

Zur Reliabilität von Schmerzmessungen

DIPLOMARBEIT

von

Carsten Maurischat

Kiel, 1995

Betreuung: PD Dr. Fehm-Wolfsdorf & Dr. Pohl

Carsten Maurischat
Ringstr. 75
24114 Kiel

Inhaltsverzeichnis

1 EINLEITUNG	1
2 THEORETISCHER UND EMPIRISCHER HINTERGRUND	4
2.1 PSYCHOBIOLOGIE DES SCHMERZES	4
2.1.1 NEUROPHYSIOLOGIE UND BIOCHEMIE DES SCHMERZES	5
2.1.2 KOMPONENTEN DES SCHMERZERLEBNISSES	8
2.2 EXPERIMENTELLE SCHMERZFORSCHUNG	11
2.2.1 METHODEN DER SCHMERZINDUKTION	12
2.2.2 VERFAHREN DER SCHMERZMESSUNG	15
2.2.2.1 Allgemeine Überlegungen zur Problematik des Messens	16
2.2.2.2 Reizabhängige Verfahren: Schwellenmessung	18
2.2.2.3 Antwortabhängige Verfahren: verbale Skalierung	21
2.3 SCHMERZMESSUNG ALS TEST	23
2.3.1 GÜTEKRITERIEN EINES TESTS	23
2.3.2 UNTERSUCHUNGEN ZUR RELIABILITÄT VON SCHMERZMESSUNGEN	27
3 ABLEITUNG DER FRAGESTELLUNG	31
4 DARSTELLUNG DER EXPERIMENTE	31
4.1 UNTERSUCHUNGSDESIGN	31
4.2 SCHMERZINDUKTIONSMETHODEN	33
4.2.1 MESSUNG DES DRUCKSCHMERZES	33
4.2.2 MESSUNG DES ISCHÄMIESCHMERZES	35
4.2.3 MESSUNG DES HITZESCHMERZES	36
4.3 DURCHFÜHRUNG DER UNTERSUCHUNGEN	38
4.3.1 VERSUCHSPERSONEN	38
4.3.2 ABLAUF	38
5 AUSWERTUNG	40

6	ERGEBNISSE	41
<hr/>		
6.1	METHODENÜBERGREIFENDE ERGEBNISSE	41
6.1.1	ÜBERPRÜFUNG DER VORAUSSETZUNGEN	41
6.1.2	VERSUCHSLEITEREFFEKTE	41
6.1.3	VERGLEICH DER METHODEN	42
6.2	ERGEBNISSE DES DRUCKSCHMERZES	43
6.2.1	ERGEBNISSE DER SCHWELLENMESSUNG	43
6.2.1.1	Einzelne Finger	44
6.2.1.2	Gemittelte Finger	45
6.2.2	ERGEBNISSE DER TOLERANZSCHWELLENMESSUNG	46
6.2.2.1	Einzelne Finger	46
6.2.2.2	Gemittelte Finger	47
6.3	ERGEBNISSE DES ISCHÄMIESCHMERZES	49
6.4	ERGEBNISSE DES HITZESCHMERZES	49
6.4.1	ERGEBNISSE DER SCHWELLENMESSUNG	50
6.4.2	ERGEBNISSE DER DISKRIMINATION	51
6.4.2.1	Vorüberlegungen zum Skalenniveau	51
6.4.2.2	Ergebnisse der Skalierung	54
7	DISKUSSION	56
<hr/>		
7.1	DISKUSSION DER METHODENÜBERGREIFENDEN ERGEBNISSE	57
7.2	DISKUSSION DER ERGEBNISSE DES DRUCKSCHMERZES	58
7.3	DISKUSSION DER ERGEBNISSE DES ISCHÄMIESCHMERZES	60
7.4	DISKUSSION DER ERGEBNISSE DES HITZESCHMERZES	60
7.5	ZUSAMMENFASSENDE DISKUSSION	62
8	ZUSAMMENFASSUNG	66
<hr/>		
9	LITERATUR	67
<hr/>		

1. Einleitung

Schmerz ist für die meisten Menschen eine meist unangenehme, aber selbstverständliche Empfindung. Es gibt nur einige wenige Menschen, die eine angeborene Schmerzunempfindlichkeit (kongenitale Analgesie) haben. Ihnen fehlt jede bewußte **Schmerzempfindung**, das heißt sie nehmen Verbrennungen oder Verletzungen nicht als unangenehm oder eben als sehr schmerzhaft wahr.

Bei der eindeutigen Mehrheit der Leute gehören Schmerzen jedoch zum Leben. Jeder kennt Schmerzen, hat sie wahrscheinlich schon in vielfältiger Hinsicht z. B. als Zahnschmerzen oder Kopfschmerzen erlebt, kann über Qualitäten oder Quantitäten berichten, hat vielleicht Angst vor ihnen und versteht vor allem Mitmenschen, die unter „denselben“ Schmerzen leiden. Ist das wirklich so? Sind es „dieselben“ Schmerzen? Wie genau können wir das Schmerzerlebnis anderer (er)fassen?

Ein Teil dieser Fragen soll in der vorliegenden Arbeit behandelt werden, die sich mit der sich mit der **Meßgenauigkeit** von Schmerzen beschäftigt. Um Schmerzen messen zu können, muß zuerst der Kontext definiert werden, in dem die Schmerzen beobachtet werden. In dieser Arbeit geht es um die Probleme, die bei der **experimentellen** Schmerzforschung auftreten, das heißt, um das Studium von schmerzhaften Reizen, die im humanen Laborexperiment gezielt induziert werden und der Erfassung der individuellen Reaktion unter weitestgehender Kontrolle externer Einflußgrößen wie sozialen oder umweltbedingten Faktoren (Murrin & Rosen, 1985).

Der Begriff „experimentell“ legt nahe, daß es noch einen nicht-experimentellen Bezugsrahmen der Schmerzforschung gibt, in dem Schmerz beobachtet werden kann. Ich spiele hier auf den **klinischen** Kontext und eine anhaltende Diskussion an, inwieweit Ergebnisse des im Labor induzierten Schmerzes in den (klinischen) Alltag übertragbar sind, also in die Umgebung, in der das Leiden unter Schmerzen und die Schmerzbehandlung im Vordergrund stehen.

Es muß zunächst zwischen experimenteller und klinischer **Schmerzdiagnostik** unterschieden werden. Experimentelle Methoden haben den Vorteil, daß der Schmerzreiz systematisch veränderbar ist, und somit Stimulus und Reiz kontrollierbar werden. Im Laborversuch ist der für die Schmerzdiagnostik wichtigste Bestandteil die **Schmerzmessung**, bei der es um das Studium der Zusammenhänge zwischen einem quantifizierbaren noxischen Reiz und dem Schmerzerleben geht. Im Gegensatz dazu geht der klinische Schmerzdiagnostik über den Ansatz der Schmerzmessung hinaus. In die klinische Diagnostik fließt zudem das psychosoziale Umfeld des Patienten ebenso mit ein, wie die Erfassung des Schmerzgeschehens auf subjektiv-verbaler, motorischer und (neuro)physiologischer Ebene (Flor, 1991).

Bereits Beecher (1959) weist in der Diskussion der experimentellen und klinischen Schmerzforschung darauf hin, daß man zwischen der bloßen Wahrnehmung eines Schmerzreizes und den „anderen Komponenten“ unterscheiden müsse, die von höheren kognitiven Funktionen ausgingen. Als Mediziner meint er, die bis dato verwendeten experimentellen Methoden hätten keine klinische Relevanz.

Gracely (1985, S. 203) sieht die eingeschränkte Gültigkeit experimenteller Methoden, die nur die Schmerzwahrnehmung untersuchen. Er sagt:

„Unlike other modalities, pain is more than a sensation. It is also an alarm signal that quickly commands attention. It produces feelings of distress, aversion and fear, and motivates escape and avoidance.“

Im klinischen Kontext sind in Schmerzen erster Linie **chronisch**, das heißt, das die Schmerzen länger als vier bis sechs Monate anhalten und nicht die Funktion eines „schnellen Alarmsignals“ haben. Die chronischen Schmerzen sind oft schlecht zu lokalisieren und zu behandeln. Begleitende Gefühle können Hilflosigkeit, depressive Verstimmungen, allgemeine Reizbarkeit und Kontrollverlust sein (Flor, 1991). Im Gegensatz dazu ist der im Labor erzeugte Schmerz im allgemeinen ein **akuter** Schmerz, der auf einen bestimmten Ort begrenzt ist. Die Intensität des Schmerzreizes hängt direkt mit der Schmerzempfindung zusammen, die nach Beseitigung des Schmerzreizes schnell wieder abklingt. Akute Schmerzen haben im Gegensatz zu chronischen eine Signal-, oder Warnwirkung, indem sie auf eine Gewebsschädigung hinweisen. Sie können aber auch eine Blutdruckerhöhung und Angstgefühle auslösen (Basler & Rehfish, 1990).

Ostkirchen (1994) stellt fest, daß im Schmerzgeschehen behaviorale, physiologische, emotionale, kognitive und motivationale Parameter in komplexer Weise zusammenhängen. Sie vertritt daher die Ansicht, daß eine systemische-dynamische Betrachtungsweise notwendig ist, um dem **Phänomen Schmerz** gerecht zu werden. Die Forscher sollten sich hinterfragen, ob physiologische Methodik nicht eher ihre persönliche Präferenz, als den Weg an sich, kennzeichnet.

Zusammenfassend wird die experimentellen Schmerzforschung kritisiert, weil sie Schmerzen erzeugt, die mit den im klinischen Alltag vorherrschenden Schmerzen nicht vergleichbar sind und weil im Labor nur ein minimaler Teil des wesentlich komplexeren Schmerzerlebens betrachtet wird. Diese Kritik führt zu der Frage, welchen Beitrag die experimentelle Schmerzforschung zum Schmerzverständnis leisten kann?

Gracely (1985) begründet den Nutzen experimenteller Forschung, indem er verschiedene Ziele nennt, die im Laborversuch erreicht werden. Zum einen konnte durch systematische Studien die Effizienz von Analgetika (Schmerzmitteln) verbessert werden, zum anderen spielt experimentell induzierter Schmerz eine wesentliche Rolle bei der Erforschung der peripheren und zentralnervösen Schmerzmechanismen, des endogenen Schmerzverarbeitungssystems und bei der Erforschung der Rolle von sensorischen und affektiven Prozessen der Schmerzwahrnehmung. Weitere Vorteile bestehen in der möglichen Analyse der Schmerzáußerungen und den sie beeinflussenden Faktoren. Auch können Reizmethoden, die dem klinischen Schmerzbild ähnlich, entwickelt werden. Schließlich ist es möglich, fundierte experimentelle Methoden mit der im Alltag vorherrschenden Schmerzempfindung zu verbinden, indem beispielsweise der alltägliche Schmerz und experimentell erzeugter Schmerz auf derselben Skala geschätzt und miteinander verglichen werden können.

Meiner Meinung nach fordert gerade das Phänomen Schmerz zu einer vielseitigen, interdisziplinären Sicht heraus. Das heißt, daß es Forschern eines bestimmten Fachgebietes zwar nicht möglich sein wird, alle Teile des Schmerzgeschehens gleich gut zu betrachten, sondern daß diese Forscher einen Bereich des Schmerzerlebens

genauer studieren müssen, um dann unter Berücksichtigung der in anderen Fachbereichen gewonnenen Erkenntnisse, praxisbezogene Untersuchungen zu entwickeln und anwendungsbezogene Aussagen treffen zu können. Denn die Schmerz Betrachtung ohne die Kenntnis physiologisch-biochemischer Abläufe ist genauso Stückwerk, wie die Annahme, durch immer exaktere Kenntnis der Mikroprozesse menschliche Sinnzusammenhänge vorhersagen zu können.

In dieser Arbeit geht es um die Ermittlung der **Reliabilität** (Meßgenauigkeit) von experimentellen Schmerzmessungen, die an zwei Untersuchungen überprüft werden soll. Denn ohne die Kenntnis der Güte der gewählten Verfahren ist es sowohl für den experimentellen als auch für den klinischen Schmerz nicht möglich, begründete Aussagen über vorgefundene Effekte, wie z. B. Schmerzwahrnehmungsänderungen oder Therapieerfolge, zu treffen, da diese Effekte ebenso gut von der Ungenauigkeit der Methodik verursacht worden sein könnten. Die Prüfung der eingesetzten Methodik ist somit notwendig, *bevor* Aussagen über mögliche Wirkungen getroffen werden.

Meine Arbeit ist folgendermaßen aufgebaut: im ersten Teil des *theoretischen Hintergrundes* referiere ich die **psychobiologischen Grundlagen** des Schmerzes.

Im zweiten Teil stelle ich verschiedene **Schmerzinduktionsmethoden** und Verfahren der **Reaktionsmessung** vor, die in der experimentellen Schmerzforschung angewendet werden und gehe allgemein auf die Problematik des „Messens“ ein.

Im dritten Teil geht es um die **testtheoretischen Anforderungen** an das Schmerzmeßinstrument. Verschiedene Studien, die diese Anforderungen prüften, werden exemplarisch dargestellt.

Im *experimentellen* Teil der Arbeit werden zwei Untersuchungen vorgestellt, die erkunden, wie die Schmerzwahrnehmung durch verschiedene Neuropeptide beeinflusst wird. Auf diese Fragestellung gehe ich in meiner Arbeit nicht ein, sondern ich befaße mich mit der Meßgenauigkeit der Verfahren, die in diesen Experimenten gewählt wurden.

Nach der **Ermittlung der Reliabilitäten** und der Ergebnisdarstellung folgt die Diskussion der verschiedenen Fragestellungen, die mit der Bestimmung und Verbesserung der Meßgenauigkeit einhergehen.

2 Theoretischer und empirischer Hintergrund

2.1 Psychobiologie des Schmerzes

2.1.1 Neurophysiologie und Biochemie des Schmerzes

Nach der Spezifitätstheorie gilt **Schmerz** als eine eigenständige, spezifische Sinnesmodalität mit einem für die Aufnahme von gewebeschädigenden Reizen („Noxen“, lat. *noxa* = Schaden, Strafe), deren Weiterleitung und zentralnervöse Verarbeitung eigenständigen nozizeptiven System. **Nozizeption** bezeichnet dabei die Aufnahme, Weiterleitung und Verarbeitung noxischer Reize. Die Nervenfasern (Axon und umgebende Schwann-Zelle) der Sensoren werden Afferenzen genannt. Sensoren, die durch noxische Reize erregt werden heißen **Nozizeptoren**. Sie sind die zahlenmäßig größte Rezeptorengruppe des somatosensorischen Systems.

Es lassen sich Nozizeptoren mit dünn markhaltigen Nervenfasern (A -Fasern, Gruppe III) von Nozizeptoren mit marklosen Nervenfasern (C-Fasern, Gruppe IV) unterscheiden. Die **A -Fasern** leiten die Erregung zwischen 2,5 m/s und 20m/s weiter. Aufgrund ihrer Funktion können A -Faser Nozizeptoren in zwei Gruppen eingeteilt werden. Die Nozizeptoren der ersten Gruppe reagieren vor allem auf intensive mechanische Reize und haben große, sich überlappende rezeptive Felder (Hochschwellige Mechanorezeptoren), die Nozizeptoren der zweiten Gruppe reagieren auf thermische und mechanische Stimuli, und besitzen kleinere rezeptive Felder (Mechanothermale Rezeptoren). Price (1998) erwähnt noch andere Rezeptoren, die selektiv auf thermische Reize reagieren (Thermosensitive Rezeptoren).

Die **C-Fasern** bilden ca. 90% der Nozizeptoren. Sie sind bedeutend langsamer, da sie die Erregung mit < 2,5 m/s weiterleiten. Sie besitzen kleine, punktförmige rezeptive Felder und reagieren **polymodal**, d. h. daß qualitativ unterschiedliche Noxen (thermische, chemische oder mechanische Reize) eine Erregung hervorrufen können.

Die Erregung erfolgt entweder durch *direkte* Reizung der Nozizeptoren, z. B. bei mechanischer Gewalteinwirkung oder bezogen auf Hitzereize bei Temperaturen ab 45 °C, oder *indirekt* über eine Vielzahl von zellulären Reaktionen, an deren Ende biochemische Stoffe (Entzündungsmediatoren) freigesetzt werden, die die Nozizeptoren entweder erregen oder deren Erregungsschwelle reduzieren. Dieser Prozeß wird **Sensibilisierung** genannt. Dadurch vermögen dann auch ursprünglich nicht noxische Reize eine Erregung hervorzurufen.

Die Nozizeptoren sind in fast allen Organen (Haut, Viscera), in der Skelettmuskulatur und auch in den Gelenken vorhanden. Histologisch betrachtet handelt es sich um freie Nervenendigungen, das heißt, daß ihre Endungen nicht zusätzlich von anderen Strukturen umgeben werden. Ihre Verteilung im Gewebe

ist unregelmäßig, so daß beispielsweise an verschiedenen Stellen der Haut unterschiedlich viele Nozizeptoren liegen.

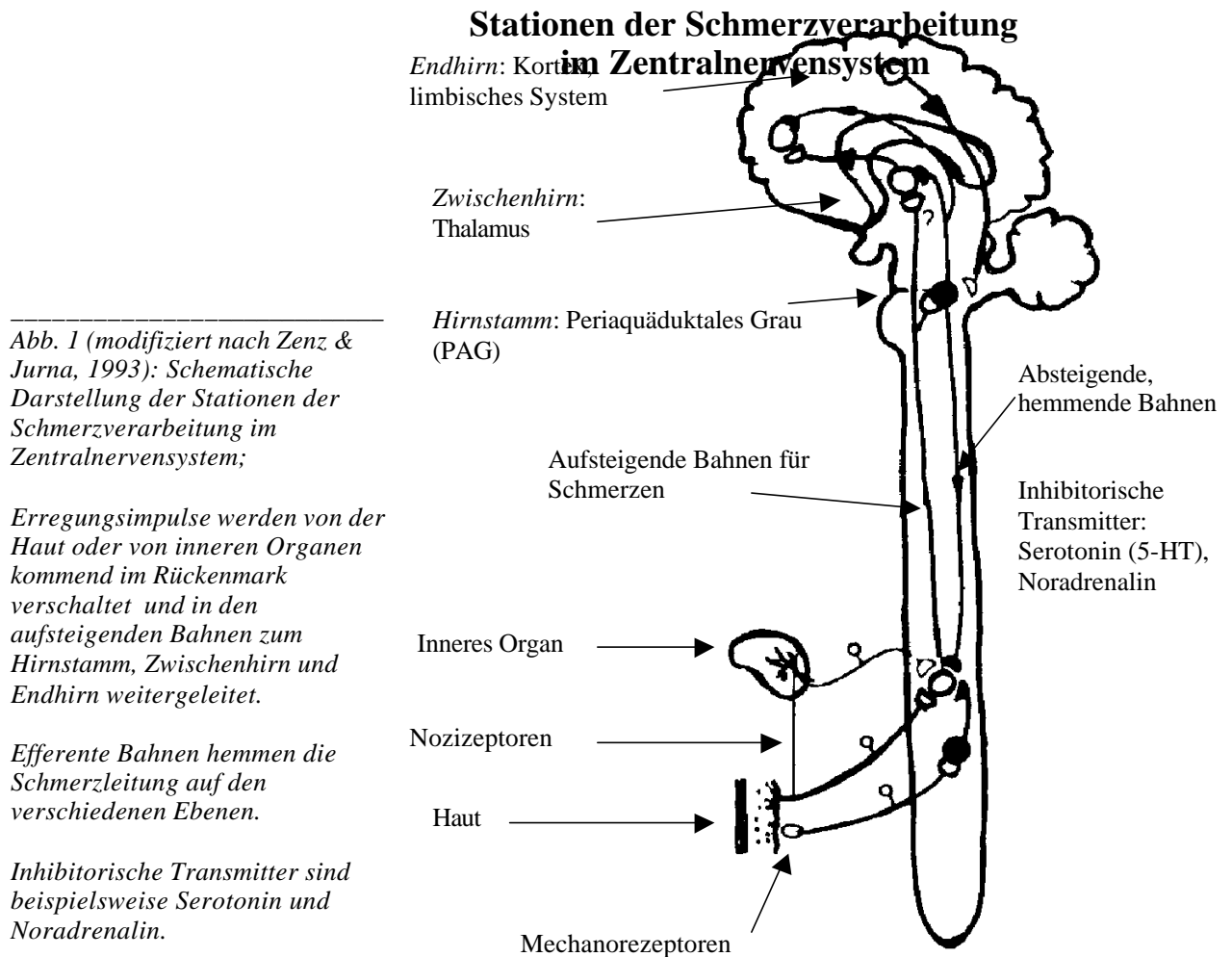
Der von der Haut(oberfläche) ausgehende Schmerz wird als Oberflächenschmerz bezeichnet. Er wird in erster Linie durch die A -Fasern vermittelt. Bei einer Verletzung setzt sofort der **erste Schmerz** ein, der als stechend oder hell beschrieben wird. Der ca. 0.5 - 1 sec. später einsetzende **zweite Schmerz** wird im Gegensatz dazu von den cutanen C-Fasern weitergeleitet. Er wird als dumpf charakterisiert. Die Nozizeptoren der Skelettmuskulatur und der Gelenke scheinen vorwiegend C-Fasern zu sein. Der von diesen tiefgelegenen Strukturen ausgehende Schmerz ist dem zweiten Schmerz ähnlich. Er wird ebenfalls als dumpf beschrieben.

Wenn eine bestimmte Schmerzinduktionsmethode ausgewählt wird, werden die verschiedenen Nozizeptoren in unterschiedlichem Ausmaß erregt. Price (1988) berichtet, daß durch ischämische Reizung bei gleichzeitiger Muskelkontraktion die tiefliegenden Nozizeptoren der Skelettmuskulatur (C-Fasern) besonders leicht erregt werden. Mechanische Reizung durch Druck erregt zuerst die hochschwelligen Mechanorezeptoren und die mechanothermale Rezeptoren, also in erster Linie A - Fasern. Aber auch die cutanen polymodalen C-Fasern können gereizt werden, mit dem Unterschied, daß die C-Fasern später reagieren und daß sich ihre Aktivität erst langsam über die Zeit summiert. Für den durch Hitze induzierten Schmerz ist gut belegt, daß der erste, akute Schmerz durch A -Fasern und der zweite Schmerz durch C-Fasern übermittelt wird. Bei der Messung des Schmerzes sollte man auch darauf achten, welchen Schmerz die Probanden einstufen, denn bereits ein kurzer Hitzereiz von 51 °C kann sowohl die Wahrnehmung des ersten Schmerzes als auch die Wahrnehmung des zweiten Schmerzes hervorrufen. Eine wiederholte Darbietung eines Schmerzreizes hat dann bereits einen paradoxen Effekt zur Folge: durch die **zeitliche Summation** sinkt die Wahrnehmungshäufigkeit und Intensitätseinschätzung des ersten Schmerzes, während die des zweiten Schmerzes steigt (Price, Hu, Dubner & Gracely, 1977; Price, McHaffie & Stein, 1992). Die Autoren sehen darin einen Indikator, daß dem ersten und dem zweiten Schmerz getrennte Schmerzmechanismen zugrunde liegen. Auch für Druckschmerz wurden zeitliche Summationseffekte belegt (Kohllöffel, Koltzenburg & Handwerker, 1991). Die unterschiedliche neurophysiologische Verarbeitung kann somit ursächlich dafür sein, daß verschiedene Induktionsmethoden verschiedene Schmerzcharakteristika hervorrufen (s. Punkt 2.2.1).

Eine andere Frage ist, welche Rolle die **räumliche Summation** bei der Auswahl von Schmerzinduktionsverfahren spielt. So könnten je nach Temperatur, Ort und Größe einer Thermode einmal mehr und einmal wenige Nozizeptoren erregt werden. Price et al. (1992) untersuchten die räumliche Summation für Temperaturen von 45 °C bis 51 °C mit Thermoden der Größen 1 bis 3 cm². Sie fanden heraus, daß die räumliche Summation zunahm, je größer die Auflagefläche und je höher die Temperatur war. In einer Replikationsstudie untersuchten Douglass, Carstens und Watkins (1992) feinere Abstufungen der Thermodengröße und die Frage, inwieweit die Platzierung der Thermoden innerhalb eines Dermatoms oder zwischen zwei überlappenden Dermatomen einen Einfluß hat. Auch diese Studie belegt, daß die Summation insbesondere bei intensiven Reizen einen Einfluß auf die

Schmerzwahrnehmung hat, allerdings unabhängig davon, in welchem Dermatom die Reize appliziert wurden.

Die **Weiterleitung** der Erregungsimpulse erfolgt nach ihrer Transformation zunächst zu den Neuronen des Hinterhorns im Rückenmark (substantia gelatinosa). Die Fasern divergieren dabei auf verschiedene Rückenmarksneurone, die ebenfalls konvergente Impulse aus den Afferenzen eines Organs, aber auch aus verschiedenen Organen erhalten (s. Abb. 1).



Die Information wird dann zu motorischen oder sympathischen Reflexen verarbeitet und/oder auf Interneurone im Tractus spinothalamicus umgeschaltet, der im Vorderseitenstrang verläuft. Über den Tractus neo-spinothalamicus laufen die Impulse zum ventro-basalen Thalamus und weiter über den Tractus thalamo-corticalis zum Gyrus postcentralis. Über diesen Weg wird die Schmerzlokalisierung vermittelt. Andere Impulse gelangen über den Tractus spinoreticularis zur Formatio reticularis, wodurch vegetative Komponenten der Schmerzreaktion (z. B. Atmung) vermittelt werden. Hier im Hirnstamm wird zudem das aufsteigende retikuläre aktivierende System (ARAS) beeinflusst, und somit die allgemeine Erregbarkeit des Cortex. Im tractus palaeo-spinothalamicus verläuft die Erregung zu den intralaminären Thalamuskernen, von wo aus Bahnen zu kortikalen Arealen und zum Limbischen System ziehen. Die

Aktivität im limbischen System wird mit der affektiven-emotionalen Schmerzkomponente der Schmerzwahrnehmung in Verbindung gebracht, der Cortex soll für die kognitive Komponente, sprich für die Bewertung, die bewußte Erkennung des Schmerzes oder für zielgerichtete Handlungen, verantwortlich sein.

Die nozizeptive Verarbeitung wird auf verschiedenen Ebenen vom Rückenmark bis zum Großhirn durch **Schmerzhemmsysteme** moduliert. Wichtige Bahnen entspringen in dem Hirnstamm (periaquäduktales Grau (PAG), Nucleus raphé magnus, Locus coeruleus) und ziehen entweder im Tractus spinothalamicus zum Rückenmark oder projizieren zum Großhirn. Diese Synapsen ermöglichen also die *efferente* Beeinflussung der Schmerzverarbeitung. Auch durch eine *afferente* Reizung anderer Neurone kann eine Schmerzhemmung ausgelöst werden. So wird bei der Stimulation von A -Fasern (z.B. des Tastsinns) die Aktivität der nozizeptiven Neurone am selben Rückenmarkssegment reduziert.

Zahlreiche (**bio-)**chemische **Substanzen** sind mittlerweile identifiziert worden, die bei dem Schmerzgeschehen zentrale oder modulierende Funktionen haben. Bei der Sensibilisierung der Schmerzrezeptoren wirken vor allem Bradykinin und Prostaglandin E (u.a. von Makrophagen synthetisiert!) erregend, Acetylsalicylsäure und Corticosteroide dämpfen dagegen die Sensibilisierung.

Die wichtigsten Substanzen, die bei der Schmerzweiterleitung auf spinaler Ebene eine exzitatorische Funktion haben, sind die Aminosäure L-Glutamat, CCK (Cholecystokinin) und das Neuropeptid Substanz P. Die Gegenspieler des Hemmsystems sind auf segmentaler Ebene Gamma-Amino-Buttersäure (GABA) und von supraspinal kommend Serotonin (5-Hydroxytryptamin), Noradrenalin und Dopamin.

Die körpereigenen Liganden (Enkephaline, Endorphine, Dynorphine) sind weitere zentral wirkende Vertreter des endogenen Schmerzkontrollsystems. Sie werden bei Streßreaktionen vermehrt ausgeschüttet und verringern gewöhnlich die Sensibilität für schmerzhafte Reize.

Interessant ist die Frage, inwieweit die verschiedenen Formen der Schmerzinduktion mehr oder weniger Streß erzeugen, und ob der Streß die Erfassung der Parameter beeinflusst.

Rezeptoren für die körpereigenen Opiate befinden sich entlang der afferenten nozizeptiven Bahnen, in den Hinterhornwurzeln des Rückenmarks, im PAG und medialen Thalamus, aber nicht im Kortex. Das hat zur Folge, daß Opiate (z.B. Morphine) nicht die Schmerzlokalisierung, also den ersten stechenden Schmerz beeinflussen, sondern nur das Schmerzerleben, also den als dumpf empfundenen tiefen Schmerz. Das heißt, daß bei einer schmerzinduzierten Streßreaktion das Schmerzhemmsystem und das Endorphinsystem (parallel dazu die erhöhte Ausschüttung von ACTH und Corticosteroiden) aktiviert werden können und so durch Streß die Erfassung des intensiven, als unangenehm empfundenen Schmerzes moduliert werden kann.

Die Erregungswahrscheinlichkeit eines Nozizeptors ist somit abhängig von der Qualität und Quantität der Noxe, dem Wirkungsort und dem Zustand des Rezeptors. Durch räumliche und zeitliche Summation können die Reizeffekte verstärkt werden. Die ungehinderte Erregungsausbreitung wird durch Schmerzhemmsysteme verhindert. Zahlreiche biochemische Substanzen wirken dabei modifizierend auf die Schmerzverarbeitung ein. Psychophysiologische Reaktionen, bei denen diese biochemischen Substanzen ausgeschüttet werden, können auf diesem Wege die Schmerzwahrnehmung beeinflussen. In diesem Sinn ist es wichtig im Experiment die Hormone, die das Schmerzempfinden modulieren, zu erfassen, um eine Einflußnahme auf die Schmerzmessung erkennen.

2.1.2 Komponenten des Schmerzerlebnisses

Was ist Schmerz?

Um auf diese Frage eine übereinstimmende Antwort zu finden, gründete sich im Jahre 1974 die „International Association for the study of Pain (IASP)“.

Kern der Definition des Begriffes „**Schmerz**“, die die IASP (1979) formulierte, ist das subjektive Erleben. Ihre Definition lautet wie folgt:

”PAIN: An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.“

Zur Schmerzwahrnehmung gehört die subjektive Erfahrung durch das Individuum. Einerseits folgen einer Gewebsschädigung nicht notwendigerweise die Schmerzen, die bei einer solchen Verletzung zu erwarten wären, andererseits können Schmerzen auch ohne die Erregung von Schmerzrezeptoren auftreten. So genügt für das Vorhandensein von „Schmerz“, wenn das Befinden mit Worten umschrieben wird, als wenn eine Gewebläsion vorläge. Ob eine Schädigung tatsächlich stattfindet ist per definitionem nicht notwendig. Zudem geht die Schmerzempfindung über eine reine Sinneswahrnehmung hinaus: sie beinhaltet gleichzeitig auch immer etwas **Affektives**, in der Regel etwas Unangenehmes.

Der Entstehung dieser Definition ging eine kontroverse Diskussion voraus (Göbel, 1988), sich auf eine gemeinsame Fassung festzulegen. So setzen die Forscher jeweils unterschiedliche Schwerpunkte: entweder stand der Gewebeschaden als ursächlicher Reiz im Vordergrund oder die Gesamtreaktion, die vom Schmerz ausgelöst wird oder aber das durch die Schmerzempfindung ausgelöste Leiden.

Die gefundene Definition beschreibt in meinen Augen den wesentlichen, gemeinsamen kleinsten Nenner der unterschiedlichen Betrachtungsmöglichkeiten. Von ihr bleiben allerdings viele Komponenten unberücksichtigt, die zur umfassenden Beschreibung und Erfassung des Schmerzgeschehens wichtig sind.

So sind Schmerzen auch stets mit **motorischen** (Flucht-, Schutzreflexe, Muskelverspannungen) und **psychomotorischen** Reaktionen (Mimik, Klagen, Weinen) und **vegetativen** Begleiterscheinungen (Schweißausbruch, Übelkeit, Änderung des Blutdrucks, der Herzfrequenz und der Atmung) verbunden. Besonders durch die motorischen und psychomotorischen Reaktionen wird die Interaktion mit der Umwelt determiniert. Je nach Kultursystem lassen sich dabei differente Funktionen der Schmerzäußerung unterscheiden. Für die Italiener hat Schmerz keine Warnfunktion, sondern ist eine Krankheit per se, in Folge der gejamert, geschrien und kräftig geweint wird. Für die Eipo in Neuguinea ist Schmerz, der z. B. durch Verletzungen entstand, völlig bedeutungslos, und hat gar keine emotionale Reaktion zur Folge (Engel, 1990).

Jede Schmerzreaktion ist Ausdruck des **sozialen Kontextes** in dem sie stattfindet. Ein Problem, das sich bei der Schmerzmessung ergeben kann, ist die Interaktion zwischen den Versuchsleitern und den Versuchspersonen. So mag das lange Aushalten von Schmerzen („der Indianer kennt keinen Schmerz“) eine soziale Funktion erfüllen, indem Stärke oder Belastbarkeit vermittelt wird. Levine und Simone (1991) untersuchten die Antwortneigung von 35 männlichen und 33 weiblichen Studenten, die nach Zufall einem männlichen oder einem weiblichen Versuchsleiter zugeteilt wurden. Die Versuchsleiter waren instruiert, sich attraktiv zu kleiden und entweder sehr maskulin oder sehr feminin zu wirken. In diesem Experiment ergab sich für die Männer eine signifikante und für die Frauen eine tendentielle Interaktion zwischen Versuchsleiter und Geschlecht. So zeigen die männlichen Studenten gegenüber einem gegengeschlechtlichen Versuchsleiter weniger Schmerzen, die weiblichen Studenten hingegen äußern bei einem männlichen Versuchsleiter mehr Schmerzen. Bei dem gleichgeschlechtlichen Versuchsleiter unterschieden sich die Studenten nicht. Die Autoren meinen, daß sich hier die sozialen Rollenklischees eines auf der einen Seite starken Machos und einer verletzlichen, zu beschützenden Frau auf der anderen Seite widerspiegeln.

Sicherlich hat die soziale Interaktion in der Versuchssituation einen Einfluß auf die zu erhebenden Daten. Daß allerdings Versuchsleiter betont geschlechtsspezifisch agieren, wird bei Routineuntersuchungen nicht die Regel sein. Die Frage, ob sich „natürlich“ verhaltende Versuchsleiter, nur aufgrund ihres Geschlechts derart Einfluß nehmen, ist noch nicht beantwortet. Es muß die Frage gestellt werden, ob das betont geschlechtsspezifische Verhalten nicht auch mit Machtausübung konfundiert ist und ob die geschlechtsspezifische Attraktivität überhaupt die entscheidende Variable darstellt. Denkbar wäre auch, daß nicht nur verschieden geschlechtliche Versuchsleiter sondern auch zwei Frauen beispielsweise von Probanden als unterschiedlich attraktiv oder dominant empfunden werden.

Schmerzen werden zudem nicht als ausschließlich gegenwärtige Erscheinung wahrgenommen, sondern im Kontext der eigenen Lebensgeschichte betrachtet bzw. bewertet. Dieser **bewertende** Aspekt ermöglicht es den Schmerz als geringfügig oder aber - prognostisch - als lebensbedrohlich einzuordnen. Inwieweit diese Einordnung geschieht, ist neben den gegenwärtigen Umständen auch beispielsweise von der (religiösen) Erziehung abhängig. So benötigen Verwundete im Krieg weitaus weniger schmerzstillende Medikamente als bei gleichartigen Verletzungen im Zivilleben gefordert werden. Juden äußern ihren Schmerz auffallend dramatisierend, wobei dieses Verhalten nicht primär durch die gegenwärtige Situation ausgelöst wird, sondern durch „die Sorge um zukünftige Auswirkungen“ (Engel, 1990, S.76).

Neben den allgemeinen Lebenssituationen, die auf die Einschätzung des Schmerzes Einfluß nehmen, gibt es auch im Alltag unterschiedliche **Tageszeiten**, zu denen der Mensch mehr oder weniger schmerzempfindlich ist. So fanden Procacci, Corte, Zoppi und Mresca (1974) bei jungen Männern und Frauen einen circadianen Rhythmus für Hitzeschmerz, der mittels Lichtbestrahlung auf der Haut des Unterarms erzeugt wurde, wobei sie vier Messungen im fünfständigem Abstand bei normaler Tagesroutine der Probanden vornahmen. Bei den Frauen zeigte sich der Rhythmus nur dann, wenn sie am selben Tag ihres Menstruationszykluses untersucht wurden. Leider wurde in dieser Studie nicht kontrolliert, ob im alltäglichen Tagesablauf Faktoren mit der Schmerzempfindlichkeit kovariieren. So wird vermutet, daß sich die Schmerzempfindung bereits mit der Nahrungsaufnahme und den damit einhergehenden hormonellen Reaktionen verändert (Albus, 1988).

Die Studie von Procacci et al. (1974) belegt zudem, daß es innerhalb der Frauengruppe und zwischen den Männern und Frauen **Geschlechtsunterschiede** bei der Schmerzwahrnehmung gibt. Auch Goolkasian (1985) zeigt, daß die Diskriminationsfähigkeit von Frauen von dem Zeitpunkt ihrer menstrualen Phase und auch von der Einnahme von Kontrazeptiva abhängig ist. Bei dem Experiment von Levine und Simone (1991) gab es unabhängig vom Geschlecht des Versuchsleiters einen signifikanten Geschlechtsunterschied beim cold-pressure-test. Die weiblichen Versuchspersonen berichteten intensivere Schmerzen als ihre männlichen Kommilitonen.

Bei anderen Untersuchungen fanden sich keine Geschlechtsunterschiede. Lautenbacher und Rollman (1993) differenzierten die Fragestellung, indem sie die Wahrnehmung schmerzhafter und nicht schmerzhafter Reize bei unterschiedlichen Schmerzinduktionsmethoden (Kälte, Hitze, elektrische Stimulation) von 20 jungen Männern und Frauen verglichen. Sie fanden nur bei der elektrischen Stimulation Geschlechtsunterschiede, nicht wie erwartet bei der Hitzereizung. Eine mögliche Erklärung wäre, daß die Größe der gewählten thermische Elektrode zu schnell zu einer räumlichen Summation führt, und Unterschiede eher durch kleinere Thermoden entdeckbar wären. Ohrbach und Gale (1989b) konnten ebenfalls keine Geschlechtsunterschiede für verschiedene Experimente mit einem Druckschmerzgerät bei klinischen und nicht-klinischen Versuchspersonen feststellen.

Anzunehmen ist, daß abhängig von der gewählten Methode Geschlechtsunterschiede in der Schmerzwahrnehmung bestehen. Solange nicht das Gegenteil für eine bestimmte Technik bewiesen wurde, sollte man diesen Faktor in jedem Fall experimentell kontrollieren.

Auf individueller Ebene wurde untersucht, welche **Persönlichkeitseigenschaften** mit der Schmerzempfindlichkeit kovariieren. In einer Übersicht weist Sternbach (1978) auf den Einfluß von Ängstlichkeit oder Depression auf die Schmerzwahrnehmung hin. So haben hochängstliche (im Sinne des trait oder state Angstkonzepts) Personen anscheinend eine niedrigere Schmerztoleranzschwelle und reagieren sensibler auf einen Schmerzreiz, während für Depressive keine einheitliche Aussage zu treffen ist. So gibt es Befunde, daß mit der Schmerzlinderung die Depression einerseits abnimmt, aber andererseits auch zunehmen kann, wenn die meist chronischen Schmerzen zurückgehen. Im Laborexperiment wird vor allem die Angst vor dem Schmerzreiz, seine Antizipation, im Vordergrund stehen. Um diese „state-Angst“ zu lindern, sollte man die Versuchspersonen mit der Versuchsanordnung vertraut machen. Hochängstliche im Sinne des „trait-Konzeptssollten sollten möglicherweise vom

Experiment ausgeschlossen werden. Abb. 2 faßt noch einmal alle Schmerzkomponenten zusammen, die am Schmerzgeschehen beteiligt sind (Göbel, 1988).

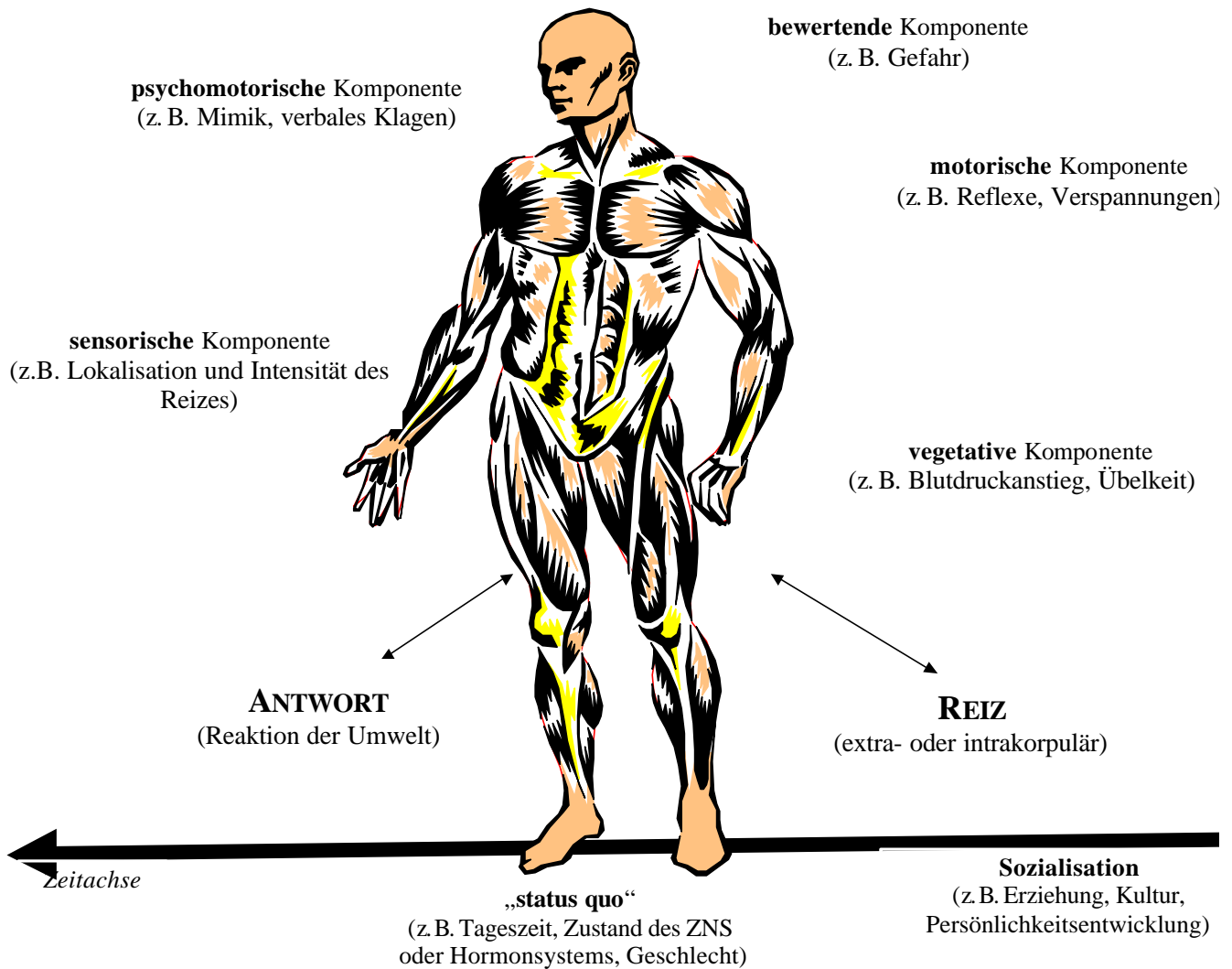


Abb. 2: Darstellung der am Schmerzgeschehen beteiligten Komponenten, die mitbestimmen, wie der Schmerzreiz vom Individuum wahrgenommen wird. Die Reaktion der Umwelt beeinflusst wechselseitig das Schmerzerlebnis. Schmerzgeschehen ist somit ein hochkomplexer, systemübergreifender Prozeß, an dem sensorisch-wahrnehmende, affektiv-emotionale und kognitiv-bewertende, vegetative, motorische und psychomotorische Elemente wechselseitig eingebunden sind, wie auch der kulturelle Einfluß, die Erziehung und der gegenwärtige Zustand des Individuums.

2.2 Experimentelle Schmerzforschung

In der experimentellen Schmerzforschung geht es darum, mit kontrollierten Reizen Schmerz zu erzeugen und die veränderte individuelle Schmerzreaktion zu erfassen. Dies geschieht unter weitestgehendem Ausschluß externer Einflußgrößen, die mit den zu beobachtenden Variablen kovariieren könnten. Chapman, Casey, Dubner, Foley, Gracely und Reading (1985) stellen in ihrem Übersichtsreferat die gängigen Ansätze der Schmerzmessung gegenüber: korrelative physiologische Untersuchungen am

Menschen (z. B. EEG Ableitungen), klinische Schmerzbeobachtungen, tierexperimentelle Untersuchungen und Humanversuche im Schmerzlabor. In dieser Arbeit geht es ausschließlich um den letzteren Ansatz, für die anderen verweise ich auf die entsprechende Literaturangabe.

2.2.1 Methoden der Schmerzinduktion

Im Laufe der Geschichte sind im Labor verschiedene Methoden der Schmerzauslösung erprobt worden, die die Frage beantworten sollten, ob dieselbe Person auf verschiedene Schmerzstimuli gleich reagiert. Dadurch wäre zum einen eine hohe Effizienz durch Ersetzung eines Verfahrens durch ein anderes möglich und zum anderen wäre durch die Annahme einer Schmerzpersönlichkeit die Schmerzreaktion einer Person vorhersagbar (Janal, Glusman, Kuhl & Clark, 1994). Andere Studien verfolgten das Ziel, durch unterschiedliche Reizverfahren den *Klinischen Schmerz* im Labor nachzustellen, um so zu gesicherten Aussagen über die Wirksamkeit von Analgetika zu gelangen (Beecher, 1959; Gracely, 1989).

Mittlerweile existiert ein breites Spektrum von einsetzbaren Schmerzreizen (Gracely, 1989; Main & Waddell, 1989; Prestele, 1969; Wolff, 1986; Zenz & Jurna, 1993), die ich aus unterschiedlichen Quellen in folgender Tabelle zusammengefaßt habe:

Tab. 1: Übersicht über die verschiedenen Schmerzinduktionsmethoden; Einander gegenübergestellt sind der methodische Vorgang, die Schmerzcharakteristik und kritische Anmerkungen.

Reizmethode	Methodisches Vorgehen	Schmerzcharakteristik	Kritische Anmerkungen
Elektrische Reize	<ul style="list-style-type: none"> • Stromfluß durch die Haut: Applikation am Finger, der Hand oder am Unterarm • Elektrische Reizung der Zähne oder direkt der peripheren und zentralen Neurone • Reizung von tiefgelegenen somat-ischen und viszeralen Strukturen 	<ul style="list-style-type: none"> • Unnatürlicher, einzigartiger Reiz, der mehr unangenehm als schmerzhaft empfunden wird, deswegen begrenztes Schmerzmodell. 	<ul style="list-style-type: none"> • Die Intensität und die Dauer des Stromflusses sind einfach zu kontrollieren, dadurch hohe Reliabilität aber wenig valide. • Gefahr der elektrischen Überlagerung, da durch den Stromfluß alle Afferenzen erregt werden. • Durch intracutane Methoden oder Reizung der Zahnpulpa können Nozizeptoren selektiv stimuliert werden.
Mechanische	<ul style="list-style-type: none"> • Nadelstiche, Druck auf die 	<ul style="list-style-type: none"> • Es wird ein breites 	<ul style="list-style-type: none"> • Die Reizkontrolle kann je

Reize	<p>Haut mit stumpfen Nadeln oder Borsten, Quetschen von Hautfalten, Dehnung von Hohlorganen</p> <ul style="list-style-type: none"> • Druck-Algorithmeter <p>(Die Druck-Algometrie hat sich zur <i>Entdeckung</i> von myofascialen Schmerzauslöschungspunkten bewährt.)</p>	Spektrum von Schmerzintensitäten unterschiedlicher Dauer erzeugt.	<p>nach Gewebeeigenschaften, Reizort und Dauer schwierig sein, da niederschwellige mechanische Afferenzen miterregt werden.</p> <ul style="list-style-type: none"> • Der Druck-Algorithmeter scheint eher unreliabel und die Applikation schwer kontrollierbar zu sein, es liegen aber auch gegenteilige Befunde vor (Fischer, 1987).
--------------	---	---	--

Reizmethode	Methodisches Vorgehen	Schmerzcharakteristik	Kritische Anmerkungen
Thermische Reize	<p>Hitzereize</p> <ul style="list-style-type: none"> • Kontakt mit heißem Wasser, einer Kontaktthermode, durch Strahlungshitze (radiant heat puls) oder Laserlicht <p>Kältereize</p> <ul style="list-style-type: none"> • <i>cold-pressure-test</i> <p>(Es werden Schmerzreize durch das Eintauchen der Hand oder anderer Gliedmaßen in eiskaltes Wasser appliziert.)</p>	<ul style="list-style-type: none"> • Bei kurzer lokaler Anwendung entsteht ein scharfer, heller Oberflächenschmerz. • Es entsteht ein kräftiger Schmerz, der schnell ansteigt und nicht lange toleriert wird. 	<ul style="list-style-type: none"> • Die Reizstärke ist gut kontrollierbar. • Bei den Kontaktmethoden können niederschwellige Thermorezeptoren miterregt werden • Bei der Lasermethode kann es durch die schnelle Erhitzung des Gewebes zu Verbrennungen kommen. • Die Methode ist wenig reliabel, bei allerdings guter Validität. • Die Wiederherstellung der Blutzirkulation nach dem Eintauchen benötigt einige Zeit, wodurch eine Mehrfachmessung erschwert ist.
Chemische Reize	<ul style="list-style-type: none"> • Auftragen von algogenen Substanzen auf die Haut. • Kantharidin-Pflaster-Methode: <p>(Das Pflaster wird am Abend vor dem Experiment auf die Haut geklebt. Es entwickelt sich eine Blase. Diese wird dann aufgeschnitten, und die sensible Haut mit der</p>	<ul style="list-style-type: none"> • Vermutlich selektive Erregung von Nozizeptoren • Der Schmerz ist einzigartig, es wird wahrscheinlich von keiner anderen Methode ein ähnlicher Schmerz erzeugt. 	<ul style="list-style-type: none"> • Geringer Grad der Reizkontrolle • Verletzung der Person wird in Kauf genommen.

Reizmethode	Methodisches Vorgehen	Schmerzcharakteristik	Kritische Anmerkungen
	<ul style="list-style-type: none"> Ischämieschmerz-Methode (submaximum effort tourniquet technique) <p>(Mit einer Blutdruckmanschette wird der Blutfluß im Oberarm unterbrochen. Der Proband wird angehalten währenddessen mit der Hand dieses Armes Muskularbeit zu verrichten, indem er z. B. einen elastischen Gegenstand wiederholt zusammendrückt. Es kommt zur Hypoxie.)</p>	<ul style="list-style-type: none"> Es wird ein kontinuierlich zunehmender, starker Muskelschmerz erzeugt. 	<ul style="list-style-type: none"> Die <i>SETT</i> wurde als <i>Referenzmodell zum Klinischen Schmerz</i> entwickelt. Die Reliabilität und Validität des Verfahrens sind umstritten.

Es wurden eine Vielzahl von unterschiedlichen Methoden entwickelt, wobei jede eine mehr oder weniger andere Schmerzempfindung auslöst. So erzeugen mechanische Reize ein großes Spektrum von Schmerzintensitäten unterschiedlicher Dauer, während der Ischämieschmerz als schwer und kontinuierlich zunehmend beschrieben wird. Jede Methodik hat sicherlich ihre Anhängerschaft. So gibt es keine einheitliche Meinung, daß die eine oder andere Methode den anderen Verfahren weit überlegen ist; jede hat ihre spezifischen Vor- und Nachteile. Bei der Auswahl mögen dann auch die jeweilige Präferenz des Experimentators (Elton, 1983), einfache praktische Gründe der Handhabung und Verfügbarkeit von Geräten oder aber die gewählte Fragestellung ausschlaggebend sein.

Erschwerend kommt hinzu, daß die Operationalisierung innerhalb eines Verfahrens nicht konstant ist. Die Hitzethermode ist bei dem einen Experiment größer, bei einem anderen kleiner. Der Druckstempel wird an verschiedenen Körperstellen angesetzt oder die Ischämieschmerzmethodik leicht variiert (Göbel & Westphal, 1989). Selbst bei Gebrauch von identischen Anordnungen gibt es Unterschiede, wenn die Meßstrategie abgewandelt wird (Elton, 1983). Auf dieses Problem der Meßstrategie, also Erfassung einer bestimmten Reaktion, gehe ich im nächsten Kapitel ein.

Durch **Korrelationsstudien** kann die Frage gelöst werden, ob unterschiedliche Schmerzinduktionsmethoden die gleiche Schmerzempfindung auslösen. Dann wäre es möglich, einen Test durch einen anderen zu ersetzen, wodurch erheblich ökonomischer und überschaubarer gearbeitet werden könnte. Nach einer Literaturdurchsicht kommen Janal et al. (1994) zu dem Schluß, daß die bisherige Forschung es verpaßt hat, eine konsistente Antwort auf die Frage zu erbringen, ob bestimmte Schmerzinduktionsmethoden sich eher ähnlich oder unähnlich sind. In ihrer eigenen Studie induzierten sie Ischämieschmerz, thermischen Schmerz und elektrischen Schmerz bei u. a. 60 jungen Männern. Die

Skalierung erfolgte über das gesamte Spektrum von geringer bis hoher Intensität. Es ergaben sich nahezu "null" Korrelationen zwischen den gewählten Methoden. Sie resümieren, daß eine Methode die andere nicht zu ersetzen vermag. Andere Ergebnisse fanden Göbel und Westphal (1989). Zur Bestimmung der Validität berechneten sie Interkorrelationen von Verfahren zur Induktion von Druck-, Kälte und Ischämieschmerz. bestimmt. Zur Erfassung der Empfindung diente das Kategorienunterteilungsverfahren, eine Skala, die aus sechs Kategorien (sehr schwacher Schmerzreiz bis sehr starker Schmerzreiz) besteht, die ihrerseits numerisch von 1 bis 10 unterteilt sind. Ihre Ergebnisse belegen teilweise hohe Korrelationen zwischen den Verfahren: Druck zu Kälte: $-.11$ bis $.51$, Druck zu Ischämie: $.25$ bis $.77$, Kälte zu Ischämie: $.25$ bis $.56$. Die Korrelationen waren zwischen den gleichen Kategoriengrenzen am höchsten. Insgesamt zeigt der intermodale Schmerzschwellenvergleich, daß die in diesem Experiment gewählten Verfahren sich mehr ähneln, als die Methoden bei Janal et. al. (1994). Insbesondere ist die hohe Korrelation von Ischämie und Druckschmerz auffällig, und belegt im Vergleich zum cold-pressure-test die beiderseitige gute Skalierungsmöglichkeit dieser Verfahren.

Wie in der Literatur beschrieben, stellen sich auf der einen Seite die verschiedenen Verfahren als unähnlich heraus, auf der anderen Seite finden sich hohe Zusammenhänge. Eine allgemeingültige Aussage bzgl. der Ähnlichkeit von Schmerzinduktionsmethoden ist somit nicht möglich. In meinen Augen besitzt jede Methode ihren eigenen eingeschränkten Gültigkeitsbereich. Deswegen sollten sich Aussagen auf das Untersuchungsdesign beziehen, und keine generellen Schlußfolgerungen über *den* Schmerz abgeleitet werden.

2.2.2 Verfahren der Schmerzmessung

Die historischen Wurzeln der Schmerzmessung liegen im Gegenstandsbereich der **Psychophysik**, die sich mit dem Übergangsverhältnis von physikalischen Reizen und der sensorischen Aufnahme und Antwort des Organismus beschäftigt. Ihr Anliegen ist es einerseits, das Verhältnis zwischen dem objektiven physikalischen Reiz und der Empfindung funktionell zu beschreiben und andererseits, Schwellen und Reizunterschiedsempfindungen bei der Schmerzwahrnehmung zu lokalisieren (Price, 1988).

Beecher (1975, S. 57) argumentiert, daß das psychophysikalische Setting zur Erfassung des Schmerzes nicht ausreicht:

„A definitive psychophysics of vision, of hearing, perhaps of taste, smell, and touch, yes, for these sensations can be turned on and off by stimulus control. But the task of developing the psychophysics of significant pain would be like that studying the „psychophysics“ of emotion or anxiety...“

Beecher regte an, die emotionale Basis des Schmerzerlebnis mitzubersichtigen, und betonte die Wichtigkeit von kontrollierten Doppel-Blind-Studien im Labor. In Folge wurden verschiedene

Instrumente zur Quantifizierung der unterschiedlichen Dimensionen von Schmerz entwickelt. Einen Überblick geben Syrjala und Chapman (1984).

In diesem Kapitel möchte ich die für diese Arbeit relevanten Verfahren der Schmerzmessung vorstellen, die sich hauptsächlich aus den Ansätzen der Klassischen Psychophysik und Neuen Psychophysik entwickelten. Ich folge dabei Gracely (1989), der die Meßverfahren in zwei Kategorien einteilt, je nachdem, ob die erzeugte *Reizintensität* oder die *verbale Antwort* auf eine vorgegebene Reizstärke als abhängige Variable betrachtet wird.

Bevor ich diese Ansätze darstelle, möchte ich noch allgemein auf die Problematik des Messens eingehen, da diese Überlegungen auch für die spezielle Aufgabe der Schmerzmessung grundlegend sind.

2.2.2.1 Allgemeine Überlegungen zur Problematik des Messens

Für eine Wissenschaft, die wie die Psychologie ihre Theorien aus der Empirie herleitet, ist die Verknüpfung von Empirie und Theorie von zentraler Bedeutung. Empirische Theorien sollen überprüfbar sein, wobei die Wahrscheinlichkeit eines Plausibilitätsrückschlusses von der Präzision der Verknüpfung abhängt. Je präziser die Verknüpfung der Empirie und der Theorie, desto unwillkürlicher die Aussagen (Steyer & Eid, 1993).

Meßmodelle explizieren die Verknüpfung der logischen Struktur eines theoretischen Begriffs mit den empirischen Begriffen. Das Meßinstrument (beispielsweise ein Testverfahren zur Ermittlung der Schmerzschwelle) definiert dabei oft erst den empirischen Begriff. So kommt es auf die Sichtweise der Wissenschaftler an, ein Konstrukt wie das der *Schmerzempfindlichkeit* als vorhanden, als existent anzusehen und zu messen oder zu meinen, daß dieses Konstrukt vielmehr erst durch die Auswahl eines spezifischen Meßinstrumentes definiert wird.

Übertragen auf das Problem der Auswahl einer Schmerzinduktionsmethode heißt das, daß durch jedes Verfahren eine jeweils andere spezifische Schmerzempfindlichkeit definiert werden würde.

Was ist messen?

"Messen ist eine Zuordnung von Zahlen zu Objekten oder Ereignissen, sofern diese Zuordnung eine homomorphe Abbildung eines empirischen Relativs in ein numerisches Relativ ist"

(Bortz, 1993, S. 19).

Was heißt das?

Bei einem Experiment werden Daten erhoben. Dieser Vorgang wird als Messung anerkannt, wenn die Zuordnung von Zahlen zu den beobachteten Merkmalen nicht beliebig, sondern nach bestimmten Regeln (homomorphe Abbildung) erfolgt, die sicherstellen, daß die Relation zwischen den Objekten erhalten bleiben. Wird also der Schmerz zweimal als „stechend“ beschrieben, so muß sich diese Übereinstimmung (Gleichheit) auch in den zugeordneten Zahlen widerspiegeln, indem „stechend“ z. B. jedesmal die Zahl „1“ zugeordnet wird.

Der Begriff „Relativ“ in der obigen Definition bezeichnet die Zusammenfassung der Objekte und ihrer Relation, also $\langle \text{Schmerz, gleich} \rangle$ wäre das empirische Relativ, $\langle R, = \rangle$ das numerische Relativ (R steht für die Menge der reellen Zahlen, aus der exemplarisch die „1“ ausgewählt wurde).

Bei der Messung wird also eine Beziehung zwischen beiden Relativen hergestellt, indem jedem Objekt eine Zahl zugeordnet wird. Diese Zuordnung oder Abbildung nennt man dann „Skala“ (Wottawa, 1977). Inwieweit wird durch die Skala die empirische Relation berücksichtigt, wie eindeutig oder „stark“ ist die Skala?

Die Skalenstärke oder das **Skalenniveau** wird von den theoretischen Voraussetzungen bestimmt. Man unterscheidet zwischen klassifikatorischen, komparatorischen oder metrischen Skalenniveaumodellen (Steyer & Eid, 1993).

Klassifikatorische Begriffe ermöglichen es, Objekte hinsichtlich ihrer Zugehörigkeit zu disjunkten Klassen zu unterscheiden (Nominalskalenniveau: Menschen mit vs. ohne Schmerzempfindung). Komparatorische Begriffe erlauben einen gerichteten Vergleich (Ordinalskalenniveau: mehr oder weniger schmerzempfindlich) und mit Hilfe der metrischen Begriffe ist zudem das genaue Verhältnis der Objekte zueinander beschreibbar (Intervallskala, Rationalskala: Person A ist doppelt so schmerzempfindlich wie Person B). Das eigentliche Ziel einer Untersuchung, die Daten auszuwerten (bestimmte statistische Verfahren setzen mindestens Intervallskalenniveau voraus) und zu interpretieren, ist also durch den Skalentyp bestimmt.

Allerdings kann das genaue Messen eine fehlende Theorie nicht ersetzen. Wenn man beispielsweise die Schmerzempfindlichkeit über die Zeit mißt, die bis zum Erreichen der Empfindung vergeht, so besitzt die abhängige Variable „Zeit“ im physikalischen Sinn sicherlich Rationalskalenniveau, im psychologischen Sinn ist allerdings fraglich, ob Person A, bei der 60 Sekunden vergehen, doppelt so schmerzempfindlich ist, wie Person B, bei der 120 Sekunden vergehen.

Ist Schmerz meßbar?

Ja, man darf bei der Messung allerdings nicht vergessen, daß immer nur bestimmte **Eigenschaften** eines Objekts betrachtet werden können. So läßt sich „der Schmerz“ per se genauso wenig

messen wie „der Mensch“, sondern immer nur bestimmte Merkmale, wie der Schmerzensausdruck, die Schmerzschwelle oder, bezogen auf den Menschen, die Körpergröße (Göbel, 1990).

Wenn man über die erhobenen Werte auf die zugrundeliegende Eigenschaft schließen möchte, so taucht die Frage auf, ob dieser gemessene Wert auch die tatsächliche Eigenschaft widerspiegelt oder ob irgendwelche Fehler das gewonnene Bild verfälschen. So können Fehlerquellen durch die Ungenauigkeit des Meßinstrumentes (z. B. variierende Druckübertragung bei mangelnder Technik) oder in der Veränderlichkeit der Versuchsbedingungen (Ermüdung der Probanden, Übermotivation, andere Versuchsleiter oder anderes Versuchsleiterverhalten, allgemeine Befindlichkeitsänderung) begründet sein. Wie diese **Meßfehlerquellen** analysiert und kontrolliert werden können, ist ein Teilgebiet der Testtheorie, auf die ich in Kapitel 2.3.1 genauer eingehe.

Wie beschrieben ist Schmerz keine eindeutige physikalische Größe, sondern ein erlebtes psychisches Phänomen. Wie ist es überhaupt möglich, physikalische Größen und psychische Phänomene in einen Meßzusammenhang zu bringen?

Die Disziplin, die sich zuerst mit der Frage des funktionellen Zusammenhangs zwischen definierten vorgegebenen physikalischen Reizen und dem psychologischen Erleben befaßte, ist die Psychophysik. Ihre Erkenntnisse wurden auf den Bereich der Schmerzmessung übertragen.

2.2.2.2 Reizabhängige Verfahren: Schwellenmessung

Der Ansatz der **klassischen Psychophysik** geht auf den Physiologen Ernst Heinrich Weber (1795-1878) zurück, der wohl die ersten Versuche unternahm, die Beziehung zwischen physikalischer Reizstärke und psychischer Empfindung zu quantifizieren (von Campenhausen, 1992). Nach dem von ihm entwickelten Weberschen Gesetz ist die Wahrnehmung des Zuwachs einer Reizintensität im Verhältnis zu einem Standardreiz konstant.

Mathematisch drückt sich die Beziehung in folgender Formel aus:

wobei k die Konstante, S die Reizintensität und ΔS die Zunahme in der Reizintensität ist. Je kleiner k ist, desto empfindlicher ist das Sinnesorgan, da bereits kleine Veränderungen in der Reizstärke wahrgenommen werden.

$$\frac{\Delta S}{S} = k ,$$

Einige Jahre später formulierte Gustav Theodor Fechner auf der Basis von Webers Ansatz eine psychophysische Funktion, die die Empfindungsstärke als proportional zum Logarithmus der Reizstärke beschreibt:

$$E = c * \ln S$$

wobei E die Empfindungsstärke, c die

Proportionalitätskonstante und S die Reizstärke ist. Einem *geometrischen* Anwachsen der Reizstärke steht somit ein *additives* Anwachsen der Empfindungsstärke gegenüber.

Für Fechner ist die „just noticeable difference“ (JND) zweier Reize die Einheit dieser Metrik, wobei die Unterschiedsschwellen über das psychische Spektrum verteilt gleich groß sein sollen. Eine Aneinanderreihung dieser JND hätte dann Intervallskalenniveau, wobei die „Breite“ der JND's Aufschluß über die Empfindlichkeit eines Sinnesorgans gibt: je schmaler, desto sensibler.

In meinen Augen ist die Annahme, daß die JND's bezogen auf das gesamte psychische Spektrum gleich groß sind, unwahrscheinlich. Die zur Wahrnehmungsveränderung notwendige Druckzunahme eines gegebenen Reizes von 1 MPa wird geringer ausfallen als die, die bei 100 MPa Ausgangsdruck notwendig wird. Es macht mehr Sinn, diese Annahme für den im Experiment beobachteten Ausschnitt aus dem theoretischen Wahrnehmungsspektrum zu machen und die Gültigkeit in diesem Bereich nachzuweisen, wenn man sich auf Weber-Fechner bezieht.

Ob der Weber-Fechner-Ansatz für den Schmerz Gültigkeit hat, wurde von Hardy, Wolff und Goodell (1952) experimentell geprüft. Die Autoren fanden für Hitzereize eine Skala, die als Maß für Schmerzintensitäten universelle Gültigkeit besitzt. Leider wurde dieser Ansatz in Folge nicht weitergeführt so daß eine weitere Überprüfung z. B. hinsichtlich anderer Schmerzinduktionsmethoden ausblieb (Heidenfelder, 1991).

Das zentrale Anliegen der klassischen Psychophysik ist es, Schwellen zu bestimmen. Bezogen auf den Schmerz wird die *absolute Schwelle* definiert als die geringste Reizintensität, mit der eine Schmerzempfindung ausgelöst werden kann (Gracely, 1994). Da der Terminus „absolut“ eine uneingeschränkte, vollkommene Größe implizieren könnte, spreche ich in Folge diesbezüglich nur von der „Schwelle“. Eine weitere Schwelle, die bei der Schmerzforschung oft erhoben wird, ist die **Toleranzschwelle**. Sie beschreibt die Reizintensität, bei der die Versuchsperson zum Abbruch der Schmerzinduktion auffordert, der Schmerz also unerträglich wird.

Drei klassische Methoden zur Ermittlung der Schmerzschwelle und Toleranzschwelle werden bis heute eingesetzt: die Grenzwertmethode (Method of Limits), das Konstanzverfahren (Method of constant stimuli) und das Herstellungsverfahren (Method of adjustment) (Göbel, 1992; Wolff, 1986).

Bei der **Grenzwertmethode** werden Reize in auf- oder absteigender Folge vorgegeben. Die Versuchsperson gibt an, ob der jeweilige Reiz schmerzhaft ist. Die Serien werden wiederholt und der Mittelwert aus allen gefundenen Schwellenreizen als Schmerzschwelle definiert.

Beim **Konstanzverfahren** werden Reize verschiedener Intensitäten in Zufallsfolge vorgegeben. Die Versuchsperson gibt auch hier an, ob der jeweilige Reiz schmerzhaft ist. Der Reizintensität, die bei 50% der Reize eine Schmerzempfindung auslöst, wird als Schwelle verwendet.

Bei der **Herstellungsmethode** nähert die Versuchsperson selbst die Reizintensität der eigenen Schwelle an, indem sie die Reizintensität langsam erhöht oder erniedrigt, bis der Schmerz eintritt oder aufhört. Die Prozedur wird wiederholt, die Ergebnisse gemittelt.

Bei den Experimenten, auf die sich diese Arbeit bezieht, wurde das Herstellungsverfahren angewandt, da dieses im Vergleich zu den anderen Methoden dem Probanden selbst eine größtmögliche Kontrolle und Sicherheit einräumt. Es ist zudem nicht so zeitintensiv wie das Konstanzverfahren.

Neben den Ansätzen von Weber und Fechner gibt es einen dritten Zugang zur Entdeckung von Schmerzschwellen, den ich der Vollständigkeit halber kurz darstellen möchte. Es handelt sich um die Theorie der Entdeckbarkeit von Signalen (**signal detection theory**, SDT) (Green & Swets, 1966). Sie geht davon aus, daß die Antwortneigung einer Person bei der Entdeckung von Schwellen mitberücksichtigt werden müsse. Das Ziel ist es, die reine Sinnesempfindung von der Antwortneigung zu trennen. Hierin liegt der grundlegende Unterschied zur klassischen Psychophysik unterscheidet, die Antwortneigung und Sinnesempfindung zusammen erfaßt. Die Grundannahme besagt, daß die Antwort auf einen gegebenen Reiz von einem Hintergrundrauschen - beispielsweise der ständigen Nervenzellaktivität - beeinflußt wird. Eine Person muß entscheiden, ob der Reiz wirklich gegeben wurde oder ob nur das Rauschen vorliegt.

Es ergeben sich somit vier Entscheidungsmöglichkeiten: 1. die Person entdeckt einen gegebenen Reiz, landet somit einen Treffer, 2. die Person entdeckt einen nicht gegebenen Reiz, schlägt fälschlicherweise Alarm, 3. die Person entdeckt einen gegebenen Reiz nicht, verpaßt diesen oder 4. die Person entdeckt einen nicht vorhandenen Reiz nicht, bleibt also korrekterweise stumm. Die Antwortneigung ist dabei variabel und hängt auch von den Konsequenzen ab, ob also z. B. eine falsche Entscheidung stark bestraft werden würde. Dann würde der Proband eher dazu neigen, falsche Aussagen zu vermeiden und nur die eindeutigsten Reize angeben. Eine implizite Annahme der SDT wäre also demnach, daß es so etwas wie eine "absolute Schwelle" gar nicht gibt und deren Messung ein Irrweg ist.

Es gab eine Vielzahl von Arbeitsgruppen, die sich mit der Übertragbarkeit der SD-Theorie in die experimentelle Schmerzforschung beschäftigten. Eine Darstellung hier würde den Rahmen dieser Arbeit verlassen. Eine gute Übersicht findet sich bei Wolff (1986). Auch Pöppel (1980) und Velden (1980) setzen sich kritisch mit der Schmerzmessung mit Hilfe der SDT auseinander. Velden referiert drei Hauptkritikpunkte:

1. Methodische Einwände: Bei vielen Untersuchungen erfolgten zuwenig Reizdarbietungen und die Versuchspersonen durchliefen keine Einübungsphasen. Somit könnten die gefundenen Werte durch eine erhebliche Zufallsstreuung beeinflusst worden sein.
2. Validität: Die SDT mißt keinen Schmerz, sondern Diskriminationsfähigkeit.
3. Effektkonfundierung: Durch die Schmerzreize wird der Aufmerksamkeitszustand der Versuchsperson beeinflusst, welcher letzten Endes für die Entdeckung eines Reizes maßgeblich wird.

Die Ergebnisse erlauben heute noch keine abschließende Beurteilung über die Nützlichkeit der SDT in der Schmerzforschung. Deswegen wurde bei den hier zu referierenden Untersuchungen auch auf die bewährteren, klassischen Verfahren zurückgegriffen.

2.2.2.3 Antwortabhängige Verfahren: verbale Skalierung

Die zugrundeliegende Sicht der **Neuen Psychophysik** sagt aus, daß ein Individuum quantitative Einschätzungen der Größe der Sinnesempfindung *direkt* vornehmen kann. Dieses Vorgehen steht insofern im Kontrast zur klassischen Methodik, als daß dort die Sinnesempfindungen *indirekt* über Reizintensitäten erfaßt werden. Göbel (1992, S. 69-70) meint, daß der Terminus „Neue Psychophysik“ ungenau ist, da bereits 1888 Merkel auf diese Verfahren hinwies.

Nach von Campenhausen (1993) gilt Stanley Smith Stevens (1906-1973) als Begründer dieses Ansatzes. Laut dem stevenschen Potenzgesetz führen gleiche Reizverhältnisse zu gleichen Wahrnehmungsverhältnissen. Die psychophysische Funktion lautet

wobei E die $E = k * S^n$ Empfindungsstärke, k eine Proportionalitätskonstante, S der Reiz und n der Exponent ist, der die Eigenart des Sinnes wiedergibt. Wäre $n = 1$, so wäre die Empfindungsstärke über das Reizkontinuum kontinuierlich gleich. Wäre hingegen $n < 1$, so nähme die Empfindungsstärke mit zunehmenden Reizen ab, bei $n > 1$ mit steigender Reizstärke zu!

Für die Diskussion der generellen Gültigkeit dieses Ansatzes sei auf die entsprechende Literatur verwiesen. Für den Schmerz wurde die Stevensche Funktion häufig überprüft. Adair, Stevens, J.C. und Marks (1968) ermittelten einen Exponenten für Druckschmerz von 1, Stevens (1977) berichtet für Schmerz, der durch elektrische Reizung erzeugt wurde, einen Exponenten von 3.5. In einer Übersicht referieren Price (1988) und Price und Harkins (1992) mehrere Studien bei denen Exponenten von 2.1 und 2.2 für die Wahrnehmung von Hitzeschmerz ermittelt wurden. Für die Empfindung des Schmerzreizes als „unangenehm“ wurden höhere Exponenten von 2.4 bis 2.7 ermittelt. Lautenbacher, Möltner und Strian (1992) ermittelten mittels einer Thermode, die am Daumenballen aufsetzte, Exponenten von 1.8 und 1.9.

Insgesamt scheint es keinen allgemeingültigen „wahren“ Exponenten für die Schmerzempfindung zu geben. Man muß die verschiedenen Induktionsmethoden getrennt betrachten. Da alle Exponenten im Bereich $\square 1$ liegen, ist festzuhalten, daß die Schmerzempfindung im Vergleich zur Reizintensität besonders im überschwelligem Bereich überproportional ansteigt. Bei der Planung von Experimenten sollte das berücksichtigt werden, damit für die Versuchspersonen bei Messungen im überschwelligem Bereich der Schmerz erträglich bleibt.

Die zur Prüfung des Stevenschen Gesetzes angewandten Verfahren sind das Größenschätzverfahren (magnitude estimation, ME) und das cross-modality-matching (CMM).

Bei der **cross-modality-matching** Methode werden Reize aus einer Sinnesmodalität (z. B. Schmerzreize) vorgegeben. Die Versuchsperson hat dann die Aufgabe, die empfundene Reizintensität in einer anderen Modalität hinsichtlich ihrer Größe anzugleichen (z. B. Helligkeit einer Lichtquelle). Nach Chapman et al. (1985) führt diese Methode zu konsistenten Antworten. Sie ist relativ unabhängig von der Antwortneigung der Probanden. Reading (1989) kritisiert, daß hier das eigentliche Ziel der einfachen Schmerzmessung zu Gunsten der Ergründung der Übereinstimmung unterschiedlicher Modalitäten vernachlässigt wird.

Das Verfahren der **Größeneinschätzung** beruht darauf, daß die Versuchsperson dem ersten Reiz eine Zahl zuordnet, die nachfolgenden Reize einer zufälligen Serie werden dann im Verhältnis zu dem ersten beziffert. Dieses Verfahren führt theoretisch zu einer Rationalskala, mit der Schmerzschwelle als einem individuellen Nullpunkt. Gracely (1994) weist darauf hin, daß dieser Punkt in der Literatur umstritten ist. Er meint, daß dieses hohe Skalenniveau in den seltensten Fällen erreicht wird, aber es könnten mehr Informationen als mit Kategorienskalen gewonnen werden.

Eine andere, *nicht* psychophysikalische, Möglichkeit der Einschätzung der Schmerzempfindung ist die Verwendung von **Kategorienskalen** (Zenz & Jurna, 1993). Die Einstufung der Schmerzempfindung erfolgt entweder numerisch z. B. durch Vorgabe einer 11 Punkte-Skala, wobei die 0 keiner Schmerzempfindung entspricht und 10 dem maximal vorstellbarem Schmerz. Es können auch Kategorien vorgegeben werden, die von „Kein Schmerz“, über „Schwacher Schmerz“ bis zu „Extrem Starker Schmerz“ reichen. Neben Intensitätsabstufungen können die Kategorien auch mit affektive Ausdrücken gekennzeichnet werden, wie z. B. „Beunruhigender Schmerz“, „Quälender Schmerz“ oder „Unerträglicher Schmerz“.

Die Kategorienskalen erreichen Ordinalskalen- bzw. Nominalskalenniveau. Sie sind einfach in der Handhabung und reagieren sensibel auf Veränderungen. Ein großer Nachteil liegt darin, daß die ausgesuchten Beschreibungen unterschiedliche menschliche Emotionen widerspiegeln können, und nicht dieselbe zugrundeliegende Schmerzdimension. Für die numerischen Kategorien taucht das Response-bias-Problem auf: bei vorwiegend starken Reizen werden diese vornehmlich im unteren Skalenbereich angesiedelt. Das Gegenteil gilt für schwache Reize: werden überwiegend schwache Reize vorgegeben so

werden diese eher im oberen Skalenbereich skaliert (Price & Harkins, 1992). Eine andere Frage greift das Problem der Kategoriengrenzen auf: differenziert eine Skala von 1 bis 4 besser als eine von 1 bis 101? Jensen (1994) verglich bei chronischen Schmerzpatienten unterschiedliche, numerische Skalen. Er fand heraus, daß eine 21-stufige Skala am sensitivsten war. Fernandez (1990) induzierte Ischämieschmerz und stellte die Frage, ob „offene“ oder „geschlossene“ Skalen zu reliableren Ergebnissen führen. Er fand heraus, daß eine offene Skala, bei der die Obergrenze nicht festgelegt wurde, die genauere Messung ermöglichte. Leider wurden in beiden Studien nicht untersucht, ob dieselbe Skala auch für unterschiedliche Schmerzinduktionsmethoden die geeignetste ist. So könnten für Hitze- und Ischämieschmerz aufgrund ihrer unterschiedlichen Schmerzcharakteristik auch unterschiedliche Skalen passend sein.

Ich habe in diesem Kapitel im Wesentlichen die Verfahren geschildert, die in den Untersuchungen, auf die sich meine Arbeit bezieht, zur Anwendung kamen. Weitere Möglichkeiten wie visuelle Analogskalen (VAS), das Kategorienunterteilungsverfahren nach Heller oder den McGill-Schmerzfragebogen können hier im einzelnen nicht aufgeführt werden, da dies den Rahmen meiner Arbeit verlassen würde. Ich verweise auf entsprechende Quellen wie Göbel (1992), Reading (1989) oder Geissner, Dalbert und Schulte (1992).

2.3 Schmerzmessung als Test

2.3.1 Gütekriterien eines Tests

Per definitionem ist ein Test ein

„...wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“

(Lienert & Raatz, 1994, S. 1).

Der Test muß dabei bestimmten Anforderungen genügen, wie sie die **Klassische Testtheorie** vorgibt. Neben dem klassischen Ansatz existieren auch andere testtheoretische Ansätze, wie der probabilistische Zugang oder die Generalisierbarkeitstheorie. Zur Diskussion der unterschiedlichen Vor- und Nachteile der Zugänge siehe Cronbach (1965), Michel und Conrad (1982) und Renkl und Helmke (1993).

Auf Grundlage der Klassischen Testtheorie soll ein Test **drei Hauptgütekriterien** genügen: er soll objektiv, valide und reliabel sein. Zudem werden vier Nebengütekriterien gefordert. Der Test soll normiert, vergleichbar, ökonomisch und nützlich sein.

Unter **Objektivität** wird der Grad verstanden, in dem die Testergebnisse vom Untersucher unabhängig sind. Verschiedene Untersucher sollten dieselben Ergebnisse ermitteln. Ein hohes Maß an *Durchführungsobjektivität* ist dann gegeben, wenn die Ergebnisse von zufälligen oder systematischen Verhaltensvariationen des Versuchsleiters nicht beeinflusst werden. Die *Auswertungsobjektivität* betrifft die Auswertung des Test nach mehr oder weniger festen Regeln. Die Übereinstimmung der Auswertung verschiedener Untersucher wird bei freien Antwortformaten (z. B. Projektive Tests) geringer sein, als bei Tests mit einem festen Auswertungsschlüssel. Die *Interpretationsobjektivität* ist dann gegeben, wenn bei Gleichheit der Ergebnisse unterschiedlicher Versuchspersonen, die Ergebnisinterpretation zu den gleichen Schlußfolgerungen führt. Bei Tests, deren Resultat an Hand einer normierten Skala interpretiert wird, ist also die Interpretationsobjektivität ganz gegeben, da das gleiche Ergebnisse bei der einen Person nicht als „hervorragend“ und bei der anderen als „schlecht“ bewertet werden kann.

Unter **Validität** wird der Grad der Genauigkeit gefaßt, mit der ein Test das mißt, was er zu messen vorgibt. Ein valider Test läßt einen fehlerfreien Rückschluß auf den Ausprägungsgrad des Persönlichkeitsmerkmals zu. Als *kriterienbezogene Validität* wird der Zusammenhang der Testmeßwerte mit einem Außenkriterium definiert, das vom Test unabhängig erhoben wird. Das Außenkriterium kann als valider Repräsentant des Persönlichkeitsmerkmals betrachtet werden. So ist es üblich, durch Korrelation der Testmeßwerte mit einem bereits validierten Test Aussagen über die Güte der kriterienbezogenen Validität zu treffen.

Die **Reliabilität** ist ein Maß für den Grad der Genauigkeit, mit der ein Test ein bestimmtes Merkmal erfaßt. Sie kennzeichnet den Grad, in dem die beobachtbaren Testergebnisse durch tatsächliche Unterschiede in den Merkmalen erklärbar sind (Michel & Conrad, 1982). Wird die Reliabilität mittels der Varianzen der klassischen Modellkomponenten („beobachteter“ Wert = „wahrer“ Wert + Fehler) berechnet, so ist die Reliabilität gleich dem Anteil der Varianz der „wahren“ Testwerte an der Varianz der „beobachteten“ Testwerte. Der Reliabilitätskoeffizient (r_{tt}) gibt den Grad der Reliabilität an.

Gewöhnlich werden drei **Reliabilitätsarten** unterschieden: die Stabilität (Vergleich nach Ablauf einer bestimmten Zeitspanne), die Äquivalenz (Vergleich der Formen) und die Interne Konsistenz (interne Widerspruchslosigkeit). Ob man als Reliabilitätsmaß einen *Stabilitätskoeffizienten* oder einen *Konsistenzkoeffizienten* wählt, hängt von den betrachteten Variablen ab, ob ein zeitkonstantes Merkmal oder ein aktuelles Merkmal getestet wird.

Zur Ermittlung der Koeffizienten gibt es verschiedene methodische Zugänge. Die Zugänge sind das Paralleltest-, das Retest- und das Halbierungsverfahren, sowie die Methode der Konsistenzanalyse.

Im *Paralleltestverfahren* werden ähnliche Tests einer Stichprobe vorgelegt und die Ergebnisse korreliert. Bei der *Retestmethode* wird derselbe Test einer Stichprobe zweimal vorgegeben und anschließend die Korrelation der beiden Testwertreihen ermittelt. Das *Halbierungsverfahren* und die *Konsistenzanalyse* ermöglichen Aussagen über die innere Konsistenz eines Meßinstrumentes. Im Halbierungsverfahren wird ein Test nach der Durchführung in zwei gleiche Teile zerlegt, die korreliert werden. Der Ansatz der Konsistenzanalyse geht davon aus, daß ein Test in so viele Teile untergliedert werden kann, wie Items vorhanden sind. Der Konsistenzkoeffizient kennzeichnet die Leistungsfähigkeit des Testinstruments, die sogenannte instrumentelle Reliabilität. Zur Ermittlung der Koeffizienten eignen sich die methodischen Zugänge unterschiedlich gut. Stabilitätskoeffizienten z. B. werden am besten im Retest-Verfahren ermittelt.

Für die Güte des Koeffizienten gilt, daß für Gruppendifferenzen ein Wert von .50, für individuelle Differenzen .70 und für Retest- oder Paralleltestreabilitäten .80 ausreichend ist. Bei einem gering reliablen Test ($.50 < r_{tt} < .70$) sollten keine individuellen Vergleiche mehr durchgeführt werden, allenfalls nur noch Gruppenvergleiche.

Wie Abbildung 3 zeigt, stehen die Hauptgütekriterien in einer wechselseitigen Abhängigkeit (nach Lienert & Raatz, 1994).

Wenn Aussagen über eines der Kriterien getroffen werden, muß auch die Gültigkeit der anderen Kriterien bedacht werden. Somit kann ein Test nicht valider sein als er reliabel ist. Die Parallel- und Retestreliabilität ist höchstens so hoch wie die Konsistenz oder die Objektivität. Ist die kriterienbezogene Validität hoch, muß auch die Objektivität und die innere Konsistenz hoch sein. Wenn eine hohe kriterienbezogene Validität ermittelt wurde, so wird auch eine hohe Objektivität und Reliabilität vorliegen. Hingegen müssen bei einer niedrigen kriterienbezogenen Validität die Objektivität und Reliabilität nicht unbedingt schlecht sein.

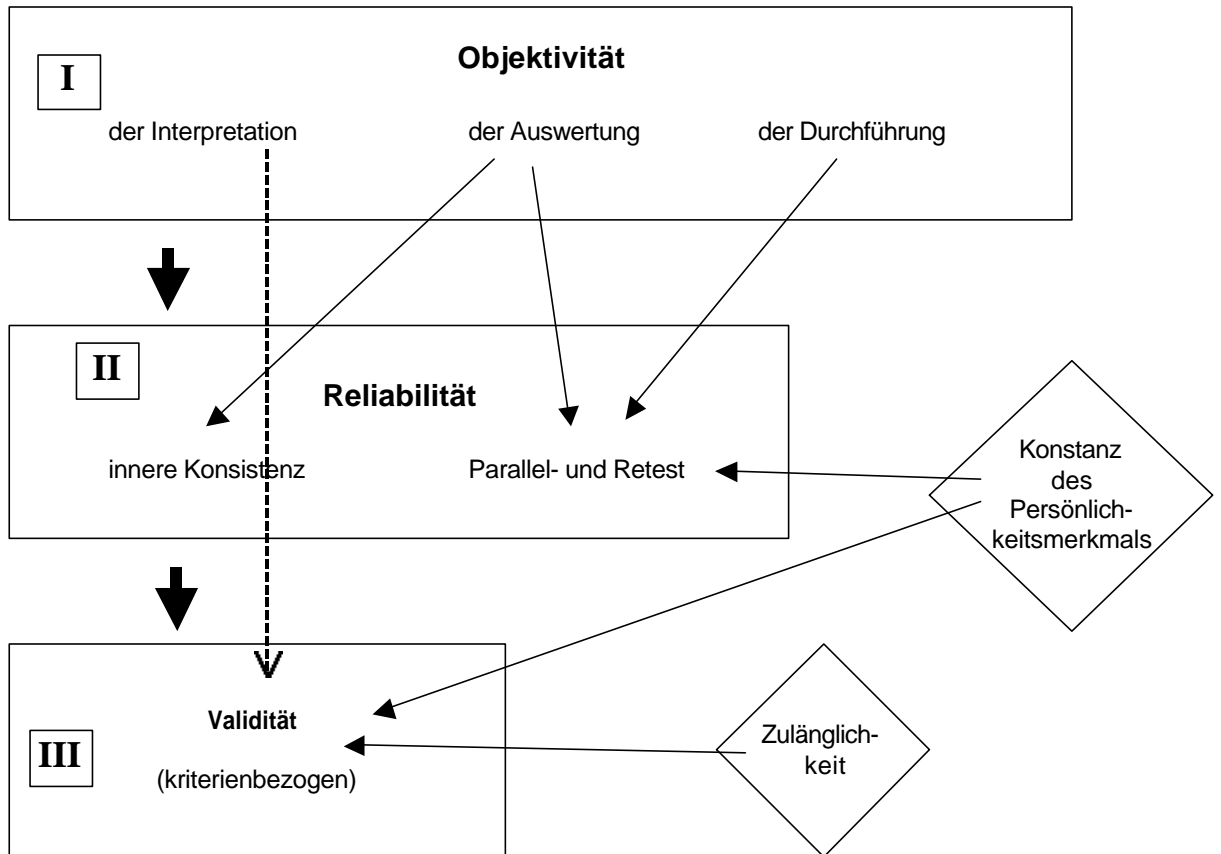


Abb. 3: Wechselbeziehungen zwischen den Hauptgütekriterien: die Gütekriterien sind hierarchisch gegliedert. Die allen übergeordnete Objektivität beeinflusst die Reliabilität und über die Interpretation die Validität. Außenkriterien wie die Zulänglichkeit (das was vom Test und einem Außenkriterium gemeinsam erfaßt wird) und der Konstanz des Persönlichkeitsmerkmals nehmen ebenfalls Einfluß auf die Validität und Reliabilität.

Die Nebenkriterien dienen zur weiteren Bewertung der Handhabbarkeit eines Tests. Die **Normierung** dient als Bezugssystem für die Einordnung individueller Ergebnisse, die **Vergleichbarkeit** dient dem Vergleich des Tests mit sich selbst (als Parallelform) oder dem Vergleich mit ähnlichen anderen Tests, die **Ökonomie** fordert eine überschaubare, einfache Durchführbarkeit, die **Nützlichkeit** verlangt als Voraussetzung zur Testplanung das Vorliegen eines praktischen Bedürfnisses, das durch keinen anderen ebenso guten Test befriedigt werden könnte.

Ich denke, daß Auch bei der Entwicklung von Tests zur Schmerzmessungen auch die Nebenkriterien beachtet werden müssen. Ein ökonomisches Verfahren entlastet in erster Linie die Versuchspersonen (in zweiter Linie die Versuchsleiter) und bei der Entwicklung neuer Verfahren sollte man in Hinblick auf die Fülle der bereits erprobten Methoden die Nützlichkeit der eigenen neuen Methodik hinterfragen.

Die Testgütekriterien wurden in erster Linie für Intelligenz-, Leistungs- und Persönlichkeitstest entwickelt, die psychologische Eigenschaften betrachten. Mit der Schmerzempfindung werden aber psychophysiologische Merkmale gemessen, woran sich die Frage anschließt, ob psychophysiologische Methoden als Test nach den testtheoretischen Kriterien bezeichnet werden können. Nach der Definition mißt ein Test die individuelle Ausprägung abgrenzbarer Persönlichkeitsmerkmale. Ich denke, daß in diesem Sinn auch verschiedene Schmerzschwellen als Merkmale eines Individuums betrachtet werden können, die testbar sind. Die Testinstrumente müssen demnach den Anforderung genügen, wie sie die klassische Testtheorie formuliert. Hacke (1983) vergleicht *neurophysiologische* Tests mit rein psychologischen Tests hinsichtlich der Gütekriterien. Er mißt dabei den neurophysiologischen Tests Vorteile zu. Hacke nennt hier eine hohe Objektivität, die durch die reizkorrelierte Messung (EEG, EMG) von Versuchspersonenvariablen relativ unabhängig ist, das hohe Validitätsniveau, das durch tierexperimentelle Außenkriterien erreicht wird und das hohe Skalenniveau, das durch Erfassung physikalischer Variablen wie der Zeit oder der Spannung erreicht wird. Er erwartet, daß gerade bei reizkorrelierten Verfahren erwarten, die Retest-Reliabilität hoch ist. Die scheinbare Genauigkeit der Messung von physikalischen Größen mag allerdings dazu führen, daß Meßunterschieden vorschnell eine Variation im Merkmal zugrundegelegt wird, und nicht ein eventuell fehlerhafter Test verantwortlich gemacht wird. Auch sei hier nochmals auf das Problem des Skalenniveaus hingewiesen, daß bei der physikalischen Härte der Skala, der Meßkontext nicht vernachlässigt werden darf.

Für *psychophysiologische* Methoden, also Verfahren in denen sowohl die subjektive Antwort der Versuchsperson als auch ihre physiologischen Merkmale erhoben werden, habe ich keine Studie gefunden, die Aussagen über die zu erwartende Qualität von Gütekriterien trifft. Ich denke, daß gerade in diesem Bereich die Chance zur Ergänzung besteht, da man sowohl über „subjektive“ Angaben als auch über die gleichzeitige „objektive“ Messungen zu Erkenntnissen gelangt, die integriert eine größere praktische Relevanz haben.

2.3.2 Untersuchungen zur Reliabilität von Schmerzmessungen

Wenn ein Testinstrument zur Schmerzmessung eingesetzt wird, ist es unabdinglich, die Qualität bzw. Reliabilität der Induktionsmethode zu überprüfen. Andererseits besteht die Gefahr, daß gefundene Effekte im Antwortverhalten auf Merkmalsveränderungen der Probanden zurückgeführt werden, obwohl sie möglicherweise auf der Ungenauigkeit oder Variabilität der Methode beruhen.

Auch Turk, Rudy und Sorkin (1993) betonen in ihrem Review über die Ergebnisse der Behandlungsmethoden von chronischen Schmerzpatienten die Wichtigkeit der Überprüfung der Reliabilität von Schmerzmeßmethoden.

"In all instances, it must be assumed, that an outcome measure is unreliable until proven otherwise."

Dies gilt in meinen Augen erst recht für die experimentelle Schmerzforschung, da hier die Kontrollierbarkeit scheinbar eine Meßgenauigkeit vorgibt, die entweder gar nicht vorhanden sein muß oder durch die experimentelle Anordnung (z.B. Anzahl der Versuchspersonen) vorgetäuscht werden kann.

Wolff (1986) kommt nach einer Literaturdurchsicht zu der Ansicht, daß man als "Daumenregel" für die **Schmerzschwelle** Reliabilitätskoeffizienten von .60 und darüber als akzeptabel, von .80 und mehr als sehr gut und solche von .90 und höher als exzellent einstufen kann. Diese sollen als *Konsistenzmaß* für die Reliabilität innerhalb einer Sitzung gelten. Für die **Schmerztoleranzgrenzen** wird eine "Daumenregel" nicht erstellt. Da aber die Schmerztoleranz gegenüber äußeren psychischen Faktoren als leichter beeinflussbar gilt (Wolff, 1986; Elton, 1983; Price, 1988) sollte für die Toleranzschwelle eine geringere Reliabilität erwartet werden.

Für die *Retest-Reliabilität* von Sitzungen, die zeitlich mindestens eine Woche auseinander liegen, wären auch geringere Werte akzeptabel. Konkrete Zahlen nennt Wolff hier allerdings nicht.

In seiner Literaturdurchsicht berichtet Wolf (1986) für die Schmerzschwelle von Reliabilitätskoeffizienten von .88 bis .92 für Strahlungshitze, .81 bis .90 für elektrische Stimulation, .65 bis .91 für Druckreize. Auch für die Schmerztoleranzschwelle wurden hohe Reliabilitäten ermittelt. Sie variieren zwischen .85 und .93 für Strahlungshitze, zwischen .81 und .86 für Druckschmerz und lagen in einer Studie bei .91 für elektrisch induzierten Schmerz.

In einer Untersuchung mit dem Druck-Algometer fanden Merskey und Spear (1964) bei 49 Studenten Retestkorrelationen von .65 für die Schmerzschwelle und .81 für eine höhere Intensität („when it hurts“). Ohrbach und Gale (1989a) testeten 10 Studenten mit einem Druck-Algometer, der auf verschiedenen Gesichtsmuskeln aufsetzte zu fünf verschiedenen Sitzungszeitpunkten. Sie fanden Reliabilitäten von .55 bis .96, abhängig von den jeweiligen Muskeln und Meßzeitpunkten. Es zeigte sich auch, daß sich mit zunehmenden Sitzungen die Reliabilität nicht in bemerkenswerter Weise steigerte.

Aufgrund der Vielzahl der Experimente wäre es unsinnig, hier die Unterschiede zwischen allen Untersuchungen aufzuzeigen und zu interpretieren. Ich habe deshalb drei Experimente herausgegriffen, die denen unserer Arbeit vergleichbar sind und die ich im Hinblick auf die spätere Diskussion genauer vorstellen möchte.

Curio (1990) maß bei 23 Studenten (22-32 Jahre) bei konstanter **thermischer Stimulation** die Aversions-, Schmerz-, Interventions-, und Toleranzschwelle, indem er ein 1x1 cm² kleines Peltierelement am Unterarm der nicht-dominanten Seite plazierte. Die Anfangstemperatur der Hitzethermode wurde der Hauttemperatur angeglichen. Als abhängige Variable wurde die Zeit gemessen, die bis zum Erreichen der jeweiligen Schwelle verging. Die Probanden drückten einen Knopf, sobald sie die jeweilige Schwelle wahrnahmen. Die Messungen wurden sechsfach durchgeführt und eine Woche später zur selben Tageszeit

(nachmittags) wiederholt. Zur Bestimmung der Retest-Reliabilität wurde der Korrelationskoeffizient nach Spearman-Brown bestimmt. Für die Konsistenz wurde Cronbach`s Alpha berechnet. Die Auswertung erfolgte getrennt für die ersten und die letzten drei Durchgänge einer Sitzung.

Für die Schmerzschwelle betrug die Retest-Reliabilität .83 für Durchgang eins bis drei und .72 für Durchgang vier bis sechs. Für die Toleranzschwelle betrug die Retest-Reliabilität 0.82 für Durchgang eins bis drei und .32 für Durchgang vier bis sechs. Die interne Konsistenz lag für die Schmerzschwelle bei .83 (Messung eins bis drei) bzw. bei .81 (Messung vier bis sechs), die der Toleranzschwelle bei .91 und .80. Die Starttemperatur nahm signifikant mit der Anzahl der Sitzungen zu. Die Reize fingen in der ersten Sitzung bei 52.7 °C an und in der 2. Sitzung 53.1 °C. Für diese Untersuchung kann die Reliabilität der Schwellen und Toleranz in den ersten drei Sitzungen als sehr gut angesehen werden. In den letzten beiden Sitzungen sinkt die Reliabilität besonders bei der Toleranzmessung. Eine Erklärung dafür gibt der Autor nicht. Es werden in dieser gut kontrollierten Studie auch keine Angaben zum Geschlecht der Probanden gemacht. Auch weitere Befunde wie die Auswertung der Konsistenzen für Sitzungen oder eine Durchschnittsberechnung für alle sechs Messungen fehlen.

Fernandez (1990) induzierte **Ischämieschmerz** bei 48 Studenten. Der Versuchsplan sah vier randomisierte Gruppen (N=12) vor, getrennt nach Geschlecht und der Meßskala, die entweder nach oben hin offen oder geschlossen war. Die Skala begann mit 0 für kein Schmerz, 1 für Schwelle und mit der Schrittweite von 1 für die weiteren Schwellen. Bei der "offenen Gruppe" war der Toleranzwert variabel. Für die "geschlossene Gruppe" galt 10 als Maß für die Toleranz. Die Empfindung brachten die Versuchspersonen zum Ausdruck, indem sie auf der neben ihnen liegenden numerischen Skala die jeweilige Stärke ankreuzten. Als abhängige Variable wurde die Zeit, die bis zum Erreichen der Schwellen verging, mit einer Stoppuhr gemessen. Nach einer Woche wurde die Sitzung wiederholt.

Als Ergebnis ergaben sich signifikant höhere Schwellen in der 2. Sitzung. Sie betragen im Mittel in der ersten Sitzung 24.35 sec. und in der 2. Sitzung 43.83 sec. Für die Schmerzschwellen lagen die Test-Retest-Korrelation (Pearson) für die Männer der "offenen" Skala am höchsten (.94), dann folgten die Frauen derselben Skala (.73), dann die Männer (.54) und Frauen (.05) der geschlossenen Gruppe. Für die Toleranzschwelle waren die Koeffizienten höher: In gleicher Reihenfolge wie oben: .98, .75, .87 und .73. Ein Ergebnis ist hier, daß die offene Skala eine genauere Messung ermöglicht. Interessant ist der Befund, daß die Schwelle weniger reliabel als die Toleranz ist. Die Autoren vermuten, daß in einem gut kontrollierten, identischen Setting die psychischen Variablen, für die die Toleranz als anfälliger gilt, nicht zur Auswirkung kommen. Die Frage, warum in von der ersten zur zweiten (identischen) Sitzung die mittlere Schmerzschwelle um nahezu 80% ansteigt bleibt von dem Autor unbeantwortet. Da die Retest-Reliabilitäten sehr hoch sind, ist zu erwarten, daß die Versuchspersonen in beiden Sitzungen im Verhältnis zum Sitzungsmittel entweder hohe oder niedrige Schwellen haben, auch wenn die absolute Schwellenhöhe ansteigt.

Göbel und Westphal (1989) untersuchten die Reliabilität von **mechanischem Druckschmerz**, einer modifizierten Methode zur Erzeugung von **Ischämieschmerz** und dem **cold-pressure-test**. Die Druckschmerzinduktion erfolgte durch Aufsetzen eines Stempels (Fläche 2.56 mm², 2.64 MPa) auf den Mittelfalangen der Finger zwei bis fünf beider Hände. Die Ischämieschmerzinduktion erfolgte durch Aufblasen der Manschette auf 200 mm Hg, die den Beginn der Reizapplikation darstellte. Gleichzeitig drückte die Versuchsperson einen Schaumstoffball rhythmisch mit der Hand, bis die Schmerztoleranzgrenze erreicht war. Beim cold-pressure-test wurde die Hand der Versuchsperson in 4 °C kaltes Wasser getaucht und verblieb dort bis zur Erreichung der Toleranz. Zur Erfassung der Empfindung diente das Kategorienunterteilungsverfahren, eine Skala, die aus sechs Kategorien (sehr schwacher Schmerzreiz bis sehr starker Schmerzreiz) besteht. Die Kategorien sind numerisch von 1 bis 10 unterteilt sind. Die Skala ist nach oben hin offen. Mit Hilfe eines verstellbaren Zeigers konnten die Versuchspersonen ihre Empfindung den Kategorien/Zahlen zuordnen. 12 Männer und 12 Frauen (20-27 Jahre) wurden zweimal morgens (7.30 bis 9 Uhr) und zweimal abends (17.30 bis 19 Uhr) untersucht. Sie nahmen also an vier Sitzungen teil, wobei diese immer einen Tag auseinanderlagen. Als Reliabilitätskoeffizient wurde Cronbach`s Alpha berechnet.

Über die vier Sitzungen hinweg ergeben sich für die Schwelle Reliabilitäten von .93 für Druck-, .95 für Ischämie- und .84 für Kälteschmerz. Für den Endpunkt der Skala (50), der mit der Schmerztoleranz vergleichbar sein soll, liegen die Konsistenzen bei .90, .96 und .92.. Die hohen Reliabilitätskoeffizienten sprechen für die Zuverlässigkeit der gewählten Methoden. Die Autoren weisen die Bezüglichkeit daraufhin, daß mit ihrer Methode die gesamte Skalenbreite gleich gut erfaßt werden könne, was dadurch erklärt werden könne, daß die Instruktion einheitlich darin bestehe, den empfundenen Schmerz zu skalieren, und nicht zwischen bloßer Wahrnehmung und Toleranz einzuschätzen, was eine komplexere, fehleranfälligere Aufgabe wäre. Zudem gibt es einen Reihenfolgeeffekt. Beim Druckschmerz werden die Probanden zur letzten Sitzung hin immer sensibler. Die Autoren werten die Druckschmerzsensibilisierung als ein Zeichen für eine gute Validität der Methode, da eine Sensibilisierung die Alltagserfahrung widerspiegelt. Ungewöhnlich ist, daß der Konsistenzkoeffizient von Cronbach`s Alpha und nicht die Retest-Korrelation über die aufeinanderfolgenden Sitzungen berechnet wurde, da die Konsistenz eher zur Ermittlung der Fehlerhaftigkeit eines Testinstruments zu einem Testzeitpunkt gewählt wird, die Retest-Korrelation dagegen die Stabilität über verschiedene Sitzungen illustriert. Ein Grund mag sein, daß die Meßzeitpunkte bei Göbel und Westphal nur einen Tag auseinander lagen, und keine Aussage über die Konstanz des Persönlichkeitsmerkmals „Schmerzschwelle“ getroffen werden sollte.

Werden alle Experimente mit ihren unterschiedlichen methodischen Zugängen verglichen, so sind nach den Kriterien von Wolff (1986) alle Methoden mit Ausnahme einer Toleranzmessung bei Curio und einer Schwellenbestimmung bei Fernandez als *akzeptabel* bis *sehr gut* anzusehen.

3 Ableitung der Fragestellung

In zwei vorliegenden Studien wurden Schmerzen mit drei verschiedenen Verfahren induziert. Es soll geprüft werden, ob die Schmerzinduktionsmethoden reliabel sind. Aufgrund der identischen Methodik ist es möglich, die beide Untersuchungen gemeinsam auszuwerten, um die Stichprobe zu vergrößern. Überprüft werden soll, ob dieses Vorgehen gerechtfertigt ist, da im experimentellen Aufbau der Untersuchungen wesentliche Unterschiede bestehen. Im intermodalen Vergleich soll die Frage beantwortet werden, ob die Schmerzinduktionsmethoden in ihrer Wirkung ähnlich sind und zur ihrer gegenseitigen Validierung genügen. Schließlich sollen mögliche Fehlerquellen aufgedeckt und Änderungen zur Verbesserung der Reliabilität vorgeschlagen werden.

4 Darstellung der Experimente

Im letzten Jahr wurden am Psychologischen Institut der Universität Kiel zwei Untersuchungen durchgeführt, die sich mit der Fragestellung beschäftigten, wie die Schmerzwahrnehmung beim Menschen durch Neuropeptide beeinflusst wird.

Im ersten Experiment (**EXP1**) „Modulation of pain perception in man by a vasopressin analogue“ geht es darum, inwieweit verschiedene Dosen Desmopressins (DDAVP) die Schmerzempfindung (Schmerzschwelle, Schmerztoleranz, Diskriminationsfähigkeit) beeinflussen (Pohl, Arnold, Schulz, Pause, Schulte, Fehm-Wolfsdorf, 1995).

Im zweiten Experiment (**EXP2**) „Analgesic effect of ceruletide in man is limited to specific pain qualities“ wurde die gleiche Frage bzgl. des Neuropeptids Cholecystokinin (CRL) gestellt (Pause, Drews, Scherhag, Pohl, Pietrowsky, Ferstl, Schulte, Fehm-Wolfsdorf, 1995).

Die beiden Untersuchungen sind im Hinblick auf die wesentlichen Parameter der Stichprobe, das Design und die Schmerzinduktionsmethoden vergleichbar. Auf die wesentlichen **Unterschiede** des Prozedere gehe ich bei den jeweiligen Punkten gesondert ein.

4.1 Untersuchungsdesign

Es handelt sich um Doppel-Blind-Versuche im Cross-over Verfahren (inkomplettes lateinisches Quadrat).

Die Versuchspersonen wurden in einer Screening-Sitzung selektiert, an die Geräte gewöhnt und mit dem Versuchsablauf vertraut gemacht. Es folgten **drei Experimentalsitzungen**. Der Abstand zwischen den Sitzungen betrug im EXP1 zwei Wochen und im EXP2 eine Woche.

In jeder Sitzung wurden zu **drei Zeitpunkten** Messungen durchgeführt. Nach einer Baseline-Messung wurde das Treatment gegeben. Hinterher erfolgten zwei Nachmessungen. In EXP1 wurde das Treatment intranasal appliziert, in EXP2 hingegen intravenös injiziert.

Kontrollvariablen waren der Blutdruck, die Hormone (ACTH, Cortisol) und die Emotionale Befindlichkeit (EWL-N)..

Die **unabhängigen Variablen** waren:

UV 1: Treatment: Placebo, niedrige Dosis und hohe Dosis des Neuropeptids

UV 2: Experimentalsitzung (S1, S2, S3)

UV 3: Meßzeitpunkt (M1, M2, M3)

UV 4: Ordnungsaspekt: Zuordnung der Personen zu den Treatment-Stufen über die Experimentalsitzungen

*Für diese Arbeit ist nur die **Baseline-Messung** (M1) wesentlich, da hier die Reliabilität der Schmerzmessung vom Treatment nicht beeinflußt ist. Ein sitzungsübergreifender Treatmenteffekt nach einer Woche bzw. nach zwei Wochen kann ausgeschlossen werden. Alle weiteren Ergebnisse und Aussagen basieren ausschließlich nur auf diesen Meßzeitpunkten der Sitzungen.*

Zur Operationalisierung des Schmerzes wurden drei unterschiedliche Induktionsmethoden eingesetzt. Als **Abhängige Variablen** wurden die Schmerzschwellen, die Toleranzschwellen und die Diskriminationsfähigkeit erfaßt.

Für die einzelnen Verfahren gilt:

- Druckschmerz		
	dzt1 =	Schmerzschwelle, Zeigefinger (Zeit in sec.)
	dzt2 =	Toleranzschwelle, Zeigefinger (Zeit in sec.)
	dmt1 =	Schmerzschwelle, Mittelfinger (Zeit in sec.)
	dmt2 =	Toleranzschwelle, Mittelfinger (Zeit in sec.)
	drt1 =	Schmerzschwelle, Ringfinger (Zeit in sec.)
	drt2 =	Toleranzschwelle, Ringfinger (Zeit in sec.)
- Ischämieschmerz	it1 =	Schmerzschwelle (Zeit in sec.)
- Hitzeschmerz	ht =	Schmerzschwelle (°C)
	h1 to h15 =	Diskriminationsreize (numerische Skala)

Abb. 4 zeigt die experimentelle Anordnung im Überblick. In der Abbildung werden nur die für diese Arbeit ausschlaggebenden Variablen aufgeführt.

	S1	S2	S3
M1	dzt1, dzt2, dmt1, dmt2, drt1, drt2 it1 ht, h1 to h15	dzt1, dzt2, dmt1, dmt2, drt1, drt2 it1 ht, h1 to h15	dzt1, dzt2, dmt1, dmt2, drt1, drt2 it1 ht, h1 to h15
M2	dito	dito	dito
M3	dito	dito	dito

Abb. 4: Überblick über die experimentelle Anordnung; In drei verschiedenen Sitzungen (S1 bis S3) wurden die Variablen (dzt1 bis h15) zu drei Meßzeitpunkten (M1 bis M3) erhoben. Für diese Arbeit ist nur die Baseline-Messung (M1) maßgeblich.

4.2 Schmerzinduktionsmethoden

Es wurden drei Verfahren zur Induktion von **Druck**-, **Ischämie**-, und **Hitze**schmerz eingesetzt, die in Folge detailliert beschrieben werden.

4.2.1 Messung des Druckschmerzes

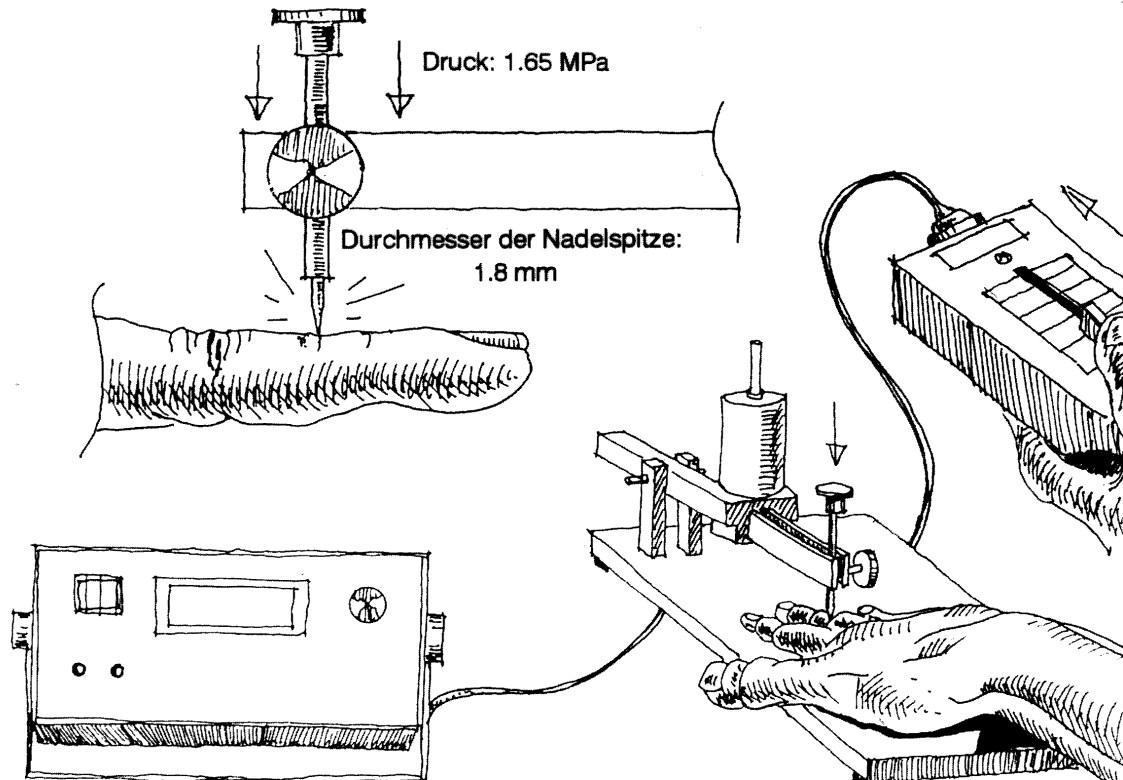


Abb. 5: Schematische Darstellung des Druckschmerzverfahrens. Eine PC-gesteuerte stumpfe Nadel drückt auf die Mittelphalanx des jeweiligen Fingers. Mit der dominanten Hand skaliert die Versuchsperson ihre Empfindung anhand einer Kategorienskala. Der Schieberegler ist ebenfalls mit dem PC verbunden.

Mit der *Herstellungsmethode* (method of adjustment) wurde die Schmerzschwelle und die Toleranzschwelle für drei Finger der nicht-dominanten Hand in der Abfolge **Zeige-**, **Mittel-**, und **Ring**finger bestimmt.

Ein computergesteuerter Arm drückte maximal 120 sec. lang mit einer stumpfen Nadel (\varnothing 1.8 mm) mit 1.65 MPa auf die Mittelphalanx eines jeden Fingers. Die Druckpunkte wurden markiert, damit im Laufe einer Sitzung für jeden der drei Meßzeitpunkte ein anderer, nicht sensibilisierter Bereich gewählt werden konnte.

Mit der dominanten Hand skalierte die Versuchsperson den Schmerz mittels eines Schiebereglers, auf dem Kategorien aufgetragen waren, die eine zunehmende Schmerzintensität widerspiegeln. Diese Skala hatte die Kategorien: „kein“, „schwacher“, „mittlerer“, „starker“ und „unerträglicher“ Schmerz (s. Abb. 6).

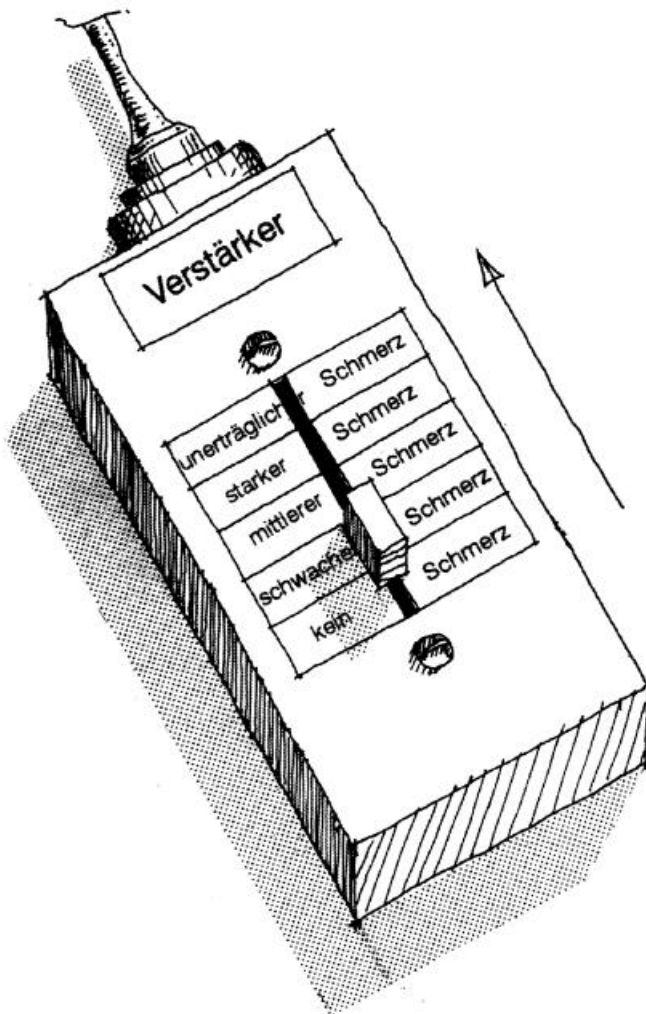


Abb. 6: Anhand der Kategorien skaliert die Versuchsperson ihre Empfindung. Die Zeit, die bis zum Erreichen einer Kategorien-grenze verging, wird von dem angeschlossenen PC erfaßt.

Der PC maß zum einen die Zeit, die bis zur Wahrnehmung des Schmerzes verging, also den Zeitpunkt an dem die Versuchsperson die Grenze zwischen „kein“ und „schwacher“ Schmerz mittels Schieber überschritt, zum anderen die Zeit bis der Schmerz als „unerträglich“ eingestuft wurde, also die Kategorie „starker“ zu „unerträglich“ überschritten wurde (Meßungenauigkeit: 0.1s). Wurde der Schmerz „unerträglich“, so hob sich automatisch der elektronische Arm.

4.2.2 Messung des Ischämieschmerzes

Mit der *Herstellungsmethode*

(method of adjustment) wurde die Schmerzschwelle bestimmt.

Dazu wurde am nicht-dominanten Oberarm eine Blutdruckmanschette angelegt und auf 200 mm Hg aufgepumpt. Gleichzeitig wurde der Sauerstoffverbrauch durch aktive Muskelbewegung erhöht, indem die Versuchsperson aufgefordert wurde, einen Softball mit der nicht-dominanten Hand rhythmisch zu pressen. Die resultierende Ischämie (Einschränkung der Durchblutungsgröße) führt dann zu einer Schmerzempfindung (s. Abb. 7).

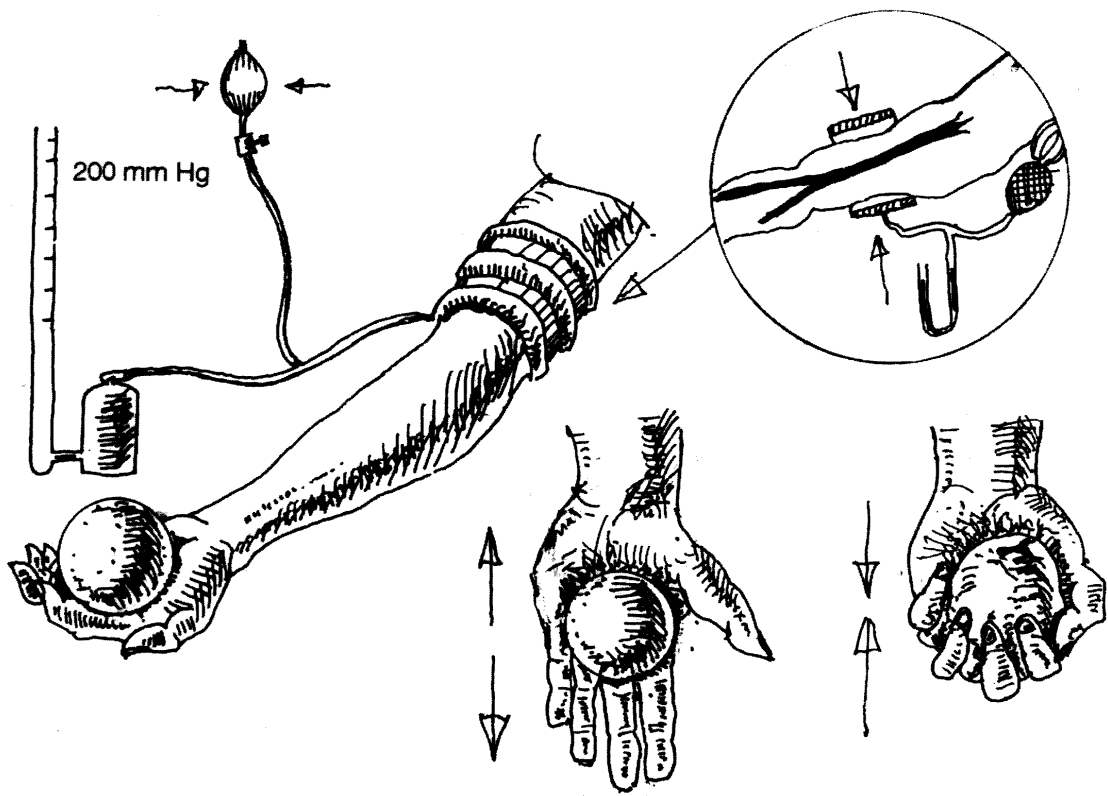


Abb. 7: Schematische Darstellung des Ischämieschmerzverfahrens; Mittels einer Blutdruckmanchette und gleichzeitiger aktiver Muskelbewegung wird eine Ischämie erzeugt.

Der Druck wurde maximal vier min. lang aufrechterhalten. Gemessen wurde die Zeit bis zur Erreichung der Schmerzschwelle anhand der gleichen Skala, die auch beim Druckschmerz angewandt wurde (vgl. Abb. 6). Die Toleranzschwelle wurde hier nicht miterhoben.

4.2.3 Messung des Hitzeschmerzes

Mit der *Herstellungsmethode* (method of adjustment) wurde die Schmerzschwelle ermittelt.

Der Proband legte dazu den Thenar (Daumenballen) der nicht-dominanten Hand auf ein Peltierelement (Thermode). Die Auflagefläche betrug 4 cm^2 (s. Abb. 8)

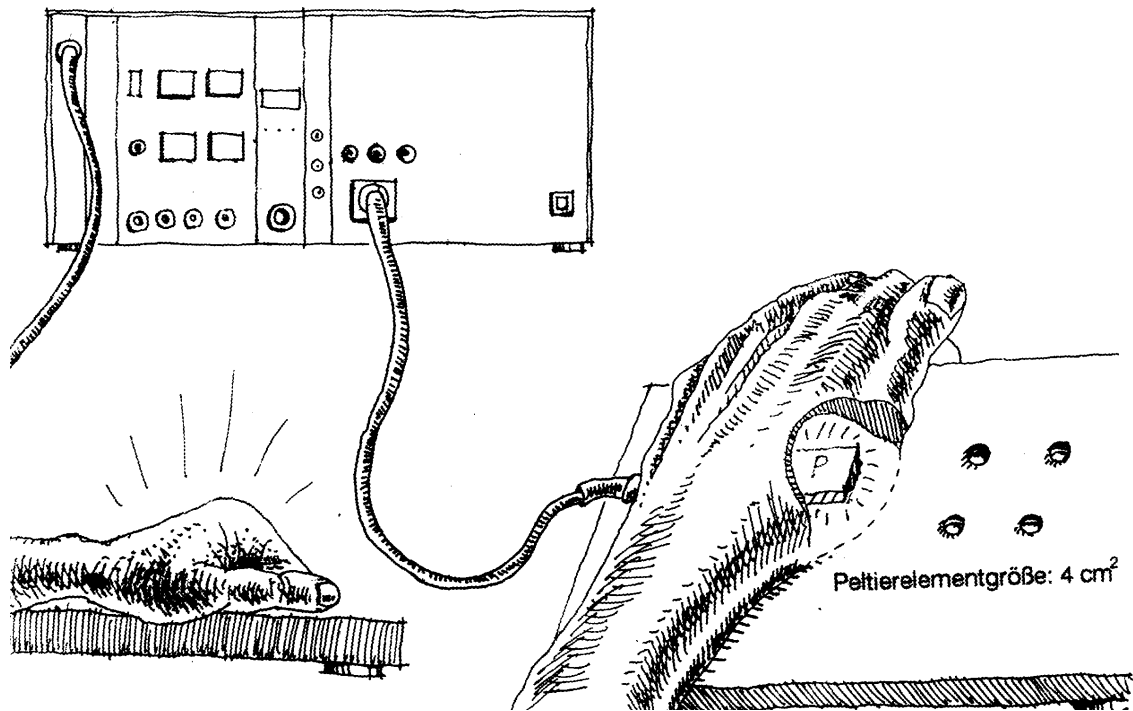


Abb. 8: Schematische Zeichnung des Hitzeschmerzinduktionsverfahrens; Am Handballen der nicht-dominanten Hand, der auf einer Thermode aufliegt, wird ein Schmerzreiz erzeugt, indem die Temperatur des Peltierelements kontinuierlich zunimmt. Die ermittelten Temperaturen werden von dem angeschlossenen PC erfaßt.

Ausgehend von 40 °C stieg der Hitzereiz kontinuierlich um 0.2 °C/sec. an. Sobald der Proband den Schmerzreiz wahrnahm, drückte er mit der anderen Hand einen Knopf und kennzeichnete so seine Schmerzschwelle. Der angeschlossene PC speicherte die zu diesem Zeitpunkt vorliegende Temperatur.

Dieses Verfahren wurde siebenmal durchgeführt. Der Durchschnitt der letzten *fünf* Durchgänge wurde dann als individuelle Hitzeschmerzschwelle genommen. Die Schwellenbestimmung wurde nur während der Baseline-Messung durchgeführt. Zu allen drei Meßzeitpunkten wurden danach Hitzediskriminierungsversuche im Größenschätzverfahren (magnitude estimation) durchgeführt.

Ausgehend von der individuellen Schwelle wurde eine zufällige Abfolge von fünf verschiedenen Hitzereizen mit einem Interstimulusintervall von 15 sec. vorgegeben. Die Reiztemperaturen betragen 0.8 °C unterhalb der Schwelle, die Schwelle selbst, 0.8 °C, 1.6 °C und 2.4 °C oberhalb der Schwelle. Die Reizdauer war konstant 2 sec. lang. Jeder Reiz wurde dreimal gegeben, so daß die Versuchspersonen insgesamt **15** Reize einschätzen mußten.

Die Einschätzung erfolgte anhand einer numerischen Analogskala, die den Versuchspersonen erläutert wurde. Die Skala geht von `0` bis `7`, wobei die Versuchspersonen mündlich instruiert wurden, den Wert `3` mit ihrer Schmerzschwelle gleichzusetzen. In diesem Sinn sollte die Einstufung `2` von einer Versuchsperson eine Schmerzempfindung unterhalb der Schmerzschwelle und eine Einschätzung `4` eine etwas über der Schwelle liegende Empfindung widerspiegeln.

4.3 Durchführung der Untersuchungen

Die Untersuchungen fanden in Räumen des Psychologischen Instituts der Universität Kiel statt.

4.3.1 Versuchspersonen

An dem **EXP1** nahmen 24 gesunde, männliche Freiwillige (Nichtraucher) teil. Sie waren zwischen 20 und 30 Jahre alt ($M=25.46$, $SD 2.57$) und wurden für ihre Teilnahme bezahlt. In der Screening-Sitzung wurden die Versuchspersonen mit den Geräten vertraut gemacht und nach verschiedenen Gesichtspunkten ausgewählt. So sollten die Probanden nicht neurotisch (FPI-R Wert > 11) sein und keine zu hohen Schmerzschwellen haben (Schwelle > 47.6 °C). Drei Personen wurden danach im EXP2 von dem weiteren Versuch ausgeschlossen. Die Versuchspersonen wurden instruiert, vor jeder Sitzung keinen Alkohol zu trinken, mindestens sechs Stunden geschlafen und 24 Stunden vorher keine Medikamente eingenommen zu haben. Zudem sollten sie drei Stunden vor Beginn keinen Sport getrieben und keine reichhaltigen Mahlzeiten zu sich genommen haben.

Am **EXP2** nahmen 25 männliche Versuchspersonen teil. Sie waren zwischen 21 und 28 Jahre alt ($M=24.4$, $SD=2.0$). Für sie galten die gleichen Kriterien wie für die Versuchspersonen von EXP1. Da die Testsitzung morgens stattfand, wurden die Männer gebeten, nach 22 Uhr des vorhergehenden Abends nichts mehr zu essen. Die Compliance wurde mit einer Blutglukosemessung kontrolliert.

4.3.2 Ablauf

Der Sitzungsbeginn lag im EXP1 am **Nachmittag** ($M= 15$ Uhr, $SD=2$ h). Es folgten die Messung der Variablen in der Reihenfolge: Blutdruck, Hormone (Blutentnahme für Cortisol, ACTH), Schmerzreize, EWL-N. Die Schmerzreize wurden immer in der gleichen Abfolge appliziert: Druckschmerz, Ischämieschmerz und Hitzeschmerz. Die Baseline-Messung erfolgte 30 min. vor, die erste Nachmessung 30 min. nach und die zweite Nachmessung 60 min. nach der Treatmentapplikation.

Im EXP2 lag der Sitzungsbeginn am **Vormittag** zwischen 7 Uhr und 9 Uhr. Die Variablen wurden hier in der Reihenfolge: EWL-N, Hormone (Speichelcortisol), Blutdruck, Schmerztest erhoben. Auch hier erfolgte die Baseline-Messung 30 min. vor der Treatmentgabe und die erste Nachmessung 30 min. später, die zweite Nachmessung allerdings 90 min. später. Die Schmerzreize wurden in der gleichen Abfolge wie in EXP1 appliziert.

Die **Versuchsleiter** waren im EXP1 zwei junge Frauen. Jeweils die Hälfte der Versuchspersonen wurde von einer der beiden getestet. Im EXP2 führten eine junge Frau und ein junger Mann die Versuche durch. Der männliche Leiter testete 13 Versuchspersonen, die Frau 12 Probanden. Zusammenfassend werden in der Tabelle 2 die Unterschiede zwischen den Experimenten gegenübergestellt.

Tab. 2: Unterschiede im experimentellen Aufbau der Experimente

	EXP1	EXP2
Treatmentgabe	intranasal (Vasopressin)	intravenös (CCK)
Erhebungsreihenfolge der Variablen	Blutdruck, Hormone, Schmerzreize, EWL-N	EWL-N, Hormone Blutdruck, Schmerzreize
Abstand zwischen den Sitzungen	14 Tage	sieben Tage
Sitzungsbeginn	nachmittags (M= 15 Uhr, SD=2 h)	vormittags 7 Uhr bis 9 Uhr
Versuchspersonen	N= 24 (M=25.46, SD 2.57) keine <i>reichhaltigen</i> Mahlzeiten vor den Sitzungen zu sich genommen	N=25 (M=24.4, SD=2.0). keine Mahlzeit nach 22 Uhr des vorhergehenden Abends mehr eingenommen
Versuchsleiter	zwei Frauen	eine Frau und ein Mann

5. Auswertung

Die Strategie der einzelnen Auswertung jedes Experimentes und der gemeinsame Auswertung beider Experimente ist identisch.

Da die Variablen zu drei aufeinanderfolgenden Sitzungen erhoben werden, können Retestrelabilitäten ermittelt werden. Für die zu einer Sitzung erfolgten Mehrfachmessungen einer Variablen soll die Konsistenz angegeben werden

- Retestrelabilitäten:

Zur Auswertung der Retestrelabilitäten für die Temperatur- und Zeitmessungen werden die Produkt-Moment-Korrelationen (Pearson) und für die Hitzediskriminationsreize Rangkorrelationen (Spearman) berechnet. Dabei werden die einzelnen Methoden, die jeweiligen Schwellen und die Größenschätzung getrennt betrachtet (vgl. Abb. 4).

Zur Mittelung (r) wurden die Korrelationskoeffizienten erst Fisher Z-transformiert, dann gemittelt und zurücktransformiert. Die Transformation ist notwendig, da nicht davon auszugehen ist, daß eine Korrelation, die in einer Untersuchung A mit .80 erhoben wurde, doppelt so hoch ist, wie eine Korrelation, die in der Untersuchung B mit .40 ermittelt wurde, da die zugrundeliegenden Verteilungen der Korrelationen in der Grundgesamtheit nicht normalverteilt sein müssen. Durch die Fisher Z-Transformation werden Koeffizienten erzeugt, deren Verteilung einer Normalverteilung entspricht und die zudem Maßzahlen einer Verhältnisskala darstellen. Die Mittelung von Korrelationskoeffizienten muß daher auf der gemeinsamen Basis der Fisher Z-Verteilung erfolgen, um nach der Rücktransformation interpretiert werden zu können.

$$Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

- Konsistenzkoeffizient

Für die Konsistenzen soll Cronbach`s Alpha ermittelt werden.

Die Konsistenz wird zum einen für den Druckschmerz ermittelt, indem die verschiedenen Finger als dreifache Realisierung einer Variablen betrachtet werden und zum anderen für die Hitzeschmerzschwelle, die fünfmal realisiert wurde (vgl. Abb. 4).

- Normalverteilung

Zur Prüfung der Variablen auf Normalverteilung wird der Einstichproben Kolmogorov-Smirnov-Test gerechnet.

- Versuchsleitereffekte

Je nach Vorliegen einer Intervall- oder Ordinalskala wird ein ungerichteter t-Test für unabhängige Stichproben oder der Mann-Whitney U-Test gerechnet.

6 Ergebnisse

Im Ergebnisteil werden alle Koeffizienten für die getrennte und gemeinsame Auswertung der Experimente in einer gemeinsamen Tabelle dargestellt.

Für alle Ergebnisse gilt folgendes Signifikanzniveau:

*** = $p < 0.001$

** = $p < 0.01$

* = $p < 0.05$

n.s.= nicht signifikant

6.1 Methodenübergreifende Ergebnisse

6.1.1 Überprüfung der Voraussetzungen

Alle Variablen wurden auf Normalverteilung getestet. Bei der getrennten Auswertung und der gemeinsamen Auswertung der Experimente kann für alle Hitzeskalierungen **keine** Normalverteilung angenommen werden. Bei der gemeinsamen Auswertung liegt zudem keine Normalverteilung vor bei der Schwellenmessung, 2. Sitzung, Ringfinger (RF) und der Schwellenmessung, 3. Sitzung, Zeigefinger (ZF) und gemittelte Finger (GF) (s. Tab.3). Für die getrennten Auswertungen kann für alle anderen Variablen Normalverteilung angenommen werden.

Tab. 3: Nichtnormalverteilte Variablen in der gemeinsamen Auswertung

Variable	M	SD	K-S Z	P (z)
RF	7.72	5.2	1.42	.03
ZF	9.89	8.44	1.45	.03
GF	9.57	7.15	1.46	.03

6.1.2 Versuchsleitereffekte

Bei allen Variablen sollten keine Unterschiede abhängig vom Versuchsleiter vorliegen. Im EXP1 und im EXP2 ergeben sich für Ischämie- und Druckschmerz keine bemerkenswerten Differenzen zwischen den Versuchsleitern. Bei den Hitzeskalierungen unterschieden sich die Versuchsleiter im EXP1 bei allen Hitzeschwellenmessung und bei allen Hitzeschwellenmittelungen, im EXP2 bei zwei Hitzereizskalierungen. Es zeigt sich eine Tendenz, daß im EXP1 bei der 1. Versuchsleiterin und im EXP2 bei der weiblichen Versuchsleiterin immer höhere Werte gemessen wurden. Alle signifikanten Vergleiche finden sich im **Anhang**. Tabelle 4 zeigt exemplarisch die signifikanten Vergleiche der Hitzeschwellenmittel zu den drei Sitzungen im EXP1.

Tab. 4: Hitzeschwellenmittel der Sitzungen im EXP1

	°C	M	SEM	t -Wert	p (t)
1. Sitzung	VL 1	46.19	.24	2.70	.02
	VL 2	44.69	.50		
2. Sitzung	VL 1	46.01	.23	2.16	.04
	VL 2	45.11	.35		
3. Sitzung	VL 1	46.29	.18	4.08	.00
	VL 2	44.31	.45		

6.1.3 Vergleich der Methoden

Die **Schwellenmessungen** der unterschiedlichen Verfahren werden nach Sitzungstermin getrennt ausgewertet und ihre Mittelungen (Fingermittel, Hitzemittel) über drei Sitzungen hinweg verglichen. Da die Toleranzschwelle nur beim Druckschmerz erhoben wurde, geht sie nicht in den intermodalen Vergleich mit ein.

Tab.5 : Korrelationen der Schwellenmessungen, getrennte Sitzungen

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	Ischämie	Hitze	Ischämie	Hitze	Ischämie	Hitze
1. Sitzung						
Druck	.18 n.s.	.32 n.s.	.28 n.s.	.44*	.14 n.s.	.33*
Hitze	-.04 n.s.		-.05 n.s.		-.02 n.s.	
2. Sitzung						
Druck	.27 n.s.	-.21 n.s.	.11 n.s.	.09 n.s.	.16 n.s.	-.07 n.s.
Hitze	.13 n.s.		-.17 n.s.		.06 n.s.	
3. Sitzung						
	Ischämie	Hitze	Ischämie	Hitze	Ischämie	Hitze

Druck	.19 n.s.	-.01 n.s.	.16 n.s.	.41*	.18 n.s.	.13 n.s.
Hitze	.10 n.s.		-.42*		-.08 n.s.	

Tabelle 5 zeigt den intermodalen Vergleich der Schmerzinduktionsverfahren für die einzelnen Sitzungen. Das Ergebnis ist für die verschiedenen Korrelationen heterogen. *Ischämie und Hitzeschmerz* sind bei beiden Experimenten zur Mehrheit unkorreliert. Die Korrelationen liegen in beiden Experimenten entweder leicht über Null (bis .13) oder korrelieren negativ (bis -.42). Für den Vergleich von *Ischämie und Druckschmerz* sind alle Korrelationen höher. Sie variieren von .16 bis .27. Für den dritten Vergleich von *Hitze und Druckschmerz* ist das Bild sowohl innerhalb eines Experimentes als auch zwischen den Experimenten sehr unterschiedlich. Im EXP2 ergeben sich in der 1. Sitzung und 3. Sitzung mit .44 und .41 die höchsten Korrelation, in der 2. Sitzung hingegen korrelieren die Methoden fast gar nicht (.09). Im EXP1 bietet sich ein anderes Bild. Hier finden sich hauptsächlich negative oder geringe Korrelationen (-.21 bis .04), während in der 1. Sitzung Hitze und Druck mit .32 korrelieren.

Tabelle 6 gibt auf Aufschluß über den intermodalen Vergleich unabhängig vom Sitzungszeitpunkt. Das Resultat ist kongruent mit den obigen Ergebnissen. Ischämie und Hitzeschmerz sind unkorreliert, Ischämie und Druckschmerz korrelieren mit .18 und .27 während Hitze und Druckschmerz im EXP gar nicht und im EXP2 hoch (.40) korrelieren.

Tab. 6: Korrelationen der Schwellenmessungen, *alle Sitzungen*

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	Ischämie	Hitze	Ischämie	Hitze	Ischämie	Hitze
Druck	.27 n.s.	.04 n.s.	.18 n.s.	.40*	.19 n.s.	.18 n.s.
Hitze	.10 n.s.		-.22 n.s.		.00 n.s.	

6.2 Ergebnisse des Druckschmerzes

An den drei Experimentalsitzungen wurde an jeweils drei Fingern [Reihenfolge: Zeigefinger (ZF), Mittelfinger (MF), Ringfinger (RF)] die Schmerzschwellen und Toleranzschwellen gemessen. Es folgen die Ergebnistabellen für Mittelwerte (**M, oberer Wert**) und Standardfehler (**SEM, unterer Wert**) und die Retest-Reliabilitäten für die einzelnen Finger, sowie für die gemittelten Finger (GF). Bei der Konsistenzenberechnung werden die drei Finger als Realisierung einer Variablen betrachtet.

6.2.1 Ergebnisse der Schwellenmessung

Beim **Vergleich der Mittelwerte** (Tab. 7) fällt im EXP1 für alle einzelnen Finger auf, daß die Schmerzschwelle von der ersten zur dritten Sitzung steigt. Im EXP2 zeigt sich diese Tendenz nicht. Beim Ringfinger sinkt die Schwelle, während sie bei den anderen Finger über die Sitzungen schwankt. Für EXP1 & EXP2 zeigt sich die Beobachtung aus EXP1 in abgeschwächter Form. Bei allen Fingermessungen von EXP1 und EXP2 fällt auf, daß die Schwellen in der 1. und 2. Sitzung des EXP2 höher liegen, als im EXP1.

Tab. 7: *M und SEM der einzelnen und der gemittelten Finger*

	EXP1 (N=24)				EXP2 (N=25)				EXP1 & EXP2 (N=49)			
	ZF	MF	RF	GF	ZF	MF	RF	GF	ZF	MF	RF	GF
1. Sitzung	6.5 1.24	7.18 1.30	5.88 1.03	6.52 1.05	10.07 1.01	10.07 1.15	8.76 .94	9.63 .93	8.32 .83	8.66 .88	7.35 .72	8.12 .73
2. Sitzung	7.60 1.13	7.19 1.03	7.94 1.3	7.58 1.04	8.08 .90	10.25 1.49	7.52 .78	8.61 .96	7.84 .71	8.75 .93	7.72 .74	8.12 .70
3. Sitzung	9.53 1.91	9.98 2.19	9.38 1.75	9.63 1.77	10.29 1.52	9.23 .87	9.05 1.25	9.5 1.09	9.89 1.2	9.6 1.15	9.21 1.06	9.57 1.02

6.2.1.1 Einzelne Finger

Tab. 8: *Retest-Reliabilitäten*

<i>Zeigefinger</i>	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.52**	.40 n.s.	.71***	.84***	.58***	.58***
3. Sitzung	.78***		.58**		.68***	
\bar{r}	.59		.72		.61	
<i>Mittel-finger</i>						
2. Sitzung	.54**	.58**	.83***	.60**	.71***	.46**
3. Sitzung	.90***		.47*		.70***	

\bar{r}	.72		.66		.63	
Ringfinger						
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.48*	.57**	.71***	.64**	.51***	.59***
3. Sitzung	.62**		.62**		.59***	
\bar{r}	.56		.66		.56	

Die **Retest-Reliabilitäten** variieren für die einzelnen Finger im EXP1 von .40 bis .90, im EXP2 von .47 bis .84. Die mittleren Korrelationen liegen im EXP1 im Bereich von .56 bis .72, im EXP2 bei .66 und .72. Bei keinem Finger finden sich sowohl bei den Sitzungsreliabilitäten als auch in den gemittelten Werten konsistent höhere Werte. Im EXP1 ergeben sich die höchsten Korrelationen zwischen der 1. und 3. Sitzung, bei EXP2 ist das Ergebnis uneinheitlich (Tab.8).

6.2.1.2 Gemittelte Finger

Die **Konsistenz** der Schwellenmessung (Tab. 9) für die Finger als Realisierung einer einzigen Variable variiert von .85 bis .90. Die beiden Experimente zeigen in allen Sitzungen keine größeren Unterschiede in ihren Werten.

Tab. 9: Konsistenz

	EXP1 (N=24)	EXP2 (N=25)	EXP1 & EXP2 (N=49)
1. Sitzung	$\alpha = .85$	$\alpha = .89$	$\alpha = .88$
2. Sitzung	$\alpha = .88$	$\alpha = .89$	$\alpha = .86$
3. Sitzung	$\alpha = .90$	$\alpha = .88$	$\alpha = .91$

Für das Mittel der Finger liegen die Reliabilitäten für EXP1 bei .64 bis .91, für das EXP2 bei .65 bis .86. Die durchschnittliche Korrelation liegt für beide Experimente bei .77. Die gemeinsame Auswertung von EXP1 & EXP2 liegt mit .73 niedriger (Tab. 10).

Tab. 10: Retest-Reliabilitäten

	EXP1	EXP2	EXP1 & EXP2
--	------	------	-------------

	(N=24)		(N=25)		(N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.67***	.64**	.86***	.77***	.75***	.67***
3. Sitzung	.91***		.65***		.76***	
\bar{r}	.77		.77		.73	

6.2.2 Ergebnisse der Toleranzschwellenmessung

Beim Vergleich der Mittelwerte ist kein Anstieg oder Abfall sowohl in der Sitzungsreihenfolge, als auch in der Abfolge der Finger ersichtlich. Bei allen Fingermessungen fällt auf, daß im EXP1 alle Toleranzschwellen in allen Sitzungen und bei allen Fingern nahezu doppelt so hoch (41.6 bis 53.61 sec.) liegen, wie in EXP2 (20.01 bis 29.60 sec.). Bei der gemeinsamen Auswertung ergeben sich Schwellen, die in etwa in der Mitte zwischen EXP1 und EXP2 zu finden sind (Tab. 11).

Tab. 11: M und SEM der Finger bei getrennter und gemeinsamer Auswertung

	EXP1 (N=24) (#N=23)				EXP2 (N=25)				EXP1 & EXP2 (N=49) (#N=48)			
	ZF	MF	RF	GF	ZF	MF	RF	GF	ZF	MF	RF	GF
1. Sitzung	45.94 5.69	46.40 [#] 4.56	41.60 4.50	45.20 4.27	26.91 2.69	26.22 2.95	24.29 3.19	25.80 2.69	36.23 3.37	35.89 [#] 3.02	32.77 2.99	35.31 2.84
2. Sitzung	53.61 4.88	47.82 [#] 4.80	46.72 5.13	49.73 4.25	24.57 3.25	25.47 3.30	20.01 2.18	23.35 2.74	38.79 3.56	36.18 [#] 3.27	33.09 3.33	36.27 3.13
3. Sitzung	50.60 4.85	47.64 5.00	45.88 4.76	48.04 4.18	29.60 3.78	24.74 2.52	24.54 2.73	26.29 2.80	39.89 3.39	35.96 3.20	34.99 3.10	36.94 2.93

6.2.2.1 Einzelne Finger

Die ermittelten Reliabilitäten (Tab. 12) der einzelnen Finger im EXP1 zwischen .54 bis .82 und im EXP2 im Bereich von .56 bis .88. Die mittleren Korrelationen variieren im EXP1 von .55 bis .73 und im EXP2 von .75 bis .82. Es fällt auf, daß die Koeffizienten im EXP2 bei allen Sitzungen und allen Fingern höher sind, als die Werte im EXP1 (Tab. 12).

Tab. 12: Retest-Reliabilitäten:

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
Zeigefinger						
2. Sitzung	.54**	.58**	.88***	.79***	.70***	.74***
3. Sitzung	.54**		.78***		.67***	
\bar{r}	.55		.82		.70	
Mittel-finger						
2. Sitzung	.60** (N=22)	.60** (N=23)	.82***	.81***	.75*** (N=47)	.74*** (N=48)
3. Sitzung	.67*** (N=23)		.56**		.73*** (N=48)	
\bar{r}	.62		.75		.74	
Ringfinger						
2. Sitzung	.82***	.70***	.80***	.78***	.83***	.79***
3. Sitzung	.67***		.69***		.74***	
\bar{r}	.73		.76		.79	

6.2.2.2 Gemittelte Finger

Die **Fingerkonsistenz** variiert von .79 bis .94 (Tab. 13). Besonders im EXP2 fällt auf, daß die Konsistenzen alle über .90 liegen und im Vergleich zum EXP1 wesentlich höher sind.

Tab. 13: Konsistenz

	EXP1 (N=24)	EXP2 (N=25)	EXP1 & EXP2 (N=49)

1. Sitzung	$\alpha = .79$	$\alpha = .90$	$\alpha = .86$
2. Sitzung	$\alpha = .82$	$\alpha = .94$	$\alpha = .91$
3. Sitzung	$\alpha = .82$	$\alpha = .92$	$\alpha = .89$

Für die gemittelten Finger ergeben sich Reliabilitäten für das EXP1 von .70 bis .78 und für das EXP2 .73 bis .90. Die durchschnittliche Korrelation liegt bei .74 (EXP1) und .83 (EXP2). Alle Reliabilitäten sind im EXP2 höher, als im EXP1. In beiden Experimenten ist die Korrelation der 1. und 2. Sitzung am höchsten (Tab. 14).

Tab. 14: Retest-Reliabilitäten

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.78***	.74***	.90***	.82***	.86***	.84***
3. Sitzung	.70***		.73***		.78***	
\bar{r}	.74		.83		.83	

6.3 Ergebnisse des Ischämieschmerzes

An den drei Experimentalsitzungen wurde jeweils einmalig die Schmerzschwelle bestimmt. Es folgen die Ergebnistabellen für Mittelwerte (M) und Standardfehler (SEM) (Tab. 15) und die Retest-Reliabilitäten (Tab.16).

Die **Mittelwerte** des EXP1 liegen im Bereich von 66.73 bis 69.88 sec. Damit sind sie deutlich höher als im EXP2, wo sich Koeffizienten von 51.76 bis 56.40 sec. ergaben.

Tab. 15: M und SEM der Schmerzschwellen

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	M	SEM	M	SEM	M	SEM
1. Sitzung	66.77	4.68	56.40	4.99	61.48	3.47
2. Sitzung	66.73	5.79	51.76	4.17	59.09	3.67
3. Sitzung	69.88	7.28	55.38	4.61	62.49	4.36

Die **Retest-Reliabilitäten** liegen für das EXP1 im Bereich von .84 bis .88, für das EXP2 im Bereich von .70 bis .94. Im Mittel betrug die Retest-Reliabilität im 1. Experiment .86 und im 2. Experiment .83. Die Koeffizienten der gemeinsamen Auswertung liegen auf Sitzungsebene zwischen den Werten der Einzelauswertung.

Tab.: 16: Retest-Reliabilitäten

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.88***	.84***	.70***	.94***	.79***	.88***
3. Sitzung	.85***		.75***		.79***	
\bar{r}	.86		.83		.82	

6.4 Ergebnisse des Hitzeschmerzes

An den drei Experimentalsitzungen wurde die Schwelle fünfmal gemessen. Das Mittel daraus wurde als individuelle Schmerzschwelle gewählt. Anschließend wurden dreimal fünf unterschiedliche Reize zur Diskrimination vorgegeben. Die Einschätzung der Reize erfolgte auf einer siebenstufigen

numerischen Skala (0-7). Es folgen die Ergebnistabellen für Mittelwerte (M) und Standardfehler (SEM), die Retest-Reliabilitäten und Konsistenzen der Schwellenmessung und der Skalierung.

6.4.1 Ergebnisse der Schwellenmessung

Die **Mittelwerte** der Sitzungen liegen sowohl für das EXP1, als auch für das EXP2 bei ca. 45 °C. (Tab. 17). Das gleiche Ergebnis zeigt auch die gemeinsame Auswertung.

Tab. 17: M und SEM der mittleren Schwellen

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	M	SEM	M	SEM	M	SEM
1. Sitzung	45.44	.31	45.20	.26	45.32	.20
2. Sitzung	45.56	.23	44.90	.27	45.22	.18
3. Sitzung	45.30	.31	45.39	.24	45.34	.19

Tab. 18: Konsistenz der fünfachen Schwellenmessung

	EXP1 (N=24)	EXP2 (N=25)	EXP1 & EXP2 (N=49)
	1. Sitzung	$\alpha = .99$	$\alpha = .98$
2. Sitzung	$\alpha = .97$	$\alpha = .99$	$\alpha = .98$
3. Sitzung	$\alpha = .99$	$\alpha = .99$	$\alpha = .99$

In Tabelle 18 sind die **Konsistenzen** der Schwellenmessungen aufgeführt. Sie liegen bezüglich der getrennten und der gemeinsamen Auswertung im Bereich von .97 bis .99.

Die **Retest-Reliabilitäten** schwanken im EXP1 von .49 bis .78, im EXP2 von .53 bis .71. Die durchschnittlichen Koeffizienten beträgt .61 im EXP1 und .64 im EXP2. Es fällt auf, daß die Korrelationen der 2. und 3. Sitzungen in beiden Experimenten vergleichsweise hoch ist (Tab. 19).

Tab. 19: Retest-Reliabilitäten:

	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.49*	.78***	.53**	.71***	.50***	.69***

3. Sitzung	.51*		.68***		.57***	
\bar{r}	.61		.64		.59	

6.4.2 Ergebnisse der Diskrimination

6.4.2.1 Vorüberlegungen zum Skalenniveau

Tabelle 20 zeigt die Mittel und den Standardmeßfehler der fünf Reize bei den einzelnen Sitzungen. Die Rangfolge der Reize wurde entsprechend ihrer Intensität weitestgehend beibehalten. Der unterschwellige Reiz (-0.8 °C), die Schwelle und der höchste überschwellige Reiz (2.4 °C) werden in allen Experimenten und bei allen Sitzungen rangskaliert. Bei den leicht überschwelligen Reizen (0.8 und 1.6 °C) geht im EXP1 bei der 2. und 3. Sitzung die Ordnung verloren. Der höhere Reiz wird im Mittel niedriger skaliert (3.43 bzw. 3.47), als der weniger intensive (3.64 bzw. 3.75).

Tab. 20: *M und SEM der mittleren Reize*

	EXP1 (N=24)			EXP2 (N=25)			EXP1 & EXP2 (N=49)		
	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung
- 0.8 °C	2.50 0.18	2.54 0.19	2.04 0.17	2.35 0.11	2.45 0.11	2.29 0.08	2.42 0.10	2.50 0.11	2.35 0.09
0.0 °C	3.49 0.19	3.35 0.13	3.49 0.16	3.15 0.11	3.03 0.09	3.15 0.13	3.31 0.11	3.18 0.08	3.31 0.10
0.8 °C	3.78 0.20	3.64 0.17	3.75 0.19	3.37 0.12	3.48 0.11	3.40 0.10	3.57 0.12	3.56 0.10	3.57 0.11
1.6 °C	4.01 0.31	3.43 0.27	3.47 0.26	4.16 0.13	4.05 0.12	4.00 0.13	4.09 0.16	3.75 0.15	3.74 0.15
2.4 °C	4.86 0.27	4.25 0.25	4.28 0.24	4.75 0.12	4.84 0.14	4.69 0.15	4.80 0.15	4.55 0.15	4.49 0.14

Beim Vorliegen einer Intervallskala sollten z. B. die Differenzen zwischen den Reizen 0.8 ° und 0 ° genauso groß sein, wie die Differenzen zwischen 0.8° und 1.6°. Tabelle 21 gibt Auskunft über die Differenzen der Skalenstufen bei den einzelnen Sitzungen. Es zeigt sich, daß jeweils am Ende der Reizskala die Differenzen deutlich größer sind (D1, D4), als im mittleren Bereich (D2, D3).

Tab. 21: *Mittlere Differenzen zwischen den Temperaturstufen, Angaben in °C*

	EXP1	EXP2	EXP1 & EXP2
--	------	------	-------------

	(N=24)			(N=25)			(N=49)		
	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung
D1	.99	.81	1.45	.80	.58	.86	.89	.68	.96
D2	.29	.29	.26	.22	.45	.25	.26	.38	.26
D3	.23	-.21	-.28	.79	.57	.60	.52	.19	.17
D4	.85	.82	.81	.59	.79	.69	.71	.80	.75

$D1=(0.0)-(-0.8)$

$D2=(0.8)-(0.0)$

$D3=(1.6)-(0.8)$

$D4=(2.4)-(1.6)$

N = 24

Hitzeskalierung - Mittel aller Sitzungen, EXP1

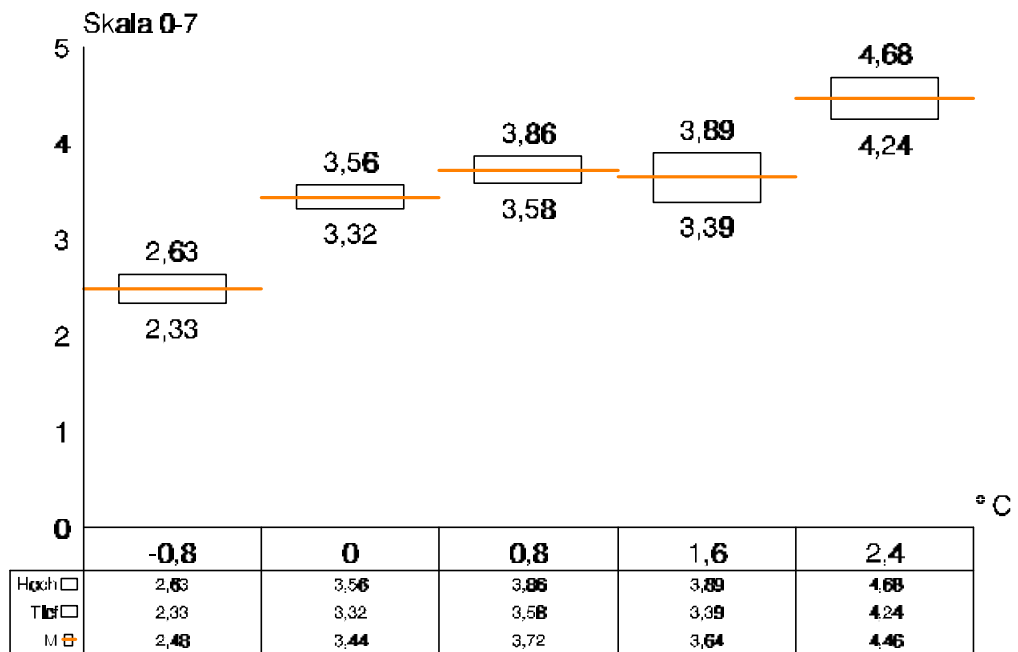


Abb. 9: Für EXP1 werden die Mittel (M) der Hitzeskalierung unabhängig vom Sitzungszeitpunkt angegeben, sowie ihre Streubreite (Hoch, Tief).

N = 25

Hitzeskalierung - Mittel aller Sitzungen, EXP2

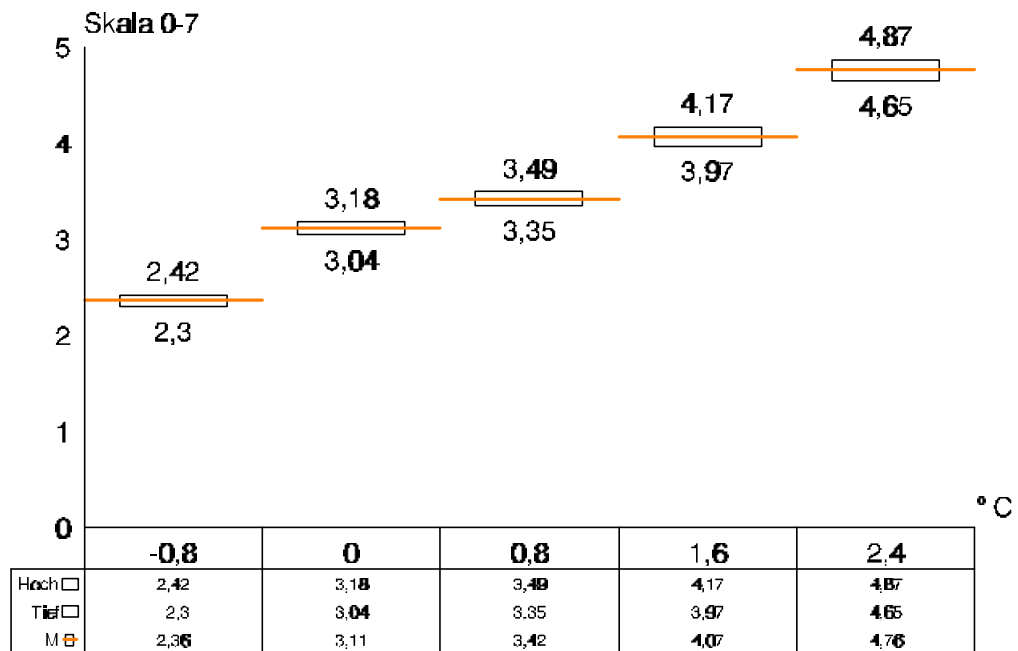


Abb. 10: Für EXP2 werden die Mittel (M) der Hitzeskalierung unabhängig vom Sitzungszeitpunkt angegeben, sowie ihre Streubreite (Hoch, Tief).

N = 49

Hitzeskalierung - Mittel aller Sitzungen, EXP1 & EXP2

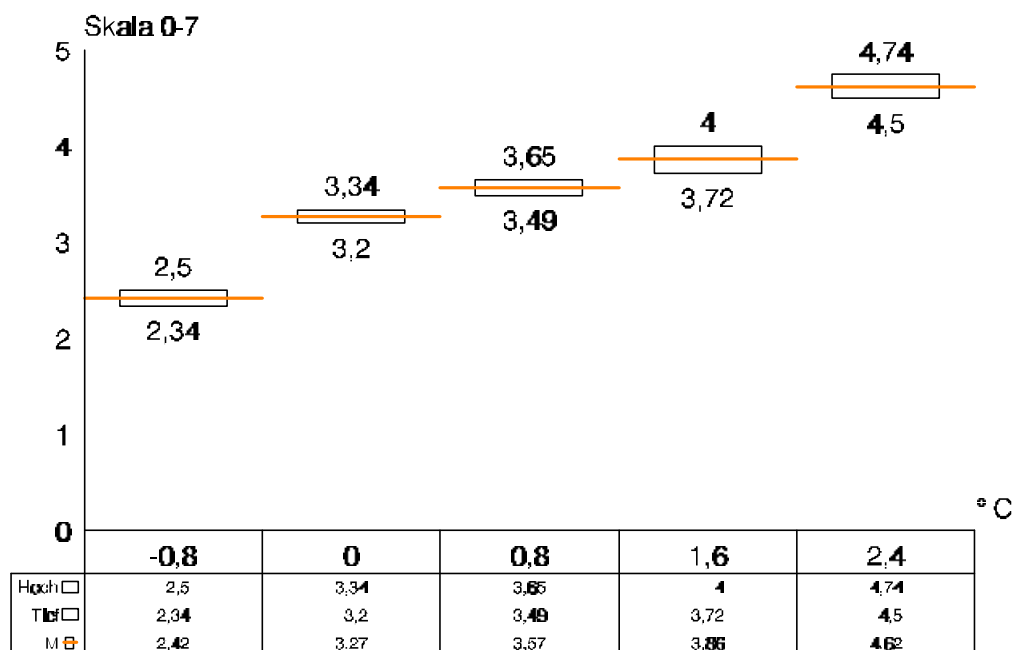


Abb.11: Für EXP1 & EXP2 werden die Mittel (M) der Hitzeskalierung unabhängig vom Sitzungszeitpunkt angegeben, sowie ihre Streubreite (Hoch, Tief).

Die Abbildungen 9 bis 11 geben unabhängig von den Sitzungen Auskunft über das Skalenniveau der Hitzediskriminationsversuche. Im EXP1 und bei EXP1 & EXP2 ist erkennbar, daß die Differenzen zwischen den Reizkategorien zum Ende der jeweiligen Skala hin größer werden. Bei EXP2 ist dieses Bild nicht so deutlich. Bis auf die Differenz zwischen der Schwelle und 0.8 °C beträgt der mittlere Anstieg zwischen den anderen Kategorien ca .7

6.4.2.2 Ergebnisse der Skalierung

Die **Konsistenzen** der Skalierungen schwanken beträchtlich (Tab. 22). Im ersten Experiment variiert sie von .54 bis .82 für -0.8 °C, von -.03 bis .57 für 0.0 °C, von .72 bis .74 für 0.8 °C, von .79 bis .91 für 1.6 °C und von .81 bis .90 für 2.4 °C. Im zweiten Experiment liegen die Konsistenzen etwas niedriger: .23 bis .66 für -0.8 °C, von -.15 bis .54 für 0.0 °C, von .49 bis .58 für 0.8 °C, von .52 bis .63 für 1.6 °C und von .69 bis .84 für 2.4 °C.

Tab. 22: Konsistenz der dreifachen Hitzeskalierung

	EXP1 (N=24)			EXP2 (N=25)			EXP1 & EXP2 (N=49)		
	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung
- 0.8 °C	.82	.54	.77	.66	.60	.23	.75	.55	.66
0.0 °C	.57	-.03	.41	.19	-.15	.54	.47	-.17	.44
0.8 °C	.77	.72	.74	.58	.49	.56	.73	.61	.69
1.6 °C	.91	.79	.86	.52	.63	.56	.84	.77	.81
2.4 °C	.90	.84	.81	.69	.71	.84	.86	.78	.77

Tab. 23: Mediane der Mediane der dreifachen Einschätzung

	EXP1 (N=24)			EXP2 (N=25)			EXP1 & EXP2 (N=49)		
	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung	1. Sitzung	2. Sitzung	3. Sitzung
- 0.8 °C	2	2	2	2	2	2	2	2	2
0.0 °C	3	3.5	4	3	3	3	3	3	3

0.8 °C	4	3.5	4	3	3	3	4	3	3
1,6 °C	4.5	3	3,5	4	4	4	4	4	4
2.4 °C	5	4	4	5	5	4	5	5	4

Bei *Ordinalskalen* kann der Medianverlauf betrachtet werden. Tabelle 23 zeigt, daß der Median des unterschwelligen Reizes konstant 2 ist. Die Schätzungen für den überschwelligen Reiz liegen bis auf zwei Ausnahmen mit 4 oder 5 oberhalb der Mediane von 1.6 °C. Die Mediane im EXP1 der mittleren Hitzereize von 0.8 °C liegen mit 3.5 bzw. 4 höher als die Einschätzungen von 1.6 °C mit 3 bzw. 3.5. Im EXP2 ist auffällig, daß die Schwelle und 0.8 °C bei allen Sitzungen mit 3 eingestuft werden.

Die **Retest-Reliabilitäten** der Skalierung variieren im Schnitt im ersten Experiment von .37 (-0.8 °C) bis .78 (1.6 °C), wobei die höchsten Koeffizienten für die überschwelligen Temperaturen (1.6 und 2.4 °C) mit .78 und .71 erreicht werden. Im zweiten Experiment streuen die mittleren Korrelationen in einem Bereich von .03 (Schwelle) bis .48 (2.4 °C). Die Höhe der Korrelationen ist für jede Reizintensität im EXP2 geringer als im EXP1. Zudem werden beim Schwellenreiz im EXP2 zwei negative Korrelationen zwischen der 1. und 3. Sitzung (-.14) und zwischen der 2. und 3. Sitzung (-.15) ermittelt (Tab. 24). Die mittleren Korrelationen der gemeinsamen Auswertung von EXP1 & EXP2 liegen immer zwischen der mittleren Korrelationen der einzelnen Auswertung. Bei der Reizintensität von 0.8 °C im EXP1 & EXP2 werden im Vergleich zu EXP1 die Korrelationen, die zwischen .30 und .40 liegen signifikant.

Tab. 24: Retest-Reliabilität:

-0.8 °C	EXP1 (N=24)		EXP2 (N=25)		EXP1 & EXP2 (N=49)	
	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.42*	.49*	.41*	.00 n.s.	.42**	.35*
3. Sitzung	.23 n.s.		.18 n.s.		.23 n.s.	
\bar{r}	.38		.20		.33	
0.0 °C						
1. Sitzung		3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.21 n.s.	.77***	.37 n.s.	-.15 n.s.	.30*	.43**
3. Sitzung	.19 n.s.		-.14 n.s.		.10 n.s.	
\bar{r}	.44		.03		.28	
0.8 °C						
1. Sitzung		3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.36 n.s.	.39 n.s.	.38 n.s.	.18 n.s.	.38**	.33*
3. Sitzung	.36 n.s.		.02 n.s.		.30*	
\bar{r}	.37		.20		.34	
1.6 °C						
1. Sitzung		3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.71***	.88***	.34 n.s.	.41*	.56***	.77***
3. Sitzung	.70***		.17 n.s.		.53***	
\bar{r}	.78		.31		.63	
2.4 °C						
1. Sitzung		3. Sitzung	1. Sitzung	3. Sitzung	1. Sitzung	3. Sitzung
2. Sitzung	.57**	.80***	.62**	.52**	.54***	.75***
3. Sitzung	.74***		.27 n.s.		.59***	
\bar{r}	.71		.48		.64	

7 Diskussion

Die Gliederung der Diskussion folgt dem Aufbau des Ergebnisteils. Zuerst werden die methodenübergreifenden Ergebnisse und dann die Ergebnisse der einzelnen Schmerzinduktionsmethoden

diskutiert. Anschließend werden alle Teile in einer zusammenfassenden Diskussion integriert und Fragestellungen diskutiert, die alle Verfahren betreffen. Im Verlauf der Diskussion werden Verbesserungsvorschläge formuliert.

7.1 Diskussion der methodenübergreifenden Ergebnisse

Die Voraussetzung der Normalverteiltheit der Variablen ist in beiden Experimenten gegeben. Bei den Berechnungen, in welchen die Variablen, die *nicht* normalverteilt sind und bei den Hitzeskalierungen dürfen also die als signifikant gewordenen Ergebnisse nicht als ungleich von Null interpretiert werden. Allerdings steht bei der Reliabilitätsbestimmung die Stärke des Zusammenhangs, also die Höhe der Koeffizienten im Vordergrund, deren Ergebnis von der Normalverteiltheit unbeeinflusst ist.

Mit der gewählten Statistik ergaben sich im EXP1 signifikante **Versuchsleitereffekte** bei allen Hitzeschwellenmessungen und den Hitzeschwellenmittel Für das zweite Experiment konnten Effekte für die Hitzeskalierung ermittelt werden. Dies zeigt, daß insbesondere das Hitzeschmerzverfahren sensibel gegenüber äußeren Einflüssen, wie der Anwesenheit verschiedener Versuchsleiter, ist. Ein geschlechtsspezifischer Einfluß, wie Levine und Simone (1991) berichteten, kann nicht nachgewiesen werden. Die gemischtgeschlechtliche Versuchsleitung im EXP2 führt zu keinen deutlicheren Unterschieden bei den Variablenausprägungen, als die gleichgeschlechtliche Versuchsleitung im EXP1. Trotzdem fällt auf, daß im EXP1 bei der 1. Versuchsleiterin die Schwellentemperatur für die Hitzereize immer höher ist, als die Schwellentemperatur, die bei der 2. Versuchsleiterin ermittelt wurde. Das hieße, daß die Schwellen bei der 1. Versuchsleiterin später wahrgenommen werden als bei der zweiten Versuchsleitern. Eine ähnliche *Tendenz* zeigt sich auch im 2. Experiment. Hier treten bei der weiblichen Versuchsleiterin (VL1). immer höhere Werte auf, als bei dem männlichen Versuchsleiter (VL 2). Es bleibt die Vermutung, daß bereits das individuelle Erscheinungsbild eines Experimentators eine Messung beeinflussen kann, wobei bei EXP2 die persönliche Erscheinung natürlich mit dem Geschlecht konfundiert ist. Insgesamt kann angenommen werden, daß *methodenspezifisch* Versuchsleitereffekte möglich sind. Allerdings sollten die gefunden Effekte nicht überbewertet werden. Durch die Halbierung der Stichprobe ist es möglich, daß den Befunden ein Teilstichprobenproblem zu Grunde liegt. In der Gruppe der ersten Versuchsleiter könnten zufällig gerade diejenigen Probanden getestet worden sein, die im Vergleich zu den anderen Teilnehmern weniger schmerzempfindlich sind. Um ein Teilstichprobenproblem zu vermeiden, ist es also wichtig, ausreichend viele Probanden zu untersuchen.

Da Schmerzinduktionsmethoden vorgeben, das gleiche Merkmal - den Schmerz - zu messen, gibt der **intermodale Vergleich** zum einen Aufschluß darüber, ob die eine Methode als Kriterium herangezogen werden, um die Validität einer anderen Methode zu stützen. und ob zum anderen gegebenenfalls aus Effizienzgründen auf eine einzige Methode zurückgegriffen werden könnte. Bei den

hier gewählten Verfahren kann nur die Schmerzschwelle als Vergleichskriterium herangezogen werden, da sie bei allen Methoden erhoben wurde. Das Ergebnis ist für die verschiedenen Korrelationen heterogen. *Ischämie und Hitzeschmerz* sind unkorreliert, *Ischämie und Druckschmerz* korrelieren höher und für *Hitze und Druckschmerz* ist das Bild sowohl innerhalb eines Experimentes als auch zwischen den Experimenten sehr uneinheitlich. Das Resultat zeigt, daß die verschiedenen Methoden in diesen Untersuchungen *nicht* denselben Schmerz erfassen und nicht zu ihrer gegenseitigen Validierung dienen können. Bei Ischämie und Hitze ist das Bild eindeutig, für Ischämie und Druck liegen die Korrelationen zwar höher, sind aber mit $< .30$ zu gering, um eine hohe Ähnlichkeit zu repräsentieren. Die gegensätzlichen Ergebnisse bei der Hitze-Druck-Korrelation macht für diesen Vergleich eine Interpretation sehr schwierig. Im EXP1 wäre die Annahme der Unkorreliertheit eher gerechtfertigt, im EXP2 eher eine hohe Ähnlichkeit. Die Ergebnisse des Ischämie-Hitzeschmerzvergleiches sind kongruent mit denen von Janal et al. (1994). Auch diese Autoren ermittelten für den intermodalen Vergleich von Ischämie und thermischen Reizen „Null“-Korrelationen. Göbel und Westphal (1989) fanden Korrelationen zwischen Druckschmerz und Ischämie Korrelationen von .25 bis .77. Diese waren abhängig von den Kategoriengrenzen. Für die Kategoriengrenze 10, die der Schmerzschwelle entsprach, betrug die Korrelation .64. Ihre Ergebnisse deuten auf einen wesentlich engeren Zusammenhang zwischen diesen Verfahren hin, wohingegen bei den hiesigen Experimenten nur ansatzweise ein Zusammenhang erkennbar ist. Zusammenfassend ist ersichtlich, daß die Methoden nicht austauschbar sind und daß eine Versuchsperson nicht generell mehr oder weniger schmerzempfindlich ist. Für umfangreichere Aussagen ist es sinnvoll, für die Schmerzbeobachtung weiterhin verschiedene Modalitäten heranzuziehen.

7.2 Diskussion der Ergebnisse des Druckschmerzes

Werden die **Retest-Reliabilitäten** der *Schwellenmessung* bei den einzelnen Fingern nach den Kriterien von Wolff (1986) beurteilt, nach denen Konsistenzen von .60 bis .80 als akzeptabel gelten und nach denen für Retest-Korrelationen auch geringere Werte geeignet sind, so hat sich die Druckschmerzmethode als sehr reliables Verfahren erwiesen. Insgesamt kann nach den genannten Kriterien auch die *Toleranzmessung* als sehr reliabel betrachtet werden. Die Retest-Reliabilität der Toleranzmessung ist im Vergleich zur Schwellenmessung im EXP2 sowohl für alle einzelnen Finger als auch für die gemittelten Finger deutlich höher. Das ist umso mehr erstaunlich, da die Toleranzschwelle als leichter beeinflussbar gilt und zudem von höheren Funktionen abhängig ist, wie von der affektiven und bewertenden Komponente (s. Price, 1988). In diesem Sinn zeigt sich in EXP1 ein anderes Bild: bezogen auf die gemittelten Finger ist die Schwellenmessung reliabler. Trotzdem scheint sich aber im experimentellen Kontext die Beeinflussbarkeit der Toleranzschwelle nicht nachteilig auszuwirken. Die Untersuchungen von Curio (1990) und Göbel und Westphal (1989) belegen, daß bei verschiedenen

Methoden die Toleranzschwelle ebenso reliabel ist wie die Schwelle. Bei Fernandez (1990) ist die Toleranzmessung sogar reliabler als die Schwellenmessung. Das Ergebnis in den hier beschriebenen Experimenten spiegelt die unterschiedlichen Befunde, die sich in der Literatur finden, wider. Es gibt daher keinen Grund, auf die Erhebung der Toleranzschwelle aus Gründen der mangelnden Reliabilität zu verzichten.

Die **Konsistenz** der *Schwellenmessung* für die Finger als Realisierung einer einzigen Variable kann als sehr gut angesehen werden. Das gleiche gilt für die *Toleranzmessung*. Besonders im EXP2 fällt auf, daß die Konsistenzen alle über .90 liegen und als „exzellent“ gelten können. Dieses Ergebnis ist dementsprechend kongruent mit der höheren Reliabilität der Toleranzschwelle im 2. Experiment. Die hohe Konsistenz ist zudem ein Indiz dafür, daß Schmerzreize an verschiedenen Fingern induziert werden können, ohne durch die größere Varianz eine zu hohe Ungenauigkeit in Kauf nehmen zu müssen. Bei der Auswertung können die einzelnen Finger als Einheit betrachtet werden, da sie als solche eine genauere Schätzung der Schwellen ergeben als dies durch die Einzelauswertung der Finger möglich ist. Das Vorgehen, die Schmerzschwelle auf Basis mehrerer Finger zu ermitteln, bietet zudem den Vorteil, daß eine Sensibilisierung eines Fingers verhindert wird.

Beim **Vergleich der Mittelwerte** der Schmerzschwelle fällt im EXP1 für alle einzelnen Finger auf, daß die Schmerzschwelle von der ersten zur dritten Sitzung steigt. Im EXP2 zeigt sich diese Tendenz nicht. Beim Ringfinger sinkt die Schwelle, während sie bei den anderen Finger über die Sitzungen schwankt. Für EXP1 & EXP2 zeigt sich die Beobachtung aus EXP1 in abgeschwächter Form. Die einzelnen Finger unterscheiden sich nicht auffällig. Trotz der immer gleichen Induktionsreihenfolge gibt es keinen Finger, der eine wesentlich höhere Schwelle hat. Beim Vergleich der Mittelwerte der Toleranzmessung ist kein besonderer Anstieg oder Abfall sowohl in der Sitzungsreihenfolge, als auch in der Abfolge der Finger ersichtlich. Die konstante Induktionsreihenfolge der Finger hat insofern keinen systematischen Effekt auf die Resultate. Es wäre trotzdem sicherer bei einem Folgeexperiment diese Fehlerquelle im voraus auszuschließen, indem man eine zufällige Wahl der Fingerreihenfolge trifft. Der Anstieg der Schmerzschwelle im EXP1 könnte einen Transfereffekt auf die Baseline-Messung der 2. und 3. Sitzung zur Folge haben. Da im EXP2 kein Anstieg erfolgte wäre denkbar, daß im EXP1 das Treatment Vasopressin ungewöhnlicherweise die Baselines der jeweils 14 Tage später folgenden Sitzungen noch beeinflußt.

Bei allen Fingermessungen fällt auf, daß alle Schwellen im EXP2 deutlich höher liegen als im EXP1. Noch stärker ist diese Diskrepanz im bei den Toleranzschwellen. Im EXP1 liegen die Toleranzschwellen in allen Sitzungen und bei allen Fingern nahezu doppelt so hoch wie in EXP2. Wenn der niedrigste Schwellenwert und die höchste Toleranz als Grundlage betrachtet werden, so ist die Streubreite der Schmerzempfindung im EXP1 wesentlich breiter (5.88 bis 53.61 sec.) als im EXP2 (8.08 bis 29.60 sec.). Über mögliche Ursachen kann nur spekuliert werden, da in der Literatur keine vergleichbaren Ergebnisse berichtet wurden. Möglicherweise ist der unterschiedliche Sitzungsbeginn der

beiden Experimente entscheidend. Demnach würde morgens die Schwelle höher liegen, während die Toleranz eher erreicht wird. Auf der anderen Seite wäre die Schwelle am Nachmittag niedriger, aber der Schmerz länger auszuhalten. Eine andere Erklärung bietet die „Sättigung“ der Versuchspersonen. Bezogen auf die klareren Resultate der Toleranzmessung sind die Probanden im EXP2, die hungrig zur Sitzung erschienen, wesentlich schmerzempfindlicher, als die Versuchspersonen im EXP1, denen kleine Mahlzeiten erlaubt war. Wie dem auch sei, ist das Ergebnis ein weiterer Beleg dafür, daß die Interpretation einer Schwelle nur aus dem Setting heraus geschehen sollte, in welchem sie gemessen wurde. Eine reliable Messung darf nicht mit der Erfassung der „wahren“ Schwellen gleichgesetzt werden.

7.3 Diskussion der Ergebnisse des Ischämieschmerzes

Die Schmerzschwelle wurde einmalig zu jeder Sitzung bestimmt. Die **Retest-Reliabilitäten** liegen zum größten Teil über .80. Nach den Kriterien von Wolff (1986) können die Koeffizienten als sehr gut bezeichnet werden. Die Ischämieschmerzmethodik sich somit reliables Verfahren erwiesen. Das ist insofern verwunderlich, als daß die Reliabilität dieses Verfahren umstritten ist (vgl. Punkt 2.2.1). Auch Fernandez (1990) fand für verschiedene Skalentypen teils sehr gute (.94), teils sehr schlechte (.54) Reliabilitäten. Er argumentierte, daß „offene“ Skalen reliabler sind als „geschlossene“ Skalen. Da aber in seiner Untersuchung der Schwellenwert für beide Skalen ein und derselbe ist, dürfte die Skalenart die Schwellenmessung nicht beeinflussen. Auch bei den hier beschriebenen Experimenten war nur der untere Skalenbereich des Schieberegler maßgeblich, so daß die Skala keine bedeutende Fehlerquelle sein kann. Göbel und Westphal (1989) fanden für den Ischämieschmerz Konsistenzen von .95. Dieses Ergebnis ist ebenfalls ein Kennzeichen für die gehobene Güte der Ischämieschmerzmethodik.

Die **Mittelwerte** des EXP1 sind deutlich höher als im EXP2. Dieses Ergebnis spricht wiederum dafür, daß auch die Höhe der Ischämieschmerzschwelle eine *relative* Größe ist. Dieses Ergebnis ist übereinstimmend mit den Resultaten der Mittelwerte bei der Druckschmerzmethodik. Es ist anzunehmen, daß die gleichen Unterschiede im experimentellen Aufbau für die Diskrepanz maßgeblich sind. Auch bei dem Ischämieschmerz wären demnach die Versuchspersonen des EXP2 weniger schmerzempfindlich, als die Probanden im EXP1.

7.4 Diskussion der Ergebnisse des Hitzeschmerzes

Die Hitzeschmerzschwelle wurde fünfmal gemessen. Die **Konsistenzen** sind allesamt exzellent. Dieses Ergebnis ist insofern nicht erstaunlich, da zwischen den Einzelmessungen nur Sekundenabstände lagen. Äußere Einflüsse sind somit fast gänzlich auszuschließen. Die Konsistenzen, die Curio (1990) in seiner Untersuchung ermittelte, lagen bedeutend niedriger bei .83 bzw. .81. Ein Nachteil bei seinem

Experiment mag die Größe der Thermode sein, die nur 1 cm² betrug, während bei dem EXP1 und dem EXP2 eine Größe von 4 cm² gewählt wurde. Wie bereits Price et al. (1992) zeigten, erhöht sich mit zunehmender Thermodengröße die räumliche Summation und damit die Wahrnehmungswahrscheinlichkeit eines Reizes. Wenn die Reize nicht zu schmerzhaft werden, hätten größere Auflageflächen gegenüber kleineren Vorteile. Curio (1990) fand im Vergleich zu den bei ihm niedrigeren Konsistenzen hohe Retest-Reliabilitäten von .72 und .83. In unseren Untersuchungen ist das überraschender Ergebnis, daß die **Retest-Reliabilitäten** relativ niedrig sind. Die durchschnittlichen Koeffizienten sind mit .61 (EXP1) und .64 (EXP2) gerade noch akzeptabel. Zusammenfassend scheint die hohe Konsistenz der Messung innerhalb einer Sitzung nicht mit einer hohen individuellen Konstanz der Schmerzschwelle einherzugehen. So ist die Schwelle im Gruppenmittel reproduzierbar, während es für den einzelnen Probanden schwierig ist, von einer hohen Schwelle in der ersten Sitzung, auf die Höhe ihrer Schwelle in einer anderen Sitzung zu schließen.

Die **Mittelwerte** der Sitzungen sowohl für das EXP1, als auch für das EXP2 im Schnitt bei ca. 45 °C. Die mittlere Schmerzschwelle ist somit bezüglich beider Experimentalgruppen sehr konstant. Dies ist übereinstimmend mit der Erkenntnis, daß im allgemeinen bei Temperaturen von 45 °C Schmerz empfunden wird, was für die Existenz einer allgemeinen Hitzeschmerzschwelle spricht, die stark an den Zustand der Nozizeptoren gekoppelt ist.

Im Vordergrund der Auswertung der **Diskrimination** stand zunächst die Frage, welches Skalenniveau die numerische Skala erreicht. Die mittleren Differenzen der Skalenstufen sind jeweils am Ende der Reizskala deutlich größer als im mittleren Bereich. Diese Bild ergibt sich sowohl für die einzelnen Sitzungen (Tab.21), als auch für der sitzungsunabhängigen Betrachtung. Im Hinblick auf das Skalenniveau liegt eindeutig *keine* Intervallskala vor. Allerdings weist der Skalenverlauf über alle Sitzungen im EXP2 (Abb. 10) daraufhin, das mit der gegebenen numerischen Skala auch Intervallskalenniveau erreichbar wäre. Sowohl auf Sitzungsebene, als auch sitzungsübergreifend wurde Ordinalskalenniveau erreicht, da die Rangordnung der Reize entsprechend ihrer Intensität weitestgehend beibehalten wurde. Ausnahmen treten ausschließlich nur bei den bei den leicht überschwellig Reizen im Exp1 auf. Auch beim Mittel über alle Sitzungen liegt im EXP1 die mittlere Schätzung der Intensitäten von 1.6 °C mit 3.39 unter dem Mittel von 0.8 °C mit 3.58 (Abb. 9). Diese Bild zeigt sich bei den gleichen Sitzungen auch im Vergleich der Mediane. Die mittlere Schätzung von 0.8 °C liegt überder Einschätzung von 1.6. In dieser Übersicht ist gut zu erkennen, daß die Schwelle und die schwellennahen Reize sehr schlecht diskriminiert werden können. So werden im EXP2 die Schwelle und 0.8 °C gleichhoch eingestuft. Eindeutige Unterscheidungen treffen die Versuchspersonen nur für den unterschwellig Reiz und zum größten Teil für den überschwelligsten Reiz. Zusammenfassen ist festzuhalten, daß nur die extremen Reize (-.08 °C , 2.4 °C) zufriedenstellend diskriminiert werden können. Für die Schwelle und die leicht überschwellig fällt es den Versuchspersonen schwer, diese klar zu unterscheiden. Hier liegt mit Sicherheit eine Quelle für Ungenauigkeiten, da die Einschätzung des Schwellenreizes nicht eindeutig ist. Möglicherweise wird im Laufe einer Sitzung von der Versuchsperson ein höherer Reiz von 08 °C als

Schwelle interpretiert und als Maßstab für die weitere Skalierung genommen, was insofern zu Fehleinschätzungen führt.

Die **Konsistenz** der Skalierung variiert außerordentlich. Allgemein können alle Koeffizienten im EXP1 bis auf die Schwelle als akzeptabel gelten, während im EXP2 dieses Urteil nur für den höchsten Reiz zutrifft. Alle anderen Konsistenzen sind nicht annehmbar. Eine übereinstimmende Schätzung gleicher Reize mit gleichen Zahlen ist also für die Probanden eine schwierige Aufgabe. Auffallend ist, daß die negativen Konsistenzen ausschließlich bei der wiederholten Schwellenschätzung auftreten. Auch daraus läßt sich schließen, daß die Wahrnehmung und Erinnerung der eigenen Schwellentemperatur bereits vage ist. Einschätzungen die sich an dieser Empfindung orientieren sollen, müßten demnach unvorhersagbar schwanken. Dieses Ergebnis ist demzufolge passend mit den Befunden zur Skalierung, wo auch eine schlechte Schwellendiskrimination ersichtlich wurde.

Die **Retest-Reliabilitäten** der Skalierung sind bis auf die Intensitäten von 1.6 °C und 2.4 °C im EXP1 insgesamt nicht akzeptabel. Allerdings weisen die teils guten Ergebnisse bei den überschwelligen Temperaturen im ersten Experiment darauf hin, daß durch Verbesserungen der Operationalisierung, indem beispielsweise höhere Reizintensitäten verwendet werden könnten, reliablere Hitzeschmerzmessungen möglich sind. Die zwei negative Korrelationen bei der Schwellenmessung weisen darauf hin, daß nicht nur der Ankerreiz inkonsistent ist, sondern daß auch die Schwelle dementsprechend schlecht reproduziert werden kann. Dies gilt für das 2. Experiment, in welchem die Retest-Reliabilität der Schwelle nahezu Null beträgt. So könnte auch erklärt werden, daß die absolute Höhe der Korrelationen bei allen Reizintensitäten bedeutend niedriger ausfällt als im ersten Experiment.

7.5 Zusammenfassende Diskussion

Als Fazit auf die Frage, ob die drei Induktionsmethoden reliabel sind, ist festzuhalten, daß die Ischämieschmerz und die Druckschmerzmethode als reliabel angesehen werden können, während die Reliabilität des Hitzeschmerzverfahrens zwar akzeptabel, aber im Vergleich zu den anderen Verfahren stark verbesserungsbedürftig ist. Trotz der Unterschiede in der experimentelle Anordnung kann dieses generelle Fazit für beide Experimente gelten. Insofern kann die eine Studie als Replikation der anderen Studie gewertet werden. Das ist insofern bemerkenswert, da Verfahren eine reliable Messung ermöglichen, obwohl die absolute Höhe der Schmerzschwellen beim Ischämieschmerz und der Druckschmerztoleranz zwischen den Experimenten stark voneinander abweichen. Beim Hitzeschmerz sind die Schwelle im Gruppenmittel in beiden Experimente fast gleich, trotzdem ermöglicht das eingesetzte Verfahren keine reliable Messung. Anscheinend ist die Hitzeschmerzschwelle weitaus sensibler und von Störquellen beeinflusbarer, als die Schwellen bei den anderen Verfahren. Dies mag mit

der neurophysiologischen Reizverarbeitung zusammenhängen. So hätten Änderungen der Hauttemperatur durch die Raumtemperatur („kalte Hände“), Jahreszeit oder angstbedingtem Schwitzen bei der Versuchsdurchführung eher Einfluß auf die thermale Schmerzschwelle, als beispielsweise auf die ischämische Schwelle, bei der vornehmlich tiefgelegene Rezeptoren in den Muskeln erregt werden. Zenz und Jurna (1993) bemerken, daß nicht die Hauttemperatur sondern die Temperatur der Nozizeptoren entscheidend für deren Erregungswahrscheinlichkeit sei. In ihren Augen ist es sinnvoll, die nozizeptive Temperatur mittels Kanülen zu bestimmen. Ein andere Erklärung für die mangelnde Reliabilität der Hitzeinduktionsmethode wäre ein Reihenfolgeeffekt. Da dieses Verfahren immer zuletzt eingesetzt wurde, ist es möglich, daß die vorhergehende Messung von Druck- und Ischämieschmerz die Messung der Hitzeschmerzschwelle systematisch beeinflusste. Um diese Gefahr auszuschließen, ist es wichtig, bei weiteren Untersuchungen, die Abfolge der Verfahren zu variieren.

Bei der Ischämie- und Druckschmerzmethode hat sich der zur genauen Einschätzung der Schmerzempfindung eingesetzte Schieberegler als ein Hilfsmittel erwiesen, mit dem die Schwellen gut skaliert werden können. Sein Vorteil besteht darin, daß durch die visuelle Rückmeldung der Skala auf dem Verstärker eine gezieltere Konzentration auf das eigene Empfinden möglich ist. So ist durch die direkte Wiedergabe der Empfindung durch die kontinuierliche Handbewegung eine unkomplizierte, leicht handhabare Aufgabe. Die Skala zur Erfassung der Hitzediskrimination hat sich als problematisch herausgestellt. Erstens ist der Grund unklar, warum die Skala von 0 bis 7 erreicht, obwohl nur fünf unterschiedliche Hitzereize vorgegeben werden. Sinnvoller wäre es ein Vielfaches der Anzahl der Reize zu wählen. Die Intensität der Reize ist im leicht überschwelligen Bereich schlecht zu unterscheiden. Hier ist es sinnvoll, Reizintensitäten zu wählen, die zum einen weiter auseinander liegen, und zum anderen *stärker* sind. Vorstellbar ist beispielsweise: -1.5°C , die Schwelle und 1.5°C , 3.0°C und 4.5°C oberhalb der Schwelle. Unter der Voraussetzung, daß die Versuchspersonen keine Schwellen über 47°C haben, wären die Probanden maximal einer Temperatur von 51.5°C ausgesetzt. Eine Reizhöhe von 51°C wurde bereits bei anderen Experimenten ohne ethische Probleme angewandt (z. B. Douglass et al., 1992; Price et al., 1992). Zudem scheint die Schwelle selbst schlecht als Ankerreiz erinnert werden zu können. Um dem Problem entgegenzutreten könnten die jeweiligen Schwellen öfters vorgegeben werden. Denkbar ist, daß vor jedem Diskriminationsvorgang die Schwelle mehrmals zur Einprägung wiederholt appliziert wird, um so die Assoziation der Intensität und Zahl zu stärken.

Eine andere Fragestellung war, ob es gerechtfertigt ist, beide Experimente zusammen auszuwerten. Wie die Ergebnisse der Mittelwerte des Druckschmerzes beider Schwellen, des Ischämieschmerzes und die meisten Reliabilitäten bei allen drei Verfahren zeigen, liegen die Koeffizienten der gemeinsamen Auswertung zum größten Teil ungefähr in ihrer Höhe zwischen der Reliabilitäten der einzelnen Experimente. Die Reliabilität der Meßmethoden verbessert sich nicht, was darauf hindeutet, daß die wesentlichen Unterschiede zwischen beiden Experimenten bestehen. Gerade diese Unterschiede aber werden durch die gemeinsame Betrachtung verschleiert, wodurch wichtige Informationen verloren gehen würden. So hat auch die Vergrößerung der Stichprobe um nahezu das

Doppelte (N=49) hat keine Auswirkung auf die Reliabilität. Im Gegenteil in Hinsicht auf die Stichprobengröße hat es sich erwiesen, daß 24 bzw. 25 Versuchspersonen vollkommen ausreichend sind.

Alle Verfahren wurden an drei aufeinanderfolgenden Sitzungen durchgeführt. Die Mehrfachmessung bei der Ischämieschmerz zu einem widersprüchliches Ergebnis Im EXP2 erhöht die dreimalige Messung die Reliabilität von .70 auf durchschnittlich .83, während die Reliabilität im EXP1 leicht von .88 auf .86 sinkt. Eine eindeutige Aussage zum Effekt Mehrfachmessung ist für diese Methode nicht möglich. Bei den anderen Verfahren fällt auf, daß nur in ganz wenigen Fällen die höchsten Retest-Reliabilitäten zwischen der ersten und zweiten Sitzung zu finden sind, sondern eher zwischen der 1. und 3. Sitzung oder zwischen der 2. und 3. Sitzung. Von daher hat sich die dritte Sitzung vorwiegend positiv auf die Reliabilitätsbestimmung ausgewirkt. Ich denke, daß generell Mehrfachmessungen eine umfangreichere Grundlage zur Bestimmung und Interpretation der Reliabilität bieten und deswegen unumgänglich sind.

Die Reliabilität ist von Objektivität und der Validität des gewählten Verfahrens abhängig. Die *Validität* ist für alle Methoden gegeben, da die drei Methoden in vielen Untersuchungen bewiesen haben, Schmerz zu erfassen. Eine genaue Aussage zur kriterienbezogenen Validität wäre durch eine korrelative Analyse mit einem Außenkriterium (z. B. EXP1 mit EXP2) möglich. Die *Durchführungsobjektivität* ist eingeschränkt. Wie beschrieben mag ein Versuchsleiter zu anderen Ergebnissen kommen, als ein anderer. Zudem sind die genauen Instruktionen den Versuchspersonen mündlich mitgeteilt worden. Dies kann durch eigene Formulierungsabwandlungen leicht zu Ungenauigkeiten führen. Sicherer wäre es, den Probanden die Instruktionen schriftlich vorzulegen und vor jeder Messung nochmals lesen zu lassen. Der Ablauf der Sitzungen ist sehr ökonomisch. Die Meßanordnung ist für die Versuchspersonen durch die Screening-Sitzung überschaubar und für die Versuchsleiter durch die standardisierten Ablauf leicht umzusetzen. Der Zeitaufwand von maximal 2,5 Stunden ist den Probanden zuzumuten, Unmut sollte nicht aufkommen.

Ich fasse noch mal alle Anregung zur Verbesserung in Stichworten zusammen:

- ☞ Einen Versuchsleiter bestimmen
- ☞ Toleranzschwellenmessung beibehalten und bei anderen Methoden miterheben
- ☞ Mehrere Finger testen und als einzelne Variable auswerten
- ☞ Zeitlichen Abstand zwischen den Sitzungen verlängern
- ☞ Fingerreihenfolge variieren
- ☞ Zustand der Hauttemperatur/Nozizeptoren prüfen, Einfluß der Raumtemperatur beachten
- ☞ Reihenfolge der Induktionsverfahren variieren
- ☞ Andere Hitzereizintensitäten wählen und ihren Abstand vergrößern
- ☞ Andere numerische Skala verwenden
- ☞ Schwellenreiz als Anker wiederholt vorgeben
- ☞ Den Probanden eine schriftliche Instruktionen vorlegen

Abschließend möchte ich auf den Anwendungsbereich der experimentellen Schmerzforschung zurückkommen. In der Klinik wird nicht die methodenspezifische Schmerzempfindlichkeit im Vordergrund stehen, sondern das Leiden unter chronischen Schmerzen. Wozu kann das Wissen um reliable Schmerzinduktionsmethoden den Menschen dienen? Nun, ich denke, daß durch die experimentelle Forschung es möglich ist, herauszufinden, welche Mechanismen der Schmerzwahrnehmung zu Grunde liegen. Neben der Probe des Wirkungsspektrums von Analgetika auf die Schmerzwahrnehmung wären weitergehende Ideen denkbar, beispielsweise die Beziehung zwischen Schmerzwahrnehmung und anderen Körperwahrnehmungen (Erschöpfung, Konzentration, Geruch) herauszuarbeiten. Dadurch könnten Zusammenhänge im Zusammenspiel der verschiedenen Sinnesmodalitäten bei der Selbstwahrnehmung erkannt werden. Der Patient könnte entdecken, daß seine Schmerzempfindung unter bestimmten Umständen weniger belastend ist und somit die Möglichkeit eigener Kontrolle erfahren.

8 Zusammenfassung

In der experimentellen wie auch in der klinischen Schmerzforschung ist es unumgänglich, die Reliabilität einer Methode zu überprüfen, bevor Aussagen über die Wirksamkeit des Verfahrens getroffen werden. Theoretischer Hintergrund dieser Arbeit ist die Psychobiologie des Schmerzes, die einsetzbaren Verfahren zur Schmerzinduktion, die sich aus der Psychophysik heraus entwickelnden Verfahren der Schmerzmessung und die Schmerzmessung unter der Perspektive der Testtheorie.

Bei zwei experimentellen Untersuchungen, die sich mit der Beeinflussung der Schmerzwahrnehmung durch Neuropeptide beim Menschen befassen, wird die Reliabilität der Schmerzinduktionsverfahren bestimmt. Als Verfahren wurden drei gängige Methoden zur Induktion von Druckschmerz, Ischämieschmerz und Hitzeschmerz eingesetzt. Die Untersuchungen sind hinsichtlich der Methoden identisch. An der einen Untersuchung nahmen 24, an der anderen 25 männliche Versuchspersonen teil. Sie kamen zu drei experimentellen Sitzungen. Beim Ischämieschmerzverfahren wurden die Schmerzschwelle bestimmt, beim Druckschmerzverfahren die Schmerzschwelle und Schmerztoleranz. Bei der Hitzeschmerzmethode wurde neben der Schmerzschwellenmessung ein Hitzediskriminationsversuch im Größenschätzverfahren durchgeführt.

Das Ischämieschmerzverfahren erwies sich als hochreliables Verfahren zur Messung der Schmerzempfindung. Das Druckschmerzverfahren dokumentierte sowohl für die Schwellenmessung als auch für die Toleranzmessung gute Reliabilität. Die Hitzeschmerzmethode kann grade noch als akzeptabel gelten. Trotz einiger hoher Reliabilitäten im ersten Experiment ist die Hitzeschmerzmethode verbesserungswürdig. Es treten große Unterschiede in der absoluten Höhe der Ischämie- und Druckschmerzschwellen zwischen beiden Experimenten auf. Diese Diskrepanzen werden unter Berücksichtigung der Unterschiede im experimentellen Anordnung diskutiert.

Mögliche Meßfehlerquellen im experimentellen Design und Anregungen zur Verbesserung der Meßgenauigkeit werden erörtert.

9. Literatur

- Adair, E.R., Stevens, J.C. & Marks, L.E. (1968). Thermally induced pain, the dol scale, and the psychophysical power law. *American Journal of Psychology*, 81 (2), 147-164.
- Albus, M. (1988). Cholecystokinin. *Progress in Neuro Psychopharmacology and Biological Psychiatry*, 12 (Suppl.), 5-21.
- Basler, H.D. & Rehfisch H.P. (1990). Schmerz und Schmerzbehandlung. In R. Schwarzer (Hrsg.), *Gesundheitspsychologie* (311-331). Göttingen: Verlag für Psychologie.
- Beecher, H.K. (1959). *Measurement of subjective response*. N.Y.: Oxford University Press.
- Beecher, H.K. (1975). Quantification of the subjective pain experience. In M. Weisenberg (Ed.), *Pain: clinical and experimental perspectives* (56-66). Sant Louis: The C. V. Mosby Company.
- Birbaumer, N. & Schmidt, R.F. (1991). *Biologische Psychologie* (2. Aufl.). Berlin: Springer.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin: Springer Verlag.
- von Campenhausen, C. (1993). *Die Sinne des Menschen* (2. Aufl.). Stuttgart: Thieme.
- Chapman, C.R., Casey, K.L., Dubner, R., Foley, K.M., Gracely, R.H. & Reading, A.E. (1985). Pain measurement: an overview. *Pain*, 22, 1-31.
- Cronbach, L.J. (1965). *Essential of psychological testing* (2nd Ed.). N.Y.: Harper & Row.
- Curio, I. (1990). Reliability of thermal pain thresholds. *Psychologische Beiträge*, 32, 16-22.
- Douglass, D.K., Carstens, E. & Watkins, L.R. (1992). Spatial summation in human thermal pain perception: comparison within and between dermatomes. *Pain*, 50, 197-202.
- Elbert, T. & Rockstroh, B. (1993). *Psychopharmakologie* (2. Aufl.). Göttingen: Hogrefe.
- Elton, D. (1983). *Psychological control of pain*. North Ryde: Grune & Stratton Australia.
- Engel, J. (1990). Kulturelle Einflüsse auf das Schmerzerleben. In R. Wörtz (Hrsg.), *Chronischer Schmerz und Psyche* (65-87). Stuttgart: Gustav Fischer Verlag.
- Fernandez, E. (1990). Artifact in pain ratings, its implications for test-retest reliability, and correction by a new scaling procedure. *Journal of Psychopathology and Behavioral Assessment*, 12 (1), 1-15.
- Fischer, A.A. (1987). Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. *Pain*, 30, 115-126.
- Flor, H. (1991). *Psychobiologie des Schmerzes*. Bern: Huber.
- Geissner, E., Dalbert, C. & Schulte, A. (1992). Die Messung der Schmerzempfindung. In E. Geissner & G. Jungnitsch (Hrsg.), *Psychologie des Schmerzes: Diagnose und Therapie* (79-97). Weinheim: Psychologie Verlags Union.
- Göbel, H. (1988). Über die Schwierigkeit einer umfassenden Definition des Phänomens Schmerz. *Der Schmerz*, 2, 88-93.
- Göbel, H. (1990). Experimentelle und klinische Algesimetrie: Beispiele angewandter Psychophysik. In R. Wörtz (Hrsg.), *Chronischer Schmerz und Psyche* (89-109). Stuttgart: Gustav Fischer Verlag.
- Göbel, H. (1992). *Schmerzmessung: Theorie, Methodik, Anwendungen bei Kopfschmerz*. Stuttgart: Gustav

- Fischer Verlag.
- Göbel, H. & Westphal, W. (1989). Experimentelle Schmerzinduktion im algosimetrischen Humanversuch: Gütekriterien. *Der Schmerz*, 3, 85-93.
- Goolkasian, P. (1985). Phase and sex effects in pain perception: A critical review. *Psychology of Women Quarterly*, 9 (1), 15-28.
- Gracely, R.H. (1985). Pain psychophysics. *Advances in Behavioral Medicine*, Vol. 1, 199-231.
- Gracely, R.H. (1989). Methods of testing pain mechanisms in normal man. In P.D. Wall & R. Melzak (Eds.), *Textbook of pain* (2nd Ed.) (257-268). N.Y.: Churchill Livingstone.
- Gracely, R.H. (1994). Studies of pain in normal man. In P.D. Wall & R. Melzak (Eds.), *Textbook of Pain* (3rd Ed.) (315-336). N.Y.: Churchill Livingstone.
- Green, D.M. & Swets, J.A. (1966). *Signal detection theory and psychophysics*. N.Y.: Wiley.
- Hacke, W. (1983). *Neurophysiologische Untersuchungsverfahren unter testtheoretischen Aspekten: Validität, Reliabilität und Objektivität ausgewählter neurophysiologischer Methoden*. Unveröffentlichte Habilitationsschrift. Medizinische Fakultät der RWTH, Aachen.
- Hardy, J.D., Wolff, H.G. & Goodell, H. (1952). *Pain sensations and reactions*. Baltimore: Williams & Wilkins Co.
- Heidenfelder, M. (1991). *Zur Quantifizierung der Schmerzempfindlichkeit: Der Einfluß der Erwartung auf die Schmerzwahrnehmung*. Unveröffentlichte Dissertation, Justus-Maximilians-Universität, Würzburg.
- IASP (1979). Pain terms: a list with definitions and notes on usage. *Pain*, 6 (Suppl.), 253-264.
- Janal, M.N., Glusman, M., Kuhl, J.P. & Clark, W.C. (1994). On the absence of correlation between responses to noxious heat, cold, electrical and ischemic stimulation. *Pain*, 58, 403-411.
- Jensen, M.P., Turner, J.A. & Romano, J.M. (1994). What is the maximum number of levels needed in pain intensity measurement? *Pain*, 58, 387-392.
- Kohllöffel, L.U.E., Koltzenburg, M. & Handwerker, H.O. (1991). A novel technique for the evaluation of mechanical pain and hyperalgesia. *Pain*, 46, 81-87.
- Lautenbacher, S. & Rollman, G.B. (1993). Sex differences in responsiveness to painful and non-painful stimuli are dependent upon the stimulation method. *Pain*, 53, 255-264.
- Lautenbacher, S., Möltner, A. & Strian, F. (1992). Psychophysical features of the transition from pure heat perception to heat pain perception. *Perception & Psychophysics*, 52 (6), 685-690.
- Levine, F.M. & De Simone, L.L. (1991). The effects of experimenter gender on pain report in male and female subjects. *Pain*, 44, 69-72.
- Lienert, G.A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5. Aufl.). Weinheim: Psychologie Verlags Union.
- Main, C.J. & Waddel, G. (1989). The assessment of pain. *Clinical Rehabilitation*, 3 (4), 267-274.
- Merskey, H. & Spear, F.G. (1964). The reliability of the pressure algometer. *British Journal of Social and Clinical Psychology*, 3, 130-136.
- Michel, L. & Conrad, W. (1982). Theoretische Grundlagen psychometrischer Tests. In K.J. Groffmann & L. Michel (Hrsg.), *Grundlagen psychologischer Diagnostik* (1-129). Göttingen: Verlag für Psychologie.
- Murrin, K.R. & Rosen, M. (1985). Pain measurement. In G. Smith & B.G. Covino (Eds.), *Acute Pain* (104-132). London: Butterworths & Co.

- Ohrbach, R. & Gale, E.N. (1989a). Pressure pain thresholds in normal muscles: reliability, measurement effects, and topographic differences. *Pain*, 37, 257-263.
- Ohrbach, R. & Gale, E.N. (1989b). Pressure pain thresholds, clinical assessment, and differential diagnosis: reliability and validity in patients with myogenic pain. *Pain*, 37, 157-169.
- Ostkirchen, G. (1994). Physiologische und psychologische Schmerztheorien - Plädoyer für Entwicklung eigenständiger medizin-psychologischer Schmerztheorien. In H. Seelbach (Hrsg.), *Rheuma, Schmerz, Psyche* (113-138). Bern: Verlag Hans Huber.
- Pause, B.M., Drews, C., Scherhag, C., Pohl, J., Pietrowsky, R., Ferstl, R., Schulte, H.M. & Fehm-Wolfsdorf, G. (1995). Analgesic effect of ceruletide in man is limited to specific pain qualities. Unveröffentlichtes Manuskript, Christian-Albrechts-Universität, Kiel.
- Pohl, J., Arnold, H., Schulz, A., Pause, B.M., Schulte, H.M. & Fehm-Wolfsdorf, G. (1995). Modulation of pain perception in man by a vasopressin analogue. Unveröffentlichtes Manuskript, Christian-Albrechts-Universität, Kiel.
- Pöppel, E. (1980). Kritische Anmerkung zum Problem der Schmerzmessung. *Medizinische Psychologie*, 6 (1/2), 214-218.
- Prestele, H. (1969). Analyse der Methoden zur Schmerzmessung. Unveröffentlichte Dissertation, Ludwig-Maximilian-Universität, München.
- Price, D.D. (1988). *Psychological and neural mechanisms of pain*. N.Y.: Raven Press.
- Price, D.D. & Harkins, S.W. (1992). Psychophysical approaches to pain measurement and assessment. In D.C. Turk & R. Melzak (Eds.), *Handbook of Pain Assessment* (111-134). N.Y.: The Guilford Press.
- Price, D.D., Hu, J.W., Dubner, R. & Gracely, R.H. (1977). Peripheral suppression of first pain and central summation of second pain evoked by noxious heat pulses. *Pain*, 3, 57-68.
- Price, D.D., McHaffie, J.G. & Stein, B. (1992). The psychophysical attributes of heat-induced pain and their relationships to neural mechanisms. *Journal of Cognitive Neuroscience*, 4 (1), 1-14.
- Procacci, P., Della Corte, M., Zoppi, M. & Maresca, M. (1974). Rhythmic changes of the cutaneous pain threshold in man. A general review. *Chronobiologia*, 1, 77-96.
- Reading, A.E. (1989). Testing pain mechanisms in persons in pain. In P.D. Wall & R. Melzak (Eds.), *Textbook of Pain* (2nd Ed.) (269-280). N.Y.: Churchill Livingstone.
- Renkl, A. & Helmke, A. (1993). Prinzip, Nutzen und Grenzen der Generalisierungstheorie. *Empirische Pädagogik*, 7 (1), 63-85.
- Schmidt, R.F. & Thews, G. (Hrsg.) (1995). *Physiologie des Menschen* (26. Aufl.). Berlin: Springer Verlag.
- Sternbach, R.A. (1978). Psychological dimensions and perceptual analysis, including pathologies of pain. *Handbook of Perception*, Vol. VIB, 231-261.
- Stevens, S.S. (1977). Psychophysik. In W. Arnold, H.J. Eysenck & R. Meili (Hrsg.), *Lexikon der Psychologie* (71-78). Herder: Freiburg im Breisgau.
- Steyer, R. & Eid, M. (1993). *Messen und Testen*. Berlin: Springer Verlag.
- Syrjala, K.L. & Chapman, C.R. (1984). Measurement of clinical pain: A review and integration of research findings. In C. Benedetti, C.R. Chapman & G. Moricca (Eds.), *Advances in Pain Research and Therapy*, Vol 7 (71-101). N.Y.: The Raven Press.
- Turk, D.C., Rudy, T.E. & Sorkin, B.A. (1993). Neglected topics in chronic pain outcome studies: determination of success. *Pain*, 53, 3-16.

- Velden, M. (1980). Messung experimentell induzierter Schmerzen mit Hilfe der Signal-Entdeckungs-Theorie. *Medizinische Psychologie*, 6, 152-162.
- Wolff, B.B. (1986). Behavioral measurement of human pain. In R.A. Sternbach (Ed.), *The Psychology of Pain* (2nd Ed.) (121-151). N.Y.: The Raven Press.
- Wottawa, H. (1977). *Psychologische Methodenlehre*. München: Juventa Verlag.
- Zenz, M. & Jurna, I. (Hrsg.) (1993). *Lehrbuch der Schmerztherapie*. Stuttgart: Wissenschaftliche Verlagsgesellschaft.