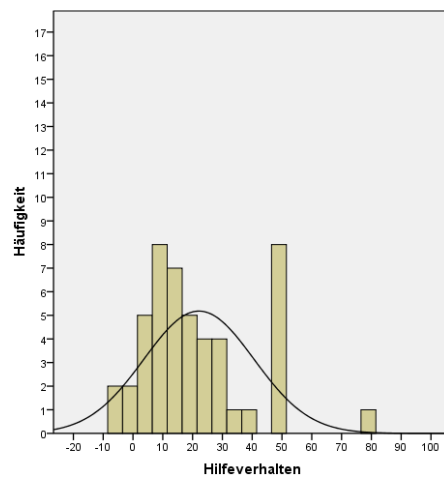


Abb. 68: Histogramm Hilfe

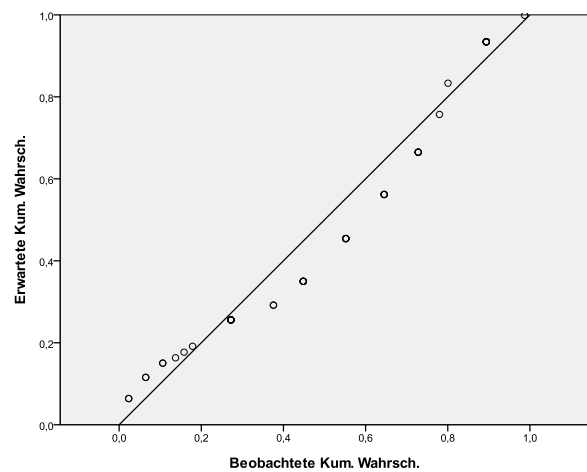


Abb. 69: Probability Plot Hilfe

Virtuelle Nothilfe

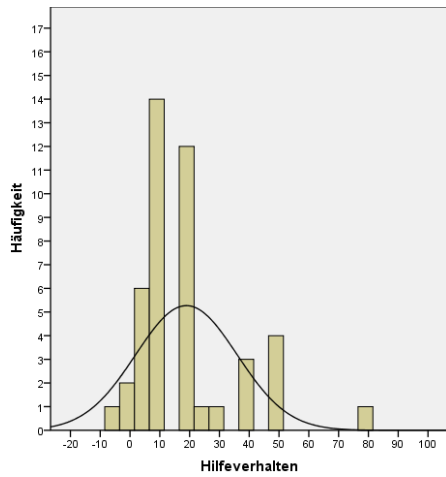


Abb. 70: Histogramm Töten

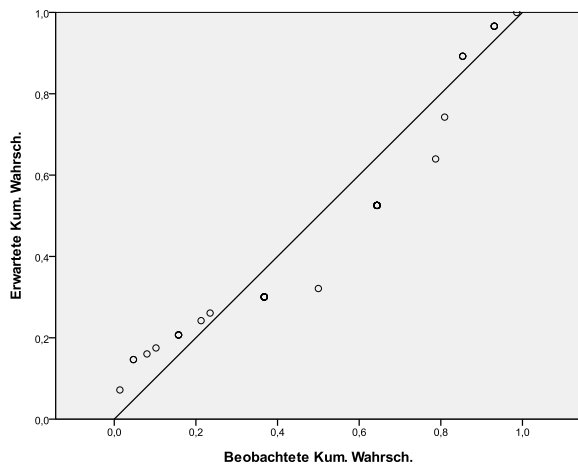


Abb. 71: Probability Plot Töten

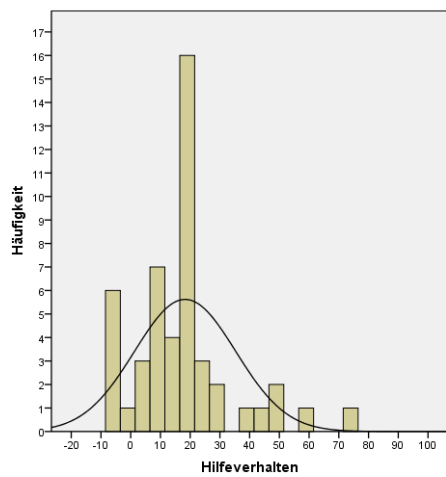


Abb. 72: Histogramm Schatzjagd

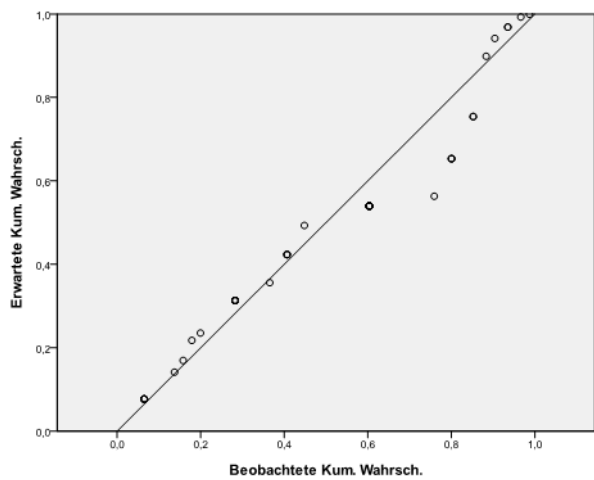


Abb. 73: Probability Plot Schatzjagd

Virtuelle Nothilfe

P. HOMOGENITÄT DER REGRESSIONEN

Tabelle P1: Kovarianzanalyse zur Prüfung der Homogenität der Regressionen mit „Gewaltverhalten“ als abhängige Variable und „Erhebungswelle“ sowie „Computergegner durchschaut“ als Kovariaten

Quelle der Varianz	SS	df	MS	F	p
Gewalt im Spiel (GiS)	0.002	1	0.002	0.001	.976
Hilfe im Spiel (HiS)	5.269	1	5.269	1.942	.165
GiS × HiS	0.179	1	0.179	0.066	.798
Erhebungswelle	16.920	1	16.920	6.236	.013
GiS × Erhebungswelle	0.288	1	0.288	0.106	.745
HiS × Erhebungswelle	5.926	1	5.926	2.184	.141
GiS × HiS × Erhebungswelle	0.345	1	0.345	0.127	.722
Computergegner durchschaut	12.958	1	12.958	4.776	.030
GiS × Computergegner durchschaut	0.516	1	0.516	0.190	.663
HiS × Computergegner durchschaut	6.761	1	6.761	2.492	.116
GiS × HiS × Computergegner durchschaut	2.064	1	2.064	0.761	.384
Fehler	472.124	174	2.713		
Gesamt	533.962	186			

Tabelle P2: Kovarianzanalyse zur Prüfung der Homogenität der Regression mit „Hilfeverhalten“ als abhängige Variable und „Spielaufgabe Realitätsnähe Grafik“, „Spielaufgabe Aktivierung (lustlos)“ sowie „Spielaufgabe Schwierigkeit“ als Kovariaten

Quelle der Varianz	SS	df	MS	F	p
Gewalt im Spiel (GiS)	1123.317	1	1123.317	4.296	.040
Hilfe im Spiel (HiS)	205.359	1	205.359	0.785	.377
GiS × HiS	1241.469	1	1241.469	4.747	.031
Spielaufgabe Realitätsnähe Grafik	1501.974	1	1501.974	5.744	.018
GiS × Spielaufgabe Realitätsnähe Grafik	20.057	1	20.057	0.077	.782
HiS × Spielaufgabe Realitätsnähe Grafik	50.940	1	50.940	0.195	.660
GiS × HiS × Spielaufgabe Realitätsnähe Gr.	651.688	1	651.688	2.492	.116
Spielaufgabe Schwierigkeit	881.150	1	881.150	3.370	.068
GiS × Spielaufgabe Schwierigkeit	509.159	1	509.159	1.947	.165
HiS × Spielaufgabe Schwierigkeit	295.413	1	295.413	1.130	.289
GiS × HiS × Spielaufgabe Schwierigkeit	497.743	1	497.743	1.903	.170
Spielaufgabe Aktivierung (lustlos)	1718.327	1	1718.327	6.571	.011
GiS × Spielaufgabe Aktivierung (lustlos)	339.453	1	339.453	1.298	.256
HiS × Spielaufgabe Aktivierung (lustlos)	50.172	1	50.172	0.192	.662
GiS × HiS × Spielaufgabe Akt. (lustlos)	184.214	1	184.214	0.704	.402
Fehler	44454.967	170	261.500		
Gesamt	54499.833	186			

Q. FORMELN

Formel Q1: Umrechnung der Effektstärke r in Hedges g

$$g = \frac{r}{\sqrt{1-r^2}} \times \sqrt{\frac{df \times (n_1 + n_2)}{n_1 \times n_2}}, df = n_1 + n_2 - 2$$

nach Rosenthal (1994, S. 236-239)

Formel Q2: t -Test zum a priori Einzelvergleich zweier Gruppen

$$t = \frac{D}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \times MS_{\text{Fehler}}}}; df = (N - J \times K)$$

nach Hussy und Jain (2002, S. 210); Bortz (1999, S. 253f.)

Formel Q3: Umrechnung der Effektstärke f^2 in η^2

$$\eta^2 = f^2 / (1 + f^2)$$

nach Cohen (1988, S. 273-289)

Formel Q4: Ermittlung der Effektstärke ω^2 in einer ANOVA

$$\omega^2 = \frac{SS_{\text{Effekt}} - df_{\text{Effekt}} \times MS_{\text{Fehler}}}{SS_{\text{Total}} + MS_{\text{Fehler}}}$$

nach Tabachnick & Fidell (2007, S. 55); Bortz (1999, S. 270)

Formel Q5: Tukey-Kramer-Test für multiple Mittelwertvergleiche

$$Q^* = \frac{|M_j - M_{j'}|}{\sqrt{\left[MS_{\text{Fehler}} \times \left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right) \div 2\right]}}, df = N - J$$

nach Diehl und Arbinger (1992, S. 226)

¹ Beispielsweise ist es unethisch, für sich selbst als Wissenschaftler Rationalität und Reflexivität zu beanspruchen, den untersuchten Personen aber diese Eigenschaften a priori und ohne Begründung abzusprechen.

² Spiele können danach unterschieden werden, an welcher Art Gerät sie gespielt werden. Im Falle eines PCs werden sie häufig „Computerspiele“ genannt, im Falle einer Spielekonsole heißen sie „Videospiele“ (Klimmt, 2006a, S. 17f.). Ursprünglich gab es zwischen diesen Spielen größere Unterschiede, da die ersten Spielekonsolen wenig leistungsfähig waren, aber über andere Arten von Eingabegeräten verfügten (z. B. Joystick, Gamepad etc.). Inzwischen sind die Unterschiede nur noch gering. Heutzutage werden aus wirtschaftlichen Gründen die meisten Spiele sowohl für den PC als auch die Spielekonsole entwickelt. Daher ist für die Spielwirkungsfor-schung eine Unterscheidung zwischen Video- und Computerspiel wenig zielführend. Als Oberbegriff schlägt Klimmt (Klimmt, 2006a, S. 17f.) den Begriff „Computerspiel“ vor, da Video- und Computerspiele beide eine Form von Computer benötigen, um verwendet werden zu können. Diesem Vorschlag wird hier gefolgt. Der Begriff „Computerspiel“ wird synonym für Video- und Computerspiele verwendet, d. h. wenn von Computerspielen die Rede ist, sind Videospiele mitgemeint.

³ Hier fasst er die sich stärker überschneidenden Genres Actionspiele, Adventurespiele, Action-Adventure-Spiele, Ego-Actionspiele und Plattformspiele zu einem einzigen Genre zusammen, da die Grenzen fließend sind und meistens keine Kriterien zu einer genauen Abgrenzung gegeben werden.

⁴ Tedeschi und Felson (1994, S. 91) kommen nach einer Sichtung der Studien zu Zillmanns Erregungstransfermodell zu dem Schluss, dass die Theorie durch die bisherigen Studien nicht richtig untersucht wurde. Die Theorie postuliert, dass eine vorhandene abklingende Erregung sich auf eine nachfolgende Provokationssituation überträgt und so den durch die Provokation erzeugten Ärger verstärkt. In den Experimenten wird jedoch zuerst eine Provokationssituation hergestellt und erst im Anschluss eine Erregung evoziert. Die Reihenfolge ist somit vertauscht.

⁵ Der Begriff „coercion“ beschreibt das Phänomen, dass eine Person eine andere zu etwas zwingt. Da das Wort „Nötigung“ ein juristischer Fachbegriff ist, scheint der Begriff „Zwang“ die universellere Übersetzung zu sein. Hier wird trotzdem der Begriff „Nötigung“ bevorzugt, da sich dieser Begriff mit „Verhalten“ zu „Nötigungsverhalten“ kombinieren lässt, ohne dass dieser zusammengesetzte Begriff missverständlich erscheint. Der Begriff „Zwangverhalten“ hingegen wäre zu leicht mit dem klinischen Begriff des „Zwangsverhaltens“ verwechselbar, während der Begriff „Zwingverhalten“ bzw. „Erzwingverhalten“ ohne Erläuterung wenig verständlich ist.

⁶ Nach Scheele und DuBois (2006, S. 406) stellen diese Ergebnisse aber keine vollständige Widerlegung des rezeptionspsychologischen Purgations-Konzeptes dar, da bisher nur Ärger/Aggression untersucht wurden, sich die aristotelische Katharsishypothese aber auf mehr als nur diese Emotion bzw. dieses Verhalten beziehen würde.

⁷ $k = 5$, $n = 483$, $r+ = .29$ (um Publikations-Bias korrigiert $.15$), Fail-Safe $N = 36$, Orwin's Fail-Safe $N = 6$, $p \leq .05$ für Begg & Mazumdar's Rangkorrelationstest, $p \leq .001$ für Egger's Regression, 95% CI = $-.06, .35$

⁸ $k = 9$, $n = 2150$, $r+ = .15$ (ohne Publikations-Bias $.06$), Fail-Safe $N = 86$, Orwin's Fail-Safe $N = 3$, $p \leq .06$ für Begg & Mazumdar's Rangkorrelationstest, $p \geq .05$ für Egger's Regression, 95% CI = $-.04, .16$

⁹ Hier werden nur die Ergebnisse für experimentelle Querschnitt-Studien berichtet.

¹⁰ Vereinzelt finden sich Abweichungen vom Standard. Bei Giancola und Zeichner (1995b) mussten die Probanden ihre Einstellungen nicht am Anfang, sondern am Ende jeder Runde vornehmen. Bei Bartholow und Anderson (2002) wurde das CRTT zweimal hintereinander durchgeführt, wobei der Gegner die Einstellungen nur im ersten Durchgang und der Proband sie nur im zweiten Durchgang vornehmen konnte (siehe dazu auch Elson, 2011).

¹¹ Wie gut schmeckt Ihnen das Getränk (1 = sehr gut; 9 = sehr schlecht)? Angenehmer Saft: $M = 1.89$; Neutraler Saft: $M = 5.52$; Scharfer Saft: $M = 8.21$. Wie unangenehm wäre es Ihrer Meinung nach, ein kleines Glas voll von diesem Saft zu trinken (1 = gar nicht unangenehm; 9 = extrem unangenehm)? Angenehmer Saft: $M = 1.47$; Neutraler Saft: $M = 3.79$; Scharfer Saft: $M = 8.32$. Wie scharf schmeckt das Getränk Ihrer Ansicht nach (1 = überhaupt nicht scharf; 9 = extrem scharf)? Angenehmer Saft: $M = 1.21$; Neutraler Saft: $M = 1.37$; Scharfer Saft: $M = 7.47$.

¹² Diese Vermutung ließe sich klären mit Hilfe der Antworten zu der Frage „Wie sehr glauben Sie, dass Ihr Gegenüber das ausgewählte Getränk (nicht) mag?“. Die Daten wurden angefragt, von den Autor/innen aber nicht zur Verfügung gestellt.

¹³ Es gibt zwar einen Zusammenhang mit Psychopathie, aber Psychopathen unterschätzen den verursachten Schmerz (Russel et al. 2002, zit. n. Ritter & Eslea, 2005, S. 412), sodass sie möglicherweise überzeugt sind, die Schüsse wären mehr oder weniger harmlos. Hier besteht eine Parallele zum „rough and tumble play“ beim Bobo Modeling (Ritter & Eslea, 2005, S. 412).

¹⁴ In der in dieser Studie verwendeten computerbasierten Form von „willingness to assist“ sind die Selbstdarstellungsmöglichkeiten gering, da der Versuchsleiter in der Erhebungssituation nicht genau erkennen kann, wie viele Aufgaben ein Proband bearbeitet hat. Er kann höchstens über die Dauer Rückschlüsse ziehen.

¹⁵ Dies entspricht auch dem Stereotyp der „damsel in distress“, d. h. einer Frau in Not, der von einem ritterlichen Mann geholfen werden muss.

¹⁶ Das Programm ist verfügbar unter <http://uk.groups.yahoo.com/group/CRTTP/>.

¹⁷ Die Kalibrierung wurde mit dem Lautstärkemessgerät „Type 2232“ von Brüel & Kjær durchgeführt, welches über einen Messbereich von 34 bis 130 dB(A) verfügt, mit einem hochsensiblen Kondensatormikrofon ausgestattet ist und damit eine Messgenauigkeit der Klasse 1 (engl. „Type 1 Precision“) erreicht. Vor der Messung wurde die Lautstärke von Windows auf 70% eingestellt. Das führte zu einer hohen Lautstärke, war aber nötig, um im CRT R&P den Wert von 95 dB(A) erreichen zu können. Danach wurde im CRT R&P die Systemlautstärke (engl. „System Volume“) auf den Wert 100 fixiert. Dann wurde der verwendete Rauschton „KipNoiseBlastStereo“ abgespielt. Dabei wurde das Messgerät mit einem Abstand von 2-3 mm an eine Muschel des Kopfhörers gehalten und so die maximale Lautstärke des Rauschtons gemessen. Die Geräuschlautstärke (engl. „Sound Volume“) wurde solange reduziert, bis auch nach drei Messungen die Maximallautstärke der jeweiligen Stufe nicht überschritten wurde. Ab Intensität 6 musste auch die Systemlautstärke reduziert werden, weil sonst die gewünschten Werte nicht erzielbar waren.

¹⁸ Das Programm basierte auf der Experimentalsoftware „E-Prime 2“.

¹⁹ Die Subskala musste zu diesem Zweck vom Englischen ins Deutsche übersetzt werden. Dabei wurde wie folgt verfahren: Zwei des Englischen mächtige Personen übersetzten die Items unabhängig voneinander. Alle Items, die nicht auf die gleiche Weise übersetzt wurden, wurden im Anschluss an eine professionelle Übersetzerin wei-

tergereicht. Deren Übersetzung stimmte in jedem Fall mit einer der bereits vorliegenden Übersetzungen überein.

²⁰ Im ersten Kategoriensystem wurde expliziert, dass nicht nur Aggression/Gewalt/Rache unter die Kategorien 4 und 5 fallen, sondern auch alle bedeutungsähnlichen Begriffe (z. B. Schadenfreude). Außerdem wurde klargestellt, dass es irrelevant ist, ob ein Proband die von ihm genannten Begriffe als abhängige oder unabhängige Variable auffasst. Im zweiten Kategoriensystem wurde bei Kategorie 4 die Anweisung ergänzt, diese Kategorie nicht zu kodieren, wenn ein Proband die Aussage erst macht, nachdem er über das Experiment aufgeklärt wurde.

²¹ „Gemütszustand“ wurde als emotionaler Zustand aufgefasst. Die Probandenaussagen „Ab dem/ Am Anfang des/ Zum Zeitpunkt des Reaktionszeitspiel/s“ wurden so gewertet, dass der Proband im Zweifelsfall den Zeitpunkt vor der ersten Runde des CRTT gemeint hat. Wenn der Proband geäußert hat, dass er ab dem CRTT bemerkt hat, dass es um Aggression geht, dann wurde davon ausgegangen, dass er das CRTT als Aggressionsmaß ansieht.

²² In Kategorie „Proband ist Gegner“ wurde ergänzt, dass hierunter auch fällt, wenn der Computer über Sieg und Niederlage entschieden haben soll, während in Kategorie „Computer ist Gegner“ ergänzt wurde, dass hierunter auch fällt, wenn der Computer die Lautstärken eingestellt haben soll.

²³ Tedeschi und Felson (1994, S. 63) unterscheiden eine schwache und eine starke Form der „demand characteristics“. Bei der starken Form erlauben Hinweisreize den Probanden, die Hypothese(n) des Experiments zu erschließen, während bei der schwachen Form die Hinweisreize den Probanden zu erkennen geben, welches Verhalten von ihnen erwartet wird, ohne dass die Probanden dadurch die Hypothese(n) erschließen.

²⁴ Alle Fragen wurden vom Englischen ins Deutsche übersetzt und dabei in der Stärke der Pol-Formulierung aneinander angeglichen.

²⁵ Der einzige Datensatz, bei dem nur fünf der sechs Banditen getötet wurden, war an dieser Stelle der Auswertung bereits eliminiert worden.

²⁶ Genau genommen hat die Kurtosis den Wert drei. Aus Gründen der einfacheren Darstellbarkeit wird in SPSS für die Kurtosis ein um drei verminderter Wert angezeigt. Dieser Darstellungsweise wird hier gefolgt.

²⁷ Dem Autor bekannte Beispiele für solche Spiele wären „Bioshock“ Teil 1 und 2, „Prince of Persia: Sands of Time“, „Mass Effect“ Teil 1 und 2 etc.