

Wie viele Guppys leben in
Santiago?
Zur Ubiquität des
Numerischen Priming
beim Ankereffekt

Dissertation

zur Erlangung des akademischen Grades
Dr. phil.
der Universität Kassel,
Fachbereich Wirtschaftswissenschaften
Fachrichtung Psychologie

vorgelegt von

Matthias Willmann

Kassel, März 2004

Dissertation eingereicht am : 22.03.2004

Tag der mündlichen Prüfung: 14.05.2004

Erstgutachter : Prof. Dr. Johannes Becker

Zweitgutachter: PD. Dr. Martin Hänze

Inhaltsverzeichnis

Einleitung	3
Theoretischer Teil	6
Der Ankereffekt	7
Urteilsheuristiken	7
Aus praktischer Perspektive	11
Aus theoretischer Perspektive	13
Zusammenfassung	15
Experimentelle Paradigmen	16
Grob-/Feinschätzungsparadigmen	16
Implizit vorgegebene Anker	20
Selbstgenerierte Anker	21
Basic Anchoring	22
Zusammenfassung	24
Drei Erklärungsmodelle	26
Unzureichende Adjustierung	26
Konversationale Inferenzen	30
Numerisches Priming	34
Zusammenfassung	38
Das Modell der selektiven Zugänglichkeit	39
Das Basismodell	39
Erweiterungen	43
Empirische Befunde	46
Theoretische Anmerkungen	57
Zusammenfassung	59
Eigene empirische Fragestellungen	61
Ankereffekte im Standardparadigma	61
Bedingungen für numerisches Priming	62

Empirischer Teil	65
Ankereffekte beim Objektvergleich	66
Vorstudie 1	68
Methode	68
Ergebnisse und Diskussion	70
Experiment 1	73
Methode	73
Ergebnisse	77
Diskussion	82
Ankereffekte durch subliminales Priming	84
Vorstudie 2	85
Methode	87
Ergebnisse und Diskussion	89
Experiment 2	91
Methode	92
Ergebnisse	96
Diskussion	100
Abschließende Diskussion	103
Erklärungsmodelle im Rückblick	104
Numerisches versus semantisches Priming	104
Zur Interaktion numerischer und semantischer Prozesse	106
Subliminale Verankerung	109
Subliminale Darbietungen	109
Aktivation als Auslöser für Ankereffekte	112
Was folgt auf Aktivierung?	113
Aktivierung vs. Aufmerksamkeit	115
Zusammenfassung	119
Literaturverzeichnis	121
Anhang	136
Vorstudie 1	137
Experiment 1	146
Vorstudie 2	152
Experiment 2	153

Einleitung

*... the nightspots spend your spirit
beat your head against the wall
two dead ends and you've
still got to choose.*

- Tom Waits -

Zu den wichtigsten kognitiven Leistungen, die das tägliche Leben in bedeutendem Maße bestimmen, gehört das Fällen von Urteilen und das Treffen von Entscheidungen, oft auch mit attraktiveren Alternativen als im obigen Beispiel.

Ein *Broker* muss sich bei der Anlage eines Geldbetrages etwa zwischen zwei verschiedenen Aktienpapieren entscheiden. Seine Wahl macht er abhängig von der von ihm erwarteten Performance der beiden Aktien. Eine *Frau* möchte sich auf einem Gebrauchtwagenmarkt nach einem neuen Auto umsehen. Um eine gewisse Kontrolle über die anstehende Geldausgabe zu behalten, setzt sie sich jedoch vorher ein preisliches Limit. Ein nordhessischer *Geschäftsmann* möchte seinem Gast aus London den Kasseler Herkules zeigen. Auf dem Weg dorthin fragt der Gast, ob das Gebäude denn höher sei als der ihm vertraute Glockenturm des Big Ben, dessen Höhe auf Nachfrage mit knapp hundert Metern angegeben wird.

Gemeinsam ist diesen drei Beispielen, dass ihnen Urteile unter Unsicherheit zugrunde liegen. Bei dem Broker resultiert die Unsicherheit aus der Abhängigkeit von unbekanntem, weil zukünftigen wirtschaftlichen Ereignissen. Bei der Frau tragen neben zukünftigen Ereignissen, die ihre finanzielle Situation beeinträchtigen könnten, auch die große Anzahl an relevanten Informationen zur Unsicherheit bei. Hierzu gehört unter anderem die Gesamtheit an Informationen über die gewünschten Details der Ausstattung und der Eigenschaften des zu erwerbenden Wagens sowie die Unsicherheit über den hierfür angemessenen Preis. Dem Geschäftsmann fehlt zur Sicherheit seines Urteils lediglich eine vielleicht schon mehrfach gehörte, aber wieder vergessene Höhenangabe in Meter.

Eine weitere Gemeinsamkeit der drei Beispiele ist, dass ihr Gegenstand jeweils Urteile über numerische Größen sind. Der Broker räsoniert über den möglichen Gewinn oder Verlust, basierend auf dem Verlauf der entsprechenden Aktienkurse. Die Frau möchte einen bestimmten Geldbetrag festlegen, der beim Kauf eines Wagens nicht überschritten werden soll. Der Geschäftsmann schließlich ist zum Vergleich zweier Größen aufgefordert. Eine davon ist ihm vorgegeben, die andere muss er mangels Kenntnis auf der Basis verfügbaren Wissens schätzen.

Das wohl bekannteste Phänomen der Urteils- und Entscheidungsforschung im Kontext numerischer Urteile ist der *Ankereffekt*. Dieser bezeichnet die Assimilation einer numerischen Schätzung oder Prognose an einen salienten Wert, dem sogenannten Anker, auch unabhängig von dessen inhaltlicher Relevanz. Der Ankereffekt gilt als eines der robustesten Phänomene im Bereich der Urteils- und Entscheidungsforschung. Die Frage nach den möglichen Ursachen für den Ankereffekt ist Gegenstand der vorliegenden Dissertation.

Die Arbeit ist folgendermaßen aufgebaut: Zunächst wird der Ankereffekt, seine inhaltliche Einordnung und seine Relevanz aus theoretischer wie praktischer Sicht diskutiert. Im anschließenden Kapitel werden die gängigen experimentellen Paradigmen vorgestellt, mit denen die Mehrzahl der vorliegenden empirischen Befunde zum Ankereffekt erhoben wurde. Ein weiteres Kapitel erläutert drei prominente Erklärungsmodelle für den Ankereffekt. Ein vierter Erklärungsansatz, der auch aufgrund der Vielzahl der hierzu vorgelegten Publikationen das momentan prominenteste Modell für Ankereffekte ist, wird in einem separaten Kapitel ausführlich diskutiert.

Zwei an den Theorieteil anschließende Fragestellungen sind dann der Gegenstand des empirischen Teils der Arbeit. Zum einen soll überprüft werden, ob Ankereffekte in einem speziellen experimentellen Paradigma, wie bisher angenommen, allein auf semantischen Prozessen basieren oder ob noch ein weiterer Prozess, das so genannte numerische Priming, am Zustandekommen des Ankereffektes beteiligt ist. Zum anderen werden Voraussetzungen für das Auftreten von Ankereffekten durch numerisches Priming untersucht, insbesondere die als relevant eingeschätzte Bedeutung der Aufmerksamkeit gegenüber dem Ankerwert. Beide Experimente verwenden neue, im Kontext von Ankereffekten bisher nicht eingesetzte experimentelle Paradigmen. Um die Bedeutung semantischer Prozesse im Zusammenspiel mit numerischen Prozessen zu untersuchen, wird ein so genannter Objektvergleich eingesetzt. Der Stellenwert von Aufmerksamkeit wird überprüft, in dem die Ankerwerte subliminal, also unterhalb der Wahrnehmungsschwelle, präsentiert werden. Beiden Experimenten ist jeweils eine Voruntersuchung vorangestellt, die notwendige Bedingungen für die entsprechenden Experimente überprüfen.

In der abschließenden Diskussion werden die Ergebnisse der beiden Untersuchungen noch einmal zusammengefasst, und deren Implikationen angesichts der bisher vorliegenden Erkenntnisse aus theoretischer wie praktischer Sicht diskutiert. Des Weiteren werden mögliche Forschungsfragen und denkbare experimentelle Untersuchungen hierzu skizziert, die an die innerhalb dieser Arbeit vorgelegten Ergebnisse sinnvoll anschließen könnten.

Theoretischer Teil

Der Ankereffekt

In diesem einführenden Kapitel soll der zentrale Gegenstand der vorliegenden Arbeit, der Ankereffekt, vorgestellt werden. Zunächst wird in einem kurzen Abriss dessen Einordnung in den Bereich der Urteilsheuristiken aus historischer und aktueller Sicht thematisiert. Im Anschluss daran wird die Bedeutung des Ankereffektes aus praktischer wie theoretischer Perspektive diskutiert.

Bereits an dieser Stelle sei darauf hingewiesen, dass Ankereffekte innerhalb der Arbeit nur im Kontext numerischer Urteile diskutiert werden (für eine Darstellung des Effektes bei nicht-numerischen Urteilen siehe etwa Keysar & Barr, 2002). Des weiteren bezieht sich die Arbeit nur auf die von Tversky und Kahneman (1974) mit diesem Terminus bezeichneten Assimilationseffekte bei Urteilen über numerische Größen. Auf die in psychophysiologischen Bereichen auftretenden Kontrasteffekte beim Urteilen (siehe etwa Helson, 1964), für die unglücklicherweise derselbe Begriff verwendet wird, kann hier ebenfalls nicht eingegangen werden (für eine Diskussion der beiden Phänomene siehe Mussweiler & Strack, 1999a).

Urteilsheuristiken

Vor etwa dreißig Jahren publizierten Amos Tversky und Daniel Kahneman ihren Artikel *Judgment under Uncertainty: Heuristics and Biases* (1974), mit dem der Grundstein für eines der bedeutendsten Forschungsprogramme innerhalb der Psychologie gelegt wurde, dem nach diesem Artikel benannten „Heuristics and Biases“-Programm.

Die zentrale These des Artikels besagt, dass in einer Vielzahl von Situationen Urteile unter Unsicherheit auf der Basis weniger so genannter heuristischer Prinzipien oder Urteilsheuristiken gefällt werden. Unter einer Urteilsheuristik wird allgemein eine Strategie verstanden, mit deren Hilfe komplexe Aufgaben, wie etwa Wahrscheinlichkeiten zu bestimmen oder Vorhersagen abzugeben, unter Rückführung auf einfachere Urteilsoperationen gelöst werden können (vgl. Tversky & Kahneman, 1974, S. 1124).

Urteilsheuristiken zeichnen sich dadurch aus, dass man einerseits mit ihrer Hilfe Urteile und Entscheidungen auch unter ungünstigen Informationskonstellationen sehr effizient und in der Regel hinreichend treffsicher fällen kann. Andererseits führt die Verwendung von Urteilsheuristiken aber unter bestimmten Bedingungen zu systematischen Fehlern (vgl. Stephan, 1999).

Entscheidend für die Relevanz dieser Heuristiken ist, dass Ihre Verwendung nicht auf deliberativen oder strategischen Prozessen über die Art der

Urteilsbildung zu basieren scheint (Gilovich & Griffin, 2002). Tversky und Kahneman (1983) sehen als Basis der Heuristiken zunächst natürliche, spontane Bewertungen, die nicht-intentional durch die Urteilsituation evoziert werden und die dadurch im weiteren häufig die Vernachlässigung anderer relevanter Informationen nach sich ziehen. Diese Position, wie auch eine Vielzahl der dazu erhobenen Befunde (für einen Überblick siehe Gilovich, Griffin & Kahneman, 2002), widersprechen der gelegentlich vertretenen Annahme, dass die Verwendung der Heuristiken und die in diesem Zusammenhang beobachteten Urteilsfehler das Resultat einer bewussten, kognitive Ressourcen sparenden Strategie sind (vgl. etwa Fiske & Taylor, 1991).

Die drei Urteilsheuristiken, die von Tversky und Kahneman in dem erwähnten Artikel diskutiert werden, sind die *Verfügbarkeitsheuristik*, die *Repräsentativitätsheuristik* und die *Ankeranpassungsheuristik*.

Die *Verfügbarkeitsheuristik* kommt bei Urteilen und Entscheidungen zum Tragen, die auf Schätzungen von Klassengrößen oder Häufigkeiten von Ereignissen basieren. Prominente Beispiele hierfür sind unter anderem Wahrscheinlichkeitsurteile, die sich an Klassengrößen oder Häufigkeiten orientieren, wie etwa die Wahrscheinlichkeit, an einer bestimmten Krankheit zu erkranken, oder die Wahrscheinlichkeit für Kurseinbrüche an der Börse (vgl. Stephan, 1999). Als Basis für die Schätzung von Klassengrößen oder Häufigkeiten dient nach der Verfügbarkeitsheuristik die Leichtigkeit, mit der es gelingt, sich an entsprechende Exemplare oder Ereignisse zu erinnern. Je leichter diese Einzelfälle zu Bewusstsein gebracht werden können, umso größer werden die entsprechenden Klassen geschätzt (ausführlicher siehe Schwarz, 1998). Die Verwendung der Verfügbarkeitsheuristik führt in vielen Fällen zu hinreichend guten Urteilen, da es häufig in der Tat leichter ist, sich an einzelne Exemplare einer größeren Klasse zu erinnern, als an einzelne Exemplare einer entsprechend kleineren Klasse (Tversky & Kahneman, 1974).

Allerdings führt die Verfügbarkeitsheuristik unter bestimmten Bedingungen auch zu systematischen Fehlurteilen. Die Bedingungen hierfür sind, dass die Leichtigkeit, sich Exemplare oder Ereignisse ins Gedächtnis zu rufen, durch Faktoren bedingt wird, die weitgehend unabhängig von den entsprechenden Klassengrößen sind. Die Vielzahl möglicher Faktoren lassen sich einteilen in erfahrungsbedingte Verfügbarkeit, gedächtnisbedingte Verfügbarkeit und einbildungskraftbedingte Verfügbarkeit (für eine weiterführende Darstellung siehe Stephan, 1999).

Die *Repräsentativitätsheuristik* kommt bei Urteilen und Entscheidungen zum Tragen, die sich auf die Wahrscheinlichkeit beziehen, mit der ein Ob-

jekt A zu einer Klasse B gehört oder ein Ereignis A durch einen Prozess B erzeugt wurde. Beispiele hierfür, die Gegenstand experimenteller Untersuchungen waren, sind etwa die Wahrscheinlichkeit, dass eine beschriebene Person einer bestimmten Berufsgruppe angehört (Kahneman & Tversky, 1972; Tversky & Kahneman, 1974) oder eine Ereignissequenz das Ergebnis eines Zufallsprozesses ist (Gilovich, Vallone & Tversky, 1985). Als Basis für derartige Wahrscheinlichkeitsschätzungen dient nach Tversky und Kahneman die Repräsentativität des Objektes oder Ereignisses für die entsprechende Klasse bzw. den entsprechenden Prozess. Unter Repräsentativität wird eine bestimmte Ähnlichkeitsbeziehung verstanden (siehe Tversky, 1977, für eine exzellente Darstellung dieser Thematik).

Auch hier gilt, dass auf der Basis der Repräsentativitätsheuristik in vielen Situationen hinreichend gute Urteile gefällt werden, da die entsprechende Ähnlichkeitsbeziehung häufig ein guter Indikator für die gesuchte Wahrscheinlichkeit ist. Systematische Fehlurteile basieren im wesentlichen darauf, dass die Information der Ähnlichkeit gemessen an normativen Standards übergewichtet wird. Dies äußert sich zum einen darin, dass prinzipiell verfügbare relevante Informationen untergewichtet, zum Teil auch völlig ignoriert werden, beispielsweise die im Kontext von Wahrscheinlichkeiten wichtigen Basis-Raten oder a-priori-Wahrscheinlichkeiten (siehe hierzu Bar-Hillel, 1980). Zum anderen basiert die Gewichtung der Ähnlichkeitsinformation, insbesondere bei der Beurteilung von Zufallsprozessen, oft auf inadäquaten Konzepten der ursächlichen Prozesse (ausführlicher hierzu siehe Gilovich et al., 1985).

Die *Ankeranpassungsheuristik* schließlich bezieht sich auf jegliche Art numerischer Schätzungen oder Prognosen, wenn im Kontext der Schätzung ein entsprechender numerischer Wert verfügbar ist. Nach Tversky und Kahneman wird der verfügbare Wert als Startpunkt für die Suche nach einer adäquaten Antwort in den Urteilsbildungsprozess integriert (ausführlicher zur Modellierung dieser Heuristik siehe S. 26 ff.). Die mögliche Urteilsverzerrung basiert darauf, dass dieser Startwert oder Anker bei der Urteilsbildung übergewichtet wird, was sich darin äußert, dass die finale Schätzung zu nahe am Anker bleibt. Darüber hinaus werden verfügbare Werte selbst dann als Anker verwendet, wenn sie keinerlei Information oder Relevanz für die zu schätzende Größe enthalten. Beobachtet werden kann dieser *Ankereffekt* daran, dass die Vorgabe verschiedener Werte zu substantiell unterschiedlichen Schätzungen führt, selbst dann, wenn der vorgegebene Wert aus Sicht der Urteiler erkennbar irrelevant für die zu schätzende Größe ist (vgl. Tversky & Kahneman, 1974).

Bereits innerhalb dieser kurzen Darstellung der drei Urteilsheuristiken ist erkennbar, dass die Ankeranpassungheuristik eine etwas andere Struktur als die Verfügbarkeits- bzw. die Repräsentativitätsheuristik besitzt.

Ein Unterschied manifestiert sich unter anderem darin, dass schon die Bezeichnung „Ankeranpassungheuristik“ in der Literatur jenseits historischer Darstellungen selten verwendet wird. Man findet in diesem Kontext stattdessen den Begriff des Ankereffektes, was im wesentlichen damit zusammenhängt, dass verschiedene Zweifel an der Adäquatheit des von Tversky und Kahneman (1974) skizzierten Prozesses zum Auftreten des Effektes existieren (vgl. auch S. 26 ff.).

In späteren Arbeiten präzisierte Kahneman (vgl. Kahneman, 2003; Kahneman & Frederick, 2002) einen weiteren Unterschied innerhalb einer spezifischeren Definition des Heuristikbegriffes. Demnach wird eine Heuristik charakterisiert durch eine Attribut-Substitution: Das Zielattribut des Urteilsobjektes wird innerhalb der Urteilsbildung ersetzt durch ein damit in Beziehung stehendes heuristisches Attribut (Kahneman, 2003, S. 707). Das Urteil wird dann auf der Basis des heuristischen Attributes gefällt. Als Essenz dieser Attribut-Substitution formuliert Kahneman (2003, S. 709), dass befragte Personen eine sinnvolle und angemessene Antwort auf eine Frage geben, die *nicht* gefragt wurde.

Bei der Verfügbarkeitsheuristik wird etwa das Zielattribut „Größe“ der Klasse ersetzt durch das heuristische Attribut „Leichtigkeit des Abrufes von Exemplaren“ der Klasse. Bei der Repräsentativitätsheuristik ist das Zielattribut die „Wahrscheinlichkeit“, dass das Objekt zur entsprechenden Klasse gehört. Ersetzt wird dieses durch das heuristische Attribut der „Repräsentativität“ des Objektes bezüglich der Klasse. Vor dem Hintergrund, dass dem Ankereffekt keine derartige Attribut-Substitution zugrunde liegt, wird dieser von Kahneman explizit aus der Gruppe der heuristikbasierten Urteilsphänomene ausgeschlossen (vgl. Kahneman, 2003, S. 707).

Es sollte allerdings angemerkt werden, dass dieser Ausschluss qua definitionem im Sinne einer Begriffspräzisierung nicht zwingend erscheint. Selbst wenn eine Konzeptionierung über die Attribut-Substitution hilfreich sein kann, sind andere Definitionen des Heuristikbegriffes denkbar, die den Bereich der Ankereffekte beinhalten (siehe etwa Stephan, 1999, S. 103 f.). Auch können die Vielzahl an Ähnlichkeiten, die Ankereffekte zu anderen heuristikbasierten Urteilsphänomenen aufweisen, als Argument angeführt werden, diese unter einem gemeinsamen Oberbegriff zu subsumieren.

Schwierigkeiten entstehen bei einer Ankereffekte beinhaltenden, *prozessorientierten* Heuristik-Definition allerdings zwangsläufig dadurch, dass über die dem Ankereffekt zugrunde liegenden Prozesse, wie die vorliegende Arbeit

später ausführlich diskutieren wird, bisher noch wenig einheitliche Positionen existieren.

Aus praktischer Perspektive

Die Relevanz des Ankereffektes aus praktischer Sicht basiert unter anderem darauf, dass er ein ausgesprochen ubiquitäres Phänomen ist, welches in einer Vielzahl von Urteilsgebieten beobachtet werden kann. Vorliegende Befunde belegen Ankereffekte bei Fragen zum Allgemeinwissen (Epley & Gilovich, 2002; Mussweiler, Förster & Strack, 1997; Russo & Shoemaker, 1989), bei Wahrscheinlichkeitsschätzungen (Chapman & Johnson, 1999; Tversky & Kahneman, 1974), bei der Bewertung von Lotterien (Chapman & Johnson, 1994), bei Preisschätzungen (Northcraft & Neale, 1987), bei Fragen nach einem angemessenen Urteil für Straftäter (Englich & Mussweiler, 2001) oder bei Prognosen über Kursverläufe im Finanzwesen (Becker & Stephan, 1994, 1996).

Darüber hinaus ist der Ankereffekt nicht beschränkt auf Laborbefunde, sondern zeigt sich auch bei relevanten Entscheidungen im Alltagsgeschehen. Einen eindrucksvollen Beleg hierfür liefert ein Feldexperiment von Erkel und Becker (2004): Im Rahmen von telefonischen Verkaufsgesprächen wurde versucht, den Referenzpunkt für die Preiswahrnehmung eines Produktes zu verankern durch die wiederholte Nennung einer Zahl. Diese Zahl hatte weder eine Relevanz für den Wert des Produktes noch war sie mit einer Währungseinheit versehen. Der Erfolg dieser Manipulation zeigt sich in einem signifikanten Anstieg der Vertragsabschlüsse in Abhängigkeit von der genannten Zahl. Durch die Konfrontation mit einem irrelevanten numerischen Wert kann also auch das Verhalten bei realen Kaufentscheidungen beeinflusst werden (für weitere Beispiele siehe etwa Nunes & Boatwright, 2001).

Neben der Ubiquität zeichnet sich der Ankereffekt auch durch eine außerordentliche Stabilität aus. Diese zeigt sich zum einen daran, dass Ankereffekte unabhängig von der Extremität des vorgegebenen Ankerwertes auftreten (Chapman & Johnson, 1994; Quattrone et al., 1984). Selbst völlig absurde Werte lösen eine Assimilation der Schätzung an die Vorgabe aus. Wegener, Petty, Detweiler-Bedell und Jarvis (2001) beobachteten beispielsweise Ankereffekte bei der Schätzung des Alters, in dem Ernest Hemingway seinen ersten erfolgreichen Roman verfasste, bei einem vorgegebenen Ankerwert von 158020 Jahren!

Zum anderen scheint der Ankereffekt auch unabhängig von motivationalen Faktoren zu sein. Wilson, Houston, Etling und Brekke (1996, Exp. 4)

zeigen etwa, dass finanzielle Anreize, ein möglichst gutes Urteil abzugeben, wie etwa durch die Auslobung eines Preises für die beste Schätzung, zu keiner Reduzierung des Ankereffektes führen.¹ Analoge Ergebnisse finden sich bei Chapman und Johnson (1995, Exp. 2).

Ein weiterer Beleg für die Stabilität des Ankereffektes ist, dass selbst Hinweise auf einen möglichen Einfluss des vorgegebenen Wertes bei einer Schätzung denselben nicht verhindert. Wilson et al. (1996, Exp. 5) verwenden sieben verschiedene Versionen von Hinweisen. Diese reichen von einer allgemeinen Bemerkung, dass Schätzungen einem Einfluss unterliegen können, bis zu der expliziten Angabe, in welcher Weise ein vorgegebener Wert eine numerische Schätzung verzerrt unter Anführung eines Beispiels nebst dem Hinweis, derartige Einflüsse zu vermeiden. Unter allen sieben Bedingungen zeigen sich signifikante Ankereffekte. Darüber hinaus existieren bei den Schätzungen keine signifikanten Unterschiede in Abhängigkeit davon, *welcher* Hinweis oder gar ob überhaupt ein Hinweis gegeben wurde.

Auch der Einsatz so genannter Decision-Support-Systeme (DSS), mit deren Unterstützung in verschiedenen Bereichen eine Optimierung von Urteilen gelingt (für einen Überblick siehe Olson & Courtney, 1992), versagt im Bereich des Ankereffektes. Wie George, Duffy und Ahuja (2000) in einer Replikation von Northcraft und Neale (1987) zeigen, führt der Einsatz von speziell für die Urteilsituation entwickelter DSS zu keinerlei nennenswerter Reduktion des Ankereffektes.

Schließlich weisen auch eine Reihe von Studien darauf hin, dass Ankereffekte unabhängig vom Expertisegrad des Urteilers auftreten. Wie Northcraft und Neale (1987) beispielsweise zeigen, treten Ankereffekte bei der Schätzung des Verkaufspreis einer Immobilie unabhängig davon auf, ob die Urteiler Studenten oder professionelle Immobilienhändler sind. In einer Studie von Stephan (2001), die sich mit einer Reihe von Urteilsfehlern im Kontext ökonomischer Szenarien beschäftigt, werden ebenfalls zwei verschiedene Stichproben von Probanden verwendet. Die erste Stichprobe besteht aus ca. 140 Studenten, die zweite Stichprobe aus ca. 140 professionellen Devisenhändlern, die im Rahmen eines Kongresses zu Finanzentscheidungen befragt werden konnten. Die Kompetenz der Devisenhändler sei dadurch belegt, dass diese nach Angabe des Kongressveranstalters ein Kapital von insgesamt 350 Milliarden Euro an internationalen Finanzmärkten verwalten. Bei der im Kontext des Ankereffektes kritischen Frage nach einer Prognose

¹Auf deskriptiver Ebene ist die Assimilation an den Ankerwert bei Auslobung des Preises sogar stärker als in der Kontrollgruppe (siehe Wilson et al., 1996, S. 396 f. für eine Diskussion diesbezüglich).

des DAX-Kurses zeigen sich wiederum sowohl bei den Studenten als auch bei den Devisenhändlern signifikante Ankereffekte in Abhängigkeit von einem vorher präsentierten Wert.

Die hier angeführten Beispiele mögen genügen, um die Relevanz des Ankereffektes aus praktischer Sicht zu verdeutlichen. Sie belegen, dass der Ankereffekt ein extrem stabiles Phänomen im Bereich der Urteilsbildung ist, dass in einer Vielzahl von Urteilsgebieten und unter variierenden Urteilsbedingungen beobachtet werden kann.

Aus theoretischer Perspektive

Neben der praktischen Relevanz hat der Ankereffekt auch aus theoretischer Sicht einen hohen Stellenwert innerhalb der Entscheidungsforschung, da er als zentrales Erklärungskonzept für eine Reihe weiterer Urteilsphänomene verwendet wird.

Ein Beispiel hierfür ist die übersteigerte Urteilssicherheit oder *Overconfidence* (Lichtenstein, Fischhoff & Phillips, 1982; Russo & Shoemaker, 1989). Eine übliche Möglichkeit, die Einschätzung der eigenen Urteilskompetenz zu überprüfen, besteht darin, nicht Punktschätzungen, sondern Intervallschätzungen über verschiedene numerische Größen zu erfragen. Die Probanden werden etwa gebeten, für jede Größe eine Untergrenze und eine Obergrenze anzugeben so, dass der wahre Wert mit neunzigprozentiger Wahrscheinlichkeit innerhalb der angegebenen Intervallgrenzen liegt. Eine adäquate Einschätzung der eigenen Urteilskompetenz zeigt sich daran, dass über verschiedene Aufgaben hinweg in etwa 90 Prozent der Fälle das Intervall die zu schätzende Größe enthält.

In verschiedenen Studien (vgl. Kiell & Stephan, 1997; Russo & Shoemaker, 1989) zeigt sich, dass trotz intensiver Erläuterung der Aufgabenstellung und insbesondere auch bei hohem Expertisegrad der Urteiler die Trefferquote drastisch unter dem vorgegebenen Wert liegt. Die von Kiell und Stephan (1997) befragten professionellen Devisenhändler etwa, die 90 %-Konfidenzintervalle für die Kursprognosen verschiedener Standarddevisen angeben sollten, erzielten lediglich eine Trefferquote von 29 Prozent.

Die Erklärung der übersteigerten Urteilssicherheit über Ankereffekte lautet, dass die Urteiler zunächst eine plausible Punktschätzung für die Urteilsgröße generieren. Diese Punktschätzung fungiert dann im folgenden als Anker für die Schätzung der Ober- und Untergrenzen. Die Assimilierung der Grenzen an diesen Anker führt dazu, dass das Intervall, gemessen am vorgegebenen Sicherheitsstandard, zu eng gewählt wird (vgl. Stephan, 1999, S. 117).

Ein weiteres Beispiel ist die systematische Verschätzung von Ereigniskombinationen. Neben der Bestimmung von Wahrscheinlichkeiten für einzelne Ereignisse ist es in vielen Situationen erforderlich, die Wahrscheinlichkeit dafür zu schätzen, dass von mehreren Einzelereignissen *wenigstens eines* eintritt oder dass *alle* Einzelereignisse *gemeinsam* eintreten. In ersterem Fall spricht man von einer *Ereignisdisjunktion*, letzteres bezeichnet man als *Ereigniskonjunktion*.

Wie Bar-Hillel (1973) in einer frühen Arbeit zeigt, werden Ereignisdisjunktionen von gleichwahrscheinlichen Einzelereignissen systematisch unterschätzt, Ereigniskonjunktionen von gleichwahrscheinlichen Einzelereignissen hingegen systematisch überschätzt. Bereits Tversky und Kahneman (1974) erklären diese Urteilsfehler unter Rückgriff auf Ankereffekte.

Als Anker fungiert hierbei die Wahrscheinlichkeit für das Eintreten eines einzelnen Ereignisses. Es ist intuitiv offensichtlich, dass das gemeinsame Eintreten mehrerer Einzelereignisse unwahrscheinlicher ist, als das Eintreten des Einzelereignisses. Ebenso gilt, dass die Wahrscheinlichkeit, dass wenigstens eines der Einzelereignisse eintritt, höher ist als die Wahrscheinlichkeit des Einzelereignisses. Ausgehend von der Wahrscheinlichkeit des Einzelereignisses muss also bei einer Ereignisdisjunktion nach unten, bei einer Ereigniskonjunktion nach oben adjustiert werden, um die gesuchte Wahrscheinlichkeit zu schätzen. Eine Assimilation der abschließenden Schätzung an die als Anker fungierende Wahrscheinlichkeit des Einzelereignisses führt auf die von Bar-Hillel beobachteten systematischen Fehler.

Ein naheliegendes Beispiel im Hinblick auf Ereigniskonjunktionen ist etwa die Angabe, in welchem zeitlichen Rahmen eine Dissertation mit einer hohen Wahrscheinlichkeit, sagen wir 90 %, abgeschlossen ist. Eine denkbare Kalkulation könnte darauf basieren, für die sieben noch fehlenden Bestandteile der Arbeit jeweils einen entsprechenden Zeitrahmen festzulegen und dann die entsprechenden Zeitangaben aufzusummieren. Die vorgegebene Frist wird jedoch nur dann eingehalten, wenn *jeder* einzelne Bestandteil der Arbeit in der dafür angenommenen Zeit fertiggestellt ist. Würde jeder Bestandteil mit einer Wahrscheinlichkeit von 90 % in der anvisierten Zeit fertig gestellt, läge die Wahrscheinlichkeit, dass alle sieben in dem entsprechenden Zeitrahmen fertiggestellt werden, bei 0.9^7 und damit knapp unter 48 Prozent² (für analoge Alltagsbeispiele mit teilweise erheblichen finanziellen Konsequenzen siehe etwa Stephan, 1999, S. 114 ff.).

²Auf eine Diskussion des Aspektes, dass manche Bestandteile auch früher fertig werden könnten, soll hier nicht eingegangen werden, zumal praktische Erfahrungen nicht für diese Möglichkeit sprechen.

Ein weiteres Urteilsphänomen, das man häufig in alltäglichen Situationen beobachten kann, ist der so genannte Rückschaufehler (auch: *Knew-it-all-along-effect*, Fischhoff, 1975; Pohl, 1992). Dieser bezeichnet die Situation, dass *nach* dem Eintreten eines Ereignisses dieses in der Rückschau, also bereits *bevor* man um den Ausgang wusste, für absehbar gehalten wird. Man überschätzt mitunter drastisch, für wie wahrscheinlich man im Vorfeld ein Ereignis gehalten hat angesichts des Wissens, dass es eingetreten ist. Typische Beispiele findet man alltäglich bei Diskussionen etwa nach einer politischen Wahl oder nach wichtigen Sportereignissen.

Ist die Urteilsgröße ein numerischer Wert wie etwa der erzielte Prozentanteil einer Partei bei einer Wahl oder die Anzahl der erzielten Tore, Punkte, Körbe o. ä. bei einem entsprechenden Sportereignis, bietet sich auch hier eine Erklärung über den Ankereffekt an. Die Schätzung oder Erinnerung des im Vorfeld antizipierten numerischen Wertes ist assimiliert an den nun bekannten richtigen Wert (für eine ausführliche Diskussion hierzu siehe Pohl, 2000)

Zusammenfassung

Als Ankereffekt wird die Assimilation einer numerischen Schätzung an einen salienten Wert bezeichnet. Ihren Ursprung hat die Forschung zum Ankereffekt im Bereich der Urteilsheuristiken von Tversky und Kahneman (1974). Im Kontext der diversen Befunde zu den verschiedenen Heuristiken wurde jedoch im späteren von Kahneman (2003) selbst in Frage gestellt, ob Ankereffekte unter die Rubrik der heuristik-basierenden Urteilsfehler fallen.

Ankereffekte zeichnen sich insbesondere dadurch aus, dass sie zu den stabilsten Urteilsanomalien im Bereich der Urteils- und Entscheidungsforschung gehören. Verschiedene Faktoren, die in anderen Bereichen zu einer erkennbaren Optimierung des Urteils führen, versagen im Bereich des Ankereffektes.

Darüber hinaus ist der Ankereffekt ein ausgesprochen ubiquitätes Phänomen und wurde in den verschiedensten Bereichen der Urteilsbildung nachgewiesen. Unter anderem findet sich eine Vielzahl von Belegen, die zeigen, dass der Ankereffekt eine wichtige Bedeutung für Urteils- und Entscheidungsbildungen in alltäglichen Situationen besitzt.

Weitere Relevanz folgt daraus, dass der Ankereffekt als Erklärungsprinzip für eine Reihe weiterer Urteilsphänomene verwendet wird. Zu nennen sind insbesondere die übersteigerte Urteilssicherheit, die systematische Verschätzung von Wahrscheinlichkeiten komplexer Ereignisse sowie der Rückschaufehler im Kontext numerischer Urteile.

Experimentelle Paradigmen

Neben der Vielzahl von Urteilsbereichen, in denen Ankereffekte beobachtet wurden, unterscheiden sich die Befunde zum Ankereffekt in der Literatur auch in Hinsicht auf die verwendeten experimentellen Paradigmen. Die Unterscheidung bezieht sich hierbei in erster Linie auf die Umstände, unter denen der Ankerwert etabliert wurde. Im Folgenden sollen diese experimentellen Paradigmen systematisch dargestellt werden, um eine Basis für die daran anschließende Diskussion über die verschiedenen theoretischen Modelle zur Erklärung des Ankereffektes zu legen.

Grob-/Feinschätzungsparadigmen

Grob-/Feinschätzungsparadigmen zeichnen sich dadurch aus, dass sie aus zwei aufeinander folgenden Aufgaben bestehen, bei denen in der ersten Aufgabe lediglich ein Urteil darüber erfragt wird, ob eine Schätzgröße über oder unter einem vorgegebenen Wert liegt. Mit dieser Aufgabe, der so genannten *Grob- oder Vergleichsschätzung* wird der genannte Wert als Anker etabliert. Die darauffolgende *Fein- oder Absolutschätzung* verlangt dann nach der Schätzung einer Größe in Form eines numerischen Wertes.

In einem der ersten und dem wohl meist zitierten Experiment zum Ankereffekt ließen Tversky und Kahneman (1974) ihre Probanden zunächst ein Urteil darüber abgeben, ob der prozentuale Anteil afrikanischer Staaten in der UNO höher oder niedriger als eine Zahl war, die mit Hilfe eines Glücksrades ermittelt wurde, auf dem die Zahlen von 1 bis 100 abgebildet waren. Im Anschluss daran sollten die Probanden den Prozentsatz afrikanischer Staaten in der UNO schätzen. Es zeigte sich ein substanzieller Einfluss der gedrehten Zahl auf das abschließende Urteil.

In der Folgezeit wurde eine Vielzahl von Untersuchungen nach diesem Schema durchgeführt (für einen Überblick siehe etwa Chapman & Johnson, 2002; Mussweiler & Strack, 1999a), so dass sich für das Paradigma im Kontext von Ankereffekten die Bezeichnung *Standardparadigma* etabliert hat. Es zeichnet sich dadurch aus, dass Grob- und Feinschätzung auf dasselbe Urteilsobjekt referieren. Darüber hinaus enthält es, wie die meisten anderen Paradigmen auch, eine Prozedur, mit Hilfe derer die Relevanz des Ankerwertes im Hinblick auf die zu schätzende Größe diskreditiert wird. Dies kann beispielsweise dadurch geschehen, dass der Ankerwert für die Probanden ersichtlich das Ergebnis eines Zufallsprozesses ist, wie im oben geschilderten Beispiel. Verschiedentlich wird die Irrelevanz des Ankers auch dadurch etabliert, dass die Versuchsinstruktionen einen Hinweis auf die Zufälligkeit

der verwendeten Vergleichswerte, gegebenenfalls mit einer Beschreibung der zugrunde liegenden Zufallsprozedur, enthalten (Strack & Mussweiler, 1997; Mussweiler & Strack, 1999b). Um weitere Erkenntnisse über die Umstände zu gewinnen, unter denen Ankereffekte im Standardparadigma auftreten, wurde in der Folge mit einer Reihe von Variationen diese Paradigmas gearbeitet.

Eine erste Variation bezieht sich auf die Extremität beziehungsweise Plausibilität der verwendeten Ankerwerte. Neben Arbeiten, in denen die Extremität der verwendeten Anker in verschiedenen Abstufungen systematisch variiert wurde (Chapman & Johnson, 1994; Quattrone et al., 1984), wurden in einer Reihe von Experimenten die Effekte von plausiblen Ankern denen von unplausiblen Ankern gegenübergestellt (Mussweiler et al., 1997; Mussweiler & Strack, 1999b, 2000a, 2001a; Wegener et al., 2001). Im Hinblick auf in diesem Bereich vorliegende Befunde und die dazugehörigen theoretischen Modelle erscheint es sinnvoll, das *Standardparadigma mit unplausiblen Ankern* als separates Paradigma zu bezeichnen.

Allerdings ist die Dichotomisierung des Kontinuums möglicher Ankerwerte in die beiden Bereiche plausible und unplausible Anker aus verschiedenen Gründen nicht unproblematisch. Zum einen ist natürlich offensichtlich, dass es interindividuelle Unterschiede hinsichtlich des Plausibilitätsbereiches für eine zu schätzende Größe gibt und auch auf individueller Ebene keine feste Grenze zwischen plausiblen und unplausiblen Werten existiert. Ein weiteres Problem hängt mit den unterschiedlichen Methoden der Klassifikation der Bereiche und der daraus resultierenden Auswahl entsprechender Ankerwerte zusammen. In einer Reihe von Arbeiten werden die Ankerwerte auf der Basis der Schätzungen von nicht verankerten Normstichproben ermittelt. Als plausible Anker verwendet man etwa Werte, die eine Standardeinheit unterhalb bzw. oberhalb des Mittelwertes der Normstichprobe liegen, die unplausiblen Anker weichen hingegen um mehr als 10 Standardeinheiten vom Mittelwert der Normstichprobe ab (Mussweiler et al., 1997; Strack & Mussweiler, 1997). Becker, Stephan und Willmann (2000) ermittelten die Grenzen der Plausibilitätsbereiche hingegen direkt in einem Vorexperiment. Andere Autoren wiederum legen die entsprechenden Werte aufgrund eigener Plausibilitätsüberlegungen fest ohne Rückgriff auf erhobene Daten (Mussweiler & Strack, 2001a; Wegener et al., 2001).

Unterschiede existieren in diesem Zusammenhang auch im Bereich der aufgetretenen Befunde: Während bei Strack und Mussweiler (1997; siehe auch Mussweiler et al., 1997; Mussweiler & Strack, 1999b) die Verwendung unplausibler Anker zu stärkeren Ankereffekten führt, liegen die Effekte bei Mussweiler und Strack (2001a) etwa auf einem vergleichbaren Niveau. We-

gener et al. (2001) hingegen beobachtete sogar schwächere Ankereffekte bei der Verwendung unplausibler Anker. Eine Erklärung für diese Diskrepanz bietet sich unter Berücksichtigung der oben erwähnten Studien von Chapman und Johnson (1994) und Quattrone et al. (1984) an. Demnach scheinen Schätzungen zwar bis weit in den Bereich unplausibler Werte hinein fast linear den vorgegebenen Ankern zu folgen, dieser monotone Trend bricht jedoch ab bzw. kehrt sich um, wenn die Ankerwerte zu extrem gewählt werden. Es scheint also auch im Bereich der Unplausibilität noch eine Grenze zu geben, die offensichtlich unplausible Werte trennt von Werten, die nicht nur als unplausibel sondern als völlig absurd empfunden werden, und dass letztere sich in ihrem Einfluss auf das Feinurteil von ersteren unterscheiden. Der Vergleich der verwendeten Werte in den oben zitierten Studien³ legt nun nahe, dass die unterschiedlichen Ergebnisse darauf basieren, dass unterschiedlich extreme unplausible Anker verwendet werden. Trotz dieser Problematik erscheint es aber generell sinnvoll, mit der Kontrastierung der Effekte von plausiblen und unplausiblen Ankern zu arbeiten, da auf diese Weise wichtige Erkenntnisse im Hinblick auf die Überprüfung verschiedener theoretischer Modelle gewonnen werden können.

Weitere Variationen des Standardparadigmas beziehen sich auf die semantische Nähe zwischen dem Urteilsobjekt der Grobschätzung und der Feinschätzung. Eine erste Abstufung wird durch einen *Dimensionswechsel* zwischen Grob- und Feinschätzung erreicht. Bei dem Dimensionswechsel referieren Grob- und Feinschätzung auf unterschiedliche Dimensionen desselben Urteilsobjektes. Beispielsweise wird zunächst gefragt, ob das Brandenburger Tor höher oder niedriger als 150 Meter ist, anschließend erfragt man eine Feinschätzung der Breite des Brandenburger Tores (Strack & Mussweiler, 1997; siehe auch Mussweiler et al., 1997; Wong & Kwong, 1999). Die Bedeutung des Dimensionswechsels für Ankereffekte und die dazugehörigen Befunde werden im Kontext der verschiedenen Erklärungsmodelle an späterer Stelle diskutiert.

Beim *Objektwechsel* beziehen sich Grob- und Feinschätzung auf unterschiedliche Urteilsobjekte. Beispielsweise folgt auf eine Grobschätzung über die Anzahl der Ärzte im lokalen Telefonbuch eine Feinschätzung der Anzahl der Staaten in der UNO (Wilson et al., 1996, Exp. 1). Im Kontext von Objektwechseln scheint es ferner für die Intensität des Ankereffektes eine

³In Mussweiler et al. (1997, Exp. 1) wurde +45 °C als hoher unplausibler Anker für die Durchschnittstemperatur in der Antarktis verwendet, in Mussweiler und Strack (2001a, Exp. 1) verwendeten dieselben Autoren +900 °C als hohen unplausiblen Anker für dasselbe Item. Wegener et al. (2001) gaben ihren Probanden etwa für die Schätzung des Alters von George Washington den Vergleichswert 167054 Jahre vor.

Tabelle 1: Grob-/Feinschätzungsparadigmen

<i>Paradigma</i>	Beispiel für die Grobschätzung
<i>Standardparadigma</i>	„Ist die Sphinx von Gizeh länger oder kürzer als 200 Meter?“
<i>Standardparadigma mit unplausiblen Anker</i>	„Ist die Sphinx von Gizeh länger oder kürzer als 20000 Meter?“
<i>Dimensionswechsel</i>	„Ist die Sphinx von Gizeh höher oder niedriger als 200 Meter?“
<i>Objektwechsel mit Skalenidentität</i>	„Ist das Brandenburger Tor länger oder kürzer als 200 Meter?“
<i>Objektwechsel mit Skalenwechsel</i>	„Kostet das Betriebssystem Windows XP mehr oder weniger als 200 Euro?“
<i>arithmetische Komparation mit Skalenidentität</i>	„Sind 200 Meter mehr oder weniger als 200 Yards?“
<i>arithmetische Komparation mit Skalenwechsel</i>	„Sind 200 australische Dollar mehr oder weniger als 200 kanadische Dollar?“

Anmerkung: Mögliche Grob-/Feinschätzungsparadigmen mit Bezug auf die Feinschätzung: „Wie lang ist die Sphinx von Gizeh in Metern?“

Rolle zu spielen, ob die numerischen Schätzungen der beiden Objekte sich auf dieselbe Skala (z.B. Längen in Meter) oder auf unterschiedliche Skalen (z.B. Länge in Metern vs. Preis in Dollar) beziehen (Becker, Stephan & Willmann, 2003). Zur Bedeutung von *Skalenwechsel* und *Skalenidentität* sei an dieser Stelle ebenfalls auf spätere Kapitel verwiesen.

Eine Art Spezialfall des Objektwechsels besteht darin, als Grobschätzungsaufgabe eine so genannte *arithmetische Komparation* (vgl. Mussweiler et al., 1997, Exp. 2) zu verwenden. Mit arithmetischer Komparation ist der Vergleich zweier Größenangaben in unterschiedlichen Einheiten gemeint (Bsp.: „Sind 10 Meter mehr oder weniger als 15 Fuß?“), auf den dann eine Feinschätzung folgt. Auch bei der arithmetischen Komparation lässt sich eine Unterscheidung dahingehend treffen, ob die zu vergleichenden Größen kompatibel mit der in der Feinschätzung sind, oder ob sie auf gänzlich andere Größeneinheiten referieren (vgl. Becker et al., 2000, Exp. 2). Eine Auflistung möglicher Grob-/Feinschätzungsparadigmen mit Beispielen ist in Tabelle 1 enthalten.

Implizit vorgegebene Anker

Dieser von Mussweiler (1997, „implicit provided anchors“, S. 17 f.) vorgeschlagene Terminus bezeichnet ein Paradigma, in dem der Ankerwert als zusätzliche Information in Bezug auf die zu schätzende Größe präsentiert wird. Der Begriff „implizit“ verweist darauf, dass dieses Paradigma im Gegensatz zum Standardparadigma keine explizite Aufforderung enthält, die zu schätzende Größe mit dem Ankerwert zu vergleichen.⁴

Das wohl bekannteste Experiment mit implizit vorgegebenen Ankern ist die Studie von Northcraft und Neale (1987). In dieser Untersuchung wurden sowohl professionelle Immobilienhändler als auch Studenten dazu aufgefordert, den Preis einer Immobilie zu schätzen. Vor der Schätzung wurden die Teilnehmer mit umfangreichen Informationen über die Immobilie ausgestattet, hierzu gehörte auch eine Besichtigung der Immobilie vor Ort. Variiert wurde lediglich der in den zehnjährigen Informationsunterlagen angegebene Listenpreis der Immobilie. Es zeigte sich, dass sowohl bei den Experten als auch bei den Studenten die abschließende Schätzung eine starke Assimilation an den angegebenen Listenpreis aufwies.

Dieser Befund scheint auf den ersten Blick wenig überraschend, wenn man davon ausgeht, dass der Listenpreis eine relevante Information in Bezug auf die zu schätzende Größe darstellt. Interessant ist das Ergebnis jedoch aus den folgenden Gründen: Zum einen gaben die Experten in einer Nachbefragung an, dass der angegebene Listenpreis keinen Einfluss auf ihr Urteil hatte. Des weiteren behaupteten befragte Immobilienhändler, dass Listenpreise, die um mehr als fünf Prozent vom Schätzwert des Gebäudes abweichen, von Experten als offensichtlich inkorrekt erkannt werden. Die verwendeten Anker wichen hingegen um bis zu zwölf Prozent vom Schätzwert der Immobilie ab und lösten dennoch Ankereffekte aus. Außerdem zeigte sich, dass die Ankereffekte bei den Experten mit gleicher Intensität auftraten, wie bei den Studenten. Unterschiede zwischen den beiden Probandengruppen ergaben sich zwar bei der Befragung, wie relevant der Listenpreis für das Urteil sei, in dem Sinne, dass Laien dem vorgegebenen Listenpreis eine höhere Relevanz für ihr eigenes Urteil einräumten. Dass trotzdem keine Unterschiede hinsichtlich des assimilativen Einflusses des Listenpreises auf das Urteil auftraten, stellt somit einen Widerspruch zu normativen Theorien der Urteilsbildung dar, da die wahrgenommene Relevanz vorliegender

⁴Mussweiler verwendet in diesem Kontext für das Standardparadigma die Bezeichnung explizit vorgegebener Anker (Mussweiler, 1997, „... anchors are *explicitly provided* ...“, S. 16).

Informationen eine wichtiger Faktor für deren Berücksichtigung bei der Urteilsbildung ist (vgl. etwa Jungermann, Pfister & Fischer, 1998).

Das Paradigma implizit vorgegebener Anker besitzt eine hohe Relevanz aufgrund seiner Alltagsnähe. Wie obiges Beispiel zeigt, können numerische Informationen, auch wenn diese innerhalb eines reichhaltigen Informationskontextes dargeboten werden, numerische Urteile beeinflussen, ohne dass die Urteiler diesen Einfluss wahrnehmen.

Selbstgenerierte Anker

Ein weiteres experimentelles Paradigma besteht darin, dass die Anker von den Urteilern selbst generiert werden. Als Beispiel hierzu diene wiederum eine Studie von Tversky und Kahneman (1974):

In dem Experiment sollten die Probanden innerhalb von fünf Sekunden das Ergebnis einer Multiplikationsaufgabe schätzen. Eine Hälfte der Probanden bekam die Aufgabe in aufsteigender Reihenfolge der Faktoren präsentiert ($1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$), die andere Hälfte in absteigender Reihenfolge ($8 \times 7 \times \dots \times 2 \times 1$). Tversky und Kahneman postulierten, dass die Urteiler bei dieser Aufgabe die ersten Multiplikationsschritte (von links nach rechts) durchführen und das Ergebnis dieser Berechnung als Anker fungiert, an den die Schätzung assimiliert ist. Demnach sollten die Schätzungen bei aufsteigender Reihenfolge der Faktoren niedriger sein als bei absteigender Reihenfolge, da die Multiplikation der ersten Faktoren im ersteren Fall zu einem geringeren Wert führt. Des weiteren sollte in beiden Gruppen das richtige Ergebnis unterschätzt werden, da in beiden Fällen das Produkt der ersten Faktoren entsprechend weit vom Produkt *aller Faktoren* entfernt ist. Beide Hypothesen wurden bestätigt (Median bei aufsteigender Reihenfolge: 512, bei absteigender Reihenfolge: 2250, richtiges Ergebnis: 40320).

Ein weiteres Beispiel soll illustrieren, mit welchen Methoden die Wahl des selbstgenerierten Ankers im Kontext von Wissensfragen manipuliert werden kann. Epley und Gilovich (2003, Exp. 2) befragten 88 amerikanische Studenten, wann die erste englische Siedlung Jamestown in Amerika gegründet wurde. Hierfür verwendeten sie zwei verschiedene Frageformate. Einmal enthielt die Frage die Formulierung „... the first settlement in the New World ...“, in dem anderen Format wurde lediglich die Bezeichnung „New World“ durch „United States“ ersetzt. Intendiert war, dass die Probanden in der ersten Version das Jahr der Entdeckung Amerikas, 1492, als Anker generieren, in der zweiten Formulierung hingegen das Jahr der Gründung der Vereinigten Staaten, 1776. In einer Nachbefragung gaben 74 % der Probanden an, bei der Beantwortung der Frage tatsächlich über den jeweils intendierten Wert

nachgedacht zu haben. Darüber hinaus zeigte sich der erwartete Ankereffekt: Enthielt die Frage die Formulierung „New World“, wurde die Gründung der Siedlung Jamestown signifikant früher geschätzt, als bei der Formulierung „United States“ (1601 vs. 1641).

Selbstgenerierte Anker zeichnen sich insbesondere dadurch aus, dass sie auf für das Urteilsobjekt relevanten Informationen basieren und der Anker von dem Urteiler von vornherein als falsch, aber im Bereich der zu schätzenden Größe liegend, klassifiziert wird (vgl. Epley & Gilovich, 2003). Darüber hinaus besitzt auch dieses Paradigma eine hohe Praxisrelevanz, da wohl in der Mehrzahl der Alltagssituationen, die eine numerische Schätzung erfordern, dieser keine Vergleichsschätzung vorangeht und häufig ein salienter Referenzwert existiert, der als Orientierung für die Schätzung dient (vgl. Epley & Gilovich, 2002, S. 139).

Basic Anchoring

Der Begriff des *Basic Anchoring* wurde von Wilson et al. (1996) eingeführt. Nach Wilson et al. bezeichnet Basic Anchoring die Beeinflussung einer numerischen Schätzung durch einen völlig uninformativen Wert, ohne dass die Urteiler dazu aufgefordert wurden, den Wert mit der zu schätzenden Größe zu vergleichen (vgl. ebd. S. 389). Ein Beispiel hierfür ist das Experiment 3 von Wilson et al. (1996). In diesem Experiment mussten die Probanden zunächst in einer vorgeblichen Studie zur Analyse von Handschriften Zahlen abschreiben. Diese Zahlen sollten als Anker für eine nachfolgende Schätzaufgabe fungieren, die als Bestandteil eines weiteren, von der Handschriftenstudie unabhängigen Experimentes präsentiert wurde.

Nach dieser Definition sind die in dem Abschnitt Grob-/Feinschätzungsparadigmen aufgeführten Objektwechsel allerdings ebenfalls Formen von Basic Anchoring, und werden von den Autoren auch in diesem Zusammenhang verwendet (siehe etwa Wilson et al., 1996, Exp. 1).

Wie aber Brewer und Chapman (2002, S. 66) bereits anmerkten, unterscheiden sich diese Urteilssituationen von der Etablierung des Ankers in einer vorgeblich unabhängigen Voraufgabe dadurch, dass explizit zum Vergleich des Ankerwertes mit einer Größe aufgefordert wird. Dies erforderte, den Anker als Quantität aufzufassen und ihn in diesem Sinne bedeutungshaltig mit einer anderen Größe in Beziehung zu setzen.⁵

⁵Nach Ansicht von Brewer und Chapman stellt aus diesem Grund bereits die Ankeretablierung in Experiment 2 der Arbeit von Wilson et al. eine Mischform zwischen Basic Anchoring und „traditional anchoring“ (Brewer & Chapman, 2002, S. 66) dar: In dieser Versuchsbedingung mussten die Probanden ihre als Anker intendierte ID-Nummern

Darüber hinaus erscheint auch die Einordnung des oben beschriebenen Dimensionswechsels nicht unproblematisch: Obwohl unter dieser Bedingung ein Vergleichsurteil erfragt wird, das den Ankerwert in Relation zum Zielobjekt der Feinschätzung stellt, fällt dieses Paradigma nach obiger Definition in den Bereich des Basic Anchoring, da beim Absoluturteil die *zu schätzende Größe* auf eine andere Dimension des Urteilsobjektes referiert.

Vor diesem Hintergrund wird der Begriff des Basic Anchoring innerhalb der vorliegenden Arbeit in einer Bedeutung verwendet, die von der ursprünglichen Definition abweicht. Als Basic Anchoring werden experimentelle Paradigmen bezeichnet, bei denen der Anker in einer Voraufgabe etabliert wird, die *keinerlei Zusammenhang zur Feinschätzung* aufweist und deren Bearbeitung *keine Schätzung einer numerischen Größe* beinhaltet.

Nach dieser Definition fällt die oben zitierte „Graphologie“-Studie von Wilson et al. (1996, Exp. 3) in den Bereich des Basic Anchoring. Gleiches gilt für das Experiment 2 derselben Arbeit. In dieser Studie war intendiert, dass die auf dem Fragebogen abgedruckte vorgebliche ID-Nummer der Probanden als Anker fungierte. Die vor der Feinschätzung durchzuführende Aufgabe bestand in den verschiedenen Versuchsbedingungen darin zu überprüfen, in welcher Farbe die ID-Nummer gedruckt war, ob sie aus vier Ziffern bestand oder ob sie größer oder kleiner als ein vorgegebener Vergleichswert war. Mit Ausnahme der Überprüfung der Zifferanzahl ergaben sich unter allen Bedingungen signifikante Ankereffekte.

Weitere Beispiele für Basic Anchoring finden sich in den Experimenten von Becker und Stephan (1994, 1996). In der Studie von 1994 mussten die Probanden zunächst mit Hilfe eines Zufallszahlengenerators eine vierstellige Zahl auswählen, die als Zielreiz für eine nachfolgende Zahlensuchaufgabe fungierte. Der Zahlengenerator war so manipuliert, dass er entweder eine hohe oder eine niedrige Zahl auswählte, die als Anker für eine spätere numerische Schätzung intendiert war. Die Zahlensuchaufgabe bestand darin, in mehreren Durchgängen zu entscheiden, ob in einem kurzzeitig dargebotenen Zahlenfeld der „ausgewählte“ Zielreiz enthalten war. Diese Aufgabe wurde den Probanden als eine Art „Gehirnjogging“ vorgestellt, deren Einfluss auf einen anschließenden Kreativitätstest untersucht werden sollte. Nach Bearbeitung des Kreativitätstests wurde die Untersuchung für beendet erklärt. Bevor die Versuchsteilnehmer jedoch den Raum verließen, wurden sie von dem Versuchsleiter überraschend gebeten, an einer kurzen Fragebogenstudie eines befreundeten Kollegen teilzunehmen. Fast alle Teilnehmer waren damit einem vorgegebenen Wert vergleichen, was lediglich ein Zurückblättern innerhalb des Fragebogens erforderte.

zu bereit. Diese Fragebogenstudie enthielt die kritische Feinschätzung. Die Probanden, hauptsächlich Studenten der Ökonomie, sollten eine Prognose für den Stand des DAX in sechs Wochen abgeben. Über eine Nachbefragung wurde sichergestellt, dass die Versuchsteilnehmer keinen Zusammenhang zwischen den beiden Untersuchungen erkannten. Es zeigten sich signifikante Ankereffekte bei der DAX-Prognose in Abhängigkeit von dem Zielreiz der Zahlensuchaufgabe.

In einer Wiederholung des Experimentes mit professionellen Devisenhändlern eines Frankfurter Bankhauses als Versuchsteilnehmer wurde dieser Effekt repliziert (Becker & Stephan, 1996). Dieser Befund erstaunt zunächst im Kontext der existierenden Literatur, da nach Wilson et al. (1996, S. 399) insbesondere im Bereich des Basic Anchoring eine höhere Kompetenz seitens der Urteiler den Ankereffekt verringert. Allerdings ist anzumerken, dass diese Position auf Befunden basiert, die im Sinne der innerhalb dieser Arbeit verwendeten Terminologie nicht dem Basic Anchoring, sondern dem Objektwechsel zuzuordnen sind. Darüber hinaus wäre es denkbar, dass sich der Einfluss von Kompetenz auf die Intensität von Ankereffekten bei allgemeinen Wissensfragen anders äußert als im Bereich von Prognosen, die insbesondere im Kontext wirtschaftlicher Kenngrößen auch für Experten Urteile unter hoher Unsicherheit darstellen (Kiell & Stephan, 1997).

Nach bisherigem Kenntnisstand ist eine notwendige Voraussetzung für das Auftreten von Ankereffekten beim Basic Anchoring, dass innerhalb der Ankeretablierung eine Auseinandersetzung mit der numerischen Größe erfolgt, die dem Wert genügend Aufmerksamkeit zukommen lässt, um seinen Einfluss auf eine nachfolgende Schätzung zu entfalten. Kritisch ist in diesem Zusammenhang die Frage nach dem Mindestmaß an notwendiger Aufmerksamkeit, um den Wert zum Anker für eine nachfolgende Schätzung werden zu lassen. Wilson et al. (1996, S. 399) diskutieren diese Frage auch bezüglich der zeitlichen und inhaltlichen Distanz zwischen Ankeretablierung und Urteil. Eine weiterführende Betrachtung, die insbesondere die Beziehung zwischen Aufmerksamkeit und Aktivierung thematisiert, ist unter anderem Gegenstand des zweiten Experimentes dieser Arbeit.

Zusammenfassung

Im vorliegenden Kapitel wurden die gebräuchlichsten experimentellen Paradigmen zum Ankereffekt anhand von Beispielen vorgestellt und diskutiert. Neben dem Standardparadigma und seinen im Abschnitt Grob-/Feinschätzungsparadigmen erläuterten Variationen, die in einem Großteil der

vorliegenden experimentellen Befunde Verwendung finden, wurden auch die weniger beforschten, aber aufgrund ihrer Alltagsrelevanz wichtigen Paradigmen mit implizit vorgegebenen und selbstgenerierten Ankern diskutiert. Abschließend wurde das Basic Anchoring Paradigma vorgestellt, mit dem erst seit etwa zehn Jahren gearbeitet wird (Becker & Stephan, 1994, 1996; Wilson et al., 1996), und dessen Befunde mitunter kontrovers diskutiert werden (Mussweiler & Strack, 2001b; Pohl, 2003).

Die Bedeutung der Differenzierung nach den verwendeten Paradigmen resultiert daraus, dass die Gemeinsamkeit lediglich in dem beobachteten Effekt, der Assimilation einer numerischen Schätzung an einen salienten Wert, besteht. Dies muss nicht notwendigerweise ein Hinweis darauf sein, dass diesem Effekt über die verschiedenen Paradigmen hinweg dieselben kognitiven Prozesse zugrunde liegen (vgl. Chapman & Johnson, 2002, S. 122; Mussweiler, 1997, S. 19). Denkbar wäre es zum einen, dass unterschiedliche Prozesse und damit unterschiedliche Erklärungsansätze für den Ankereffekt existieren, die sich in eindeutiger Weise den verschiedenen experimentellen Paradigmen zuordnen lassen. Zum anderen wäre es aber auch möglich, dass in verschiedenen experimentellen Paradigmen die gleichen Prozesse in unterschiedlichen Gewichtungen zum Ankereffekt in dem jeweiligen Paradigma beitragen.

Bevor auf diese Frage jedoch ausführlicher eingegangen werden kann, sollen zunächst die bisher existierenden Erklärungen für den Ankereffekt unter Berücksichtigung der verwendeten Paradigmen vorgestellt und diskutiert werden.

Drei Erklärungsmodelle für den Ankereffekt

In diesem Kapitel sollen zunächst drei Erklärungsmodelle für den Ankereffekt präsentiert werden, die *unzureichende Adjustierung*, *konversationale Inferenzen* und *numerisches Priming*. Ein weiteres Modell, das die Diskussion der vergangenen Jahre dominiert hat, das Modell der selektiven Zugänglichkeit, wird im Anschluss in einem separaten Kapitel ausführlich dargestellt und diskutiert.

Unzureichende Adjustierung

Das Modell der unzureichenden Adjustierung wurde von Tversky und Kahneman selbst in ihrem grundlegenden Artikel zu Urteilsheuristiken (Tversky & Kahneman, 1974) als Erklärung für den Ankereffekt vorgeschlagen. Es postuliert, dass Urteiler bei numerischen Schätzungen zunächst von einem Startwert ausgehen, und diesen dann adjustieren, um zu einer Antwort bezüglich der erfragten Größe zu gelangen. Als Startwert für diesen Prozess fungiert nach Tversky und Kahneman der etablierte Ankerwert. Der Adjustierungsprozess wird im Allgemeinen als unzureichend angenommen. Daraus resultiert, dass verschiedene Ankerwerte unterschiedliche Schätzungen evozieren, die jeweils an diesen Startwert assimiliert sind (vgl. Tversky & Kahneman, 1974, S. 1128). Eine zentrale Frage im Hinblick auf dieses Modell ist, *warum* der Adjustierungsprozess im Allgemeinen unzureichend ist (siehe Lopez, 1982).

Eine Klasse von Erklärungsansätzen hierzu bezieht sich auf die Unsicherheit der Urteiler. Quattrone et al. (1984) nehmen beispielsweise an, dass die Urteiler bezüglich der zu schätzenden Größe lediglich über ein Plausibilitätsintervall verfügen, d. h. einen Bereich möglicher Werte, die sie im Hinblick auf die zu schätzende Größe für plausibel halten. Der Adjustierungsprozess ausgehend vom Anker, so argumentieren Quattrone et al. weiter, bewege sich dann in Richtung dieses Intervalls, und breche ab, sobald ein Wert innerhalb des Intervalls erreicht ist. Hieraus resultiert, dass hohe Anker eine Schätzung evozieren, die am oberen Ende des Plausibilitätsintervalls liegt, niedrige Anker hingegen eine Schätzung am unteren Ende dieses Intervalls. Ein formales Modell, das diese Idee aufgreift, wurde im Kontext von Präferenzumkehrungen von Busemeyer und Goldstein (1992; siehe auch Busemeyer & Townsend, 1993) entwickelt.

Eine Kritik an dieser Modellierung des Adjustierungsprozesses bezieht sich darauf, dass man auf diese Weise Ankereffekte mit unplausiblen Ankern, nicht jedoch die vielfach beobachteten Ankereffekte mit plausiblen Ankern

erklären könne (Strack & Mussweiler, 1997; Mussweiler et al., 1997). Wenn der vorgegebene Anker bereits innerhalb des Plausibilitätsbereiches liege, fehle sowohl die Notwendigkeit einer Adjustierung, als auch ein Abbruchkriterium für den Prozess. Diese Kritik basiert allerdings auf der impliziten Annahme, dass die Plausibilitätsbereiche intra- und interindividuell hinreichend stabil sind, und insbesondere nicht einem Einfluss des Ankers unterliegen. Wie jedoch schon die Befunde von Quattrone et al. (1984, Exp. 2) im Kontext unplausibler Anker nahelegen, scheint das Abbruchkriterium für den Adjustierungsprozess und damit die entsprechende Grenze des Plausibilitätsbereiches eine Funktion der Ankerextremität zu sein. Dies wird daran ersichtlich, dass mit zunehmender Extremität des Ankers auch die Schätzungen und damit die postulierte Grenze des Plausibilitätsbereiches extremer werden. Was die Daten nun im Falle zunehmend unplausiblerer Anker nahelegen, kann auch im Kontext a priori plausibler Anker angenommen werden, nämlich dass der Plausibilitätsbereich ex post konstruiert wird (vgl. Kahneman & Miller, 1986) und damit einem Einfluss des Ankers unterliegt. Insbesondere da in der Beantwortung der Grobschätzung im Standardparadigma der Anker explizit als zu hoch bzw. zu niedrig klassifiziert wird, erfordert die Feinschätzung eine Adjustierung dieses vorgegebenen Wertes. Es liegt also nahe, dass demzufolge auch a priori plausible Vergleichswerte außerhalb des ex post konstruierten Plausibilitätsbereichs liegen (siehe auch Wegener et al., 2001).

Eine weitere Klasse von Erklärungsansätzen dafür, dass die Adjustierung unzureichend ist, referiert darauf, dass der Adjustierungsprozess aufwändig ist und kognitive Ressourcen erfordert. Infolgedessen würde ein Mangel an Anstrengung oder ein Mangel an kognitiven Ressourcen seitens des Urteilers dazu führen, dass der Adjustierungsprozess zu früh abgebrochen wird und damit das finale Urteil zu nahe am Anker bleibt (vgl. Chapman & Johnson, 2002). Evidenz für diesen Ansatz liefern Befunde, in denen gezeigt wurde, dass sich Urteilsverzerrungen verstärken bei Urteilen, die auf Adjustierungs- und Informationsintegrationsprozessen basieren, wenn mithilfe von „Cognitive-Load“-Prozeduren die kognitiven Ressourcen der Urteiler eingeschränkt werden (vgl. Gilbert, Miller & Ross, 1998; Gilbert, Pelham & Krull, 1988; Kruger, 1999).

Allerdings zeigt sich bei numerischen Schätzungen im Kontext von Ankereffekten, dass verschiedene Effekte, die dieser Erklärungsansatz nahelegt, nicht auftreten: Wenn ein Mangel an kognitivem Aufwand verantwortlich für den unzureichenden Adjustierungsprozess ist, müssten Ankereffekte ausbleiben oder sich zumindest abschwächen in Situationen, in denen beispielsweise durch Belohnungen Anreize zur aufwändigeren Bearbeitung gegeben werden.

Die Befunde von Stephan (1992), Kiell (1995), Wilson et al. (1996) wie auch bereits von Tversky und Kahneman (1974) zeigen, dass dies nicht der Fall ist. Falls der Adjustierungsprozess aufgrund eines Mangels an kognitiven Ressourcen unzureichend ist, wäre zu erwarten, dass Ankereffekte bei Experten geringer ausfallen, da diese auf Wissensbasen zurückgreifen können, die den kognitiven Aufwand verringern und zu ökonomischen Urteilsbildungsprozessen führen sollten. Auch hier widersprechen die vorliegenden Daten dieser Annahme (Northcraft & Neale, 1987; Kiell & Stephan, 1997; Stephan, 1992).

Ein neuerer Ansatz (Epley & Gilovich, 2003) diskutiert die Möglichkeit, den unzureichenden Adjustierungsprozess mit der Regret-Theorie (Kahneman & Tversky, 1982; Miller & Taylor, 1995) bzw. dem „Omission-bias“ (Ritov & Baron, 1990, 1992) in Verbindung zu bringen. Befunde aus diesem Gebiet zeigen, dass Fehler in der Tendenz weniger stark bedauert werden, wenn sie auf dem Unterlassen von Handlungen, denn auf Handlungen basieren. Beispielsweise wird ein Verlust am Aktienmarkt weniger schmerzhaft empfunden, wenn er auf dem Beibehalten der besessenen Wertpapiere basiert, als wenn er das Resultat eines Wechsels zu anderen Wertpapieren ist (Kahneman & Riepe, 1998). Vor dem Hintergrund, dass ein Adjustierungsprozess aus der Sukzession einzelner Anpassungsschritte besteht, leiten Epley und Gilovich nun eine Parallele zwischen Unterlassen von Handlungen vs. Handlungen und unzureichender vs. übertriebener Adjustierung her. Ob ein ähnliches Muster des Bedauerns allerdings auch im Bereich von unzureichender versus übertriebener Adjustierung existiert, und damit die Tendenz zur unzureichenden Adjustierung erklärt werden kann, ist bisher noch nicht untersucht worden (siehe Epley & Gilovich, 2003).

Abgesehen von der Frage, warum der Urteilsprozess unzureichend ist, wäre es für das Modell der unzureichenden Adjustierung darüber hinaus vorteilhaft, direkte Evidenz dafür zu präsentieren, dass Ankereffekte auf kognitiven Prozessen basiert, die eine Adjustierung beinhalten. Eine Reihe von Untersuchungen bediente sich hierbei, einem Vorschlag von Lopez (1982) folgend, Methoden, mit denen sich einzelne Prozesse der Urteilsbildung erfassen und verfolgen lassen. Eine Möglichkeit hierfür bietet der Einsatz speziell zu diesem Zweck entwickelter Experimentalsoftware, wie beispielsweise des Computerprogrammes *Mouselab* (Johnson, Payne, Schkade & Bettman, 1989).

In einer Studie von Schkade und Johnson (1989), in der diese Software zum Einsatz kam, zeigte sich, dass der Zusammenhang zwischen der Intensität des Adjustierungsprozesses und der Intensität des Ankereffektes, wie er im Modell der unzureichenden Adjustierung postuliert wird, nicht auftrat.

Eine Replikation dieser Studie von Chapman und Johnson (unveröffentlichte Daten, siehe Chapman & Johnson, 2002, S. 129) bestätigte diesen Befund.

Ein weiteres Argument gegen das Modell der unzureichenden Adjustierung basiert auf Daten, die zeigen, dass vorgegebene Ankerwerte einen Einfluss auf Urteile auch in Abwesenheit der Möglichkeit eines postulierten Adjustierungsprozesses haben. Sowohl Stephan (1993) wie auch Jacowitz und Kahneman (1995) wiesen nach, dass im Standardparadigma der Anker bereits das Urteil bei der Grobschätzungsaufgabe beeinflusst, also *bevor* der postulierte Adjustierungsprozess beginnt. Dieser so genannte *primäre Ankereffekt* (vgl. Stephan, 1993) äußert sich darin, dass der Anteil der Probanden, die sich bei der Grobschätzung für eine der beiden extremen Kategorien entscheiden (oberhalb des hohen bzw. unterhalb des niedrigen Ankers), signifikant größer ist, als die entsprechenden Anteile der Feinschätzungen in einer Kontrollgruppe, die nicht mit einem Vergleichsurteil konfrontiert wurden. Jacowitz und Kahneman (1995) beispielsweise verwendeten die 15. und 85. Perzentile einer Kontrollgruppe als niedrige bzw. hohe Anker im Standardparadigma. Es zeigte sich, dass unter der hohen Ankerbedingung der Anteil der Probanden, die in der Grobschätzung den hohen Anker als zu niedrig klassifizierten bei 27 %, also deutlich oberhalb der 15 % in der Kontrollgruppe, lag.⁶ Mit anderen Worten, die Präsentation des Ankers in der Grobschätzungsphase scheint die Plausibilitätsverteilung der zu schätzenden Größe bereits zu verzerren, *bevor* die Urteiler mit der Feinschätzungsaufgabe, die den postulierten Adjustierungsprozess auslöst, konfrontiert wurden.

Das Modell der unzureichenden Adjustierung wurde in der nachfolgenden Literatur häufig selbst als unzureichend klassifiziert und fand aus diesem Grund meist nur im Kontext eines historischen Abrisses der Literatur zum Ankereffekt Erwähnung (vgl. Strack & Mussweiler, 1997; Mussweiler, 1997; Chapman & Johnson, 1999). Dies liegt zum Teil an den oben diskutierten Widersprüchen zu empirischen Befunden. Ein weiterer Grund ist darin zu sehen, dass das Modell in seiner ursprünglichen Kennzeichnung (Tversky & Kahneman, 1974) aufgrund einer fehlenden Spezifizierung der zugrunde liegenden Prozesse und der damit verbundenen Schwierigkeit, empirisch überprüfbare Hypothesen jenseits des Ankereffektes selbst herzuleiten, von verschiedenen Autoren eher als Prozessbeschreibung denn als Prozessklärung angesehen wird (vgl. Gigerenzer, 1996; Strack & Deutsch, 2002).

Allerdings wird das Modell der unzureichenden Adjustierung in neueren Publikationen von Epley und Gilovich (2002, 2003) als Erklärung für Anker-

⁶Ein analoger Effekt unter der niedrigen Ankerbedingung blieb bei Jacowitz und Kahneman aus, wurde aber von Stephan (1993) beobachtet.

effekte im Kontext selbstgenerierter Anker diskutiert (siehe auch S. 21). Soll beispielsweise geschätzt werden, aus wie vielen Mitgliedsstaaten die Vereinigten Staaten im Jahre 1880 bestanden, fungiert nach Epley und Gilovich die den (amerikanischen) Probanden vertraute, heutige Anzahl von 50 Staaten als Anker, von dem ausgehend nach unten adjustiert wird.

In einer Reihe von Experimenten konnten Epley und Gilovich (2002, 2003) Evidenz dafür liefern, dass bei Schätzaufgaben, zu denen wie im obigen Beispiel ein den Urteilern bekannter Referenzwert existiert, dieser als Startwert für einen Adjustierungsprozess dient, auf dem das finale Urteil basiert. Außerdem wiesen Epley und Gilovich nach, dass dieser Adjustierungsprozess im Allgemeinen unzureichend ist. Darüber hinaus zeigten die Autoren, dass Prozeduren, mit denen die Intensität eines Adjustierungsprozesses beeinflusst werden sollte, einen Einfluss auf Schätzungen bei selbstgenerierten Ankern hatten, Schätzungen im Standardparadigma von diesen Prozeduren jedoch unbeeinflusst blieben. Epley und Gilovich werten dies als Evidenz dafür, dass unzureichende Adjustierung die Ursache für Ankereffekte bei selbstgenerierten Ankern ist, die Ankereffekte im Standardparadigma jedoch auf anderen kognitiven Prozessen basieren.

Zusammenfassend ist also festzuhalten, dass das Modell der unzureichenden Adjustierung trotz verschiedener theoretischer Unzulänglichkeiten und der oben geschilderten empirischen Kritik auch weiterhin als möglicher Kandidat für ein Erklärungsmodell des Ankereffektes, zumindest bei selbstgenerierten Ankern, angesehen werden muss.

Konversationale Inferenzen

Ein zweiter Erklärungsansatz attribuiert den Ankereffekt auf konversationale Inferenzen im Sinne des Modells der Konversationsnormen von Grice (1975). Im Zentrum dieses Ansatzes steht die Maxime der Quantität (*maxim of quantity*, Grice, 1975). Hiernach interpretieren Personen innerhalb eines Konversationsaktes die Beiträge der Teilnehmer unter der Annahme, dass diese so informativ wie notwendig gestaltet sind, aber nicht darüber hinaus gehen, d. h. den Gesprächspartner nicht mit unnötigen Informationen ausstatten (vgl. Schwarz, 1999). Die Anwendung dieser Norm auf die standardisierte Untersuchungssituation könnte es nun bedingen, dass der Anker als relevante Information in Bezug auf die zu schätzende Größe angesehen wird (vgl. Bless, Strack & Schwarz, 1993; Schwarz, 1994, 1996).

Wenn beispielsweise im Standardparadigma dazu aufgefordert wird, die zu schätzende Größe mit einem vorgegebenen Wert zu vergleichen, könnten die Urteiler den Vergleichswert als Hinweis auf den Bereich, in dem die zu

schätzende Größe liegt, interpretieren. Der Assimilationseffekt basiert also in diesem Fall darauf, dass die Urteiler den vorgegebenen Vergleichswert, unabhängig von dessen normativer Relevanz, als zusätzliche, von dem Versuchsleiter zur Verfügung gestellte relevante Information bewerten, und in den zur Urteilsbildung notwendigen Informationssuchprozess integrieren.

Das Modell der konversationalen Inferenzen ist in der Tat ein plausibler Kandidat zur Erklärung einer Reihe von experimentellen Befunden im Kontext des Ankereffektes. Wird ein ersichtlich nicht irrelevanter Wert als zusätzliche Information zur Verfügung gestellt (Burton, Lichtenstein & Herr, 1993; Northcraft & Neale, 1987) oder kann im Kontext realitätsnaher experimenteller Szenarien die Irrelevanz eines vorgegebenen Vergleichswertes nicht hinreichend etabliert werden (Mussweiler, Strack & Pfeiffer, 2000, Exp. 1), erscheint es nicht unplausibel, den Effekt auf Urteilsbildung im Sinne von konversationalen Inferenzen zu interpretieren.

Auch bei den mit dem Ankereffekt verwandten Effekten von Referenzpunkten auf die Urteilsbildung (Kahneman, 1992) bietet sich eine Erklärung unter Rückgriff auf Konversationsnormen an. Referenzpunkte, beispielsweise im Kontext von Gegenangeboten bei Verhandlungen (Galinsky & Mussweiler, 2001) oder im Kontext von Preiswahrnehmungen bei Kaufentscheidungen (Burton et al., 1993; Stephan & Willmann, 2002), entfalten ihren Einfluss über die wahrgenommene Relevanz des Referenzwertes. Wird bei Produktangeboten neben dem aktuellen Verkaufspreis ein vorheriger, deutlich höher liegender Preis oder eine ebenfalls höher liegende Preisempfehlung des Produktherstellers dargeboten, fungieren diese Referenzwerte nach Burton et al. (1993) als Anker für einen internen Preisstandard des Produktes. Die Bewertung des Verkaufspreises basiert dann auf einem Vergleich mit dem derart verankerten Standard.

Schwierigkeiten bereiten dem Modell der konversationalen Inferenzen Experimente, bei denen der Vergleichswert klar erkennbar durch einen Zufallsprozess zu Stande kam. Beispiele hierfür sind die erwähnten Befunde, bei denen der Anker mit Hilfe eines Glücksrades (Tversky & Kahneman, 1974), durch das Ziehen einer Karte (Cervone & Peake, 1986; Wilson et al., 1996) oder auf der Basis der Telefonnummer der Probanden (Russo & Shoemaker, 1989) bzw. deren Versicherungsnummer (Chapman & Johnson, 1999) ermittelt wurde. In diesem Zusammenhang sollte jedoch unterschieden werden zwischen der normativen Relevanz des Ankers in Bezug auf das Urteil und dem subjektiv empfundenen Informationsgehalt des Ankers aus Sicht der Probanden. In zwei Experimenten von Chapman und Johnson (1999, Exp. 3, Exp. 4) wurden die Vergleichswerte für jeden Probanden individuell auf der Basis seiner Versicherungsnummer („social security number“) er-

mittelt. Trotzdem gaben in der Nachbefragung 34 % (Exp. 3) bzw. 15 % (Exp. 4) der Probanden an, dass der Vergleichswert *informativ* in Bezug auf das zu fällende Urteil sei. Allerdings traten in beiden Experimenten auch unter den Probanden, die die Irrelevanz des Vergleichswertes bestätigten, substantielle Ankereffekte auf.

Eine weitere Kritik am Modell der konversationalen Inferenzen (vgl. Mussweiler, 1997, S. 24; Mussweiler et al., 1997, S. 596) bezieht sich auf die Befunde mit implausiblen Ankern. Ein Vergleichswert von 45000 Kilometer kann, so Mussweiler, nur schwer als informativ für die Längenschätzung eines Flusses angesehen werden, der in etwa 1100 Kilometer lang ist. Jedoch lässt sich auch hier ein Gegenargument konstruieren, dessen empirische Überprüfung allerdings bislang noch aussteht: Ein unplausibler hoher Anker wie im obigen Beispiel könnten von den Urteilern als Hinweis interpretiert werden, das Urteilsobjekt sei hinsichtlich der zu schätzenden Dimension im oberen Bereich seiner Objektkategorie (im Beispiel: Flüsse), entsprechendes gelte für unplausible niedrige Anker.

Die deutlichsten Befunde, die gegen das Modell der konversationalen Inferenzen sprechen, finden sich im Bereich des Basic Anchoring (Becker & Stephan, 1994, 1996; Wilson et al., 1996). Wird der Anker in einer vorgeblich unabhängigen Voraufgabe etabliert und wird die Unabhängigkeit der beiden Aufgaben *aus Sicht der Probanden* in einer Nachbefragung abgesichert, lässt sich kein Argument für eine Beteiligung der im Modell postulierten Prozesse am Auftreten eines Ankereffektes konstruieren.

Eine direkte Überprüfung des Modells der konversationalen Inferenzen im Kontext einer „willingness to pay“ (WTP) - Befragung stammt von Kahneman und Knetsch (1993). Ausgangspunkt der Untersuchung ist eine Telefonumfrage aus 1984, in der die Befragten angeben sollten, wie viel Dollar sie für die Durchführung eines ökologischen Programmes zu zahlen bereit wären. Variiert wurde der in der Eingangsfrage („Wären Sie bereit, X Dollar zu zahlen?“) genannte Betrag (niedriger Anker: 25 \$, hoher Anker 200 \$).

Es zeigen sich substantielle Effekte in Bezug auf die WTP in Abhängigkeit vom Ankerwert. Eine naheliegende Erklärung im Sinne der konversationalen Inferenzen würde lauten, dass die Befragten den in der Eingangsfrage genannten Wert als informativ im Hinblick auf die WTP in der Bevölkerung ansehen, und diese Information bei ihrer Antwort berücksichtigten.

Kahneman und Knetsch überprüfen diese Hypothese in einem Experiment, bei dem sie nach der Eingangsfrage *nicht* die persönliche Zahlungsbereitschaft, sondern ein Urteil über die durchschnittliche Zahlungsbereitschaft in der Bevölkerung erfragt haben. Es zeigt sich, dass die in der Eingangsfrage

etablierten Anker hierbei keinerlei Einfluss auf die Schätzung der Zahlungsbereitschaft in der Bevölkerung hat (Kahneman & Knetsch, 1993, Exp. 1).

Dieser Befund ist problematisch für das Modell der konversationalen Inferenzen, da keine Belege für das Modell nachgewiesen worden sind, *obwohl* die experimentellen Bedingungen dafür ausgesprochen günstig sind.

Zum einen sind keinerlei Bemühungen unternommen worden, die Relevanz der Ankerwerte zu unterminieren. Zum anderen spricht innerhalb des Szenarios einer Telefonumfrage aus Sicht des Befragten einiges dafür, dass der genannte Wert eine Relevanz für die erfragte Größe hat. Es ist nämlich davon auszugehen, dass Initiatoren einer Befragung eine hinreichende Variation bei der Beantwortung der Eingangsfrage erwarten, was nur dann der Fall ist, wenn der Vergleichswert im Bereich des Zentrums der WTP-Verteilung in der Bevölkerung liegt.⁷ Wenn sich unter derartigen Bedingungen keinerlei Hinweise auf den Einfluss konversationaler Inferenzen nachweisen lassen, erscheint es wenig plausibel anzunehmen, dass diese Prozesse unter weit ungünstigeren Umständen am Zustandekommen des Ankereffektes beteiligt sind.

Neben den Widersprüchen zu verschiedenen experimentellen Befunden, existiert im Hinblick auf das Modell der konversationalen Inferenzen auch ein konzeptionelles Problem. In der grundlegenden Arbeit zu Urteilsheuristiken wird unter dem Ankereffekt zunächst nur die Assimilation einer numerischen Schätzung an einen verfügbaren Wert verstanden (Tversky & Kahneman, 1974). In späteren Arbeiten (Kahneman, 1992; Kahneman & Knetsch, 1993) wird diese Definition dahingehend eingeschränkt, dass unter dem Ankereffekt die Assimilation einer numerischen Schätzung an einen Wert verstanden wird, der in Bezug auf die Urteilssituation *erkennbar irrelevant* ist.⁸ Die Irrelevanz des Ankers bezieht sich also nicht auf eine normative Bewertung der Urteilssituation, sondern auf die subjektive Bewertung seitens des Urteilers. Legt man nun letztere Definition zugrunde (siehe auch Chapman & Johnson, 2002), ist das Modell der konversationalen Inferenzen *per definitionem* keine Erklärung für den Ankereffekt, weil es auf die vom Urteiler wahrgenommene Relevanz des Ankers referiert.

In der Diskussion zum Ankereffekt in der neueren Literatur spielt das

⁷„The reasoning is that it is foolish to ask a survey question that everyone will answer alike, and that the investigator who framed the survey is probably no fool.“ (Kahneman & Knetsch, 1993, S. 3)

⁸Nach dieser engeren Definition gehören die vorgestellten Paradigmen mit implizit vorgegebenen bzw. selbstgenerierten Ankern nicht in den Bereich der Ankereffekte, was möglicherweise ein Grund dafür darstellt, warum diese Paradigmen weniger intensiv beforscht wurden.

Modell der konversationalen Inferenz eine eher geringe Rolle. Seine Bedeutung für die aktuelle Forschung besteht in erster Linie darin, mit experimentellen Paradigmen zu arbeiten, die es ermöglichen, konversationale Inferenzen als Alternativerklärung zu den im Zentrum der Fragestellung stehenden Modellen weitestgehend auszuschließen.

Numerisches Priming

Ein dritter Erklärungsansatz attribuiert den Ankereffekt auf numerisches Priming (Becker & Stephan, 1994, 1996; Becker et al., 2000; Jacowitz & Kahneman, 1995; Kahneman & Knetsch, 1993; Wilson et al., 1996; Wong & Kwong, 2000). Demnach wird durch die Präsentation des Ankerwertes ein spezifisches numerisches Konzept aktiviert. Die Assimilation der Schätzung an den Anker basiert dann in erster Linie auf der Aktivierung dieses Konzeptes. Hinsichtlich der Spezifizierung dieses Ansatzes existieren verschiedene Positionen, die sich in der Modellierung des Primingprozesses, des aktivierten Konzeptes und des Prozesses, mittels dessen sich das aktivierte Konzept auf die Feinschätzung auswirkt, voneinander unterscheiden.

Der Primingprozess lässt sich zum einen modellieren im Sinne des Standard-Priming (Higgins, 1996). Demnach erhöht die Präsentation eines Reizes das Aktivationspotential der mentalen Repräsentation des entsprechenden Konzeptes. Darauf basierend, dass dieser Zustand erst allmählich abklingt, besitzt dieses Konzept zum Zeitpunkt einer nachfolgenden Urteils-situation genügend Residualaktivierung, was die Wahrscheinlichkeit erhöht, in der nachfolgenden Urteils-situation verwendet zu werden (vgl. Wilson et al., 1996).

Eine andere Form des Primingprozesses, der im Kontext des Ankereffektes diskutiert wird (Kahneman & Knetsch, 1993), ist das sogenannte *Backward-Priming* (vgl. Kahneman & Miller, 1986; Kahneman, Treisman & Gibbs, 1992). Die zentrale Idee des Backward-Priming ist, dass der Aktivationsprozess, auf dem ein möglicher Primingeffekt basiert, nicht von der Präsentation des entsprechenden Reizes ausgeht, sondern von der Urteils-situation, in der der Primingeffekt üblicherweise beobachtet wird. Die Präsentation eines Reizes ist zwar eine notwendige Voraussetzung dafür, dass dieser in einer nachfolgenden Urteils-situation aktiviert werden kann; die Aktivierung wird aber ausgelöst durch die nachfolgende Urteils-situation und kontrolliert durch ihre Charakteristika (vgl. Kahneman & Knetsch, 1993, S. 10). Stellt man sich beispielsweise vor, dass in einer Aufgabe sowohl ein Geldbetrag in Euro als auch eine Längenangabe in Metern genannt wird, würde nach dem Backward-Priming-Modell durch eine nachfolgende Frage über einen Preis

in Euro der genannte Geldbetrag aktiviert, durch eine nachfolgende Frage über eine Größe in Metern die genannte Längenangabe.

Der Unterschied zwischen Standard-Priming und Backward-Priming wird von verschiedenen Autoren (vgl. Mussweiler, 1997; Mussweiler et al., 1997; Wilson et al., 1996) im Kontext des numerischem Priming als nicht besonders relevant erachtet. Er erweist sich jedoch als kritisch in Bezug auf Erklärungsansätze für das Auftreten und Ausbleiben von Ankereffekten unter den verschiedenen Variationen des Grob-/Feinschätzungsparadigmas, insbesondere im Kontext von Objekt- und Skalenwechsel (vgl. Becker et al., 2003; Kahneman & Knetsch, 1993; Mussweiler & Strack, 2001b).

Im Standard-Priming ist die durch die Präsentation des Ankerwertes erhöhte Aktivierung der Auslöser für Ankereffekte. Das Auftreten und Ausbleiben des Ankereffektes unter verschiedenen experimentellen Paradigmen lässt sich somit allenfalls mit Hilfe von Zusatzannahmen über die Intensität der durch die Reizpräsentation ausgelöste Aktivierung erklären. Derartige Zusatzannahmen erscheinen im Bezug auf die existierenden experimentellen Befunde häufig wenig plausibel.

Aus dem Backward-Priming-Modell hingegen lassen sich Vorhersagen ableiten, unter welchen Umständen ein vorher präsentierter numerischer Wert zum Anker für ein numerisches Urteil werden kann, die auf die Ähnlichkeit zwischen den Charakteristika der Urteilsituation und der Reizpräsentation referieren. Demnach sind beim Vergleich zwischen verschiedenen experimentellen Paradigmen umso stärkere Ankereffekte zu erwarten, je ähnlicher Urteilsituation und der Kontext der Ankerpräsentation sind, beispielsweise im Hinblick auf Merkmale wie die verwendeten Skalen und Dimensionen oder auch der inhaltliche Kontext. Diese Vorhersage deckt sich mit einer Vielzahl von Befunden aus der Literatur (vgl. Chapman & Johnson, 1994; Kahneman & Knetsch, 1993; Mussweiler & Strack, 2001b; Strack & Mussweiler, 1997; für ein Gegenbeispiel siehe etwa Wong & Kwong, 2000).

Voraussetzung für die beim Backward-Priming postulierten Prozesse ist natürlich, dass der präsentierte numerische Wert *mitsamt seines Präsentationskontextes* enkodiert wird. Diese Voraussetzung erscheint trivial vor dem Hintergrund der Spreading-Activation-Theorie (Collins & Loftus, 1975; für eine aktuelle Darstellung siehe etwa Reisberg, 2001) als weitgehend akzeptierte Modellierung kognitiver Prozesse. Sie führt aber zu der kritischen Frage nach der spezifischen Struktur des via Standard- oder Backward-Priming aktivierten Konzeptes, auf dem der Assimilationseffekt basiert.

Nach der Modellierung von Kahneman und Knetsch (1993) besteht dieses Konzept aus dem numerischen Wert *und* eventuellen Attributen wie der angegebenen Skala bzw. Maßeinheit. Wong und Kwong (2000) hingegen be-

schreiben die mentale Repräsentation des Ankers zunächst zwar ebenfalls als Kombination des numerischen Wertes einschließlich eventueller Affixe wie Vorzeichen oder Maßeinheit. Die Verbindungen zwischen diesen Bestandteilen wird aber als oberflächlich postuliert. Als das den Ankereffekt auslösende Konzept wird ausschließlich die mentale Repräsentation des numerischen Wertes angesehen (Wong & Kwong, 2000, S. 317 f.; für die Diskussion einer ähnlichen Problematik siehe etwa Tenpenny, 1995). Wong und Kwong (2000) präsentieren eine Reihe von Experimenten, die diese Theorie stützen; Befunde, die der Theorie widersprechen, werden ebenfalls diskutiert.

Neben der Bedeutung des Kontextes der Ankerpräsentation ist ein weiterer wichtiger Aspekt die Frage nach dem Inhalt des aktivierten numerischen Konzepts. In einer Reihe von Publikationen (Jacowitz & Kahneman, 1995; Kahneman & Knetsch, 1993; Wilson et al., 1996) wird hierbei ausschließlich auf die als Anker etablierte Zahl referiert, die aufgrund ihrer Aktivierung mit einer höheren Wahrscheinlichkeit als mögliche Antwort für eine nachfolgende Schätzaufgabe in Betracht gezogen wird. Dieser Ansatz wird von verschiedenen Autoren kritisiert vor dem Hintergrund der Implikation, dass die präsentierte Zahl dann auch öfter als Antwort für die Feinschätzung gegeben werden sollte, was jedoch nicht der Fall ist (vgl. Mussweiler, 1997, S. 36).

Eine alternative Theorie, die sowohl im Einklang mit dem Spreading-Activation-Modell, als auch mit Befunden zur mentalen Repräsentation von Zahlen (für einen Überblick siehe Dehaene, 1999) steht, würde lauten, dass durch die Präsentation des Ankers nicht nur exklusiv diese Zahl, sondern auch der umliegende Zahlbereich aktiviert wird (siehe auch Wong & Kwong, 2000, S. 329). Für diesen Ansatz sprechen Befunde, in denen der Ankereffekt ausgelöst wurde durch die wiederholte Auseinandersetzung mit verschiedenen Zahlen innerhalb eines engeren Intervalls, wie beispielsweise bei Wilson et al. (1996, Exp. 3; für eine kritische Diskussion dieses Befundes siehe Brewer & Chapman, 2002).

Die abschließende Frage in Bezug auf das Modell des numerischen Primings lautet nun, auf welche Weise das aktivierte numerische Konzept eine nachfolgende numerische Schätzung beeinflusst. In vielen Publikationen wird in diesem Kontext lediglich postuliert, das aktivierte numerische Konzept gehe aufgrund seiner „... erhöhten Zugänglichkeit in die Berechnung des absoluten Urteils mit ein ...“ (Mussweiler et al., 1997, S. 593), oder werde infolge dessen als plausible Antwort für die nachfolgende Schätzaufgabe in Betracht gezogen (Mussweiler, 1997, S. 25; Wilson et al., 1996, S. 388). Eine derartige Darstellung erscheint problematisch, da sie einen eher bewussten, evaluativen Integrationsprozess suggeriert. Insbesondere im Zusammenhang

von Ankereffekten bei unplausiblen Ankerwerten stellt sich die Frage, wie ein Wert, der im Hinblick auf das Schätzwerturteil als unplausibel postuliert wird, im Kontext des Feinurteils als plausible Antwort in Betracht gezogen werden kann.

Eine Darstellung, die den Aspekt der Bewusstheit bei der Integration des Ankerwertes in das Schätzwerturteil ausführlicher thematisiert, findet sich bei Jacowitz und Kahneman (1995). Sie postulieren, dass unter Unkenntnis der wahren Ausprägung der zu schätzenden Größe eine intuitive Schätzung die Aktivierung möglicher Antworten evoziert, die *implizit, automatisiert* und *unkontrolliert* abläuft (vgl. ebd. S. 1162). Die Antwort wird dann als eine Art Kompromiss angenommen, bei der *jeder* aktivierte Wert, mit einer entsprechenden Gewichtung versehen, in den Urteilsprozess integriert wird (Strack, 1992; Wilson & Brekke, 1994). In Anlehnung an Gilbert (1990), wonach die anfängliche Reaktion auf jede Behauptung zunächst den Glauben an die Richtigkeit der Behauptung beinhaltet, postulieren Jacowitz und Kahneman ähnliche Prozesse im Kontext des Ankereffektes. Die Frage beispielsweise, ob der Amazonas länger oder kürzer als 5000 Meilen ist, kann demnach zumindest einen oberflächlichen Glauben an beide Möglichkeiten evozieren, wodurch der Ankerwert zu einer zumindest kurzzeitig in Betracht kommenden Antwort wird. Dieser Prozess wird als automatisiert angenommen, so dass auch das Erwägen des Ankers als möglicher Antwort nicht auf einer bewussten Evaluation des Ankerwertes im Kontext der Schätzaufgabe basiert, sondern dieser Evaluation vorausgeht (vgl. Gilbert, 1990).

Zusammenfassend ist festzuhalten, dass numerisches Priming eine Vielzahl von Befunden im Kontext von Ankereffekten erklären kann, bei denen alternative Erklärungsansätze weitgehend auszuschließen sind (Becker & Stephan, 1994, 1996; Becker et al., 2000; Wilson et al., 1996; Wong & Kwong, 2000). Auch bei Befunden, die sich mit Hilfe anderer Modelle erklären lassen, bietet sich eine alternative Erklärung via numerischem Priming an (ausführlicher hierzu siehe S. 57 ff.). Schwierigkeiten existieren im Hinblick auf diskrepante Befunde bei experimentellen Paradigmen mit Dimensions-, Skalen- und Objektwechsel bzw. beim Basic Anchoring (vgl. Pohl, 2003), die wie dargestellt im Kontext divergierender Modellierungen der einzelnen Prozessbestandteile diskutiert werden. Allerdings ist anzumerken, dass diese Befunde auch von alternativen Modellen nicht zufriedenstellend erklärt werden können.

Zusammenfassung

Die Beiträge, welche die drei vorgestellten Modelle zur Erklärung des Ankereffektes leisten, lassen sich folgendermaßen zusammenfassen:

Unzureichende Adjustierung scheint die Basis für Ankereffekte bei selbstgenerierten Anker zu sein. Obwohl diesem Paradigma aufgrund seiner hohen Praxisrelevanz sicherlich eine besondere Bedeutung zukommt, ist es in der Literatur zum Ankereffekt eher randständig behandelt worden. In der Mehrzahl der Befunde wurden mit anderen experimentellen Ansätzen, wie beispielsweise Grob-/Feinschätzungsparadigmen, Ankereffekte beobachtet. Wie insbesondere die Experimente von Epley und Gilovich (2002, 2003) nahelegen, scheint das Erklärungsmodell der unzureichenden Adjustierung jedoch auf die Verwendung selbstgenerierter Anker beschränkt. Im Bereich anderer experimenteller Paradigmen spricht wenig dafür, dass die in dem Modell postulierten Prozesse am Zustandekommen des Ankereffektes beteiligt sind.

Das Modell der konversationalen Inferenzen referiert auf die wahrgenommene Relevanz des Ankerwertes in Bezug auf das Urteil. Damit erscheint es als plausible Erklärung in Situationen, in denen eine Relevanz des Ankers naheliegend ist. Allerdings kann es keine Ankereffekte erklären, bei denen diese Möglichkeit definitiv ausgeschlossen werden kann, wie beispielsweise beim Basic Anchoring. Seine Bedeutung für die aktuelle Forschung besteht darin, soweit möglich mit experimentellen Paradigmen zu arbeiten, die dieses Modell als Erklärung ausschließen. Allerdings liegt wenig direkte Evidenz dafür vor, dass ohne entsprechend sorgfältige Vorkehrungen konversationale Inferenzen einen Beitrag zum Ankereffekt leisten würden.

Von den drei vorgestellten Erklärungsansätzen ist numerisches Priming das Modell, welches neben dem im folgenden Kapitel präsentierten Modell der selektiven Zugänglichkeit die Diskussion um Ankereffekte jenseits von selbstgenerierten Anker dominiert. Für numerisches Priming sprechen eine Reihe von Befunden, insbesondere unter Bedingungen, bei denen kein inhaltlicher Zusammenhang zwischen der Ankeretablierung und dem Urteil besteht. Allerdings existiert bisher kein einheitliches Modell des numerischen Priming, welches auf der Basis spezifizierter Prozesse die Vielzahl der zum Teil diskrepanten Befunde unter Berücksichtigung entsprechender Randbedingungen erklären kann. Ob und in wie weit numerisches Priming am Zustandekommen des Ankereffektes in anderen experimentellen Paradigmen, insbesondere im Standardparadigma beteiligt ist, ist unter anderem Gegenstand des ersten Experimentes dieser Arbeit.

Das Modell der selektiven Zugänglichkeit

Das Modell der selektiven Zugänglichkeit („Selective Accessibility Model“, kurz: SAM) wurde von Mussweiler und Strack zunächst ausschließlich als Erklärungsmodell für Ankereffekte im Standardparadigma vorgeschlagen (Mussweiler, 1997; Strack & Mussweiler, 1997). In späteren Arbeiten versuchten die Autoren dann unter Rückgriff auf verschiedene Zusatzannahmen, mit diesem Modell auch Ankereffekte in anderen experimentellen Paradigmen zu erklären (Mussweiler & Strack, 2000a, 2001a, 2001b).

In diesem Kapitel soll zunächst das ursprüngliche Modell vorgestellt werden. Im Anschluß daran werden die von Mussweiler und Strack vorgeschlagenen Erweiterungen des SAM im Hinblick auf andere experimentelle Paradigmen in Form des so genannten Zwei-Stufen-Modells beschrieben. In einem weiteren Abschnitt wird versucht, eine systematische Zusammenstellung der bisher zum SAM vorliegenden experimentellen Befunde zu geben. Vor der Zusammenfassung des Kapitels werden verschiedene Aspekte der theoretischen Konzeption des Zwei-Stufen-Modells diskutiert.

Das Basismodell: Hypothesenkonformes Testen und semantisches Priming

Das SAM in seiner ursprünglichen Version postuliert, dass Ankereffekte im Standardparadigma auf der Kombination zweier fundamentaler Mechanismen basieren, nämlich *hypothesenkonformes Testen* (Trope & Liberman, 1996) und *semantisches Priming* (Higgins, 1996) (für eine vergleichbare Konzeption siehe Chapman & Johnson, 1994, 1999).

Bei der Beantwortung der Grobschätzungsaufgabe wird nach Mussweiler und Strack zunächst die Hypothese getestet, dass die zu schätzende Größe dem vorgegebenen Vergleichswert entspricht. Wird etwa gefragt, ob die Elbe kürzer oder länger als 3000 km sei, führt dies dazu, dass die Probanden die Hypothese: „Die Elbe ist 3000 km lang“, überprüfen (Mussweiler et al., 1997, S. 594). Hierbei wird auf eine „positive Teststrategie“ (Klayman & Ha, 1987; Snyder & Swann, 1978; Wason, 1960) zurückgegriffen, d. h. es werden selektiv Informationen generiert, die mit dieser Hypothese konform sind. Nach Mussweiler und Strack führt dies dazu, dass die Urteiler auf der Basis von semantischem und episodischem Wissen versuchen, ein mentales Modell (Johnson-Laird, 1983) des Urteilsobjektes zu konstruieren, dessen zu schätzende Ausprägung dem Anker entspricht (Mussweiler & Strack, 1999a, S. 146).

Die Konsequenz des hypothesenkonformen Testens muss jedoch nicht

notwendigerweise darin bestehen, die getestete Hypothese zu bestätigen (vgl. Trope & Liberman, 1996). In der Regel werden die Vergleichswerte ja so extrem gewählt, dass nach Beurteilung der gesammelten Evidenz die Hypothese, das Urteilsobjekt entspreche dem vorgegebenen Vergleichswert, mit hoher Wahrscheinlichkeit verworfen wird (Mussweiler, 1997, S. 33).

Die positive Teststrategie führt allerdings zur gezielten Generierung ankerkonsistenter Information. Dies erhöht die Zugänglichkeit dieser Information im Gedächtnis über den Zeitraum der Grobschätzungsphase hinaus, so dass sie in einer nachfolgenden Urteilsituation leichter zur Urteilsfindung herangezogen werden kann als Wissensseinheiten, die schwerer zugänglich sind, da sie nicht in einer vorangehenden Situation generiert wurden. Dieser Prozess wird von Mussweiler und Strack als semantisches Priming bezeichnet (vgl. Mussweiler, 1997; Mussweiler & Strack, 1999a, 1999b; Strack & Mussweiler, 1997).

Ob und in welcher Weise die generierten ankerkonsistenten Informationen die nachfolgende Feinschätzung beeinflussen, hängt von der *Anwendbarkeit* und der *Repräsentativität* der Information in Bezug auf das Absoluturteil ab. Im Einklang mit bestehenden Konzeptionen des semantischen Priming (vgl. Higgins & Brendl, 1995; Strack, 1992) determinieren diese beiden Faktoren die Intensität und die Richtung des Ankereffektes via semantischem Priming.

Die Anwendbarkeit von Informationen wird nach Higgins (1996) bestimmt durch das Ausmaß an Überschneidung mit dem Urteilsobjekt. Bezogen auf die Urteilsituation im Grob-/Feinschätzungsparadigma bedeutet dies, dass die leichter verfügbare, weil im Kontext der Grobschätzung generierte Information nur in dem Maße für die Feinschätzung verwendet wird, in dem sie auf die Feinschätzung bezogen werden kann. Das Standardparadigma liefert hierfür die besten Voraussetzungen, da beide Fragen auf dasselbe Objekt und dieselbe Urteilsdimension referieren. Sämtliche relevanten Informationen, die beispielsweise zur Testung der Hypothese, die Elbe sei 3000 km lang, generiert werden, lassen sich bei einer geforderten Längenschätzung der Elbe anwenden. Bezieht sich demgegenüber die komparative Frage auf die Verschmutzung der Elbe, die Feinschätzung jedoch auf ihre Länge, so kann die zur Bearbeitung der Vergleichsaufgabe generierte Information, etwa über die Industrieansiedlungen oder die Farbe des Wassers, nur in geringem Maße zur Längenschätzung herangezogen werden. Sie wird daher das Absoluturteil auch kaum beeinflussen (Mussweiler et al., 1997, S. 595). Die Anwendbarkeit bestimmt somit das *Ausmaß der Verwendung* zugänglicher Informationen.

Auf welche Weise die verwendete Information die Feinschätzung beeinflusst, hängt von ihrer Repräsentativität (vgl. Strack, 1992) ab. Ist die In-

formation repräsentativ für das zu beurteilende Objekt, wird sie als Basis für das Urteil herangezogen. Dies führt zu einer Assimilation des Urteils an die entsprechende Information. Ist die Information hingegen nicht repräsentativ, wird sie als Vergleichsstandard verwendet. Hieraus resultiert eine Kontrastierung des Urteils von der Information. Bezogen auf das Beispiel der Längenschätzung der Elbe würde letzteres bedeuten, dass im Falle einer vorangehenden Grobschätzung, deren Gegenstand der Mississippi ist, die über den Mississippi generierte Information als Vergleichsstandard herangezogen wird. Da die Elbe im Vergleich zu dem Mississippi ein relativ kleiner Fluss ist, würde hieraus eine kürzere Schätzung der Elbe resultieren (vgl. Mussweiler et al., 1997, S. 595). Der beschriebene Objektwechsel führt nach diesen Annahmen also nicht zu einem Assimilationseffekt, sondern zu einem Kontrasteffekt.

Allerdings bezieht sich diese Annahme auf eine Kontrastierung vom Urteilsobjekt der Grobschätzung *in Bezug zur unbeeinflussten Feinschätzung*. Werden wie im obigen Beispiel vor der Frage nach der Länge der Elbe Informationen über den Mississippi generiert, fällt die Schätzung der Elbe nach dem SAM geringer aus, als wenn *keine Informationen über den Mississippi* generiert wurden. Die Bezeichnung Kontrasteffekt referiert also nicht auf die unter diesen Bedingungen verwendeten Anker, sondern lediglich auf die beiden Urteilsobjekte (für eine eingehendere Diskussion dieser Problematik siehe S. 48).

Ist die während der Grobschätzung generierte ankerkonsistente Information bezüglich des Urteilsobjektes der Feinschätzung sowohl anwendbar als auch repräsentativ, führt dies dazu, dass bei der Suche nach Informationen zur Beantwortung der Feinschätzung verstärkt auf dieses Wissen zurück gegriffen wird. Die Feinschätzung basiert demnach in erster Linie auf ankerkonsistenter Information, wodurch das Urteil in Richtung des Ankers verzerrt wird (Mussweiler & Strack, 1999a, S. 147).

Eine zusammenfassende Darstellung dieses Modells ist in Abbildung 1 enthalten. Das SAM bezieht sich in dieser Form ausschließlich auf verschiedene Grob-/Feinschätzungsparadigmen. Die Vorhersagen, die Mussweiler und Strack aus ihrem Modell ableiten (vgl. etwa Mussweiler, 1997; Mussweiler et al., 1997; Mussweiler & Strack, 1999a; Strack & Mussweiler, 1997), sind folgende:

Das SAM postuliert zunächst Ankereffekte im Standardparadigma. Werden im Kontext von Dimensionswechseln in der Grobschätzungsphase Informationen generiert, die sich nur in geringem Maße auf die zu schätzende Dimension beim Feinurteil anwenden lassen, folgen allenfalls schwache Ankereffekte. Im Extremfall, wie im oben geschilderten Beispiel, bleiben nach

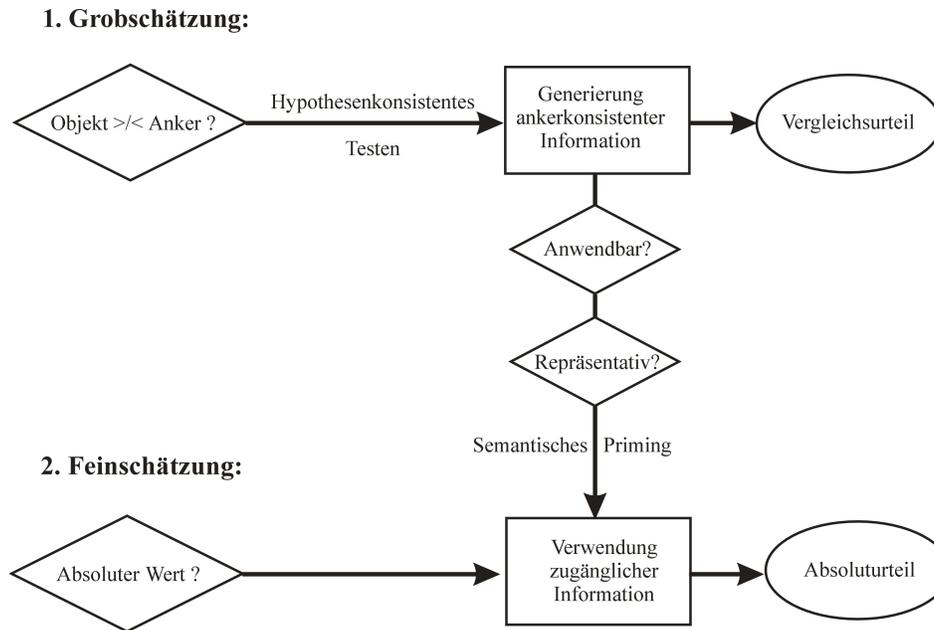


Abbildung 1: Das Modell der selektiven Zugänglichkeit (nach Mussweiler und Strack, 1999a, S. 155); nähere Erläuterungen im Text.

dem SAM Ankereffekte gänzlich aus. Selbiges gilt, wenn die Anwendbarkeit der generierten Information auf andere Weise verhindert wird, beispielsweise durch einen Objektwechsel oder bei der arithmetischen Komparation. Eine weitere Vorhersage bezieht sich auf einen Spezialfall des Objektwechsels. Stehen die beiden verwendeten Urteilsobjekte in einer Beziehung zueinander, die es gestattet, das Objekt der Grobschätzung als Vergleichsstandard für das Urteilsobjekt der Feinschätzung zu verwenden, können nach dem SAM die oben beschriebenen Kontrasteffekte auftreten. Dies ist nach Strack und Mussweiler dann der Fall, wenn beide Objekte derselben Kategorie angehören und sich hinsichtlich des zu schätzenden Merkmals stark voneinander unterscheiden.

Die Wirkung unplausibler Anker im Standardparadigma, wie auch das Auftreten von Ankereffekten in anderen experimentellen Paradigmen lassen sich mit dem Modell in der bisherigen Form zunächst nicht erklären. Um diese Lücke zu schließen, haben Mussweiler und Strack verschiedene Zusatzannahmen postuliert, die im folgenden Abschnitt diskutiert werden.

Erweiterungen: SAM als Kern eines Zwei-Stufen-Modells

In einer Reihe von Untersuchungen wurde gezeigt, dass auch durch die Verwendung völlig unplausibler Vergleichswerte im Standardparadigma Anker-effekte auslöst werden, die mitunter sogar extremer ausfallen, als bei plausiblen Vergleichswerten (vgl. Chapman & Johnson, 1994; Mussweiler & Strack, 1999b, 2000a, 2001a; Quattrone et al., 1984). Diese Befunde lassen sich mit dem SAM in seiner bisher vorgestellten Form nicht erklären. Zum einen erscheint es wenig plausibel, dass Urteiler beispielsweise die Frage: „Ist die Elbe länger oder kürzer als 45000 km?“ bearbeiten, indem sie aus der Frage die Hypothese: „Die Elbe ist 45000 km lang“, generieren und diese einer hypothesenkonformen Testung unterziehen. Zum anderen lassen sich auch beim besten Willen keine Argumente *für* diese Hypothese generieren, deren erhöhte Zugänglichkeit eine nachfolgende Feinschätzung in Richtung des Ankerwertes verzerren könnte.

Ähnliche Schwierigkeiten bereiten dem SAM experimentelle Paradigmen, bei denen kein expliziter Vergleich zwischen dem Ankerwert und der zu schätzenden Größe durchgeführt wird, wie beispielsweise beim Basic Anchoring oder bei implizit vorgegebenen Ankern. Kennzeichnend für diese Paradigmen ist, dass vor der Feinschätzung keine Informationen generiert werden, die auf das zu schätzende Objekt anwendbar sind. Daher besteht auch hier keine Möglichkeit, dass die durch die Feinschätzung ausgelöste Suche nach relevanten Informationen dem Einfluss leicht zugänglicher ankerkonsistenter Informationen unterliegt.

Für derartige Urteilsituationen postulieren Mussweiler und Strack nun ein Zwei-Stufen-Modell (vgl. Mussweiler & Strack, 1999a, 1999b). Im einzelnen nehmen die Autoren an, dass durch die Frage nach einer Feinschätzung in diesem Fall nicht sofort die Suche nach relevanten Informationen ausgelöst wird, sondern zunächst ein so genannter *impliziter Vergleich* (Mussweiler & Strack, 1999b, S. 157).⁹ Als Evidenz für die Plausibilität dieser Annahme zitieren sie Befunde, wonach Urteile im wesentlichen auf komparativen Prozessen basieren (Festinger, 1954; Helson, 1964; Kahneman & Miller, 1986).

Zur Durchführung dieses impliziten Vergleiches suchen die Urteiler in einem ersten Schritt nach einem geeigneten Vergleichsstandard in Form eines numerischen Wertes (Mussweiler & Strack, 1999a, S. 156 f.). Die Wahl die-

⁹Um möglichen Verständnisproblemen vorzubeugen, sei angemerkt, dass mit dem Begriff „implizit“ hier lediglich angezeigt werden soll, dass keine explizite Aufforderung gegeben wurde, einen Vergleich durchzuführen, wie beispielsweise im Standardparadigma. Eine weitergehende Bedeutung des Begriffes wie etwa im Sinne der kognitionpsychologischen Terminologie (vgl. etwa Baddeley, 1997, S. 351 ff.) ist meines Erachtens nicht intendiert.

ses Vergleichswertes kann durch drei Mechanismen geleitet werden. Als erste Möglichkeit diskutieren Mussweiler und Strack konversationale Inferenzen. Ist wie im Beispiel implizit vorgegebener Anker ein für die Urteilsituation relevanter Wert salient, kann dieser als Vergleichsstandard für den impliziten Vergleich verwendet werden. Als zweite Möglichkeit nennen Mussweiler und Strack die erhöhte Zugänglichkeit eines numerischen Wertes, also de facto numerisches Priming. Erfordert eine Voraufgabe die Auseinandersetzung mit einem numerischen Wert, erhöht dies die Zugänglichkeit des Wertes. Dadurch steigt die Wahrscheinlichkeit, das bei der Suche nach einem Vergleichsstandard dieser Wert gewählt wird, auch wenn kein inhaltlicher Zusammenhang zwischen der Voraufgabe und Urteilsobjekt der Feinschätzung besteht. Die dritte von Mussweiler und Strack vorgeschlagene Möglichkeit bezieht sich auf das Standardparadigma mit unplausiblen Ankern und bedarf einer etwas ausführlicheren Erläuterung.

Im Standardparadigma mit unplausiblen Ankern wird Grobschätzung nach Mussweiler und Strack (2000a) auf der Basis kategorialen Wissens beantwortet. Um zu entscheiden, ob die Elbe länger oder kürzer als 45000 km ist, ist es nicht notwendig, spezifisches Wissen über das zu schätzende Exemplar zu generieren. Stattdessen reicht es aus, auf vorhandenes Wissen über die Kategorie des Urteilsobjektes, in diesem Fall die Kategorie „Flüsse“, zurückzugreifen. Mit Hilfe des kategorialen Wissens kann nun entschieden werden, dass der Vergleichswert zu hoch in Bezug auf die gesamte Kategorie ist, da kein Fluss auf der Erde eine Länge von 45000 km erreicht. Hieraus resultiert dann die Antwort auf die Vergleichsfrage für ein Exemplar dieser Kategorie. Die Priorität von kategorialem Wissen gegenüber exemplarischem Wissen ist ein in anderen Urteilsbereichen abgesicherter Befund, falls ersteres zur Urteilsfindung hinreichend ist (Fiske & Neuberg, 1990).

Um innerhalb dieses Prozesses einen numerischen Wert ins Spiel zu bringen, der als Vergleichsstandard für den impliziten Vergleich herangezogen werden kann, postulieren Mussweiler und Strack für die Beantwortung der Grobschätzung unter Rückgriff auf kategoriales Wissen im einzelnen die folgenden Schritte: Im Falle eines hohen unplausiblen Ankers generieren die Urteiler mit Hilfe von kategorialem Wissen einen exklusiven Wert als Obergrenze für die Kategorie. Analoges gilt für unplausible niedrige Anker. Zur Bestimmung dieses Wertes wird ein Adjustierungsprozess, ausgehend vom vorgegebenen Anker, angenommen. Die Grobschätzung wird dann beantwortet auf der Basis eines Vergleiches zwischen dem unplausiblen Anker und der selbstgenerierten Kategoriegrenze (Mussweiler & Strack, 2001a, S. 149). Als Vergleichsstandard für den durch Feinschätzung ausgelösten impliziten Vergleich wird in diesem Fall die selbstgenerierte Kategoriegrenze, von Musswei-

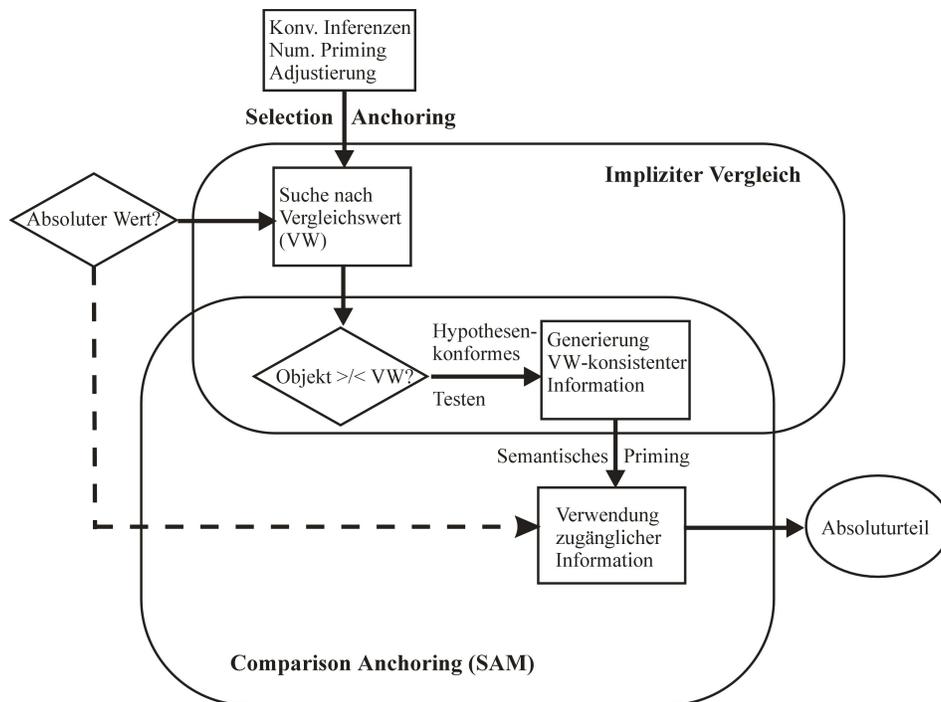


Abbildung 2: Die Integration des SAM mit anderen Modellen; nähere Erläuterungen im Text.

ler und Strack auch häufig als Grenze des Plausibilitätsintervalls bezeichnet (Mussweiler & Strack, 2000a, S. 1040), verwendet.

Den Einfluss, den vorgegebene numerische Werte auf die Wahl des Vergleichsstandards haben können, bezeichnen Mussweiler und Strack als *selection anchoring* (vgl. Mussweiler & Strack, 2001b, S. 252). Dieser Effekt kann nach Meinung der Autoren je nach dem vorliegenden experimentellen Paradigma auf konversationalen Inferenzen, numerischem Priming oder Adjustierung¹⁰ ausgehend von einem unplausiblen Anker basieren.

Allerdings wird die Auswahl des Vergleichswertes als nicht hinreichend erachtet, um eine nachfolgende Schätzung zu beeinflussen (vgl. Mussweiler

¹⁰Mussweiler und Strack (2001a, S. 149) sprechen in diesem Zusammenhang von *unzureichender* Adjustierung. Dieser Begriff scheint bezogen auf die postulierten Prozesse falsch gewählt, da das Ziel des Adjustierungsprozesses die Suche nach der entsprechenden Kategoriegrenze ist. Diesbezüglich ist der Adjustierungsprozess aber nicht unzureichend. Mussweiler und Strack behaupten ja explizit, dass auch unterschiedlich extreme unplausible Anker zu derselben Kategoriegrenze führen (Mussweiler & Strack, 2001a, S. 151).

& Strack, 2001b, S. 252). Notwendig hierfür ist ein Prozess, den Mussweiler und Strack als *comparison anchoring* bezeichnen. Hierunter verstehen die Autoren nun die Prozesse des SAM (vgl. Mussweiler & Strack, 1999a, 1999b, 2001b). Im einzelnen würde dies bedeuten, dass Urteiler den gewählten Vergleichsstandard zunächst hypothesenkonform testen und dabei selektiv Argumente dafür generieren, dass die zu schätzende Größe dem Vergleichsstandard entspricht. Diese Informationen erhalten dadurch eine höhere Zugänglichkeit. Bei der Suche nach Informationen über die zu schätzende Größe, auf deren Basis das Feinurteil gefällt wird, gehen sie dann aufgrund ihrer Zugänglichkeit in stärkerem Maße ein, und verzerren so das Urteil in Richtung des Vergleichswertes.

Eine Zusammenfassung dieser Annahmen ist in Abbildung 2 dargestellt.¹¹ Auf diese Weise erklären Mussweiler und Strack Ankereffekte jenseits des Standardparadigmas durch eine Integration des SAM mit den bereits diskutierten Modellen zum Ankereffekt.

Konversationale Inferenzen, numerisches Priming oder Adjustierung ausgehend vom unplausiblen Anker bilden je nach Urteilssituation die Basis für das selection anchoring bei der Wahl des Referenzstandards für den postulierten impliziten Vergleich. Diese Prozesse sind allerdings nicht hinreichend zum Entstehen von Ankereffekten. Sie stellen lediglich eine notwendige Voraussetzung dar, um auf einer zweiten Stufe, dem comparison anchoring, mit Bezug auf den Ankerwert Prozesse zu postulieren, die denen des SAM entsprechen. Als Resümee konstatieren Mussweiler und Strack: „... anchoring effects are semantic in nature ...“ (Mussweiler & Strack, 2001a, S. 159).

Empirische Befunde zum SAM

In diesem Abschnitt sollen die zentralen empirischen Befunde, die Mussweiler und Strack zur Stützung des SAM präsentiert haben, vorgestellt und zum Teil auch im Kontext der Befunde anderer Forschungsgruppen diskutiert werden.

Ankereffekte bei Grob-/Feinschätzungsparadigmen Eine erste empirische Argumentationlinie hat darin bestanden zu zeigen, dass die im Standardparadigma auftretenden Ankereffekte ausbleiben, wenn die Anwend-

¹¹Es sei darauf hingewiesen, dass Mussweiler und Strack an keiner Stelle ihrer Arbeiten die zweite Stufe ihres Modelles in der hier dargestellten Weise spezifizieren. Diese Spezifizierung ist jedoch meines Erachtens die einzig denkbare Interpretation der von Mussweiler und Strack in diesem Kontext verwendeten Formulierungen (siehe hierzu auch die Ausführungen auf S. 57 ff. dieser Arbeit).

barkeit der bei der Grobschätzung generierten Argumente minimiert wird. Hierzu präsentieren Mussweiler und Strack eine Reihe von Experimenten, in denen die Effekte im Standardparadigma kontrastiert werden mit denen in den vorgestellten Variationen desselben, nämlich Dimensionswechsel, Objektwechsel und arithmetischer Komparation.

Die präsentierten Daten sprechen eindeutig für die Annahmen des SAM. Signifikante Ankereffekte bleiben aus, wenn zwischen Grob- und Feinschätzung die Urteilsdimension wechselt (Strack & Mussweiler, 1997, Exp. 1; Mussweiler et al., 1997, Exp. 3), Grob- und Feinschätzung auf unterschiedliche Urteilsobjekte referieren (Strack & Mussweiler, 1997, Exp. 2) oder als Grobschätzung eine arithmetische Komparation (Mussweiler et al., 1997, Exp. 2) präsentiert wird.

Allerdings sind die damaligen ersten Befunde zum SAM aus heutiger Sicht aus zwei Gründen kritisch zu sehen. Zunächst hat sich in einer Reihe von Experimenten anderer Forschungsgruppen gezeigt, dass auch beim Objektwechsel, beim Dimensionswechsel oder bei der arithmetischen Komparation signifikante Ankereffekte auftreten (vgl. Becker et al., 2000; Wilson et al., 1996; Wong & Kwong, 1999, 2000). Diese Studien entsprechen nicht nur dem Aufbau nach den Experimenten von Strack und Mussweiler, sondern sind teilweise sogar direkte Replikationen davon (Becker et al., 2000, Exp. 1; Wong & Kwong, 1999, zitiert nach: Wong & Kwong, 2000).

Als Anlass für ihre Replikation nennen Wong und Kwong (2000, S. 329 f.) die geringen Stichprobengrößen in den Experimenten von Strack und Mussweiler. Im replizierten Experiment 1 (Strack & Mussweiler, 1997) befragten die Autoren 32 Probanden in einem 2×2 -"between-subjects" Design, was zu einer Zellenstichprobengröße von acht Personen führt.¹² Wong und Kwong spekulieren nun, dass das beobachtete Ausbleiben von Ankereffekten unter der Bedingung Dimensionswechsel das Resultat eines Stichprobenfehlers sein könnte. In ihrer Replikation mit einer größeren Probandenanzahl zeigt sich, dass unter der Bedingung Dimensionswechsel nicht nur signifikante Ankereffekte auftreten, sondern diese sogar vergleichbare Effektstärken aufweisen wie im Standardparadigma. Wong und Kwong (2000, S. 330) werten dies als Bestätigung ihrer Annahme. Analoge Befunde zeigen auch Becker et al. (2000, Exp. 1) bei der arithmetischen Komparation in einer Replikation von Mussweiler et al. (1997, Exp. 2) mit einer höheren Probandenanzahl.

¹²Diese Zellengrößen basieren darauf, dass die Daten über zwei weitere zweistufige Versuchsbedingungen aggregiert wurde. Berücksichtigt man diese, befragten Strack und Mussweiler unter jeder Versuchsbedingung exakt zwei Probanden (vgl. Strack & Mussweiler, 1997, S. 439 f.).

Ein weiterer Grund für die Skepsis besteht darin, dass Mussweiler und Strack zu einem späteren Zeitpunkt die bereits geschilderte modifizierte SAM-Version vorstellen. Dieses Modell nimmt explizit Bezug darauf, Ankereffekte beispielsweise auch beim Objektwechsel zu erklären. In späteren in diesem Kontext präsentierten Experimenten werden Ankereffekte in diesem Paradigma dann auch von Mussweiler und Strack (2001b, Exp. 3) beobachtet.

Von besonderem Interesse ist ein Befund von Strack und Mussweiler (1997, Exp. 2) aus den früheren Arbeiten, da dieser als Überprüfung der geschilderten Repräsentativitätsannahme präsentiert wird. Die Autoren untersuchten den Einfluss von Anker beim Objektwechsel, wenn die beiden verwendeten Urteilsobjekte aus derselben Kategorie stammen, innerhalb dieser Kategorie jedoch maximale Unähnlichkeit bezüglich der zu schätzenden Dimension aufweisen. Basierend auf ihren Ausführungen postulieren Strack und Mussweiler in diesem Fall Kontrasteffekte. Die vorliegenden Daten entsprechen dieser Hypothese, was auch in späteren Publikationen (vgl. Mussweiler, 1997; Mussweiler & Strack, 1999a) als Beleg für diesen Bestandteil des Modells diskutiert wird.

Obschon auf den ersten Blick überzeugend, werfen die Daten bei genauerer Betrachtung doch einige Fragen auf. Wie bei der Darstellung des Modells (vgl. S. 41) bereits erwähnt, bezieht sich der postulierte Kontrasteffekt auf eine Kontrastierung vom Urteilsobjekt der Grobschätzung im Verhältnis zur unbeeinflussten Feinschätzung. Der in Experiment 2 (Strack & Mussweiler, 1997) beobachtete Befund ist jedoch ein Kontrasteffekt in Bezug auf die in der Grobschätzung verwendeten Anker. Wurde in der Grobschätzung die Durchschnittstemperatur der Antarktis mit -20 °C (hoher Anker) verglichen, fiel die Feinschätzung der Durchschnittstemperatur auf Hawaii geringer aus, als wenn in der Grobschätzung der niedriger Anker -50 °C verwendet wurde.

Geht man davon aus, dass das Zielobjekt der Grobschätzung unter beiden Ankerbedingungen hinreichend verschieden vom Zielobjekt der Feinschätzung ist, wären unter beiden Bedingungen Kontrasteffekte im Vergleich zur unbeeinflussten „Hawaii“-Schätzung zu erwarten. Falls Unterschiede zwischen den beiden Ankerbedingungen auftreten, müssten diese, so lässt sich argumentieren, auf einen Assimilationseffekt hinauslaufen. Eine Entfernung vom mentalen Modell der Antarktis, das unter Berücksichtigung des hohen Ankers konstruiert wurde, sollte nämlich zu höheren Schätzungen der Temperatur auf Hawaii führen, als eine Entfernung von dem mentalen Modell, das auf den niedrigen Anker referiert.

Das berichtete Datenmuster ließe sich mit den Modellannahmen nur

dann erklären, wenn man annimmt, dass die mentalen Modelle, die bezüglich desselben Objektes mit unterschiedlichen Vergleichswerten generiert werden, zu *qualitativ* unterschiedlichen Einflüssen bezüglich seiner Verwendung als Vergleichsstandard führen. Ein Beispiel hierfür wäre, dass das mentale Modell der Antarktis unter der Bedingung „niedriger Anker“ zu Kontrasteffekten im Vergleich zur unbeeinflussten Schätzung führt, unter der Bedingung „hoher Anker“ jedoch nicht. Derartige Überlegungen lassen sich aber ebenso wie der im Modell ursprünglich postulierte Kontrasteffekt (vgl. Mussweiler et al., 1997, S. 595) anhand der vorliegenden Daten nicht überprüfen, da keine Kontrollgruppen erhoben wurden.

Wong und Kwong (2000, S. 317) weisen auf die Möglichkeit hin, dass der in diesem Experiment beobachtete Kontrasteffekt in Wahrheit ein Assimilationseffekt sein könnte. Vom Standpunkt ihrer Modellierung des numerischen Priming wirkt der etablierte numerische Wert *ohne* seiner Präfixe und Suffixe als Anker für eine nachfolgende Schätzung. Bezogen auf die „Hawaii“-Schätzung würde dies bedeuten, dass durch den im Kontext der Antarktis verwendeten hohen Anker ($-20\text{ }^{\circ}\text{C}$) der niedrige numerische Wert „20“ etabliert wird. Analog wird durch den niedrigen Anker ($-50\text{ }^{\circ}\text{C}$) der hohe numerische Wert „50“ etabliert. Eine umfangreiche experimentelle Überprüfung (Wong & Kwong, 2000, Exp. 3) mit fast dreihundert Probanden bestätigt diese Position.

Testung der Selektivitäts-Hypothese Die fundamentale Annahme des SAM besteht in der so genannten Selektivitäts-Hypothese (vgl. Mussweiler & Strack, 1999a, S. 147). Sie besagt, dass Urteiler aus der Grobschätzungsaufgabe die Hypothese, die zu schätzende Größe entspreche dem vorgegebenen Vergleichswert, generieren, und darauf hin selektiv Argumente generieren, die für diese Hypothese sprechen.

Um die Verwendung einer positiven Teststrategie während der Grobschätzung zu überprüfen, führten Mussweiler und Strack (1999b) zunächst zwei Experimente durch, in der mit verschiedenen Formulierungen der Grobschätzungsfrage gearbeitet wurde. Im ersten Experiment bekam eine Hälfte die Vergleichsaufgabe in dem Format: „Ist das Zielobjekt *größer* als der Anker?“, während die andere Gruppe gefragt wurde, ob das Zielobjekt *kleiner* als der Anker sei. Nach Mussweiler und Strack würden die Probanden die jeweiligen Fragen zunächst in die beiden Hypothesen: „Das Zielobjekt ist größer als der Anker“, bzw.: „Das Zielobjekt ist kleiner als der Anker“, überführen. Würden diese Hypothesen dann mit Hilfe einer positiven Test-

strategie bearbeitet, müssten die darauf folgenden Absolutschätzungen im ersteren Fall höher ausfallen als im zweiten.

Die erhobenen Daten bestätigten diese Hypothese. Sowohl unter der hohen als auch unter der niedrigen Ankerbedingung fielen die Absolutschätzungen höher aus, wenn in der Grobschätzung gefragt wurde, ob das Zielobjekt größer als der Anker sei.¹³

Das zweite Experiment dieser Arbeit widmete sich der Frage, ob die ursprüngliche Grobschätzung des Standardparadigmas, wie vom SAM behauptet, in die Hypothese, die zu schätzende Größe entspreche dem vorgegebenen Anker, überführt wird. Der Aufbau des Experimentes entsprach dem Experiment 1 mit anderen Frageformaten in der Grobschätzung. Neben der üblichen Grobschätzungsaufgabe wurde in der zweiten Versuchsbedingung gefragt, ob die zu schätzende Größe in etwa dem Anker entspreche. Beispielsweise musste ein Teil der Probanden von der Feinschätzung der Elbe angeben, ob die Elbe *länger oder kürzer als 890 km* sei. Der andere Teil bekam stattdessen die Frage, ob die Elbe *etwa 890 km lang* sei.

Nach der Argumentation von Mussweiler und Strack müssten beide Frageformate dazu führen, dass die Probanden die Hypothese, die zu schätzende Größe entspreche dem Ankerwert, konfirmatorisch testen. Da die Feinschätzung nach dem SAM auf der Zugänglichkeit der während der Grobschätzung generierten Argumente basiert, sollte lediglich die Wahl der Anker, nicht aber das Frageformat einen Einfluss auf die Feinschätzungen haben. Wiederum bestätigten die Daten diese Hypothese. Es ergaben sich keine Unterschiede bei den Feinschätzungen in Abhängigkeit vom verwendeten Frageformat.

Ein weiterer experimenteller Ansatz zur Überprüfung der Selektivitätshypothese bestand in der Kombination des Standardparadigmas mit Worterkennungsaufgaben. Gegenstand von Worterkennungsaufgaben ist es, so schnell wie möglich zu entscheiden, ob eine präsentierte Buchstabenkombination ein Wort darstellt. In einer Reihe von Untersuchungen wurde gezeigt, dass ein Wort umso schneller korrekt klassifiziert wird, je enger die Beziehung des dargebotenen Begriffes zu vorher aktivierten Konzepten ist (für einen Überblick hierzu siehe Neely, 1991).

¹³Allerdings überrascht, dass sich der Einfluss der während der Grobschätzung postulierten Prozesse zwar auf die Beantwortung der Feinschätzung, nicht jedoch auf die Beantwortung der Grobschätzung auswirkt. Der Anteil der Probanden, die das Zielobjekt oberhalb des hohen Ankers einschätzen, ist völlig unabhängig davon, ob die Antwort auf eine Bestätigung (Frage: „Ist das Zielobjekt größer als der Anker?“) oder Zurückweisung (Frage: „Ist das Zielobjekt kleiner als der Anker?“) der getesteten Hypothese hinausläuft. Selbiges gilt für die niedrige Ankerbedingung (siehe Mussweiler & Strack, 1999a, S. 144 f.).

Darauf aufbauend untersuchten Mussweiler und Strack (2000a), ob die Bearbeitung einer Grobschätzungsaufgabe die Klassifizierung von Begriffen erleichtert, die eine enge Beziehung zu Informationen aufweisen, die nach dem SAM während der Grobschätzung generiert werden. Wird etwa bei der Grobschätzung der Jahresdurchschnittstemperatur in Deutschland der Vergleichswert 5 °C präsentiert, sollten sich die darauf hin generierten Informationen vermehrt auf kalte Temperaturen beziehen. In einer daran anschließenden Worterkennungsaufgabe sollte dies dazu führen, dass beispielsweise Begriffe wie „Winter“, „Schnee“, „Ofen“ usw. schneller klassifiziert werden als Worte, die keine Beziehung zum Konzept „kalte Temperaturen“ aufweisen. Analoges gelte bei einem Vergleichswert von 20 °C für das Konzept „warme Temperaturen“.

Die Daten zweier Experimente (Mussweiler & Strack, 2000a, Exp. 1, Exp. 2) bestätigen diese Annahme. Unter beiden Ankerbedingungen wurden Begriffe schneller klassifiziert, wenn sie eine semantische Beziehung zu ankerkonsistenten Informationen aufwiesen. Obwohl die postulierten Unterschiede bei den Reaktionszeiten nur schwach waren, liefert dies zusätzliche Evidenz für die Selektivitätshypothese.¹⁴

Testung der Zugänglichkeits-Hypothese Die Zugänglichkeitshypothese besagt, dass die Feinschätzung primär auf während der Grobschätzung generierter ankerkonsistenter Information basiert, da diese zum Zeitpunkt der Feinschätzung eine höhere Zugänglichkeit besitzt.

Die empirischen Unterstützung dieser Hypothese besteht aus einer Vielzahl von Experimenten (Mussweiler, 1997, Exp. 1–3; Mussweiler & Strack, 1999b, Exp. 3; Mussweiler & Strack, 2000a, Exp. 4, Exp. 5; Strack & Mussweiler, 1997, Exp. 3), die im Wesentlichen den selben Aufbau besitzen und der selben inhaltlichen Logik folgen. Variiert wurden in diesen Experimenten die Ankerhöhe (hoch vs. niedrig) und die Plausibilität des Ankers (plausibel vs. unplausibel) in einem vollständigen 2 × 2-Design. Als abhängige Variablen dienten die *Bearbeitungszeiten* der Grobschätzung und der Feinschätzung.

Nach dem SAM wird bei der Grobschätzung mit plausiblen Ankern eine elaborierte Teststrategie verwendet und ein ankerkonsistentes mentales Modell des Urteilsobjektes konstruiert (vgl. Mussweiler & Strack, 1999a, S. 150). Die Grobschätzung mit unplausiblen Ankern wird hingegen auf der Basis kategorialen Wissens in weniger aufwändiger Weise beantwortet.

¹⁴Die Signifikanzniveaus der Paarvergleiche (semantische Beziehung: ja/nein) lagen bei *einseitiger Testung* bei 0.05, 0.07, 0.07 bzw. 0.26; Effektstärken wurden nicht angegeben.

Mussweiler und Strack postulieren daher, dass eine Grobschätzung mit plausiblen Ankern mehr Zeit benötigt als mit unplausiblen Ankern. Unter der Bedingung „plausibler Anker“ wird nach dem SAM bei der Feinschätzung auf das Wissen zurückgegriffen, das während der Grobschätzung aktiviert wurde, während unter der Bedingung „unplausible Anker“ zu diesem Zeitpunkt kein exemplar-spezifisches Wissen vorliegt. Mussweiler und Strack leiten hieraus die Hypothese her, dass nach der Vorgabe von plausiblen Ankern weniger Zeit für die Feinschätzung benötigt wird als nach Vorgabe von unplausiblen Ankern.

Die Daten sämtlicher oben angegebener Experimente in diesem Paradigma bestätigen die beiden Hypothesen. Während die Grobschätzung bei Vorgabe von unplausiblen Ankern schneller beantwortet wird als bei Vorgabe von plausiblen Ankern, findet sich bei den Feinschätzungen das umgekehrte Datenmuster.

Auch wenn diese Befunde im Einklang mit den Annahmen des SAM stehen, erscheint es jedoch fraglich, ob sie einen direkten Beleg für das Modell liefern. Dass eine vergleichsweise einfache Frage („Ist die Elbe länger oder kürzer als 45000 km?“) schneller beantwortet wird, als eine entsprechend schwerere Frage („Ist die Elbe länger oder kürzer als 890 km?“), lässt sich auch ohne die spezifischen Modellannahmen begründen. Berücksichtigt man, dass bei dem plausiblen Anker während der Grobschätzung (zeit-)intensiver über das Urteilsobjekt nachgedacht wird, erscheint es ebenfalls naheliegend, dass sich dieser Aufwand positiv auf die Bearbeitung der Feinschätzung auswirkt. Auch hierfür sind die Annahmen des Modells nicht zwingend erforderlich.

Darauf hinweisend, dass diese Befunde nur sehr indirekte Unterstützung für das SAM liefern (Mussweiler & Strack, 1999b, S. 147) haben Mussweiler und Strack in einem Experiment das oben beschriebene Design um einen weiteren Faktor ergänzt (Mussweiler & Strack, 1999b, Exp. 3). Während eine Hälfte der Probanden siebzehn Paare von Grob- und Feinschätzungsaufgaben unter den ursprünglichen Bedingungen präsentiert bekam, musste die andere Hälfte zusätzlich die jeweiligen Grobschätzungen unter Zeitdruck bearbeiten. Alle Aufgaben wurden auf einem Computer-Display präsentiert. Unter der Bedingung „Zeitdruck“ wurden die Probanden darauf hingewiesen, dass ihnen zur Beantwortung der Grobschätzungen nur fünf Sekunden zu Verfügung stehen, nach Ablauf dieser Zeit würde das Display automatisch die nächste Frage anzeigen. Für die Bearbeitung der Feinschätzungen gab es keine Zeitvorgaben.

Mussweiler und Strack (1999b, S. 147 f.) argumentieren, dass mit der Bedingung „Zeitdruck“ die Zugänglichkeit in direkterer Weise manipuliert

werden kann. Unter Zeitdruck werden während der Grobschätzung weniger Informationen generiert. Dies verringert den Umfang an zugänglicher Information während der Feinschätzung, was sich in längeren Bearbeitungszeiten bei der Feinschätzung niederschlagen sollte.

In den Daten von Experiment 3 zeigt sich, dass die Bearbeitungszeiten für die Feinschätzung signifikant länger waren, wenn die Grobschätzung unter Zeitdruck bearbeitet wurde. Neben dem Faktor „Zeitdruck“ war auch der Faktor „Plausibilität“ signifikant. Allerdings ergaben sich keine Interaktionseffekte zwischen diesen beiden Faktoren. Die Bearbeitungszeiten verlängerten sich also unter Zeitdruck unabhängig davon, ob plausible oder unplausible Anker vorgegeben waren.

Dieser Befund steht im Widerspruch zum getesteten Modell, da der geringere Umfang an zugänglicher Information verantwortlich für die längere Bearbeitungsdauer sein sollte. Bei unplausiblen Ankern werden aber nach dem SAM während der Grobschätzung keine Informationen generiert, die im Kontext der Feinschätzung Verwendung finden. Daher dürfte sich der Zeitdruck auch nicht auf die Bearbeitungszeiten der Feinschätzung bei unplausiblen Ankern auswirken.

Eine nahe liegende Erklärung für dieses Ergebnis ist, dass die längeren Bearbeitungszeiten unter der Bedingung „Zeitdruck“ das Resultat einer kognitiven Ausgleichsstrategie sind. Wenn bei der sukzessiven Bearbeitung zweier Aufgabentypen im Schema: „A, B, A, B, A, B, . . .“, die A-Aufgaben unter Zeitdruck bearbeitet werden müssen, erscheint es naheliegend anzunehmen, dass unabhängig vom Inhalt der Aufgaben der Zeitdruck quasi durch eine Art „mentales Durchatmen“ bei den B-Aufgaben kompensiert wird, was sich in längeren Bearbeitungszeiten der B-Aufgaben im Vergleich zu einer Kontrollgruppe ohne Zeitdruck-Bedingung niederschlägt.

Mussweiler und Strack sprechen diese Möglichkeit zwar an, halten sie aber für unwahrscheinlich. Als Begründung hierfür wird angegeben:

... the fact that plausibility and the time pressure manipulation have parallel effects suggests that both are mediated by a similar mechanisms, namely the differential generation of knowledge that is relevant for the subsequent absolute judgment (Mussweiler & Strack, 1999b, S. 151).

Diese Position erscheint aus drei Gründen kritisch: Zum einen ist das Argument, ähnliche Ergebnisse seien ein Indikator für ähnliche Ursachen, insbesondere angesichts nahe liegender Alternativerklärungen nicht sonderlich stark. Zum anderen ist der Schluss letztendlich zirkulär, da die Befunde unter der Bedingung „Zeitdruck“ erhoben wurden, um die Interpretation der

Befunde bei plausiblen vs. unplausiblen Ankern zu stützen, weil diese von den Autoren nur als sehr indirekte Belege für das SAM angesehen werden. Schließlich fehlt auch eine Spezifizierung, wie sich der Zeitdruck bei unplausiblen Ankern auf die Bearbeitungszeiten der Feinschätzung auswirken könnte, derart, dass dies im Einklang mit dem SAM steht.

Neben der Relevanz dieser Experimente für die Selektivitätshypothese, werden die oben beschriebenen Befunde auch im Hinblick auf das erweiterte SAM diskutiert. Die dazu gehörigen Interpretationen werden zusammen mit anderen Befunden im folgenden Abschnitt vorgestellt.

Testung des Zwei-Stufen-Modells Die von Mussweiler und Strack vorgelegten empirische Befunde zu dem Zwei-Stufen-Modell als Erweiterung des SAM beziehen sich nahezu ausschließlich auf das Standardparadigma mit unplausiblen Ankern. Neben der direkten Überprüfung einzelner Bestandteile und Annahmen werden von den Autoren auch zwei Ergebnisse bereits vorliegender Experimente, die im Kontext anderer Fragestellungen durchgeführt wurden, als Evidenz für das Zwei-Stufen-Modell angeführt.

Zum einen handelt es sich hierbei um den oben dargestellten Befund bezüglich der Bearbeitungsdauer von Grob- und Feinschätzung bei unplausiblen Ankern. Die kürzeren Bearbeitungszeiten bei der Grobschätzung wie auch die längeren Bearbeitungszeiten bei der Feinschätzung stehen nach Mussweiler und Strack (1999a, 2000a) im Einklang mit den Prozessen, die das Zwei-Stufen-Modell postuliert. Das andere Ergebnis, das in einer Reihe von Experimenten (vgl. Chapman & Johnson, 1994; Mussweiler et al., 1997; Mussweiler & Strack, 1999b; Strack & Mussweiler, 1997) beobachtet wurde, ist, dass unplausible Anker häufig zu stärkeren Ankereffekten führen als plausible Anker (für ein Gegenbeispiel siehe Wegener et al., 2001). Diesen Befund erklären Mussweiler und Strack ebenfalls über das Zwei-Stufen-Modell. Als Vergleichswert fungiert nach diesem Ansatz die entsprechende Grenze des Plausibilitätsintervalls, welche per definitionem extremer als ein plausibler Anker ist. Demzufolge werden beim hypothesenkonformen Testen auch Argumente generiert, die für eine extremere Ausprägung der zu schätzenden Größe sprechen, woraus extremere Absolutschätzungen resultieren (vgl. Mussweiler & Strack, 2000a). Andere Autoren hingegen diskutieren dieses Ergebnis als Konsequenz von numerischem Priming (vgl. Becker et al., 2003).

In zwei Experimenten überprüften Mussweiler und Strack (2001a) einen Aspekt des Zwei-Stufen-Modells für unplausible Anker in direkter Form. Es handelt sich hierbei um die Hypothese, dass unterschiedlich extreme unplausible Anker zu ähnlichen Absolutschätzungen führen. Diese von den Auto-

ren als zentral bezeichnete Annahme basiert darauf, dass unabhängig von der Extremität des unplausiblen Ankers in beiden Fällen die entsprechende Grenze des Plausibilitätsbereiches als Standard für den impliziten Vergleich gewählt werden sollte. Die auf der Basis kategorialen Wissens ermittelte Grenze des Plausibilitätsbereiches sollte ihrerseits unabhängig von dem vorgegebenen unplausiblen Anker sein (vgl. Mussweiler & Strack, 2001a, S. 151).

In beiden Experimenten zeigte sich, dass unterschiedlich extreme unplausible hohe Anker zu nahezu identischen Absolutschätzungen führten, während mit plausiblen hohen und niedrigen Ankern der ursprüngliche Ankereffekt repliziert wurde. Allerdings wirft dieser an sich klare Befund einige Fragen auf. Zum einen wurde von Becker et al. (2000) bereits vorher ein Experiment mit identischem Aufbau und Fragestellung durchgeführt, in dem sich die Feinschätzungen in Abhängigkeit von der Extremität des unplausiblen Ankers signifikant unterschieden.¹⁵

Wie lassen sich diese unterschiedlichen Ergebnisse erklären? Eine nahe liegende Erklärung für die Diskrepanz zwischen den Befunden von Mussweiler und Strack (2001a) und Becker et al. (2000) referiert auf die von Mussweiler und Strack verwendeten unplausiblen Ankern (vgl. auch Becker et al., 2003).

Mussweiler und Strack verwendeten Anker, die nicht auf der Basis von Vorexperimenten ausgewählt wurden und nicht nur unplausibel, sondern außerordentlich extrem waren. Beispielsweise wurden als hohe unplausible Anker für die Schätzung der Durchschnittstemperatur in der Antarktis die Werte +700 °C bzw. +900 °C vorgegeben. Dies verwundert umso mehr, als dass Mussweiler und Strack in einer früheren Arbeit (Mussweiler et al., 1997) für dasselbe Item einen deutlich weniger extremen Wert (+45 °C) als unplausiblen hohen Anker verwendeten. Dieser Wert besaß darüber hinaus den Vorteil, dass seine Unplausibilität durch eine Voruntersuchung nachgewiesen wurde.

Es erscheint offensichtlich, dass die von Mussweiler und Strack verwendeten Anker aus dem Bereich nicht nur unplausibler, sondern völlig absurder Vergleichswerte stammen. Wie bereits auf S. 17 f. diskutiert unterscheiden sich derartige Vergleichswerte hinsichtlich ihrer Wirkung auf eine nachfolgende Feinschätzung jedoch von der Wirkung lediglich unplausibler Anker.

Weitere Evidenz für diese Erklärung folgt daraus, dass den unplausiblen

¹⁵Ein „Hand-out“ dieses Experimentes wurde Mussweiler ca. sechs Monate vor der Erscheinung ihrer Arbeit auf Anfrage zugesandt. Das Experiment wurde allerdings im Rahmen ihrer Arbeit weder diskutiert noch erwähnt.

Ankern in beiden Experimenten Absolutschätzungen folgten, die auf demselben Niveau wie unter der Bedingung „hoher plausible Anker“ lagen.¹⁶ Dieser Befund steht im Gegensatz zu einer Vielzahl der von Mussweiler und Strack angeführten Experimente, wonach unplausible Anker, im Einklang mit dem Zwei-Stufen-Modell wie auch dem numerischem Priming, zu extremeren Absolutschätzungen führen als plausible Anker.

Zusammenfassend ist also festzuhalten, dass die Ergebnisse von Mussweiler und Strack (2001a) keinen überzeugenden Belege dafür darstellen, dass unterschiedlich extreme unplausible Anker prinzipiell zu Ankereffekten vergleichbarer Intensität führen, da sie, wie oben dargestellt, unter sehr speziellen Bedingungen aufgetreten sind und eine Generalisierung aus diesem Grund nicht naheliegend erscheint. Dass unterschiedlich extreme unplausible Anker grundsätzlich zur Generierung derselben Kategoriegrenze führen, ist damit ebenfalls nicht nachgewiesen. Es stellt sich auch generell die Frage, warum die Generierung einer Kategoriegrenze, die ja letztendlich ebenfalls eine numerische Schätzung darstellt, unabhängig von Faktoren sein sollte, deren Einfluss auf numerische Schätzungen als gesichert angesehen wird.

In einer weiteren Arbeit befassen sich Mussweiler und Strack (2000a) mit einem anderen Bestandteil des Zwei-Stufen-Modells, nämlich der Annahme, dass Grobschätzungen mit unplausiblen Ankern auf der Basis von kategorialen Wissen beantwortet werden. Ist dies der Fall, so müsste die höhere Zugänglichkeit kategorialen Wissens dazu führen, dass eine anschließende Feinschätzung der entsprechenden Kategoriegrenze schneller beantwortet wird als nach der Grobschätzung eines Exemplars mit einem plausiblen Anker. Wird in der Grobschätzung etwa erfragt, ob der Mississippi länger oder kürzer als 45000 km ist, sollte eine anschließende Feinschätzung nach der *maximalen Länge von Flüssen auf der Erde* schneller abgegeben werden, als wenn bei der Grobschätzung der Länge des Mississippi der Vergleichswert 2000 km verwendet wird.

Mussweiler und Strack (2001a, Exp. 4, Exp. 5) überprüften diese Annahme in zwei Experimenten. In beiden Fällen zeigte sich das postulierte Datenmuster. Eine Feinschätzung der Kategoriegrenze wird schneller be-

¹⁶Mussweiler und Strack (2001a, S. 154 f.) berichten zwar in beiden Experimenten einen signifikanten Effekt des Hauptfaktors Plausibilität. Dieser ist jedoch trivial und kann nicht als Evidenz für den üblichen Plausibilitätseffekt dienen, da hier der Einfluss zweier unplausibler hoher Anker mit dem Einfluss zweier plausibler Anker verglichen wird, von denen der eine ein hoher und der andere ein *niedriger* Anker ist. Vergleicht man hingegen den Einfluss der beiden unplausiblen hohen Anker mit dem adäquaten Referenzpunkt, nämlich dem Einfluss des plausiblen hohen Ankers, wird ersichtlich, dass die Absolutschätzungen in den drei Bedingungen nahezu identisch sind.

arbeitet, wenn bei der Grobschätzung eines entsprechenden Exemplars ein unplausibler Anker vorgegeben wurde. Allerdings liefern diese Befunde lediglich Evidenz dafür, dass bei der Bearbeitung der Grobschätzungsaufgabe mit unplausiblen Ankern auf kategoriales Wissen zurückgegriffen wird. Die spezifischen Annahmen des Zwei-Stufen-Modells für die Grobschätzung mit unplausiblen Ankern, wonach auf der Basis kategorialen Wissens eine entsprechende Grenze des Plausibilitätsintervalls konstruiert wird, können durch diese Ergebnisse nicht gestützt werden.

Erklärungsbedürftig erscheint insbesondere, dass bei expliziter Erfragung der Kategoriegrenze die entsprechenden Schätzungen einem extremen Einfluss durch die vorangegangene Grobschätzung unterliegen. Die maximale Länge, die Flüsse auf der Erde erreichen können, wird im Mittel auf 5560 Meilen (etwa 9000 km) geschätzt, wenn bei der vorangehenden Grobschätzung des Mississippi ein plausibler Anker vorgegeben wurde (vgl. Mussweiler & Strack, 2001a, Exp. 4). Bei Vorgabe des unplausiblen Ankers ist die Schätzung der Kategoriegrenze im Mittel mehr als doppelt so hoch (15872 Meilen \approx 25500 km). Wenn unter diesen Bedingungen die generierten Kategoriegrenzen derart differieren, warum sollten dann die im Rahmen eines impliziten Vergleichs generierten Kategoriegrenzen unabhängig von äußeren Einflüssen sein?

Neben den oben dargestellten Arbeiten präsentieren Mussweiler und Strack keine weiteren experimentellen Befunde, die die Annahmen des Zwei-Stufen-Modells stützen. Insbesondere finden sich auch keine Hinweise darauf, wie die Autoren sich die zweite Stufe der hypothesenkonsistenten Testung und des daran anschließenden semantischen Priming als Folge eines impliziten Vergleiches im Detail vorstellen und auf welche Weise sie diesen für empirisch prüfbar halten. Dies ist aber aus theoretischer Sicht der entscheidende Aspekt des Modells, da es sich im Wesentlichen lediglich dadurch von bereits existierenden Modellen des Ankereffektes unterscheidet.

Theoretische Anmerkungen zum Zwei-Stufen-Modell

Mit der Erweiterung des SAM in Form des beschriebenen Zwei-Stufen-Modells versuchen Mussweiler und Strack, Ankereffekte auch jenseits des Standardparadigmas im Kern auf das SAM zurückzuführen. Das Modell wird in einer Reihe von Arbeiten diskutiert (vgl. Mussweiler & Strack, 1999a, 1999b, 2000a, 2001a, 2001b), allerdings fehlen für dieses Modell bisher stützende empirische Belege. Die im letzten Abschnitt dargestellten Untersuchungen beziehen sich lediglich auf eine Voraussetzung für die postulierten Prozesse, nämlich in welcher Weise ein Vergleichswert für den impliziten Ver-

gleich bei Vorgabe von unplausiblen Ankern ins Spiel kommt. Die Ergebnisse stehen zum Teil im Widerspruch mit den Annahmen des Modells. Für den zentralen Bestandteil des Modells, wonach sämtliche Prozesse, die das SAM für die zwei Fragen des Standardparadigmas postuliert, nun unter entsprechenden Bedingungen sukzessiv durch die Frage nach einer Feinschätzung ausgelöst werden, fehlt bisher jegliche empirische Evidenz.

Aus theoretischer Sicht stellt sich neben einer Reihe konzeptioneller Schwierigkeiten (für eine eingehendere Diskussion siehe Becker et al., 2003) insbesondere die Frage, wie ein semantisches Priming innerhalb dieses Modells realisiert werden soll. Voraussetzung für einen Priming-Prozess ist die Existenz zweier getrennter Phasen: Eine erste Phase, in der bestimmte Wissensseinheiten aufgrund einer Auseinandersetzung mit entsprechenden Informationen oder Reizen eine erhöhte Zugänglichkeit erfahren, und eine zweite Phase, in der sich die erhöhte Zugänglichkeit dieser Information auf die Bearbeitung einer Aufgabe auswirkt (vgl. Baddeley, 1997; Higgins, 1996; Reisberg, 2001).

Während diese beiden Phasen in dem SAM beim Standardparadigma den beiden Schätzaufgaben zugeordnet sind, muss man bei dem Zwei-Stufen-Modell von der Annahme ausgehen, dass die Urteiler zunächst im Rahmen des impliziten Vergleiches hypothesenkonforme Informationen in Bezug auf den Vergleichswert generieren. Dieser Prozess soll durch die Frage nach einer Feinschätzung ausgelöst werden. Nachdem die Urteiler während des impliziten Vergleiches nun selektiv Informationen generiert haben, die für eine Ausprägung der Schätzgröße im Bereich des Vergleichsstandards sprechen, müsste im Anschluss eine zweite Phase beginnen, in der sich die Urteiler auf die Aufgabenstellung besinnen, eine Schätzung abzugeben, und hierfür, analog dem SAM im Standardparadigma, relevante Informationen suchen. Die während des impliziten Vergleiches generierten Informationen sollen nun aufgrund ihrer Zugänglichkeit bei dieser Suche in stärkerem Maße eingehen, und die Schätzung in Richtung des Vergleichswertes verzerren. Abgesehen davon, dass für diese Prozesse keinerlei empirische Evidenz vorliegt, mutet die Postulierung zweier aufeinanderfolgender Phasen der Informationsgenerierung, zunächst bezogen auf einen Vergleichswert, dann unabhängig von diesem Wert, deren Abfolge durch *eine* Fragestellung ausgelöst wird, und bei der die zweite Suche durch die während der ersten Suche generierten Informationen im Sinne eines *Primingprozesses* beeinflusst wird, nicht sonderlich plausibel an.

Man könnte an dieser Stelle einwenden, dass die Festlegung auf einen Primingprozess in Form zweier Phasen nicht von zentraler Bedeutung für das Modell ist, sondern lediglich der Einfluss generierter semantischer In-

formation, die für den Vergleichsstandard als mögliche Ausprägung des Urteilsobjektes spricht. Den Prozess des Generierens entsprechender semantischer Informationen ließe sich alternativ auch in Form eines „cued-retrieval“-Prozesses (vgl. etwa Baddeley, 1997) modellieren, bei dem der Vergleichswert die Funktion des „retrieval cues“ übernimmt. Im einzelnen würde dies bedeuten, dass zunächst nur *eine* Phase der Generierung urteilsrelevanter Informationen postuliert wird. Der im Kontext der Aufgabe etablierte numerische Wert könnte nun in Analogie zu bekannten Befunden aus dem Bereich der Gedächtnispsychologie (für eine Zusammenfassung siehe Schwartz & Reisberg, 1991, Kap. 8) dazu führen, dass semantische Informationen leichter zugänglich sind, die eine Assoziation zu den beiden aktivierten Konzepten, dem Urteilsobjekt und dem Ankerwert, beinhalten. Die explizite Durchführung eines Vergleiches zwischen Urteilsobjekt und Ankerwert wäre zwar denkbar, ist in diesem Zusammenhang aber nicht mehr zwingend notwendig.

Eine derartige Interpretation des Zwei-Stufen-Modells erscheint allerdings aus zwei Gründen problematisch: Zum einen stellt sich die Frage, was das Modell in dieser Form noch mit dem ursprünglichen SAM zu tun hat, und ob es gerechtfertigt ist, hierbei von einer Variante des SAM zu sprechen. Zum anderen wäre das Modell damit theoretisch wie empirisch letztlich nicht unterscheidbar von numerischem Priming. Das Modell des numerischen Priming besagt ja nicht, dass während des Urteilsprozesses keinerlei semantische Informationen bezüglich des Urteilsobjektes generiert werden, sondern dass der Einfluss auf die Schätzung, wie auf S. 34 ff. dargestellt, primär auf der Zugänglichkeit des entsprechenden Wertes basiert. Eine eher unspezifische Annahme dahingehend, dass ausgelöst durch die Zugänglichkeit des Ankerwertes auch ankerkonsistente semantische Informationen im Rahmen der Urteilsbildung generiert werden, scheint zum Erkenntnisgewinn im Rahmen des Modelles wenig beizutragen. Dies gilt umso mehr, als dass es weiterhin unklar wäre, ob diese Informationen, so sie denn generiert werden, die Basis für das numerische Urteil darstellen, oder eher eine Art „Rationalisierung“ des bereits gefällten Urteils, wie es aus anderen Bereichen der Urteilsbildung bekannt ist (vgl. etwa Zajonc, 1980).

Zusammenfassung

Das Modell der selektiven Zugänglichkeit von Mussweiler und Strack gilt weithin als etabliertes Modell zur Erklärung von Ankereffekten (vgl. Ariely, Loewenstein & Prelec, 2003; Chapman & Johnson, 2002; Epley & Gilovich, 2002). In seinem Kern besagt es, dass Ankereffekte nicht auf der Basis von

numerischem Priming, sondern von semantischem Priming zu Stande kommen. Bei einer zusammenfassenden Betrachtung entsteht folgendes Bild:

Für die ursprüngliche Version des SAM, die Ankereffekte im Standardparadigma erklärt, liegt eine umfassende empirische Evidenz vor. Obwohl einzelne Ergebnisse wie oben dargestellt durchaus offen für alternative Erklärungsansätze sind, und insbesondere die Interpretation der frühen Experimente, in denen verschiedene Grob-/Feinschätzungsparadigmen gegenübergestellt wurden, angesichts neuerer Befunde (vgl. etwa Becker et al., 2000; Mussweiler & Strack, 2001b; Wong & Kwong, 2000) fraglich erscheint, spricht unter Berücksichtigung der Gesamtheit der vorliegenden Befunde vieles dafür, dass die im SAM postulierten Prozesse am Zustandekommen des Ankereffektes im Standardparadigma beteiligt sind.

Für das Zwei-Stufen-Modell, mit dem Ankereffekte jenseits des Standardparadigmas unter Rückführung auf das SAM erklärt werden sollen, liegt bisher keine empirische Evidenz vor. Darüber hinaus erscheint dieses Modell auch aus theoretischer Sicht sehr spekulativ, und es ist unklar, ob das Modell in seiner bisherigen Form, wenn es über die zwei Stufen beispielsweise ein Zusammenspiel von numerischem Priming und semantischem Priming postuliert, einen Erklärungsbeitrag leistet, der über eine Erklärung via numerischem Priming hinausgeht. Es erscheint daher naheliegend, das SAM als Erklärungsmodell zunächst auf den Bereich zu beschränken, für den es auch ursprünglich intendiert war (vgl. Mussweiler, 1997; Strack & Mussweiler, 1997), nämlich für *Ankereffekte im Standardparadigma*.

Im Bereich des Basic Anchoring, wie auch bei anderen Grob-/Feinschätzungsparadigmen bietet sich nach bisherigem Kenntnisstand numerisches Priming als Erklärungsmodell an. Eine Integration semantischer Prozesse in dieses Modell setzt voraus, diese zunächst zu spezifizieren und anschließend einer empirischen Überprüfung zu unterziehen. Wenn jedoch numerisches Priming als Erklärungsmodell für Ankereffekte in anderen experimentellen Paradigmen in Betracht gezogen werden muss, stellt sich die Frage, ob diese Prozesse nicht auch neben den im SAM postulierten einen Beitrag zum Ankereffekt im Standardparadigma leisten.

Eigene empirische Fragestellungen

Im Folgenden sollen die beiden Fragestellungen vorgestellt werden, die Gegenstand der Experimente des empirischen Teils dieser Arbeit sind. Das erste Experiment bezieht sich unmittelbar auf eine Position innerhalb des Modells der selektiven Zugänglichkeit und versucht diese anhand einer Gegenüberstellung des Standardparadigmas mit einem eigens hierfür entwickelten experimentellen Paradigma zu überprüfen. Das zweite Experiment nimmt auf einen generelleren Aspekt Bezug, nämlich auf die Frage nach den Voraussetzungen dafür, dass ein numerischer Wert als Anker im Hinblick auf eine nachfolgende Schätzaufgabe wirkt.

Ankereffekte im Standardparadigma: SAM – allein zu Haus?

Die Ausführungen im letzten Kapitel legen dar, dass das SAM eine plausible und empirisch abgesicherte Erklärung für Ankereffekte im Standardparadigma liefert, hingegen in anderen experimentellen Paradigmen sich alternative Modelle, vorzugsweise das des numerischen Priming, anbieten. Die sich unmittelbar daran anschließende Frage, die ebenfalls schon im letzten Kapitel formuliert wurde, ist, ob Ankereffekte im Standardparadigma allein auf den im SAM postulierten Prozessen basieren, oder ob neben diesen auch numerisches Priming an den beobachteten Effekten beteiligt ist.

Mussweiler und Strack (2001b, S. 251) vertreten diesbezüglich die Position, dass rein numerische Einflüsse beschränkt sind auf Urteilsituationen, in denen semantische Prozesse nicht operieren können. Infolgedessen leistet numerisches Priming nach Meinung der Autoren keinen Beitrag am Zustandekommen des Ankereffektes im Standardparadigma. Diese Position erscheint aus theoretischer Sicht kritisch, da unter numerischem Priming ein automatisierter, auf reiner Aktivierung basierender Prozess verstanden wird, dessen auslösender Stimulus zunächst nur der präsentierte numerische Wert ist. Dass das Ablaufen dieses Prozesses verhindert wird durch zusätzliche eher deliberative Prozesse wie der Aquirierung und Bewertung semantischer Information, erscheint aus diesem Grund wenig plausibel.

Darüber hinaus tragen auch die von Mussweiler und Strack (2001b) hierzu präsentierten Experimente wenig zur Klärung dieser Frage bei, da hier wiederum nur Ankereffekte im Standardparadigma mit denen unter der Bedingung Objektwechsel verglichen werden, mit dem bekannten Ergebnis, dass im Standardparadigma stabilere Effekte zu beobachten sind.

Die Frage, ob neben dem SAM auch numerisches Priming am Zustandekommen des Ankereffektes im Standardparadigma beteiligt ist, verlangt

allerdings einen anderen experimentellen Ansatz. Hierfür müsste das Standardparadigma verglichen werden mit einer Variation dieses Paradigmas, welche zwei Eigenschaften aufweist: Zunächst sollten die experimentellen Bedingungen gewährleisten, dass sämtliche semantischen Prozesse, die nach dem SAM ursächlich für Ankereffekte sind, in analoger Weise ablaufen. Gleichzeitig sollte unter dieser Bedingung jedoch ein möglicher Einfluss rein numerischer Prozesse im Vergleich zum Standardparadigma minimiert werden. Wenn unter diesen Umständen Ankereffekte auftreten, die denen im Standardparadigma entsprechen, wäre dies ein Hinweis darauf, dass in der Tat nur semantische Prozesse für den Ankereffekt im Standardparadigma verantwortlich sind. Würde die Reduzierung numerischer Einflüsse hingegen zu einer Reduzierung des Ankereffektes führen, spräche dies für eine Beteiligung von numerischem Priming an Ankereffekten im Standardparadigma.

Das erste Experiment der vorliegenden Arbeit, *Ankereffekte beim Objektvergleich*, überprüft diese Fragestellung.

Bedingungen für numerisches Priming: Aufmerksamkeit und Aktiviation

Wie bereits dargestellt existieren eine Reihe von Arbeiten, die Ankereffekte zumindest jenseits des Standardparadigmas auf numerisches Priming zurückführen (vgl. etwa Becker et al., 2000; Brewer & Chapman, 2002; Wilson et al., 1996; Wong & Kwong, 2000). Auf der Basis dieser Arbeiten lassen sich anhand der präsentierten Befunde Aussagen über spezifische experimentelle Bedingungen herleiten, die günstig für das Auftreten von Ankereffekten als Folge von numerischem Priming sind. Einen besonderen Stellenwert diesbezüglich besitzt die Frage nach den notwendigen Bedingungen für die Ankeretablierung.

Aufbauend auf die innerhalb dieses Bereiches fundamentale Publikation von Wilson et al. (1996) wird der Grad der *Aufmerksamkeit*, die dem Ankerwert entgegengebracht wird, als zentrale intervenierende Variable angesehen (vgl. auch Chapman & Johnson, 2002, S. 123). In dem bereits auf S. 22 f. dieser Arbeit beschriebenen Experiment 3 von Wilson et al. (1996) zeigt sich, dass ein signifikanter Ankereffekt auftritt, wenn die Probanden vor der Schätzung fünf, jeweils sieben Zahlen enthaltende Seiten kopiert haben, nicht jedoch, wenn sie nur eine Seite kopiert haben. Dieses Ergebnis wird von den Autoren damit erklärt, dass den Zahlen unter letzterer Bedingung zu wenig Aufmerksamkeit entgegen gebracht wurde, so dass diese nicht ihre Wirkung auf eine nachfolgende Schätzaufgabe entfalten konnten. In der abschließenden Diskussion des Artikels wird diese Thematik im Kontext der

anderen Experimente noch einmal aufgegriffen. Explizit wird hierbei die Frage aufgeworfen, welcher Grad an Aufmerksamkeit einem numerischen Wert entgegen gebracht werden muss, damit dieser Wert im Kurzzeitgedächtnis einen Einfluss auf eine nachfolgende Schätzung hat (vgl. Wilson et al., 1996, S. 399).

Bezugnehmend auf die Aktivierung eines numerischen Konzeptes als Auslöser von numerischem Priming (vgl. S. 34 ff.) muss jedoch gefragt werden, ob es im Sinne einer Präzisierung der Bedingungen für Ankereffekte nicht notwendig ist, zu unterscheiden zwischen dem Grad an Aufmerksamkeit und dem Grad an Aktivierung, die der mentalen Repräsentation des numerischen Konzeptes zuteil wird. Unter Aufmerksamkeit wird eine kognitive Resource verstanden, die intentional zugeteilt wird und deren Zuteilung von Bewusstsein begleitet ist (vgl. Baddeley, 1997). Wird einer Information ein entsprechendes Maß an Aufmerksamkeit zuteil, führt dies natürlich zur Aktivierung entsprechender mentaler Konzepte. Bedeutsam ist die Unterscheidung zwischen Aufmerksamkeit und Aktivierung aus erklärungstheoretischer Sicht dann, wenn einem numerischen Konzept auch genügend Aktivierung zukommen kann, um als Anker für eine nachfolgende Schätzung zu wirken, *ohne* dass dies einen bewusstseinsbegleitenden Prozess wie den der Aufmerksamkeitszuwendung erfordert. Aus dem Nachweis dieser Möglichkeit würde bezugnehmend auf die oben erwähnte, von Wilson et al. aufgeworfene Frage die radikale Antwort resultieren, dass *keinerlei* Aufmerksamkeit notwendig ist, ja dass schon die Konzeption eines Wertes im Kurzzeitgedächtnis obsolet ist.

Ein extrem starker Beleg hierfür würde darin bestehen zu zeigen, dass die Darbietung numerischer Werte ein nachfolgendes Urteil beeinflusst, ohne dass die Urteiler über eine bewusste Wahrnehmung darüber verfügen, dass ihnen numerische Werte präsentiert wurden. Der Überprüfung dieser Möglichkeit widmet sich das zweite Experiment der vorliegenden Arbeit, *Ankereffekte durch subliminales Priming*.

Wären unter derartigen Bedingungen Ankereffekte nachzuweisen, hätte dies neben theoretischen auch eine Reihe praktischer Implikationen. Es muss zwar davon ausgegangen werden, dass es sich bei subliminalen Reizdarbietungen außerhalb des Versuchslabors um ein seltenes, möglicherweise gar nicht-existentes Phänomen handelt (vgl. Bargh, 1992). Allerdings legen eine Reihe von Untersuchungen nahe, dass es sich bei den unter diesen Umständen ausgelösten kognitiven Prozessen um dieselben handelt, die auch unter solchen Umständen auftreten, die eine hohe Alltagsrelevanz aufweisen (vgl. etwa Bargh, 1992; Merikle & Joordens, 1997). Eine ausführliche Diskus-

sion dieser Thematik erfolgt im Anschluss an das zweite Experiment dieser Arbeit.

Empirischer Teil

Ankereffekte beim Objektvergleich

Gegenstand des ersten Experimentes ist eine Untersuchung des Zusammenspiels von semantischer und numerischer Information beim Zustandekommen des Ankereffektes im Standardparadigma. Nach dem SAM (vgl. etwa Mussweiler, 1997) gehen diese Effekte darauf zurück, dass die Vergleichsfrage das Generieren von semantischer Information evoziert, die dafür spricht, dass das Zielobjekt hinsichtlich der gefragten Dimension in etwa dem Vergleichswert entspricht. Da diese Information nun zum Zeitpunkt der darauf folgenden Absolutfrage aufgrund ihrer vorherigen Aktivierung eine höhere Zugänglichkeit besitzt, beeinflusst sie nach dem SAM die Feinschätzung in Richtung des Ankers. Einen über die beschriebenen Prozesse hinausgehenden Einfluss der Präsentation eines numerischen Wertes innerhalb der Vergleichsfrage ist in diesem Modell nicht enthalten (Mussweiler & Strack, 2001b). Die Frage, ob neben semantischem Priming auch numerisches Priming am Zustandekommen des Ankereffektes *innerhalb dieses experimentellen Paradigmas* beteiligt ist, soll im Folgenden überprüft werden.

Die Methode, Einflüsse von semantischem und numerischem Priming innerhalb dieses Paradigmas zu separieren, soll darin bestehen, neben den üblichen Vergleichsfragen im Standardparadigma, bei der numerische Größen als Vergleichsstandard fungieren, auch solche zu verwenden, bei denen das zu schätzende Objekt hinsichtlich der relevanten Ausprägung mit anderen eher vertrauteren Objekten verglichen werden soll.

Die Logik, auf deren Basis dieses Experiment zur Trennung des Einflusses von semantischer und numerischer Information entwickelt wurde, ist die folgende: Wenn Ankereffekte lediglich auf den beschriebenen Prozess des hypothesenkonformen Testens während der Vergleichsfrage zurückgehen, sollte es im Hinblick auf das Generieren entsprechender Information keine Rolle spielen, ob man das Zielobjekt explizit mit einer numerischen Größe vergleicht oder mit einem vertrauten Objekt, das hinsichtlich der relevanten Ausprägung der numerischen Größe entspricht. Soll beispielsweise die Länge eines Maulwurfs in Zentimeter geschätzt werden, mögen die vorausgehenden Grobschätzungen im Standardparadigma danach fragen, ob ein Maulwurf länger oder kürzer als 9 Zentimeter (niedriger Anker) bzw. 30 Zentimeter (hoher Anker) ist. Alternativ dazu kann man fragen, ob ein Maulwurf länger oder kürzer als eine Zigarretten-Schachtel (niedriger Anker) bzw. die Seitenhöhe eines DinA4-Blattes (hoher Anker) ist. Unter der Voraussetzung, dass die Urteiler die beiden Vergleichsobjekte hinsichtlich ihrer Größe hinreichend homogen im Bereich der explizit vorgegebenen numerischen Werte einordnen, müsste nach dem SAM in beiden Fällen Ankereffekte vergleich-

barer Intensität auftreten, da bei der Suche nach konfirmatorischer Information für die Hypothese, dass die Länge eines Maulwurfs in etwa 9 Zentimeter bzw. der Länge einer Zigarrettenschachtel entspricht, dieselben Argumente generiert werden sollten. Wären die Effekte stärker, wenn in der Vergleichsaufgabe explizit ein numerischer Wert präsentiert wird, würde dies für eine Beteiligung von numerischem Priming am Zustandekommen des Ankereffektes sprechen.

Ein möglicher Einwand gegen diese Interpretation könnte lauten, dass beim hypothesenkonformen Testen unter der Bedingung Objektvergleich, sowohl konfirmatorische Information im Hinblick auf das Zielobjekt, als auch im Hinblick auf das Vergleichsobjekt generiert werden könnte. Verlangt man beispielsweise vor einer Absolutschätzung der Länge eines Maulwurfs ein Urteil darüber, ob ein Maulwurf länger oder kürzer als eine Zigarrettenschachtel ist, würde nach dem SAM die Hypothese: „Ein Maulwurf ist etwa so lang wie eine Zigarrettenschachtel“, durch die Suche nach hypothesenkonformer Information überprüft. Diese kann sich nun im Gegensatz zu der Situation beim Vergleich mit einer numerischen Größe sowohl auf die Länge eines Maulwurfs, als auch auf die Länge einer Zigarrettenschachtel, beziehen. Da letztere aber nicht auf das Zielobjekt der Absolutschätzung anwendbar ist, könnten schwächere Effekte unter der Bedingung Objektvergleich auch darauf zurückgehen, dass der Anteil der während des Vergleichsurteils generierten semantischen Information, die im Kontext des Absoluturteils *anwendbar* ist, geringer ausfällt, als beim Vergleich mit einer numerischen Größe.

Sollte diese alternative Erklärung zutreffen, müsste sich dies allerdings auch in einer verzerrten Schätzung der Vergleichsgröße gegenüber eine Probandengruppe, die keinen Vergleich durchgeführt hat, niederschlagen. Um diese Möglichkeit zu kontrollieren, werden sowohl von den Probanden, die einen Vergleich mit der Zielgröße durchführen, als auch von Probanden, die diesen Vergleich nicht durchführen, am Ende der Befragung Schätzungen der entsprechenden Vergleichsgrößen erhoben. Die größere zeitliche Distanz zwischen dem Vergleichsurteil und der Schätzung der Vergleichsgröße kann aus Sicht des SAM nicht als Erklärung für ein Ausbleiben von Effekten des Vergleichs auf die Schätzung der Vergleichsgrößen herangezogen werden, da, wie Mussweiler argumentiert, Ankereffekte, die auf semantischen Informationen basieren, eine extreme Stabilität aufweisen, also beispielsweise selbst dann auftreten, wenn die zeitliche Distanz zwischen Grobschätzung und Feinschätzung eine Woche beträgt (Mussweiler, 2001).

Vorstudie 1

Ziel dieser Vorstudie ist es, geeignete Größen auszuwählen, die bei Experiment 1 als Ziel- und Vergleichsobjekte verwendet werden können. Da die Zielobjekte auch in dem zweiten Experiment eingesetzt werden sollen, sollten die verwendeten Größen in der entsprechenden Einheit im einstelligen Bereich mit zentraler Tendenz um fünf herum eingeordnet werden (zur Begründung hierfür siehe die Ausführungen auf S. 85 f.). Darüber hinaus sollte in der Zielstichprobe (Studenten der Universität Kassel) eine relativ hohe Unsicherheit über die wahre Ausprägung der als Zielobjekte verwendeten Größen existieren. Als Vergleichsobjekte werden im Gegensatz dazu Größen gesucht, die von den Probanden mit relativ hoher Sicherheit im Bereich um eins bzw. um neun herum eingeordnet werden, um eine Vergleichbarkeit mit den entsprechenden numerischen Ankern „1“ bzw. „9“ zu gewährleisten.¹⁷

Methode

Vorgehen Vor dem Hintergrund obiger Überlegungen wurde mit zwei Klassen von Größen als Basis für mögliche Kandidaten für Ziel- und Vergleichsobjekte gearbeitet. Zum einen handelte es sich hierbei um die Einwohnerzahl von mehr oder minder vertrauten Städten, gemessen in Millionen Einwohner. Dabei wurde angestrebt, vertraute Städte auszuwählen, die im Bereich um eine Millionen bzw. neun Millionen Einwohner geschätzt werden, und weniger vertraute Städte auszuwählen, die im Bereich um etwa fünf Millionen Einwohner geschätzt werden, da erstere als Vergleichsobjekte, und letzte als Zielobjekte für Experiment 1 fungieren sollen. Die zweite Klasse von Größen bestand aus der Länge (bzw. dem Durchmesser) von Objekten, gemessen in Zentimeter. Auch hier wurde angestrebt, die Objekte derart auszuwählen, dass vertraute Objekte in der Tendenz am Rand, und weniger vertraute Objekte in der Tendenz im Zentrum des einstelligen Zahlbereiches eingeordnet werden.

Auf der Basis dieser Überlegungen wurde ein Fragebogen konzipiert, der aus jeder Klasse zehn Urteilsobjekte enthielt. Des weiteren wurden fünf Schätzaufgaben aufgenommen, bei denen *Anzahlen* im einstelligen Zahlbereich zu schätzen waren, die im Kontext der Untersuchung als Füller-Items dienen. Die Schätzungen wurden im „forced-choice“-Format erfragt, d. h. es

¹⁷Es wurde angestrebt, neben den Zielobjekten auch die verwendeten Anker in den beiden Experimenten konstant zu halten. Die Vorstudie 2, auf deren Basis die numerischen Anker ausgewählt wurden, war zum Zeitpunkt der Planung dieser Studie bereits abgeschlossen.

wurden explizit die Werte von eins bis neun als mögliche Antworten vorgegeben, mit der Anweisung, sich für eine der vorgegebenen Antworten zu entscheiden, also keine „Zwischenwerte“ anzukreuzen.

In einer kurzen Nachbefragung sollten die Probanden jeweils die zwei Aufgaben, die sie am leichtesten fanden und die zwei Aufgaben die sie am schwierigsten fanden, benennen. Über diese Befragung sollen zusätzliche Informationen über den Schwierigkeitsgrad der Schätzaufgaben erfasst werden. Des Weiteren sollten die Probanden, falls sie bei irgendwelchen Fragen im Zweifel waren, ob die zu schätzende Größe im Bereich zwischen eins und neun liegt, diese ebenfalls benennen.

Um die Motivation, an der Untersuchung teilzunehmen und möglichst sorgfältige Schätzungen abzugeben, zu erhöhen, wurde die Bearbeitung des Fragebogens mit einem Gewinnspiel verbunden. Unter den fünf Probanden, die die besten Schätzungen abgegeben, wurde eine Person ausgewählt, die einen Geldpreis in Höhe von 50 DM erhielt.

Material In der Untersuchung wurde ein siebenseitiger Fragebogen verwendet, auf dessen Deckblatt zunächst die vorgebliche Zielsetzung der Untersuchung, das Austesten von Frageformaten zur Entwicklung eines Allgemeinbildungstests, angegeben war. Des Weiteren bestand die Instruktion aus einer ausführlichen Erläuterung des Frageformates und des Procederes für das Gewinnspiel. Auf den folgenden fünf Seiten waren die 25 Schätzaufgaben aufgelistet. Die letzte Seite enthielt die oben angegebene, aus drei Fragen bestehende Nachbefragung und die Erhebung der Demographia. Von dem Fragebogen wurden vier verschiedene Versionen angefertigt, die sich lediglich in der Abfolge der Schätzfragen unterschieden, um eventuelle Reihenfolgeeffekte zu kontrollieren. Ein Exemplar des Fragebogens ist im Anhang, S. 137, enthalten.

Probanden An der Vorstudie nahmen 80 Studenten der Universität Kassel teil (53 männlich; 27 weiblich). Das Durchschnittsalter in der Stichprobe betrug 27 Jahre (SD: 4.2). Die Probanden wurden auf dem Universitätsgelände angesprochen und gefragt, ob sie Interesse hätten, an einer kurzen Befragung teilzunehmen. Die Zeitdauer der Untersuchung wurde mit etwa 10 Minuten angegeben. Es wurde darauf hingewiesen, dass die Möglichkeit bestünde, ein Preisgeld von 50 DM zu gewinnen. Die Bearbeitung der Fragebögen wurde im Einzelversuch durchgeführt, während dem Ausfüllen des Fragebogens war eine Versuchsleiterin anwesend, die den korrekten Ablauf überwachte und ggf. Fragen beantwortete.

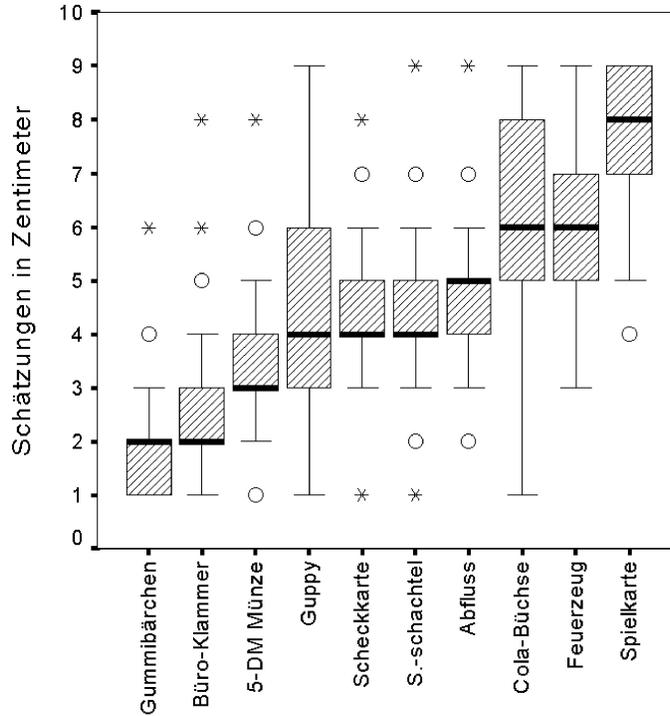


Abbildung 3: Boxplots der Längenschätzungen, Ausreißer sind mit einem Kreis, Extremwerte mit einem Stern gekennzeichnet, $N = 75$.

Ergebnisse und Diskussion

In die Auswertungen sind nur diejenigen Probanden eingegangen, die *alle* Schätzaufgaben beantwortet haben. Die Stichprobengröße reduziert sich damit auf 75 Personen. Hinsichtlich der vier verschiedenen Versionen des Fragebogens haben sich keine systematischen Zusammenhänge zwischen der Abfolge der Fragen und der einzelnen Schätzungen ergeben, so dass bei der Darstellung der Ergebnisse dieser Faktor nicht mehr berücksichtigt wird.

Die Ergebnisse bei den Längenschätzungen sind in Abbildung 3 dargestellt, weitere Auswertungen hierzu finden sich in Tabelle 8 im Anhang, S. 144. Auf der Basis dieser Ergebnisse ist zunächst als Zielitem die Schätzung des Guppys in Zentimetern ausgewählt worden. Dieses Item zeichnet sich zum einen dadurch aus, dass die abgegebenen Schätzungen ohne Ausreißer oder Extremwerte den gesamten Bereich möglicher Antworten

von 1 bis 9 abdecken, also eine breite Variation aufweisen und die zentrale Tendenz im Zentrum des Intervalls einstelliger Zahlen liegt. Darüber hinaus ist dieses Item in der Nachbefragung von 23 % der Probanden als eine der beiden schwierigsten Schätzaufgaben benannt worden, wohingegen lediglich 5 % der Befragten Zweifel geäußert haben, dass diese Größe im Bereich einstelliger Zahlen liegt (siehe Anhang, Tabelle 8, S. 144).

Als Vergleichsitem unter der niedrigen Ankerbedingung ist die Länge eines Gummibärchens in Zentimeter ausgewählt worden. Der Interquartilsbereich liegt bei diesem Item zwischen einem und zwei Zentimetern und das Item ist von 27 % der Befragten als eine der beiden einfachsten Schätzaufgaben benannt worden. Für die hohe Ankerbedingung scheint das Item „Länge einer Spielkarte in Zentimeter“ die beste Wahl für den Objektvergleich zu sein. Der Interquartilsbereich liegt zwischen sieben und neun Zentimetern, das Item ist von *keinem* Proband als eine der beiden schwierigsten und von 5 % der Probanden als eine der beiden einfachsten Aufgaben benannt worden. Neun Probanden haben Zweifel geäußert, ob die Länge einer Spielkarte im Bereich zwischen einem und neun Zentimetern liegt. Vor dem Hintergrund, dass es sich hierbei allerdings um eine relativ vertraute Größe handelt, ist nicht davon auszugehen, dass die Schätzungen dieser Probanden in einem offenen Frageformat sonderlich weit von der Obergrenze von neun Zentimetern entfernt gewesen wären.

Die Ergebnisse bei den Schätzungen der Einwohnerzahl von Städten in Mio. Einwohner sind in Abbildung 4 dargestellt, analog finden sich weitere Auswertungen hierzu im Anhang in Tabelle 9, S. 145. Als Zielitem dieser Kategorie ist auf der Basis der Ergebnisse die Einwohnerzahl von Santiago in Mio. Einwohner ausgewählt worden. Die Kriterien hierfür waren wiederum, dass die zentrale Tendenz der Schätzungen im Zentrum des Bereichs einstelliger Werte liegt, die Schätzungen eine breite Streuung aufweisen und das Item in der Nachbefragung von keinem Probanden als einfache, jedoch von 19 % der Befragten als schwierige Aufgabe benannt worden ist.

Die Auswahl der Vergleichsitems auf der Basis der Ergebnisse gestaltet sich in dieser Kategorie erheblich schwieriger als bei den Längenschätzungen. Das Item mit den im Mittel niedrigsten Schätzungen, die Einwohnerzahl von Dublin, erscheint als Vergleichsitem für die niedrige Ankerbedingung ungeeignet, u. a. da es von den Probanden in der Tendenz als eher schwierige Aufgabe eingestuft wird (vgl. Tabelle 9, S. 145). Die Schätzung der Einwohnerzahl von Hamburg ist zwar von 15 % der Probanden als eine der beiden einfachsten Aufgaben benannt worden, mit einer mittleren Schätzung von 2.6 Mio. und einem Interquartilsbereich, der den angestrebten Wert von 1 Mio. nicht enthält, erscheint es allerdings ebenfalls ungeeignet, um im Haupt-

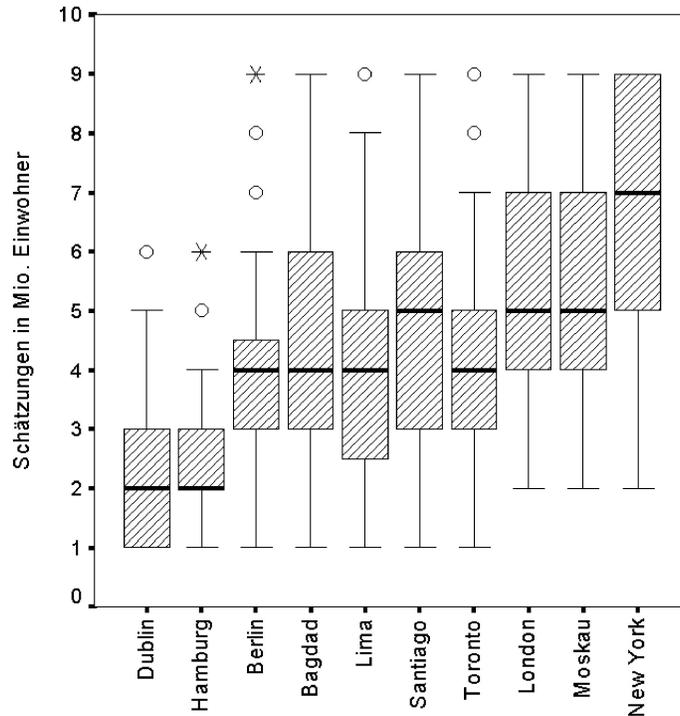


Abbildung 4: Boxplots der Einwohnerschätzungen, Ausreißer sind mit einem Kreis, Extremwerte mit einem Stern gekennzeichnet, $N = 75$

experiment als Vergleichsobjekt zu fungieren. Vor dem Hintergrund, dass die Vergleichsobjekte in dem Hauptexperiment ebenfalls geschätzt werden, und somit eine weitere Kontrolle bezüglich der angemessenen Realisierung der Versuchsbedingungen im Sinne der relevanten Fragestellung existiert, ist entschieden worden, ein in dieser Vorstudie nicht getestetes Vergleichsobjekt zu benutzen. Unter der Annahme, dass es sich um eine für die Probanden sehr bekannte Stadt handele, die hinsichtlich ihrer Einwohnerzahl kleiner als Hamburg eingeschätzt würde, wurde die Stadt Köln als Vergleichsitem für die niedrige Ankerbedingung ausgewählt.

Im Hinblick darauf, dass Schätzungen hinsichtlich ihrer zentralen Tendenz deutlich im oberen Bereich des einstelligen Zahlenspektrums liegen sollten, kommt unter den getesteten Items lediglich die Einwohnerzahl von New York als Vergleichsobjekt für die hohe Ankerbedingung in Frage. Allerdings weisen die Schätzungen bei diesem Item in der Stichprobe eine hohe

Variation auf und immerhin 23 % der Probanden äußerten Zweifel, dass die Einwohnerzahl von New York in Mio. Einwohner im einstelligen Zahlbereich liegt (siehe Tabelle 9, S. 145). Da es aber schwierig erscheint, eine Metropole mit etwa 9 Mio. Einwohnern zu finden, die einer deutschen Probandengruppe vertrauter als New York ist, ist wiederum vor dem Hintergrund der oben angesprochenen Kontrollmöglichkeit im Hauptexperiment entschieden worden, die Stadt New York als Vergleichsobjekt für die hohe Ankerbedingung zu verwenden. Die Alternative, stattdessen die Einwohnerzahl eines vertrauten Landes wie beispielsweise der Schweiz zu benutzen, ist verworfen worden, da der Vergleich zwischen einer Stadt und einem Land aufgrund der unterschiedlichen Assoziationen die Möglichkeit birgt, Kontrasteffekte auszulösen (vgl. hierzu auch Mussweiler, 2003).

Experiment 1

In dem nun folgenden Experiment soll überprüft werden, ob es qualitative Unterschiede hinsichtlich des Ankereffektes bei der Schätzung von Größen gibt in Abhängigkeit davon, ob diese in der Vorfrage mit numerischen Größen gemäß dem Standardparadigma verglichen werden, oder mit vertrauten Vergleichsobjekten, die hinsichtlich der relevanten Ausprägung in etwa den numerischen Werten entsprechen.

Die verwendeten Zielgrößen in diesem Experiment sind die Länge eines Guppys in cm, sowie die Einwohnerzahl der chilenischen Hauptstadt Santiago in Mio. Einwohner. Als niedrige Anker im Standardparadigma werden die Werte „1 cm“ bzw. „1 Mio.“ verwendet, als hohe numerische Anker die Werte „9 cm“ bzw. „9 Mio.“. Unter der Versuchsbedingung „Objektvergleich“ werden für das Zielitem *Guppy* die Länge eines Gummibärchens bzw. die Länge einer Spielkarte als niedriger resp. hoher Anker verwendet. Für die Schätzungen der Einwohnerzahl von *Santiago* werden die Einwohnerzahlen von Köln und New York als niedriger bzw. hoher Anker verwendet.

Methode

Material Zur Durchführung des Experimentes wurden vier verschiedene Versionen eines Fragebogens entworfen. Jeder Fragebogen enthielt 25 Fragen, 19 davon waren Füller-Fragen, die in allen vier Versionen identisch waren und die keine Schätzung einer numerischen Größe zum Gegenstand hatten. Die sechs kritischen Fragen waren folgendermaßen auf die Fragebögen verteilt: In der Frage 5 wurde zu einem Objektvergleich zwischen einem der beiden verwendeten Zielobjekte und einem entsprechenden Vergleichsobjekt

Tabelle 2: Fragebogenversionen zu Experiment 1

		Version A	Version B	Version C	Version D
Objekt- vergleich	Frage 5	Guppy <> Gummibär?	Santiago <> Köln?	Guppy <> Spielkarte?	Santiago <> New York?
	Frage 6	FS Guppy	FS Santiago	FS Guppy	FS Santiago
Standard- Paradigma	Frage15	Santiago <> 1 Mio.?	Guppy <> 1 cm?	Santiago <> 9 Mio.?	Guppy <> 9 cm?
	Frage 16	FS Santiago	FS Guppy	FS Santiago	FS Guppy
Vergleichs- objekt 1	Frage 22	FS New York	FS Spielkarte	FS Köln	FS Gummibär
Vergleichs- objekt 2	Frage 25	FS Gummibär	FS Köln	FS Spielkarte	FS New York

Anmerkung: FS: Feinschätzung

jekt aufgefordert, daran anschließend erfolgte in Frage 6 die Feinschätzung des entsprechenden Zielobjektes. Das Fragenpaar 15 und 16 enthielt eine Anordnung gemäß dem Standardparadigma mit dem jeweils anderen Zielitem. Innerhalb eines Fragebogens wurde bei beiden Aufgaben entweder mit zwei niedrigen oder mit zwei hohen Anker gearbeitet. In Frage 22 wurde die Feinschätzung eines Vergleichsobjektes erfragt, mit dem *kein* Vergleich durchgeführt wurde. Hierbei wurde stets nach dem Vergleichsobjekt gefragt, dass der entgegengesetzten Ankerbedingung bezüglich des Zielobjektes der Frage 16 zugeordnet war. In Frage 25 wurde eine Feinschätzung des Vergleichsobjektes erfragt, mit dem in Frage 5 der Vergleich durchgeführt wurde.

Auf diese Weise wurde realisiert, dass jeder Proband sowohl innerhalb des Standardparadigmas als auch innerhalb des als Objektvergleich bezeichneten Paradigmas jeweils eine der beiden Zielgrößen schätzen musste. Des Weiteren wurden von jedem Proband Feinschätzungen von zwei der vier in diesem Experiment verwendeten Vergleichsobjekte erfragt. Das erste Vergleichsobjekt wurde von der entsprechenden Person nicht verwendet; in der letzten Frage des Fragebogen musste das Vergleichsobjekt geschätzt werden,

mit dem in Frage 5 ein Vergleich durchgeführt wurde. Zur Übersicht sind die vier Versionen noch einmal in Tabelle 2 dargestellt.

Die 25 Fragen waren auf fünf Seiten verteilt. Hierbei wurde darauf geachtet, dass die aufeinander folgenden Grob- und Feinschätzungsaufgaben jeweils auf separaten Seiten abgedruckt waren. Zusammen mit dem Deckblatt, das die Versuchsinstruktionen enthielt und einer abschließenden Erhebung einige demographischer Angaben bestand der Fragebogen somit aus sieben Seiten. Ein Exemplar des verwendeten Fragebogens ist im Anhang, S. 146, beigelegt.

Probanden An dem Experiment nahmen 120 Studenten der Universität Kassel teil, davon waren 86 männlich. Das durchschnittliche Alter betrug 26 Jahre (SD: 4.2). Die Probanden wurden auf dem Universitätsgelände angesprochen und gefragt, ob sie Interesse hätten, an einer kurzen Fragebogenstudie teilzunehmen. Der zeitliche Aufwand wurde mit etwa fünfzehn Minuten angegeben, die Probanden erhielten für ihre Teilnahme keine finanzielle Entschädigung. Das Experiment wurde im Einzelversuch durchgeführt, während der Bearbeitung des Fragebogens war eine Versuchsleiterin anwesend, die ggf. Fragen beantwortete, und den Ablauf des Experimentes überwachte.

Experimentelles Design und Hypothesen Bezüglich der Schätzungen der Zielgrößen entspricht das vollständige experimentelle Design einem 2 (*Anker*: niedrig vs. hoch) \times 2 (*Item*: Guppy vs. Santiago) \times 2 (*Vergleich*: Objektvergleich vs. Standardparadigma) – Plan. Der Faktor *Anker* wurde hierbei „between-subjects“ variiert, die beiden Faktoren *Item* und *Vergleich* waren in einem „two period crossover“ – Design angeordnet (vgl. Cotton, 1989).

Vor dem Hintergrund der Überprüfung der Manipulationsbedingung, die wie im Folgenden geschildert getrennt nach betrachteter Zielgröße erfolgt und als Voraussetzung für die Interpretation der Daten dient, erscheint es allerdings sinnvoll, dieser Trennung auch bei dem Design in Bezug auf die Zielgrößen Rechnung zu tragen. Das Design reduziert sich somit für jede Zielgröße auf einen 2 (*Anker*: niedrig vs. hoch) \times 2 (*Vergleich*: Objektvergleich vs. Standardparadigma) – Plan, bei dem beide Faktoren „between subjects“ variiert wurden.

Die erste Hypothese in Bezug auf die Schätzungen der verwendeten Zielobjekte postuliert eine Replikation des üblichen Ankereffektes:

H.1.1: Die Absolutschätzungen der Zielobjekte sind unter den hohen Ankerbedingungen höher als unter den niedrigen Ankerbedingungen.

Von entscheidender Relevanz für die zentrale Fragestellung des Experimentes ist die zweite Hypothese, die auf die differentiellen Unterschiede des Ankereffektes in Abhängigkeit von der Art der Vergleichsaufgabe Bezug nimmt. Sie lautet:

H.1.2: Unter der Bedingung Objektvergleich fallen die postulierten Ankereffekte geringer aus, als im Standardparadigma.

Um Unterschiede hinsichtlich der Intensität der Effekte in Abhängigkeit von der Art der Vergleichsaufgabe (Standardparadigma vs. Objektvergleich) im oben beschriebenen Sinne interpretieren zu können, ist allerdings zunächst zu überprüfen, ob die Manipulationsbedingung erfolgreich war. Hierzu gehört zum einen, dass die Vergleichsobjekte hinsichtlich der relevanten Ausprägungen von den Probanden hinreichend homogen im Bereich der im Standardparadigma verwendeten numerischen Anker eingeordnet wurden.

Des weiteren sollte mit Hilfe geeigneter Korrelationstechniken überprüft werden, ob ein Zusammenhang zwischen den Schätzungen der verwendeten Vergleichsobjekte und den Schätzungen der Zielgrößen *innerhalb einer Ankerbedingung* besteht. Ein signifikanter Zusammenhang innerhalb einer Ankerbedingung in Verbindung mit einer starken Variation bei den Schätzungen der Vergleichswerte, wäre ein deutlicher Beleg dafür, dass die Schätzungen der Zielobjekte unter der Bedingung Objektvergleich nicht in dem ursprünglich intendiertem Sinne mit den Schätzungen im Standardparadigma vergleichbar wären. Wenn nämlich starke interindividuelle Unterschiede hinsichtlich Vorstellungen der Vergleichsobjekte existieren, die sich auf die Schätzungen der Zielobjekte auswirken, würde dies dafür sprechen, dass die Probanden innerhalb einer Ankerbedingung mit ein- und demselben Vergleichsobjekt auf unterschiedliche Größenbereiche verankert werden und sich diese Unterschiede auf die Schätzungen der Zielobjekte auswirken.

Darüber hinaus ist zu überprüfen, ob die Durchführung eines Vergleichs einen Einfluss auf die Schätzung des Vergleichsobjektes hat. Dies könnte gegebenenfalls ein Hinweis darauf sein, dass sich der in der H.1.2 postulierte Effekt in der oben beschriebenen Weise erklären ließe, derart dass beim Objektvergleich hypothesenkonforme Argumente generiert werden, die sich *sowohl* auf das Zielobjekt *als auch* auf das Vergleichsobjekt beziehen. Hieraus

würde entsprechend der Logik des SAM folgen, dass die Schätzungen der Vergleichsobjekte unter der Bedingung, dass ein Vergleich durchgeführt wurde, in Richtung der Größen der Zielobjekte verzerrt sind, und stärkere Ankereffekte im Standardparadigma darauf basieren, dass sich die während der Grobschätzung generierten hypothesenkonformen Argumente in diesem experimentellen Paradigma ausschließlich auf die Zielobjekte beziehen.

Es erscheint in diesem Zusammenhang sinnvoll, die Manipulationsbedingungen zunächst getrennt für jedes der beiden verwendeten Zielobjekte zu überprüfen, und im Anschluß daran die Hypothesen in Bezug auf die entsprechenden Zielobjekte getrennt zu testen, sofern die Überprüfung der Manipulationsbedingungen zufriedenstellend verlief.

Unter Berücksichtigung der Separierung nach Art des Zielobjektes entspricht das experimentelle Design im Hinblick auf die Schätzungen der Vergleichsobjekte jeweils einem $2 \text{ (Vergleichsobjekt: hoher Anker vs. niedriger Anker)} \times 2 \text{ (Vergleich durchgeführt: ja vs. nein)}$ – Plan. Den vier Gruppen sind jeweils separate Stichproben zugeordnet.

Ergebnisse

Überprüfung der Manipulationsbedingung Die Überprüfung der Manipulationsbedingung erfolgt wie geschildert getrennt nach der Art des Zielobjektes.

Bei dem Zielobjekt *Guppy* wurde unter der niedrigen Ankerbedingung die Länge eines Gummibärchens in Zentimeter als Vergleichsobjekt verwendet, unter der hohen Ankerbedingung die Länge einer Spielkarte in Zentimeter. Das Gummibärchen ist von den Probanden, die einen Vergleich mit dem Zielobjekt durchführten, im Mittel auf 1.92 cm geschätzt worden (SD: 0.94; $n = 30$), die Schätzungen der Kontrollgruppe liegen im Mittel bei 1.62 cm (SD: 0.52; $n = 30$). Der Unterschied zwischen den beiden Bedingungen wird von einem U-Test¹⁸ als nicht signifikant ausgewiesen ($U = 378$; $p = .28$).

Die Länge einer Spielkarte ist von den Probanden, die einen Vergleich mit dem Zielobjekt durchführten, im Mittel auf 9.92 cm geschätzt worden (SD: 2.97; $n = 30$), in der Kontrollgruppe liegen die Schätzungen im Mittel bei 10.90 cm (SD: 7.79; $n = 30$).¹⁹ Auch in diesem Fall ist der Unterschied

¹⁸Da die Voraussetzungen für parametrische Verfahren nicht erfüllt waren, und es sich lediglich um den Vergleich zwischen zwei Stichproben handelte, wurde ein verteilungsfreies Verfahren verwendet.

¹⁹Die Mittelwertdifferenz von einem Zentimeter (und die Differenz bei den Standardabweichungen) geht auf einen Ausreißer in der Kontrollgruppe zurück (Schätzung: 50 cm);

zwischen den Schätzungen unter den beiden Bedingungen nicht signifikant ($U = 412.5$; $p = .57$).

Die mittleren Schätzungen der Vergleichsobjekte liegen also bei den Probanden, die einen Vergleich durchgeführt haben, in etwa in der Höhe der explizit vorgegebenen numerischen Anker (niedriger Anker: 1 cm vs. 1.92 cm; hoher Anker: 9 cm vs. 9.92 cm). Die Abweichungen von den numerischen Ankern sind nicht nur gering, sondern weisen unter beiden Ankerbedingungen mit nahezu identischer Intensität in dieselbe Richtung. Darüber hinaus ergeben sich keine signifikanten Unterschiede bei den Schätzungen der Vergleichsobjekte in Abhängigkeit davon, ob ein Vergleich mit dem Zielobjekt durchgeführt wurde.

Zur Überprüfung, ob unter der Bedingung Objektvergleich die Variation in der Vorstellung der Größe der verwendeten Vergleichsobjekte einen Einfluss auf die Schätzungen des Zielobjektes hat, ist eine Partialkorrelation zwischen den Schätzungen der verwendeten Vergleichsobjekte und des Zielobjektes berechnet worden. Nach Herausparsialisierung des Faktors Anker liegt die Korrelation bei $r = 0.06$ ($p = .63$). Es ergibt sich also kein Zusammenhang zwischen den Schätzungen von Vergleichs- und Zielobjekt innerhalb der Ankerbedingungen. Neben den geringen Streuungen bei den Schätzungen der Vergleichsobjekte liefert dies zusätzliche Evidenz dafür, dass die Manipulationsbedingung bezüglich des Zielobjektes Guppy beim Objektvergleich in angemessener Weise realisiert wurde.

Ein anderes Bild ergibt sich bei der Analyse der Vergleichsobjekte für das Zielobjekt *Santiago*. Das Vergleichsobjekt unter der niedrigen Ankerbedingung, die Einwohnerzahl von Köln, ist von den Probanden, die einen Vergleich durchgeführt haben, im Mittel auf 2.22 Mio. geschätzt worden (SD: 2.05; $n = 30$). Die Kontrollgruppe hat die Einwohnerzahl von Köln im Mittel auf 4.99 Mio. (SD: 14.56; $n = 29$) geschätzt.²⁰ Der Unterschied zwischen den Schätzungen unter den beiden Bedingungen ist wiederum nicht signifikant ($U = 376$; $p = .37$).

Das Vergleichsobjekt unter der hohen Ankerbedingung, die Einwohnerzahl von New York, ist von den Probanden, die einen Vergleich durchgeführt haben, im Durchschnitt auf 14.83 Mio. geschätzt (SD: 18.77; $n = 29$) worden. Die Kontrollgruppe hat diese Größe lediglich im Mittel auf 6.77 Mio. (SD:

schließt man diesen von der Analyse aus, liegen die Schätzungen in der Kontrollgruppe im Mittel bei 9.55 cm (SD: 2.54).

²⁰Die Mittelwertdifferenz und die Differenz bei den Standardabweichungen geht auf einen Ausreißer in der Kontrollgruppe zurück (Schätzung: 80 Mio.). Schließt man diesen von der Analyse aus, liegen die Schätzungen in der Kontrollgruppe im Mittel bei 2.31 Mio. (SD: 2.01).

3.60; $n = 29$) geschätzt. Der Unterschied zwischen den beiden Bedingungen ist signifikant ($U = 263.5$; $p < .015$), weist jedoch in die entgegengesetzte Richtung, wie unter der Assimilationshypothese vermutet. Bei den Probanden, die bei dem Zielobjekt Santiago einen Objektvergleich durchgeführt haben, weisen die mittleren Schätzungen der Vergleichsobjekte also eine erhebliche Distanz zu den im Standardparadigma verwendeten numerischen Werten auf (niedriger Anker: 1 Mio. vs. 2.22 Mio.; hoher Anker: 9 Mio. vs. 14.83 Mio.). Darüber hinaus streuen die Schätzungen der verwendeten Vergleichsobjekte relativ stark. Die Schätzungen der Einwohnerzahl von Köln reichten bei den Probanden, die einen Vergleich durchgeführt haben, von 0.5 Mio. bis 10 Mio. Bei der Einwohnerzahl von New York variierten die entsprechenden Schätzungen zwischen 2 Mio. und 100 Mio.

Zur Überprüfung des Einflusses der Variation bei den Schätzungen der Vergleichsobjekte auf die Schätzung des Zielobjektes ist wiederum eine Partialkorrelation berechnet worden. Nach Herausparsialisierung des Faktors Anker ergibt sich ein hochsignifikanter Zusammenhang zwischen den Schätzungen der verwendeten Vergleichsobjekte und des Zielobjektes ($r = 0.91$; $p < .0001$). Zusammenfassend lässt sich also festhalten, dass die Manipulationsbedingung bezüglich des Zielobjektes Santiago misslungen ist. Die Daten lassen sich somit im Hinblick auf die zentrale Fragestellung des Experimentes nicht interpretieren.

Schätzungen Die Verteilungen der Schätzwerte weisen, wie bei Daten mit natürlichem Nullpunkt und einer zentralen Tendenz, die nahe am Nullpunkt liegt, zu erwarten ist, eine starke Asymmetrie auf. Um die Daten zu normalisieren und die Varianzhomogenität zu erhöhen sind die Rohdaten zunächst einer Wurzeltransformation (vgl. Winer, 1971) unterzogen, und die so transformierten Daten als abhängige Variablen für die parametrischen Verfahren verwendet worden.

Zunächst ist eine zweifaktorielle Varianzanalyse mit den beiden Faktoren *Anker* und *Vergleich* bezüglich der transformierten Guppy-Schätzungen durchgeführt worden. Es ergibt sich ein signifikanter Effekt des Hauptfaktors Anker ($F(1, 115) = 21.07$; $p < .001$; $r = 0.39$).²¹ Die Probanden haben unter der niedrigen Ankerbedingung geringere Schätzungen abgegeben als unter der hohen Ankerbedingung (1.51 vs. 2.02, transformierte Daten). Die Hypothese H.1.1 ist somit bestätigt worden. Der Hauptfaktor *Vergleich* erweist

²¹Die Effektstärken werden in Form von Korrelationskoeffizienten angegeben. Nach Cohen (1988) können r s von .5, .3 und .1 als Indikatoren für starke, mittlere bzw. schwache Effekte angesehen werden.

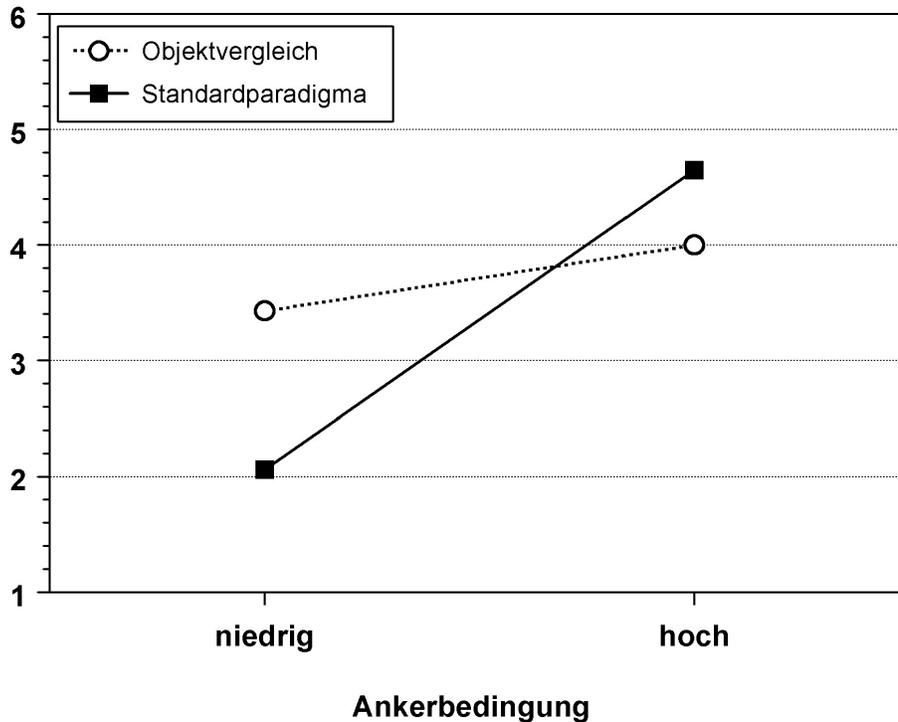


Abbildung 5: Mittlere Guppy-Schätzungen in Zentimeter, getrennt nach Versuchsbedingung.

sich als nicht signifikant ($F(1, 115) = 0.38; p > .50$). Die mittleren Schätzungen unter der Bedingung Objektvergleich liegen auf einem vergleichbaren Niveau, wie im Standardparadigma (1.79 vs. 1.73, transformierte Daten).

Darüber hinaus ergibt sich eine schwach signifikante Interaktion zwischen den Faktoren *Anker* und *Vergleich* ($F(1, 115) = 3.09; p = .082$). Die Ankereffekte sind im Standardparadigma stärker, als unter der Bedingung Objektvergleich (siehe Abb. 5). Zur Überprüfung dieses Interaktionseffektes sind getrennt nach Vergleichsbedingung geplante Kontraste berechnet worden. Im Standardparadigma zeigt sich ein sehr signifikanter Effekt des Faktors *Anker* ($t(50) = 5.36; p < .001; r = 0.58$). Unter der Bedingung Objektvergleich fällt dieser Effekt deutlich schwächer aus ($t(42.28) = 1.76; p = .086; r = 0.23$).²² Zur weiteren Überprüfung der Hypothese H.1.2 sind die Effektstärken in z-Werte transformiert worden, um

²² Aufgrund nicht vorliegender Varianzhomogenität werden bei dem t-Test die entsprechend korrigierten Statistiken berichtet.

Tabelle 3: Mittlere Schätzungen der Einwohner von Santiago

	Objektvergleich	Standardparadigma
niedriger Anker	2.154 (5.09)	1.983 (4.72)
hoher Anker	3.527 (18.57)	2.813 (8.24)

Anmerkung: Mittelwerte der transformierten Schätzungen getrennt nach Versuchsbedingung, in Klammern mittlere Schätzungen der Rohdaten, $n = 27$ bis 30 pro Zelle.

den Unterschied auf Signifikanz zu überprüfen.²³ Der entsprechende z-Test weist den Unterschied in den Effektstärken unter den Bedingungen Objektvergleich vs. Standardparadigma als signifikant aus ($z = 2.31$; $p_{\text{eins.}} = .01$). Dies bestätigt die Hypothese H.1.2.

Eine zweifaktorielle Varianzanalyse bezüglich der transformierten Santiago-Schätzungen führt ebenfalls zu einem signifikanten Effekt des Faktors *Anker* ($F(1, 110) = 17.58$; $p < .001$; $r = 0.37$). Unter der niedrigen Ankerbedingung ist die Einwohnerzahl von Santiago in Mio. Einwohner geringer eingeschätzt worden, als unter der hohen Ankerbedingung (2.07 vs. 3.16, transformierte Daten). Darüber hinaus ergibt sich ein schwach signifikanter Effekt des Hauptfaktors *Vergleich* ($F(1, 110) = 2.83$; $p = .095$; $r = 0.16$). Unter der Bedingung Objektvergleich fallen die Schätzungen höher aus, als im Standardparadigma (2.81 vs. 2.41, transformierte Daten). Die Interaktion zwischen den beiden Hauptfaktoren sind nicht signifikant ($F(1, 110) = 1.07$; $p = .30$). Die letzten beiden Befunde lassen sich vor dem Hintergrund der Überprüfung der Manipulationsbedingung damit erklären, dass unter der Bedingung Objektvergleich die Schätzungen des Zielobjektes nahezu linear den Annahmen über die Größe der Vergleichsobjekte folgten. Da diese insbesondere unter der hohen Ankerbedingungen stark variierten, finden sich analoge Variationen auch in der Schätzung des Zielobjektes wieder, was zu dem entsprechenden Datenmuster führt.

Die Schätzungen der Einwohnerzahl von Santiago sind, getrennt nach Versuchsbedingung, in Tabelle 3 dargestellt.

²³Für Details dieser Prozedur siehe etwa McNemar (1962), S. 139 f.

Diskussion

Die beiden Hypothesen H.1.1 und H.2.2 sind in dem vorliegenden Experiment anhand zweier verschiedener Zielobjekte überprüft worden. Bei dem zweiten Zielobjekt, der Einwohnerzahl von Santiago, ist die Manipulationsbedingung aus den angesprochenen Gründen, die unabhängig von dem theoretischen Ansatz der Untersuchung sind, misslungen. Die Daten lassen sich somit im Hinblick auf die zentrale Fragestellung nicht interpretieren. Daher beschränkt sich die Diskussion auf die Ergebnisse bei dem ersten Zielobjekt.

Diesbezüglich sind beide Hypothesen bestätigt worden. Von entscheidender Bedeutung für die hier diskutierte Fragestellung ist die Bestätigung der Hypothese H.1.2, wonach unter der Bedingung Objektvergleich signifikant schwächere Ankereffekte aufgetreten sind als im Standardparadigma.

Die Interpretation dieses Befundes ist die Folgende: Nach dem SAM kommt der Präsentation eines numerischen Wertes bei der Grobschätzungsaufgabe keine *unmittelbare* Bedeutung für das Auftreten von Ankereffekten im Standardparadigma zu (Mussweiler & Strack, 2001b). Der Effekt basiert nach diesem Modell ausschließlich darauf, dass bei hinreichender Ähnlichkeit zwischen dem Urteilsobjekt und dem Vergleichsstandard auf der relevanten Urteilsdimension (vgl. Mussweiler, 2003) beim Vergleichsurteil selektiv Information über das Urteilsobjekt generiert werden, die für eine Ausprägung im Bereich des Vergleichsstandards sprechen. Nicht der Vergleichsstandard an sich, sondern ausschließlich das in Bezug auf diesen Standard generierte Wissen *über das Urteilsobjekt* beeinflusst über seine erhöhten Zugänglichkeit das anschließende Feinurteil.

Diese Position impliziert, dass unter den beiden in Experiment 1 verwendeten Paradigmen Ankereffekte vergleichbarer Intensität auftreten sollten. Die Suche nach Informationen über den Guppy, die dafür spricht, dass er etwa 9 Zentimeter lang ist, oder dass er so lang ist wie ein Vergleichsobjekt, von dem bekannt ist, dass es etwa 9 Zentimeter lang ist, müsste demnach die Aktivierung derselben Wissenseinheiten evozieren.

Die vorliegenden Daten widersprechen dieser Implikation. Sie liefern somit Evidenz dafür, dass beim Standardparadigma im Gegensatz zum Objektvergleich neben den oben beschriebenen Prozessen noch weitere beteiligt sind. Der Vergleich der beiden experimentellen Paradigmen legt aus theoretischer Sicht nahe, dass diese durch die Präsentation eines numerischen Wertes im Standardparadigma ausgelöst werden, und damit, dass numerisches Priming am Zustandekommen des Ankereffektes im Standardparadigma beteiligt ist.

Denkbare Einwände gegen diese Interpretation könnten darauf rekur-

rieren, dass die höhere Ambiguität des Vergleichstandards beim Objektvergleich der Auslöser für die Verwendung einer positiven Teststrategie ist, bei der weniger oder qualitativ andere Informationen bezüglich des Zielobjektes generiert werden als im Standardparadigma. Die Basis für die Ankereffekte wären demnach in beiden Paradigmen die Prozesse des SAM, die qualitativen Unterschiede würden aus den Unterschieden beim Ablauf der postulierten Prozesse resultieren. Obwohl dies sicherlich nicht grundsätzlich auszuschließen ist, sprechen verschiedene theoretische Überlegungen in Verbindung mit einer Reihe von vor diesem Hintergrund in der Untersuchung erhobenen Befunde gegen diese Möglichkeit:

Eine Vertauschung der Zuordnung von Zielobjekt und Vergleichsstandard unter der Bedingung Objektvergleich ist aufgrund der starken Vertrautheit der verwendeten Vergleichsobjekte, die auch über die Voruntersuchung abgesichert wurde, auszuschließen. Die bereits angesprochenen Möglichkeit, dass sich hypothesenkonforme Informationen unter der Bedingung Objektvergleich im Gegensatz zum Standardparadigma sowohl auf das Zielobjekt wie auch auf das Vergleichsobjekt beziehen können, erscheint aus diesem Grund ebenfalls wenig plausibel. Darüber hinaus weisen die erhobenen Schätzungen der Vergleichsobjekte keine Unterschiede auf in Abhängigkeit davon, ob mit dem Objekt ein Vergleich durchgeführt worden ist oder nicht. Dieser Befund spricht ebenfalls dagegen, dass während des Vergleichs hypothesenkonforme Argumente bezüglich des Vergleichsobjektes generiert wurden, da sich dies nach dem SAM in verzerrten Schätzungen der Vergleichsobjekte niederschlagen müsste. Ein weiterer Einwand im Kontext der Ambiguität des Vergleichsobjektes könnte darauf verweisen, dass in Abhängigkeit von variierenden Vorstellungen über die Größe des Vergleichsobjektes, die Urteiler unter ein und derselben Ankerbedingung auf verschiedene Größenbereiche verankert werden, und dies einen entsprechenden Einfluss auf die Feinschätzungen hat. Die Daten bei dem Zielitem „Einwohner von Santiago“ stellen einen Beleg für diese Möglichkeit dar. Bezogen auf die hier diskutierten Befunde muss diese Möglichkeit jedoch ebenfalls zurückgewiesen werden. Zum einen weisen die verwendeten Vergleichsobjekte nur sehr geringe Streuungen auf. Zum anderen zeigen die berechneten Partialkorrelationen, dass innerhalb der Ankerbedingungen keinerlei Zusammenhang zwischen den Schätzungen von Ziel- und Vergleichsobjekt besteht.

Das zusammenfassende Resümee des vorliegenden Experimentes lautet also, dass Ankereffekte auch im Standardparadigma zumindest anteilig auf numerischem Priming basieren.

Ankereffekte durch subliminales Priming

Gegenstand des zweiten Experimentes dieser Arbeit ist die Frage, ob es für das Auftreten von Ankereffekten im Basic-Anchoring-Paradigma notwendig ist, dass dem dargebotenen numerischen Wert genügend Aufmerksamkeit entgegengebracht wird, wie Wilson et al. (1996) postulieren, oder ob ein dargebotener Wert auch zum Anker für eine nachfolgende Schätzung werden kann, *ohne* dass diesem Wert seitens des Urteilers Aufmerksamkeit zuteil wird. Überprüft werden soll die letztere Möglichkeit mittels eines experimentellen Paradigmas, bei dem die als Anker intendierten Werte in einer vorgeblich unabhängigen Voraufgabe *subliminal*, also unterhalb der Schwelle zur bewussten Wahrnehmung präsentiert werden.

Obwohl der mögliche Einfluss subliminaler Reize über Jahrzehnte hinweg kontrovers diskutiert wurde (für einen Überblick siehe Loftus & Klingger, 1992), existiert mittlerweile eine umfangreiche Anzahl von Befunden, die einen derartigen Einfluss auf das Verhalten belegen. Eine Reihe von Experimenten (vgl. etwa Foster & Davis, 1984; Marcel, 1983) zeigt die üblichen Bahnungs- und Inhibitionseffekte, wenn vor der Darbietung des Zielreizes ein reizkonsistenter bzw. -inkonsistenter Stimulus subliminal dargeboten wird, was zunächst auf eine semantische Verarbeitung der subliminal dargebotenen Stimuli hinweist.

Bedeutsam im Kontext der vorliegenden Fragestellung sind Befunde, in denen ein Einfluss subliminal dargebotener Stimuli auf die Urteilsbildung nachgewiesen wird. Bargh und Pietromonaco (1982) sowie Devine (1989) zeigen beispielsweise, dass mittels subliminaler Darbietungen von Begriffen Kategorien aktiviert werden können, was sich auf eine anschließende Bewertung ambivalenter Personenbeschreibungen in Richtung der aktivierten Kategorie auswirkt. Gabrielcik und Fazio (1984) weisen als Beleg für die Verfügbarkeitsheuristik (vgl. hierzu S. 8) nach, dass sich die Schätzung von Worthäufigkeiten in hypothesenkonformer Weise verändert, wenn vorher eine Reihe von Exemplaren der entsprechenden Wortklasse subliminal dargeboten wurde.

Eine wichtige Frage in Bezug auf das geplante Vorhaben ist, ob auch die subliminale Darbietung von Zahlen zu einer Aktivierung entsprechender numerischer Konzepte führt. Evidenz hierfür folgt aus einer Reihe neuerer Arbeiten von Dehaene und Kollegen (Dehaene et al., 1998; Greenwald, Abrams, Naccache & Dehaene, 2003; Naccache & Dehaene, 2001). Der Aufbau dieser Experimente folgt denen zum Nachweis von Bahnungs- und Inhibitionseffekten: Die Probanden sollen möglichst schnell entscheiden, ob eine dargebotene Zahl kleiner oder größer als ein vorgegebener Vergleichswert ist.

Kurz vor der Präsentation dieses Zielreizes wird zusätzlich subliminal eine Zahl dargeboten. Es zeigen sich die üblichen Primingeffekte in Abhängigkeit von der Relation des subliminal dargebotenen Wertes zum Vergleichsstandard: Wird ein Urteil darüber erfragt, ob die als Zielreiz dargebotene Ziffer kleiner oder größer als 5 ist, erfolgt die korrekte Klassifizierung signifikant schneller, wenn die subliminal dargebotene Zahl dieselbe Größenrelation zur 5 aufweist, wie der Zielreiz (Dehaene et al., 1998). Von besonderer Bedeutung hierbei ist, dass diese Effekte auch dann auftreten, wenn die subliminal dargebotenen Zahlen nicht zu der Gruppe der verwendeten Zielreize gehören (Naccache & Dehaene, 2001). Dies spricht gegen einen Effekt auf der Basis von reinem Reiz-Reaktions-Lernen, und belegt die konzeptuelle Verarbeitung der subliminal dargebotenen Werte.

Ob der Einfluss dieser Verarbeitung subliminal dargebotener Zahlen auf Primingeffekte der oben dargestellten Weise beschränkt ist, oder sich auch bei weit komplexeren Urteilen wie im Kontext des Ankereffektes zeigt, soll nun überprüft werden.

A priori wurde hierbei festgelegt, nur einstellige Zahlen als subliminal präsentierte Werte zu verwenden. Die Problematik bei mehrstelligen Zahlen besteht darin, dass bisher wenig Erkenntnisse darüber existieren, wie diese bei einer subliminalen Darbietung enkodiert werden, respektive welche numerischen Konzepte durch derartige Darbietungen aktiviert werden. Wird beispielsweise die Zahl 769 mehrfach subliminal dargeboten, könnte dies zu einer Aktivierung des Zahl 769 führen, alternativ wäre auch eine Aktivierung der Zahlen 76 oder 69, gegebenenfalls auch der einzelnen Ziffern 7, 6 und 9 denkbar. Wie Greenwald et al. (2003) in einer Reihe von Experimenten zeigen, kann sogar über die Kontextbedingungen manipuliert werden, ob Probanden bei *subliminalen* Darbietungen zweistelliger Zahlen auf die erste oder die zweite Ziffer fokussieren. Aus diesem Grund erfolgt die Einschränkung auf den einstelligen Zahlbereich.

Vorstudie 2

Vor jedem Experiment, in dem mit subliminal dargebotenen Stimuli gearbeitet wird, muss als wichtige Vorentscheidung die Darbietungsdauer des Bildmaterials festgelegt werden, die gewährleistet, dass das Material nicht bewusst wahrgenommen wird. Die Auswahl einer geeigneten Darbietungsdauer ist Gegenstand dieser Vorstudie.

Eine Recherche in der existierenden Literatur zu subliminalen Darbietungen zeigt, dass über die verwendeten experimentellen Paradigmen zur Bestimmung der Schwelle zwischen bewusster und nicht-bewusster Wahr-

nehmung wenig einheitliche Vorstellungen existieren. Sie reichen von einfachen Recall-Tests, in denen die Probanden beispielsweise die kurzzeitig dargebotenen Worte benennen sollen (Bargh & Pietromonaco, 1982; Devine, 1989) über Recognition-Tests, in denen Listen von Reizen vorgelegt werden, mit der Bitte, die in der Experimentalphase als subliminal intendiert dargebotenen Reize zu markieren (Devine, 1989), bis hin zu Vorexperimenten, in denen mit der Methode der Signalentdeckungstheorie (vgl. etwa Velden, 1982) gearbeitet wird (Dehaene et al., 1998).

Cheesman und Merikle (1984, 1986; Merikle & Cheesman, 1986) schlagen in diesem Kontext eine Unterscheidung zwischen *objektiver* und *subjektiver* Wahrnehmungsschwelle vor. Als unterhalb der objektiven Wahrnehmungsschwelle wird ein Niveau der Darbietungsintensität bezeichnet, wenn „forced-choice“-Befragungen wie beispielsweise Signalentdeckungsexperimente indizieren, dass Probanden etwa nicht zwischen Durchgängen mit und ohne Reizdarbietung diskriminieren können. Der Begriff der subjektiven Wahrnehmungsschwelle bezieht sich auf eine höhere Reizintensität, nämlich ab der Probanden in Befragungen von einer bewussten Wahrnehmung des Reizes berichten.

In neueren Arbeiten (vgl. Dehaene et al., 1998; Naccache & Dehaene, 2001) werden Bahnungs- und Inhibitionseffekte nachgewiesen in einem Paradigma, dem die objektive Wahrnehmungsschwelle zugrunde liegt. Nach Greenwald (1992) hingegen zeigen sich stabile Befunde bezüglich des Einflusses subliminaler Darbietungen nur dann, wenn die Reize oberhalb der objektiven aber unterhalb der subjektiven Wahrnehmungsschwelle dargeboten werden. Merikle (1992) kritisiert die Verwendung der objektiven Wahrnehmungsschwelle als Kriterium für die Subliminalität dargebotener Reize auch aus konzeptioneller Sicht. Diese Operationalisierung setze nämlich voraus, dass eine korrekte Klassifizierung, beispielsweise bei Signalentdeckungsexperimenten, ausschließlich von bewusster Informationsverarbeitung geleitet wird. Unter der Annahme, dass auch nicht-bewusste Prozesse zur korrekten Klassifizierung beitragen, führt die Verwendung der objektiven Wahrnehmungsschwelle dazu, den Nachweis nicht-bewusster Prozesse zu führen in einem experimentellen Paradigma, welches dieselben a priori zumindest anteilig unterbindet.

Vor dem Hintergrund, dass die Unterscheidung zwischen bewussten und nicht-bewussten Prozessen entscheidend an Relevanz gewinnt, wenn bewusste und nicht-bewusste Prozesse zu qualitativ unterschiedlichen Konsequenzen führen, anstatt dass nicht-bewusste Wahrnehmung lediglich eine „schwächere“ Form von bewusster Wahrnehmung ist (vgl. Dixon, 1971; Merikle, 1992; Shevrin & Dickman, 1980), schlagen Merikle und Kollegen

(Merikle, 1992; Reingold & Merikle, 1990; Merikle & Daneman, 1998) folgende Methodik vor: A priori werden die Bereiche bewusster und nicht-bewusster Wahrnehmung auf der Basis introspektiver Berichte der Probanden bestimmt (subjektive Schwelle). Ergänzend hierzu werden die Befunde, dass Reizdarbietungen oberhalb und unterhalb dieser Schwelle zu qualitativ unterschiedlichen Konsequenzen führen, als zusätzliche Validierung für diese Klassifizierung heran gezogen.

Eine derartige Vorgehensweise setzt allerdings voraus, dass unterschiedliche Konsequenzen in Abhängigkeit davon, ob Reize nicht-bewusst oder bewusst wahrgenommen werden, Gegenstand der entsprechenden Untersuchung sind. In der vorliegenden Situation ist dies nicht der Fall. Vor der Alternative stehend, welche der beiden Wahrnehmungsschwellen als Kriterium für Subliminalität im Kontext dieser Untersuchung herangezogen werden soll, wurde entschieden, das strengere Kriterium der objektiven Wahrnehmungsschwelle zu verwenden, und in der Vorstudie zur Ermittlung der Darbietungsdauer mit der Methodik der Signalentdeckungstheorie in Anlehnung an Dehaene et al. (1998) zu arbeiten.

Method

Vorgehen Die Untersuchung wurde im Einzelversuch mit Hilfe eines Personal-Computers durchgeführt. Jeder Proband hatte zwei Sitzungen an verschiedenen Tagen zu absolvieren. Beide Sitzungen hatten denselben Aufbau. Zunächst wurden die Probanden mit dem genauen Ablauf des Experimentes vertraut gemacht. Ihnen wurde gesagt das es sich um eine Voruntersuchung zur Ermittlung von Darbietungsdauern für ein Experiment mit subliminalen, d. h. „unterschwellig“ Reizen handele. Ihre Aufgabe würde darin bestehen zu entscheiden, ob innerhalb eines Darbietungsdurchgangs ein bestimmter Reiz enthalten war, oder nicht. Jeder Durchgang begann damit, dass zunächst ein Kreuz auf der Mitte des Bildschirms präsentiert wurde, welches die Probanden fixieren sollten. Im Anschluß daran wurde nach etwa zwei bis sieben Sekunden (zufallsabhängig, durch das Programm gesteuert) entweder rechts oben, rechts unten, links oben oder links unten, ausgehend vom Fixationskreuz ein Reiz dargeboten. Die Aufgabe der Probanden bestand darin, zunächst zu entscheiden, ob der Reiz rechts oder links vom Fixationskreuz dargeboten wurde. Hiefür sollten sie die beiden oberen, mit „rechts“ und „links“ beschrifteten Tasten einer separaten, mit vier Tasten versehenen Tastatur benutzen. Im Anschluß daran sollten die Probanden entscheiden, ob lediglich ein als „Maske“ bezeichnete Reiz dargeboten wurde, oder ob vor der Maske zusätzlich eine der vier Ziffern 1,

4, 7 oder 9 präsentiert worden war. Hierfür sollten die Probanden die mit „ja“ beschriftete untere linke Taste drücken, wenn sie meinten, eine Ziffer gesehen zu haben, ansonsten die mit „nein“ beschriftete untere rechte Taste. Sowohl die vier Ziffern als auch die Maske wurden den Probanden vorher auf dem Bildschirm präsentiert. Den Probanden wurde des weiteren mitgeteilt, dass die Darbietungszeiten bei den Ziffern variieren würden, so dass in manchen Durchgängen die Ziffern gut zu sehen seien, in anderen Durchgängen weniger gut.

Nach dem Erhalt der Instruktionen, wurden die Probanden gebeten, vor dem Bildschirm Platz zu nehmen. Dann wurde zunächst der Abstand von Auge zu Bildschirm vermessen und die Sitzposition gegebenenfalls so korrigiert, dass der Abstand während der Versuchsdurchführung bei allen Probanden 63 cm betrug. Im Anschluß daran wurde von dem Versuchsleiter ein Probedurchlauf gestartet, der dazu diente, die Probanden mit der Apparatur vertraut zu machen. Der Probedurchlauf enthielt 16 Darbietungen, hiervon acht mit und acht ohne Präsentation einer Ziffer. Sowohl jede der vier Ziffern als auch jede der vier Darbietungsdauern wurde zweimal verwendet, die Abfolge der Durchgänge war zufällig und wurde durch das Programm bestimmt. Die Daten des Probedurchlaufs wurden nicht gespeichert.

Nachdem die Probanden den Probedurchgang absolviert hatten und bestätigten, den Versuchsablauf verstanden zu haben, wurden in vier Blöcken die Daten erhoben. Jeder Block bestand aus 16 Darbietungen, wiederum acht mit und acht ohne Präsentation einer Ziffer. Jede Ziffer wurde zweimal dargeboten. Die Darbietungsdauer der Ziffern blieb während eines Blockes konstant. Nach jedem Block war eine Pause, bevor der Versuchsleiter nach ca. anderthalb Minuten den nächsten Block startete. Sowohl die Abfolge der einzelnen Darbietungen als auch die Abfolge der vier Blöcke, denen jeweils eine der vier Darbietungsdauern zugeordnet war, wurde zufällig durch das Programm bestimmt. Auf diese Weise wurden in den zwei Sitzungen von jedem Proband insgesamt 128 Messungen erhoben.

Apparatur und Stimuli Die Untersuchung wurde mittels eines Personal-Computers durchgeführt. Der Versuchsablauf wurde mithilfe der Computersoftware Experimental Run Time System (ERTS, Beringer, 1999) realisiert. An dem Computer war die für das Programm ERTS entwickelte Tastatur EXKEY mit vier, entsprechend dem Versuchsablauf beschrifteten Tasten angeschlossen, über die die Probanden ihre Antworten eingaben.

Die Reize (siehe Anhang, S. 152) wurden auf einem 15"-Monitor mit einer Bildwiederholungsrate von 59 Hz dargeboten. Sämtliche Reize wurden in

schwarz vor einem hellgrauen Hintergrund präsentiert. Die dargebotenen Ziffern hatten eine Höhe von 10 mm, die Breite variierte zwischen 4 und 9 mm. Die verwendete Maskierung hatte eine Höhe von 12 mm und eine Breite von 11 mm. Das Fixationskreuz hatte eine Höhe und Breite von jeweils 8 mm. Die Ziffern und die Maskierung wurden an den erwähnten Positionen im Abstand von 48 mm ausgehend vom Fixationskreuz dargeboten. In Verbindung mit einem „Auge-Bildschirm“-Abstand von 63 cm war hierdurch gewährleistet, dass die Zielreize innerhalb des parafovealen Sichtfeldes dargeboten wurden (vgl. Bargh & Pietromonaco, 1982). Die vier Darbietungsdauern, mit denen in dieser Untersuchung gearbeitet wurde, waren 17, 33, 50 und 67 Millisekunden. Sie entsprechen bei einem 59 Hz- Bildschirm einer Darbietung innerhalb von einer, zwei, drei oder vier Bildwiederholungsraten. Die Maske folgte unmittelbar auf die Präsentation der Ziffer. Ihre Darbietungsdauer betrug 253 Millisekunden.

Probanden An der Voruntersuchung nahmen 10 Studenten der Universität Kassel (5 männl., 5 weibl.) als Probanden teil. Das Durchschnittsalter der Teilnehmer betrug 26 Jahre. Alle Versuchsteilnehmer berichteten, dass sie ggf. mit Brille normalsichtig seien. Die Probanden wurden auf dem Universitätsgelände angeworben und erhielten für ihre Teilnahme keine Entlohnung.

Ergebnisse und Diskussion

Die Ergebnisse einer ersten Analyse, getrennt nach prozentualer „Hit-Rate“, „False Alarms“ und dem daraus resultierenden Diskriminanzmaß d' , sind in Tabelle 4 dargestellt. Wie erwartet steigt die Diskriminationsfähigkeit mit zunehmender Darbietungsdauer. Auffällig ist jedoch, dass schon bei der kürzesten Darbietungsdauer von 17 Millisekunden, obwohl die „Hit-Rate“ bei lediglich 29.4 % liegt, ein signifikanter Unterschied zwischen den Antworten bei Anwesenheit und Abwesenheit der Zielreize vorliegt ($\chi^2 = 7.03$; $p = .008$).

Eine weitergehende Analyse der Daten ergibt, dass die korrekte Klassifizierung der Reize stark von den jeweils dargebotenen Ziffern abhängt. Auf der Basis einer anhand der deskriptiven Datenstruktur (siehe Anhang, Tabelle 10, S. 152) naheliegenden Unterteilung der Reize in die Gruppen {1, 9} bzw. {4, 7} sind hierzu weitere χ^2 -Tests berechnet worden, die sich lediglich auf die Antworten („ja“ vs. „nein“) bei Anwesenheit der Reize beziehen.

Es zeigt sich, dass die Ziffern „1“ bzw. „9“ signifikant schlechter korrekt klassifiziert werden als die Ziffern „4“ und „7“. Aggregiert über alle Dar-

Tabelle 4: Ergebnisse von Vorexperiment 2

	Darbietungsdauer in msek.			
	17	33	50	67
Hit rate (%)	29.4	54.4	68.8	88.8
False alarms (%)	16.9	16.9	18.1	8.1
d'	0.42*	1.07**	1.40**	2.61**

Anmerkung: n=160 pro Zelle; *: $p < 0.05$; **: $p < 0.001$ (χ^2 -Test)

bietungsdauern sind erstere in 50.9 %, letztere in 69.7 % der Fälle erkannt worden ($\chi^2 = 23.5$; $p < .001$). Dieser Trend wird auch in separaten Analysen für jede einzelne Darbietungsdauer bestätigt ($\chi^2 = 5.09$; 18.37 ; 5.70 bzw. 2.25 bei 17, 33, 50, bzw. 67 msek.; $p = .024$; $< .001$; $= .017$, bzw. $= .133$ resp.).²⁴

Als Erklärung für diesen Unterschied erscheint es naheliegend anzunehmen, dass aufgrund der spezifischen Struktur der Maske die Ziffern „1“ und „9“ effektiver maskiert wurden, als die Ziffern „4“ und „7“.

Auf der Basis dieser Ergebnisse sind noch einmal separate Analysen für die beiden Gruppen von Reizen im Signalentdeckungs-Paradigma berechnet worden. Die Ergebnisse sind in Tabelle 5 dargestellt. Als Referenzgröße der „False Alarms“ für die Berechnung der Diskriminanzmaße wurde für beide Gruppen die entsprechenden Häufigkeiten aus den Gesamtdaten verwendet, die aus diesem Grund auch in Tabelle 5 noch einmal aufgeführt sind.

Bei einer Darbietungsdauer von 17 Millisekunden sind die Ziffern „1“ und „9“ in 21.3 % der Fälle von den Probanden korrekt klassifiziert worden. Die Quote der „False-Alarms“ liegt innerhalb der entsprechenden Durchgänge bei 16.9 %. Ein χ^2 -Test, bezogen auf diese Darbietungsdauer, führt zu einem Wert von $\chi^2 = 0.682$, $p = .41$.

Es kann demnach mit hinreichender Sicherheit davon ausgegangen werden, dass bei einer Darbietungsdauer von 17 Millisekunden die Probanden über keine bewusste Wahrnehmung der Ziffern „1“ und „9“ in dem hier

²⁴Der geringere Unterschied bei der längsten Darbietungsdauer basiert auf einem Deckeneffekt, da hier fast alle Reize korrekt klassifiziert wurde (vgl. Anhang, Tabelle 10, S. 152).

Tabelle 5: Ergebnisse von Vorexperiment 2 getrennt nach Reizen

Reize		Darbietungsdauer in msec			
		17	33	50	67
1 und 9	False alarms ^a (%)	16.9	16.9	18.1	8.1
	Hits ^b (%)	21.3	37.5	60.0	85.5
	d'	0.16	0.64**	1.17**	2.43**
4 und 7	Hits ^b (%)	37.5	71.3	77.5	92.5
	d'	0.64**	1.52**	1.67**	2.84**

Anmerkung: ^a: n=160 pro Zelle; ^b : n=80 pro Zelle, **: $p < 0.001$ (χ^2 -Test)

behandelten experimentellen Paradigma verfügen.²⁵ Auf der Basis dieser Ergebnisse ist entschieden worden, in dem nachfolgend geschilderten Experiment zur Wirkung subliminaler Anker mit einer Darbietungsdauer von 17 Millisekunden zu arbeiten und die Werte „1“ und „9“ als niedrigen bzw. hohen Anker zu verwenden.

Experiment 2

Mit den im Vorexperiment 2 ermittelten Darbietungsdauern und Reizen soll nun überprüft werden, ob subliminal dargebotene numerische Werte als Anker für nachfolgende Schätzaufgaben fungieren.

Das Experiment besteht aus zwei Phasen. In der ersten Phase wird der subliminale Anker im Kontext einer Reaktionsaufgabe etabliert (vgl. Bargh & Pietromonaco, 1982; Devine, 1989). Die relevanten Schätzaufgaben werden in der zweiten Phase innerhalb eines sogenannten Allgemeinbildungstests dargeboten. Als Zielitems werden wiederum Schätzungen der Länge eines Guppys in Zentimeter und der Einwohnerzahl von Santiago in Mio. Einwohner erfragt.

Um diese Versuchsanordnung für die Probanden möglichst plausibel erscheinen zu lassen, werden sie dahingehend informiert, dass mit dem Experiment mögliche Interaktionseffekte untersucht werden sollen, die bei der auf-

²⁵Zur Verwendung von Signifikanztest zur Bestätigung der H_0 vgl. etwa Bortz, 1993, S. 155.

einanderfolgenden Bearbeitung unterschiedlicher Aufgaben auftreten. Des weiteren wird in diesem Kontext erläutert, dass die beiden zu bearbeitenden Aufgaben auf einem Pool von acht verschiedenen Aufgaben stammen, und von der Versuchsperson mit Hilfe eines Zufallsgenerators selbst ausgewählt werden. Neben dem Aspekt der Plausibilität der Versuchsanordnung bietet dieses Szenario die Möglichkeit, sich innerhalb einer Nachbefragung in für die Teilnehmer nachvollziehbarer Weise nach Einschätzungen über Zusammenhänge zwischen den beiden Aufgaben, Einflüsse der ersten auf die zweite Aufgabe und Auffälligkeiten bei der Bearbeitung der Aufgaben zu erkundigen.

Methode

Vorgehen Das Experiment wurde im Einzelversuch mit Hilfe eines Computers durchgeführt. Den Probanden wurde in den Instruktionen mitgeteilt, dass es Gegenstand des Experimentes sei, Interaktionseffekte zu untersuchen, die bei der aufeinanderfolgenden Bearbeitung verschiedener Aufgaben, zwischen denen kein inhaltlicher Zusammenhang besteht, auftreten.

Mit der Begründung, diese Effekte systematisch untersuchen zu wollen, sollte zunächst jeder Proband mittels eines Zufallsgenerators aus einem Pool von acht verschiedenen Aufgaben durch Drücken einer Taste zwei Aufgaben auswählen, die dann in der ausgewählten Reihenfolge bearbeitet werden mussten. Ohne Wissen der Probanden fiel die „Wahl“, unabhängig vom Tastendruck der Probanden, stets auf dieselben zwei Aufgaben. Die erste Aufgabe wurde als Reaktionstest vorgestellt, die zweite Aufgabe als kurzer Test zur Allgemeinbildung.

Bei dem Reaktionstest bestand die Aufgabe darin, möglichst schnell die Position eines kurzzeitig dargebotenen Reizes zu erkennen. Zunächst wurde in der Mitte des Bildschirms ein Fixationskreuz dargeboten, im Anschluß daran wurde nach zwei bis sieben Sekunden (zufallsabhängig, durch das Programm gesteuert) entweder links unten, links oben rechts unten oder rechts oben ein den Probanden in der Einleitung des Experimentes bereits vorgestellter Reiz präsentiert.

Die Probanden sollten lediglich entscheiden, ob der Reiz links oder rechts vom Fixationskreuz dargeboten wurde. Zur Beantwortung sollten die linke und rechte Shift-Taste der Tastatur verwendet werden, die hierfür mit einem „L“ bzw. einem „R“ markiert waren. Die Aufgabe bestand aus 96 Durchgängen. Den Probanden wurde aus motivationalen Gründen mitgeteilt, dass es sich um 80 Durchgänge handeln würde, von denen die ersten zehn Probedurchgänge seien, die dazu dienen, sich mit der Apparatur ver-

traut zu machen, und nicht in die Auswertung eingingen. Des weiteren wurden die Probanden darauf hingewiesen, zu Beginn eines jeden Durchgangs das Kreuz in der Mitte des Bildschirms zu fixieren und beide Zeigefinger auf die Antwort-Tasten zu legen, da dies die besten Voraussetzungen für eine möglichst schnelle und korrekte Klassifizierung liefere.

Im Rahmen dieser Aufgabe wurde der subliminale Anker etabliert. Der vorgebliche Zielreiz diente als Maskierung für die Darbietung einer Ziffer, die als Anker für die nachfolgend geschilderten Schätzaufgaben fungieren sollte. Auf der Basis der Ergebnisse des Vorexperimentes 2 wurden die Werte „1“ und „9“ als niedriger bzw. hoher Anker verwendet, die Darbietungsdauer der Ziffern betrug 17 msek., Positionierung und Maskierung der Ziffern entsprach dem Aufbau des Vorexperimentes.

Im Anschluß an diese Aufgabe wurden innerhalb des sogenannten Allgemeinbildungstests acht Aufgaben zur Beantwortung vorgelegt. Hierbei handelte es sich um vier Füller-Items qualitativer Natur, die im Multiple-Choice-Format an Position 2, 3, 5 bzw. 8 dargeboten wurden, und zwei weiteren Fragen an Position 6 und 7, die Gegenstand eines anderen Experimentes waren und für das hier beschriebene Experiment ebenfalls als Füller-Items dienten.

Bezüglich des subliminalen Priming wurde mit den bereits vorgestellten Zielitems gearbeitet, die entsprechenden Fragen lauteten:

Guppy: „Wie lang ist ein Guppy (Zierfischart) in cm?“

Santiago: „Wie viele Einwohner hat die chilenische Hauptstadt Santiago in Mio. Einwohner?“

Eine Hälfte der Probanden bekam als erste Frage das Item *Guppy*, die andere Hälfte das Item *Santiago*. Das jeweils andere Item wurde als viertes Item nach zwei Füllerfragen vorgelegt, um beim Auftreten eines Primingeffektes bei der ersten Frage überprüfen zu können, ob dieser Effekt nur kurzfristiger Natur ist oder auch eine größere zeitliche Distanz zur Priming-Phase überdauert.

Sämtliche Instruktionen wurden den Probanden über am Computerbildschirm angezeigte Texte vorgegeben. Eine Versuchsleiterin war während des gesamten Experimentes anwesend, um eventuelle Fragen zu beantworten und den korrekten Ablauf des Experimentes zu überwachen. Zu Beginn des Experimentes wurde der „Auge-Bildschirm“-Abstand der Probanden von der Versuchsleiterin vermessen, und die Sitzposition gegebenenfalls so korrigiert, dass der Abstand während des Experimentes bei allen Probanden 63 cm betrug.

Am Ende des Experimentes wurde den Probanden noch ein kurzer Fragebogen zur Bearbeitung vorgelegt. In diesem Fragebogen wurden zunächst die Personendaten erhoben, anschließend waren die beiden bearbeiteten Aufgaben („Reaktionstest“ und „Allgemeinbildungstest“) hinsichtlich ihres Schwierigkeitsgrades zu beurteilen. Des Weiteren wurden die Probanden nach ihrer Einschätzung bezüglich eines Einflusses der ersten Aufgabe auf *ihre Leistungen* bei der zweiten Aufgabe befragt, bezüglich weiteren Zusammenhängen zwischen den beiden Aufgaben, sowie nach allgemeinen Besonderheiten im Experiment, die ihnen aufgefallen waren. Mit diesen Fragen sollte überprüft werden, ob die Manipulationsbedingung des subliminalen Priming angemessen realisiert wurde.

Anschließend bedankte sich die Versuchsleiterin bei den Probanden und zahlte das Versuchsgeld aus.

Apparatur und Stimuli Die Untersuchung wurde mittels eines Personal-Computers durchgeführt, der Versuchsablauf wurde mithilfe der Computersoftware Experimental Run Time System (ERTS, Beringer, 1999) realisiert. Die in der ersten Versuchsphase präsentierten Reize (siehe Anhang, S. 152) stimmten mit denen im Vorexperiment überein. Sämtliche Instruktionen und die in der zweiten Versuchsphase präsentierten Fragen sind im Anhang, S. 153 bzw. S. 157 enthalten. Die Darbietung erfolgte in schwarz vor einem hellgrauen Hintergrund auf einem 15"-Monitor mit einer Bildwiederholungsrate von 59 Hz.

Die Eingaben der Probanden erfolgte mittels einer IBM-Computer-Tastatur. In der ersten Versuchsphase benutzten die Probanden die linke und rechte Shift-Taste, die hierfür farblich markiert waren. In der zweiten Versuchsphase wurden die Probanden aufgefordert, sämtliche Eingaben mit Hilfe des Nummernblocks der Tastatur vorzunehmen, der zu diesem Zweck ebenfalls markiert war.

Die Nachbefragung erfolgte mittels eines Fragebogens, der im Anhang, S. , enthalten ist. Bei der Nachbefragung wurden zunächst als Personendaten Alter, Geschlecht und Studienfach erfragt. Der empfundene Schwierigkeitsgrad der beiden Aufgaben wurde mit Hilfe von Likert-Skalen mit sieben Stufen und verankerten Endpunkten („sehr einfach“, bzw. „sehr schwierig“) erfasst. Der Einfluss der Bearbeitung der ersten Aufgabe auf die Leistungen in der zweiten Aufgaben wurde im „forced-choice“-Format mit den drei Alternativen positiver / negativer / kein Einfluss erhoben. Die beiden letzten Fragen des Fragebogens bezogen sich auf andere Zusammenhänge bzw. auf Auffälligkeiten im Experiment, und dienten der weiteren Überprüfung der

Subliminalität der dargebotenen Reize. Die Probanden erhielten hierbei die Gelegenheit, im offenen Format ihre Anmerkungen vorzunehmen.

Probanden An dem Experiment nahmen 83 Personen (48 Männer, 35 Frauen) teil. Bis auf zwei Schüler waren sämtliche Probanden Studenten der Universität Kassel, das Durchschnittsalter betrug 25 Jahre. Alle Probanden berichteten, dass sie ggf. mit Brille normalsichtig seien. Die Probanden wurden vor der Universitätsmensa angeworben und erhielten für ihre Teilnahme an dem Experiment 10 DM.

Experimentelles Design und Hypothesen Das vollständige experimentelle Design entspricht einem 2 (*Anker*: niedrig vs. hoch) \times 2 (*Item*: Guppy vs. Santiago) \times 2 (*Position*: 1. Frage vs. 4. Frage) – Plan mit Nachbefragung. Den Bedingungen des Faktors *Anker* wurden separate Gruppen von Probanden zugeordnet, die beiden Faktoren *Item* und *Position* waren in einem „two period crossover“ – Design angeordnet (vgl. Cotton, 1989).

Die Auflistung der Hypothesen, die innerhalb des Experimentes getestet werden sollten, beginnt zunächst mit der Überprüfung der Manipulationsbedingung, d. h. des subliminalen Priming. Die erste Hypothese lautet daher:

H.2.1: Die Darbietung der Ziffern innerhalb der ersten Versuchsphase bleibt von den Probanden unbemerkt.

Diese Hypothese ist zentral für das gesamte Experiment. Ergänzend zu den Befunden des Vorexperimentes 2 wurde zur weiteren Überprüfung den Probanden über die letzten beiden Fragen der Nachbefragung Gelegenheit gegeben, zu möglichen Einflüssen der ersten Aufgabe auf die zweite Aufgabe, und anderen Auffälligkeiten innerhalb des Experimentes Stellung zu nehmen. Auf eine explizite Nachbefragung wurde verzichtet, da diese eine zu hohen Aufforderungscharakter besitzt, und damit zu verzerrten Ergebnissen führen könnte.

Unter der Voraussetzung, dass die H.2.1 erfüllt ist, bezieht sich die nächste Hypothese auf die zentrale Fragestellung des Experimentes.

H.2.2: Probanden, die in der Priming-Phase mit dem hohen Anker konfrontiert wurden, geben bei der ersten Frage höhere Schätzungen ab, als Probanden, die mit dem niedrigen Anker konfrontiert wurden.

Über einen möglichen Einfluss der Wahl des Items oder Interaktionseffekte zwischen Anker und Item wurden im Vorfeld keine Hypothesen formuliert.

Für den Fall, dass die H.2.2 bestätigt werden sollte, sollte über eine Analyse der Schätzungen bei der vierten Frage überprüft werden, ob sich der Einfluss des subliminalen Primings auch auf Urteile auswirkt, die in größerer zeitlicher Distanz zur Priming-Phase gefällt werden. Die Hypothese hierzu lautet:

H.2.3: Probanden, die in der Priming-Phase mit dem hohen Anker konfrontiert wurden, geben auch bei der vierten Frage höhere Schätzungen ab, als Probanden, die mit dem niedrigen Anker konfrontiert wurden.

Über einen möglichen Einfluss der Wahl des Items oder Interaktionseffekte zwischen Anker und Item wurden auch bei der H.2.3 im Vorfeld keine Hypothesen formuliert.

Sollte neben der Hypothese H.2.2 auch die Hypothese H.2.3 bestätigt werden, müsste im weiteren mit Hilfe entsprechender Korrelationstechniken überprüft werden, ob das subliminale Priming als Ursache für einen Anker-effekt bei der vierten Frage angesehen werden kann, oder ob die Probanden in erster Linie durch ihre eigene Schätzung bei der ersten Frage verankert wurden (siehe hierzu Becker et al., 2003, Exp. 3).

Ergebnisse

Die zentralen Hypothesen des Experimentes beziehen sich auf die Schätzungen der Probanden unter den verschiedenen experimentellen Bedingungen. Um mögliche Befunde jedoch im Sinne der Fragestellung zu interpretieren, muss zunächst gewährleistet sein, dass in der Nachbefragung keinerlei Hinweise enthalten sind, die darauf hindeuten, dass die Probanden die in der ersten Aufgabe enthaltene Priming-Prozedur als solche erkannt haben, oder gar die als subliminal intendierten Darbietungen erkennen konnten. Demzufolge beginnt die Analyse der Ergebnisse zunächst mit der Auswertung der Nachbefragung, bevor die Schätzungen der Probanden analysiert werden.

Nachbefragung Zunächst sei erwähnt, dass die Probanden die „Reaktionsaufgabe“ als deutlich einfacher eingestuft haben als der „Allgemeinbildungstest“ (Mittelwerte: 1.70 vs. 4.96, $t(82) = -18.9$; $p < .001$). Hierbei ist zu berücksichtigen, dass nach dem Schwierigkeitsgrad der Aufgaben gefragt wurde, nicht danach, wie anstrengend die Aufgaben empfunden wurden.

Ein Einfluss der ersten Aufgabe auf die Leistungen in der zweiten Aufgabe ist wie erwartet von der Mehrheit der Probanden (81.9 %) nicht gesehen worden. Bei den verbliebenen 18.1 % der Befragten überwog die Meinung,

der Einfluss auf die zweite Aufgabe sei negativ (13.3 % vs. 4.8 %). Kommentare bei den folgenden beiden Fragen lassen vermuten, dass dieser Einfluss von den entsprechenden Probanden in erster Linie auf abnehmende Konzentrationsfähigkeit und die Tendenz auch bei den Allgemeinbildungsfragen möglichst schnell zu antworten, zurückgeführt wurde.

Die im Kontext der Manipulationsbedingung relevante Frage, ob nach Meinung der Probanden andere Zusammenhänge zwischen den beiden Aufgaben bestünden, ist ebenfalls von der Mehrheit der Teilnehmer (89.2 %) verneint worden. Unter den genannten Zusammenhängen sind in erster Linie die oben erwähnten („abnehmende Konzentrationsfähigkeit“ / „Tendenz, auch bei Aufgabe 2 möglichst schnell zu antworten“) zu nennen, neben verschiedenen eher diffusen Angaben, wie „Reaktion vs. Wissen“ oder „weiß nicht, welcher“.

Von der Möglichkeit, andere Kommentare zu dem Experiment abzugeben, hat in der letzten Frage nur eine Minderheit der Teilnehmer Gebrauch gemacht (24.1 %). Diese beziehen sich in erster Linie auf Aufgabe 1, die von den entsprechenden Probanden als sehr lang, sehr anstrengend oder langweilig eingestuft worden ist. Ein weitere Klasse von Kommentaren bezieht sich auf Kritik an der Qualität des Monitors. Dies kann damit zusammenhängen, dass die Bildwiederholungsrate von 59 Hz, gemessen am Standard aus der Zeit der Versuchsdurchführung, sehr niedrig ist und von den Probanden, die im Umgang mit besseren Geräten vertraut waren, als störend empfunden worden ist. Nicht auszuschließen ist allerdings auch, dass einigen Probanden die kurzzeitige Darbietung der Ziffern in der ersten Experimentalphase als „Flimmern“ wahrgenommen haben. Eine Auflistung sämtliche Kommentare zu den letzten beiden Fragen der Nachbefragung ist im Anhang, S. 161, enthalten.

Für die experimentelle Fragestellung entscheidend ist, dass kein Proband einen Kommentar geäußert hat, der auch bei weitest gehender Interpretation eine Vermutung beinhaltet, dass innerhalb der ersten Aufgabe im Hinblick auf die Bearbeitung der zweiten Aufgabe eine gezielte Manipulation vorgenommen worden sei, oder gar innerhalb der ersten Aufgabe andere Reize als der angebliche Zielreiz, d. h. die Maskierung, dargeboten worden sind. Die Hypothese H.2.1 kann somit als bestätigt angesehen werden.

Schätzungen Wie auch in Experiment 1 weisen die Verteilungen der Schätzwerte eine starke Asymetrie auf. Eine Wurzeltransformation, analog dem Vorgehen in Experiment 1, führt zu keiner nennenswerten Verbesse-

Tabelle 6: Mittlere Schätzungen bei der ersten Frage

	alle Daten ^a	getrennt nach Item ^b	
		Guppy	Santiago
niedriger Anker	0.559 (5,20)	0.565 (5.19)	0.552 (5.22)
hoher Anker	0.698 (6,75)	0.755 (8.15)	0.643 (5.41)

Anmerkung: Mittelwerte der logarithmierten Schätzungen getrennt nach Versuchsbedingung, in Klammern mittlere Schätzungen der Rohdaten. ^a: n = 40 bzw. 41 pro Zelle; ^b: n = 19 bis 21 pro Zelle

rung der Datenstruktur hinsichtlich der Normalität.²⁶ Aus diesem Grund sind die Daten einer Logarithmus-Transformation (Winer, 1971) unterzogen worden, die zu deutlich besseren Ergebnissen hinsichtlich der Normalität der Daten führt ($z = 0.90$; $p = .394$ bei Frage 1, bzw. $z = 1.12$; $p = .160$ bei Frage 4). Die logarithmierten Daten werden als abhängige Variablen für die parametrischen Verfahren verwendet.

Vor dem Hintergrund der hierarchischen Ordnung der Hypothesen H.2.2 und H.2.3 wird auf eine Analyse entsprechend des vollständigen Versuchsplans verzichtet. Es wird stattdessen zunächst eine separate Analyse bezüglich der Schätzungen bei der ersten Frage durchgeführt, im Anschluß daran wird die Schätzungen bei der vierten Frage analysiert. In den Analysen reduziert sich damit das experimentelle Design auf jeweils einen 2 (*Anker*: niedrig vs. hoch) \times 2 (*Item*: Guppy vs. Santiago) – Plan, bei dem beide Faktoren „between subjects“ variiert wurden, mit den logarithmierten Schätzungen als abhängiger Variable.

Bei der Analyse der Schätzungen bei Frage 1 ist eine Person von der Analyse ausgeschlossen worden, da ihre Schätzung deutlich außerhalb des Bereiches plausibler Schätzwerte liegt (Einwohner von Santiago: 600. Mio > Mittelwert + 6 Standardeinheiten), und mit hoher Wahrscheinlichkeit auf einen Eingabefehler oder ein falsches Verständnis der Fragestellung zurückzuführen ist. Ein weitere Person weist bei Frage 1 einen fehlenden Wert auf, so dass die nun folgende Analyse auf den Daten der verbliebenen 81 Personen basiert.

²⁶Kolmogorov-Smirnov-Test bei Frage 1: $z = 1.98$; $p = .001$ (Rohdaten) gegenüber $z = 1.33$; $p = .059$ (Wurzeltransformation); bei Frage 4: $z = 2.14$; $p < .001$ gegenüber $z = 1.66$; $p = .008$

Zunächst ist eine zweifaktorielle Varianzanalyse mit den beiden Faktoren *Anker* und *Item* bezüglich der logarithmierten Schätzungen durchgeführt worden. Es zeigt sich ein im Hinblick auf die einseitige Fragestellung signifikanter Effekt des Faktors *Anker* ($F(1, 77) = 3.30; p < .074; r = 0.21$). Personen, die während des subliminalen Primings mit dem niedrigen Anker konfrontiert worden sind, geben niedrigere Schätzungen ab, als Personen unter der hohen Ankerbedingung (0.559 vs. 0.698, logarithmierte Daten). Der Faktor *Item* erweist sich als nicht signifikant ($F(1, 77) = 0.65; p > .42$). Die Schätzungen der Länge eines Guppys in Zentimeter liegen somit in einem vergleichbaren numerischen Bereich, wie die Schätzungen der Einwohnerzahl von Santiago in Mio. Einwohner.

Die Interaktion zwischen den beiden Faktoren *Anker* und *Item* verfehlt ebenfalls deutlich das Signifikanzniveau ($F(1, 77) = 0.42; p > .52$). Obwohl der Ankereffekt auf deskriptiver Ebene bei dem Item „Guppy“ stärker ausfällt als bei dem Item „Santiago“ (vgl. Tabelle 6), erweisen sich diese Unterschiede als statistisch nicht signifikant.

Zur weiteren Absicherung dieses Befundes ist der Einfluss des subliminalen Primings auf die Schätzungen noch mit Hilfe eines nicht-parametrischen Verfahrens überprüft worden. Hierzu sind zunächst die Daten getrennt nach Guppy- bzw. Santiago-Schätzung z-transformiert worden, um eine Konfundierung mit dem Faktor *Item* auszuschließen. Anschließend ist ein U-Test der so erhaltenen z-Werte bezüglich des Faktors *Anker* durchgeführt worden. Es zeigt sich, dass der Einfluss des subliminalen Ankers auch auf der Basis dieses Verfahrens als signifikant ausgewiesen wird ($U = 625; p_{\text{eins.}} < .033$). Die Hypothese H.2.2 kann somit als bestätigt angesehen werden.

Bei den Schätzungen der an Position 4 präsentierten Fragen ergibt sich ein anderes Bild. Zunächst erscheint es sinnvoll im Sinne einer der voranstehenden Analyse äquivalenten Behandlung der Daten, zwei Ausreißer von der Analyse auszuschließen. Die entsprechenden Schätzungen sind zwar in Absolutwerten nicht so extrem, wie bei dem oben erwähnten Ausreißer, liegen jedoch in Relation zu der Verteilung der Schätzwerte in einem vergleichbaren Bereich (Länge Guppy: 55 cm > Mittelwert + 5.7 Standardeinheiten; Einwohner Santiago: 170 Mio. > Mittelwert + 6 Standardeinheiten).

In eine zweifaktorielle Varianzanalyse der logarithmierten Schätzungen der verbliebenen 81 Probanden ist der Hauptfaktor *Anker* nicht signifikant ($F(1, 77) = 0.33; p > .56$). Die Werte unter den beiden Ankerbedingungen liegen zwar im Mittel im postulierten Trend (0.590 unter der niedrigen vs. 0.620 unter der hohen Ankerbedingung, logarithmierte Daten), dieses Datenmuster scheint aber zufällig zu Stande gekommen zu sein, was auch die

Tabelle 7: Mittlere Schätzungen bei der vierten Frage

	alle Daten ^a	getrennt nach Item ^b	
		Guppy	Santiago
niedriger Anker	0.590 (5.37)	0.484 (4.05)	0.686 (6.57)
hoher Anker	0.620 (5.51)	0.502 (4.00)	0.745 (7.11)

Anmerkung: Mittelwerte der logarithmierten Schätzungen getrennt nach Versuchsbedingung, in Klammern mittlere Schätzungen der Rohdaten. ^a: n = 42 bzw. 40 pro Zelle; ^b: n = 19 bis 22 pro Zelle

„Trendumkehr“ beim Vergleich zwischen logarithmierten Daten und Rohdaten beim Guppy-Item (siehe Tabelle 7) nahelegt.

Überraschend ist innerhalb der Schätzungen bei der vierten Frage ein signifikanter Effekt des Faktors *Item* aufgetreten ($F(1, 77) = 10.91$; $p < .01$). Die Schätzungen der Länge eines Guppys in Zentimeter lagen deutlich unter den Schätzungen der Einwohnerzahl von Santiago in Mio. (0.493 [4.02] vs. 0.713 [6.82]; logarithmierte Daten [Rohdaten]). Die Interaktion zwischen den beiden Faktoren *Anker* und *Item* verfehlt wiederum deutlich das Signifikanzniveau ($F(1, 77) = 0.09$; $p > .76$).

Zur weiteren Überprüfung der im Kontext der Hypothese H.2.3 relevanten Fragestellung ist der Einfluss des Faktors *Anker* zusätzlich mit Hilfe eines nicht-parametrischen Verfahrens getestet worden. Analog dem Vorgehen bei Frage 1 sind die Schätzungen zunächst getrennt nach Item z-transformiert worden. Auch bei diesem Verfahren zeigt sich kein signifikanter Einfluss des Faktors *Anker* auf die Schätzungen ($U = 731$; $p_{\text{eins.}} > .20$). Die Hypothese H.2.3 muss somit zurückgewiesen werden.

Diskussion

Die Hypothesen des zweiten Experimentes sind nur zum Teil bestätigt worden. Zunächst zeigt sich ein signifikanter Effekt des subliminalen Priming auf die Schätzaufgaben bei der ersten Frage. Dies ist ein Beleg dafür, dass auch subliminal dargebotene numerische Werte Ankereffekte auslösen können und bestätigt damit die zentrale Hypothese des Experimentes. Damit ein präsentierter numerischer Wert zum Anker für eine nachfolgende Schätzaufgabe wird, ist es nicht zwingend notwendig, dass diesem Wert seitens des Urtei-

lers Aufmerksamkeit zuteil wird (vgl. Wilson et al., 1996). Selbst Werte, bei denen die Urteiler im extremen Fall über keinerlei Bewusstsein darüber verfügen, dass eine Darbietung stattgefunden hat, können also nachfolgende Schätzungen beeinflussen.

Demgegenüber steht der Befund, dass bei der vierten Frage des vorgeblichen Allgemeinbildungstests ein Ankereffekt infolge des subliminalen Priming ausgeblieben ist. Dies deutet zunächst darauf hin, dass die in diesem experimentellen Paradigma beobachteten Ankereffekte eine deutlich geringere Stabilität aufweisen, als in vielen anderen Arbeiten zum Ankereffekt. Allerdings ist es unklar, auf welche Ursachen die Fragilität des Effektes zurückgeht.

Ein denkbarer Grund wäre, dass Ankereffekte, die allein auf der Aktivierung eines numerischen Wertes basieren, generell eine geringere Stabilität aufweisen als Effekte, bei denen zusätzlich semantische Prozesse die Assimilation verstärken. Diese Position wird unter anderem vertreten von Brewer und Chapman (2002), die in einer Reihe von Experimenten einige Befunde von Wilson et al. (1996) repliziert haben. Demgegenüber stehen allerdings die auf S. 23 f. diskutierten Experimente von Becker und Stephan (1994, 1996). Die Autoren zeigen in zwei Experimenten, dass eine im Kontext einer Voraufgabe etablierte Zahl eine nachfolgende Schätzung auch dann beeinflusst, wenn die Phase der Ankeretablierung und die kritischen Schätzung zusätzlich durch die Bearbeitung einer fünfminütigen Distraktoraufgabe separiert werden. Vor dem Hintergrund dieser Befunde erscheint es wenig plausibel anzunehmen, dass die beobachtete Fragilität des Ankereffektes im vorliegenden Experiment allein auf dem zugrunde liegenden Prozess des numerischen Priming basiert.

Eine nahe liegendere Erklärung würde auf die geringe Aktivierung des numerischen Konzeptes infolge des subliminalen Priming referieren. Die Aktivierung reichte aus, um einen signifikanten Effekt bei der ersten Aufgabe zu bewirken, nicht jedoch um eine weitere Schätzung zu beeinflussen, die drei Aufgaben später abgegeben wurde. Fraglich erscheint jedoch, ob dieser Befund dahingehend interpretiert werden kann, dass Effekte im vorliegenden Paradigma nur dann auftreten, wenn die Schätzung unmittelbar auf die subliminale Ankeretablierung folgt, oder dass aus den Ergebnissen ein generelles Maß für die zeitliche Stabilität von Ankereffekten durch subliminale Anker hergeleitet werden kann.

Schließlich muss angemerkt werden, dass auch die erste Schätzung nicht unmittelbar auf die Ankeretablierung folgte. Nach der letzten Darbietung wurde zunächst die vorgebliche Reaktionsaufgabe für beendet erklärt. Im Anschluss daran mussten die Instruktionen für den sogenannten Allgemein-

bildungstest gelesen und verstanden werden. Erst nachdem dies geschehen war und eventuelle Fragen an die Versuchsleiterin beantwortet waren, wurde die erste Schätzaufgabe präsentiert. Die daraus resultierende zeitlichen Distanz der ersten Frage zur Priming-Phase war insgesamt gesehen damit nur unwesentlich kürzer als bei der vierten Frage, was es unwahrscheinlich erscheinen lässt, dass allein aufgrund der zeitlichen Distanz kein Ankereffekt bei der vierten Frage beobachtet wurde, auch wenn diese Möglichkeit nicht grundsätzlich ausgeschlossen werden kann.

Plausibler erscheint dahingehend, dass neben der zeitlichen Distanz insbesondere die Auseinandersetzung mit numerischen Werten bei der ersten Schätzung diesen Ankereffekt verhindert hat. Es ist davon auszugehen, dass bei der ersten Frage zunächst verschiedene Werte als mögliche Antwort erwogen wurden, bevor die Schätzung abgegeben wurde. Die in diesem Kontext aktivierten Werte könnten den Ankereffekt dadurch verhindert haben, dass der subliminal aktivierte Anker in Bezug auf die Schätzung bei der vierten Frage seine Exklusivität als aktiviertes numerisches Konzept verliert.

Zusammenfassend ist festzuhalten, dass in Experiment 2 der erste Nachweis dafür erbracht wurde, dass Ankereffekte auch durch subliminal präsentierte numerische Werte ausgelöst werden können. Zur Frage der Stabilität dieser Effekte unter Berücksichtigung der angesprochenen Randbedingungen ist es allerdings notwendig, weitere Experimente mit diesem Paradigma durchzuführen.

Abschließende Diskussion

Erklärungsmodelle im Rückblick

Ein Schwerpunkt der vorliegenden Arbeit besteht darin, weitere Erkenntnisse über die Bedeutung der existierenden Modelle zur Erklärung des Ankereffektes zu gewinnen.

Für den ursprüngliche Ansatz von Tversky und Kahneman (1974), wonach Ankereffekte auf unzureichender Adjustierung basieren, liegt deutliche empirische Evidenz innerhalb des Paradigmas von selbstgenerierten Ankern vor (Epley & Gilovich, 2002, 2003). Gleichzeitig zeigen Epley und Gilovich jedoch, dass der Prozess der unzureichenden Adjustierung keinen Beitrag am Zustandekommen von Ankereffekten jenseits dieses experimentellen Paradigmas zu leisten scheint.

Das Modell der konversationalen Inferenzen nach Grice (1975) referiert auf die wahrgenommene Relevanz des Ankerwertes aus Sicht der Urteiler. Seine Erklärungskraft ist damit beschränkt auf Urteilssituationen, die, ob schon von praktischer Relevanz, nur einen geringen Bereich innerhalb der experimentellen Paradigmen im Kontext der Untersuchungen zum Ankereffekt darstellen. Außerdem sprechen selbst vorliegende empirische Befunde unter Bedingungen, die für das Modell günstig sind, nicht für die dort postulierten Prozesse (vgl. hierzu S. 32 f. dieser Arbeit). Darüber hinaus stellt dieser Erklärungsansatz, wie Chapman und Johnson (2002, S. 122) anmerken, auch aus theoretischer Sicht einen eher uninteressanten Fall dar, unter anderem deshalb, weil der Ankereffekt damit letzten Endes auf eine Art „Demand“-Effekt (vgl. Schwarz, 1994; Schwarz & Bless, 1992) reduziert wird.

Damit verbleiben als Erklärung für die Vielzahl von beobachteten Ankereffekten jenseits der oben dargestellten Paradigmen lediglich zwei Modelle: Numerisches Priming und das Modell der selektiven Zugänglichkeit. Deren Zusammenspiel im Standardparadigma war Gegenstand des ersten Experimentes dieser Arbeit.

Numerisches versus semantisches Priming

Ausgangspunkt des ersten Experimentes war die von Mussweiler und Strack vertretene Position, wonach semantisches Priming der zentrale Prozess für Ankereffekte sei, und numerisches Priming allenfalls dann eine Rolle spielen könnte, wenn semantische Einflüsse unterbunden werden (vgl. etwa Mussweiler & Strack, 2001b, S. 251). Die Dominanz des SAM wird von den Autoren unter anderem auch mit der Dominanz des Standardparadigmas im Hinblick auf das Phänomen des Ankereffektes begründet. Die in diesem Paradigma beobachteten Effekte gehen nach Meinung von Musswei-

ler und Strack allein auf die im SAM postulierten Prozesse zurück. Die vorliegenden Daten aus Experiment 1 widerlegen diese Position. Es wurde gezeigt, dass auch Ankereffekte im Standardparadigma nicht allein auf den semantischen Prozessen des SAM basieren, sondern im Zusammenspiel mit numerischem Priming entstehen. In Verbindung mit anderen Befunden (Becker et al., 2000; Wilson et al., 1996; Wong & Kwong, 2000) entsteht damit folgendes Bild: Nicht semantisches, sondern numerisches Priming ist der zentrale, im Sinne von ubiquitäre, Prozess beim Zustandekommen von Ankereffekten; semantische Prozesse können diese Effekte verstärken, *wenn* die entsprechenden Voraussetzungen dafür gegeben sind.

Neben den experimentellen Befunden sprechen auch die theoretischen Konzeptionen der beiden Modelle für letztere Interpretation. Während als Auslöser von numerischem Priming der automatisierte Prozess der Aktivierung des numerischen Konzeptes aufgrund der Konfrontation mit einem entsprechenden Reiz angesehen wird (vgl. Wong & Kwong, 2000), basiert das postulierte semantische Priming auf eher deliberativen Prozessen wie dem absichtsgeleiteten Generieren von Informationen (vgl. Mussweiler, 1997). Aus dieser Sicht erscheint es ebenfalls wenig plausibel anzunehmen, dass die als automatisiert postulierten, auf numerische Konzepte bezogenen Aktivationsprozesse dadurch ausbleiben, dass die Möglichkeit besteht, semantisches Wissen zu prozessieren.

Diese Überlegung führt zu der wichtigen Frage der Interaktion von semantischem und numerischem Priming beim Ankereffekt. Die aus theoretischer Sicht nahe liegende Annahme ist zunächst, dass beim Zusammenspiel numerischer und semantischer Prozesse stärkere Ankereffekte auftreten, als wenn der Assimilationseffekt im Wesentlichen lediglich auf einem der beiden Prozesse basiert. Diese Annahme, die letztendlich auch die Voraussetzung für die Planung des Experimentes 1 zur Überprüfung der entsprechenden Fragestellung darstellt, ist durch das präsentierte Datenmuster bestätigt worden.²⁷

Eine weiter reichende Interpretation der Daten scheidet daran, dass die verwendeten Paradigmen keine strikte Prozess-Dissoziation zulassen. Es ist zwar offensichtlich, dass beim Objektvergleich ein möglicher Einfluss von numerischem Priming im Vergleich zum Standardparadigma minimiert worden ist. Allerdings wird bezüglich des Objektvergleiches nicht der Anspruch erhoben, dass damit quasi „process-pure“ semantische Primingprozesse evoziert

²⁷Dieser Schluß ist nicht zirkulär, da die Annahme der Aggregation der Effekte lediglich eine Voraussetzung *für das Design* des Experimentes war, aufgrund der eindeutigen Zuordnung der Prozesse zu den Versuchsbedingungen aber durch das Experiment selbst falsifizierbar gewesen wäre, also somit keine a-priori-Annahme im logischen Sinne darstellt.

werden. Denkbar wäre beispielsweise, dass zumindest ein Teil der Urteiler bei diesem Paradigma die Grobschätzung beantwortet, in dem zunächst der vertraute Vergleichsstandard in eine numerische Größe transferiert wird, bevor dann ein Vergleich zwischen Zielobjekt und dem entsprechenden Wert durchgeführt wird. Auf diese Weise besteht die Möglichkeit, dass auch beim Objektvergleich während der Grobschätzung durch den Vergleichsstandard numerische Konzepte aktiviert werden. Demzufolge lassen sich aus dem Vergleich der Effektstärken in dem vorliegenden Experiment keine quantitativen Aussagen über die Anteile der beiden Prozesse am Ankereffekt im Standardparadigma herleiten.²⁸ Die Frage nach denkbaren experimentellen Plänen, die weiteren Aufschluß über die Interaktion von numerischen und semantischen Prozessen beim Ankereffekt geben könnten, soll im nächsten Abschnitt ausgeführt werden.

Experimentelle Paradigmen zur Interaktion numerischer und semantischer Prozesse

Das Experiment 1 dieser Arbeit belegt zwar das Zusammenspiel von numerischen und semantischen Prozessen beim Ankereffekt im Standardparadigma, zur weiteren Untersuchung des Zusammenspiels wäre es allerdings wünschenswert, über ein experimentelles Paradigma zu verfügen, innerhalb dessen die beteiligten Prozesse in Opposition zueinander gestellt werden können. Ein experimenteller Ansatz hierzu könnte den folgenden Aufbau besitzen:

Als Zielitem verwendet man etwa eine Aktie mit abfallendem Trend und momentanen Wert von etwa 50 Euro. Die Schätzung bezieht sich auf eine Kursprognose, ebenfalls in Euro, für etwa acht Wochen nach dem Erhebungszeitpunkt. Die Probanden sollten eine entsprechende Sachkenntnis bezüglich des Börsengeschehens aufweisen. Vor den Feinschätzungen werden getrennten Versuchsgruppen zwei verschiedene Grobschätzungsaufgaben präsentiert, die beide auf eine Veränderung entsprechend einer Trendfortsetzung (fallender Kurs) referieren. Als Vergleichsstandard werden allerdings nicht mögliche Kursstände in Euro vorgegeben, sondern mögliche Kursveränderungen in dem für börsenerfahrene Urteiler vertrautem Format der prozentualen Veränderung, ausgehend von aktuellen Kurs.

Nach der Präsentation des aktuellen Kursverlaufes mit einem momenta-

²⁸Eine zusätzliche Problematik in diesem Zusammenhang besteht natürlich darin, dass die Aggregation der Effekte nicht notwendig streng additiv sein muss. Da aber keinerlei Befunde oder Modelle existieren, an die eine Diskussion dieser Frage sinnvoll anschließen könnte, sei diese Problematik hier nur der Vollständigkeit halber erwähnt.

nen Stand von ca. 50 Euro könnten die zwei verwendeten Grobschätzungsaufgaben also lauten:²⁹

Glauben Sie, dass der Kurs der ...-Aktie in den kommenden acht Wochen um mehr als **30 %** (*alternativ: 40 %*) fallen wird?

- Ja, der Kurs wird um mehr als **30 %** (*alternativ: 40 %*) fallen.
- Nein, der Kurs wird nicht um mehr als **30 %** (*alternativ: 40 %*) fallen.

Im Anschluss an die Beantwortung der Grobschätzungsaufgabe wird dann als Feinschätzung eine Prognose für den Stand der Aktie in acht Wochen, ausgehend vom Untersuchungszeitpunkt, in Euro erfragt.

Die Dissoziation der postulierten semantischen und numerischen Prozesse besteht nun darin, dass einerseits beziehend auf das SAM unter der Bedingung B die extremeren Argumente für einen Kursrückgang generiert werden sollten. Im Einklang mit den Erkenntnissen über die Sensibilität des semantischen Priming schon gegenüber leichten Variationen des Frageformats (vgl. Mussweiler & Strack, 1999b) sollte sich dies darin äußern, dass unter der Bedingung B *niedrigere* Feinschätzungen abgegeben werden, als unter Bedingung A.

Andererseits wird nach dem Modell des numerischen Priming durch die wiederholte Darbietung des Vergleichswertes das entsprechende numerische Konzept aktiviert. Bezugnehmend auf das verwendete Frageformat der Feinschätzung (Kursprognose in Euro) und das Urteilsobjekt (Aktie mit fallendem Kurs und momentanen Wert von etwa 50 Euro) sind die dargebotenen numerischen Werte so gewählt, dass basierend auf numerischem Priming unter der Bedingung B *höhere* Schätzungen abgegeben werden sollten, als unter Bedingung A.

Mit diesem experimentellen Design lässt sich zunächst nur überprüfen, welcher der beiden Priming-Prozesse sich im Hinblick auf die Urteilsbildung durchsetzt. Interessant erscheint dieses Design aufgrund seines einfachen Aufbaus dadurch, dass es die Möglichkeit offeriert, über die Aufnahme weiterer Faktoren das Zusammenspiel von numerischen und semantischen Prozessen in Abhängigkeit von externen Bedingungen in ökonomischer Weise zu untersuchen.

²⁹Die Vergleichsfrage mit der Vorgabe „30 %“ wird im Folgenden als „Bedingung A“ bezeichnet, die Vergleichsfrage mit der Vorgabe „40 %“ als „Bedingung B“.

Eine dieser externen Bedingungen, die Gegenstand verschiedener Untersuchungen zum Ankereffekt gewesen ist (vgl. Northcraft & Neale, 1987; Stephan, 2001; Wilson et al., 1996), ist beispielweise die Kompetenz der Urteiler in Bezug auf den Urteilsgegenstand. Auf der Phänomen-Ebene sprechen ein Reihe von Befunden dafür, dass der Expertisegrad der Urteiler einen geringen Einfluss auf die Intensität von Ankereffekten hat (siehe hierzu auch S. 12 ff.). Denkbar wäre allerdings ein Einfluss des Expertisegrades auf die dem Effekt zugrundeliegenden Prozesse. Beispielsweise ist es aus theoretischer Sicht nahe liegend, dass die Fähigkeit und Intensität zum Generieren hypothesenkonformer Informationen, welche die Voraussetzungen für die im SAM postulierten Primingprozesse darstellen, vom Expertisegrad des Urteilers abhängen (vgl. Mussweiler, 1997). Daran anschließend könnte die Hypothese formuliert werden, dass Ankereffekte umso stärker auf semantischen Prozessen basieren, je kompetenter die Urteiler im Bereich des Urteilsgegenstandes sind, bei geringem Expertisegrad der Ankereffekt jedoch im wesentlichen durch numerisches Priming evoziert wird.

Eine Überprüfung dieser Hypothese könnte auf der Basis des dargestellten Versuchsplans durch Hinzunahme des Expertisegrades als weiteren Faktor erfolgen. Bezüglich dieses Designs würde die Hypothese eine disordinale Interaktion zwischen dem Expertisegrad und der Grobschätzungsaufgabe vorhersagen. Bei kompetenten Urteilern sollten unter Bedingung B die Schätzungen geringer ausfallen, bei Laien hingegen unter Bedingung A.

Eine hiermit verwandte Fragestellung, die sich in Anlehnung an gängige Zwei-Prozess-Theorien zur sozialen Kognition (vgl. Chaiken, 1980; Kahneman & Frederick, 2002; Petty & Cacioppo, 1986; Sloman, 2002) anbietet, bezieht sich auf die Art der Verarbeitung beim Groburteil und den daraus resultierenden Priming-Prozessen. Auch wenn eine strikte Zuordnung von semantischem und numerischem Priming in die Kategorien *zentrale* bzw. *periphere* Verarbeitung (vgl. Petty & Cacioppo, 1986) nicht angemessen erscheint, sprechen die theoretischen Konzeptionen dafür, dass numerisches Priming eher auf einer peripheren Verarbeitung basiert (vgl. Wong & Kwong, 2000), im SAM hingegen eher zentrale bzw. elaborative Verarbeitungsprozesse zum Tragen kommen.

Die Manipulation des Verarbeitungsmodus innerhalb des dargestellten Versuchsplans könnte über einen *Divided vs. Full Attention* - Faktor (vgl. etwa Merikle & Joordens, 1997) erfolgen.³⁰ Wiederum würde die skizzierte

³⁰Dieser Ansatz basiert auf der Annahme, dass elaborative Verarbeitung in höherem Maße Aufmerksamkeit erfordert, und insofern durch die Reduktion von Aufmerksamkeit der Anteil elaborativer Prozesse an der Urteilsbildung reduziert wird.

Hypothese bezüglich eines zweifaktoriellen Designs eine disordinale Interaktion zwischen der Grobschätzungsaufgabe und der Aufmerksamkeitsmanipulation postulieren. Eine Bestätigung der Hypothese würde empirische Evidenz für die dargestellte Zuordnung der Priming-Prozesse in die entsprechenden Verarbeitungsmodi liefern und in diesem Kontext zum Erkenntnisgewinn über die Bedingungen, unter denen Ankereffekte bei Berücksichtigung der zugrunde liegenden Informationsverarbeitungsprozesse auftreten, beitragen.

Neben den beiden skizzierten Varianten wären sicherlich noch weitere Faktoren denkbar, deren Interaktion mit den beiden dem Ankereffekt zugrunde liegenden Mechanismen sich in diesem Ansatz überprüfen ließe. Außer extern zu manipulierenden Bedingungen, wie beispielsweise Zeitdruck, wären auch Persönlichkeitsmerkmale wie „Need for Cognition“ nach Petty und Cacioppo (1986) oder einzelne Dimensionen der „Big Five“ (McCrae & Costa, 1987) als Gegenstand möglicher Experimente denkbar (für eine Diskussion des Einflusses von Persönlichkeitsmerkmalen auf Urteilsverzerrungen siehe Kiell, 1995).

Die Befunde des ersten Experimentes dieser Arbeit liefern erstmalig Evidenz dafür, dass im prominentesten experimentellen Paradigma zum Ankereffekt, dem Standardparadigma, nicht ein sondern zwei Prozesse zum Auftreten des Ankereffektes beitragen. Das Zusammenspiel dieser beiden Prozesse, ob mit Hilfe der oben skizzierten oder anderer experimenteller Paradigmen, scheint ein fruchtbares Feld für weitere experimentelle Forschung, um Aufschluss über das Phänomen des Ankereffektes zu gewinnen.

Implikationen der subliminalen Verankerung

Im zweiten Experiment der vorliegenden Arbeit wurde nachgewiesen, dass auch subliminal dargebotene numerische Werte Ankereffekte bei nachfolgenden numerischen Schätzurteilen auslösen können. Dieses Ergebnis hat eine Reihe von Implikationen im Hinblick auf den Kenntnisstand zum Einfluss subliminaler Darbietungen, auf das Modell des numerischen Priming und im Hinblick auf die praktische Bedeutsamkeit für die Bildung numerischer Urteile in Alltagssituationen.

Subliminale Darbietungen

Eine zentrale Fragestellung im Bereich der Forschung mit subliminalen Stimuli ist die nach der analytischen Kapazität und Komplexität nicht-

bewusster Prozesse und als Konsequenz hieraus der Tragweite subliminaler Einflüsse auf Verhalten (Loftus & Klinger, 1992; Greenwald, 1992).

Die grundlegende Fähigkeit zur Verarbeitung subliminal dargebotener Stimuli gilt mittlerweile als etabliertes Phänomen und wurde in einer Vielzahl von Untersuchungen im Kontext von Bahnungs- und Inhibitionseffekten nachgewiesen (für einen Überblick siehe Bornstein & Pittman, 1992). Diese reichen von Einflüssen neutraler Stimuli auf der Basis von Reiz-Reaktions-Lernen (siehe Kiebele, 2001, für eine Darstellung im sportwissenschaftlichen Kontext) bis zu Effekten, die eine semantische Verarbeitung subliminal dargebotener Begriffe belegen (Marcel, 1983). Einen weiteren Schritt im Hinblick auf die Tragweite subliminaler Einflüsse liefern Befunde, bei denen durch die Aktivierung von relevanten Konzepten und Stereotypen über subliminal dargebotene Begriffe konkretes Verhalten der Probanden beeinflusst wird. Dies zeigt sich etwa bei Reaktionen auf unangenehme Situationen, wenn ein relevantes Konzept, wie beispielsweise *Feindseligkeit*, subliminal aktiviert wird (Bargh, Chen & Burrows, 1996, Exp. 3). Auch zeigen sich Einflüsse auf die Beurteilung ambivalenter Urteilsobjekte, wenn ein hinsichtlich der Ambivalenz relevantes Konzept subliminal aktiviert wird (Bargh & Pietromonaco, 1982; Devine, 1989; Erdley & D'Agostino, 1988).

Das Experiment 2 der vorliegenden Arbeit schließt zunächst unmittelbar an letzteren Ansatz an. Die Ambivalenz des Urteilsobjektes ist offensichtlich bei der Schätzung einer nicht bekannten Größe, ebenso wie die Relevanz des subliminal dargebotenen Wertes im Hinblick auf das Urteil. Allerdings gehen die Befunde von Experiment 2 in mehreren Aspekten über bereits vorliegende Ergebnisse zum Einfluss subliminaler Darbietungen auf die Urteilsbildung hinaus.

Die Erweiterungen basieren in erster Linie darauf, dass nicht Begriffe sondern numerische Werte als subliminale Stimuli verwendet wurden. Voraussetzung für einen Einfluss auf die Urteilsbildung hierbei ist zunächst, dass eine konzeptuelle Verarbeitung dieser Stimuli stattfindet. Die Befunde von Dehaene und Kollegen (Dehaene et al., 1998; Greenwald et al., 2003; Naccache & Dehaene, 2001) belegen diese Möglichkeit, waren dem Autor zum Zeitpunkt der Planung des Experimentes 2 allerdings nicht bekannt. Des weiteren geht ein Einfluss auf die Urteilsbildung bei numerischen Schätzungen hinsichtlich der Komplexität deutlich über die Bahnungs- und Inhibitionseffekte in den Arbeiten von Dehaene und Kollegen hinaus, so dass ein Einfluss auf die Urteilsbildung, auch unter Voraussetzung der konzeptuellen Verarbeitung numerischer Werte nicht evident ist.

Dies gilt insbesondere, da die numerischen Werte im Gegensatz zu den Stimuli, die in vorliegenden Experimenten zum Einfluss subliminaler Dar-

bietungen auf die Urteilsbildung verwendet wurden (für einen Überblick siehe Wyer, 1997), keine affektive Komponente besitzen. Wie Erdley und D'Agostino (1988) zeigen, basieren die Effekte subliminaler Darbietungen bei Stimuli, die eine affektive Valenz aufweisen, *im Bereich der Urteilsbildung* zwar nicht auf einem generellen affektivem Priming. Dass diese Komponente aber im Bereich subliminaler Darbietungen einen hohen Stellenwert besitzt, zeigen unter anderem Experimente von Bargh, Chaiken, Raymond und Hymes (1996), in denen Bahnungseffekte beobachtet wurden, auch wenn Prime und Zielreiz außer eine positiven Valenz keine semantischen Gemeinsamkeiten besitzen.

Weitere Evidenz für die Bedeutung der Valenz folgt aus Befunden von Bargh, Litt, Pratto und Spielman (1989, zitiert nach: Bargh, 1994). Den Probanden wurden kurzzeitig Begriffe dargeboten, die Darbietungsdauer wurde hierbei systematisch variiert. Die Aufgabe bestand darin, in einer „forced-choice“-Situation entweder ein Urteil über die Valenz (positiv – negativ) des Begriffes abzugeben, oder bezüglich zweier vorgegebener Worte zu entscheiden, welches der beiden ein Synonym für den kurzzeitig dargebotenen Begriff ist. Mit sinkender Darbietungsdauer verschlechterte sich wie erwartet die korrekte Klassifizierung. Es zeigte sich jedoch, dass bei Darbietungsdauern, bei denen die semantische Klassifizierung ersichtlich zufällig erfolgte, die evaluative Klassifizierung überzufällig häufig richtig war. Die Probanden verfügten demnach über einen Zugriff auf evaluative Information bezüglich der Reize, ohne einen Zugriff auf deren semantischen Inhalt zu besitzen. Dieser *automatic-evaluation-effect* (Bargh, 1994) demonstriert nicht nur die Bedeutung der affektiven Komponente im Kontext subliminal dargebotener Begriffe, sondern wirft auch Fragen im Hinblick auf die Etablierung der Darbietungsdauer für subliminale Primingprozeduren auf (für einen Überblick siehe Bargh, 1994; Bargh & Chartrand, 2000).

Vor diesem Hintergrund stellen die Ergebnisse aus Experiment 2 einen ergänzenden Befund im Hinblick auf die Tragweite subliminaler Darbietungen auf die Urteilsbildung dar. Sie zeigen, dass obwohl die affektive Komponente einen wesentlichen Stellenwert bezüglich der automatisierten Verarbeitung subliminaler Stimuli besitzt, das Vorhandensein einer entsprechenden Valenz keine Voraussetzung für einen Einfluss subliminaler Reize auf die Urteilsbildung darstellt. Auch affektfreie Stimuli wie numerische Werte können via subliminaler Darbietungen ein nachfolgendes Schätzurteil beeinflussen.

Aktivierung als Auslöser für Ankereffekte

Das Modell des numerischen Priming besagt, dass Ankereffekte auf der Aktivierung eines spezifischen numerischen Konzeptes basieren, die durch die Präsentation des Ankerwertes ausgelöst werden (vgl. etwa Jacowitz & Kahneman, 1995; Wilson et al., 1996; Wong & Kwong, 2000). Wie auf S. 34 ff. dargestellt, existieren eine Reihe experimenteller Befunde, die dieses Modell stützen. Der Ansatz dieser Experimente besteht darin, dass mit Hilfe mehr oder weniger aufwändiger Methoden versucht wird, mögliche Alternativerklärungen, die auf bewusste Verarbeitungsstrategien anstatt auf einen nicht-bewussten Aktivationsprozess referieren, weitest gehend auszuschließen.

Doch selbst wenn bei entsprechender Sorgfalt ein Einfluss bewusster Prozesse bei der Ankeretablierung unwahrscheinlich erscheint (vgl. Becker & Stephan, 1994, 1996; Wilson et al., 1996, Exp. 3), ist es natürlich prinzipiell nicht möglich, diese auszuschließen, so lange die Aktivierung des Konzeptes selbst an bewusste Prozesse gebunden ist. Dies gilt umso mehr, als dass das Modell des numerischen Primings in experimentellen Untersuchungen meist in Opposition zu spezifischen anderen Modellen gebracht wird (vgl. Wong & Kwong, 2000), und nicht generell in Opposition zu Einflüssen bewusster Prozesse, die die logische Negation zum postulierten nicht-bewussten Aktivationsprozess darstellen.

In diesem Kontext stellt das zweite Experiment dieser Arbeit zunächst einen wesentlichen Beleg für das Modell des numerischen Priming an sich dar. Als Analogie hierzu diene die bekannte Arbeit von Bargh und Pietromonaco (1982). Die Autoren belegen mit dieser ersten Studie zum subliminalen Priming von Persönlichkeitsmerkmalen, dass die ursprünglichen Befunde von Higgins, Rholes und Jones (1977) und Srull und Wyer (1979, 1980) nicht auf bewussten Kognitionen der Versuchsteilnehmer, sondern auf der reinen Aktivierung der entsprechenden Konzepte basieren (für eine Diskussion hierzu siehe Bargh & Chartrand, 2000).

Entsprechend wird in Experiment 2 die Methode des subliminalen Priming verwendet um nachzuweisen, dass allein die Aktivierung des numerischen Konzeptes hinreichend für das Auftreten von Ankereffekten ist. Da die Probanden über keinerlei Bewusstsein verfügen, *dass* ihnen ein numerischer Wert dargeboten wurde, können auf diese Weise auch sämtliche bewussten Prozesse im Kontext der Ankeretablierung als Quelle für die beobachteten Ankereffekte ausgeschlossen werden.

Auf diese Weise liefert das Experiment 2 einen ersten ultimativen Beleg

für die zentrale Annahme des numerischen Priming, wonach Ankereffekte allein auf der Aktivierung eines numerischen Konzeptes basieren können.

Was folgt auf Aktivierung?

Die Ergebnisse aus Experiment 2 zeigen zwar, *dass* Ankereffekte durch die Aktivierung eines numerischen Konzeptes ausgelöst werden können. Bezüglich der Frage *auf welche Weise* diese Aktivierung das nachfolgende Schätzwert beeinflusst, sind sie jedoch zunächst nur schwer zu interpretieren. Hierzu existierende Ansätze wurden bereits im theoretischen Teil dieser Arbeit vorgestellt (vgl. S. 34 ff.). An dieser Stelle soll nun ein neueres Modell aus einem noch unveröffentlichten Manuskript von Skurnik, Schwarz, Park und Galinsky (2000) im Kontext der vorliegenden Ergebnisse diskutiert werden.

Skurnik et al. erklären den Ankereffekt beim Basic Anchoring Paradigma über implizite Attributionen bezüglich der Zugänglichkeit des Ankerwertes. Nach dieser Position bewirkt die Auseinandersetzung mit einem numerischen Wert in einer vorgeblich unabhängigen Aufgabe zunächst, dass der so aktivierte Wert aufgrund seiner erhöhten Zugänglichkeit während der anschließenden Suche nach einem Schätzwert in das Bewusstsein tritt. Entscheidend für den Einfluss auf die Schätzung ist nun, ob dieser Vorgang auf die vorherige Präsentation des Wertes attribuiert wird, oder fälschlicherweise als Reaktion auf die Frage nach einer Schätzung verstanden wird.

Im letzteren Fall treten nach Skurnik et al. Assimilationseffekte auf, da der Gedanke an den Wert im Kontext der Fragestellung als Information dahingehend interpretiert wird, dass es sich um eine mögliche und sinnvolle Antwort auf die Schätzaufgabe handelt. Im ersteren Fall werden Kontrasteffekte postuliert. Diese werden allerdings nicht auf bewusste Korrekturstrategien seitens der Urteiler zurückgeführt (vgl. Strack, Schwarz, Bless, Kübler & Wänke, 1993), sondern auf die simple Vermeidung, den Wert als mögliche Antwort in Betracht zu ziehen (vgl. Skurnik et al., 2000, S. 5, S. 10 f.). In beiden Fällen werden die entsprechenden Kognitionen nicht als explizite, sondern als implizite Prozesse postuliert.

Skurnik et al. (2000) überprüfen diese Hypothese mit zwei Experimenten, die sich des experimentellen Paradigmas von Wilson et al.'s „Handschriftenstudie“ bedienen (vgl. Wilson et al., 1996, Exp. 3, für eine Darstellung siehe S. 22 f.). Eingebettet wird dieses Paradigma in einen Recognition-Test zur Wiedererkennung vorgegebener Aussagen. Anhand der Leistungen im Recognition-Test werden die Versuchsteilnehmer mittels eines Mediansplits in zwei Gruppen eingeteilt. Diese Klassifizierung dient neben der Ankerbedingung (hoch vs. niedrig) als zweiter Faktor für die Analyse der abgege-

benen Schätzungen. In beiden Experimenten ist allein die Interaktion zwischen Ankerbedingung und Wiedererkennungsleistung im Recognition-Test signifikant. Separate t-Tests zeigen, dass bei Probanden mit schlechter Wiedererkennungsleistung die üblichen Assimilationseffekte auftreten. Bei den Probanden mit guten Leistungen im Recognition-Test treten hingegen Kontrasteffekte auf. Skurnik et al. werten dies als Bestätigung für ihr Modell. Dieses Ergebnis ist in zweierlei Hinsicht von Interesse.

Zum einen bietet es einen vielversprechenden Ansatz für die Frage nach der Fragilität von Basic Anchoring Effekten (Brewer & Chapman, 2002; Pohl, 2003). Wie Pohl (2003) ausführt, sind die Umstände, unter denen Basic Anchoring Effekte auftreten oder ausbleiben, bisher noch weitgehend unklar. Bezugnehmend auf das Modell von Skurnik et al. könnte die Fähigkeit der korrekten Attribution für die Zugänglichkeit des numerischen Wertes als entscheidende intervenierenden Variable ausgemacht werden. Diese kann zunächst innerhalb der verwendeten Stichproben, unabhängig vom experimentellen Paradigma, variieren oder auch von den spezifischen Bedingungen der Ankeretablierung abhängen. Ist der Anteil der Probanden, die die Zugänglichkeit korrekt attribuieren gering, führt dies zu einem stabilen Effekt beim Basic Anchoring, mit steigender Anzahl verschwindet dieser Effekt.³¹

Zum anderen offeriert das Ergebnis von Skurnik et al. eine interessante Interpretation im Hinblick auf die Ankereffekte in Experiment 2 dieser Arbeit. Generell zeigen eine Reihe von Experimenten, dass subliminales Priming zu schwächeren Effekten als supraliminales Priming führt, was als Evidenz für eine schwächere Aktivierung der entsprechenden Konzepte im ersteren Fall interpretiert wird (für einen Überblick siehe Bargh & Chartrand, 2000). Demzufolge liegt nahe, dass auch mit Hilfe der in Experiment 2 verwendeten Primingprozedur eine eher geringe Aktivierung der numerischen Konzepte ausgelöst wird. *Wenn* diese allerdings ausreicht, den numerischen Wert im Kontext der Schätzaufgabe ins Bewusstsein treten zu lassen, existiert keine andere Möglichkeit, als dies fälschlicherweise auf die Fragestellung zu attribuieren, da die Probanden keinen Zugriff auf die wahre Quelle für die Zugänglichkeit des Wertes, nämlich die subliminalen Darbietungen, besitzen. Auf diese Weise führt die Aktivierung via subliminalem Priming nach

³¹In dem von Pohl (2003) gegeben Abriss über die variierenden Befunde beim Basic Anchoring finden sich allerdings keine Kontrasteffekte. Dies könnte entweder damit zusammen hängen, dass aufgrund der verwendeten Paradigmen der Anteil korrekt attribuierender Probanden innerhalb der Stichproben hinreichend weit von 100 % entfernt bleibt. Alternativ könnte auch die gängige Publikationspraxis dafür verantwortlich sein, dass derartige Befunde, so sie existieren, das Licht der Öffentlichkeit nicht erreichen.

dem Modell von Skurnik et al. im Gegensatz zu anderen Basic-Anchoring-Paradigmen also zwangsläufig zu Assimilationseffekten.

Es sollte allerdings angemerkt werden, dass bezüglich des Modells der impliziten Attribution und dessen empirischer Verifikation noch einige offene Fragen existieren. Insbesondere ist unklar, warum die Leistungen im Recognition-Test ein Indikator für die Art der impliziten Attribution sind.³² Die Argumentation bezieht sich nicht auf eine generelle Quellen-Zugänglichkeit, wie Skurnik et al. (2000, S. 5) betonen. Auch zeigen Nachbefragungen bei beiden Experimenten, dass alle Probanden unabhängig von ihren Ergebnissen im Recognition-Test im Stande waren, sich an den vorgegebenen Ankerwert zu erinnern.

Zusammenfassend ist also festzuhalten, dass das Modell von Skurnik et al. im Kontext des numerischen Priming einen interessanten Ansatz für die Frage bietet, *auf welche Weise* ein aktivierter numerischer Wert den Assimilationsprozess auslöst. Allerdings sind sowohl auf theoretischer als auch auf empirischer Seite weitere Arbeiten notwendig, um diese Frage abschließend zu beantworten.

Aktivierung vs. Aufmerksamkeit: Praktische Implikationen

Eine zentrale Frage im Hinblick auf das zweite Experiment dieser Arbeit bezieht sich auf die Position von Wilson et al. (1996), wonach die intervenierende Variable für das Auftreten von Ankereffekten beim Basic Anchoring der Grad an Aufmerksamkeit sei, die dem Ankerwert entgegengebracht wird. Nach Wilson et al. (1996, S. 399) ist, insbesondere wenn Ankeretablierung und Schätzung als separate Aufgaben präsentiert werden, ein entsprechend hohes Mass an Aufmerksamkeit gegenüber dem Ankerwert notwendig, damit dieser einen Einfluss auf eine nachfolgende numerische Schätzung entfaltet.

Experiment 2 widerlegt diese Position und zeigt, dass selbst in dem extremen Fall, bei dem die Urteiler nicht nur keine Aufmerksamkeit auf einen dargebotenen Wert verwenden, sondern noch nicht einmal wissen, *dass* eine Darbietung stattgefunden hat, dieser Wert dennoch Ankereffekte auslösen kann. Dieser Befund hat aus theoretischer Sicht eine Reihe von Implikationen. Neben den bereits Dargestellten sei im Kontext der Diskussion von

³²Ein Recognition-Test misst zwar nach Mandler (1980) sowohl das explizite als auch das implizite Gedächtnis in dem Sinne, dass die richtige Klassifizierung sowohl auf freiem Erinnern als auch korrekter Attribuierung für die Vertrautheit der dargebotenen Begriffe an die frühere Präsentation basieren kann. Neben der aus der Konfundierung von expliziten und impliziten Gedächtnistestung resultierenden Problematik erscheint auch eine Gleichsetzung der impliziten Gedächtnistestung beim Recognition-Test und dem, was Skurnik et al. als implizite Attribution skizzieren, nicht zwingend.

Wilson et al. (1996, S. 399 f.) insbesondere dessen Modellierung des Ankers als einem Wert im Kurzzeitgedächtnis („... the number in short-term memory ...“, ebd. S. 399) angesprochen, die angesichts der vorliegenden Ergebnisse weniger angemessen erscheint, als eine Modellierung über ein Gedächtnismodell der Aktivationsausbreitung (für eine aktuelle Darstellung siehe Reisberg, 2001).

Berücksichtigt man allerdings, dass die Aktivierung eines numerischen Konzeptes bereits in gängigen Darstellungen von numerischem Priming als Voraussetzung expliziert wird (siehe etwa Wong & Kwong, 2000), und sich die Position von Wilson et al. auch als Verweis darauf verstehen lässt, an welche Bedingungen eine entsprechende Aktivierung in der Regel geknüpft ist, so stellt sich die Frage nach den *praktischen Implikationen* des Befundes, unter Laborbedingungen eine entsprechende Aktivierung und damit einen Anker-effekt via sublimalem Priming ausgelöst zu haben. In enger Verbindung hiermit steht die generelle Frage nach der praktischen Bedeutsamkeit von Experimenten mit subliminalen Darbietungen.

Nach Bargh (1992, 1994) basiert das Interesse an subliminalen Darbietungen aus sozialpsychologischer Sicht nicht auf der Bedeutsamkeit subliminaler Einflüsse auf das Verhalten *per se*, sondern auf der Möglichkeit, Einflüsse von Stimuli zu untersuchen unter der Bedingung, dass das Individuum keinerlei Einflüsse auf das Verhalten wahrnimmt. Im Hinblick auf die Konsequenzen der Stimulusdarbietung auf das Verhalten sei es irrelevant, ob diese subliminal oder supraliminal dargeboten werden.³³ Entscheidend sei vielmehr, ob sich das Individuum bewusst ist über die Art und Weise, wie der Stimulus interpretiert und kategorisiert wird, und wie dieses Bewusstsein sein weiteres Verhalten bestimmt (Bargh, 1992, S. 237). Dadurch, dass subliminal dargebotene Stimuli nicht bewusst prozessiert werden können, bietet sich die Möglichkeit, über den Vergleich mit supraliminalen Darbietungen die Bedingungen zu identifizieren, unter denen supraliminal dargebotene Stimuli Verhalten beeinflussen, ohne dass dies von Intentionen oder der Wahrnehmung eines Einflusses begleitet wird.

Die Implikationen dieser Position wurden bereits dargestellt dahingehend, dass die Verankerung von numerischen Urteilen über subliminale Darbietungen entsprechende Aufschlüsse über die zugrunde liegenden Prozesse liefert.

Eine hierauf aufbauende Position mit konkreteren praktischen Konse-

³³Diese Position bezieht sich natürlich auf den qualitativen Aspekt des Einflusses. Bezüglich der quantitativen Aspekte ist bekannt, dass subliminale Darbietungen generell zu schwächeren Effekten führen als supraliminale Darbietungen (vgl. Bargh & Chartrand, 2000).

quenzen beziehen Merikle und Joordens (1997). Wie Merikle und Kollegen (Cheesman & Merikle, 1984, 1986; Merikle, 1992) in einer Reihe von Untersuchungen zeigen, führen Darbietungen von Stimuli bei verschiedenen Aufgaben zu qualitativ unterschiedlichen Ergebnissen in Abhängigkeit davon, ob die Stimuli sub- oder supraliminal dargeboten werden.

Merikle und Joordens (1997) replizieren einige dieser Ergebnisse und erheben zusätzliche Befunde über den Einfluss supraliminal dargebotener Stimuli in Abhängigkeit davon, ob diese mit geteilter Aufmerksamkeit oder mit voller Aufmerksamkeit verarbeitet werden. Über die verschiedenen Aufgabenstellungen hinweg zeigt sich, dass subliminal dargebotene Stimuli dieselben Einflüsse auf Urteile und Verhalten haben wie Stimuli, die zwar supraliminal dargeboten werden, allerdings ohne dass diesen Darbietungen die volle Aufmerksamkeit seitens der Probanden zuteil wird.

Merikle und Joordens werten diese Befunde als Evidenz dafür, dass Wahrnehmung mit bzw. ohne Bewusstsein und Wahrnehmung mit bzw. ohne Aufmerksamkeit äquivalente Wege zur Beschreibung derselben zugrunde liegenden Prozessdissoziation sind (Merikle & Joordens, 1997, S. 219). Neben denkbaren Konsequenzen dieser Position aus theoretischer Sicht folgt hieraus insbesondere, dass Studien mit subliminal dargebotenen Stimuli Aufschlüsse über den Einfluss der Stimuli geben, wenn diese zwar supraliminal präsentiert, aber seitens des Individuums nicht beachtet werden.

Vor diesem Hintergrund haben die Ergebnisse aus dem zweiten Experiment der vorliegenden Arbeit eine bedeutsame praktische Implikation. In den bisher vorliegenden Befunden zum Ankereffekt finden sich keine Experimente, in denen ein Ankerwert supraliminal dargeboten wird und gleichzeitig über entsprechende Manipulationen sichergestellt ist, dass die Urteiler dem Wert allenfalls ein geringes Maß an Aufmerksamkeit zuwenden. Im Einklang mit der Position von Wilson et al. (1996) zum Stellenwert der Aufmerksamkeit für die Aktivierung des numerischen Konzeptes wird in Experimenten zum Ankereffekt im Gegensatz dazu stets größte Sorgfalt darauf verwandt, dass Urteiler die Ankeretablierung mit größtmöglicher Aufmerksamkeit verfolgen.

Ungeachtet der Etablierung dieses Vorgehens im Bereich der experimentellen Forschung und der unzweifelhaften Bedeutung dieser und anderer Befunde für Erkenntnisse über die Determinanten der Urteilsbildung stellt sich jedoch die Frage nach der Alltagsnähe der Urteilsituationen, die Gegenstand dieser Experimente sind. Außerhalb des psychologischen Labors sollte es eher die Ausnahme sein, dass Urteiler einen bestimmten Wert mehrfach in einem Zahlenfeld identifizieren müssen, bevor sie eine Prognose über den DAX-Kurs abgeben (Becker & Stephan, 1994, 1996), oder über mehrere

Seiten hinweg bestimmte vorgegebene Zahlen abschreiben müssen, bevor sie eine numerische Größe schätzen (Skurnik et al., 2000; Wilson et al., 1996). Selbst bezüglich des prominentesten experimentellen Paradigmas im Bereich des Ankereffekts könnte gefragt werden, wie häufig einer numerischen Schätzung in alltäglichen Situationen wohl die Bitte um einen Vergleich mit einem erkennbar irrelevanten Vergleichsstandard vorangestellt ist (vgl. hierzu auch Epley & Gilovich, 2003).

Wenn Urteile über numerische Größen *in alltäglichen Situationen* einem Einfluss durch eine vorherige Konfrontierung mit einem numerischen Wert unterliegen sollten, stellt sich die Frage nach dem Einfluss der zahlreichen und häufig nur beiläufig wahrgenommenen Zahlen in Alltagsgeschehen, ohne dass ein Experimentator zu Seite steht mit der Bitte, diese in jeder erdenklichen Weise zu prozessieren. Der beiläufige Blick auf die Zimmernummer beim Betreten eines Raumes, das Wählen einer nachgeschlagenen Telefonnummer, die nach der Eingabe vergessen scheint, oder die Zahlen, die beim Durchblättern der morgendlichen Zeitung überflogen werden, sind Beispiele für derartige Situationen.

Die Position von Merikle und Joordens (1997) über die Parallelität zwischen Wahrnehmung ohne Bewusstsein und Wahrnehmung ohne Aufmerksamkeit legt nun nahe, dass paradoxerweise gerade das Experiment 2 dieser Arbeit, welches die wohl artifiziellste Form der Ankeretablierung innerhalb dieses Forschungsbereiches beinhaltet, am ehesten Evidenz für einen derartigen Einfluss numerischer Werte in Alltagsgeschehen liefert.

Man könnte diesbezüglich einwenden, dass verschiedene Befunde, in denen ein Ankereffekt ausgeblieben ist (vgl. etwa Wilson et al., 1996, Exp. 3, „one-page“-Bedingung; Pohl, 2003), gegen diese Möglichkeit sprechen. Allerdings beinhalten diese Experimente ebenfalls eine Ankeretablierung derart, dass der Wert, wenn auch nur kurzzeitig, im Fokus der Aufmerksamkeit steht. Insbesondere das dargestellte Modell der impliziten Attribution von Skurnik et al. (2000) legt aber nahe, dass dies nicht grundsätzlich gegen die Möglichkeit des Einflusses eines präsentierten, aber nicht beachteten Wertes spricht.

Zurückkehrend auf die Position von Wilson et al. (1996) liegt dieser zugrunde, dass der Grad an Aktivierung eines numerischen Konzeptes mit dem Grad der Aufmerksamkeit ansteigt, die einem numerischen Wert zuteil wird. Wilson et al. zeigen, dass erst bei entsprechend intensiver, bewusster Auseinandersetzung mit einem numerischen Wert die Aktivierung ausreicht, um Ankereffekte auszulösen.

Die vorliegenden Daten zum subliminalen Priming von Ankerwerten weisen aber in Verbindung mit den hierzu diskutierten Theorien auf die

Möglichkeit hin, dass auch eine geringere Aktivierung entsprechender numerischer Konzepte zu Ankereffekten führen kann, *gerade wenn diese nicht* auf einer bewussten, aufmerksamkeitsgeleiteten Verarbeitung des Ankerwertes basiert. Da schon aus der Begrenztheit der Aufmerksamkeit als psychische Ressource folgt, dass ein Großteil der Umweltwahrnehmung ohne Aufmerksamkeitszuwendung stattfindet, ist die Frage nach dem Einfluss derart verarbeiteter Informationen von hoher Relevanz für die Urteilsbildung in alltagsnahen Situationen.

Zusammenfassung

In den vergangenen dreißig Jahren seit Erscheinen des Artikels von Tversky und Kahneman (1974) wurden eine Vielzahl von experimentellen Befunden präsentiert, die Erkenntnisse über Bedingungen und zugrunde liegenden Prozesse des Ankereffektes zum Gegenstand haben. Die vorliegende Arbeit erweitert diesen Bestand zunächst um zusätzliche Erkenntnisse in Verbindung mit neuartigen experimentellen Paradigmen.

Es konnte Evidenz dafür geliefert werden, dass Ankereffekte im prominentesten und am intensivsten beforschten Paradigma, dem Standardparadigma, nicht wie angenommen allein auf dem Prozess der selektiven Zugänglichkeit (Mussweiler, 1997) basieren, sondern dass auch numerisches Priming zum Effekt in diesem Paradigma beiträgt. Dies wirft eine Reihe von Fragen bezüglich des Zusammenspiels der beiden Prozesse auf. Exemplarisch sei hier die Frage erwähnt, ob die beobachtete Stabilität des Ankereffektes im Standardparadigma unter variierenden Bedingungen ein Beleg für die Stabilität der zugrunde liegenden Prozesse ist, oder ob sie auf einem Trade-off der beteiligten Prozesse basiert. Sollte letzteres zutreffen, würde dies unter anderem eine erneute Auseinandersetzung mit den Bedingungen, die auf der Effektebene bereits als irrelevant klassifiziert wurden, erfordern.

Bezogen auf das Modell des numerischen Priming konnte mit Hilfe einer subliminalen Ankeretablierung gezeigt werden, dass Ankereffekte allein durch die Aktivierung numerischer Konzepte evoziert werden können. Dieser Befund ist zunächst der unzweifelhafteste Beleg für die Kernaussage des Modells, wonach die Assimilation auf der Aktivierung numerischer Konzepte basiert. Darüber hinaus zeigt das Ergebnis im Gegensatz zu bisherigen Positionen, dass Aufmerksamkeitszuwendung keine notwendige Bedingung für die Verzerrung numerischer Urteile durch vorher präsentierte Werte darstellt. Obschon dieser Einfluss nur unter den in Verbindung mit subliminalen Darbietungen notwendig artifiziellen Bedingungen identifiziert wurde, wei-

sen neuere Positionen, sowohl bezüglich des Ankereffektes als auch bezüglich der Verarbeitung subliminaler Darbietungen, auf die Möglichkeit hin, dass dieses Phänomen nicht auf das Versuchslabor beschränkt sein muss.

Das zusammenfassende Fazit dieser Arbeit ist also, dass sie einige Antworten offeriert, die zum Verständnis des Ankereffektes bei der Urteilsbildung beitragen. Wie üblich im Kontext wissenschaftlicher Bemühungen weisen die gegebenen Antworten jedoch den Weg auf für eine Vielzahl weiterer Fragen.

We are still confused, but on a higher level.

- Florian Illies -

Literaturverzeichnis

- Ariely, D., Loewenstein, G. & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, 118, 73–105.
- Baddeley, A. D. (1997). *Human memory: Theory and practice*. Hove, UK: Psychological Press.
- Bargh, J., Litt, J., Pratto, F. & Spielman, L. (1989). On the preconscious evaluation of social stimuli. In A. Bennett & K. McConkey (Eds.), *Cognition on individual and social contexts* (pp. 357–370). Amsterdam: Elsevier.
- Bargh, J. A. (1992). Does subliminality matter to social psychology: Awareness of the stimulus versus awareness of its influences. In R. F. Bornstein & T. S. Pittman (Eds.), *Perception without awareness: Cognitive, clinical, and social perspectives* (pp. 236–255). New York, NY: The Guilford Press.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency and control in social cognition. In R. S. Wyer & T. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum.
- Bargh, J. A., Chaiken, S., Raymond, P. & Hymes, C. (1996). The automatic evaluation effect: Unconditionally automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*, 32, 185–210.
- Bargh, J. A. & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253–285). New York, NY: Cambridge University Press.

- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- Bargh, J. A. & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trade information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, 43, 437–449.
- Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance*, 9, 396–406.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211–233.
- Becker, J. & Stephan, E. (1994). *Urteilsverzerrungen durch Priming von Ankern*. Beitrag zum 39. Kongress der Deutschen Gesellschaft für Psychologie, Hamburg.
- Becker, J. & Stephan, E. (1996). *Ankereffekte: Das Zusammenspiel zwischen expliziten und impliziten kognitiven Prozessen*. Beitrag zum 40. Kongress der Deutschen Gesellschaft für Psychologie, München.
- Becker, J., Stephan, E. & Willmann, M. (2000). *Lassen sich Ankereffekte nur durch semantisches Priming auslösen? Zur Rolle von numerischem Priming bei der Entstehung von Ankereffekten*. Beitrag zum 42. Kongress der Deutschen Gesellschaft für Psychologie, Jena.
- Becker, J., Stephan, E. & Willmann, M. (2003). *Are anchor effects only based on arguments? On the contribution of numeric priming to anchoring*. Unpublished manuscript.
- Beringer, J. (1999). *Experimental Run Time System (Version 3.28)*. Frankfurt a. M.: BeriSoft Cooperation.
- Bless, H., Strack, F. & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, 23, 149–165.
- Bornstein, R. & Pittman, T. S. (Eds.). (1992). *Perception without awareness: Cognitive, clinical, and social perspectives*. New York, NY: The Guilford Press.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. ed.). Berlin: Springer.

- Brewer, N. T. & Chapman, G. B. (2002). The fragile basic anchoring effect. *Journal of Behavioral Decision Making*, *15*, 65–77.
- Burton, S., Lichtenstein, D. R. & Herr, P. M. (1993). An examination of the effects of information consistency and distinctiveness in a reference-price advertisement context. *Journal of Applied Social Psychology*, *23*, 2074–2092.
- Busemeyer, J. R. & Goldstein, W. M. (1992). Linking together different measures of preference: A dynamic model of matching derived from decision field theory. *Organizational Behavior and Human Decision Processes*, *52*, 370–396.
- Busemeyer, J. R. & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.
- Cervone, D. & Peake, P. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgment and behavior. *Journal of Personality and Social Psychology*, *50*, 492–501.
- Chaiken, S. (1980). Heuristic vs. systematic information processing and the use of source vs. message cues in persuasion. *Journal of Personality and Social Psychology*, *39*, 752–766.
- Chapman, G. B. & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*, 223–242.
- Chapman, G. B. & Johnson, E. J. (1995). *Anchoring, confirmatory search, and the construction of values*. Working paper No. 95-017.
- Chapman, G. B. & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, *79*, 115–153.
- Chapman, G. B. & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgment of belief and value. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120–138). Cambridge, UK: Cambridge University Press.
- Cheesman, J. & Merikle, P. M. (1984). Priming with and without awareness. *Perception and Psychophysics*, *36*, 387–395.

- Cheesman, J. & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology*, *40*, 343-367.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, *82*, 407-428.
- Cotton, J. W. (1989). Interpreting data from a two period crossover design. *Psychological Bulletin*, *106*, 503-515.
- Dehaene, S. (1999). *Der Zahlensinn oder Warum wir rechnen können*. Basel, CH: Birkhäuser.
- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., Moortele, P.-F. & LeBihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597-600.
- Devine, P. G. (1989). Stereotypes an prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 680-690.
- Dixon, N. F. (1971). *Subliminal perception: The nature of a controversy*. New York, NY: McGraw-Hill.
- Englich, B. & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the court room. *Journal of Applied Social Psychology*, *31*, 1535-1551.
- Epley, N. & Gilovich, T. (2002). Putting adjustment back in the anchoring and adjustment heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 139-149). Cambridge, UK: Cambridge University Press.
- Epley, N. & Gilovich, T. (2003). Are adjustments insufficient? *Personality and Social Psychology Bulletin* (*in press*).
- Erdley, C. A. & D'Agostino, P. (1988). Cognitive and affective components of automatic priming effects. *Journal of Personality and Social Psychology*, *54*, 741-747.

- Erkel, S.-E. & Becker, J. (2004). *Anchoring beyond judgment: Lassen sich Kaufentscheidungen durch willkürliche numerische Werte beeinflussen?* Beitrag zum 44. Kongress der Deutschen Gesellschaft für Psychologie, Göttingen.
- Festinger, L. (1954). A theory of social comparison. *Human Relations*, 7, 117–140.
- Fischhoff, B. (1975). Hindsight \neq foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Fiske, S. T. & Neuberg, S. L. (1990). A continuum of impression formation, from category based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). New York, NY: Academic Press.
- Fiske, S. T. & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York, NY: McGraw-Hill.
- Foster, K. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 680–698.
- Gabrielcik, A. & Fazio, R. (1984). Priming and frequency estimation: A strict test of the availability heuristic. *Personality and Social Psychology Bulletin*, 10, 85–89.
- Galinsky, A. D. & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology*, 81, 657–669.
- George, J. F., Duffy, K. & Ahuja, M. (2000). Countering the anchoring and adjustment bias with decision support systems. *Decision Support Systems*, 29, 195–206.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky (1996). *Psychological Review*, 103, 592–596.
- Gilbert, D. T. (1990). How mental systems believe. *American Psychologist*, 46, 107–119.

- Gilbert, D. T., Miller, A. G. & Ross, L. (1998). Speeding with Ned: A personal view of the correspondence bias. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: the legacy of Edward E. Jones* (pp. 5–66). Washington, DC: American Psychological Association.
- Gilbert, D. T., Pelham, B. W. & Krull, D. S. (1988). On cognitive business: When person perceivers meet person receivers. *Journal of Personality and Social Psychology*, *54*, 733–740.
- Gilovich, T. & Griffin, D. (2002). Introduction - heuristics and biases: Then and now. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.
- Gilovich, T., Griffin, D. & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.
- Gilovich, T., Vallone, R. P. & Tversky, A. (1985). The hot hand in basketball: on the misperception of random sequences. *Cognitive Psychology*, *17*, 295–314.
- Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist*, *40*, 766–779.
- Greenwald, A. G., Abrams, R. L., Naccache, L. & Dehaene, S. (2003). Long-term semantic memory versus contextual memory in unconscious number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 235–247.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.
- Helson, H. (1964). *Adaptation level theory: An experimental and systematic approach to behavior*. New York, NY: Harper.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York, NY: The Guilford Press.

- Higgins, E. T. & Brendl, C. M. (1995). Accessibility and applicability: Some "activation rules" influencing judgment. *Journal of Experimental Social Psychology, 31*, 218–243.
- Higgins, E. T., Rholes, W. S. & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141–154.
- Jacowitz, K. E. & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*, 1161–1166.
- Johnson, E. J., Payne, J. W., Schkade, D. A. & Bettman, J. R. (1989). *Monitoring information acquisitions and decisions: The Mouselab system. Mouselab user's manual*. Fuqua School of Business, Duke University. First Edition 1985, revised 1989.
- Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Jungermann, H., Pfister, H. R. & Fischer, K. (1998). *Die Psychologie der Entscheidung: Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag.
- Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes, 51*, 296–312.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*, 697–720.
- Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge, UK: Cambridge University Press.
- Kahneman, D. & Knetsch, J. (1993). *Anchoring or shallow inference: The effect of format*. Unpublished manuscript.
- Kahneman, D. & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.
- Kahneman, D. & Riepe, M. (1998). Aspects of investor psychology. *Journal of Portfolio Management, 24*, 52–65.

- Kahneman, D., Treisman, A. & Gibbs, B. J. (1992). The reviewing of object files: Object specific integration of information. *Cognitive Psychology*, 24, 175–219.
- Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kahneman, D. & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, UK: Cambridge University Press.
- Keysar, B. & Barr, D. J. (2002). Self-anchoring in conversation: Why language users do not do what they "should". In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, UK: Cambridge University Press.
- Kiebele, A. (2001). Wie lässt sich die Verarbeitung nicht-bewusst wahrgenommener Reize methodisch belegen? *ITES E-Journal: Motor Control and Learning*, URL: <http://www.uni-saarland.de/fak5/swi-eu/deu/portal.htm>.
- Kiell, G. (1995). *Motivation und Fähigkeit als Moderatorvariablen bei Entscheidungsverzerrungen durch Urteilsheuristiken*. Unveröffentlichte Diplomarbeit, Universität zu Köln.
- Kiell, G. & Stephan, E. (1997). *Urteilsprozesse bei Finanzanlageentscheidungen von Experten. Abschlussbericht einer experimentellen Studie mit professionellen Devisenhändlern*. Forschungsbericht des Instituts für Wirtschafts- und Sozialpsychologie der Universität Köln.
- Klayman, J. & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypotheses testing. *Psychological Review*, 94, 211–228.
- Kruger, J. (1999). Lake wobegon be gone! The "below-average-effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77, 1121–1134.
- Lichtenstein, S., Fischhoff, B. & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, UK: Cambridge University Press.

- Loftus, E. F. & Klinger, M. R. (1992). Is the unconscious smart or dumb? *American Psychology, 6*, 761–765.
- Lopez, L. L. (1982). *Toward a procedural theory of judgment*. OND final report, University of Wisconsin.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252–271.
- Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*, 197–237.
- McCrae, R. & Costa, P. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 56*, 81–90.
- McNemar, Q. (1962). *Psychological Statistics* (3rd ed.). New York, NY: Wiley.
- Merikle, P. & Cheesman, J. (1986). Consciousness is a "subjective" state. *Behavioral and Brain Sciences, 9*, 42–43.
- Merikle, P. M. (1992). Perception without awareness: Critical issues. *American Psychologist, 40*, 792–795.
- Merikle, P. M. & Daneman, M. (1998). Psychological investigations of unconscious perception. *Journal of Consciousness Studies, 5*, 5–18.
- Merikle, P. M. & Joordens, S. (1997). Parallels between perception without attention and perception without awareness. *Consciousness and Cognition, 6*, 219–236.
- Miller, D. T. & Taylor, B. R. (1995). Counterfactual thought, regret, and superstition: How to avoid kicking yourself. In N. J. Rose & J. M. Olson (Eds.), *What might have been: The social psychology of counterfactual thinking* (pp. 305–331). Hillsdale, NJ: Lawrence Erlbaum.
- Mussweiler, T. (1997). *A selective accessibility model of anchoring: Linking the anchoring heuristic to hypotheses consistent testing and semantic priming*. Lengerich, Germany: Pabst.
- Mussweiler, T. (2001). The durability of anchoring effects. *European Journal of Social Psychology, 31*, 431–442.

- Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, *110*, 472–489.
- Mussweiler, T., Förster, J. & Strack, F. (1997). Der Ankereffekt in Abhängigkeit ankerkonsistenter Information: Ein Modell selektiver Zugänglichkeit. *Zeitschrift für experimentelle Psychologie*, *44*, 589–615.
- Mussweiler, T. & Strack, F. (1999a). Comparing is believing: A selective accessibility model of judgmental anchoring. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 10, pp. 135–167). Chichester, UK: Wiley.
- Mussweiler, T. & Strack, F. (1999b). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, *35*, 136–164.
- Mussweiler, T. & Strack, F. (2000a). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology*, *78*, 1038–1052.
- Mussweiler, T. & Strack, F. (2001a). Considering the impossible: Explaining the effects of implausible anchors. *Social Cognition*, *19*, 145–160.
- Mussweiler, T. & Strack, F. (2001b). The semantics of anchoring. *Organizational Behavior and Human Decision Processes*, *86*, 234–255.
- Mussweiler, T., Strack, F. & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *9*, 1142–1150.
- Naccache, L. & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*, 215–229.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–337). Hillsdale, NJ: Lawrence Erlbaum.
- Northcraft, G. B. & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*, 84–97.

- Nunes, J. C. & Boatwright, P. (2001). *Incidental prices and their effect on consumer willingness to pay*. Working paper, University of Southern California.
- Olson, D. L. & Courtney, J. F. (1992). *Decision support models and expert systems*. New York, NY: MacMillan.
- Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York, NY: Academic Press.
- Pohl, R. F. (1992). Der Rückschaufehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen. *Kognitionswissenschaft*, 3, 38–44.
- Pohl, R. F. (2000). SARA- ein kognitives Prozeßmodell zur Erklärung von Ankereffekt und Rückschaufehler. *Kognitionswissenschaft*, 9, 77–92.
- Pohl, R. F. (2003). *Fragile, but real: Basic anchoring*. Paper presented at the 19th conference on Subjective Probability Utility and Decision Making (SPUDM), Zürich, CH.
- Quattrone, G. A., Lawrence, C. P., Warren, D. L., Souza-Silva, K., Finkel, S. E. & Andrus, D. E. (1984). *Explorations in anchoring: The effect of prior range, anchor extremity, and suggestive hints*. Unpublished manuscript, Stanford University.
- Reingold, E. M. & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language*, 5, 9–28.
- Reisberg, D. (2001). *Cognition: Exploring the science of mind* (2nd ed.). New York, NY: Norton and Company.
- Ritov, I. & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, 3, 263–278.
- Ritov, I. & Baron, J. (1992). Status quo and omission bias. *Journal of Risk and Uncertainty*, 5, 49–61.
- Russo, J. E. & Shoemaker, P. J. H. (1989). *Decision traps: The ten barriers to brilliant decision-making and how to overcome them*. New York, NY: Simon & Schuster.

- Schkade, D. A. & Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, *44*, 203–231.
- Schwartz, B. & Reisberg, D. (1991). *Learning and memory*. New York, NY: Norton and Company.
- Schwarz, N. (1994). Judgment in social context: Biases, shortcomings, and the logic of conversation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 125–162). San Diego, CA: Academic Press.
- Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Hillsdale, NJ: Lawrence Erlbaum.
- Schwarz, N. (1998). Accessible content and accessible experience: The interplay of declarative and experimental information in judgment. *Personality and Social Psychology Review*, *2*, 87–99.
- Schwarz, N. (1999). Self reports: How the questions shape the answers. *American Psychologist*, *54*, 93–105.
- Schwarz, N. & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgments. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217–245). Hillsdale, NJ: Lawrence Erlbaum.
- Shevrin, H. & Dickman, S. (1980). The psychological unconscious: A necessary assumption for all psychological theory? *American Psychologist*, *35*, 421–434.
- Skurnik, I., Schwarz, N., Park, D. & Galinsky, A. (2000). *Assimilation and contrast in the anchoring effect: Implicit attributions about accessibility*. Unpublished manuscript.
- Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). Cambridge, UK: Cambridge University Press.
- Snyder, M. L. & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, *33*, 1202–1212.

- Srull, T. K. & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: *Journal of Personality and Social Psychology*, *37*, 1660–1672.
- Srull, T. K. & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and judgments. *Journal of Personality and Social Psychology*, *38*, 841–856.
- Stephan, E. (1992). *Zur Rolle von Urteilsheuristiken bei finanziellen Entscheidungen*. Beitrag zum 38. Kongress der Deutschen Gesellschaft für Psychologie, Trier.
- Stephan, E. (1993). *If we're so dumb, how come we made it to the moon? - Der Einfluss von Moderatorvariablen auf Urteilsheuristiken* [The influence of moderator variables on judgmental heuristics]. Paper presented at the 3rd Congress of the Swiss Society for Psychology, Zurich.
- Stephan, E. (1999). Die Rolle von Urteilsheuristiken bei Finanzentscheidungen: Ankereffekte und kognitive Verfügbarkeit. In L. Fischer, T. Kutsch & E. Stephan (Eds.), *Finanzpsychologie* (pp. 101–134). München: R. Oldenbourg Verlag.
- Stephan, E. (2001). *The use of judgmental heuristics in experts and laypeople: How big is the difference?* Paper presented at the 26. IAREP conference in Bath.
- Stephan, E. & Willmann, M. (2002). Barbaren, Broker, Teppichhändler: Ankereffekte bei Finanzentscheidungen. *Wirtschaftspsychologie*, *4*, 8–21.
- Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249–275). Hillsdale, NJ: Lawrence Erlbaum.
- Strack, F. & Deutsch, R. (2002). Urteilsheuristiken. In D. Frey & M. Irle (Eds.), *Theorien der Sozialpsychologie* (Vol. 3, 2nd ed., pp. 352–384). Bern, CH: Verlag Hans Huber.
- Strack, F. & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, *73*, 437–446.

- Strack, F., Schwarz, N., Bless, H., Kübler, A. & Wänke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology*, *23*, 53–62.
- Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, *2*, 339–363.
- Trope, Y. & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational factors. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239–270). New York, NY: The Guilford Press.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
- Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgments. *Psychological Review*, *90*, 293–315.
- Velden, M. (1982). *Die Signalentdeckungstheorie in der Psychologie*. Stuttgart, Germany: Kohlhammer.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.
- Wegener, D. T., Petty, R., Detweiler-Bedell, B. T. & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, *37*, 62–69.
- Wilson, T. D. & Brekke, N. C. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*, 117–142.
- Wilson, T. D., Houston, C., Etling, K. M. & Brekke, N. C. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *4*, 287–402.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York, NY: McGraw-Hill.

- Wong, K. F. E. & Kwong, J. Y. Y. (1999). *Is semantic similarity between the comparative and the numerical question necessary for the occurrence of anchoring effects?* Manuscript submitted for publication.
- Wong, K. F. E. & Kwong, J. Y. Y. (2000). Is 7300 m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior and Human Decision Processes*, 82, 314–333.
- Wyer, R. S. (Ed.). (1997). *Advances in social cognition, Vol. 10: The automaticity of everyday life*. Mahwah, NJ: Lawrence Erlbaum.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151–175.

Anhang

Vorstudie 1

Fragebogen der Vorstudie 1

Instruktion

Vielen Dank, dass Sie sich bereit erklären, an dieser Befragung teilzunehmen. Das Ziel dieser Untersuchung ist es, einen Test über das Allgemeinwissen von Studenten zu entwickeln. Wir möchten hierzu in Vortests verschiedene Frageformate testen.

In diesem Fragebogen geht es darum, für verschiedene Größen aus den Bereichen Geographie, Biologie, Politik und Alltagswissen möglichst gute Schätzungen abzugeben. Einige Fragen sind leicht, andere sind schwer zu beantworten. Bitte beantworten Sie **alle** Fragen. Wenn Sie sich nicht sicher sind, geben Sie die Antwort, die Sie für die plausibelste halten.

Um Ihnen die Beantwortung der Fragen zu erleichtern, werden sämtliche Fragen in einem einheitlichen "multiple-choice"-Format präsentiert.

In jeder Frage wird Ihnen zunächst die zu schätzende Größe und die entsprechende Einheit (z.B. „Höhe einer Litfaßsäule in m“) vorgegeben. Als mögliche Antworten werden Ihnen bei jeder Frage die Zahlen von 1 bis 9 angeboten. Sie sollen nun diejenige Zahl ankreuzen, die Ihrer Ansicht nach der zu schätzenden Größe am ehesten entspricht. Bitte entscheiden Sie sich für eine der vorgegebenen Antworten, kreuzen Sie keine „Zwischenwerte“ an. Sind Sie beispielsweise der Ansicht, dass der Wert der zu schätzenden Größe in der entsprechenden Einheit 4,7 ist, kreuzen Sie bitte die 5 an, sind Sie der Ansicht, dass der Wert 4,3 ist, kreuzen Sie bitte die 4 an.

Damit Sie für Ihre Mitarbeit eine Belohnung erhalten können, haben wir diese Befragung mit einem **Gewinnspiel** gekoppelt. Voraussetzung für die Teilnahme ist, dass alle Fragen beantwortet wurden. Unter den **fünf besten Teilnehmern** wird dann eine Person ausgelost, die einen Preis in Höhe von **50 DM** erhält.

Hierzu ist jeder Fragebogen am linken oberen Blattrand und auf dem Coupon am unteren Blattrand mit einer Code-Nummer versehen. Trennen Sie bitte den Coupon ab, und bewahren Sie ihn auf, wenn Sie an dem Gewinnspiel teilnehmen möchten.

Die Codenummer des Gewinners wird ab dem 29.08.00 im FB03-Psychologie, Holländische Str. 36-38 (INCON-Gebäude) im 2. Stock an der Zimmertür des Raumes 2114 ausgehängt. Der Gewinner kann sich dann in diesem

Raum melden und unter Vorlage des Coupons den Preis entgegennehmen. Falls Sie im Falle eines Gewinns persönlich benachrichtigt werden wollen, teilen Sie dies bitte der Versuchsleiterin mit, und notieren Sie Ihre Telefonnummer oder E-Mail-Adresse auf dem unteren Blattrand.

Haben Sie noch irgendwelche Fragen ?

Wenn nicht, möchten wir Sie bitten, mit dem Ausfüllen des Fragebogens zu beginnen.

Schätzaufgaben

Die 25 Schätzaufgaben waren auf fünf Seiten verteilt. Bei der Konstruktion der vier verschiedenen Fragebögen wurde darauf geachtet, dass auf jeder Seite zwei Längenschätzungen, zwei Einwohnerschätzungen und eine Füllfrage enthalten war und dass zwei aufeinander folgende Fragen aus verschiedenen Gebieten waren. Darüber hinaus war die Abfolge der Fragen zufällig.

Längenschätzungen

LÄNGE EINES HARIBO-GUMMIBÄRCHENS IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

LÄNGE EINER HANDELSÜBLICHEN BÜROKLAMMER IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

DURCHMESSER EINES 5-DM-STÜCKS IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

MAXIMALE LÄNGE EINES AUSGEWACHSENEN GUPPYS (ZIERFISCHART) IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

BREITE (KURZE SEITE!) EINER SCHECKKARTE IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

LÄNGE EINER HANDELSÜBLICHEN STREICHHOLZSCHACHTEL IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

DURCHMESSER EINES WASCHBECKENABFLUSSES IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

DURCHMESSER EINER COCA-COLA-BÜCHSE IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

LÄNGE EINES HANDELSÜBLICHEN EINWEGFEUERZEUGES IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

LÄNGE EINER SPIELKARTE (SKAT-SPIEL) IN CM:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

Einwohnerschätzungen

ANZAHL DER EINWOHNER VON DUBLIN IN MIO:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

ANZAHL DER EINWOHNER VON HAMBURG IN MIO:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

ANZAHL DER EINWOHNER VON BERLIN IN MIO:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

ANZAHL DER EINWOHNER VON BAGDAD IN MIO:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

ANZAHL DER EINWOHNER DER PERUANISCHEN HAUPTSTADT LIMA IN MIO:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

ANZAHL DER EINWOHNER DER CHILENISCHEN HAUPTSTADT SANTIAGO IN MIO:

<input type="checkbox"/>								
1	2	3	4	5	6	7	8	9

ANZAHL DER EINWOHNER VON TORONTO IN MIO:

1 2 3 4 5 6 7 8 9

ANZAHL DER EINWOHNER VON LONDON IN MIO:

1 2 3 4 5 6 7 8 9

ANZAHL DER EINWOHNER VON MOSKAU IN MIO:

1 2 3 4 5 6 7 8 9

ANZAHL DER EINWOHNER VON NEW YORK IN MIO:

1 2 3 4 5 6 7 8 9

Füller-Fragen

REGIERUNGSDAUER DES EHEMALIGEN US-PRÄSIDENTEN J.F. KENNEDY
IN JAHREN:

1 2 3 4 5 6 7 8 9

ANZAHL DER REICHSPRÄSIDENTEN DER WEIMARER REPUBLIK:

1 2 3 4 5 6 7 8 9

MAXIMALE LEBENSERWARTUNG EINER MAUS IN JAHREN:

1 2 3 4 5 6 7 8 9

ANZAHL DER LANDESMINISTERIEN IN DER HESSISCHEN LANDESREGIE-
RUNG:

1 2 3 4 5 6 7 8 9

REGIERUNGSDAUER DER EHEMALIGEN BUNDESKANZLERS BRANDT IN JAH-
REN :

1 2 3 4 5 6 7 8 9

Nachbefragung

Zum Abschluss der Untersuchung haben wir noch einige Fragen zu Ihren Antworten und Ihrer Person:

NENNEN SIE BITTE DIE ZWEI FRAGEN, DIE SIE AM LEICHTESTEN FANDEN:

1) _____

2) _____

NENNEN SIE BITTE DIE ZWEI FRAGEN, DIE SIE AM SCHWERSTEN FANDEN:

1) _____

2) _____

WAREN SIE BEI IRGENDWELCHEN FRAGEN IM ZWEIFEL, OB DIE ZU SCHÄTZENDE GRÖSSE WIRKLICH ZWISCHEN 1 UND 9 LIEGT ?

Nein

Ja, und zwar bei der/den Frage(n) _____

IHR ALTER: _____

IHR GESCHLECHT: männlich

weiblich

VIELEN DANK FÜR IHRE TEILNAHME AN DIESER BEFRAGUNG!

Ergebnisse von Vorstudie 1

Tabelle 8: Statistiken der Längenschätzungen in Zentimeter

Item	MW	SD	Diff. (einfach/schwierig)	Zweifel
Gummibärchen	1.8	0.8	19 (20/1)	2³⁴
Büroklammer	2.5	1.1	12 (12/0)	0
5 DM Münze	3.3	1.0	15 (15/0)	0
Guppy	4.3	2.0	-15 (2/17)	4³⁵
Scheckkarte	4.4	1.1	7 (9/2)	0
Streichholzschachtel	4.5	1.3	11 (11/0)	1
Abfluss	4.6	1.3	7 (8/1)	0
Cola-Dose	6.2	1.8	3 (4/1)	2
Feuerzeug	6.3	1.4	5(5/0)	1
Spielkarte	7.6	1.4	4(4/0)	9³⁶

Anmerkungen: MW = Mittelwert; SD = Standardabweichung; Diff.= Differenz: Anzahl der Nennungen „besonders einfach“ - „besonders schwierig“ (Angaben in Klammern); Verwendete Items in Fettdruck, N = 75

³⁴Beide Probanden gaben eine Schätzung von 1 cm ab, es ist also davon auszugehen, dass die Zweifel sich darauf beziehen, ob ein Gummibärchen die Größe von einem Zentimeter überhaupt erreicht.

³⁵Alle vier Probanden gaben eine Schätzung im Bereich von 4 - 5 cm ab. Die Zweifel scheinen sich daher eher auf eine generelle Unsicherheit bzgl. der Länge eines Guppys zu beziehen.

³⁶Acht der neun Probanden gaben im forced-choice-Format eine Schätzung von 9 cm ab. Es ist demnach davon auszugehen, dass die entsprechenden Probanden im offenen Format einen höheren Wert angegeben hätten.

Tabelle 9: Statistiken der Einwohnerschätzungen in Mio. Einwohner

Item	MW	SD	Diff. (einfach/schwierig)	Zweifel
Dublin	2.2	1.2	-4 (1/5)	0
Hamburg	2.6	1.2	10 (11/1)	0
Berlin	3.7	1.5	14 (14/0)	0
Toronto	3.8	1.7	-7 (1/8)	0
Lima	3.9	2.0	-17 (2/19)	2
Bagdad	4.5	2.0	-9 (0/9)	0
Santiago	4.8	2.0	-14 (0/14)	2
London	5.4	2.1	-2 (3/5)	4
Moskau	5.4	1.9	-2(1/3)	5
New York	6.8	2.0	0(4/4)	17

Anmerkungen: MW = Mittelwert; SD = Standardabweichung; Diff.= Differenz: Anzahl der Nennungen „besonders einfach“ - „besonders schwierig“ (Angaben in Klammern); Verwendete Items in Fettdruck, N = 75

Experiment 1

Fragebogen zu Experiment 1

Instruktion - bitte sorgfältig durchlesen!

Vielen Dank, dass Sie sich bereit erklären, an dieser Befragung teilzunehmen. Das Ziel dieser Untersuchung ist es, einen Test zum Allgemeinwissen von Studenten zu entwickeln. Wir möchten hierzu in mehreren Vortests verschiedene Frageformate testen.

Die Fragen, die wir Ihnen in dieser Untersuchung stellen, stammen aus den Bereichen Geographie, Biologie, Sport, Politik, Literatur, Unterhaltung und Alltagswissen. Die verwendeten Frageformate sind Multiple-Choice-Fragen, Vergleichsfragen und offene Fragen. Sowohl die Antwortalternativen bei den Multiple-Choice-Fragen als auch die Vergleichsobjekte bei den Vergleichsfragen sind mit einem Computer-Programm nach dem Zufallsprinzip zusammengestellt worden. Sie müssen daher nicht immer sehr plausibel sein. Bitte stören Sie sich nicht daran.

Einige Fragen sind leicht, andere schwer zu beantworten.

Bitte beantworten Sie **alle Fragen** in der **vorgegebenen Reihenfolge**. Wenn Sie sich nicht sicher sind, geben Sie die Antwort, die Sie für die plausibelste halten.

Haben Sie noch irgendwelche Fragen ?

Wenn nicht, möchten wir Sie bitten, mit dem Ausfüllen des Fragebogens zu beginnen.

a) Wie heißt der längste Fluß Afrikas ?

- Niger Nil
 Kongo Sambesi

b) Ist ein Eishockey-Puck leichter oder schwerer als eine Computer-Diskette ?

- Leichter als eine Computer-Diskette.
 Schwerer als eine Computer-Diskette.

c) Wie heißt die Hauptstadt der Schweiz ?

Antwort: _____

d) Welche dieser Sprachen gehört nicht zu den indogermanischen ?

- Englisch Japanisch
 Griechisch Russisch

e) *Version A bzw. C:*

Ist ein Guppy (Zierfischart) kürzer oder länger als ein Gummibärchen
(*alternativ:* eine Spielkarte) ?

- Kürzer als ein Gummibärchen (eine Spielkarte).
 Länger als ein Gummibärchen (eine Spielkarte).

Version B bzw. D:

Hat die chilenische Hauptstadt Santiago weniger oder mehr Einwohner
als Köln (*alternativ:* New York)?

- Weniger Einwohner als Köln (New York).
 Mehr Einwohner als Köln (New York).

f) *Version A bzw. C:*

Wie lang ist ein Guppy in cm ?

Antwort: _____ cm

Version B bzw. D:

Wie viele Einwohner hat Santiago in Mio. Einwohner ?

Antwort: _____ Mio.

g) Wie heißt die Landeshauptstadt von Mecklenburg-Vorpommern ?

- Rostock Schwerin
 Stralsund Ludwigslust

h) Wer spielte sowohl in „Saturday Night Fever“ als auch in „Pulp Fiction“ eine Hauptrolle ?

Antwort: _____

i) Wer schrieb den Roman „Der Idiot“ ?

- Vladimir Nabokov Franz Kafka
 Günter Grass Fjodor Dostojewski

j) Ist Frankreich kleiner oder größer als Spanien ?

- Kleiner als Spanien.
 Größer als Spanien.

k) Wie heißt der Bundesminister für Verkehr ?

- Reinhard Klimmt Werner Müller
 Kurt Bodewig Walter Riester

l) Begann die Kanzlerschaft von Helmut Schmidt vor oder nach der ersten bemannten Mondlandung ?

- Vor der ersten bemannten Mondlandung.
 Nach der ersten bemannten Mondlandung.

m) Der Sänger welcher Band hieß Jim Morrison ?

- The Doors Oasis
 Spandau Ballet The Who

n) Wie hieß der erste Bundespräsident der Bundesrepublik Deutschland ?

Antwort: _____

o) *Version A bzw. C:*

Hat die chilenische Hauptstadt Santiago weniger oder mehr als 1 Mio. (*alternativ: 9 Mio.*) Einwohner ?

- Weniger als 1 Mio. (9 Mio.) Einwohner.
 Mehr als 1 Mio. (9 Mio.) Einwohner.

Version B bzw. D:

Ist ein Guppy (Zierfischart) kürzer oder länger als 1 cm (*alternativ: 9 cm*) ?

- Kürzer als 1 cm (9 cm).
 Länger als 1 cm (9 cm).

p) *Version A bzw. C:*

Wie viele Einwohner hat Santiago in Mio. Einwohner ?

Antwort: _____ Mio.

Version B bzw. D:

Wie lang ist ein Guppy in cm ?

Antwort: _____ cm

q) Welcher Autohersteller verwendete schon 1913 ein Montagefließband?

- Audi Fiat
 Ford Toyota

r) Wie heißt die offizielle Währung Mexikos ?

Antwort: _____

s) Lebte Charles Dickens vor oder nach der Französischen Revolution ?

- Vor der Französischen Revolution.
 Nach der Französischen Revolution.

t) In welcher Stadt fanden die Olympischen Sommerspiele 2000 statt ?

Antwort: _____

u) Worüber berichten die Artikel im Feuilleton einer Zeitschrift?

- Innenpolitik Sport
 Kunst und Kultur Wirtschaft und Börse

v) *Version A bzw. C:*

Wie viele Einwohner hat New York (*alternativ: Köln*) in Mio. Einwohner ?

Antwort: _____ Mio.

Version B bzw. D:

Wie lang ist eine Spielkarte (*alternativ: ein Gummibärchen*) in cm ?

Antwort: _____ cm

w) Ist Außenminister Fischer jünger oder älter als Mick Jagger ?

- Jünger als Mick Jagger.
 Älter als Mick Jagger.

x) Welche Nation wurde 1950 Fußball-Weltmeister ?

- Uruguay England
 Italien Frankreich

y) *Version A bzw. C:*

Wie lang ist ein Gummibärchen (*alternativ: eine Spielkarte*) in cm ?

Antwort: _____ cm

Version B bzw. D:

Wie viele Einwohner hat Köln (*alternativ: New York*) in Mio. Einwohner ?

Antwort: _____ Mio.

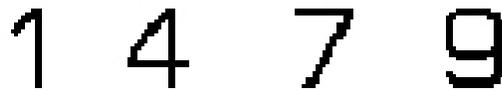
Vorstudie 2

Verwendete Reize in Vorstudie 2

Fixationskreuz und Maskierung:



Ziffern:



Ergebnisse von Vorstudie 2

Tabelle 10: Prozentanteil der "ja"-Antworten in Abhängigkeit von der darbotenen Ziffer und der Darbietungsdauer

Ziffer	Darbietungsdauer in msek.			
	17	33	50	67
1	17,5	35,0	75,0	92,5
4	35,0	67,5	65,0	95,0
7	40,0	75,0	90,0	90,0
9	25,0	40,0	45,0	77,5

Anmerkung: n=40 pro Zelle

Experiment 2

Instruktionen zu Experiment 2

Die Instruktionen zu Experiment 2 wurden, wie im Folgenden dargestellt, auf separaten Seiten auf dem Bildschirm dargeboten. Das Weiterblättern erfolgte jeweils mit der „Return“-Taste.

Sämtliche Texte wurden in schwarzer Schrift vor hellgrauem Hintergrund dargeboten.

Einleitung

Vielen Dank dafür, dass Sie sich bereit erklären, an diesem Experiment teilzunehmen.

Gegenstand dieses Experimentes ist es, Wechselwirkungen bei der Bearbeitung von Aufgaben zu untersuchen, die unterschiedliche Fähigkeiten und Fertigkeiten des Versuchsteilnehmers ansprechen.

Aus der Forschung ist bekannt, da die Leistung bei der Bearbeitung einer Aufgabe davon abhängen kann, welche Aufgaben vorher bearbeitet wurden, auch wenn zwischen den beiden Aufgaben kein inhaltlicher Zusammenhang besteht.

Um diese Zusammenhänge systematisch zu untersuchen, haben wir das folgende Experiment entwickelt.

Weiter mit <RETURN> !

Wir verwenden in diesem Experiment eine Auswahl von acht verschiedenen Aufgabentypen.

Diese Auswahl umfasst Reaktionsaufgaben, Wahrnehmungsaufgaben, aber auch Aufgaben zum logischen Denken, zum Allgemeinwissen oder Urteile über verschiedene Objekte.

Um anhand der Leistungen präzise Rückschlüsse auf mögliche Interaktionseffekte ziehen zu können, bearbeitet jeder Versuchsteilnehmer nur zwei dieser acht Aufgaben. Diese werden mittels eines Zufallsgenerators von dem Versuchsteilnehmer ausgewählt.

Weiter mit <RETURN> !

Das Experiment besteht also aus drei Phasen:

In der ersten Phase wählen Sie mittels eines Zufallsgenerators zwei Aufgaben aus, die Sie bearbeiten sollen.

Die darauffolgenden beiden Phasen bestehen jeweils aus der Erläuterung und der Durchführung der von Ihnen ausgewählten Aufgaben.

Das gesamte Experiment dauert, je nach Auswahl der Aufgaben und Bearbeitungsgeschwindigkeit, zwischen 10 und 20 Minuten.

Wenn Sie noch irgendwelche Fragen bezüglich des Experimentes haben, wenden Sie sich bitte an die Versuchsleiterin.

Um das Experiment zu beginnen, drücken Sie bitte die RETURN-Taste.

Durch das Drücken der <RETURN>-Taste nach Darbietung der letzten Seite wurde der Zufallsgenerator gestartet.

Dieser besaß folgenden Aufbau:

In der oberen Bildschirmhälfte erschien zentriert der Text „Auswahl der ersten Aufgabe“. Darunter stand zentriert „Aufgabe:“. Dahinter liefen un-systematisch und mit kurzen Darbietungsdauern die Buchstaben von „A“ bis „H“ durch das Bild. Auf Drücken einer Taste stoppte der Durchlauf bei dem Buchstaben „F“. Im zweiten Durchgang wurde auf analoge Weise die Aufgabe D „ausgewählt“.

Nach einer weiteren Aufforderung, die <RETURN>-Taste zu drücken, erschien die unten dargestellte Instruktion zur Aufgabe F.

Instruktion: Priming-Phase

AUFGABE F:

In dieser Aufgabe geht es darum, möglichst schnell die Position eines kurzzeitig dargebotenen Reizes zu erkennen.

Ihnen wird zunächst in der Mitte des Bildschirms ein Fixationskreuz dargeboten. Etwa zwei bis sieben Sekunden später erscheint links unten, rechts unten, links oben oder rechts oben kurzzeitig der unten abgebildete Reiz.



Sie sollen lediglich entscheiden, ob der Reiz links oder rechts vom Fixationskreuz dargeboten wurde. Drücken Sie bitte die mit einem „L“ markierte, linke Shift-Taste, wenn Sie meinen, da der Reiz links vom Fixationskreuz dargeboten wurde und die mit einem „R“ markierte, rechte Shift-Taste, wenn Sie meinen, da der Reiz rechts vom Fixationskreuz dargeboten wurde.

<Weiter mit RETURN !>

Es geht in dieser Aufgabe darum, sowohl MÖGLICHST SCHNELL als auch MÖGLICHST AKKURAT zu antworten. Sie haben für die Beantwortung maximal 1,5 Sekunden Zeit, danach erfolgt die nächste Darbietung.

Um den dargebotenen Reiz möglichst gut zu erkennen, sollten Sie zu Beginn eines Durchgangs das Kreuz in der Mitte des Bildschirms fixieren. Des weiteren möchten wir Sie bitten, beim Beginn der Aufgabe Ihre Zeigefinger auf die Antworttasten zu legen und den ausgemessenen Abstand zum Bildschirm einzuhalten.

Diese Aufgabe besteht aus 80 Darbietungen. Die ersten 10 Darbietungen dienen dazu, Sie mit der Apparatur und dem Versuchsablauf vertraut zu machen, Ihre Ergebnisse werden erst ab dem elften Durchgang gespeichert.

<Weiter mit RETURN !>

Haben Sie noch irgendwelche Fragen ?

Wenn nicht, möchten wir Sie bitten, durch Drücken der

<RETURN> - Taste

den Versuchsdurchgang zu starten.

Instruktion: „Allgemeinbildungstest“**AUFGABE D:**

Diese Aufgabe ist ein kurzer Test zur Allgemeinbildung. Es werden Ihnen Fragen aus verschiedenen Wissensgebieten vorgelegt, die Sie mit Hilfe des NUMMERNBLOCKES der Tastatur beantworten sollen.

Manche Fragen sind "Multiple-choice"-Aufgaben, bei denen Ihnen verschiedene Antwortmöglichkeiten vorgegeben werden. Hinter jeder Antwortmöglichkeit steht in Klammern eine Ziffer. Bitte markieren Sie die Antwort, die Ihrer Meinung nach die richtige ist, durch Drücken der entsprechenden Ziffer.

Bei anderen Aufgaben sollen Sie eine Schätzung für eine numerische Größe abgeben. Benutzen Sie zur Beantwortung dieser Fragen bitte ebenfalls den Nummernblock der Tastatur.

SÄMTLICHE ANTWORTEN MÜSSEN DURCH DRÜCKEN DER
"ENTER"-TASTE BESTÄTIGT WERDEN !

Wenn Sie hierzu noch Fragen haben, wenden Sie sich bitte an die Versuchsleiterin, andernfalls starten Sie den Durchgang durch Drücken der "Enter"-Taste.

Aufgaben im „Allgemeinbildungstest“

Die Fragen wurden in der vorliegenden Reihenfolge separat auf dem Bildschirm zentriert dargeboten. Bei den Schätzfragen erschienen die eingegebenen Ziffern und ggf. der Dezimalpunkt auf der vorgegebenen Linie. Bei den „Multiple-choice“-Fragen bewirkte das Drücken der entsprechenden Nummern-Taste, dass die dazugehörige Antwort in rot erschien. Alle Antworten mussten durch Drücken der „Enter“-Taste bestätigt werden.

- a) Wie lang ist ein Guppy (Zierfischart) in cm ? (*alternativ*: Frage d))

_____ cm

b) Wie heißt die Landeshauptstadt von Mecklenburg-Vorpommern ?

Rostock (1) Schwerin (2)
Stralsund (3) Potsdam (4)

c) Wie heißt der Bundesminister für Verkehr ?

Reinhard Klimmt (1) Werner Müller (2)
Kurt Bodewig (3) Walter Riester (4)

d) Wieviele Einwohner hat die chilenische Hauptstadt Santiago in Mio. Einwohner ? (*alternativ*: Frage a))

_____ Mio.

e) Welche Nation wurde 1950 Fußballweltmeister ?

Uruguay (1) Brasilien (2)
Italien (3) Frankreich (4)

f) Sind 10 (*alternativ*: 200) Yards (amerikanische Längeneinheit) mehr oder weniger als 10 (200) Meter ?

Mehr als 10 (200) Meter. (1)
Weniger als 10 (200) Meter. (2)

g) Wie lang ist die Sphinx von Gizeh in Meter ?

_____ Meter

h) Wer schrieb den Roman „Der Idiot“ ?

Vladimir Nabokov (1) Franz Kafka (2)
Günter Grass (3) Fjodor Dostojewski (4)

Fragebogen für die Nachbefragung

Zum Abschluß des Experimenten benötigen wir noch einige Angaben zu Ihrer Person und Ihrer Einschätzung der durchgeführten Aufgaben.

1) Ihr Alter: _____

2) Ihr Geschlecht: männlich
 weiblich

3) Ihr Studienfach? _____

4) Wie beurteilen Sie die **erste** Aufgabe hinsichtlich ihres Schwierigkeitsgrades auf einer Skala von 1 (sehr einfach) bis 7 (sehr schwierig)?

1	2	3	4	5	6	7
sehr						sehr
einfach						schwierig

5) Wie beurteilen Sie die **zweite** Aufgabe hinsichtlich ihres Schwierigkeitsgrades auf einer Skala von 1 (sehr einfach) bis 7 (sehr schwierig)?

1	2	3	4	5	6	7
sehr						sehr
einfach						schwierig

6) Glauben Sie, dass sich die Bearbeitung der ersten Aufgabe auf Ihre Leistungen in der zweiten Aufgabe ausgewirkt hat?

Nein.
 Ja, positiv.
 Ja, negativ.

7) Bestand Ihrer Einschätzung nach irgendein anderer Zusammenhang zwischen den beiden Aufgaben?

Nein.

Ja, und zwar: _____

8) Sind Ihnen bei der Durchführung des Experimentes noch andere Dinge aufgefallen, die Sie uns mitteilen möchten?

Nein.

Ja : _____

**Haben Sie nochmals vielen Dank dafür,
dass Sie an diesem Experiment
teilgenommen haben!**

Ergebnisse bei den offenen Fragen der Nachbefragung

In Klammern sind die Anzahlen der Probanden angegeben, soweit mehr als eine Person diese Antwort gab.

Frage 7:

Bestand Ihrer Einschätzung nach irgendein anderer Zusammenhang zwischen den beiden Aufgaben?

Nein (74)

Nachlassende Konzentration (3)

Bei Aufgabe 2 schnellere Antworten gegeben (2)

Bei Aufgabe 2 manchmal intuitiv geantwortet

Dichotomisiertes Drücken : Längenangaben

Reaktion vs. Wissen

Weiß nicht welcher

Frage 8:

Sind Ihnen bei der Durchführung des Experimentes noch andere Dinge aufgefallen, die Sie uns mitteilen möchten?

Nein (63)

Aufgabe 1: anstrengend / lang / langweilig (6)

Schlechter Monitor (5)

Aufgabe 2: schwierige Fragen (2)

Aufgabe 1: entspannend

Aufgabe 1: Systematik in der Symbolabfolge?

Aufgabe 1: Einmal falsch gedrückt

Aufgabe 2: gehetzt gefühlt durch Aufgabe 1

Allgemeinwissen schlecht

Reihenfolge der Aufgaben lieber umgekehrt

Menschen sind unterschiedlich begabt

Hiermit versichere ich, dass ich die vorliegende Dissertation selbständig und ohne unerlaubte Hilfe angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden.

Matthias Willmann